Anovel: Sound Reading

An explorative study on perception and design of aural and textual interplay in multimodal narrative presentations

Author: Rasmus Juhlin

Master Thesis, 15 credits, advanced level (ME620A)
Media Technology: Strategic Media Development
Malmö University
Supervisor: Fredrik Rutz
Examiner: Maria Engberg

Date of submission: 22 May 2019
Abstract

Through a three-part process employing research through design, this study has explored narratives being presented through a multimodal technical medium consisting of both textual and stream (sound) components. It has examined an existing application, Booktrack, and through developing three separate prototypes, has sought to identify and understand how one might approach a narrative when constructed using the aforementioned components. Specifically, it has explored the stream (sound) design and the perception of novels and fiction texts when presented through both text and sound. Taking on a perspective with its origin in sound studies, the study has identified five general and three specific stream design guidelines for working with sound in relation to text. Moreover, it has indicated that contextually appropriate streams’ presences alongside written text affect how a reader visualizes the narrative. Further, it has explored narrative design with both the textual and stream components in mind. Thereby, it posits a venue where the multimodality of the presentation might be used to expand the narrative presentation, using both the text and the stream as tools to further the narrative. However, it also identifies the similarities of the narrative presentation with the traditional novel and similar silent texts as being an indictment of concern. Namely, participants of the study expressed the stream’s intrusion and impact upon their immersion and visualization of written stories as both immersion enhancing and as forcibly guiding their imagination. In a society where multimodal presentations are available through phones, tablets, and other devices, it seems plausible a multimodal presentation, such as the one explored, might constitute the next step in everyday presentations. Take news articles, advertisements, and information brochures as a few tangible areas where this kind of presentation might be employed in the close future.

Keywords: Audionarratology, Booktrack, Design practice, Multimodality, Narrative, Remediation, Sound studies
Acknowledgements

First and foremost, I want to express my gratitude to all my teachers, lecturers, and otherwise involved people at the strategic media development program, you’ve all helped and made this thesis happen. I would especially like to thank my supervisor Fredrik Rutz for putting up with me opening every discussion with: “ok, so I have changed some things and this time I know what I’m doing”. I also want to give my examiner Maria Engberg credit for suggesting the name “Anovel” to frame the narrative presentations explored in the study.

I would like to thank all my superb classmates and friends for being such great supports throughout the process. Further, a big thank you to all participants and respondents of the study. Finally, a worthy mention goes to the university’s library and its staff, as I essentially lived there for the duration of the thesis.

I owe all of you my sincerest gratitude, and I wish you all the best.

Thank you!
# Table of Contents

1. **Introduction** .................................................................................................................. 5  
   1.1 Aim of study ................................................................................................................. 6  
   1.2 Research question ........................................................................................................ 7  
   1.3 Thesis overview ............................................................................................................ 7  

2. **Theoretical framework** ................................................................................................. 8  
   2.1 Audionarratology and Narrative .............................................................................. 8  
   2.2 Mediums, Remediation and Multimodality ............................................................. 10  
   2.3 The history of Booktrack .......................................................................................... 14  
   2.4 Sound studies and Streams ...................................................................................... 15  

3. **Methodology** ................................................................................................................ 18  
   3.1 Scientific approach ...................................................................................................... 18  
   3.2 Research design .......................................................................................................... 18  
   3.3 Semi structured qualitative studies ......................................................................... 21  
   3.4 Stream notation and examination ........................................................................... 22  
   3.5 Interviews .................................................................................................................. 26  
   3.6 Focus group interview ............................................................................................... 26  
   3.7 Opinion survey .......................................................................................................... 26  
   3.8 Sampling and participants ........................................................................................ 27  
   3.9 Prototype: development, materials, and tools ........................................................ 28  

4. **Part 1: Prestudy** ............................................................................................................ 30  
   4.1 Examination procedure ............................................................................................. 30  
   4.2 Expert interview procedure ..................................................................................... 31  
   4.3 Learnings from the prestudy ..................................................................................... 32  
   4.4 Summary of Part 1 learnings .................................................................................... 38  

5. **Part 2: First prototypes** ............................................................................................... 40  
   5.1 Prototype design and intention ............................................................................... 41  
   5.2 Stream design ........................................................................................................... 42  
   5.3 Focus group procedure ............................................................................................. 43  
   5.4 Survey procedure ..................................................................................................... 44  
   5.5 Learnings from the first prototypes ......................................................................... 44  
   5.6 Summary of Part 2 learnings .................................................................................... 48  

6. **Part 3: Final prototype** ............................................................................................... 49  
   6.1 Designing the narrative ............................................................................................ 49  
   6.2 Final survey procedure ............................................................................................ 53  
   6.3 Expert revisit interview procedure ......................................................................... 54  
   6.4 Learnings from the final prototype ......................................................................... 54  
   6.5 Summary of Part 3 learnings .................................................................................... 59  

7. **Discussion** ................................................................................................................... 60  
   7.1 Multimodal perception ............................................................................................. 60  
   7.2 The narrative presentation’s components’ relationship ......................................... 63  
   7.3 Designing for and defining the narrative’s presentation ....................................... 64  
   7.4 Stream design principles ......................................................................................... 66  

8. **Conclusion** ................................................................................................................... 69  
   8.1 Limitations ................................................................................................................ 71  
   8.2 Future studies ............................................................................................................ 72  

**References** .................................................................................................................... 74  

**Appendices** .................................................................................................................... 78  
Appendix A: Stream notation .............................................................................................. 78  
Appendix B: Consent form .................................................................................................. 80  
Appendix C: Expert semi structured interview questions ............................................. 82  
Appendix D: Opinion survey questions ............................................................................ 84  
Appendix E: Prototype links ............................................................................................. 88
Table of figures

Figure 1: Thesis overview structure ................................................................. 7
Figure 2: Hallet’s multimodal constitution of the fictional world ......................... 13
Figure 3: Double diamond structure and this study’s double diamond merge ............ 19
Figure 4: Research design flow ........................................................................ 19
Figure 5: Part-chapters’ structure .................................................................... 20
Figure 6: Notation form excerpt ........................................................................ 23
Figure 7: Pinocchio notation scan excerpt ........................................................ 25
Figure 8: Booktracks’ stream remediation and general purpose trends .................. 33
Figure 9: Pinocchio stream spectrogram analysis ............................................... 34
Figure 10: First prototypes’ stream structure and design ..................................... 42
Figure 11: Final prototype’s stream structure and design .................................... 50
Figure 12: Components’ relationship across the narrative’s sections (example) ....... 51
Figure 13: How stream affected immersion, visualization, and mood charts ........... 57
Figure 14: Stream mention among participants ................................................ 58
Figure 15: Stream and imagination inclusive adaptation of Hallet’s constitution of the fictional world 62
Figure 16: Multimodal syuzhet balance scale ................................................... 66
Figure 17: Notation form ................................................................................... 79
Figure 18: Final prototype cover image ............................................................. 88
Figure 19: QR code to the final prototype ......................................................... 88

Table of tables

Table 1: Examined booktracks ........................................................................... 31
Table 2: Stream design principles ................................................................... 68
1. Introduction

Throughout history, humans have had an inclination to gather and retell events, memories, and stories. Towards this end, media (plural medium) have been and will be created. In contemporary society, we find print books, e-books, movies, and audiobooks as examples of such media. Coming from a sound design background, I wanted to write a thesis connected to sound in some way. The initial idea surrounded exploring dynamic and adapting sound environments applied to, what I called, unstructured narratives. Essentially, building towards how to design and how it was perceived to have a dynamic sound environment applied to an improvised free form told story or similar. However, I came to realize this idea built upon a preposition assuming there existed, what I then simply framed as, “structured narratives”. In essence, this meant to refer to a narrative that was written, established, or simply; followed an established fabula in some way. Looking into this, and then particularly on sound being used in conjunction with written narratives (as this seemed a logical origin for my initial idea), I found essentially nothing.

In 2016 Mildorf and Kinzel defined the field of “Audionarratology” (Mildorf & Kinzel, 2016). Intending to encompass studies of sound, narrative, and sound’s purpose in presentations across essentially any medium, this field means to encapsulate studies of sound in various contexts. Encompassing everything from sound’s technical qualities to its semantic properties, the field’s creation and broadly defined span emphasize the lack of cohesive studies on the subject (Mildorf & Kinzel, 2016, pp. 9, 14).

Digital devices in contemporary society might be built to serve a general or a specific purpose. Among the more prevalent ones today are smartphones, tablets, and computers. The rapid development of such devices has enabled the consumption of digitally presented content almost wherever we go. As a consequence, devices capable of presenting both read and heard content are accessible on an unprecedented level in contemporary society.

The modality of mediums vary, i.e. the number of possible and potentially used channels for presentation differ. For example, the traditional novel is often considered “monomodal” (Hallet, 2014, p. 152) whereas a movie using both visual and aural (sound) components is multimodal in its presentation. However, as Hallet (2014) exemplifies, the monomodality of the read book is not necessarily as straightforward as may be claimed. Indeed, the inclusion of images posit a multimodal way of presenting the otherwise written narrative in the book. This is especially
noticeable when an image is used as an active tool to describe “this is what it looked like”, rather than being used to re-represent, illustrate, that which is textually described (Hallet, 2014).

Returning to the topic of sound, I wanted to explore how sound might be used alongside text in a narrative. As earlier, I found little related to sound design and the text to sound relationship, indicating the need for a study. However, I did find Booktrack, a commercially available application which included e-books with “cinematic soundtracks” (Booktrack, 2019a). Because of the rarity of studies on the specific subject of text and sound interplay, this study took shape as an attempt to understand and explore how it was perceived. Further, it wanted to exemplify how one might approach the sound design when dealing with a narrative presentation through both sound and text.

1.1 Aim of study
This study examines how multimodal narratives consisting of textual and aural (sound) components are perceived. Further it explores and defines core sound design principles when working with sound in a presentation using these two components. As such, the study subscribes to the idea that the content of a medium might be understood as a medium in itself (McLuhan, 2001, p. 8). Thereby, it operates with an understanding that differentiates between the “technical (presentation) medium” and the “narrative (presented) medium”. This allows the study to discuss of the components (text and sound) that constitute the narrative, as well as the presentation of the narrative.

Further, this study explores what purpose sound might have and bring to otherwise silently read texts. As such the study situates itself in the recently defined field of “Audionarratology” (Mildorf & Kinzel, 2016). Thereby, it aspires to expand the knowledge in a narrow field in which the purpose and presence of sound is brought to the fore. Finally, this study wants to act as an inspirational node for future studies within the aforementioned field and the, in large, unexplored area of this particular multimodal presentation of content.
1.2 Research question
Because of the mentioned aims, two intertwined research questions were defined.

RQ1: How is the presentation of a narrative affected and perceived when presented through a multimodal medium consisting of a textual and an aural stream component?

RQ2: What are suitable guidelines and sound design principles when creating the stream component of this medium?

1.3 Thesis overview
The thesis has been divided into eight chapters which each contains several sections. Because the study follows an intertwined process (see section 3.2 Research design) and is presented concurrently, a brief elaboration was considered necessary. The structure is as follows: Chapter 1 introduced the topic and research question considered during the study. Thereafter, Chapter 2 serve to establish the theoretical framework for the study, situating it and explaining all concepts used throughout. Chapter 3 justifies the choice of methods, the participant sampling, ethical consideration and describe the research design further. Following this, Chapters 4, 5, and 6 each describe the procedure and findings of Part 1, Part 2, and Part 3 of the study respectively. As the different parts were dependent on one another’s findings, this was determined the best way to show how the study’s process was interlinked throughout. In Chapter 7, the discussion of the overall findings is constructed. It elaborates upon them in relation to both the research question(s), the identified themes of the study and the theoretical background. Finally, Chapter 8 states the conclusions of the study, followed by limitations and suggestions for future steps, both in direct correlation with the subject presented here and related inquiries. Presented below in Figure 1 is an illustration of the thesis to visually establish the chapter flow and structure.

![Figure 1: Thesis overview structure.](image-url)
2. Theoretical framework

This background and prior research chapter of this study serves to establish the terminology used and explain the interdisciplinary concepts in the study. It follows a general structure to first establish the field in which the study wants to situate itself. Second, a description of what the theory of remediation is and how it will be used in the study alongside the concept of multimodality. Then, the history of the application “Booktrack” is briefly explained. Because this application has presented a tangible approach of a technical medium presenting narratives through both text and sound, this section meant to contextualize it as the narratives presented there were subjected to examination through the study. In relation to these areas and the aural focus of the research, the next section regards sound studies and introduces the concept of auditory streams which will be used in place of “sound” and similar terms throughout the study.

2.1 Audionarratology and Narrative

The subject area of this research regards the interplay of text and sound in fiction novels, with an emphasis on the “aural” (sound) component. Therefore, it situates itself in the recently defined field of “Audionarratology” (Mildorf & Kinzel, 2016). Further, the anchor of the study is established and defined as the narrative itself (Herman, 2009, p. 73), which thereafter act as a node throughout the study.

2.1.1 Audionarratology

Mildorf and Kinzel (2016) present several events of importance for both narratology and sound studies through the latter part of the 20th century. Proposing an “acoustic turn” in narratology, similar to the linguistic and visual turns (see Bachmann-Medick, 2016), they criticize and distance themselves from what others have defined as a visual bias in past studies within narratology research (see Schweighauser 2013, p. 476). Similar perspectives can be observed within studies of digitally presented narratives, such as studies on hypertext (see: Bell, 2010; Bolter, 2011), hypermedia (see Bell, 2010; Murray, 2017), or within the field of interactive digital narratives (see Koenitz, 2017). While present in many cases, the sound component is rarely discussed beyond being mentioned as existing in context. In relation to hypermedia, Bell points out that “because [hypermedia] utilize additional media such as sound and visual images they will likely require additional media-specific tools if their analysis is to be comprehensive” (Bell, 2010, pp. 188-189). This notion indicates the overlook may reside in an awareness that insufficiently equipped interdisciplinary studies may result in vague, or risk, inaccurate results.
Mildord and Kinzel define the term of “Audionarratology” as a means to provide an umbrella term for studies which “[…] explore sound and their relation to narrative structure” (Mildorf & Kinzel, 2016, p. 12). As such, the term attempts to define itself in an area ambiguously situated somewhere in between sound studies and narratology. Further, the field encompass most any medium where sound is present and includes studies of sounds on any level. From encoded (language) and embodied (music) sound (Schafer, 2005), to the semantics, technical aspects, electro-acoustic manipulations and presentations of audio. Finally, the perception and recognition of sounds, and even studies surrounding indeterminable noise are also considered subjects situated in this field (Mildorf & Kinzel, 2016, pp. 12-13). Thereby, they want to shift the emphasis from visual to aural and encourage studies where sound constitute the main subject of inquiry.

They further recognize that the field is interdisciplinary at its core, and as such, one has to consider the entangled disciplines throughout studies’ processes (Mildorf & Kinzel, 2016, p. 18). Because the field operates along several possible narrative study trajectories, they propose the main foundation for studies should take a semiotic approach, i.e. what does the sound mean and how should it be approached in context (Mildorf & Kinzel, 2016, pp. 13-15). This defines the essential premise behind audionarratology, which is that “[…] sound, voice and music carry (narrative) meaning in their own right” (Mildorf and Kinzel, 2016, p. 18).

As a researcher in this field, one also need to acknowledge that while sound may constitute the main focus of study, other components and the narrative presentation itself carry weight appropriate for the medium in question. Simply put, in certain media (radio plays, podcasts, etc.) the aural occupies a larger narrative role than in others (e.g. movies and games). Thereby, the narrative impact and subsequent narrative “allowance” of the aural component varies thereafter (Mildorf & Kinzel, 2016, pp. 13, 18; Skoulding, 2016).

2.1.2 Narrative

A narrative might be defined in several different ways and through narratological studies, several attempts have been made to define the word. In the largest sense, it might be construed according to Melberg as “[a] representation of temporal development. It is the representation of events in time” (Melberg, 2009, p. 248). This idea is supported by Abbot who defines narrative to be the “[…] principal way in which our species organizes its understanding of time” (Abbot, 2005, p. 3), building towards explaining events that will, have or are occurring (Abbot,
This approach correlates with Tecklenburg who in turn argues most theories surrounding narrative focus on an end product (Tecklenburg, 2014, p. 90). Meaning, as a subject of a study, “narrative” represents neither a “narration taking place within a narrative”, nor the “narration of the narrative”, but rather the description of events that is narrated. Further, Melberg posits a narrative should be considered “[…] an instrument for distributing and elaborating the perspectives that can be adopted on a given set of events” (Melberg, 2009, p. 247). However, a narrative also requires the inclusion of “experientiality” (Caracciolo, 2014; Fludernik, 1996). This quality means to distinguish a “story” narrative from the orderly description of, say, a recipe or otherwise schematic instruction (Fludernik, 1996; Herman, 2009).

Derived from Russian formalism, the terms fabula and syuzhet (also sjuzhet, sujet) become relevant for discussing the narratives encountered through this study (Herman, Jahn, Ryan, 2010, pp. 157, 535). The first, fabula, refers to the underlying core, what, of the narrative. It is described as “a series of logically and chronologically related events [emphasis added] that are caused or experienced by actors” (Bal, 1999, p. 5). Events in this sense refer to the transitions from one state in a narrative to the next. Thereby, they construe an overall linear perspective on the narrative’s progression (Bal, 1999, p. 5; Grabes, 2014). In comparison, syuzhet describe how the fabula is composed and presented when mediated. These three concepts (fabula, events, and syuzhet) are raised as they will be used to discuss both the examined and the designed narratives in the study. Further, the word narrative is considered synonymous with both “story”, fictional or otherwise, and the concept of the “narrative medium” (as will be explained in the next section) throughout the study.

2.2 Mediums, Remediation and Multimodality

This section aims to establish the main theory used throughout the study alongside the dual-purpose terminology of the term “medium”, and the concept of multimodality and how it is relevant for the study.

2.2.1 Mediums

Medium as a concept has been defined in various ways, depending on both context and purpose (Wolf, 2011, p. 159). Throughout this study, the word has two applications (the technical medium and the narrative) and therefore an elaboration on the subject was considered necessary. The common definition of the word regards the platform of presentation of any content. Essentially, it is simple defined as “[…] a way of communicating information” (Oxford
Dictionary, 2019). Elleström defines the technical medium as “any object, physical phenomenon or body that mediates, in the sense that it ‘realizes’ and ‘displays’ [content]” (Elleström, 2010, p. 30). This definition only regards that which enables the realization and presentation of content, be it the material of a sculpture or the screen of a phone. Building on this, the term will be applied in a broader sense in this study, meaning to represent both physical objects, materials, and devices (hardware) and when relevant, their digital counterparts (software) as well.

McLuhan posits that anything (content) being presented and communicated might be considered as a medium in itself, being that which carries the inherent meaning of said content (McLuhan, 2001, p. 8). Thereby, the word medium is applicable to the narrative itself and allows discussion of it and the defined components (fabula, events, and syuezhet) independent of the technical medium. This distinction is raised to differentiate between the presenting and the presented, as the study deals with both in an interleaved manner. Hereafter, the use of the word medium therefore refer to either the technical medium or the narrative (medium).

2.2.2 The theory of Remediation

Content often shift from one technical medium to another. This transition is what Bolter and Grusin describe as “remediation” (Bolter & Grusin, 1999, pp. 53-62). It is a term similar to repurposing but focused on content that exchanges mediums. In the vaguest sense, anything being mediated might be understood as experiencing remediation, being inspired by, or adoptions of existing presentations (Bolter & Grusin, 1999, pp. 15, 44-45). Tangible examples of remediation include; a book becoming a movie, a story being visually represented in a painting, and so on. In contemporary society, digital media has enabled remediation on an unprecedented scale. Simply put, the ways to distribute, create and thereby mediate content, stories, and events has never been as accessible as now.

The purpose behind mediation, and concurrently remediation, of content resides in creating authentic (re)presentations for an observer or end-user (Bolter & Grusin, 1999, p. 53). As stated, the term can be seen as an adaption of the word “repurposing”, being focused on the transferal of content between different mediums. Bolter and Grusin explain the term is derived from the latin word remederi which means “to heal, to restore to health” (Bolter & Grusin, 1999, p. 59). As such, the concept of remediation is closely linked to revitalization and thereby to presenting novel and authentic versions of existing content. In turn, authenticity refers to the ability to
evoke immediacy for an observer, striving towards providing an experiential quality akin to experiencing whatever is represented firsthand. This notion of remediation, immediacy, and authenticity can be observed in both historic artworks striving towards realism and the photorealism of today’s computer graphics. Specifically, the goal in these instances are to achieve what Bolter and Grusin define as *transparent immediacy* (Bolter & Grusin, 1999, pp. 21-31). In essence, transparency refer to the technical medium itself being hidden as part of the presentation. An example would be a realism artist painting in such a way that they remove their brushstrokes, or a camera removing the human factor from the depiction altogether (Bolter & Grusin, 1999, pp. 25, 28).

However, transparent immediacy is not the only approach for presentation of content. Often used in conjunction with the aforementioned term, is *hypermediacy* (Bolter & Grusin, 1999, p. 31-44). As opposed to transparent immediacy, hypermediacy depends on multiple parts constituting a presentation and acts towards making each part visible (Bolter & Grusin, 1999, pp. 33-34). However, both terms strive towards achieving and eliciting immediacy for the observer of any mediated content. In the context of this study, the different terms were considered alongside one another. Throughout, the study revolves around the concept of remediation and how the two components in the multimodal presentation regard each other.

### 2.2.3 Multimodality

Books, novels and other versions of written content is created and designed using a multitude of components that together constitute the presentation the written text. Among these are typography, layout, color, and most important for this study, images. Despite this, the traditional novel is generally considered “monomodal” (Hallet, 2014, p. 152). Drawing on a few examples, Hallet describes what he calls a “multisemiotic narrative”, in which images are used alongside the written (textual) narrative. Rather than being superfluous to the text, he highlights cases where the images complement the written, therethrough providing insights that the text might be unable to describe on its own (Hallet, 2014, p. 156).

Essentially, these narratives are built through several components (text and images) and via conscious usage of the components as narrative tools, the narratives might be observed to become multimodal in their structure and presentation (Hallet, 2014, p. 159). Multimodality is, according to Gibbons, inherent to the multisensory experience of living (Gibbons, 2014). The implication being that a multimodal medium utilizes several channels for presenting content.
However, Hallet problematizes the notion of a multimodal medium as becoming more complicated as well, meaning that experiencing a multimodal medium becomes a “multiliterate act” (Hallet, 2014, p. 168). In essence, the ability to understand and make sense of what is presented becomes dependent on the observer/user/reader’s ability to comprehend the individual components (see Figure 2). Further, this means understanding is dependent on the experiences, cultural appropriations, as well as personal perceptive abilities to make sense of the presentation accordingly (Hallet, 2014; Ryan, 2015).

As a concept, multimodality may refer to more than multi component presentations. Elleström builds a definition of the term by breaking down and defining the essential components of any one medium. More complex but similar to Hallet’s definition, these are defined as the four modalities that construe any medium and serve as a model for comparing, understanding, and discussing intermedial relations (Elleström, 2010, pp. 11-13, 35-36). On the other hand, Kress defines multimodality as a “field in which semiotic work takes place, a domain for enquiry, a description of the space and of the resources that enter into meaning in some way or another” (Kress, 2011, p. 38). Unlike both Hallet and Elleström, Kress’ take on the term deals not only with technical medium’s composition and relationships. Instead, alongside social semiotics, it encapsulates several disciplines to provide a why to the technical medium’s what.

Because of the focus and scope of this study, the concept and term multimodality here draw mainly on Hallet’s definition. Thereby, it may be framed using Gibbons description of multimodal literature as something “[…] that feature a multitude of semiotic modes in the communication and progression of [narratives]” (Gibbons, 2015, p. 420). Hereafter, comparisons are made to “silent texts” on several occasions. This means to differentiate
between the multimodal narratives explored in the study and narrative content mediated through text alone. In this respect, silent text means to frame and focus on conventionally typeset text and includes printed, digitized, and wholly digital texts. A general descriptor and denominator for these silent texts is their purpose of presenting information in an understandable and non-obtrusive way. For comparative reasons, they might be considered as monomodal and as striving to achieve immediacy (Bolter & Grusin, 1999, p. 21) in the most straightforward way possible for readers.

Finally, as the narratives examined in this study include both read and heard content it was decided to designate the observer of the medium as a “reader”. In part, this term was chosen due to the definition found in *the living handbook of narratology* being “a reader is a decoder, decipherer, [and] interpreter [...] of any text in the broad sense of signifying matter” (Prince, 2013). Alongside this, previous usage of the term, describing both reading and listening (see Have & Pedersen, 2015, p. 28; and, Rubery, 2011, p. 12), supported the rationale of using it to encompass both read and heard consumption of content through this study.

2.3 The history of Booktrack

In 2011, localized in New Zealand, the Booktrack application was released. Sporting a novel approach towards silent reading, the service enabled reading with soundtracks accompanying the reader’s pace. In the context of multimodality, these “booktracks” (name derived from ‘book’ and ‘soundtrack’) consisted of two components, one textual and one aural. Whenever a user would read one of these soundtracked e-books, the underlying system would play sound effects, ambiances, and cinematic music alongside the written text.

Under the hood, the booktracks rely on a “sing-along” system, in which an either visible of invisible tracker attempts to match the reader’s pace. Working on a page-by-page basis, it iteratively changes the speed of this tracker to, after a while, correspond to the individual reader’s average pace.

Inherently, the novelty of the multimodal presentation, and the aural intrusion into the sphere of silent texts, was met with both interest and critique. Despite this, or perhaps because of this, the e-book booktracks seemingly have been neglected and dismissed over the 8 years since the applications release (latest soundtracked e-book was released by booktrack in 2016). Instead, purely aural but similarly “soundtracked” audiobooks have become increasingly available
through the same Booktrack application. However, it is the original e-book booktracks that are of interest for this study. Because they posed a tangible and commercially available example of the multimodal technical medium, an examination of how the narrative had been approached in them was conducted. In addition, the existing e-book booktracks (hereafter only booktracks) provided a starting point for understanding how to regard the stream (sound) design in relation to the textual component of the medium (see section 3.4 Stream notation and examination).

2.4 Sound studies and Streams

Because this study encompasses sound elements on several layers, a clear definition was deemed necessary. In general sound studies, the simple word “sound” is ambiguous. It may refer to music, noise, voice, or any inherent quality or component found in an aural event (Mildorf & Kinzel, 2016, p. 11; Shafer, 1977). The terms sound and audio are therefore considered interchangeable and refer to either a general (heard) aural event, or to whatever aurally is being discussed currently. Further, sound effects, ambiences, and music constitute the three types of sounds relevant throughout the study.

The overarching definition for the aural component of the narrative presentation in the study will be referred to as the “stream” exclusively. This term is derived from the term “auditory streams” (Bregman, 2002, p. 220). Similar to the term “soundscape” (an “aural landscape”), coined by Schafer (1977), auditory stream (hereafter only referred to as streams) strive to define and encompass the whole of an aural environment. However, soundscapes in its original application refer to real-world environments. Therefore stream was considered more generally applicable, including non-environmental parts as well.

However, Schafer (1977) defines several concepts and ways of identifying and categorizing sounds in soundscapes. These provide a framework and multi layered set of terms against which the concept of streams might be discussed as well. Below are several terms borrowed from Schafer’s soundscape analysis described briefly as a point of reference for the reader of this study. Alongside these, two sources connected to movie sound and sound design are also described (see Murch, 2005; and, Chion & Gorbman, 1994).
2.4.1 Sound study terms

*High- and low-fidelity Soundscapes* (Schafer, 1977, pp. 43-44)

Essentially, High and Low fidelity soundscapes roughly correlate to Silent and Noisy aural environments. A high fidelity soundscape is one where individual sounds are more clearly audible (a rural area), where a low fidelity one instead have a higher level of noise present in the environment (an urban area). Further, areas generally vary between high and low fidelity over time (e.g. at night the general environment is more silent, allowing more detailed sounds to be audible more clearly, making it more high fidelity).

*Keynote sounds, Signals, and Soundmarks* (Schafer, 1977, pp. 9-19)

When identifying sounds in soundscapes, each particular aural event (i.e. singular sound) can be assigned a level of generality. (1) Is the sound a general geographic even occurring in a particular environment (i.e. waves at the sea, or birds in a forest), then it is a *Keynote sound*. (2) A *Signal* indicates a particular and specific sound event that incites the listener’s attention, such as an alarm or similar event. Finally, (3) *Soundmarks* are sounds unique to a particular environment (e.g. a church bell). Further, specific acoustic qualities may also be classified as soundmarks (e.g. a church’s echoey interior.).

*Referential aspect classification* (Schafer, 1977, pp. 139-144)

Schafer posits several ways of structuring and classifying sounds. Among these are the “classification according to referential aspects” of a given sound event (i.e. what does the listener identify the sound as). In its simplicity, this system is built around six main categories of sounds, each of which contain between zero to twelve subcategories. For example, among the first tier of categories you might find *Natural sounds* which in turn host ‘sounds of water’ and ‘sounds of air’, among others. Depending on the sound, there also exist a third layer within these, but this level was not regarded during this study.

*Encoded and Embodied sounds* (Murch, 2005)

Unlike the previous terms mentioned, these are derived from film sound. Used to classify the attentive requirements of a listener, Encoded and Embodied sounds are defined as opposites on a spectrum in which any sound might be placed. Encoded sounds require more attention and effort from the listener to make sense (language being the distinct example). Embodied sound instead requires little to no attentive or conscious listening to make sense (music being
exemplified). Simply put, encoded sound requires context and more attention to make sense than does embodied.

Diegetic and non-diegetic sounds (Chion & Gorbman, 1994, pp. 73, 79-82)
Diegesis refer to the level of presentation of an object in relation to a narrative. A diegetic sound takes place within the narrative and is thereby able to be perceived by characters in a story, for example. On the contrary, a non-diegetic sound does not take place in the storyworld or narrative itself, rather it is something that is placed outside the diegesis of the narrative. As a tangible example; orchestral music during a movie is most likely non-diegetic as it is only audible for the observer of the movie. However, should the music originate from an orchestra currently present in the story it would instead be diegetic in nature.
3. Methodology

This chapter describes the epistemological approach to the research, the research design and concurrently, the different parts of the study. Finally, the methods of inquiry employed have been described and motivated alongside participant sampling and relevant ethical considerations through the study.

3.1 Scientific approach

The study’s evaluation relies on subject’s meaning making and interpretations, which situate it as employing a relativist approach, leaning towards symbolic interactionism (Grey, 2014). The overall structure borders on the concept of research through design (Zimmerman, Forlizzi, & Evenson, 2007; Zimmerman, Stolterman, & Forlizzi, 2010) and draws upon both it and the “Double diamond” (2018) for structuring the design process (see next section).

Frayling (1993) defines three perspectives of research in the arts: Research into, for, and through design; two of which are relevant for this study. Namely, (1) research for design and (2) research through design. What characterizes the first (for design) is the intention of advancing practice, often taking shape in frameworks, recommendations, and methods of design (Zimmerman, Stolterman, & Forlizzi, 2010, p. 313), which corresponds to one of the research questions of this study. Further, Zimmerman and Forlizzi state that the second (through design) differs from its two siblings in that “[…] it is an approach to doing research. It can result in knowledge for design and into design” (Zimmerman & Forlizzi, 2014, p. 169). Finally, as artefacts in research through design serve the double purpose of product and means of understanding (Zimmerman, Stolterman, & Forlizzi, 2010), prototypes were used during two stages of the study (see Chapter 5 and Chapter 6).

3.2 Research design

Because the study intended to use research through design to achieve research for design, i.e. a framework for design practice when approaching this particular medium; a process involving prototypes was implied. However, due to the existing application of Booktrack, in which an approach towards the medium was presented, a thorough examination of existing examples was conducted prior to any development. Therefore, the overall research design was adapted from
A standard “Double diamond” (2018) structure. The research design is presented below and aims to illustrate the structure of the study.

A regular double diamond process would not be interleaved as the one shown at the bottom in Figure 3. However, in a normal double diamond research process, the first diamond would culminate into an identified problem and design decision, which then the second would build solutions for. Instead, this illustration means to show how the study was conducted in an interleaved manner. In essence, it shows how the design decision at the mid-way double diamond spread into both adjacent areas. Further, this allowed the description of the study to be divided into the following three parts: (1) Prestudy, (2) First prototypes, and (3) Final prototype. All parts served to increase the understanding of the medium, though the second and third might equally be attributed to exploring the medium as well. In Figure 4, a flow chart of the entire process has been provided. This chart means to further elucidate how the study research design took shape, what methods were used where, and how each part connected from beginning to end. As opposed to an iterative prototype development process, the prototypes in this study served to explore
particular subjects. Hereafter, the purpose behind each part of the study is described in more
detail, alongside the reasoning as to how they tie together in relation to the research question
and the aim of the study. However, this intends to only briefly explain the structure and methods
used through each part. To elaborate, each method used is explained and motivated later in this chapter, while the procedure regarding the application of each method is described in detail
during each Part-chapter (Chapter 4, 5, and 6).

Finally, Figure 5 illustrates how each upcoming “Part” Chapter is structured, consisting of four areas each. Hereafter, each Part chapter’s contents are described briefly to outline the study’s composition.

3.2.1 Prestudy (Part 1)

After the initial literature review was completed, the first step was to perform a critical review
and examination of the Booktrack soundtracked e-books, the “booktracks”. To do this, a way
to identify and notate the stream content was devised, building upon the theory of remediation
and several concepts derived from sound studies. Further, a temporal analysis was conducted
on a few of the selected booktracks. This was done to visualize the stream and examine how
and if there existed any apparent patterns in the overall stream design.

Further, the prestudy enlisted experts to try out the existing Booktrack application. Striving
towards an understanding of the medium and the specifics of the stream design, participants at
this stage were recruited due to their experience and proficiency in sound design. The same
experts were enlisted again during Part 3.

3.2.2 First prototypes (Part 2)

Because a prototype was to be developed towards the end of the study, the second part built
towards that end. To enable further understanding of the multimodal presentation of the
narrative and to begin to answer the research question, two smaller prototypes were developed.
Considering several aspects brought forth from the booktracks examination and the expert
interviews, these explored the core stream design on an existing fiction text.

The main focus for the prototypes was to explore the stream design in relation to the written
content, with the narrative being the main topic during the following focus group interview.
Four practitioners in sound design were recruited for the focus group interview. Further, a survey focusing on the narrative and its presentation was distributed to the participants. The focus group participants were assigned one of the prototype versions to read, striving towards an even distribution of the versions. Prior to the focus group session, the survey was submitted individually to enable individual thoughts before discussion. In addition, the survey was distributed to six external participants (not present for the focus group). These were also randomly allotted one version of the prototypes, keeping the version distribution even. The external participants were, unlike the aforementioned experts and focus group participants, not proficient or affiliated with sound design, striving to reduce possible aural bias among the enlisted participants.

3.2.3 Final prototype (Part 3)

The final part of the study gathered all pre-study and first prototype findings and converted them into a final proof-of-concept prototype. This was an attempt to both concretize a possible design practice approach, and further explore how narrative design with intent, considering both the read and the heard at the story’s ideating stage, was perceived by readers of the narrative. It was thereafter publicized and distributed to participants on a convenience basis. A focused survey served as the main tool for gathering data at this stage, focusing on the participant’s perception of the narrative.

Alongside the survey, the experts from the pre-study was enlisted for interviews regarding the prototype design, focusing on the stream component. In addition, they were also asked to participate in the survey.

3.3 Semi structured qualitative studies

Because a large part of the study’s empirical data was collected through interviews and written surveys, a framework for how to regard the data was considered vital. Semi structured qualitative studies (SSQS) involve iterative and systematic coding of verbal data and due to its semi structured nature, it might also be supplemented by other kinds of data (Blandford, 2013). Because the study intended to use interviews and focus group sessions as a way to understand and answer the research questions, surveys would be employed to supplement these. Similar to grounded theory, the first part of the study identified four themes that intended to guide the research (Blandford, 2013; Charmaz, 2008, Grey, 2014). These served as nodes for examination of varying importance through concurrent parts of the study.
Further, as there is no correct way to gather data in semi structured inquiries, the data gathering should be approached and adapted with the research and study subject in mind (Blandford, 2013). Indeed, the most appropriate method should be used, but less efficient methods does not inherently create fundamental issues or fallibility within the study (Woolrych, 2011; Willig, 2013). Due to these loose reigns regarding the methods used, a transparent description of the research process is of outmost importance to ensure the validity of such studies (Blandford, 2013). Therefore, this study strives towards complete transparency, meaning to enable scrutiny of the study itself and to show a detailed recollection meant to inspire future studies on the topic.

In SSQS, the underlying purpose should always be kept close in mind and guide the collection of data through a study (Rogers, et al., 2011). Further, participant inclinations, attitude, pay-off, etc. should be considered as such aspects might create an unawares bias in the study results (Blandford, 2013). Finally, the number of participants in SSQS studies are often left undefined due to the inherent iterative and ongoing process underlying such studies, meaning participants should be enlisted as required (Blandford, 2013; Charmaz, 2006, p. 113). This was considered an important trait for this study, as the obscurity of the technical medium alongside the research’s focus indicated several stages of participant involvement.

Several tools might be used for recording data in SSQS that help make the research more systematic (Grey, 2014; Blandford, 2013; Corbin & Strauss, 2008; Rogers et al., 2011, p. 227). However, tools should act as supports for the data gathering, and not define the process themselves (Corbin & Strauss, 2008, p. xi). Ensuring “good” qualitative research have been defined by Yardley (2000, p. 219) as considering four main characteristics: (1) Sensitivity to context, (2) commitment and rigor, (3) transparency and coherence, and finally (4) impact and importance. Throughout this study, these characteristics were regarded as a baseline for achieving a thorough and valid study.

3.4 Stream notation and examination
Due to the lack of existing methods dealing with interleaved (read) text and (heard) sound, a way to code the stream contents in relation to the text and the overall narrative was devised during the study. The stream notation and examination were conducted in several steps. Main considerations were the occurrence of individual sounds, their type (sound effect, ambience, music), diegesis, and the semantic- and general purpose of their application. Drawing on
multiple ways to identify and classify sound in an aural environment, three main sources were employed for the stream notation system and structure. These sources were: Schafer (1977), Chion (Chion & Gorbman, 1999), and Murch (2005). As stated previously in the thesis the word “stream” refers to the whole of the aural (sound) component of the narrative presentation. Alongside these sound study sources, Bolter and Grusin’s (1999) concept of remediation functioned as an addition, focus and limitation for the notation scope. Namely, the theory of remediation enabled framing and thereby the examination of how the stream and text components related to one another within, and in relation to, the overall narrative.

A key quality for the notation system was that it, like stenography (shorthand) writing, needed to be conducted quickly and without effort. Thereby, the narrative design (text and stream content) might be experienced as intended by the narrative designer. In Figure 6 below, and excerpt of the used form is presented (see Appendix A for the complete notation form).

**Figure 6: Notation form excerpt**

On the top of the page, the name of the booktrack would be noted, alongside the release year. If there were several A4’s of notation, the current page number was noted at the top as well. All examined booktracks were read on a computer web browser with headphones (relevant as the “page change” of the notation as the booktracks was dependent on the device used).

Because the goal was to note as many things with as little effort as possible, the system operated on several layers. First, each cluster of (three) lines represent a level of remediation, top to
bottom: literal, interpreted, and appended. These terms derived from how the stream sound was connected to the written text at the point in which it occurred. The top (literal) line represented a level of remediation in which the stream sound was also described in the text, thereby being literally applied based on the written. The middle (interpreted) meant a stream sound was implied through the context of the narrative, but was never explicitly explained or described in the text. Also, these kinds of sounds might border on the last line category, the Appended. This final line represented a level in which the stream contents was not in any way found in the written, meaning something appended in relation to it. Often this regarded non-diegetic and thematic sounds in the stream. Further, each cluster of lines belong to a sound type (Sound effect, Ambience, Music). What was written during the notation was the numbers 1 (world/event) or 2 (emotion/mood), each indicating the establishing purpose of the sound in relation to the written and overall narrative.

The notation was conducted in a left to right manner, meaning the occurrence of sounds created a visible pattern in relation to each other. Further, the cross ‘x’ symbol indicated the end of an ongoing sound present in the stream, often prior to something new starting thereafter. The long line ‘|’ represented noted changes in the narrative or significant stream alterations identified by the researcher, indicating different “sections” of the narrative (i.e. the environment in the narrative changes for some reason). In relation to narratology, these sections and changes roughly correlate to the “events” found in the fabula as presented through the syuezhet of the narrative. The ‘^’ symbol referred to when pages were changed during reading, this meant sections might be revisited and roughly re-experienced afterwards if something was of concern (do note these should only be accurate on the same platform as the researcher used, i.e. a computer’s web browser). The ‘S’ was put into place when either the ambience or music was silent for some time, do note this was in relation to one another as the streams very rarely were completely silent. Finally, the large circle indicated several sounds (ambiences mostly) being interleaved (i.e. playing at the same time). Below in Figure 7, an excerpt from the opening of the Pinocchio booktrack (Booktrack, 2019c) is shown.
Figure 7: Pinocchio notation scan excerpt

The sounds in the stream were identified using Schafer’s referential aspects of classification (Schafer, 1977, pp. 139-144). This meant sounds were noted on a fairly general basis, yet was differentiated enough to be separated. It should be noted that the sounds classification was not noted during the notation. Further, because the difference between a sound effect and ambience might be extremely vague, the decision was made to note ambiguous sounds as an ambience if recurring sound effects were identified as playing on a loop. Further, barring non-looping recurring sound effects (such as several “footsteps” in a row), all sounds were noted each time they were identified, even though the same sound might have played earlier in the stream during an earlier section of the narrative or similar.

A sound was classified as music if it included either clearly audible melodic or tonal components, or if an instrument was identified by the researcher. In large, this study regarded ambiguous (ambience/music) sound events as follows: If there was a distinct melody, instrument, or ongoing beat, it was designated as Music. If the component was tonal but low key, blending into the ambience and environment, it was designated as Ambience.

Do note that other researchers performing the same analysis might interpret the stream differently, both in what they note and how they interpret the purpose of the stream sounds.
3.5 Interviews
Semi-structured interviews are performed using a predefined structure regarding a subject or particular topic. In general, interviews are qualitative in nature and are regarded as an important and valid method of inquiry in research (Grey, 2014). Because of the obscurity of the medium explored through this study, interviews were considered the best way to gather data and experiences from related fields and similar mediums of presentation. Through approaching the interviews in a semi-structured manner, the structure itself encouraged reflection and discussion on related subjects as well, leading towards an understanding consisting of multiple pieces from similar topics. This was considered important as discussions during more open-ended interviews allow for deeper insights that enable the expansion into topics not raised by the questions themselves (Grey, 2014; Walliman, 2017, p. 114).

Experts in the field of sound design were recruited as the initial source for insights regarding the stream design in the narrative presentation. The same experts were enlisted twice over the course of the study. First as an external source of knowledge that meant to act as both inspiration and critique regarding the study, the technical medium, and the narrative presentation in general. The second time was at the end of the study, where they were part of the evaluation of the final prototype and concurrently, the study itself.

3.6 Focus group interview
After a test session of two booktrack inspired prototypes (see Chapter 5), a focus group interview was conducted. An advantage of focus group interviews is that they allow for both open and focused discussions while also providing an environment where different perspectives might thrive (Grey, 2014, p. 250). This session was based on the interview questions posed during the first expert interview. However, the emphasis during the focus group interview was on the narrative and encouraging the participants to discuss how they experienced both it and the technical medium. The researcher acted as a mediator through the discussion, rekindling it with questions when it waned.

3.7 Opinion survey
As a way of collecting individual thoughts on the various prototypes throughout the study, surveys were developed both as a complement and as the main tool for gathering data (Grey, 2014). During the study, surveys were employed twice. The first application was in conjunction with the focus group interview and acted as an addition to it. The second time it served as the
main evaluation for the final prototype during Part 3 of the study. In both cases it meant to gather participants’ individual thoughts on the narrative presentation and the technical medium presenting it.

The surveys were made anonymously and were distributed through various channels by the researcher on a convenience basis. To ensure the validity of the survey results, filtering questions were applied to identify outliers that could be ignored. Some of these questions regarded everyday media consumption, while others ensured the submitter actually had experienced the prototype prior to answering the survey. The two surveys were focused on the particular prototype preceding them, thereby providing data relevant to the current stage of the study (Grey, 2014, p. 375).

3.8 Sampling and participants
Due to the nature of the study, the importance of participant’s previous knowledge and experience varied over the different parts of the study. However, the sampling was made on a convenience basis throughout and participants were recruited accordingly. In every case, participants were presented with the subject of the study, the effort required from them, and the purpose of their participation (be it insights, testing or otherwise). Participants were informed that any and all participation were voluntary and anonymous.

Where applicable, the recruiting was focused on gathering participants whose experiences and previous knowledge were relevant to the inquiry at hand. Thereby, although by convenience, the recruitment leaned towards a purposeful sampling (Patton, 2002, p. 273; Creswell, 2015, p. 203).

Further, as the interview and focus group sampling was made to consciously strengthen the results in particular areas, they might also be attributed as judgmental sampling (Marshall, 1996, p. 523). In these instances, this was done to achieve relevance and reliability upon which wider conclusion might be drawn (Blandford, 2013; Marshall, 1996). Finally, the goal of these specific samples was not to ensure the representativeness of a population. Rather, it was to ensure the relevance of the participant’s backgrounds and experiences in accordance to the study’s aim (Charmaz, 2008, p. 83).
3.5.1 Ethical considerations

All participants interviewed were given a consent form to sign alongside the information that they could withdraw from the study at any point without needing to provide a reason for doing so. Further, for every recorded session, the first questions posed regarded whether the participant(s) allowed the data gathered to be used for the study and if the participation was voluntary. See Appendix B for the consent form used. Further, any questions regarding the consent from was clarified and participants were made aware of their rights according to the GDPR law (GDPR, 2019).

Further, to ensure the privacy of all participants, all gathered data was processed by the researcher alone.

During Part 2 (first prototypes) of the study, an existing novel was used as the textual component of the narrative presentation. The chosen novel was “The eye of the world” by Robert Jordan (1990). This prototype was never publicized outside a controlled setting to avoid any possible copyright infringements. For the final prototype, the entirety (text and stream) was created by the researcher to enable publicizing the prototype.

3.9 Prototype: development, materials, and tools

This section serves to briefly explain the tools used and development of the prototypes throughout the study, as it is not discussed otherwise. As previously stated, this study meant to explore a medium consisting of a presentation of both text and sound. Therefore, the main tools for creating the prototypes was Microsoft Word for word processing, and the DAW (Digital Audio Workstation) Pro tools for stream development and sound design. The parts were combined using the web-based application Booktrack Studio (Booktrack, 2019b) and published through the same. Because the Booktrack Studio provided the tools (albeit limited) to create and present narratives consisting of both text and sound, it was used as the development platform for the prototypes. This allowed the study to focus on the perception and the sound design of the narrative presentation without developing a technical medium platform from scratch suited for the same purpose.

For the first prototypes, two excerpts from Robert Jordan’s book “The eye of the world” (1990, pp. ix-x, 1-5) were merged and used as the textual component.
All sounds used in the streams was either recorded during the study or created from a sound library that was the property of the researcher prior to the study.

During the revision of this thesis, the Booktrack Studio was disabled. Because of this, links provided may be faulty. A description on how to reach the prototypes developed through the study can be found in Appendix E.
4. Part 1: Prestudy

A traditional approach to research through design would include creating and iterating a prototype based on an initial idea, set requirements, and possibly limitations. Because the Booktrack application presented a commercially available approach to the medium under scrutiny in the study, it was subjected to examination. This would serve to create an initial understanding of the medium and the baseline for designing the upcoming narrative presentations. Thereby, it meant to guide the study’s upcoming prototype developments.

The examination served to familiarize the researcher with the medium of the study and simultaneously provide inspiration for what the upcoming prototypes might look like. Being the most tangible area for application of the theory of remediation, this first part built towards an empirical foundation upon which the upcoming two parts could rest.

In addition to the examination, two experts were recruited and asked to read a few of the booktracks prior to an interview session focused on the stream design and general approach towards the medium itself.

4.1 Examination procedure

Eight booktracks of varying length were examined in the study. Because of the temporal nature of the stream and the booktracks depending on the reader keeping a steady pace for synchronizing sounds, a notation system was devised that would allow for multi layered notation while simultaneously reading (see section 3.4 Stream notation and examination). The booktracks chosen were picked considering their length and “popularity”. In addition, the choices had different original authors and the booktrack versions was created by “Booktrack” only. They further span over three years (released 2013, 2014, and 2015), striving towards gaining a longitudinal sample of the booktracks’ stream design.

The notation served to identify what type, occurrence and purpose stream sounds had in relation to the written and overarching narrative. It also provided an empirical overview of the sounds used in the Booktrack designed streams. Further, the notation enabled the researcher to experience the narrative as a whole while concurrently gathering the data, allowing for a deeper understanding of the medium itself.
In addition, the stream of a few booktracks was recorded by the researcher. This was done to later observe and visually examine the experienced stream via the use of a spectrogram (described in detail in section 4.3.2 Temporal analysis findings). In Table 1 below, each booktrack examined in the study is presented. However, because all booktracks can be found through the Booktrack website or application, only the ones actively mentioned in the text has a reference. It should be noted that during the revision of this thesis, the Booktrack website was altered and the Booktrack Studio feature disabled. However, the booktracks remain accessible through the Booktrack application.

<table>
<thead>
<tr>
<th>Name</th>
<th>Release</th>
<th>Length (words)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A reflection</td>
<td>2015</td>
<td>255</td>
<td>-</td>
</tr>
<tr>
<td>20,000 Leagues under the sea (excerpt)</td>
<td>2013</td>
<td>2785</td>
<td>-</td>
</tr>
<tr>
<td>Little red riding hood</td>
<td>2013</td>
<td>1415</td>
<td>-</td>
</tr>
<tr>
<td>Hansel and Gretel</td>
<td>2013</td>
<td>2922</td>
<td>-</td>
</tr>
<tr>
<td>Sherlock Holmes: The adventure of the speckled band</td>
<td>2013</td>
<td>9896</td>
<td>-</td>
</tr>
<tr>
<td>Pinocchio</td>
<td>2014</td>
<td>8561</td>
<td>Booktrack (2019c)</td>
</tr>
<tr>
<td>Rikki Tikki Ravi</td>
<td>2013</td>
<td>5951</td>
<td>Booktrack (2019d)</td>
</tr>
<tr>
<td>A Descent into the Maelstrom</td>
<td>2015</td>
<td>7075</td>
<td>-</td>
</tr>
<tr>
<td>The Oval Portrait</td>
<td>2015</td>
<td>1293</td>
<td>-</td>
</tr>
<tr>
<td>Ms. Found in a bottle</td>
<td>2015</td>
<td>4225</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1: Examined booktracks.

4.2 Expert interview procedure

Because narrative presentations consisting of streams superimposed on text was considered obscure at the time of the study, two experts with experience in related aural fields were enlisted. The experts in this study are referred to as Expert A, and Expert B. Both experts were male and proficient within music composition, critical listening, and sound design. Expert A, aged 43, had been working with music composition and critical listening, and sound design. Expert B, aged 34, had a more technical background with interest for music production and had also worked actively with listening and environmental audio.

Because the literacy extension in the medium was aural in nature (operating under the belief the streams were added to the text), the experts experience with sound design, music, and listening meant they were able to accurately discern and describe the aural components of the stream. Thereby, they were able to communicate and provide both detailed and valuable insights into the medium based on their brief experiences interacting with it.
The interviews themselves took place at the experts’ respective workplaces, which the researcher visited for the duration of the interviews. Each session begun with the researcher asking demographic questions before going into detail on the stream and medium. Thereafter, the participants were given time to read two booktracks before the interview continued. In both interviews, the participants chose one booktrack themselves and were then also asked to read the “Rikki Tikki Tavi” booktrack (see Booktrack, 2019d). This booktrack was chosen because the researcher had found it to be of good quality among the examined booktracks, allowing for the stream design and application in the medium to be discussed without focusing on the negative aspects a poor design might have entailed. The experts read the booktracks using a tablet and headphones provided by the researcher.

Prior to the interviews, each expert signed and received a consent form and agreed to be recorded for the study (see Appendix B for consent form used). Finally, the interviews followed a semi-structured format, based around the questions found in Appendix C. Both experts were encouraged to elaborate on topics of interest throughout the sessions.

4.3 Learnings from the pre-study

Hereafter, the findings from the different methods used in the prestudy are listed individually. Finalizing the chapter is a section summarizing the findings from each individual method, leading into the next part of the study.

4.3.1 Stream notation and analysis findings

The stream notation and analysis indicated clear patterns throughout the Booktrack stream design. Firstly, the classification of sounds were contextually related to the narrative, meaning the sound employed we applied based on the environment described or implied through the text. Generally, most sounds were observed as Natural Sounds (due to the most prevalent ambiances; wind, rain, and animal sounds, falling into that category), unless the narrative established a particular environmental context in which such sounds were omitted (e.g. inside, underground, etc.). Further, in relation to the remediation of the narrative, a clear trend among the types of sound (sound effects, ambiances, music) was identified. In Figure 8 below, the sum total of the ten tracks are presented. Individually, the same trends could be observed, which in
turn indicate a distinct approach towards the stream design in relation to the written and overall narrative.

As seen in the above tables, each level of remediation was dominated by one type of sound. Further, as can be seen in the stream sound purpose table, music was almost exclusively meant to evoke an emotion for the reader or a particular mood. In most cases, establishing mood was considered synonymous with playing thematic melodic music throughout various parts of a booktrack. Exclusively, sounds used leaned towards the embodied category on the sound spectrum, rarely requiring effort to process as a reader. However, as the sound effects predominantly were applied literally alongside the written, this was observed to help the reader make sense of more encoded sounds, providing context through the text. Music was further only applied on a non-diegetic level, meaning to affect the reader’s state of mind in relation to the written. To caption the level of remediation in relation to sound types: (1) sound effects were attached to literally written cues, (2) ambiances were almost exclusively interpreted from the narrative context, and (3) music was appended in a non-diegetic manner, striving towards evoking emotions and thematic moods through the stream. Further, music was generally constructed with a clearly audible beginning, middle, and end, resulting in a clear dissonance in relation to the narrative if the reader did not keep up with the pace of the booktrack. This was not as clearly observed when low-key music was used and subsequently did not cause an issue in those cases.

The overall stream design adhered to a design similar to high fidelity soundscapes (Schafer, 1987, pp. 43-44). Even though some narratively implied stream environments might have been interpreted as low fidelity, the stream design most always allowed key sound events to emerge.
audibly throughout. In cases where this differed, music was the prevalent culprit dominating the stream in particularly action filled narrative sections. Akin to film’s cinematic sound design, this indicate an awareness towards the stream design through which the most appropriate and purposeful sounds were prioritized for the presentation (i.e. sounds that are supposed to be heard are applied, not everything audible in a given narrative environment). However, outside keeping the stream’s aural qualities perceptible, most sounds were presented as either mirrored aural representations of that which was written, or as a general interpretation of the environment. The key here is that most sounds served the purpose of representing what was happening in the written, rather than contributing, or independently being used, to affect the narrative. In some cases, attributes such as a “heavy” door was accentuated in the stream design. However, rarely did even such examples do more than flavor the written. This might attributed to the inherent approach of the booktracks, being essentially remediated into “stream enhanced versions” of existing literature. In particular, the Pinocchio booktrack (Booktrack, 2019c) included several sounds indicating events found in “cartoon” themed sound design. As an example, a sound was applied to Pinocchio’s nose when it was growing, both indicating the written event alongside the thematic qualities of the stream design, situating the mood of the written.

4.3.2 Temporal analysis findings

The “Pinocchio” (Booktrack, 2019c) and “Rikki Tikki Tavi” (Booktrack, 2019d) booktracks’ streams were recorded alongside notating and reading. Above in Figure 9 is the experienced stream of the Pinocchio booktrack. The recording was loaded into a spectrogram tool to provide another perspective into the overall stream design.

Figure 9: Pinocchio stream spectrogram analysis.

The “Pinocchio” (Booktrack, 2019c) and “Rikki Tikki Tavi” (Booktrack, 2019d) booktracks’ streams were recorded alongside notating and reading. Above in Figure 9 is the experienced stream of the Pinocchio booktrack. The recording was loaded into a spectrogram tool to provide another perspective into the overall stream design.
The two sections represent the two stereo channels (left above, right below). The black and white gradient represent the amplitude of any given frequency (white loud, black silent). Finally, the X-axis represent time (left start, right end) and the Y-axis (per channel) span the frequencies 20 Hz (low bottom) to 20 KHz (high top).

What can be easily identified is the stream’s ongoing nature alongside it spanning across the entire bandwidth of frequencies throughout. This shows similarities to white noise, which support the streams presence as an, in large, ambience in relation to the text and narrative. There are four distinct breaks in the stream visualization, all of which triggered from the researcher pausing the stream during the notation or from the researcher not changing pages quickly enough. That being said, it can be observed that the stream was very rarely “silent” (in comparison to surrounding stream material) for an extended amount of time.

4.3.3 Expert interview findings

Following is the most relevant data from the expert interviews presented. The chosen excerpts focus on what correlated and what differed between the two experts. Meaning, their experience with, expressed preferences for, and their thoughts on the technical medium, the narrative, and the stream design was the focus of the interviews.

When asked about their first impressions after experiencing the booktracks and the narrative presentation therein, both experts were positive towards the notion of the medium. When asked if anything stood out in particular, both experts reacted to the music and the usage of it in the booktracks. Expert A explained:

[…] but what I experienced now was, that I became a bit distracted. There was some staccato music to incite tension, and then you are supposed to read at the same time, I found that it felt strange.

Similarly, Expert B said:

I believe it becomes a problem when the music is supposed to tell you *something*, which you already know… *Now* there should be tension, *now* scary […]. A kind of drone as music might be enough. […] and then there is that, there is really no need to play music all the time. It is just like in movies and games, it is being overused.

Indeed, both experts believed that music should be applied either with strong intention or be versatile and nondescript, with the main purpose to set the mood Further, Expert A noted:
... and music is so sensitive, because music is semantic, it is about describing emotions...

Expert A continued on the topic of possible issues with semantics in sounds:

... [And] if there is an inherent semantic meaning, it should be used for a purpose and not as an ambient noise. Because then it will be noticed. [...] [So] if you are not trying to describe something very specific... simply put, you need to be careful with it.

Because there existed a strong critique to the musically abundant booktrack design, the experts were asked what kind of sounds they would prefer in the stream. Both experts unanimously said environmental ambiences, as exemplified by Expert B:

[But] ambiences are the most important, because they give you so much. You are treated to an [aural] environment... so this is how this place sounds like [...].

Further, both experts pointed to the importance of cultural awareness when dealing with ambiences themselves in streams. As said by Expert A:

Yes, when changing environments, I think you quickly understand that something is different. Though, I also think it is very culturally stipulated, a Swedish park does not sound like and American one, for example.

On the narrative presentation and the interplay between the text and the stream, the participants were asked if they saw it possible for the stream to carry more narrative weight. Even further, they were asked if they thought parts of the written might be replaced with purely stream content, to which Expert A responded:

I’m thinking that it would be interesting to see what happens when you remove [textual] descriptions, will the text be as good then? [...] Then, what happens to the format itself? It becomes something else... I think that it would be required that you write books for this kind of format.

When presented with the same notion, Expert B said:

Yes, that is what I think. I can imagine authors for children’s books considering adaptations of already existing and easier to read literature. [And] given the digital world we live in, perhaps this could get more children to read... by creating a larger experience.

However interested in the study’s topic of inquiry, both experts expressed reluctance towards the medium outside its novelty, comparing it with reading silent texts. As an excerpt, Expert A noted after finishing the second booktrack:
I think this one worked better, at least compared to the earlier one. But I am not sure if I would keep reading this way.

On the synchronizing sound effects in the booktracks, Expert A raised an important issue:

I became stressed. Because I knew it was supposed to synchronize with my pace and then the played sounds didn’t, I felt that I was at the wrong place [in the text]. However, [towards the end] once the system started to recognize my pace, and I calmed down and didn’t think about it.

Further, on the narrative and the streams impact upon the perception, Expert A said:

I think the aural environment established the scene. I mean, now I remember things from the story that usually are not so clear. […] And we associate sounds with things… If you are at a specific spot… it creates strong associations, some sounds… settle. So, if you design sound well for this [narrative presentation], I think you can achieve it remaining in memory longer, at the very least.

Expert B speculated on the technical medium on how customizability might provide enough interest and versatility to medium:

Some people want music when they read, it is a good way to isolate yourself. I sometimes have trouble reading when the environment is too noisy, so this might be a way to isolate yourself and simultaneously enhance the immersion. And then I think that if this would become something of an established medium, that if you could choose your experience, I want music, or I want low key music, or maybe only ambiences.

The overall notion of the medium was received with interest and curiosity by the experts. Through this, three themes of interest were identified surrounding the technical medium and the narrative presentation. Firstly, (1) the stream purpose in relation to the written text and the underlying sound design in the stream, second, (2) how the reader perceives the narrative and concurrently the stream components themselves as enhancing or impeding. Finally, the third (3) theme regarded user interaction with the technical medium and technical aspects of the stream design (including crossfades, layers and similar). The first and second theme was closely connected to the research question of the study and further, to the field of audionarratology. In part, the third correlated similarly to the aim of the study. However, the concept of user interaction and concurrently the user experience of the presenting technical medium implied a development and focus on product creation, which was not the intent of the study. Because of this, it was in large disregarded in relation to the study evaluation and success. However, because it was identified as a distinct theme, it was a part of the prototype developments, surveys, and later interviews in the study as essentially a subcategory to the research questions.
Overall, the learnings gained provided a solid framework for how to regard the stream design, both in terms of what sounds to focus on, but also in terms of the semantic qualities inherent in the sounds themselves. This led into the second part of the study, in which two different streams were applied to the same text to explore how a reader’s perception of the narrative was affected by it. Hereafter follows a summary of the overall learnings from the Prestudy.

4.4 Summary of Part 1 learnings

Themes identified as core for the study:

- Main themes
  - The stream purpose in relation to the written text and the underlying stream design.
  - How the narrative presentation is perceived and concurrently the stream components as enhancing or impeding the written.

- Sub themes
  - User interaction with the technical medium.
  - Technical aspects or the stream (playback, crossfades, etc.).

On the stream design:

- Do not overuse music.
- If using music, avoid letting it dominate the narrative and what the reader should feel. This might cause the reader to be distracted from the narrative being presented.
- Ambiences serve the best general stream component and might be applied for a long time without annoying the reader as it shares qualities similar to white noise.
- Sounds have cultural connotations, and if not geographically identifiable, it might imply specific things on an individual basis.
- Stream design should surround that which is relevant for the narrative, striving towards a high fidelity stream.
- Use looping ambiences and music to avoid the temporality of the stream to interfere with and possibly determine the reader’s pace.
On sound effects and synchronization:

• The booktracks used sound effects that “synchronized” with the reader’s pace. This was raised by the experts as potential cause for confusion, or even as stressful, as readers are aware that they should have reached or be at a particular point in the text. Simultaneously, this mismatch was raised as creating anticipation for what was coming. While this poses an interesting topic for discussion, the notion of synchronizing sounds also suggests an exploration of user experience and concurrently implies the development and focus on the technical medium, which was not the intent of this study. Therefore, the concept of precisely synchronized sounds (as found in the booktracks) were omitted throughout this study.
5. Part 2: First prototypes

While a thorough examination of an existing application provides useful insights, it can never completely replace learnings gained from prototyping, as is inherent to research through design. Therefore, a two stage prototyping process was applied during the study. This chapter describes the first initial prototypes during this process. At this stage, the prototypes were meant to elaborate on the Part 1 findings in a tangible way, while simultaneously building towards answering the research questions and inspire the Part 3 prototype design.

Furthermore, as the prototypes were put together using Booktrack Studio, some inherent issues were identified and will be presented briefly. The toolkit was meant to create booktracks as presented by Booktrack. Therefore, it was in large reliant of two features for interaction with the “stream timeline”. This meant there was only two ways to dynamically affect and change the stream using the Booktrack Studio. Namely, these were: (1) During intentional (chapter) page changes, and (2) via triggering from the aforementioned “sing-along tracker”.

Because the intent behind the study at this stage was not to explore how dynamic and automatic triggering of sound effects and stream changes felt for readers, the second was, in large, neglected through the study. Instead, all prototypes throughout the study had users reduce the tracker’s “words per minute” (WPM) in the prototypes to the minimum of 50. This served two purposes. Firstly, this reduction meant the sing along tracker was slow enough to avoid accidental triggering of stream changes, allowing the reader to be essentially free to read at whatever pace suited them. Second, it created a situation where the sing-along tracker could consciously be used as a tool. Because the tracker was automatic in both its playback and how it handled fades of sounds (predicting that in X seconds it will reach the end of a sound meaning it would trigger a fade out ahead of time to avoid clipping the end), this enabled the researcher to use the trigger as a pseudo-timer for how and where fades took place. The duration of fades could be set separately, though needed to be adjusted on the text-playback layer to suit and avoid artefacts being created (clicks, pops, and unnaturally quick fades akin to cuts). Further, the knowledge that the tracker might work as a roughly controlled trigger meant the stream could include sound effects that would trigger on a delay, thereby creating a more dynamic aural environment.
Returning to the (1), the page changes, this was used alongside the adaption of the tracker. Because the booktracks were presented in a reactive manner, suited for web, phone, or tablet presentation, the text written in one chapter would be separated into pages beyond the designer’s control. However, the inclusion of chapters made possible a page change defined by the designer. Indeed, this allowed the researcher to design streams in conjunction with the identified narrative sections present in the story. As a tradeoff, this meant some pages and chapters would vary in length and textual representation in the presentation (as the reactive nature decided how it the text was presented).

5.1 Prototype design and intention
A key finding from the Prestudy was the prevalence and utility of ambiences and ambient sounds in the streams. Given the interpreted nature of these, it was assumed a change in the stream therefore might affect how the written was perceived by the reader. Further, as was noted in the expert interviews, the usage of synchronizing sounds might cause confusion or stress whilst reading. Therefore, precise synchronizations were omitted completely, instead striving for loose synchronization points, meaning to create and add dynamism to the stream independent from the reader’s pace.

Due to the scope of the study, an existing text was used as the textual component for these prototypes. This allowed the first prototypes to focus on the stream design and further, on how the overall narrative was perceived in comparison between the two versions. Further, it allowed the researcher to engage with the narrative in a practical way, leading towards Part 3 and the final prototype’s development.

For deciding which text to pick, a set of criteria were defined as follows: (1) the written narrative persisted over several “sections”, in which the potential aural environment changed to a large extent. These sections roughly correlate to the events of a fabula. (2) The text included a section which narratively could be described as an aurally “silent” one, meaning to explore whether an obvious amplitude decrease detracted from the experience when applied with intent. (3) The textual ambience descriptions should be general and few, allowing them to be removed and effectively replaced by the stream. Further this third criteria would allow multiple versions of a stream to be designed without unnecessarily altering large portions of the written text. The text chosen consisted of two concurrent excerpts from the book “The wheel of time” by Robert
Jordan (1990, pp. ix-x, 1-5). Further, the first excerpt for the prototype was cut short to create a shorter and more cohesive syuzhet of the narrative’s fabula.

5.2 Stream design

The chosen text contained two distinct section in the narrative. The first was a battle aftermath inside a large palace building, and the second was an outdoors snowy environment. Further, the second section contained a part in which the protagonist’s perception would “drift”, thereby providing a possible subsection in which “silence” might be introduced in the stream. Below in Figure 10, the two different stream designs are visually shown in relation to the identified narrative sections.

![Figure 10: First prototypes' stream structure and design.](image)

As can be seen, most all sounds used were looping ones. This was done to ensure the stream temporality was independent from the reader’s pace. Further, at almost all points in time, several layers of sound was interleaved (i.e. playing on top of one another, at the same time). This technique is often utilized in sound design for games, referred to as “layering”. Essentially,
several looping sounds of (often) varying duration play at the same time. This results in an ever changing aural environment, as opposed to using a single looping sound containing all layers, which would lead to a more easily identifiable looping stream.

Further, the major difference between the two streams can be seen during the second section. The first (top) version aligned with what was implied through the text, while the second replaced and added several sounds in the sections. Links to the two first prototypes can be found in Appendix E.

5.3 Focus group procedure
Similarly to the expert interviews, the study second part focused on the stream design alongside the perception of the overarching narrative presentation. Therefore, the ability to accurately convey the aural experience was deemed an important quality for the enlisted participants. Because of this, each of the four focus group participants were involved with and had experience with sound design similarly to the experts. Hereafter, each of the four participants will be referred to as FG1 through FG4. The focus group was greeted and introduced in detail about the study together. Thereafter they were briefed of the session and asked to sign a form of consent prior to participating in the actual session. Each participant received a form of consent signed by both the researcher and the participant in question (see Appendix B). The session main structure followed: (1) Read one of the prototype versions (divided equally between the participants), (2) complete an individual survey (see next section), and (3) focus group interview and discussion regarding the medium, stream, and the participants’ perception of the narrative in general. For the purposes of clarity, FG1 and FG2 both had the first version (original interpretation) of the prototype, whereas FG3 and FG4 had the second alternate version (changed and added sounds). It was pointed out that the participants could take their time to read through the prototype and it was explained the pace of reading was not being noted for the study.

The participants had been asked to bring their own devices for reading along with headphones for listening. The researcher provided access to headphones which were borrowed to one participant during the session.
5.4 Survey procedure
In addition to the focus group, a survey was created. Mainly, it consisted of “motivate your answer” questions intended to gather individuals’ opinions on topics prior to the focus group discussion. The idea was to enable an unfiltered channel for every participant regarding various topics. Therefore, the survey was made anonymously to further encourage critique and honest opinions. In addition to the focus group participants, the prototype(s) were distributed to six external (non-focus group) participants, meaning the survey had ten respondents in total. The external participants received the same survey but did not come together for a focus group interview session. Because both the experts and the focus group consisted of participants connected to the field of sound design and sound studies, the external participants were intended to counteract possible aural bias in the gathered data thus far. Further, as the final part of the study would employ a more refined prototype and survey, this initial survey also served as a first iteration of the second one.

5.5 Learnings from the first prototypes
Following is the most prevalent findings and epiphanies stemming from the focus group and survey results. They are treated and presented in separate sections, where the focus group participants are defined as FG1 through FG4, and the survey respondents (SR) are noted concurrently in the text.

5.5.1 Focus group findings
Several aspects of the experience and the potential of the narrative presentation was raised by the participants. The stream design was noted as efficient as an additional channel affecting how the participants perceived the written text. One participant noted:

FG1: I believe I heard fire in the background, [...] the sound gave the impression that the room was dark.

Similarly, another participant noted the details in the stream as being a key component for adding to and affecting their perception of the narrative environment:

FG3: I found that, as soon as there is such a large reverb present, my first thought was that the walls were made of stone. [That] it was some kind of castle or similar medieval building.

This notion ties into the concept of soundmarks (Schafer, 1977, p. 9), attributing the reverb in the stream assigning specific environment to the narrative. After a discussion regarding the
general stream design, a distinct difference between the groups were identified. While both expressed that they enjoyed the presentation overall, the first (original) group noted how the stream felt static from time to time. The following discussion culminated in the participants speculating around how they would approach the stream design themselves in relation to the narrative presentation and the technical medium. Focusing on the sections in the narrative, one excerpt from the conversation follows:

FG1: Like this, it was a bit static. And it became hard to connect it [the stream] to what was happening in the text, but maybe if half of it [the text] could be assigned one playback [stream] each?

The discussion regarded how the textual and aural interplay took shape in the technical medium, and thereby, how it might be approached in the narrative presentation (syuzhet) design. This issue was identified in relation to the first version only, which in its presentation was more low fidelity than the second. This indicate that high fidelity streams are preferred over low fidelity ones, although the overall level of detail in relation to the narrative was similar. Eventually, the participants agreed that streams should be designed with intent for the textually present narrative sections. Further, this discussion led into what actually differed between the two stream versions. From a brief comparison, the participants believed the second version was more dynamic and that “one-shot” sound effects might be employed to increase the variation in the stream (as had been done in the second, but not the first). However, the generality of the stream and the level of detail present seemingly achieved to affect how the written in both versions. One excerpt follows:

FG4: I think that it captured the environment and the moment in a general sense. There never was too much sound, so that you would focus on it rather than [the text]. This sets the tone of the page, and in that respect, I think it was good.

Building on this, the participants agreed that ambiances and possibly tonal drones would serve as the core components for stream design. Further, when speculating around sound effects being applied, they concluded that synchronization needed to be precise and match with the reader’s pace if representing something written in the text. Simultaneously, this conversation sparked the idea that sound effects might not need to be described in the text at all, and concurrently not need to depend on the synchronization with the reader’s pace. Furthermore, there was an ambiguity identified in what was considered a sound effect and an ambience. Here follows an excerpt of the conversation between the participants and the researcher (R):
FG2: So, if sound effects could be used, you should avoid applying anything too specific. It is one thing when it regards something universal, like fire, as opposed to a scream or similar. Then maybe you [a reader of the narrative] think that, this is not how I imagined the scream.

R: So, the purpose is to enhance, but simultaneously avoid altering the reader’s experience, imagination?

FG2: Yes.

FG4: However, if it [the narrative] is located in a dungeon somewhere, in some seventeenth century castle, you could apply them…

R: … in the ambience?

FG4: Exactly.

R: Does it differ then, between using such sounds to establish a mood, as opposed to using it to represent a specific someone screamed?

FG3: Precisely.

In general, the technical medium was considered an alternative to silent texts by the participants. Further, they unanimously thought it would be the most suited for use at home due to the more controlled aural environment there, as exemplified by the following quote:

FG3: When regarding these kinds of formats [books, audiobooks, etc.], many use them to and from work when external [aural] sources might affect the initial placement and design of sounds [in the stream]. I would imagine that it [the technical medium] is better suited for at home use, where nothing external might bother you.

Finally, the participants believed the stream had affected how they had perceived the narrative. The following quote is brought forward due to the potential of future studies highlighted through it:

FG4: For me, it [the stream] enhanced the experience. I have dyslexia and it is easy to lose track when reading. And then I have to remember where I was, what I was reading, what was happening, and where they were heading. It was like, even though I drifted, I still remained in the moment, and for me, that was very nice.

5.5.2 First survey findings

The survey participants numbered ten, including the focus group participants. The following answers have been noted as SRn (where the n is a number iteratively increased per answer). The noted answers have been chosen to exemplify the correlations and differences between the
respondents. Overall, the stream presence was considered an immersion enhancement alongside the written text, with one participant noting:

SR1: I felt like the first one played out inside a house with a fireplace in the background. It felt really dark. The second one was clearly outside since you could hear a strong wind.

A question posed asked the respondents to describe the narrative over the two sections in a free text answer. Throughout, all participants provided a summary of the written textual narrative, leaving out the stream component entirely. One of the shorter answers follows:

SR2: In the first one there was a man who was looking for his wife. And in the second, there were two men walking and one of them saw that they were being followed, or so he thought.

This might have been attributed to several things. Both in terms of the deliberately vague question, and the general notion of the medium itself. Being so closely related to silent texts, in both the way the user interacts with the medium (only changing pages from time to time) and in the text’s role to carry the majority of the narrative descriptions, it is understandable the textual component might constitute what is recollected and retold. A longer example of the answer follows, to exemplify the indication further:

SR3: First. Two men in the midst of a room in the aftermath of a battle, one in distress over a dead woman who was one of the men’s wife. The other had lost his memories according to the first man. The two of them were in disagreement over the situation at hand. Second. A father and son riding along a trail with goods, heading towards a city and an upcoming festivity. The son gets disturbed to the core by a mysterious rider traveling behind them. Stumbling, the son looks away for a moment and the rider is then gone. His father does not disbelieve him but sees no traces of the rider.

Despite this, there seemed to exist a positive attitude towards the stream’s presence in the overall experience. However, the imagination of the reader was brought up as an issue several times. This suggested that the readers’ perspectives on the medium was heavily influenced by silent texts, in which the written alone serves to evoke the imagination. Building on this, the notion of the stream therefore seemingly imposed upon some participants’ expectations of such “read mediums”. An excerpt raising the issue reads:

SR4: [the stream was] bad in [the] way that if my imagination is interpreting it [the written] in a different way, you can’t get away from that [aural] suggestion.
While the impact of the stream was considered an issue for the imagined in some cases, it also improved upon it in others:

SR5: It sure did give a more intense reading, you could feel the changes in the environment and also the different emotions of the characters which made it more relatable. You really felt the difference from just reading a regular book, where it sometimes can be difficult to understand certain scenarios or even scene changes.

The word “intense” was interpreted as an adaption of immersive by the researcher, based on the text following the word. This quality of immersion was further raised by several of the survey respondents, as exemplified in this final quote of the survey findings:

SR6: The sound definitely made it easier to build the environment that the stories took place in, and also helped with the immersion a lot more. I have never read a book with ambience sound added to it, so this was a really interesting first try […]

Indeed, the novelty of the technical medium itself exemplifies in some answers directly, and was implied in others through the comparisons between the medium and silent texts. Building on the answers surrounding the perception and the summary of the narrative, the final part of the study took shape as a way to further explore how the stream might be used. Thereby, the final prototype would be designed with the intent to create a narrative presentation in which both text and stream had narrative impact and affect.

5.6 Summary of Part 2 learnings

- The stream does affect how the written text is perceived and visualized, at least when they make sense in the narrative.
- The presence of the stream was raised as an imposer into the imagination of the reader.
- Sound effects should be applied with intent in instances where they do not interfere with the reader’s imagination.
- Sound effects, not recurring sounds, might be used to diversify the ambient stream design.
- Sound effects applied and established with intent may carry narrative impact understood by readers.
- High fidelity streams are preferred over low fidelity ones.
6. Part 3: Final prototype

For the final part of the study, a more elaborate prototype was developed. The findings from the first and second part constructed the framework which defined the stream design. Further, they had raised queries guiding the narrative design and the presentation components’ relationship. Essentially, the final prototype intended to focus on developing the entirety of the fabula and syuzhet of the narrative, rather than designing a stream for an existing text. The evaluation of this final prototype was made using a refined survey alongside a revisit to the experts recruited during the prestudy (Part 1).

6.1 Designing the narrative

Because the intention was to actively consider both components presenting the narrative (stream and text), both would be created from scratch during development. An existing text might have been adapted for the purpose, but it was rejected as an option. To elaborate; an adaption of an existing text would require a good understanding of the existing narrative, and because a large part of the prototype regarded designing the stream alongside the text, a bottom up development was chosen. Further issues might have surrounded copyright infringements given the prototype would be publicized after development, adding towards the reasoning of creating the whole as part of the process. It should be emphasized that the purpose of the prototype was to explore a possible approach when designing for the presentation of the narrative and not the user interaction with the technical medium. Thereby the perception of the narrative was observed further alongside exploring the stream design. Building on this notion, the following sections describe how several key parts defined the shape of the narrative design, and concurrently, the final prototype.
As can be seen in Figure 11, a similar structure to the first prototypes was employed. However, during this iteration, sections were divided not only for narrative purpose, but also for solely stream changes. In addition, there exist three instances of “bell” throughout the stream, representing a sound effect that was intended to serve as an aurally presented event of the fabula of the narrative (see section 6.1.2 Design aspect 2: Stream’s active narrative impact). A link to the final prototype alongside some additional information can be found in Appendix E. In Figure 12 below, an example of how the stream and the text components fit into and relate to one another through the various sections is presented.
6.1.1 Design aspect 1: Introducing hypermediacy

The main focus in the stream remained on creating and using ambiances to enhance the written text. So far in the study, most ambience crossfades (i.e. when one ambience is replaced by another) had been made on a section to section basis. In the booktrack versions, these changes were tied to the system’s tracker, triggering changes at certain points, whereas the first prototypes operated on a section-to-section system. Further, the length of the crossfade in the prototypes was considered both too short and too long, depending on the readers’ individual pace, meaning the ambience changing did not correspond with the expectations of the reader. The notion of creating automatic crossfades of “reader optimized duration” align with Bolter and Grusin’s concept of transparent immediacy (Bolter & Grusin, 1999, pp. 21-31). However, making the process automatic is only one of several ways to achieve immediacy. Therefore, this prototype would attempt to flip the auto-crossfade mentality and embrace the hypermediacy (Bolter & Grusin, 1999, pp. 31-44) inherent to the multi-component presentation. Thereby, the presentation wanted to embrace the shortcomings of the stream section fades and thereby allow for sections where the aural change was highlighted, rather than hidden.
To explore this alternative to automatic changes, “transition sections” were introduced as parts of the syuzhets (see Figure 11, p. 50). The idea was an attempt at allowing the reader to pause for a moment, and only provide a textual description of the transition. This textual component would support the aural transition with a simple description such as “[present]” or “[2 days ago]”. Thereby, the upcoming section would be introduced through the stream environment during this transition step. Smaller changes were kept on an automatic basis alongside sound effects that were applied to bolster the dynamism of the stream.

6.1.2 Design aspect 2: Stream’s active narrative impact
As stated, the booktracks and the second prototypes’ stream designs served mainly as illustrations of the text. As such, they were never used as a narrative tools outside of establishing the mood of the narrative (e.g. Pinocchio; see Booktrack, 2019c). Alongside this, the findings from the second part indicated that the written was considered the main narrative describer. Therefore, it was decided that a core component of the narrative, an important event in the fabula, would be presented through the stream during the final prototype. Further, if the stream was to carry understandable narrative impact, the surrounding elements needed to be constructed to help the recipients make sense of the stream sound events. This notion and condition was decided upon based in both audionarratology (Mildorf & Kinzel, 2016, p. 14), and multimodal literacy (Hallet, 2014, p. 168).

With this idea in place, it was decided that some kind of sound signal (Schafer, 1977, p. 9) would be used, as such sounds inherently signify an event calling for attention (i.e. an alarm of some kind). The fictional story was decided to take place in a medieval setting for two reasons. Firstly, this allowed the general stream to consist of relatable, and high fidelity (Schafer, 1977, pp. 43-44), aural environments. Secondly, this would enable usage of a sound akin to a church bell as the attentive, signal, sound. By defining the environment and the sound signal (hereafter bell), this allowed using the inherent aural and semantic qualities of similar bell sounds. Further, to ensure the bell making sense in the context of the narrative, it was decided it needed to be heard three times throughout the unfolding (syuzhets) of the story. Firstly to establish the bell sound event and its presence in the both the narrative and the stream. Secondly, remind the reader of the context and implication of the bell. Finally, apply the sound in the stream alone, leaving the reader to interpret it on their own alongside the text. Further, rather than synchronizing the sound to a particular written event, the written would react to the bell instead. This meant that rather than explaining “A bell rung” and synchronizing a sound, a character in
a dialog might react with “Did you hear the bell earlier”, or similar after the bell had rung. This served two purposes, (1) to establish the sound as a diegetic as well as a non-diegetic (Chion & Gorbman, 1994, pp. 73, 79-82; Halliwell, 2013) event for the reader, and (2) make the reader aware that the stream is connected to the written text. Finally, this meant that the final time the bell rung, it would only be reacted to according to what had been established beforehand in the story. In the context of a narrative, this bell might represent a multitude of things. It was decided the three bells would signify a countdown of some kind, meaning the final (third) bell indicated the last strike before something ended, happened, or would happen, in the story.

6.2 Final survey procedure
Because the final prototype was to be published in a public setting, a refined survey was developed alongside the prototype. Mainly, the questions were multiple-choice to allow for an effortless submission. However, one key question was kept as a fill in form. Namely, asking the respondent to briefly explain why the protagonist “wakes up and leaves” towards the end of the story. This signified the moment in which the bell would have sounded the third time. Further it allowed a focused question which might either be described by the textual reasons (where is the protagonist going) and the stream triggering reasons (why leave now) alongside the overall narrative reasons (why is there a need to leave at all). Further, because the overall aim of the questions was to examine the experience as a whole, a few questions regarded comprehension of the text, one of which was the aforementioned “waking and leaves” question. This was applied as a way to identify and exclude data from respondents who had skimmed or skipped through the narrative. Because the comprehension of the stream and textual components were of interest, this was deemed necessary. It should also be noted, that these filtering questions were based in the written text and not in the stream, as the earlier prototypes had indicated the text was the most apparent descriptive component of the medium. The exception was the mentioned fill-in form, which was consciously created with room for interpretation.

The survey was sent along with the prototype through several channels and was also linked through hyperlink and QR code on the Booktrack website (where it was published). In addition, the survey was distributed to both experts (see next section).
In total, 12 answers were collected through the survey, out of which 3 were ignored from failing the eligibility questions in the survey. This left 9 respondents as viable for processing and analysis.

6.3 Expert revisit interview procedure
Alongside the survey, a revisit to the original experts was arranged. Prior to the session, the final prototype and survey was sent to the experts for reading and individual evaluation. As per the first interview sessions, the researcher visited each expert at their respective workplaces. Because both experts already had signed a consent form valid throughout the study, their rights were reiterated and they were only asked if they still gave their consent and wanted to participate. As the first sessions, these interviews were recorded for later processing by the researcher. Finally, the interviews followed a semi-structured format, based around the questions found in Appendix C and the experts were encouraged to elaborate on topics of interest.

6.4 Learnings from the final prototype
Hereafter follows the last findings from the study. First, the expert interview findings are presented in a similar manner to the previous iteration. However, to differentiate between the two expert interview sessions, the experts will hereafter be referred to Expert C and Expert D, not necessarily with correlation to the earlier interviews. Thereafter, the findings from the final survey consisting of 9 respondents (which passed the filtering test) is presented.

6.4.1 Expert revisit findings
From the final interviews, a few insights were gathered that surround the design approach of the final prototype. Building towards an understanding of the multimodal presentation, here follows some excerpts surrounding the prototype, narrative design, and how the stream affected their perception of the story.

Starting on the narrative presentation, Expert D highlighted both why the hypermediated crossfades were good and when they might become an issue:

I think it [page turning crossfades] worked really well, having the crossfade transitions take a few seconds. And then I also appreciated that there was smaller changes as well. Because you start to expect changes every time if you start using them [page turns] as triggers for change. […] And that is why I found the smaller changes positive.
On the same topic, Expert C wished for an audible break before the narrative stream initially started (prior to it, some thematic music was playing).

I would have wanted to music to fade out completely, allow for silence for a few seconds and then have the ambience start. To establish a sort of, now it begins. [...] As it is currently, it all flows into one another.

Leaving these aspects, both experts discussed the streams impact from different perspectives. On the level of detail in the presented stream and the notion that the stream might become oversaturated with narrative intent, Expert D said:

I don’t think it needs to be any more significant [than this]. [...] [And] if the sound is meant to represent or support the [written] environmental description, you might trap yourself. Because such [textual] descriptions might include so many details a sound cannot describe.

Indeed, the stream cannot describe everything and the multimodal presentation of the narrative probably needs to be explored further. Expert D continued:

I think it is inherently dependent on what kind of text it is. Or whether you write with the medium in mind [or not]. It would be interesting to experiment with the disposition of stream sound and written environmental descriptions. But I can imagine certain stories are more suited [for this kind of presentation].

In the same mind, Expert C was discussing how the medium might actually be closer related in its presentation to movies than silent texts:

It is not an audiobook, it is something more specific... It would have been intriguing to apply a cinematic mentality to this medium, as movie narratives work with image and sound. [...] I did experience how the sound affected how I imagined the environment when reading. [...] It becomes like you apply sound to the imagined image, rather than the screen image.

When asked to elaborate, Expert C explained:

Because if you show two people a text that describes a forest or a fire, if you attach this description to a sound... You might affect the mental image, causing everyone to imagine it more similarly. [...] As in movies, that you might use the sound as one of the narrative channels to [for example] establish feelings. And thereby, write more focused around characters, events, and elaborations rather than describing the surrounding [aural] environment.
Building on this subject, both experts took note of a particular section in the narrative, attributing the low frequencies present in the ambience to create a more dynamic and affective experience. Expert D noted:

There existed this dynamism in the sound. There was this wind sound, and then something darker, lower in frequency was introduced when I’d read a bit. That worked really well.

Expert C reacted to the same kind of sound, expressing it enhanced the mood of the narrative but also might have been supported further with more tonal elements:

During the most intense fire, there was something low key… it highlighted a dark mood. Perhaps a single tonal, and possibly dissonant, drone might have enhanced that feeling even more.

During one section of the story, a carefully constructed “walla” was applied in the stream (i.e. a crowd of people speaking without anything being particularly identifiable). Expert D raised how the nondescript nature was good for the narrative:

I was trying to figure out what language they were speaking, but I couldn’t, which was good. That would probably situated the narrative geographically, and that might have caused a dissonance in regards to the story itself.

In relation to the intent and merge between the stream and text, both experts saw how they had been used to together further the narrative, as Exemplified by Expert C:

I thought the bell worked as a narrative tool. […] In the story, there was supposed to be three rings of the bell, and then they would release the hounds after them, or whatever creatures it was. […] I think it has the potential to remind the reader about… what is about to happen, that something is about to happen.

6.4.2 Second survey findings
Based in the previous survey and in relation to the research question, the final surveys aimed to further elaborate on how the narrative was perceived, and how respondents observed and recollected it. Firstly, the response rate and filtering process is described. Thereafter, the answers considered are presented alongside some correlations that could be seen between the different responders.
Post filtering invalid answers, a total of 9 respondents were included in the analysis and findings. As can be seen in Figure 13, most respondents believed the stream’s presence increased their immersion in the story. However, a few did not believe it affected how they imagined the overarching story, which indicate a minor discrepancy in the responses. This might be derived from the lack of thorough definition of the word “immersion” in the question, leaving it open to interpretation. However, it might also be attributed to what FG4 noted during the Part 2 findings, that the sound helped the respondents “remain” in the narrative when “drifting”, thereby acting as a kind of auxiliary immersion boundary perhaps not consciously considered during the actual reading.

![Figure 13: How stream affected immersion, visualization, and mood charts.](image)

Along similar lines, the respondents indicated they believed the stream supported and affected the mood present in the narrative. In large, this was assumed to correlate to the low frequency sound as noted by the experts, indicating the streams ability to define a mood even without tonal musical elements. Closely related to the immersion quality, the stream’s impact of the perceived mood might therefore be considered an inherent attribute of the component in relation to the narrative presentation.

From the fill in question on “why the protagonist wakes up and leaves”, three camps of answers could be identified. In comparison with the previous (Part 2) survey, the presence of these camps itself indicate the stream was perceived more consciously when respondents read. The camps were identified as follows: (1) stream perceived as present with impact on the narrative, (2) implication the stream was perceived as having impact though not noted in the answer literally, and (3) overall objective of the narrative stated. In essence, the first and second might both be grouped as being (interpreted from their answers) conscious of the stream, while the third does not necessarily regard the stream’s presence as clearly. As stated in Section 6.1 (Designing the narrative) and Section 6.2 (Final survey procedure), the question was designed
to allow for interpretation from the respondents. Out of the answers, four from the first camp explicitly mentioned the bell (at said narrative point only presented in the stream) in their answers. Further, the one response from the second camp implicitly explained the circumstances surrounding the bell specifically in the narrative, describing both the pre- and post-bell events and actions of the protagonist. Excerpt from said respondent:

“He was waiting for Tyll, but she had not yet arrived. So he carried on towards the Summit!”

The rest of the answers focused on the general plot going on at that point in the story, focusing on the protagonist being hunted and needing to meet up with a friend. As can be seen in Figure 14, the conscious design of the narrative presentation did lead to a more conscious perception of both stream components in the medium, as opposed to the first prototypes.

While this neither guarantees the respondents embracing or neglecting the stream component, it does signify that stream components might be used to actively further the narrative. Pursuing the query of how narratives presented through the medium is perceived, this clearly show that the stream component not only (as mentioned during interviews and focus group) might flavor existing text, but also was perceived as an integral part in the narrative presentation as a whole. Further, this indicated that designed narratives for this technical medium has the potential to

---

**Figure 14: Stream mention among participants.**

---

While this neither guarantees the respondents embracing or neglecting the stream component, it does signify that stream components might be used to actively further the narrative. Pursuing the query of how narratives presented through the medium is perceived, this clearly show that the stream component not only (as mentioned during interviews and focus group) might flavor existing text, but also was perceived as an integral part in the narrative presentation as a whole. Further, this indicated that designed narratives for this technical medium has the potential to
stimulate both the read and heard alongside one another in ways that readers understand and can make sense of.

6.5 Summary of Part 3 learnings

- Crossfades on page changes was appreciated but also triggered expectations that something would happen every time. Therefore, avoid changing the stream to a large extent every time, or vary the significance if changes are applied.
- If the stream is going to be used actively to further narrative, it needs to be established for the reader at some point. If an inherent semantic component is used, it needs to be established in the context of the narrative.
- Ambient (general) streams might be employed to guide the imagination of several readers to similar visualizations, while also not imposing too much detail into their imagination.
- Use of culturally identifiable sounds, including language, should be approached with care and intent. The cultural implications both inherent to the sound itself and the reader’s interpretation of the sound need to be considered.
- Designing a narrative for the multimodal presentation indicate a need for awareness of both the text’s and the stream’s presence and purpose in the narrative presentation. Consciously approaching them both indicate both being perceived and becoming more clearly regarded in relation to one another (i.e. readers are, or become, more aware of both what they read and what they hear).
- Conscious design with the components in mind propose the possibility of using both in conjunction with one another, rather than one supporting the other as an addendum.
- The stream should tell what is suitable in context and complexity, as the written is able to tell so much the sound is unable to. However, the stream might also tell details unable to be presented through the text, i.e. this is what it actually sounds like.
- The multimodal narrative presentation as explored here might be considered closer to movies than actual books in the way the two components interplay.
7. Discussion

This study has engaged with a novel multimodal narrative presentation composed of text and sound. Both through examination of existing “booktracks” and through explorative prototypes, it has attempted to increase the understanding of the technical medium and the narrative presented therein. Furthermore, the study engaged with how and what “readers” of this medium perceive and further what purpose the aural (sound) “stream” might serve in the presentation of a fictional narrative. Alongside this, the study has assumed an aural perspective and thereby focused on the stream component of the multimodal presentation. As a result, it has defined a core set of guidelines for designing and approaching the stream of the medium (presented in section 7.4 Stream design principles).

To a large extent, the discussion is reflective in nature. Because the medium explored through the study was considered obscure, this means to elaborate upon the findings while also situate the researcher’s perspective and reasoning prior to the drawn conclusions of the study. Hereafter the study’s themes, research questions, and established theoretical framework act as nodes around which the discussion revolve.

7.1 Multimodal perception

Throughout the study, the perception of the narrative has been a key component, both from being present in the research question and by becoming identified in one of the study’s themes. In actuality, to understand the narrative presentation, and concurrently to experience any content fulfilling the requirements of the technical medium (more on this later in this chapter), readers are required to perceive and consume both the written and the heard. Simple as may seem, the issue of understanding becomes prevalent when considering that aural associations might be both culturally unique and fundamentally defining in their simplicity. In essence, prior experiences of the reader play a large role in how well they understand the stream itself (Huwiler, 2016; Skoulding, 2016, pp. 102-103). Further, multisemiotic literacy (Hallet, 2014) becomes a tangible issue when designing streams with particular intent for the narrative, i.e.: Will the reader understand the stream, comprehend and interpret it as I, as a designer, intend and further how do they actually interpret it, and why is that so? Prefacing even understanding, as raised by Mildorf and Kinzel (2016, p. 14), do readers even want to understand and make sense of it?
From the gathered data, it has been understood that most participants perceived the stream as an immersion enhancer in relation to the written narrative. In contrast to the booktracks examined, the study distanced itself from using music to evoke and emphasize emotions and action for the reader. Still, the streams’ design was considered affecting not only the immersion, but the mood of the overall narrative by participants across all parts of the study. In essence, the perception of the medium itself require several components to work in tandem, and these components are perceived to a lesser or greater extent regardless of the reader’s affiliation towards the medium and prior experiences. However, several participants raised the issue of their imagination versus the stream’s implications. This might be an indictment of the similarities to silent texts, in which the reader’s imagination is the sole interpreter of the narrative, but should not be neglected.

Indeed, this notion instead poses an interesting stream-to-user relationship in how the narrative is perceived. As noted, the presence of the stream raised concerns that participants either expected or experienced their imagination being breached by the aural component. Conversely, the purpose and potential of the stream was expressed to aurally elaborate on the textually presented narrative. Inherently, these two opinions, in some cases raised by the same participants, implied that readers of the medium affirm the streams presence and potential impact upon their perception of the narrative. Building on this, the perception of the narrative seemingly requires several interdependent aspects to be understood as intended.

Hallet (2014) defines a structure for how a multimodal novel’s components and reader comprehension act together to create the multimodal medium. An adaption of this structure is presented below in Figure 15 and serves to exemplify how and what is required to compose an experience in the medium presented throughout this study.
Figure 15: Stream and imagination inclusive adaption of Hallet’s constitution of the fictional world.

As can be seen, the required elements to comprehend, and design for, this particular medium might take many shapes according to this loose framework (as will be discussed further in the upcoming section 7.3 Designing for and defining the narrative’s presentation) The perception and interpretation is further affected by the willingness to understand, experiences, and culture surrounding the reader (Chion & Gorbman, 1994, pp. 20-21; Kuzmičová, 2016; Huwiler, 2016; Mildorf & Kinzel, 2016; Skoulding, 2016, p.102). In essence, the modalities present when designing define whether the narrative presentation subscribes to the explored medium, and subsequently build towards listing what comprehension abilities are required of the medium’s readers. Drawing on Chion’s discussion surrounding movies and audiovisual dissonance (Chion & Gorbman, 1994, pp. 37-39), it becomes interesting to further explore how obvious dissonance between the narrative presentation’s components are perceived.

When queried during the study, most participants believed the stream served a supportive purpose in relation to the narrative. This was considered an indictment from the booktrack approach to the medium, which acted as a basis for the first prototypes. Because the stream was developed through interpretation of the existing written text, it was assumed the inherent quality of the pre-written was granted more narrative allowance in the presentation. However, through the final prototype, an attempt to direct the reader’s attention onto the stream content was made successfully. Designed in a way to first establish a stream sound event, and thereafter subsequently only use the stream to convey the implication of a narrative event, the following survey showed that half of the respondents reflected on the stream sound events in their responses. Compared to the first prototypes’ survey where no participants mentioned the stream in their responses, this was significant. This is not to say others did not take note of the stream,
but the responses exemplified the preference towards the written as carrying the main narrative responsibility. However, this does show that content for the technical medium contain the potential to both establish and use the stream as more than the aforementioned mood-establishe and mood-enhancer most participants accredited it as. It should be noted that most participants only experienced one iteration of prototypes, meaning further exposure to the medium might have increased the perceptive ability of all components simply through prolonged experience of the medium. Further, prolonged exposure might provide participants to transcend the initial novelty of the medium, possibly providing more accurate result on a longitudinal scale. However, due to the study length and the obscurity of the technical medium alongside the limited access to such narrative presentations, this was considered an acknowledged shortcoming of the study.

7.2 The narrative presentation’s components’ relationship

The study focused on the presentation of fiction novels adapted, remediated, to the particular technical medium of the study. From the prestudy and concurrent findings, the components defining the studied medium (text and sound) together constitute a foundation for describing narrative presentations on several layers. Through the presence of the stream component, aural semiotics might be used and established in relation to what is written, thereby telling stories through a multimodal presentation. From this, it seems as though the stream itself posit an optional channel and tool for further establishing emotional, environmental, and cultural contexts present in a narrative.

Because of the temporal nature of the streams, and the self-regulated pace of the reading, there exist a distinct difference in the modalities consisting the narrative presentation. Because of this, stream designs in relation to the written ought to be operating on an undefined timeline, utilizing layers, loops, and sound effects to create a dynamic aural environment that is contextually appropriate. Situating oneself opposing this view, one might posit what happens when a stream is designed without connection to the written and further, how is the narrative perceived in such circumstances? As indicated through this study, a general, or generic enough, stream ought to simply flavor the text, in some way affecting how the written is imagined by the reader.

Further, the intent behind the stream usage itself arguably present different possible ways in how the narrative presentation (syuezhet) might be designed. In comparison with silent texts,
the stream might simply posit an addendum to the existing text (as exemplified by the booktracks). In contrast, a mentality to use the stream as an active tool to aurally explain actions or events as “this is how it sounded” differ on a fundamental level. These approaches exhibit similarities from the components that construe them, but also express differences in the purpose behind and the narrative allowance of the two components consisting the narrative presentation.

7.3 Designing for and defining the narrative’s presentation
As a consequence of working towards identifying stream design principles, the study had to tangle with designing for the narrative as whole as well. Because the inclination of participants, and researcher initially, to compare the narrative presentation and the technical medium with silent texts, a query on the subject was slowly but surely highlighted. Namely, what actually defines the medium? And to what is it comparable to?

From the presence of the text, a clear similarity to silent texts can be identified. Supported by the booktracks’ origins (being derived from existing novels), a presupposition of what the technical medium was, and might be, became presented and inferred. Given this notion, the narrative as presented through both text and sound would rarely expand beyond that of “a stream applied to an existing text”. In this context, the aural stream component would never constitute more than an addition or appendage to the written.

Inherently, the stream component carries the potential to expand the narrative, by extension of the written, and to some extent, through affecting the emotional state of the reader. However, given the ambiguous frame for what actually constitute the explored medium, one might argue that any presentation including the two modalities (text and sound) construe a version of the medium (i.e. listening to music while reading arguably is a possible form of the medium). In this sense, conscious narrative and stream design becomes superfluous if the goal is to simply present the novelty of the medium. While conscious design might help to contextualize the written and the overall narrative, a narratively independent stream ought to also, if allowed, impact the reader’s state of mind while reading. Simultaneously, as demonstrated through the study, the possibility of using the two modalities to further the narrative might be considered incentive enough to justify the multimodal presentation as independent in relation to silent texts (i.e. print books, e-books, etc.).
To elaborate further, an analogy will be described. Imagine a narrative being presented. This narrative might be remediated into several mediums, recreated and adjusted to the technical medium at hand. Inherently, each medium is construed by its components through which the presentation is perceived. What then happens if some components are not used? In essence, the main difference becomes simply that the presentation does not utilize the potential of the chosen technical medium. To solidify, the narrative might be presented through either a book or as a movie. However, while both might enlist textual descriptions, most likely the movie will be more dependent on pictorial and aural elements for presenting the narrative. While the medium of film is not defined by what it presents, the components available for presenting does define how one (usually) approaches working with the medium. In direct correlation to this study’s medium, this idea implies that narratives presented become part of this particular medium if they consist of both read text and heard sound, however the narrative design may regard the two components. This kind of idea indicate there existing a multitude of possible approaches to the design and the presentation of the narrative through these two modalities. To contextualize, the possible approaches to the narrative design might be placed on a scale describing the syuzhet balance between the two modalities presenting the narrative (see Figure 16, next page). In this sense, the examined booktracks and the first prototypes of this study might be situated on the far end of such a scale. Thereby, this situates their presentation as mainly requiring the textual component, whereas the stream then act as a supplement that may or may not be omitted entirely. During Part 3 of the study, the developed prototype was an attempt at working with the stream more actively in the narrative. Thereby, it might be considered to be situated more centrally (albeit not by much) on this scale. Finally, one might define a subject for further exploration by asking: what kind of narrative presentation situates itself on the opposite side of such a scale?
Following the same line of thought, one might elaborate further with: what differentiates this medium from others? And; what makes it unique in terms of components and presentation? In essence, these questions both find footing in the stream component’s presence, and thereby the inherent attributes and possibilities made available through it. In relation to the field of audionarratology (Mildorf & Kinzel, 2016), the purpose of the stream in the narrative ought to vary from one presentation to the next. In this sense, the booktracks’ streams held an overall purpose of establishing a mood and often used music to elicit action or a particular emotion in relation to the text. Similarly, the final prototype utilized the stream to mainly establish the aural environment present in the narrative (something that was also done in booktracks), with the main difference that the narrative design, fabula and syuzhet, regarded both the stream and text components at the ideating stage of the narrative development. As a result, the prototype became equally an attempt to create content with the technical medium’s modalities in mind, as it became a different take on exploring the stream purpose and how the narrative presentation was perceived by readers.

7.4 Stream design principles

As mentioned, the first and foremost quality of any stream design, is that of contextual intent. In this sense, a stream ought to enhance the narrative in a positive way as long as it makes sense for the reader of the medium. Therefore, the focus here will mainly regard the more technical and aural approach to the actual design of the stream, rather than the stream in relation to the narrative in context.

As previously mentioned, the stream’s presentation and consumption inherently differ from the textual component in that it is temporal at its core. Essentially, streams are designed to be
experienced over time. However, in its interrelationship with the textual component, the streams should not be dependent on a set timeline in their design. This sparks interest as the narrative itself evolves in a strict linear manner (considering the syuzhet as “set” from start to finish in the existence of the text), yet the stream needs to be designed to be, in large, non-linear. Rather, the stream needs to support what might be called a temporally undefined linear development. Essentially, the stream design follows the narrative context, yet also needs to support both rapid and slow consumption rates (i.e. reader’s pace) of the narrative.

The approach found in the booktracks and the learnings throughout the explorative parts has indicated that high fidelity streams were to be preferred over low fidelity ones. In essence, similarly to any general sound design approach (for movies and games ex.), only that which is relevant should be presented through the stream. This also ties into the cultural appropriations of designers and readers both. Purposeful design, being aware of the semantics and possible implications of stream sounds equally serve as risks and as powerful tools for situating and describing the narrative through the stream. A further issue highlighted throughout the study was the intrusion the stream had upon the reader’s imagination. While the notion of the stream itself was deemed interesting and immersive, the expressed intrusion both emphasize the similarity to silent reading and the potentially diminishing affect a poorly designed stream might have on a reader’s immersion. Indeed, several participants expressed concern for the stream’s aural intrusion into their imagination, indicating there is a delicate balance between what guides the imagination and what decides, or forces, it. Because of this, a more low-key stream design, meaning a less effortful to understand was preferred by several of the study participants, more indicative than descriptive. Further, this indicated the participants leaned towards, using Schafer’s term, a more embodied stream design (Schafer, 2005), requiring less effort to comprehend.

The stream design applied through the study has aimed at general, high fidelity, and recognizable aural environments. Striving towards designs not subjected to issues surrounding multimodal literacy (Hallet, 2014). This arguably generic approach never caused any participants to express confusion or query what the stream was representing. Further, the prototypes were all set in either recent contemporary or a medieval fantasy setting, meaning most sounds present might be identified in the everyday lives of participants. This choice was made to enable understanding of the stream, though it might therefore also be queried: what would happen when presented with more distant environment (in relation to the “present
world”), such as a futuristic science fiction setting or similar? Will then generic, high fidelity streams be as easily understood by the reader, or might the fictional nature of the setting itself require more dense low fidelity streams as a result?

| General guidelines                                                                 | 1. Contextually appropriate to the narrative  
|                                                                                   | 2. Consider cultural and semiotic implications of applied sounds both from the designers and the target audience’s perspectives.  
|                                                                                   | 3. Looping sounds constitute the core of ongoing streams.  
|                                                                                   | 4. Consider the stream fidelity, what purpose and what is needed from the stream content.  
|                                                                                   | 5. Ensure the audience literacy via either low-key generic and general ambient approaches, or through establishing the stream components through the narrative. |
| Sound type specific in this particular narrative presentation                     | 1. Ambiences are the foremost tool to expand the narrative, guide the mental visualization, and enhance the immersion of the reader.  
|                                                                                   | 2. Music and tonal components posit effective tools to affect the emotions of the reader and the mood of the narrative.  
|                                                                                   | 3. Sound effects might be presented to diversify the stream, represent written events, or emphasize aural attributes of audible events. Might be employed to signify both diegetic and non-diegetic events, both present within and outside of the presented narrative. |

*Table 2: Stream design principles.*
8. Conclusion

The study has engaged with the presentation, design, and perception of an obscure medium consisting of text and sound. Through the study, several themes have been used to guide the process and thereby answer the research questions.

In relation to the main research question (RQ1), it has been confirmed that the narrative is subjected to different interpretations through the perception of the stream in conjunction with the written text. Essentially, the stream design might affect how a reader visualizes the narrative in their mind. However, this assumes the readers both want to and are being able to comprehend what is presented through the stream. Further, the technical medium is proposed to be viewed as its own medium (such as movies or games), being similar but explicitly different in its presentation when compared to silent texts. Further, the two intertwined components indicate a component relationship more similar to the audio-visual of movies than the monomodality present in silent texts. Still, the latter has been the main subject for comparisons by both the researcher and participants through the study. Indeed, even at this stage the similarities to silent texts is used as a node to describe the studied medium, being composed of text with the added novelty of the stream.

However, the stream does enable an environment of presentation in which emotions, mood and general aural experience can be expressed to a reader. The technical aspects of the stream design were raised by study participants on several occasions and pose another area, alongside user interaction with the technical medium, to explore further. As mentioned, the narrative presentation explored throughout this study has regarded the approach and creation of fiction novels for and presented through both text and sound. Thereby, it was defined that for this particular instance of the medium, the stream design and purpose serve to establish the aural environment and further expand upon what is written. Through the final prototype (see Chapter 6), an approach to the overall narrative design confirmed that the stream’s presence has the potential to be used as a separate channel for presentation with impact and consequences for the overall perceived narrative.

In regards to the second research question (RQ2), it seems as though suitable guidelines would differ depending on where the narrative is situated on the scale previously mentioned and shown in Figure 16 (see page 66). However, some broad guidelines can be stated regardless of this.
Firstly (1), the stream should be contextually appropriate in relation to the overall narrative. Second (2), the cultural and semiotic meanings of enlisted sounds should be considered throughout, both from the designer’s perspective and the target audience’s. Third (3), looping sounds constitute the core of the inherently temporal stream component. They should be employed accordingly as stated in the first and second pointers. Fourth (4), consider the fidelity required and purpose of the applied sounds in the stream. Finally, fifth (5), ensure the literacy of the reader of the narrative through either a low-key ambient approach or by establishing stream components meaning through the medium itself.

In the case of this study, the stream design has regarded how fiction novels have been and might be presented through this particular technical medium. Therefore more specific stream guidelines have also been identified alongside the broader ones presented. Expanding upon the defined five requirements, a design approach to the three types of sounds (ambiences, music, and sound effects) in relation to this version of the narrative presentation has been defined. Firstly (1), ambiences are at the first and foremost efficient stream component to expand the written, guide the imagination and mental visualization, thereby enhancing the immersion of the presented story. Second (2), music and tonal components present efficient tools available on both diegetic and non-diegetic levels of the narrative. Further, due to the inherent semantic and emotional affect of music, it is suited to establish the mood of the narrative and further to describe emotions present in the story. However, it is vital to consider the third general requirement (using loops) for music, as composed music with a beginning and end might not correlate in duration to that of a reader’s pace. Still, depending on purpose in the narrative, this might not be an issue. Third (3), sound effects represent an aural cue for explaining particular attributes of both textually described events and established aural ones. Depending on the narrative’s position on the aforementioned scale, sound effects might be aural representations of written text, or act alongside the text and the overall narrative to further the story. However, the second option require the sound effect to be established either through the narrative itself, or through using the inherent semantic components of the sound itself.

As a final note. The study has dealt with the presentation of fiction narratives through a medium composed and reliant of textual and aural components for presenting its contents. Through the process, it has realized and revealed the multimodal presentation is not as cut and dry as might be assumed. Indeed, the narrative presentation explored here only constitute a particular niche on a larger scale inclusive of similar presentations and approaches of the medium. Further, it
has raised the cultural connotations connected to the stream components, the application of them, and the literacy of the reader of a narrative as intertwined components determining how the narrative is perceived. Building on this, the study dealt with the presentation of fiction novels and wants to serve as a first step towards understanding the possible approaches and application of this kind of narrative presentation in contemporary society. Take news articles, advertisements, and information brochures as a few examples of tactile areas where this multimodal presentation might be employed in the future. In a society where multimodal presentations are available through phones, tablets, and other devices, it seems plausible a multimodal presentation such as the one explored here might constitute the next step in narrative presentations.

8.1 Limitations

The findings of the study were drawn from a small set of the population (n < 25). Thereby, it is hard to state outright that the stream always impacts how the narrative is perceived and further that the medium’s novelty has a positive reception overall. Further, most participants only experienced one or two presentations of the multimodal medium, indicating prolonged exposure might have yielded different results alongside stronger validity. Moreover, the study employed a mixture of judgmental sampling on a convenience basis and pure convenience sampling. Because the medium in question was considered obscure, this was done to bring in experiences from related fields to construct an understanding of the medium. Simultaneously, it should be emphasized that the sampling was not meant to represent a larger population at the stage of the study, though some steps were taken to reduce potential aural bias in the results.

Due to the examination of the booktracks in Part 1, the study was shaped surrounding only the narrative presentation when composite of and presented through text and sound. As a consequence, comparisons to established media (plural medium) such as audiobooks, radio plays, movies, and video games where similar narrative presentations exist were largely omitted through the study. Another approach might have incorporated Elleström’s definition of multimodality and used it to discuss the modes and modality of the medium (Elleström, 2010, p. 36) and thereby build a greater understanding of the medium itself (for an example of a similar study using this approach, see Have & Pedersen, 2015).

The study was further focused on the overall perception of the medium, leaning towards the aural aspect and the design of the stream. However, as the medium in question consisted of two
main components (text and sound), the stream focus does create a situation in which the text’s impact, structure and presentation is neglected to some extent. Due to the study’s goal and the researcher’s current lack of linguistic knowledge (discourse, terminology, etc.) this was deemed an issue better left untampered with at this stage. Indeed, the multimodal narrative presentation might have been further defined if text linguistics was included in the analysis and examination.

Further, as indicated in the discussion, some texts might be suited to different applications in the medium and even within the “genre” explored (fiction novels). To exemplify, both author writing style and the “focalization” (Abbot, 2005) of the written narrative constitute tangible features providing another layer for defining the medium. Further, the disposition between written and stream descriptions of the aural environment was not considered either. However, as a way to combat this possible issue within the scope of the study, the gathered text’s used different authors for both the examinations (Part 1) and prototypes (Part 2).

In the study’s prototypes’ streams, music was in large omitted. Based on the Part 1 learnings, this was a conscious choice but it should be mentioned that extended usage of music in the medium might yield different results. Further, as mentioned on linguistics and texts, the balance of the stream’s internal components (sound effects, ambiences, music) might define yet another level of analysis which was not conducted in this study.

8.2 Future studies

Throughout and post-conclusion of the study, three distinct future steps was identified. The first regards the interaction with the technical medium and expands upon the notion of automated stream playback in the medium found in the used dependent sing-along tracker of the booktracks. Namely, the accessibility of eye tracking technology poses the option to explore how readers perceive a medium in which streams actually follow their reading pace. Further, this implies a development of a prototype which also might be used to explore user interaction and stream customization in the narrative presentation.

The second regards the choice of “original” texts used. Through this study, fiction novels or various authors was used to examine the core of the medium. As mentioned, the writing style, stream and text disposition pose tangible subjects for further studies. Alongside these, the choice of text and purpose of the text itself posit further exploration. In essence, how is the medium perceived when applied to advertisements, newsletters, academic literature, etc.?
Further, this also builds towards defining the overall medium itself, as they likely vary in their placement on the scale presented in Figure 16 (see page 66).

Thirdly, the occurrence of reading difficulties, dyslexia, posit another venue to explore. What stream design guidelines are suitable in such circumstances, do they differ from the ones established here? And if they do, why? Further, on the multimodal components allowance; does the disposition change, or might it be changed from text to sound to ease reading? Bringing this back to the first future step, how does eye tracking impact the reading experience, and how does it correlate to the stream design and perception of the readers?

On the topic of the text and sound relationship, the notation system developed and used in the study is in need of further testing, evaluation, and development. Through the study, it proved its potential in distilling the design of the read and the heard into a visual overview of the two components relationship. With potential for improvement, the notation system might pose a valuable tool for the design and understanding of similarly presented narratives. Further, because the system was developed in the way it was, it could be transferred and adapted into a digital notation tool, inclusive of written classifications of sounds (which was omitted during the study) and timestamps, for example. Further, it might be incorporated into a spectrogram visualization of a stream alongside the textual component as well. Thereby, a digitization of the system might provide a more insightful and even less obtrusive way of notating an ongoing aural environment.

Finally, this study wanted to serve as a starting point with its foothold in design and perception, striving towards future studies on this particular, and similar, subjects. The findings might therefore act as a node for both further exploration in larger projects and as inspiration for similarly sized ones. Overall, the development, perception and possible applications of a multimodal narrative presentation in daily life warrants future studies. Whether from social, cultural, technological, or economical perspectives, the applications and implications of such a presentation in society posit a subject to be explored further.
References


Appendicies

Appendix A: Stream notation
Because of the obscurity of the medium, a way of notating the stream components was devised by the researcher. Using the theory of remediation, soundscape classifications, and several sound design concepts, the narrative acted as a node for examining how the stream was designed in relation to the text. This was intended to provide an empirical basis and a truthful depiction of the existing booktracks’ stream design. The purpose behind the notation system was to create an unobtrusive way to note a stream’s aural components alongside reading a booktrack. Because the stream inherently is temporal and that the booktracks are dependent on the reader following a particular pace, this system was designed to be used during “live” booktrack streams. As one might argue, recording the streams does enable an examination at one’s own pace after reading a booktrack. However, unless recording both the text and sound this would disconnect the intertwined components of text and sound, which in large were dependent on one another to make sense in the context of the narrative. Therefore, a “live” notation was considered the most efficient and authentic way to capture the stream’s purpose in relation to the textual component while subsequently increasing the researcher’s understanding of the medium itself.

The used notation form can be found on the next page. The usage of it is described in section 3.4 (Stream notation and examination) in the thesis.
Figure 17: Notation form
Appendix B: Consent form

STUDY SUMMARY
This study aims to explore the medium of soundtracked e-books to identify what practices for design exist and how the audio correlate to literal remediation of a written narrative. Further, it wants to investigate how the medium is perceived and if there are applicable approaches and possible best practices for the interplay between audio and text not found in existing material. The data gathered will consist of recordings (audio) and notes taken by the researcher. On the following page you will be asked to give your consent for this. If you cannot, you do not need to give an explanation.

FORESEEABLE RISKS AND DISCOMFORT
There are no foreseeable risks or discomforts associated with this study. Should you nonetheless feel discomfort or decide that you do not want to participate in the study (interview or otherwise), you are free to revoke your consent and initiate the termination of your participation.

CONSENT AND DATA TREATMENT
You are free to withdraw your consent for usage at any time for parts, or the entirety, of the data gathered through your participation. You do not need to give an explanation for withdrawal. Further, if there are any errors or misinterpretations within the data or presentation of the data, you have the right to have these errors corrected. Should there arise conflicts regarding the data collected through your participation, or if you have any qualms with the researcher’s treatment of the data, you are encouraged to contact the Data Protection Officer at Mau (dataskyddsombud@mau.se). Further, you have the right to contact Datainspektionen in Sweden regarding your data treatment, should the need arise.

BENEFITS OF PARTICIPATION
There is no monetary or otherwise material reward for participating in this study. However, I hope that you will find it rewarding to participate in pushing the knowledge barrier of this particular field of research and help me understand and explore current practices and approaches towards soundtracked e-books.

CONFIDENTIALITY AND DATA PROCESSING
The collected data will remain confidential and will be processed by the researcher alone. Digital data will be stored on a secure device protected by password. Storage of the data will be active until the thesis, of which this study is part, is passed and published- at which point all data gathered will be deleted permanently. Articles, reports and presentations by the researcher on this study will not use participants’ actual names. Further, records that identify the participants will not be publicly accessible.

CONTACT PERSON
If you have any questions about this research, call or write:
Xxxxxx xxxxxx
Mobile: xxx-xxxxxxx
xxxxxx@mail.com

VOLUNTARY PARTICIPATION/ WITHDRAWAL
Research volunteers have rights. Taking part in this research project is completely voluntary. If volunteers choose to not take part, they will have no penalty.

SIGNATURES
Volunteers will receive a signed copy of this form.

MONTH DD, YYY
CONSENT FORM

Name of Researcher: xxxxxxx xxxxxx

Please mark box if you agree

1. I confirm that I have read and understood the information sheet dated MONTH DD, YYYY for the above study. I have had the opportunity to consider the information, ask question and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without need to give any reason.

3. I understand that any information given by me may be used in future reports, articles, or presentations by the researcher.

4. I understand that my name will not appear in any reports, articles, or presentations.

5. I agree to take part in the above study.

6. I agree to the interview / focus group / consultation being audio recorded.

________________________________________  __________________________  __________________________
Name of Participant  Date  Signature

Xxxxxxx xxxxxxx  __________________________  __________________________
Researcher  Date  Signature

MONTH DD, YYYY
Appendix C: Expert semi structured interview questions

The two semi structured interviews conducted with experts structure is presented here. The first one also served as a basis for the focus group interview.

First expert interview questions (Part 1)

Demographics and background
- Age and Gender?
- Current occupation?
- Background regarding audio, music, and similar throughout different mediums?

Previous knowledge in regard to the research question
- Do you read books/e-books or listen to audiobooks normally? When, where and why?
- Had you experienced any mediums similar to soundtracked e-books before?
- What were your initial reaction to using the medium presented through Booktrack? Focus being on the actual ‘reading’ not the app in general.
- What booktracks did you ‘read’?

User experience regarding aural perception
- How did you perceive the general application of sound in the booktrack(s)? Was there something that stood out in particular that you remember? Was it something good or bad? Why did it stand out?
- How did you perceive the general quality of the sound in the booktrack(s)?
- How did you perceive the audio alongside the written narrative? Enhance/Impede? Was there circumstances where application of sounds were better or worse? Did it distract of draw your attention from the narrative? Why?
- What purpose do you think the booktrack team had when applying their sounds? If divided into categories of sound effects, ambiances and music?
- Did you ever experience silence when reading?
- Do you have any thoughts on the ambiances you heard?
- Do you think textual cues are necessary for aural sense-making in this medium? Why/why not?
- Do you have a preference to what purpose sounds might have in this medium?
- Did you ever experience listening fatigue when reading? If so, what circumstance?
- Do you think certain sounds are more suited for aural retelling in a medium like this? Why?

User experience regarding user interface and interaction
- What device were you using to experience the booktrack application? What kind of listening (headphones, speakers, etc.) did you use? Why?
- Did you use visible or invisible ‘sing-along’ tracker when reading? Why (not)?
- Do you think this choice may have affected your perception of the medium?
- Do you think familiarity with the medium would help comprehension?

Other thoughts
- Any other thoughts on the medium?
- Generally, what was good/bad in the experience overall?
Second expert interview questions (Part 3)

Intro
- Do you still consent to participating in the study?
- You are free to withdraw any and all participation at any time.
- Have you read or used any booktrack related thing since last we met?
- Have you read the prototype?
- Do you think the sound had a more or less impact for the story in this version, compared to the booktracks?

General perception
- Was there something that stood out in particular that you remember?
- How did you perceive the general quality of the sound in the booktrack(s)?
- How did you perceive the audio in the prototype? Crossfades, levels, other?
- Page changes, crossfades?
- Was there circumstances where you thought the sound had a narrative purpose?

Other thoughts
- Do you think including the stream in the narrative design as an active tool justifies it as a medium of its own, something different to regular books or audiobooks?
- The prototype has been designed with the stream in mind alongside the writing of the text. A purpose for this was to break away from the traditional book narrative. What are your thoughts on the prototype narrative presentation overall?
- Generally, what was good/bad in the experience overall?
- Any other thoughts?
Appendix D: Opinion survey questions

The two surveys’ questions employed are presented hereafter.

First survey (Part 2)

Age:
- [Fill in form]

Gender:
- Male
- Female
- Other

Do you read books or e-books? If so, how often and in what circumstance?
- [Fill in form]

Do you listen to audiobooks? If so, how often and in what circumstance?
- [Fill in form]

Do you consume other kinds of media regularly? [checking several possible]
- Music
- Podcasts
- Movies/Series
- Video games
- Radio plays
- Other [Fill in]

Which story did you read now?
- Study Excerpt 2
- Study Excerpt 3

On what device did you read the story?
- Phone
- Tablet
- Computer

What kind of listening did you use?
- Headphones/Headset
- External monitors/speakers
- Built in (internal) speakers of device

The story plays out in two main areas, did you find them easy to differentiate?
- Yes
- No

Can you write a short summary of the story and key details that you remember?
- [Fill in form]

If you think of the environments present in the story briefly, what do you remember/imagine visually and how would you describe them? (Added prior to the last 4 responses to see if a more leading question raised the stream in the narrative compared to the previous summary question. Added mainly to gain an understanding for how to phrase the second survey question: Towards the end... etc.)
- [Fill in form]

Do you think the sound affected your perception of the story?
- Scale 1-5 [Not at all - Very much]

How did you experience the sound quality?
How did you experience the sound playback?
- Scale 1-5 [Very bad - Very good]

Did you experience anything in particular that you thought was good or bad? Anything that stood out that you remember in particular?
- [Fill in form]

What did you think of the presentation of the medium itself? (i.e. the way you interacted with the text, the layout, etc.)
- [Fill in form]

Do you think you would use this kind of medium in the future?
- Yes
- No
- Maybe

Why, why not? In what circumstance?
- [Fill in form]

Anything you would like to add (on the story, medium, or this survey)?
- [Fill in form]
Second survey (Part 3)

Age:
- [Fill in form]

Gender:
- Male
- Female
- Prefer not to say
- Other [fill in]

Country of residence:
- [Fill in form]

What kind of listening device did you use?
- Headphones
- External speakers/monitors
- Built in speakers (phone, tablet etc.)
- Other [fill in]

The story was about… *(filtering question)*
- ...a young man on a hunting trip who watches a village in the distance while contemplating about cicadas. A travelling tanner was robbed and shot by some bandits.
- ...a girl who falls in love with the tanner in her village. He is being bullied by the village's hunter, who in turn courts the girl of the story.
- ...a craftsman selling leather trinkets during a festival. He then meets a friend in the forest after being forced to leave his home due to some outsiders.
- ...a lady hunter who sells pelts and fresh meat at a festival. There she encounters a friend and they discuss the value of good fur.

Which of the following best capture your perception of the story?
- a) I found the story easy to follow and think the text was the main reason for this.
- b) I found the story easy to follow and think the sound was the main reason for this.
- c) Both A and B together.
- d) I found the story hard to follow and think the text was the main cause.
- e) I found the story hard to follow and think the sound was the main cause.
- f) Both D and E together.

Which of the following best capture your general perception of the sound in the medium?
- a) I think the sound improved the narrative.
- b) I think the sound was an interesting addition, but it was not needed.
- c) I do not think the sound added anything, but it did not ruin the narrative either.
- d) I think the sound distracted and impeded my perception of the story.

How interesting did you find the story to read?
- Scale 1-4 [Not at all - Very much]

Do you think the sound increased or decreased your level of immersion in the story?
- Scale 1-4 [Decrease - Increase]

Do you think the sound affected how you visualized the story when reading?
- Scale 1-4 [Not at all - Very much]

Did the sound affect how you perceived the mood throughout the story?
- Scale 1-4 [Not at all - Very much]

How likely are you to read stories in a similar fashion in the future?
- Scale 1-4 [Not likely - Very likely]

Towards the end, the protagonist wakes up and leaves. Why?
- [Fill in form]
How would you improve the presentation of the story?
   - [Fill in form]

Anything you would like to add?
   - [Fill in form]
Appendix E: Prototype links

The study’s prototypes were published and tested through Booktrack Studio (Booktrack, 2019b). However, only the final prototype (Part 3) was made public. It was named “The tanner and the hunter” as it generally described the story. The final prototype can be found and read through either a web browser or the Booktrack application. The only prerequisites to experience it is a device (phone, tablet, or computer, etc.) and some kind of listening (headphones, monitors, internal speakers, etc.).

The cover illustrates the opening scene of the story.

Because the browser-based version of the Booktrack Studio was disabled during the revision of the thesis, the links below are faulty but kept in the hope they will be viable once (if) the Booktrack Studio becomes available again.

Therefore, the only way to access the prototype is through the downloadable application “Booktrack”. Search for “the tanner and the hunter” (with quotation marks) among the community created stories (likely, this will only be an option once a “normal” search has been conducted, as the app them prompts for continued search among the community content). The first prototypes are not accessible through the application.

The QR code and link below leads browser version of the final prototype, as presented and used in the study. Below these the links to the first prototypes (Part 2) are included. These two are only accessible through these links.

Final: https://www.booktrack.com/content/read/9d67ecb72d634dd5a9b5430249cd10bc

First Prototypes:
Original interpretation: https://booktrack.com/content/read/81e675988249464fabf3436c040055f6
Alternate version: https://booktrack.com/content/read/048cb0411f1542afbd180c6d9e29e6f2