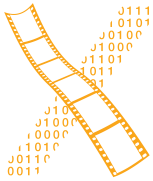

Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

SoundMuse: Contextualized Sound Cues for Creative Writing

Bachelor's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Nestan Gujejiani

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 17.12.2024
Submission date: 17.04.2025

Contents

List of Figures and Tables

Abstract

Creativity plays a central role in our lives, enabling us to develop new ideas, solve problems and express ourselves. In recent years, AI (Artificial Intelligence) has been increasingly used to support creative processes. Especially in collaborative systems, where AI can act as a subtle partner that complements, rather than replaces, human creativity.

This paper focuses on creative writing, a process that is often hindered by challenges such as concentration difficulties or writer's block. Previous research has shown that music and ambient sounds can positively influence writing. However, such effects are not always consistent - sounds can be distracting or fail to set the desired mood. This raises the question of whether soundscape that is dynamically tailored to the written content could improve the writing experience.

To explore this idea, we developed the SoundMuse prototype, a system that uses AI to generate and play context-specific auditory cues in real time. By analyzing the evolving text, the system continuously extracts semantic and emotional cues and translates them into relevant audio feedback. The system uses OpenAI's GPT-4 to detect such cues and generate corresponding tags, which are then used to retrieve fitting audios from existing databases.

We conducted a small preliminary study to investigate how such adaptive soundscapes affect the writing process. Initial results suggest that if the audio transitions are smooth and thematically appropriate, such systems can promote immersion and creative flow. At the same time, the study underlines the importance of a considered design to avoid distractions.

This work contributes to the field of human-computer interaction by exploring how AI-driven, context-aware auditory support can enrich creative writing process.

Überblick

Kreativität spielt eine zentrale Rolle in unserem Alltag. Sie ermöglicht es uns, neue Ideen zu entwickeln, Probleme zu lösen und uns auszudrücken. In den letzten Jahren wird KI (Künstliche Intelligenz) zunehmend zur Unterstützung kreativer Prozesse eingesetzt. Vor allem kollaborative Systeme zeigen dabei Potenzial, als subtile Partner menschliche Kreativität zu erweitern, statt sie zu ersetzen.

Diese Arbeit konzentriert sich auf kreatives Schreiben — einen Prozess, der oft durch Herausforderungen wie Konzentrationsschwierigkeiten oder Schreibblockaden verhindert wird. Frühere Untersuchungen zeigen, dass Musik und Klänge das Schreiben positiv beeinflussen können. Solche Effekte sind jedoch nicht immer konsistent - Klänge können ablenken oder die gewünschte Stimmung verfehlen. Dies wirft die Frage auf, ob eine dynamisch an den geschriebenen Inhalt angepasste Geräuschkulisse das Schreiberlebnis verbessern könnte.

Um diese Frage zu untersuchen, haben wir das SoundMuse-Prototyp entwickelt, ein System, das mithilfe von KI kontextspezifische akustische Reize in Echtzeit erzeugt und abspielt. Durch die Analyse des entstehenden Textes extrahiert das System kontinuierlich inhaltliche und emotionale Hinweise und interpretiert sie in passende Audioimpulse. Das System nutzt OpenAIs GPT-4, um solche Hinweise zu erkennen und entsprechende Tags zu generieren, die dann verwendet werden, um passende Audios aus bestehenden Datenbanken abzurufen.

Wir haben eine kleine Vorstudie durchgeführt, um zu untersuchen, wie sich adaptive Soundscapes auf kreatives Schreiben auswirken. Ergebnisse deuten darauf hin, dass solche Systeme Immersion und kreativen Fluss fördern können, wenn Audioübergänge flüssig und thematisch passend sind. Die Studie zeigt zudem die Bedeutung durchdachter Systemgestaltung zur Vermeidung negativer Effekte.

Diese Arbeit trägt zum Bereich der Mensch-Computer-Interaktion bei, indem sie untersucht, wie KI-gesteuerte, kontextbewusste auditive Unterstützung kreative Schreibprozesse bereichern kann.

Acknowledgments

I would like to thank Prof. Dr. Jan Borchers and Prof. Dr. Ulrik Schroeder for examining my thesis.

I am truly grateful to Paul Preuschoff for the committed and motivational supervision of my thesis.

A big thank you to all the people of i10 for always being helpful, answering my questions, and providing thoughtful suggestions along the way.

Finally, I am very grateful to my friends and family for their general support, encouragement, and patience throughout this journey.

Conventions

Throughout this thesis we use the following conventions:

The thesis is written in American English.

The first person is written in the plural form.

Unidentified third persons are referred to in the plural form.

Numbers are written as figures for values 10 and above, and spelled out in words for values from zero through nine.

DEFINITION:
Definitions are set off in orange boxes.

Chapter 1

Introduction

1.1 Creativity and collaborative AI

AI can assist creative processes without replacing them.

Creativity is a central part of human experience and plays a very important role in many areas of our lives, such as art, science [Kirsch et al., 2016], and everyday life [Putney et al., 2024]. It enables us to generate new ideas, innovate, and express ourselves uniquely. The ability to develop new ideas is important for both social progress and individual expression [Villalba, 2008]. Traditionally, creativity has been an exclusively human trait, but in recent years, artificial intelligence (AI) has developed as a powerful tool that can support creative processes [O'Toole and Ágnes Horvát, 2024].

Definitnion: Creativity

CREATIVITY:
Creativity refers to the human ability to develop ideas or artifacts that are both novel, whether personal or historical, and valuable in different ways [Boden, 2009].

Differentiation between autonomous and collaborative AI

In AI-assisted creativity, one can distinct between autonomous systems, which perform tasks completely independently, and collaborative systems, which act as assistants and accompany the creative process. The integration of collaborative systems, in particular into creative pro-

cesses, open new possibilities for different fields, as they do not take over the task but support the creative process.

AI has the potential to help overcome writing challenges.

In this thesis, we focus on creative writing as a form of creativity. Creative writing is a complex process that is usually characterized by the author's personal experience and inner thoughts. Writers often struggle with challenges such as writer's block, lack of motivation, or difficulties in maintaining focus [Baverstock and Steinitz, 2019]. This is where AI can offer an interesting opportunity to support creative work.

1.2 Music and Ambient Sounds in Creativity

Music can promote creative processes.

One such supportive element is music, which plays a central role in most people's lives. Many people not only listen to music passively but also use it actively to support creative processes, especially when writing. There are research results that indicate that music can promote creative processes such as design [Zhou et al., 2020], creative cognition [Eskine et al., 2018], or creativity in the workplace [Wu, 2024]. In creative writing in particular, music could help to create an immersive atmosphere, favor the flow state, increase concentration and productivity, or even release writer's block.

Ambient sound can promote creative processes.

In addition to music, ambient sounds have been shown to positively impact creativity. Studies indicate that background ambient sounds can enhance creative performance, such as generating more ideas in creative tasks [Mones and Massonnié, 2022]. However, the type of noise also plays a significant role, with some research suggesting that it may not always have a positive impact on creativity [Awada et al., 2022]. This suggests that a carefully tuned auditory environment can optimally stimulate the creative process.

Little research exists on adaptive music and sounds for writing.

Even though the influence of music and ambient sounds on creative writing has already been investigated, there has been little research on how dynamically generated context-

sensitive music and sounds can influence this process. To address this, our work attempts to fill the gap by investigating how music and sounds that dynamically adapt to the written content can support the creative writing process. Static playlists or random background tracks may not always align with the evolving context of a narrative, or they do not correspond to the emotional and thematical shifts in a text. This is why exploring how adaptive auditory cues can enhance the creative process might be worthwhile.

Scope: Exploring AI-driven adaptive soundscapes for creative writing

Building on this, this thesis aims to explore the intersection of creativity and AI, specifically focusing on creative writing. This work explores how AI can support the writing process through dynamically generated and context-sensitive soundscapes. The scope of this thesis is centered around the design, development, and iterative refinement of a prototype that integrates these soundscapes into the writing process.

Findings: Success in system adaptation, while refinement needed

The iterative user studies revealed that the system successfully responds to narrative context. However, the studies also identified areas where the soundscapes need better alignment with the text's emotional tone, indicating a need for further refinement.

Findings: Contextualized soundscapes can enhance the writing experience.

The developed prototype was additionally experimentally explored in a small-scale study, providing early insights into its effectiveness. The study suggests that dynamically adapting soundscapes can create an immersive atmosphere that enhances the writing experience and fosters a deeper emotional connection to the text. Furthermore, the contextualized soundscapes have the potential to inspire creativity through the diverse auditory stimuli they provide.

1.3 SoundMuse

SoundMuse as a prototype for AI-driven audio cueing while writing

A core contribution of this thesis is the development of SoundMuse, an AI-driven system that reacts to the user's writing and plays appropriate music and sounds. SoundMuse extends the scope of existing approaches, as it not only passively provides music and sound but responds in

real-time to the dynamics of the writing process by adapting the soundscape to the writing topic and the corresponding scenes. How SoundMuse is designed is explained in detail in Chapter 3.

1.4 Outline

This thesis begins with a review of existing literature on creativity and collaborative AI, as well as the influence of music and ambient sounds on creative processes, especially creative writing, and the existing work on dynamic auditory cues in digital environments (Chapter 2).

Introduction of SoundMuse, its implementation details, and development phases

In Chapter 3, we introduce the developed prototype *SoundMuse*. We outline the motivation, summarize insights from an expert interview, and give a comprehensive overview of the system. The implementation is described in detail, covering the underlying logic for audio selection and playback. Additionally, we present the iterative user studies (Section 3.5), which were conducted in several phases. While the initial studies focus on assessing technical stability, the later studies explore the impact of the generated soundscapes on the creative writing process. After describing the methodologies, we present the results and key insights.

Introduction of the preliminary user study

Chapter 4 presents a small-scale preliminary user study investigating how dynamically generated soundscapes influence the creative writing experience. We describe their methodology and findings in detail.

The thesis concludes with a discussion of findings, suggestions for improvement, and perspectives for future research.

Chapter 2

Related Work

This chapter presents the current state of research, bringing together relevant work from different research areas to provide a foundation for this thesis. Firstly, a review of work on music, as well as ambient sounds in creativity is considered, followed by a discussion of the use of AI in creativity and existing examples of dynamic music in digital environments. Finally, existing research gaps are identified and addressed in this thesis.

2.1 Creativity and Collaborative AI

As mentioned in Chapter 1, creativity is an essential part of our lives, and AI offers exciting opportunities to foster it. In this context, we are looking primarily at collaborative AI, which acts not as a substitute but as a supporting element. AI can improve the generation of ideas in group processes. AI-supported brainstorming tools, such as interactive tables or large-format displays, facilitate the structuring and development of ideas without losing the advantages of direct interpersonal communication [Hilliges et al., 2007]. In close collaboration with humans, AI can also be used in so-called mixed-initiative interfaces, where it makes suggestions, and produces, evaluates, and modifies creative content [Deterding et al., 2017].

AI enhances group creativity and helps boost creative learning.

The potential of AI-supported tools is also evident in the area of collaborative idea generation, where AI analyzes social interactions and supports groups in transforming individual ideas into coherent collaborative concepts [Shin et al., 2023]. In addition, AI is increasingly being used in education, where it promotes creative processes by enabling new learning methods and supporting creative skills [Marrone et al., 2024].

AI can accelerate creative workflows without loss in quality.

The importance of AI for creative tasks can also be seen in specific applications, such as the automated editing of film trailers. Studies have shown that AI can speed up the creative process by reducing time and effort without compromising the quality of the results [Smith et al., 2017].

Thesis focus: dynamic soundscape as creative support while writing

These diverse applications suggest that AI can support creative processes and even permanently change the way people develop and implement ideas together. This thesis specifically investigates how AI can dynamically interact with human creativity in the context of writing, using music to foster the creative work.

2.2 Music and Ambient Sounds in Creativity

As mentioned in Chapter 1, there is a handful of literature on how music and ambient sounds can affect creative work in general. Zhou et al. [2020] found that fast music improves the creative thinking of designers, especially their fluency and flexibility. Compared to conditions without music, fast music led to significantly better results in creative design tasks, while slow music did not have a similar positive effect. In their study, Eskine et al. [2018] further showed that listening to music can influence creative cognition, this effect being explained by the stimulation of mood and arousal. Referring to Wu [2024], the use of music in a working environment significantly increases the creativity of employees on both a cognitive and emotional level.

Creativity is not only affected by music but also by the broader acoustic environment. Research suggests that ambient sound can affect creative processes in different ways. Moderate levels of environmental sounds have been shown to enhance creative performance by increasing cognitive processing difficulty, which promotes abstract thinking, while very low or very high sound levels tend to hinder creativity Mehta et al. [2012]. Additionally, exposure to ambient sound can boost idea generation in creative tasks, particularly for individuals with high cognitive flexibility, though it does not necessarily improve originality Mones and Massonnié [2022]. White noise, depending on its volume, has also been found to impact cognitive performance and creativity, with lower levels enhancing sustained attention and reducing stress, while higher levels improve working memory but increase stress Awada et al. [2022]. These findings highlight the complex role of ambient sound in shaping creative performance.

2.3 Music and Ambient Sounds in Writing

Soundscapes shape writing mood, focus, and flow.

Beyond general creativity, research has also examined how auditory stimuli affect the writing process. Various studies show how auditory stimuli can influence the writing process and the associated cognitive processes.

Different sound types have various effects on writing.

Rowell and Flick [2019] investigated the effects of three different soundscapes — silence, ambient noise, and music — on the writing process. Participants were asked to complete writing tasks in these environments. Surveys conducted before and after the task provided information on how they perceived the impact of soundscapes on their cognitive and creative processes. The results showed that sound, or the absence of it, had a significant impact on writing performance. The authors reported that sounds often distracted them, causing them to lose their train of thought and changing their general mood. Above all, ambient noise was perceived as particularly distracting. The study makes it clear that different acoustic accompaniment can have a significant impact on writing. This suggests that the auditory

cues shape the writing experience but can also present challenges depending on the person and context.

Acoustic feedback can enhance writing fluency and perception

Kim et al. [2012] analyzed the effect of acoustic feedback (monaural and stereo) compared to no feedback during writing, specifically tasks that involved tracing Chinese characters. It showed that the use of musical feedback or other acoustic signals led to a higher number of characters and left fewer negative impressions overall during the writing process. This indicates that acoustic stimuli can not only stimulate writing but can also positively influence the perception of the process.

Research highlights the potential of sensory stimuli in supporting creative writing and raises important questions about the role of auditory input in isolation. Gonçalves et al. [2017] investigated how the combination of olfactory and auditory stimuli supported the creative writing process. The results showed that this multi-sensory input promoted creativity significantly more than conditions in which no stimuli were used. It was particularly interesting that the effect of the stimuli was significant for both calming and alarming cues. However, there was no similar effect when only auditory stimuli were used, which raised the question of whether music alone can enhance the creative process or whether the effect is rather triggered by a combination of different sensory perceptions.

2.4 Environmental Cues and Creative Writing

Impact of environmental cues, priming, and conformity on creativity

In their study, Nelson and Guegan [2019] do not directly analyze the influence of auditory cues on writing but rather explore the impact of other environmental cues on creative production, delivering interesting outcomes. In their work, the authors discuss the conformity effect, which suggests that people tend to replicate noticeable features in examples of creative solutions, even when they aim to be original. Additionally, they highlight priming, which refers to the process where exposure to one stimulus unconsciously ac-

tivates related concepts or knowledge, making them more accessible for use in later, unrelated contexts, suggesting that this can influence individual behavior or thought patterns.

Findings from the study indicate that:

- Participants exposed to environments without creatures produced more drawings than those in environments with creatures.
- Exposure to specific environmental elements (like creatures) led participants to create works that were less original but aligned with the environment they were exposed to.

Priming and conformity effects influence creativity

The results confirm that priming and conformity effects played a role in influencing the participants' creative output, underlining the significant impact that contextual elements, such as the environment or visual stimuli, can have on the creativity and originality of individuals. These findings raise the question of whether auditory stimuli might similarly prime creative processes or reinforce certain patterns in writing.

2.5 Dynamic Musical Cues in Creative Environments

Dynamic soundscapes in creative processes remain underexplored.

Exploring research on dynamic soundscapes and their influence on creative processes, we found a study that examined context-based music composition for tabletop role-playing games [Ferreira and Whitehead, 2021]. The examination showed that music in short video formats was able to enrich the gaming experience by enhancing scene transitions and aligning with participants' preferences. The participants in the study preferred the system's choices, indicating that the generated soundscape was more effective in complementing the scenes. However, questions remained as to how such systems behave during longer sessions and whether the positive effect can be maintained over longer

periods of time. The study highlights the need for further research into the interplay between auditory cues and context, particularly in prolonged creative processes, to understand how soundscapes can be sustainably integrated into creative environments.

In summary, previous research shows that music and ambient sounds can influence creative writing. However, the effect of context-sensitive, dynamically generated soundscapes during creative writing remains largely unexplored. In this work we aim to explore this largely unexamined area.

Chapter 3

SoundMuse

This chapter describes the development of the SoundMuse prototype. It covers the core components of the system, the details of the system implementation, and important design decisions during the development process. The work evolved through several stages, beginning with an expert interview, followed by the initial system design. The development process then progressed through an iterative implementation cycle, as illustrated in Figure 3.1.

3.1 Motivation and Purpose

As mentioned in previous chapters, this thesis focuses on exploring the effects of dynamic and contextualized auditory cues on creative writing. As there has been no previous work that has directly addressed this topic, we have developed SoundMuse as a prototype to facilitate research in this area.

SoundMuse: Adaptive audio environment for creative writing

The core idea behind SoundMuse is to provide an adaptive audio environment that responds to the content being written, that does not require manual selection of audio tracks, and aims to create a seamless auditory background that adapts to the writer's narrative. SoundMuse is primarily intended to serve as an experimental system to explore how

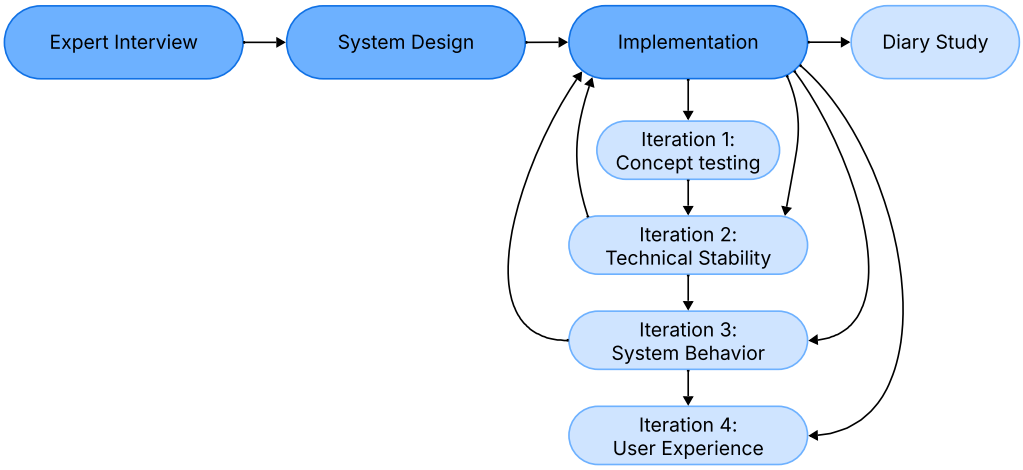

Figure 3.1: The key milestones of the SoundMuse development process, including expert interviews, system design, iterative implementation process, exploratory user study, and the diary study. The figure illustrates the methodical structure of prototype development and the thesis work around it.

contextualized auditory cues influence the creative writing process. By integrating automated auditory cueing, we can analyze how writers respond to audio stimuli associated with the dynamics of writing and whether certain soundscapes help or hinder their creative process. SoundMuse also provides a basis for further research in the field of HCI (Human-Computer Interaction), in particular, to understand how adaptive real-time feedback mechanisms can support creative tasks.

3.2 Expert Interview

Expert interview: Getting insights before implementation

To better understand how auditory stimuli can support the creative writing process, we decided to conduct an expert interview with a professional writer before the implementation process. The interview aimed to explore the role of music and sounds in the writing process and refine the design of the prototype to better suit writers' needs. An interview guide was designed as the basis for the approximately

60-minute conversation (see Appendix A). The interview was recorded, transcribed, and analyzed.

Insights: Distinction between creative and revision phases, auditory cues can aid writer's block; Personalization should be considered

One insight the interviewee described was the distinction between two writing phases: a spontaneous, creative phase and an analytical revision phase. This led to the realization that SoundMuse can be particularly supportive in the first phase of the writing process, as creativity is the main focus here. The interviewee also noted that auditory stimuli could help overcome writer's block by altering the atmosphere and providing novel creative impulses. The interview also suggested that individual preferences play a significant role, as the expert favored classical music as background, emphasizing the need for personalization.

Study design propositions: productivity, mood, motivation, user experience

In addition to the implementation, propositions were given on study design. As creativity and writing quality are difficult to measure objectively, the interviewee suggested evaluating productivity by comparing writing performance and gaining qualitative insights into mood, motivation, and user experience through self-assessments and open feedback.

3.3 System Overview

SoundMuse is a modular system that consists of four core components: the *text editor*, the *text analysis module*, the *audio retrieval module*, and the *playback mechanism*. These essential components are shown in Figure 3.2. Additionally, the data pipeline is further detailed in Figure 3.3, which illustrates the process from text recognition through the analysis and selection of audio data to the playback phase.

The *text editor* provides a simple interface where users can write freely while SoundMuse continuously analyzes the evolving text. The user interface offers a minimalist control panel, allowing users to enable and disable ambient sounds and background music independently. Additionally, users can manually update the current audio. As shown in Figure 3.4, enabling or disabling audio requires clicking the "Music" and/or "Ambient Sound" buttons. To

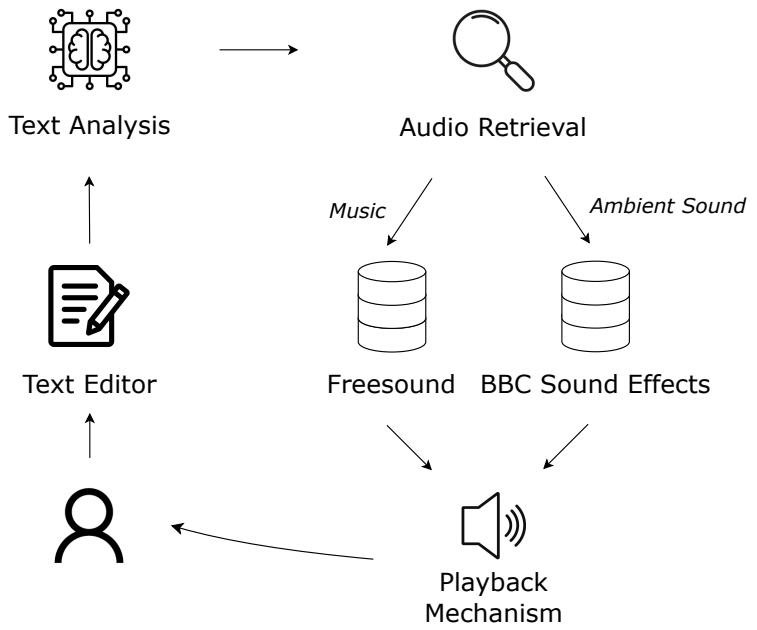

Figure 3.2: The figure illustrates the architecture of SoundMuse with its core components responsible for text processing, audio retrieval, and playback, showcasing how different parts contribute to the overall functionality of the system.

refresh the current audio, the user can type "- -" (two minus signs), triggering the playback of new background music and ambient sound.

User interface for seamless interaction

The *text analysis module* processes the written text using AI, identifying key themes, mood, context, and narrative atmosphere. Based on this analysis, it generates descriptive tags for music and ambient sound, depending on which mode is currently enabled.

AI-driven text analysis for context-sensitive audio

The *audio retrieval module* searches external databases for audio tracks that match the tags generated by AI - Freesound[1] for background music and BBC Sound Effects[2] for ambient sounds. Here, AI assists in selecting the most suitable audio for the current text segment from several

[1] `https://freesound.org/` *Accessed: April 15, 2025*

[2] `https://sound-effects.bbcrewind.co.uk/` *Accessed: April 15, 2025*

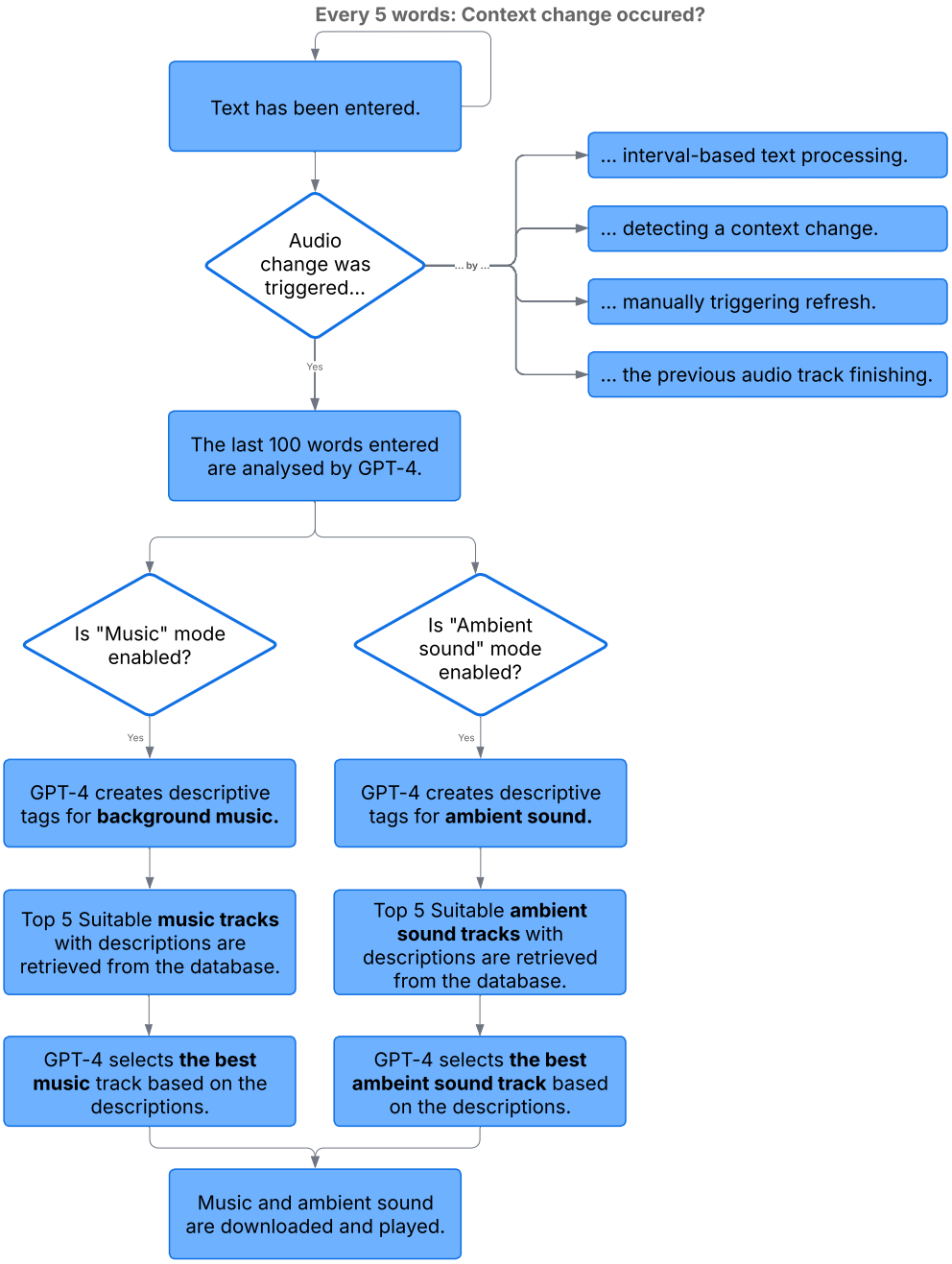

Figure 3.3: The diagram illustrates the automated process of analyzing textual input and selecting contextually appropriate audio. It showcases when audio retrieval is triggered, GPT-4's generation of descriptive tags, and how the most suitable audio is selected for playback.

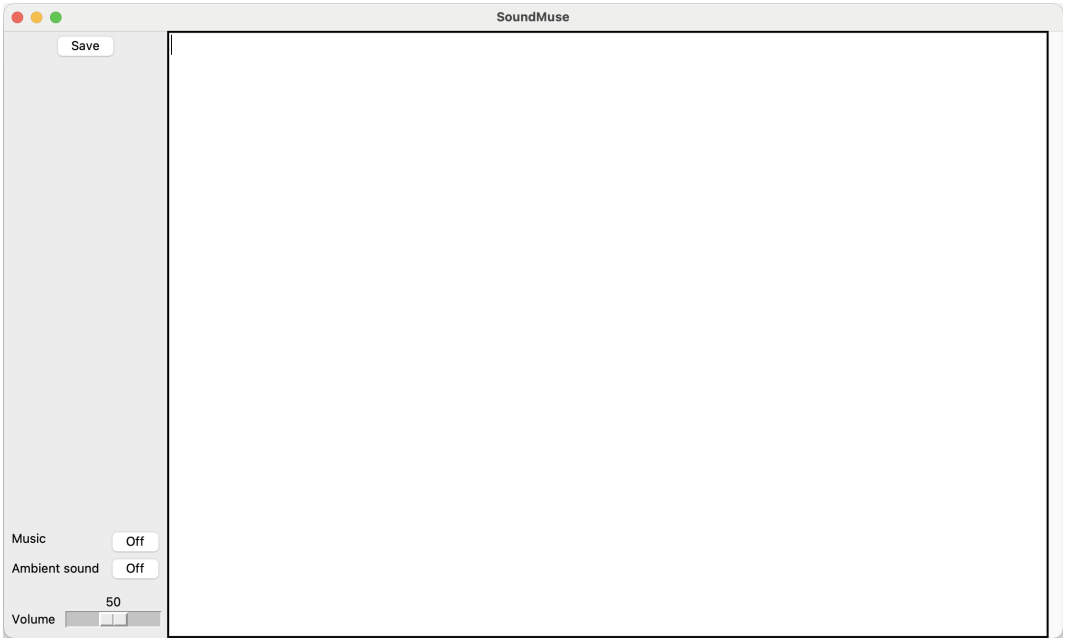

Figure 3.4: The figure illustrates the SoundMuse interface, showing the control panel where users can toggle music and ambient sounds, save the current text, and adjust the volume.

available options by looking at the audio descriptions and incorporating them into its selection.

Seamless playback

The *playback mechanism* ensures that ambient sounds and background music can play simultaneously while maintaining smooth transitions.

Together, these components form an adaptive system that supports creative writing by dynamically adjusting the soundscape to the evolving text. A detailed explanation of the system's implementation is provided in Section 3.4.

3.4 Implementation Details

3.4.1 Text Analysis Module

The *text analysis module* is responsible for processing the written text and generating meaningful descriptors that are used to select appropriate audio. This module is built on OpenAI's GPT-4[3], an NLP (Natural Language Processing) model, which is used here to analyze the text in real-time.

Interval-based text processing for efficient analysis

The system carries out an analysis at defined word intervals. This interval is defined as a parameter in the code. Currently, the interval is set to 30 words, meaning that after every 30 words, the system triggers GPT-4 to analyze the written text.

Dynamic context change detection for audio adaptation

Despite the interval-based text processing, there is a separate check for context change detection at shorter intervals. After every five typed words, the current text segment (100 words up to the current cursor position) is analyzed to determine if the current audio description tags match the new context. This analysis is again performed by GPT-4, and if a mismatch is found, GPT-4 is asked to generate new tags that describe music and ambient sound more suited to the updated context. This ensures that the system can instantly detect and adapt to changes in the narrative, maintaining a dynamic and immersive sound experience. The choice of five words offers a good compromise between a fast response time to the context change and the avoidance of overloading GPT-4 with too many requests. More information about the decision to use fixed word intervals and the secondary check for immediate context shifts can be found in Section 3.6.

Ensuring continuous audio playback

If the system detects that no audio is currently playing, even though ambient sound or music is enabled, it automatically selects new audio - ambient sound for the ambient sound mode and background music for the music mode.

[3] `https://openai.com/index/gpt-4/` *Accessed: April 15, 2025*

Four triggers for an audio change

Ultimately, four different events can trigger an audio change, provided that music or ambient sound is enabled:

- reaching the word interval,
- triggering a refresh by typing "- -",
- detecting a context change,
- detecting that no audio is currently playing.

Analysis window: 100 words

As soon as an audio change is triggered, the system takes into account the last written text — 100 words up to the current cursor position — for each GPT-4 analysis. The choice of processing 100 words at a time ensures sufficient context for analysis while keeping the input size convenient, avoiding unnecessary processing overhead for GPT-4.

Contextual tagging for audio selection

Based on the analysis, GPT-4 generates descriptive tags that capture the essence of the recent text segment. These tags are used to retrieve either background music or ambient sound, depending on the active mode. Below is an example of the prompt used for generating background music descriptions:

"Analyze the text provided by mood, context or narrative atmosphere. Create a concise music piece description, fulfilling the following criteria:

- Generate a new music description that suits the text, avoid reusing previous tags.
- Clear description of music that suits the text.
- Focus on specific musical elements that can be searched for directly in the Freesound database.
- Ensure that the description refers strictly to musical elements (melody, harmony, rhythm, and instruments) and does not include ambient sounds, environmental noises, or sound effects.
- Avoid overly long descriptions - keep the query concise and optimized for Freesound searches.

- Adapt the description to optimally support the mood, scene, or action in the section.
- Format the output as a list element of one description, e.g., ["<description as a string>"]. The description string contains a maximum of three words and only letters, no punctuation or commas.
- Present the description in English! Even if the Input is in German.
- Do not provide any additional explanations, headers, or unrelated details.

Here are the latest text sections and the previously generated description tags: {previous_text_and_tags}.
Prioritize generating a new, diverse description."

This formulation of the prompt for GPT-4 is based on several considerations derived from the goals of the project and the requirements of the Freesound database, ensuring both efficient retrieval and diverse, relevant music descriptions.

Ensuring variation in soundscape

The instruction to generate new and varied musical descriptions ensures a continuously evolving soundscape, preventing monotony. During the initial testing phase, we observed that the system occasionally repeated the same audio tracks over time, disrupting the dynamic progression of the auditory experience. To address this, we introduced the requirement to avoid reusing previous description tags. Additionally, repetitive music pieces can become predictable and distracting, potentially disrupting the creative flow. This approach aligns with the intended goal of providing dynamic auditory accompaniment.

Ensuring clarity and precision in the music description for accurate retrieval

The requirement for clear and specific music descriptions ensures accurate retrieval of suitable pieces from the Freesound database. To enhance clarity, descriptions are limited to a maximum of three words and focus exclusively on musical elements such as melody, harmony, rhythm, and instrumentation. This approach is intended to refine the search process, ensuring that only musical pieces —

rather than ambient sounds — are selected when the "Music" mode is on. The formatting of the music description for targeted searches in the Freesound database ensures that the output can be directly extracted as a list element and integrated into the code.

BBC Sound Effects prompt: Focus on enhancing ambient atmosphere

Comparing the above prompt with the one designed for ambient sound retrieval from the BBC Sound Effects database, the main difference lies in their focus. While the Freesound prompt exclusively targets musical characteristics, the BBC Sound Effects prompt identifies environmental and action-related sounds that enhance ambience. The instructions to "identify the points where background ambient soundscapes would enhance the atmosphere or complement the actions described" and "focus on sensory elements (e.g., sounds or actions) that could be paired with audio" reflect this focus.

BBC Sounds Effects: Avoiding arbitrary sound selection

Additionally, the criterion "If no suitable sound effects or sensory elements are identified in the text, provide no response at all" was added after testing revealed that GPT-4 sometimes generated arbitrary results, even when there was no clear cue for ambient sounds in the text.

3.4.2 Audio Retrieval Module

Audio retrieval from external databases

The *audio retrieval module* provides for selecting the suitable sounds and music based on the tags generated by the *text analysis module*. The system queries the two external audio databases, Freesound and BBC Sound Effects, using the tags generated by the *text analysis module*. Each tag corresponds to a specific query that the system uses to find relevant audio clips. We have two different ways to retrieve the audios from each database, with Freesound providing background music and BBC Sound Effects being used for ambient sounds.

Freesound API integration for background music retrieval

To interact with the Freesound API[4], HTTP requests are sent to the Freesound server to query audio tracks that match provided tags. The Freesound APIv2 offers several

[4] `https://freesound.org/docs/api/` *Accessed: April 15, 2025*

search methods, with the most straightforward being the Text Search resource[5], which we use in SoundMuse. The search is based on keywords. In our case, these keywords are the description tags generated by the *text analysis module*, which are directly used to find matching audios in the Freesound database. Additionally, we use the rating_desc parameter provided by Freesound APIv2[6] to sort the results by the average rating given to the audio tracks, ensuring that the highest-rated matches are prioritized. We select the five highest-rated matches from the search results.

BBC Sound Effects integration for ambient sound retrieval

The BBC Sound Effects database is integrated by querying a CSV file containing metadata for all available samples from this database. The description tags generated by GPT-4 are used as search terms to find matching sounds from the BBC Sound Effects database. The search is based on the descriptions of the samples in the CSV file, and the number of matches with the search terms is taken into account.

Prioritizing relevance in audio selection logic

As a result, for both the BBC Sound Effects and Freesound databases, the system retrieves the top five matching results based on the search criteria. The descriptions of these five samples are then passed to GPT-4, which analyzes their relevance to the current text segment and selects the most suitable audio track.

3.4.3 Playback Mechanism

Playback mechanism: Audio player integration

For audio playback, we use the Python library Pygame, specifically pygame.mixer[7] module. We chose this module as it allows simultaneous playback on two channels and offers separate control over both systems. One channel is used for Freesound audio, and the other for BBC Sound Effects. Pygame also supports smooth transitions between audio tracks, such as fade-in and fade-out, which was par-

[5] `https://freesound.org/docs/api/resources_apiv2.html#text-search`
Accessed: April 16, 2025

[6] `https://freesound.org/docs/api/resources_apiv2.html`
Accessed: April 15, 2025

[7] `https://www.pygame.org/docs/ref/mixer.html?highlight=channel#pygame.mixer` *Accessed: April 15, 2025*

ticularly important to avoid abrupt changes that could disrupt the writing flow. This playback mechanism ensures that both background music and ambient sounds can be played simultaneously and controlled independently.

Audio normalization for balanced playback

The system ensures that the volume levels are balanced, ensuring that neither overpowers the other and preventing distractions or concentration disturbances. We use the pyloudnorm[8] library to apply peak normalization, scaling the audio so that its loudest point reaches a specified dB level (defaulted to -1.0 dB). This process maintains consistent audio levels during playback.

3.4.4 User Interface

UI design using Tkinter

The UI (User Interface) for SoundMuse was created using Tkinter[9], a Python library for building simple graphical interfaces. The UI follows a minimalist design to ensure that the focus of the user remains on the writing process while providing intuitive controls for managing the soundscape.

Audio control features

The interface includes a text field for writing and buttons for activating and deactivating background music and ambient sounds. Each audio mode ("Music" and "Ambient Sounds") has a corresponding button labeled "On" or "Off", allowing the user to enable or disable them individually (see Figure 3.4). The user can click on these buttons to switch the audio playback. Additionally, a "Save" button is provided to store the written text. A volume control slider is also implemented, allowing the user to fine-tune the volume.

Refresh trigger for updating audio

The last interaction element is the ability to type "- -" into the text editor to manually update the current audio. This provides a simple and non-disruptive way to refresh the soundscape without interrupting the writing flow. If both background music and ambient sound are activated at the same time, typing "- -" will update both. Unlike the three

[8] `https://pypi.org/project/pyloudnorm/` *Accessed: April 15, 2025*

[9] `https://docs.python.org/3/library/tkinter.html` *Accessed: April 15, 2025*

other audio update triggers, the refresh trigger ensures a different prompting strategy. GPT-4 is explicitly asked to "Generate a new sound description that suits the text, avoid reusing previous tags", "prioritize generating a new, diverse description" while providing the previous tags list in the request prompt. This approach encourages variation in the audio selection, preventing repetitive patterns, as the act of manually triggering a refresh implies an intentional desire for a change in the auditory experience.

3.4.5 Error Handling and Logging

The system is designed to appropriately handle errors such as GPT-4 response errors, download errors, missing search results, or exceeding the GPT-4 token limit. Errors that occur during runtime are displayed in the terminal and stored in a JSON log file. This logging mechanism not only helps track recurring issues and allows for systematic troubleshooting and improvements but also records timestamps of when each song was played about specific text segments. This is particularly useful in user studies as it allows a detailed analysis of audio playback and provides insights into system performance.

3.5 Iterative Design Approach

Iterative implementation approach: DIA cycle

The implementation process of SoundMuse followed the DIA (Design, Implement, Analyze) cycle[10], a design and evaluation approach that consists of three stages: design, implement, and analyze. With each iteration, the prototype is gradually improved by refining the design, addressing major issues early on, and increasing the technical complexity of the implementation. The major design flaws are addressed first, while minor refinements follow later. The process involved four iterations:

[10] https://hci.rwth-aachen.de/public/Lectures/DIS1/2020_21_ws/DIS1%20-%202020%20-%20L08%20DIA%20Cycle,%20Observing%20Users,%20Brainstorming,%20and%20Storyboards.pdf
Accessed: April 15, 2025

- **Iteration 1: Concept Testing**: Identification of early problems and testing the basic functionality
- **Iteration 2: Technical Stability**: Iterative testing and debugging to ensure system robustness, involving external users
- **Iteration 3: System behavior**: Observing how the system behaves and reacts to writing
- **Iteration 4: User experience**: Studying how experienced creative writers interact with SoundMuse and how it influences their writing process
- A **diary study** with a professional writer

While the first three iterations focused on system development, the final iteration marked a shift from implementation to evaluation. At this stage, SoundMuse was considered feature-complete, and small-scale observational and experimental studies were conducted to explore its impact on creative writing. This chapter covers the first three iterations, which were primarily concerned with the implementation process. In the following, all three iterations are presented, each building on insights from testing and analysis that informed subsequent design and implementation steps. Table 3.5 and Table 3.6 provide an overview of these initial iterations.

	Aim	**Participants**	**Duration**	**Environment**
Iteration 1	Internal testing and system debugging	-	Different	-
Iteration 2	Technical stability observation	4	15 min	Lab environment
Iteration 3	Improving system behavior	2	15 min	Participants' home and writer's office

Table 3.5: The table visualizes the first three study iterations, their aims, participant numbers, duration, and environmental settings, highlighting the differences between them.

3.5.1 Iteration 1: Initial Testing

Initial phase without involving external users

The initial iteration phase was conducted without users, using a selection of around ten excerpts from existing creative texts, books, essays, and tales. We manually typed these text segments into the prototype to observe how the system reacts to different content.

Parallel processing for responsive UI

A challenge we encountered early on was the long loading time when downloading audios, which disrupted writing in the text editor. To solve this, we implemented a separate thread for the download. This allows the user to continue writing while the sounds are downloading in the background.

GPT-4 output refinement

We also observed that GPT-4 occasionally returned additional explanations or irrelevant details instead of a concise list of descriptive tags. We fixed this by adding an instruction to the prompt: "Do not include explanations, context, or unrelated details—only return the keywords as specified."

Improved playback with WAV files

Another issue occurred when playing music and ambient sounds at the same time. The original fade-out and fade-in implementation in Pygame caused short delays and stuttering. As a solution, we switched to use WAV files since they are uncompressed and can be played immediately without the additional processing time needed for MP3 decompression, resulting in smoother playback.

Audio selection timing improvement

During this initial phase, we also addressed minor issues, such as the sound being selected too early, as soon as the user began typing, which led to arbitrary selections. This was corrected by implementing a minimum word count of seven words before starting to select and play audios.

3.5.2 Iteration 2: Technical Stability

This iteration set the focus on system stability

In the second iteration, four users participated to test the system's stability. Most of the participants were students with a technical background. The studies were conducted

in a quiet laboratory setting, where the participants typed on a laptop and listened to the audio cues generated by SoundMuse through its internal speakers. In the beginning, participants reviewed and signed a consent form, which was explained to them in detail. They then wrote for about 15 minutes while we directly observed system performance, taking structured notes on technical issues and usability, and collecting brief feedback afterward. The significant observations concerned:

- occurrence of technical issues: connection drops, latency, crashes, or other unexpected errors,
- audio transitions accuracy,
- volume levels,
- user behavior and reactions.

Technical challenges - short audio loops, unresponsiveness, language issues, and playback interruptions

The results of this study iteration showed some technical challenges. It revealed that short audio tracks (e.g., 16 seconds long) were frequently replayed because users did not write fast enough to trigger a context change, causing identical tags to be generated and the same tracks to be retrieved repeatedly. To counteract this and create a more varied soundscape, we modified the prompt by including information about previously generated tags. This allowed GPT-4 to avoid repeating the same tags and instead suggest new ones.

Handling system unresponsiveness to too many GPT-4 requests

Furthermore, the system occasionally became unresponsive due to an excessive number of GPT-4 requests. For further studies, we therefore set the limit higher and supplemented the code so that the requests pause briefly when the limit is reached, as the limit resets every minute.

Language inconsistency in tag generation

Another technical limitation was language inconsistency in tag generation, as some tags were produced in German instead of English. Since the database contains only English tags, this prevented the system from retrieving appropriate audio tracks. We solved this problem by making a corresponding adjustment to the prompt: "Present the description in English! Even if the Input is in German."

	Results
Iteration 1	Long downloading times disrupted writing.
	GPT-4 occasionally returned explanations instead of tag lists.
	Playback of music and ambient sounds caused stuttering.
	Sounds were triggered too early when starting writing.
Iteration 2	Replay of short tracks due to slow writing speed.
	System unresponsiveness and threading issues.
	Tag language inconsistency (German tags).
	Abrupt audio termination.
	Need for volume control.
Iteration 3	Too fast audio transitions for writing speed.
	SoundMuse helped maintain writing flow.
	Desire for automatic switching between music and ambient sounds.
	Lack of pause indicators made the system feel unresponsive; cues could improve usability.
	Suggested fewer track transitions during focus.
	Observer presence disrupted immersion; familiarity improved experience.
	Focus improved with familiarity with SoundMuse.

Table 3.6: The table visualizes the main findings from the first three study iterations, providing a concise summary.

Abrupt audio termination issue

A minor issue occurred when an audio was abruptly terminated before the maximum playback time[11] was reached. This resulted in an undesirable experience due to the dynamics of the soundscape and the interruption of the writing flow. To prevent this, we implemented an additional detection mechanism that recognizes whether no sound is currently playing (depending on the active mode - Music or Ambient sound). If no audio is active, a new one is automatically loaded and played.

[11] Initially, a maximum playback time was a specific number of seconds set in the code, specifying how long a track should continue playing before a new one was triggered.

3.5.3 Iteration 3: System Behavior

Iteration 3: Focus on system stability and user feedback

The third iteration focused on assessing overall system behavior. As in the second iteration, data was collected through direct observation and user feedback. Participants reviewed and signed a consent form before starting, which was explained to them in detail. This time, both participants were experienced in creative writing and wrote for approximately 25 minutes. One study took place in the participant's home, while the other was conducted at the writer's office. The significant observations concerned:

- audio selection and its relevance to the writing text,
- appropriateness of transition timings,
- overall satisfaction with the writing experience using the prototype.

In this iteration, two users participated. Both participants had experience in creative writing: one was a student, and the other was the professional writer who had previously participated in the expert interview.

Two settings on Participant's request - with and without conductor

During their session, one participant requested to complete the test again after some time, asking the conductor to leave the room to explore the experience independently without being observed. In response to this spontaneous request, the study design was adjusted on the spot to allow the participant to continue testing the prototype without the conductor present. As a result, we also adjusted the study design in the subsequent iteration (see Section 4.1.1), and the conductor remained outside the room most of the time, allowing participants to test the prototype without direct observation.

System too quick to adapt to writing speed

In this iteration, a key finding was that the soundscapes changed too quickly in relation to the users' writing speed, as transitions were initially triggered by fixed time intervals. In response, we adapted the system so that the transitions are based on the number of words written rather than time. This way, the audio changes can be better adapted to the user's individual writing speed.

Suggestion to automate sound transitions

Another insight was the potential to reduce manual interaction with the prototype. One user suggested that switching between music and ambient sounds should not require user input, as manual adjustments could interrupt the writing process. Instead, an automatic system that adapts based on writing behavior could enhance focus by minimizing distractions.

Avoiding additional feedback to prevent distraction

Additionally, we observed that the absence of a clear indication for pauses between audio transitions could lead to the perception that the system was unresponsive. Implementing a visual cue to signal these pauses could improve user experience. However, we have chosen not to add any noticeable clues that could themselves be distracting.

Distinction between deep focus and breaks

One participant proposed that track changes should occur less frequently during deep focus states and more often during longer pauses, helping to maintain the writing flow while providing increased auditory stimulation during more inactive moments. In response, we have implemented a basic detection mechanism that currently only distinguishes between pause and focus mode (see Section 3.6).

Observer presence was distracting, better immersion when writing alone

A particularly relevant aspect of this iteration was that the professional writer tested SoundMuse twice - once with the observer present and once alone. This provided an interesting comparison regarding the influence of external factors on the writing experience. During the first trial, the user found the observer's presence unnatural and somewhat distracting. In contrast, in the second trial, they were more immersed in the writing process and less consciously aware of the soundscape, though they still found the auditory cues beneficial. This suggests that an undisturbed environment plays a significant role in maintaining engagement with the writing task. Additionally, in the first trial, the user momentarily shifted focus to the system when noticing audio track changes, asking, "What is happening right now?" However, in the second trial, their familiarity with the prototype appeared to contribute to a more seamless experience, allowing them to concentrate better on writing.

User found SoundMuse immersive.

During the study, it was also noted that the prototype helped maintain a consistent environment and hypothesized that this immersive experience could contribute to writing flow by creating a productive workspace.

3.6 Further Design Decisions

During the development of Soundmuse, numerous key decisions shaped the project's direction. This chapter describes these considerations and the arguments behind them.

Exploring AI-generated and database-driven approaches to audio selection

At the project's outset, we explored different approaches to audio generation. One option was using an existing audio database, while another involved generating soundscapes dynamically using AI. A fully AI-generated solution would have had the advantage of seamlessly adapting the soundscape to the text in real-time, similar to the approach used by the game mastering tool Syrinscape[12]. However, no sufficiently advanced models currently exist that can generate continuous, contextually fitting music in real-time. As an alternative, we considered AI-generated audio tracks, leveraging existing models like Meta's MusicGen[13] or Stable Audio Open[14], which generate music based on text descriptions. However, these models proved inadequate for our needs due to their insufficient quality and limited duration (e.g., Stable Audio Open's tracks lasted only 47 seconds). Consequently, we adopted a database-driven approach, selecting Freesound for background music and the BBC Sound Effects database for ambient sounds. This decision was based on the strengths of each platform: the BBC Sound Effects database provides a diverse collection of high-quality ambient sounds, while Freesound offers a broader selection of background music.

Beyond sound sourcing, we also needed to decide on the most effective approach for text analysis. We considered

[12] `https://syrinscape.com/` *Accessed: April 15, 2025*
[13] `https://musicgen.com/` *Accessed: April 15, 2025*
[14] `https://stability.ai/news/introducing-stable-audio-open` *Accessed: April 15, 2025*

Choosing between a local NLP model and an external API for text analysis

two options: training a local NLP model and integrating an external API. Given the high computational and data requirements of training a local model, we chose to integrate OpenAI's API instead. Early tests demonstrated that GPT-4 could reliably analyze text and generate contextualized sound tags, making it a suitable solution for our system.

Integrating both background music and ambient sounds in SoundMuse

We decided to implement both background music and ambient sounds in SoundMuse as research on both provides promising potential (see Chapter 2). Each offers distinct benefits that are worth investigating in the context of creative writing and AI-generated contextualized auditory cues. Additionally, every user has unique needs and preferences when it comes to auditory accompaniment. Furthermore, since users have individual preferences, the system allows them to switch between these two options according to their writing needs.

Shift from time-based to word-based audio intervals to better match the writing flow

During the initial implementation, the system was configured to change the audio after a fixed time interval, which could be customized by the user through the interface (with a default of 60 seconds if no value was specified). This approach was initially chosen for its simplicity. However, during the third iteration (see Section 3.5.3), it became evident that users type at different speeds, which meant that a fixed time interval for audio changes could be either too short or too long, depending on the individual typing pace. More importantly, the flow of ideas and the development of the narrative in writing do not necessarily correlate with a specific time duration. In contrast, the number of words written is a better indicator of the progression of the content. Therefore, we decided that implementing a word-based interval would offer a more context-sensitive approach.

The decision to fixed word intervals and secondary check for immediate context shifts

During the testing phase, we also experimented with detecting context changes rather than relying on fixed word intervals to trigger audio changes. The idea was to change the music and ambient sounds only when a context shift was detected, regardless of the number of words. However, these test results were unsatisfactory, as GPT-4 often failed to detect any context changes, leading to periods where the audio remained unchanged for extended dura-

tions or, in some cases, paused completely after one audio track finished. This made the system less responsive and diminished its effectiveness in supporting the writing process. As a result, we reverted to the previous approach of using a fixed 30-word interval, ensuring that the system would always change the soundscape at least every 30 words, providing a more responsive experience. Additionally, we introduced a secondary check after every five words to account for more immediate context shifts, particularly in cases where significant changes occurred within the 30-word window. This dual approach ensured both timely detection of clear context changes and a more continuous and adaptive auditory environment for the user.

Distinguishing between pause and focus modes

Based on the feedback received regarding the differentiation between typing pauses and being in a focused writing flow state (see Section 3.5.3), we recognized the potential benefits of distinguishing between these two states to better align the system's response to the user's writing activity. In response, we implemented two operating modes: *Pause Mode* and *Focus Mode*. *Pause Mode* is activated when the user stops writing for five seconds, while *Focus Mode* is triggered when the user is actively writing without any pauses. At this stage, there are no functional differences between the two modes other than the detection mechanism. Specifically, in *Pause Mode*, the system does not actively monitor context changes or track whether the word count exceeds the preset interval.

Chapter 4

Evaluation

As described in Chapter 3, the final phase of the iterative study shifted toward prototype evaluation. It included a small-scale exploratory user study (Iteration 4) and a diary study with a professional writer. See Table 4.1 and Table 4.3 for an overview of these studies.

4.1 Procedure

The user study (Iteration 4) and the diary study both aimed to evaluate the prototype but differed in scope and setup. The following sections describe the respective procedures in more detail.

4.1.1 Iteration 4: User Experience

Iteration 4: Semi-structured interviews and questionnaires; longer testing time; lab environment

During the fourth iteration, semi-structured interviews (see Appendix B) and questionnaires (see Appendix C) were used to gather insights from creative writers on how the generated auditory cues influenced their writing process. Studies were carried out in person, with participants asked to write a creative text. They had the option to either continue working on their project or start with four provided

sentences for inspiration (see Appendix D). The idea of providing story starters was inspired by the results of the expert interview with the professional writer, who noted that they frequently use such practices in their creative writing teaching. The study was conducted in a laboratory environment, as illustrated in Figure 4.2. Unlike previous iterations, a dedicated mono sound system was used instead of the laptop's internal speakers to provide a more consistent auditory experience.

	Aim	Participants	Duration	Environment
Iteration 4	The influence of SoundMuse on creative writing	4	60-90 min	Laboratory setting
Diary Study	The influence of SoundMuse on creative writing for a professional writer	1	a few hours	The writer's office

Table 4.1: The table visualizes the first three study iterations, their aims, participant numbers, duration, and environmental settings, highlighting the differences between them.

Participants: Demographic information and background gathered

In this study phase, we collected demographic data, which involved four participants with an average age of 30.75 years. Three of the participants were students, two of whom held a Bachelor's degree, while two had completed their high school diploma. One participant was a Marketing Manager, with the remaining participants being students. The professional writer, who participated in both the third iteration and the diary study, was 57 years old. The participants reported different writing habits: two wrote daily, one twice a week, and one once a week. Their creative works included genres such as fantasy, magical realism, horror, erotica, and poetry. All participants stated that they wrote creatively regularly. All participants reported listening to music or ambient sounds at least occasionally while writing, except the professional writer, who stated that they do not listen to anything while writing. The favorite genres varied, including chamber music, ambient, pop, lo-fi hip-hop, and rap beats.

The participants were informed that the study would be recorded and that the recordings would not be published.

Figure 4.2: The study setting of the fourth iteration, showing the workspace setup used for the SoundMuse preliminary user study.

Lab setting, observations and data collection through recorded interviews

They were also notified that the study conductor would check 10-15 minutes before the writing session ended. Before starting, they were given instructions on how to use the prototype. At the beginning of the study, the conductor provided technical support to ensure that the system was working properly, after which they were left alone in the room. During the writing process, screen and voice recordings were made to capture the interactions of the participants with the user interface and their verbal reflections. In the last 10-15 minutes, the study conductor observed the writing process of the participants. Afterward, they completed a questionnaire and took part in a short interview discussing their experience and the influence of the soundscapes on their writing. The interview was recorded. The entire session was designed to last approximately 120 minutes, with the writing phase taking up most of this time.

Questionnaire design: influence of audio environment, immersion, flow, enjoyment, and system usability

Questionnaires completed after writing sessions were used to analyze participants' experiences with SoundMuse. We focused on understanding how the audio environment influenced writing behavior, the perception of scenes, its emotional impact, as well as whether SoundMuse was helpful and, if so, in what specific ways. Building on the insights gained from the expert interview (see Section 3.2), we explored topics such as immersion [Rigby et al., 2019], flow

[Jackson and Marsh, 1996], enjoyment [Chen et al., 2021], and system usability [Brooke, 1996]. To investigate these areas, we referred to standardized user questionnaires commonly used in similar studies, adjusting them to meet our specific needs. The responses were collected using a 5-point Likert scale, where 5 represented the highest agreement ("Stimme stark zu") and 1 the lowest ("Stimme gar nicht zu"). The final questions in the questionnaire addressed demographic information and participants' general experiences with creative writing, as well as writing with music or sounds.

4.1.2 Diary Study

The diary study revealed how SoundMuse fits into a professional writer's work

After the fourth iteration, a professional writer participated in a diary study to investigate the use of the prototype in their natural working environment. The writer, who had previously participated during Iteration 3, expressed interest in exploring how SoundMuse might influence their creative process in a typical working environment. Unlike typical diary studies, the writer tested SoundMuse for just one day, for several hours, rather than over an extended period. During this time, the writer used the system without being observed and kept notes throughout the process. The system was used with internal speakers. After use, they were interviewed about their experience and provided feedback based on the notes they had taken during the use. The diary study was especially valuable as it provided insight into how the system fits into a real-world creative writing process, specifically from the perspective of a professional writer, as opposed to an observed setting.

4.2 Results

The following section outlines the results of the two evaluation studies. It begins with findings from the exploratory user study (Iteration 4), followed by the results of the diary study.

	Results
Iteration 4	Increased writing fluency over time
	Request for Immediate Audio Stopping
	Desire for customization of sound types
	Repeated exposure may enhance writing fluency
Diary Study	Overall enjoyment
	SoundMuse integrated well into workflow
	Increased immersion and productivity
	Suggestion for audio cue personalization
	Need for better adaptation to writing speed

Table 4.3: The table visualizes the main findings from iteration 4 and the diary study, providing a concise summary.

4.2.1 Iteration 4: User Experience

Auditory cues influenced emotions.

In the fourth iteration, the four participants described a variety of impressions of the auditory cues during the writing process. Three participants reported that they had a noticeable influence on their emotions during the writing process. One person stated that the auditory cues directly influenced their emotions. Another described how the background music made them feel more emotional and more connected to the story and their imagination. Another person described how the auditory cues had triggered a certain mood in them, which then flowed into the text. This aligns with the values from the questionnaire, where the statement "The acoustic accompaniment influenced my emotions during the writing process" received high agreement ratings (5, 4, 5, 5). One person also stated that auditory cues had motivated them to write.

Auditory cues inspired writing.

Two participants described the soundscapes as inspiring. At moments when they were stuck in their writing, they allowed themselves to be guided by the auditory cues or felt inspired to think about what they had heard and integrate it into the story. This was supported by the questionnaire results, where the statement "The acoustic accompaniment

inspired my writing" had an average value of 4.25, indicating a perceived inspirational effect.

Auditory cues influenced writing behavior.

About the influence on writing, several participants described that suitable audios made writing easier, especially when expressing emotions such as sadness or pleasure. It was emphasized that the auditory cues facilitated writing when it matched the content. One person reported that they consciously tried to let auditory cues affect their current state of mind, which in turn influenced their writing style. Another person described how the soundscape helped them to create a clear picture "in their mind's eye". Sometimes, it felt as if they were being guided by the sounds, as if the soundscape was "drawing them into a new world". This feedback is supported by the questionnaire, where the statement "The acoustic accompaniment has influenced my writing behavior" received average value of 4, suggesting an overall influence on the writing process.

Auditory cues created atmosphere.

Three participants described how the soundscape helped them imagine scenes more vividly. One person said that they were able to close their eyes inwards and thus put themselves in the described environment particularly well. However, the questionnaire responses regarding the statement "The acoustic accompaniment supported the atmosphere I wanted to create in my text" were more neutral (4, 4, 3, and 3), indicating that the support may not have been as strong for all participants.

Auditory cues were sometimes distracting.

Some auditory cues were perceived as disturbing. One person reported that they became rather confusing as the process progressed and unexpected noises disrupted the flow of writing. Another participant noted that certain types of background music, such as "cheeky" tracks, were perceived negatively. In another case, the repeated playback of a track was described as unpleasant. One person described how the background music did not always match their current mood and that they sometimes needed something other than what was currently playing. It was also mentioned that the conscious perception of the soundscape occasionally made the writing process more difficult. The questionnaire feedback ("The acoustic accompaniment disturbed my writing.") reflected this, with more neutral

scores of 4, 3, 3, and 3, suggesting that while occasional distraction occurred, it was not considered a major issue.

Auditory cues mostly fit, but some mismatches occurred.

In the interview, we also asked whether the users felt the acoustic accompaniment matched the context of their writing. Participants emphasized that although there were moments when the soundscape matched the text well, there were also occasional phases when it was a less precise fit. One user noted that acoustic cues often fitted the content well, e.g., when heartbeat sounds were used in scenes describing such sounds. Another participant mentioned that occasional unexpected sounds, such as car and city noises, did not match the written content.

Technical challenges revealed during this Iteration.

This iteration of the study also revealed some minor technical challenges. One participant mentioned how the refresh trigger responded too slowly and caused confusion. The user preferred to manually turn the music and ambient sounds off and on again instead to achieve the same desired result. Another issue was related to the log file, which caused an error during one trial, resulting in a shortened test time. This issue has been addressed and resolved by improving the handling of the log file. In addition, a delay in the appearance of the text on the screen during typing was observed, which only occurred for the first participant.

Additional features were requested.

In response to the question of whether they would have liked additional features for the prototype, two participants expressed a desire for smoother transitions between audio tracks without delays so that the experience would feel more dynamic. Another suggestion was the possibility of adding a dividing line after which the system would no longer analyze the previous content so that users could start new text without having to delete previous content. In addition, a desire was expressed to add alternative back-

ground sounds, such as alpha waves[1] or ASMR[2], instead of specific audio cues.

The questionnaire results additionally revealed that the participants were mostly aware of the auditory cues while writing (average 4.25) but gave mixed ratings to the reaction of the soundscape to the text (average 3.25). There was a desire for more control over the choice of music and sound (average 3.75). The ambient sounds were felt to be less helpful (average 2.5), while the music was considered helpful (average 4.25). Overall, the auditory cues were rated as useful (average 3.75), and the willingness to recommend the prototype to others was at an average of 3.5.

The results are summarized and visualized in Figure 4.4, which presents a bar chart of the average ratings of auditory experience. To illustrate user experience aspects such as immersion, flow, enjoyment, and system usability, we present heatmaps in Figure 4.5. Here, darker colors indicate more favorable responses. For questions with negative wording, we inverted the Likert scores (e.g., 1 became 5, 2 became 4, etc.) so that higher values always reflect more positive outcomes. Inverted items are marked with an (I) at the end of the corresponding question. The columns in heatmaps represent individual participants' responses.

To summarize the findings based on the feedback and questionnaires, the overall ratings for helpfulness, immersion, and user experience were as follows:

- Immersion: High ratings, particularly in situations where the soundscapes matched the writing content.
- Flow: Generally positive, but with some variation based on sound preferences.

[1] Alpha waves are brain oscillations in the frequency range of 8-12 Hz that are typically associated with relaxed alertness and reduced cognitive effort [Malik and Amin, 2017].

[2] ASMR (Autonomous Sensory Meridian Response) describes a tingling, pleasant sensation that some people experience in response to certain auditory or visual stimuli such as whispering or tapping. It is often accompanied by feelings of calm and relaxation [Mahady et al., 2023].

Figure 4.4: The figure displays average ratings on a 5-point Likert scale for selected questions from the questionnaire used in iteration 4. It provides a clear overview of how participants evaluated specific aspects of the auditory experience.

Immersion

3	2	2	4	I put a lot of effort into my writing. (I)
4	5	4	4	I felt motivated to write.
5	5	4	4	The writing process captured my attention.
5	4	4	4	I was concentrated on the writing process.
3	4	4	4	I would like to use the prototype for writing again.

Flow

5	4	4	2	My attention was completely focused on writing.
4	4	5	3	I really enjoyed the experience.
4	5	4	4	I wasn't worried about my performance.
5	3	4	3	I lost track of time while writing.
5	3	4	3	I was completely absorbed in writing.

Enjoyment

5	5	5	4	I had fun writing.
5	5	4	5	I have the feeling that I was successful in my writing.
3	4	4	4	I would like to use the prototype for writing again.

System Usability

5	5	5	4	I found the system easy to use.
3	3	5	2	I found the system too inconsistent. (I)
3	3	3	5	I think I would like to use this system frequently.

Figure 4.5: The heatmaps display participant ratings across four areas: immersion, flow, enjoyment, and usability, with each column representing an individual participant. Immersion and enjoyment received consistently high ratings, while flow and system usability showed more variability.

- Enjoyment: Mixed, with some participants noting how enjoyable the auditory cues were for enhancing creativity.
- System Feedback: The system's usability and its integration of sound feedback were generally rated as functional, though some participants expressed a desire for more flexibility and control over the music selection.

4.2.2 Diary Study

Writing process: free writing combined with structured reflection using SoundMuse

In the interview with the writer, the participant described the approach to the writing process while using SoundMuse. The writing began without a fixed plan, with a few keywords being noted down along the way. These keywords served as reference points for further development, combining free writing with structured reflection during the trial of the prototype.

Longer audio durations preferred, short pauses felt disruptive.

A preference was expressed for longer play durations of audio clips when they fit well with the writing, although it was later realized that the soundscape was no longer consciously noticed after a while. Furthermore, the short pauses between audio changes were perceived as somewhat disruptive.

Repeated audio cues were distracting; chaotic sounds felt disruptive.

It was noted that some audio clips seemed to repeat, which was distracting, especially given the musical inclination of the participant, who tends to notice patterns such as recurring chords and harmonies. Certain sounds (e.g., the sounds of screaming monkeys), due to their chaotic nature, were found out of place and disruptive.

Prototype responded to text input, but this was not noticed most of the time.

At some point, the user tested whether the prototype would respond to specific words by deliberately typing something to see how it would affect the sounds. It was confirmed that the prototype adapted its audio cues in response to input, although it had not been previously noticed whether the audio changes were contextually aligned with the content of the writing during use.

SoundMuse was overall helpful for maintaining focus, but not intended for long-term use.

Overall, the experience of using the prototype was described as beneficial, particularly in terms of creating an atmosphere that supported the writing process. The prototype helped to create an atmosphere that facilitated writing and made it easier to stay at the desk and engage with the process. Although the writer did not envision the prototype to become part of a permanent routine, it was felt to be useful. However, it was acknowledged that the experience could have been different if there had been writer's block or other major writing challenges. The participant mentioned that SoundMuse provided additional support without addressing fundamental issues, but it still proved beneficial to the writing process as a whole.

Chapter 5

Discussion

Taken together, the results show both the strengths of the current prototype in promoting creativity and the limitations that need to be overcome in order to realize its full potential.

Soundscapes can support emotional alignment and engagement in writing.

Our study results expand on previous research findings by showing that dynamically generated sound cues can influence emotions during creative activities [Mones and Massonnié, 2022], specifically in the context of creative writing. These findings suggest that an adaptive soundscape can place writers in different emotional states that align with their text, potentially strengthening their emotional connection to it and promoting the creative writing process. The data also suggests that contextualized auditory cues can have a motivational effect, helping writers stay engaged with their work.

Immersion: Suitable soundscapes support immersion and ease the creative process.

The results also suggest that context-specific auditory stimuli can create an atmosphere and environment that match the writer's imagination. This could facilitate immersion in the writing situation and positively influence the creative process. Based on participant feedback, an appropriate soundscape can even ease the writing process by fostering a more immersive creative state.

A new and interesting finding is that the dynamic changes in the auditory cues also influenced what participants

Creative stimulus: Audio cues inspire but may influence content direction.

wrote about. The evolving soundscape can serve as a source of inspiration by providing new impulses for writing. Some participants stated that certain sounds and music pieces gave them new ideas or stimulated their imagination. This suggests that a dynamic soundscape can act as a creative stimulus. This observation aligns with the findings of Nelson and Guegan [2019], showing that environmental cues can impact the orientation of creative production. This positive effect, however, also raises the question of how far such cues might subtly guide creative decisions. The inspirational influence of the soundscape can thus be ambivalent - on the one hand, it supports creativity, but on the other hand, it may distract authors from their original intentions.

Writer's block: Dynamic soundscapes can stimulate writing.

This inspiring effect of soundscapes also revealed practical potential during the study. As the professional writer noted, tools like SoundMuse could help overcome writer's block. It can provide a continuous stream of sensory impulses that stimulate the imagination and emotional connection to the text, making it easier to develop new ideas and stay engaged.

Perception of sounds is subjective and needs to be considered.

Our results additionally suggest that music and ambient sounds are perceived differently by individuals, supporting the argument that the relationship between music and creativity is complex and subjective [Xiao et al., 2023]. These findings highlight the importance of considering individual differences when designing supportive systems for creative writing and other creative tasks.

Technical challenges: Improvements in audio-text matching

Another important aspect of our study concerns the technical challenges of adapting the soundscape to the text. The participants expressed different opinions about the correspondence of music and ambient sounds with the written content. This indicates that the system still needs to be optimized or that individual differences in perception play a role.

We observed an overall stronger preference for background music over ambient sounds. This may indicate that background music is generally more favorably received. However, it could also be due to the selection of ambient sounds, which were perceived as too specific. This raises the pos-

sibility that either the sound description tags generated by GPT-4 were not optimal or that the search and retrieval process within the database needs further refinement.

Interaction issues and need for customization based on user preferences.

Some usability challenges during interaction with the system, such as lag during typing or slow reaction of refresh trigger, point to the need for refinements for a more seamless interaction. In addition, user suggestions for potential additional features show that it would be conceivable to offer customizable functions that adapt to individual preferences or writing styles.

Generalization: soundscapes may support creativity beyond writing.

Beyond the specific context of creative writing, the findings offer insights into the broader relationship between sensory input and creative processes. The study shows that context-sensitive auditory cues can influence emotional engagement, direct the flow of ideas, and enhance motivation, all of which are key components of creativity [Mastria et al., 2019; Boden, 2009; Zhou et al., 2020]. This suggests that similar systems could support other creative domains, such as design, painting, or any other creative task where emotional resonance and imagination also play a central role.

Other sensory modalities can have similar effects on creativity.

If auditory cues can serve as a creative stimulus, it is plausible that other sensory modalities might have comparable effects. Visual stimuli, for instance, have been shown to influence creative imagination [Nelson and Guegan, 2019], while olfactory cues can evoke memories and emotional states that may inspire creative thought [Gonçalves et al., 2017]. These observations support the idea that creativity is not just a mental process but is influenced by how we perceive and interact with our environment [Malinin, 2016]. This suggests that designing multisensory creative environments and tools could enhance creative processes by fostering imagination, flow state, and emotional engagement.

Chapter 6

Summary and Future Work

In this concluding chapter, we summarize the key findings and contributions of this work. We close by suggesting possible approaches for future work that build on the insights gained here.

6.1 Summary and Contributions

We developed a tool for auditory support of creative writing.

In this thesis, we developed the SoundMuse prototype, which aims to support the creative writing process through the dynamic provision of background music and ambient sounds. The aim was to investigate how context-sensitive and dynamic auditory cues can influence the creative writing process.

GPT-4 generates suitable music and ambeint sound descriptions to the context to retrieve the fitting audios.

At the beginning of the work, we developed a design concept that allows music and ambient sounds to be generated in real-time based on the analysis of the written text. For this, we used a text analysis based on OpenAI's GPT-4 to recognize context changes in the text and generate appropriate music and ambient sound descriptions, taking the emotional tone and narrative content into account. These descriptions were then used to search for and play rele-

vant sounds from the Freesound and BBC Sound Effects databases. We implemented an interactive mode that allows users to decide for themselves whether they want to listen to music, ambient sounds, or a combination of both while writing.

We tested the system iteratively: from stability to user feedback.

We tested and refined the system in an iterative study that involved observing user reactions and behavior while writing with the SoundMuse prototype, as well as gathering user feedback. The iterative process was structured to begin with a focus on technical stability, followed by system behavior evaluation, and concluded with experimental small-scale user studies, including a diary study to capture user experiences and feedback over time.

Auditory cues increase focus and emotional engagement while writing.

Our findings suggest that context-sensitive music and sound can support the writing process, enhancing creativity and maintaining focus during writing. Users reported that auditory cues helped them stay engaged with their work, and many found that the auditory environment positively influenced their emotional connection to the text, in some cases even guiding and inspiring the content.

Findings: SoundMuse inspired new ideas and influenced writing direction.

These findings are particularly significant considering the central role that creativity plays in various areas of life [Kirsch et al., 2016; Putney et al., 2024]. If an AI-driven tool like SoundMuse can help unlock creative potential, it has the potential to support not just writers but other creative professionals, too.

Subtle audio transitions and better coordination with the narrative are still required.

While the system proved beneficial, we also identified some challenges, such as the need for more subtle audio transitions. Additionally, the system did not always align perfectly with the narrative of the text, and some music or sound pieces were found to be distracting.

Contribution: AI-generated soundscapes as a tool for creativity beyond writing.

By combining theoretical research and practical implementation, this work has contributed to the understanding of how AI in the context of music and sound can serve as a tool to enhance creativity. The findings highlight the potential of AI-driven auditory cues to influence emotional engagement, idea flow, and motivation during creative writing, as well as other creative tasks. By exploring how col-

laborative AI can support the creative process through personalized and context-sensitive stimuli that adapt to the individual's needs, this work provides new insights into the role of AI and auditory stimuli in enhancing and supporting creativity.

6.2 Future Work

Although this work already makes an important contribution to the study of the influence of music and sound on the creative writing process, there are several opportunities for future research. In the following, we outline some recommendations and opportunities for enhancing SoundMuse, as well as directions for future research.

Improving text-to-audio relevance through logging and user feedback

One of the key areas for improvement lies in the system's ability to analyze text and generate appropriate sound cues. Our findings suggest that the system sometimes fails to generate contextually relevant audios. Future work could focus on investigating this issue more deeply by analyzing log entries to examine which descriptive tags are being generated for specific contexts and which audio files are retrieved as a result. Additionally, user feedback could be collected to identify moments when the audio is perceived as inappropriate, e.g., by observing when users choose to refresh it. This information could then be analyzed to understand user behavior and identify potential solutions for improving text analysis and sound relevance.

Investigating the influence of soundscape on narrative direction

A different avenue for future exploration would be a deeper investigation into how auditory cues influence the content itself. Specifically, it would be valuable to examine whether certain music or sounds manipulate the direction of the narrative and if this effect is potentially disruptive. This could help clarify whether auditory cues can be strategically used to enhance storytelling or if they risk altering the creative process in ways that could be perceived as manipulative.

To improve the system's functionality and user experience, several technical improvements can be made:

Sound Selection
Instead of letting users choose between the two modes, the system could automatically detect whether music, ambient sound, or a combination of both best suits the current writing context and adapt accordingly to the flow of the text to avoid disruptions in the writing process.

Download Optimization
Reducing the download time for audio files would allow the system to respond more promptly to changes in the written text, ensuring that the sound updates align more closely with the user's writing flow.

Mode-Specific Behavior
Differentiating system behavior between *Pause Mode* and *Focus Mode* could enhance the user experience by adjusting the sound environment according to the user's current needs.

Autocorrect
Implementing an autocorrect feature could further benefit the writing process.

Personilization
Another direction for improvement is enhancing the system's ability to adapt to individual preferences. One approach could involve allowing users to set the tone for their writing by generating a main prompt that enables them to specify their genre or the kind of auditory accompaniment they prefer (e.g., classical music). This would allow the system to tailor its suggestions to individual needs.

Exploring reasons for the lower effectiveness of ambient sounds

It would also be valuable to examine why ambient sounds, in particular, seem less effective in the current version of the system. One potential solution could be improving the search strategy to better match the context of the writing. Another possibility is to limit the use of ambient sounds, playing them only at specific points in the text where they are contextually appropriate, rather than continuously throughout the writing process. Alternatively, this could reflect a general user preference, which could be further explored in large-scale studies.

Conducting extended user experience studies

In addition to these technical improvements, expanding the user experience studies would be beneficial. While the current study has provided valuable preliminary insights into the impact of the system on the creative writing process, further research could examine whether these effects persist over time and across different user groups.

A promising direction would be to conduct a larger-scale user study with a greater number of participants, employing a more differentiated set of evaluation criteria. For instance, future studies could systematically investigate aspects such as flow, immersion, and enjoyment, as our initial results suggest that SoundMuse positively influences these dimensions. This could be supported by validated standard questionnaires. In addition, a more extensive longitudinal diary study could provide insights into how the system is used over time. Participants could work with the prototype on a daily basis over a longer period and document their experiences in different writing contexts. This would allow researchers to analyze the effects of contextualized soundscapes in individual writing routines.

Such studies could also explore how personal factors, such as writing habits, mood, or creative goals, interact with the auditory cues, providing a richer understanding of how collaborative AI tools like SoundMuse can adapt to their users.

Exploring other creative disciplines

Beyond the context of creative writing, SoundMuse could be explored in other creative disciplines to investigate how context-sensitive auditory support might enhance different forms of creativity, offering deeper insights into how AI-driven soundscapes can influence creative processes.

To conclude, this work demonstrates how AI can actively support creative processes by responding to the creator's input and adapting to contextual and emotional shifts. Rather than simply acting as a technical tool, AI can take on a more active, supportive role in the creative process. At the same time, our work provides valuable insights into how music and sound can positively influence creativity, especially in writing. Together, these findings provide a

solid foundation for future research and further exploring the interplay between AI, sound, and human creativity.

Appendix A

Expert Interview Guide

Erfahrung

Zum Einstig eine Frage (gerne kurz), um Ihre bisherige Erfahrung einzuordnen: Wie lange schreiben Sie schon, und welche Art von Texten verfassen Sie hauptsächlich?

1. **Schreibprozess/-phasen (:45) 10 min**
    - Welche Phasen durchlaufen Sie typischerweise in Ihrem Schreibprozess und variieren diese je nach Stimmung, Thema oder andere Faktoren?
    - Lesen Sie zwischendurch, während Sie schreiben, oder versuchen Sie, erstmal ununterbrochen zu schreiben?
    - Sind bestimmte Schreibphasen oder Abläufe bei anderen Schriftsteller:innen bekannt, die Sie selbst interessant oder hilfreich finden?
    - Welches Medium nutzen Sie normalerweise zum Schreiben? Digital oder auf Papier? Falls digital: Gibt es bestimmte Programme oder Texteditoren, die Sie bevorzugen?
    - In welcher Schreibphase denken Sie wäre der Tool-Einsatz am sinnvollsten?
2. **Ideenfindung (:52) 7 min**
    - Wie gehen Sie bei der Ideenfindung vor? Gibt es bestimmte Inspirationsquellen (oder Routinen), die Ihnen dabei helfen?
    - Gibt es Techniken oder Ansätze anderer Schriftsteller, die Sie kennen?
    - Was empfinden Sie allgemein als hilfreich bei der Entwicklung neuer Ideen? Gibt es z.B. Tipps, die Sie Ihren Studierenden in Ihrem Kurs mit auf den Weg geben?
3. **Blockaden (:59) 7 min**
    - Erleben Sie manchmal Schreibblockaden? Wie gehen Sie damit um? Gibt es Strategien, die für Sie besonders gut funktionieren?
    - Wie relevant sind Schreibblockaden für Schriftsteller?
    - Was erwarten Sie von einem Tool, das Blockaden beeinflussen könnte? Wie könnte es helfen, solche Blockaden zu lösen?

4. **Umgebung (:05) 6 min**
    - Welche Rolle spielt Ihre Umgebung beim Schreiben? Gibt es bestimmte Orte, Geräusche oder Stimmungen, oder Menschen, die hilfreich sind?
    - (Wie könnten Sie sich vorstellen, dass ein Tool eine unterstützende Umgebung schaffen könnte?
      Welche Eigenschaften wären dafür wichtig? (bezogen auf Art von Musik, oder z. B. eine Kontroll-Funktion, was die Klänge manuell verändert, ...))
5. **Auditive Reize (:17) 12 min**

    - Spielt Musik in Ihrem Schreibprozess eine Rolle?
      Wenn ja, welche Art von Musik hören Sie, und warum?
      Wenn nein, könnten Sie sich vorstellen, dass Musik eine unterstützende Wirkung haben könnte? Wie? Kennen Sie andere Schriftsteller:innen, die gerne Musik beim Schreiben hören?

    - Wie stehen Sie zu anderen auditiven Reizen, wie atmosphärischen Klängen (z. B. Naturgeräusche, Stadtgeräusche)? Schreiben Sie manchmal bewusst außerhalb, z.B. in einem Café oder in der Natur? Können solche Klänge hilfreich sein oder lenken sie eher ab?

    - Was denken Sie über den Unterschied zwischen atmosphärischen Klängen und Musik in dem Tool? Welche Reize wären besser, um den Schreibprozess zu unterstützen?

    - Stellen Sie sich das Tool vor: wie würde es aussehen und was würde es können? Wie wäre ein ideales Tool für Sie und was wäre ein ganz furchtbares Tool?
      (Was macht das, wie funktioniert das und andererseits wie wirkt das auf Sie?)

    - Welche Chancen sehen Sie bei dem Tool? Oder keine (gerne ehrlich)?

    - Welche Gefahren sehen Sie bei dem Tool?

    - Was denken Sie zu der Klang-Auswahl? Wie stellen Sie sich vor, dass die Klänge ausgewählt werden, passend zu dem Geschriebenen?
      Wie oft findet Musikwechsel statt?

6. **Studie (:27) 10 min**
    - Haben Sie Vorschläge, welche Faktoren (z. B. Produktivität, Stimmung, Textqualität, Kreativität…) man nach Fertigstellung des Tools in einer Studie untersuchen könnte?
    - Wie könnten man das bewerten?
    - Was erwarten Sie, was wir herausfinden in einer Studie, (wie der Einfluss auditiver Reize auf das kreative Schreiben sein könnte)? Irgendwelche Vermutungen?

Zum Schluss:

- Gibt es noch etwas, das Sie zum Thema kreatives Schreiben oder auditive Unterstützung hinzufügen möchten?
- Wenn Sie an die Zukunft des kreativen Schreibens denken, wie sehen Sie die Rolle von Technologie und künstlicher Intelligenz in diesem Bereich?

Appendix B

User Study Interview Guide

1. Bezug auf das Beobachtete während der Studie

2. **Wahrnehmung der akustischen Begleitung**
    - Wie hast du die akustische Begleitung beim Schreiben empfunden?
    - Hat sie den Schreibprozess eher unterstützt oder gestört?

    - **Falls sie gestört hat**:
        - In welchen Momenten war das der Fall?
        - Gab es einen bestimmten Grund dafür?
        - Hast du darauf reagiert, z.B. durch den Refresh-Trigger oder den On/Off-Button?

    - **Falls sie unterstützend war**:
        - In welchen Momenten war das der Fall?
        - Inwiefern?

3. Hast du einen Unterscheid zum Schreiben ohne akustische Begleitung gemerkt?

4. **Inhaltlicher Einfluss**
    - Hattest du das Gefühl, dass die Tonbegleitung dein Schreiben beeinflusst hat?
    - Hat sie das Setting oder die inhaltliche Entwicklung des Textes verändert oder beeinflusst?
    - Falls ja, an welchen Stellen?

5. Kannst du dir vorstellen, dass das Tool helfen könnte, Schreibblockaden zu lösen?

6. Hättest du dir weitere Funktionen (Features) in dem Tool gewünscht? Welche?

7. Gab es technische Probleme bei der Nutzung des Tools?

Appendix C

Questionnaire

Fragebogen für die Nachuntersuchung

Dieser Fragebogen dient dazu, Ihre Erfahrungen mit der akustischen Begleitung beim Schreiben zu untersuchen. Bitte geben Sie im Folgenden an, inwieweit Sie den angegebenen Aussagen zustimmen oder nicht zustimmen. Kreuzen Sie die entsprechenden Kästchen an.

	Stimme stark zu (5)	**Stimme zu (4)**	**Neutral (3)**	**Stimme nicht zu (2)**	**Stimme gar nicht zu (1)**
1.1 Die akustische Begleitung hat mein Schreibverhalten beeinflusst.	○	○	○	○	○

Falls ja, bitte erläutern Sie:

1.2 Die akustische Begleitung hat meine Emotionen während des Schreibens beeinflusst.	○	○	○	○	○

Falls ja, bitte erläutern Sie:

1.3 Die akustische Begleitung hat die Atmosphäre unterstützt, die ich in meinem Text erzeugen wollte.	○	○	○	○	○

Falls ja, bitte erläutern Sie:

1.4 Die akustische Begleitung hat beim Schreiben gestört.	○	○	○	○	○

Falls ja, bitte erläutern Sie:

1.5 Die akustische Begleitung hat mich beim Schreiben inspiriert.	○	○	○	○	○

Falls ja, bitte erläutern Sie:

	Stimme stark zu (5)	**Stimme zu (4)**	**Neutral (3)**	**Stimme nicht zu (2)**	**Stimme gar nicht zu (1)**
2.1 Ich konnte mir die Szenen oder Ereignisse im geschriebenen Text durch die Geräusche besser vorstellen.	○	○	○	○	○
2.2 Beim Schreiben habe ich die akustische Begleitung die ganze Zeit wahrgenommen.	○	○	○	○	○
2.3 Die akustische Begleitung reagierte gut auf den geschriebenen Text.	○	○	○	○	○
2.4 Ich hätte mir mehr Kontrolle über die Auswahl der akustischen Begleitung gewünscht.	○	○	○	○	○
2.5 Die Klänge („Ambient sound“) waren hilfreich.	○	○	○	○	○
2.6 Die Musik („Music“) war hilfreich.	○	○	○	○	○
2.7 Die akustische Begleitung allgemein war hilfreich.	○	○	○	○	○
2.8 Ich würde das Tool an Andere weiterempfehlen.	○	○	○	○	○
3.1 Meine Aufmerksamkeit war ganz auf das Schreiben gerichtet.	○	○	○	○	○
3.2 Ich habe die Erfahrung sehr genossen.	○	○	○	○	○
3.3 Ich war nicht besorgt über meine Leistung.	○	○	○	○	○
3.4 Ich habe beim Schreiben die Zeit aus den Augen verloren.	○	○	○	○	○
3.5 Ich war völlig in das Schreiben vertieft.	○	○	○	○	○
4.1 Ich habe mich beim Schreiben sehr angestrengt.	○	○	○	○	○
4.2 Ich fühlte mich motiviert beim Schreiben.	○	○	○	○	○
4.3 Der Schreibprozess fesselte meine Aufmerksamkeit.	○	○	○	○	○
4.4 Ich habe mich auf den Schreibprozess konzentriert.	○	○	○	○	○

4.5 Ich würde das Tool gerne wieder zum Schreiben verwenden.	○	○	○	○	○

5.1 Ich hatte Spaß am Schreiben.	○	○	○	○	○
5.2 Ich habe das Gefühl, dass ich beim Schreiben erfolgreich war.	○	○	○	○	○

System	**Stimme stark zu (5)**	**Stimme zu (4)**	**Neutral (3)**	**Stimme nicht zu (2)**	**Stimme gar nicht zu (1)**
6.1 Ich fand das System einfach zu bedienen.	○	○	○	○	○
6.2 Ich fand das System zu inkonsistent.	○	○	○	○	○
6.3 Ich denke, dass ich dieses System häufig nutzen möchte.	○	○	○	○	○

Allgemeine Fragen

7.1 Welche Art von Texten schreiben Sie normalerweise?

	Täglich (6)	Zweimal die Woche (5)	Einmal die Woche (4)	Einmal im Monat (3)	Einmal im Jahr (2)	Seltener (1)
7.2 Geben Sie an, wie oft Sie kreativ schreiben:	○	○	○	○	○	○

	Immer (6)	Meistens (5)	Etwa Hälfte der Zeit (4)	Gelegentlich (3)	Selten (2)	Nie (1)
7.3 Geben Sie an, wie oft Sie beim Schreiben Musik oder andere Geräusche hören:	○	○	○	○	○	○

Falls ja, welche Art von Musik oder Geräuschen:

Demografisches Profil

Alter:

Geschlecht:

Höchster Abschluss:

Beruf:

Appendix D

Story Starters

Inspirationen für den Schreibbeginn:

- Als ich durch den Wald hinter unserem Garten ging, glaubte ich ein Flüstern zu hören, aber es war niemand in Sicht. ...

- Als wir dort ankamen, war die Schatzkammer leer. ...

- Er konnte sich nicht erinnern, wie er dorthin gekommen war, beschloss aber, dem Weg zu folgen, um zu sehen, wohin er führte. ...

- In einer versteckten Stadt, in der jeder eine einzigartige Superkraft hat, entdeckt Zahra, dass sie die Einzige ist, die normal ist. ...

Appendix E

Iteration 4 - Informed Consent Form

Formular zur informierten Zustimmung

Nutzerstudie: Kreatives Schreiben und Auditive Reize

Studienleiter: Nestan Gujejiani
RWTH Aachen University
E-Mail: nestan.gujejiani@rwth-aachen.de

Zweck: Ziel dieser Studie ist es, die Auswirkungen der akustischen Begleitung durch SoundMuse auf den kreativen Schreibprozess zu beobachten und zu analysieren. Die Ergebnisse dieser Untersuchung werden in eine Bachelorarbeit einfließen.

Verfahren: Die Studie wird in Präsenz durchgeführt. Du wirst gebeten, einen kreativen Text zu schreiben.
Falls du zustimmst, wird währenddessen eine Bildschirmaufnahme von dir bei der Interaktion mit der Benutzeroberfläche aufgezeichnet. Deine Stimme wird dabei ebenfalls aufgezeichnet. Du wirst benachrichtigt, wenn die Aufnahme beginnt.
Am Anfang der Untersuchung ist die Studienleiterin für technische Unterstützung im Raum anwesend. Die letzten 10-15 Minuten wird den Schreibprozess von der Studienleiterin beobachtet.
Nach dem Schreibeprozess wirst du gebeten, einen Fragebogen auszufüllen und anschließend an einem kurzen Interview teilzunehmen. Falls du zustimmst, wird das Interview ausgezeichnet.
Die gesamte Untersuchung soll ca. 120 Minuten dauern. Du kannst jederzeit eine Pause machen oder die Studie abbrechen.

Vertraulichkeit: Als Teilnehmer*in bleibst du anonymisiert. Die gesammelten Informationen werden ausschließlich für wissenschaftliche Zwecke im Rahmen einer Bachelorarbeit und möglicher Folgepublikationen verwendet.

Kosten und Entschädigungen: Es wird keine Entschädigung gezahlt.

☐ Ich habe die Informationen in diesem Formular gelesen und verstanden.

☐ Die Informationen in diesem Formular wurden mir erklärt.

☐ Ich bin mit der Bildschirmaufnahme während des Schreibens einverstanden.

☐ Ich bin mit der Aufzeichnung des Interviews einverstanden.

Name des Teilnehmers	Unterschrift des Teilnehmers	Datum
	Unterschrift der Studienleiterin	Datum

Bibliography

[1] Mohamad Awada, Burcin Becerik-Gerber, Gale Lucas, and Shawn Roll. Cognitive Performance, Creativity and Stress Levels of Neurotypical Young Adults under Different White Noise Levels. *Scientific Reports*, Vol. 12(1), August 2022. doi.org/10.1038/s41598-022-18862-w.

[2] Alison Baverstock and Jackie Steinitz. What Makes a Writer? How Do Early Influences Shape, and Working Habits Develop, Those Who Write? *Publishing Research Quarterly*, Vol. 35(3):327–351, May 2019. doi.org/10.1007/s12109-019-09660-w.

[3] Margaret A. Boden. Computer Models of Creativity. *AI Magazine*, Vol. 30(3): 23–34, September 2009. doi.org/10.1609/aimag.v30i3.2254.

[4] John Brooke. SUS: A Quick and Dirty Usability Scale. In Patrick W. Jordan, Bruce Thomas, Ian L. McClelland, and Bernard Weerdmeester, editors, *Usability Evaluation In Industry*, pages 189–194. Taylor and Francis, November 1996.

[5] Cheng Chen, Susanne Weyland, Julian Fritsch, Alexander Woll, Claudia Niessner, Alexander Burchartz, Steffen E. C. Schmidt, and Darko Jekauc. A Short Version of the Physical Activity Enjoyment Scale: Development and Psychometric Properties. *International Journal of Environmental Research and Public Health*, Vol. 18(21), October 2021. doi.org/10.3390/ijerph182111035.

[6] Sebastian Deterding, Jonathan Hook, Rebecca Fiebrink, Marco Gillies, Jeremy Gow, Memo Akten, Gillian Smith, Antonios Liapis, and Kate Compton. Mixed-Initiative Creative Interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 628–635, New York, NY, USA, May 2017. Association for Computing Machinery. doi.org/10.1145/3027063.3027072.

[7] Katherine E. Eskine, Ashanti E. Anderson, Madeline Sullivan, and Edward J. Golob. Effects of Music Listening on Creative Cognition and Semantic Memory Retrieval. *Psychology of Music*, Vol. 48(4), November 2018. doi.org/10.1177/0305735618810792.

[8] Lucas Ferreira and Jim Whitehead. MTG: Context-Based Music Composition for Tabletop Role-Playing Games. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume Vol. 13, pages 286–288, June 2021. doi.org/10.1609/aiide.v13i1.12914.

[9] Frederica Gonçalves, Diogo Cabral, Pedro Campos, and Johannes Schöning. I Smell Creativity: Exploring the Effects of Olfactory and Auditory Cues to Support Creative Writing Tasks. In Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler, editors, *Human-Computer Interaction - INTERACT 2017*, volume Vol. 10514, pages 165–183. Springer International Publishing, September 2017. doi.org/10.1007/978-3-319-67684-5_11.

[10] Otmar Hilliges, Lucia Terrenghi, Sebastian Boring, David Kim, Hendrik Richter, and Andreas Butz. Designing for Collaborative Creative Problem Solving. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*, C&C '07, pages 137–146, New York, NY, USA, June 2007. Association for Computing Machinery. doi.org/10.1145/1254960.1254980.

[11] Susan A. Jackson and Herbert W. Marsh. Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport & Exercise Psychology*, Vol. 18(1):17–35, 1996. doi.org/10.1123/jsep.18.1.17.

[12] Junghyun Kim, Tomoko Hashida, Tomoko Ohtani, and Takeshi Naemura. Effects of Auditory Feedback for Augmenting the Act of Writing. In *Proceedings of the 3rd Augmented Human International Conference*, AH '12, New York, NY, USA, March 2012. Association for Computing Machinery. doi.org/10.1145/2160125.2160138.

[13] Christiane Kirsch, Todd Lubart, and Claude Houssemand. *Hybrid Approach to Creativity: Arts, Science, and Everyday Life*. LAP LAMBERT Academic Publishing, 01 2016. ISBN 978-3-659-85355-5.

[14] Aidan Mahady, Marcel Takac, and Alexander De Foe. What is Autonomous Sensory Meridian Response (ASMR)? A Narrative Review and Comparative Analysis of Related Phenomena. *Consciousness and Cognition*, Vol. 109, March 2023. doi.org/10.1016/j.concog.2023.103477.

[15] Aamir Saeed Malik and Hafeez Ullah Amin. Chapter 1 - Designing an EEG Experiment. In Aamir Saeed Malik and Hafeez Ullah Amin, editors, *Designing EEG Experiments for Studying the Brain*, pages 1–30. Academic Press, 2017. doi.org/10.1016/B978-0-12-811140-6.00001-1.

[16] Laura H. Malinin. Creative Practices Embodied, Embedded, and Enacted in Architectural Settings: Toward an Ecological Model of Creativity. *Frontiers in Psychology*, Vol. 6, January 2016. doi.org/10.3389/fpsyg.2015.01978.

[17] Rebecca Marrone, David Cropley, and Kelsey Medeiros. How Does Narrow AI Impact Human Creativity? *Creativity Research Journal*, July 2024. doi.org/10.1080/10400419.2024.2378264.

[18] Serena Mastria, Sergio Agnoli, and Giovanni Emanuele Corazza. How Does Emotion Influence the Creativity Evaluation of Exogenous Alternative Ideas? *PLOS ONE*, Vol. 14(7):1–16, July 2019. doi.org/10.1371/journal.pone.0219298.

[19] Ravi Mehta, Rui(Juliet) Zhu, and Amar Cheema. Is Noise Always Bad? Exploring the Effects of Ambient Noise on Creative Cognition. *Journal of Consumer Research*, Vol. 39(4):784–799, March 2012. doi.org/10.1086/665048.

[20] Precious Mones and Jessica Massonnié. What Can You Do with a Bottle and a Hanger? Students with High Cognitive Flexibility Give More Ideas in the Presence of Ambient Noise. *Thinking Skills and Creativity*, Vol. 46, December 2022. doi.org/10.1016/j.tsc.2022.101116.

[21] Julien Nelson and Jérôme Guegan. "I'd like to be under the sea": Contextual Cues in Virtual Environments Influence the Orientation of Idea Generation. *Computers in Human Behavior*, 90:93–102, January 2019. doi.org/10.1016/j.chb.2018.08.001.

[22] Katherine O'Toole and Emőke Ágnes Horvát. Extending Human Creativity with AI. *Journal of Creativity*, Vol. 34(2), August 2024. doi.org/10.1016/j.yjoc.2024.100080.

[23] Henry Putney, Sarah Silver, Paul J. Silvia, Alexander P. Christensen, and Katherine N. Cotter. Why does Creativity Foster Well-Being? Autonomy, Competence, and Relatedness during Everyday Creative Activities. *Journal of Research in Personality*, Vol. 113, December 2024. doi.org/10.1016/j.jrp.2024.104552.

[24] Jacob M. Rigby, Duncan P Brumby, Sandy J. J. Gould, and Anna L Cox. Development of a Questionnaire to Measure Immersion in Video Media: The Film IEQ. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '19, pages 35–46, New York, NY, USA, June 2019. Association for Computing Machinery. doi.org/10.1145/3317697.3323361.

[25] Christina Rowell and Amy Flick. "I could get lit to Madonna:" Soundscapes & the First Year Writer's Composing Process. *Computers and Composition*, Vol. 54, December 2019. doi.org/10.1016/j.compcom.2019.102513.

[26] Joon Gi Shin, Janin Koch, Andrés Lucero, Peter Dalsgaard, and Wendy E. Mackay. Integrating AI in Human-Human Collaborative Ideation. In *Extended*

Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, CHI EA '23, New York, NY, USA, April 2023. Association for Computing Machinery. doi.org/10.1145/3544549.3573802.

[27] John R. Smith, Dhiraj Joshi, Benoit Huet, Winston Hsu, and Jozef Cota. Harnessing A.I. for Augmenting Creativity: Application to Movie Trailer Creation. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, pages 1799–1808, New York, NY, USA, October 2017. Association for Computing Machinery. doi.org/10.1145/3123266.3127906.

[28] Ernesto Villalba. On creativity : Towards an Understanding of Creativity and its Measurements. *Publications Office of the European Union*, 2008. doi.org/10.2788/2936.

[29] Ruijun Wu. Exploring the Emotional, Cognitive, and Background Music Use: Their Influences on Employee Creativity. In *2024 International Conference on Language Technology and Digital Humanities (LTDH)*. Institute of Electrical and Electronics Engineers, July 2024. doi.org/10.1109/LTDH64262.2024.00034.

[30] Xinyao Xiao, Junying Tan, Xiaolin Liu, and Maoping Zheng. The Dual Effect of Background Music on Creativity: Perspectives of Music Preference and Cognitive Interference. *Frontiers in Psychology*, Vol. 14, October 2023. doi.org/10.3389/fpsyg.2023.1247133.

[31] Chuyi Zhou, Chunlei Chai, Jinlei Shi, and Jing Liao. Creativity Performance Following Exposure to Music. In *2020 13th International Symposium on Computational Intelligence and Design (ISCID)*, pages 322–325. Institute of Electrical and Electronics Engineers, December 2020. doi.org/10.1109/ISCID51228.2020.00078.

Index

Typeset April 17, 2025