https://projects.id.tue.nl/id/xQIKxz

A/Muse: Designing for shared autonomy over AI-based music recommendations

Berg, J. v.d.[1], Blom, M.O.[1], Meijer, F.A.[1], Liu, X.[1]
[1] University of Technology, Eindhoven, Netherlands

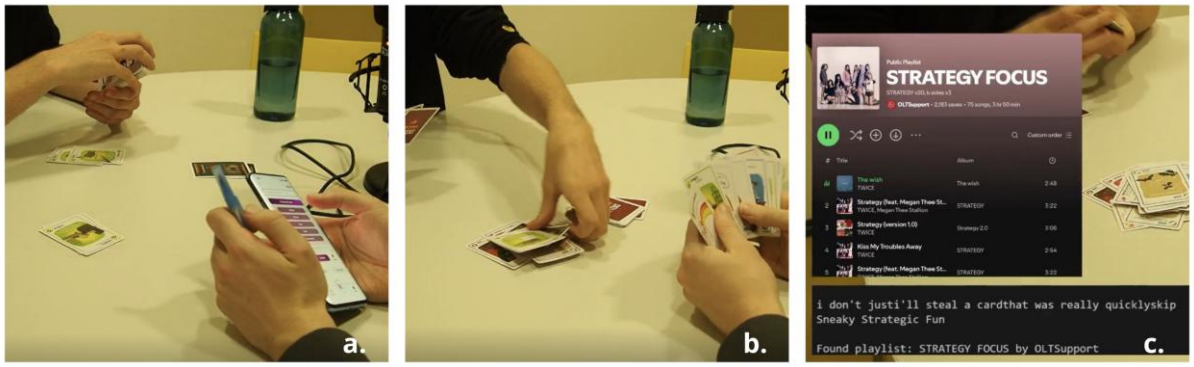


Fig. 1. Overview of the A/Muse concept. (A) The user initiates the context and starts the app. (B) The participants play their game as normal while conversation snippets get picked up. (C) The system transforms the contextual input into an ambiance fitting playlist title that is then looked up and played.

Abstract: This paper investigates the impact of an AI-driven, context-based music selection application on social dynamics and user autonomy in tabletop game settings. Insights are offered about the balance between AI automation, user control, shared autonomy and immersion in social music interactions. Through the development of A/muse, a speech-based playlist selection and playback tool, an exploration is made on how automatic playlist selection influences the perceived user experience. Using a combination of Google Web Speech, Gemini 1.5 Flash and the Spotify Web AI, the conversation of the users is transcribed, analyzed on sentiment and a suitable playlist name is recommended and played locally on Spotify. Within the user test participants were asked to participate in a game of Exploding Kittens to experience the design and evaluated on this using the User Experience Questionnaire and semi-structured interviews. The preliminary findings suggest that context-based music selection can enhance the perceived game experience on several levels.

Keywords: music, social bonding, autonomy, AI recommendations

1 Introduction

Music has become ubiquitous; we encounter it anywhere and at any time and for many people music is inseparable from their daily lives. It is played in the supermarket, at the hairdresser, during work, in tv shows and at concerts. Music has a profound effect on social settings, psychological well-being, concentration and even alleviating stress and anxiety. [1] It also has the ability to move us, both physically and emotionally. [2] With technological advancements, music has made its way from the hands that play the instruments to the devices in our ears. Now, with thousands of songs at the press of a screen people often ponder or discuss what to play or let the application provide suggestions and if we do not want to choose playlists are generated automatically.

Lately, the use of artificial intelligence (AI) has made its way into our listening behavior as well. Algorithms devise what we might want to listen to based on our likings, but are also able to generate sheet music, lyrics or even music itself by means of different models and techniques. Three main classes of research emerge with regard to music; recommendation, classification and generation. [3] In the future we might not even be able to hear what music was composed with the help of AI. And your music app always has something to recommend to you. This poses the question, what happens to our listening autonomy? If 'the algorithm' keeps reinforcing itself, it might become difficult to grasp why something is suggested, and it might even limit our musical exploration. [4]

Innovations quickly change the musical landscape, but music has always influenced us. Music offers the ability to convey and create atmosphere and influence how we feel [5] and can impose many emotions. In turn, humans are good at perceiving these emotions in music as shown by Cowen in an interactive music map. [5, 6] It is believed that interpersonal synchronization and the release of endorphins are at the heart of what makes music facilitate connection, even while passively listening. [7] But with the high subjectivity of music preferences social frictions can occur. [8] People in social settings want to belong, relating to relatedness, but if there is a mismatch in music preference this feeling can quickly deteriorate. Even though talking about music preferences is shown to increase cohesion, [7] autonomy and pick over the music at social gatherings is often a topic of discussion and tension.

In this project we explored, designed and synthesized a speech-based music queueing application, that implements AI to automatically pick and play ambiance-fitting music. Specifically, the social setting of table-top games was chosen. This context introduces dynamic social and emotional aspects, making it an interesting practice to research the implementation of democratized, automatic music selection. Therefore, the research question is; *What is the effect of context-based, automatic music selection on the perceived game experience?* This design report briefly discusses the state-of-the art of musical influences on social settings and its impact on expression and perceived atmosphere. Thereafter, an overview of the design process of our speech-based auto-playing application, A/Muse, is shown, highlighting difficulties, opportunities and argumentation for design choices. The concept and empirical results of this pilot study aim to contribute to the fields of service design and applied psychology to probe how AI can be applied to retain user autonomy by using a context-based system that streamlines music streaming in social settings.

2 Related work

Music plays a crucial role in daily life, serving various psychological and social functions. It aids in emotion regulation, social cohesion, and communication. For instance, individuals use music to manage their moods, strengthen group bonds, and engage in interaction. Understanding the multifaceted roles of music can enhance our comprehension of human behavior and social dynamics [10].

Background music, a subsection of music, can be applied to many situations to provide a comfortable atmosphere and immerse the listeners in their environment. Research indicates that background music can influence

attention and concentration. A study found that music with lyrics significantly impairs concentration, making instrumental music preferable for maintaining attention and performance in work environments [11]. Tabletop games and conversations are two settings that can benefit greatly from music playing in the background while the participants are engaged in their activity.

2.1 Background music in tabletop games

A study on the effect of soundtracks on tabletop game experiences showed an increase in enjoyment, tension and atmosphere surrounding the game experience by the participants [12]. Furthermore, the research discusses how these insights could guide the design of future tabletop games and their accompanying soundtracks, as well as inform further research into player experiences [12].

The context of tabletop games has an overlap with the research done into video games [13]. By providing sound effects for actions and dynamic music based on events, immersion can be increased and thus the enjoyment of participants [13]. This extends, though to a lesser extent, to tabletop games, studies have shown. To further increase this immersion, they proposed to make the audio interaction based, such as making sound when certain actions are taken, or elements of the game are interacted with. An example described with the game chess, is to make the music based on the first moves, sound effects for subsequent moves and special sound effects for checking an opponent. [13] Overall, studies show there tend to be an increase in enjoyment, tension and atmosphere for tabletop games when combined with background music and sound effects.

One specific context where immersion and atmosphere specifically are of great importance is role-playing games. Research performed by Ferraira et al. presents *Bardo Composer*, an AI system that generates background music for tabletop RPGs based on player speech and emotional cues. It uses speech recognition and a neural model to create music that matches the game's emotional tone, with a study showing it effectively conveys emotions. This approach shares similarities with dynamic background music in video games, where music adapts to gameplay. However, tabletop games typically lack real-time interaction with the environment, so *Bardo Composer* addresses this by using player speech to select music, creating a more responsive auditory experience in a non-digital setting [14].

2.2 Background music in conversation

Conversations are a different situation, with different criteria and atmosphere. It has no pre-determent genre or set topic to base the music selection around and can change more dramatically. Research has also been done into the impact of a variety of music on spoken word conversations. The results indicated that both higher music complexity and the presence of lyrics negatively impacted spoken-word recognition. Spoken-word recognition becomes more difficult with high-complexity background music and even more so when music contains sung lyrics. While complexity increases masking effects, lyrics contribute more significantly to interference by adding linguistic information. The impact is strongest at lower signal-to-noise ratios, meaning background music complexity matters more in louder environments like bars than in quieter ones like restaurants [15]. A study found that classical, jazz, and popular music enhanced perceived atmosphere and increased spending on main meals, suggesting that background music can improve customer experience and boost restaurant sales [2]. Additionally, self-reported listening difficulties further reduce word recognition, highlighting individual differences in speech perception in noisy settings [15]. These findings suggest that in social environments like restaurants, cafés, and bars, selecting background music with lower complexity and without lyrics may facilitate better conversation and speech comprehension.

Background music is also commonly used in educational and workplace settings to enhance concentration, reduce stress, and create a comfortable atmosphere. A study by found that its impact on productivity varies based on individual traits, task type, and environment—while some benefit from improved focus and efficiency, others find it distracting, highlighting the role of personal and contextual factors [1].

3 Design implementation

The chosen concept for this project is a context-based, AI-powered music playing service. This system plays tailored playlists while playing table-top games, considering both the game being played and the conversational context of the participants. The design is implemented through the underlying processes outlined in the system architecture, figure 2, and is intended as a mobile application, or extension to existing services. The system provides shared autonomy, in an automated matter, as music is selected based on the conversational context. While playing the system continues to listen to the conversations. When the atmosphere has changed based on the conversation a new song/playlist is played. See figure 1 for an overview of the concept in practice. The explanatory video can be found in the deliverable folder.

3.1 System

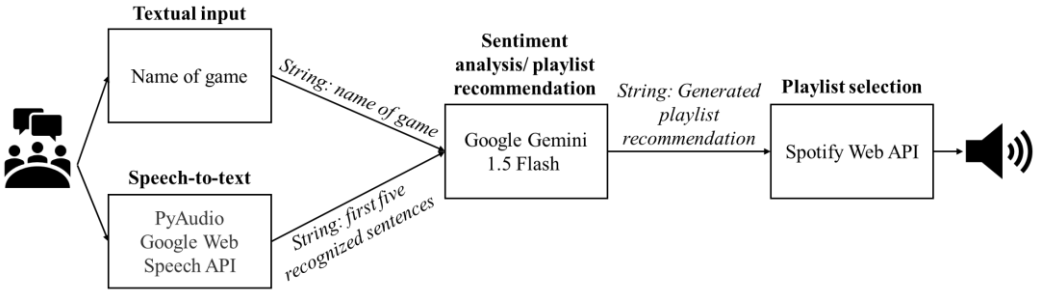


Fig. 2. System architecture with integrated models. Schematic explaining the background process from speech in conversation to background music.

The system consists of three core functionalities that use contextual data to stream ambience fitting music. See appendix C for the code link.

Automatic Speech Recognition: Google Web Speech
The python speech-recognition library [16] provides the base, enabling many different speech recognition models. The base Google Web Speech API provides fast inference, and the system is configured to save streaming ASR results in a string of five sentences to be used in determining sentiment and context. [17] To increase reliability only recognized sentences will be used.

Context to playlist: Gemini 1.5 Flash
Google Gemini 1.5 Flash takes the speech data and the context input, i.e. the game that is being played, and devises a playlist that captures the sentiment and activity of the input. Gemini 1.5 Flash [18], based on Google's transformer model, provides a powerful way to transform the contextual input into a fitting playlist name. The high token size in addition to its lighting fast responses were very suitable for this project.

It was extremely important that Gemini's output could directly be used as input into the Spotify Web API. Therefore, several (system) prompts were tested. The final system prompt and instructions, appendix D, focusses on considering the context and speech input, devising sentiment and context and creating a suitable maximum five-word output, eliminating unrelated utterances of Gemini.

Playlist selection and streaming: Spotify Web API
To select a playlist for background music, the Spotify Web API [19] was used. The recommendation generated as output from Gemini serves as an input, and a playlist with a similar name is selected from the collection of playlists created by Spotify's users. Once a suitable playlist is identified, the model initiates playback locally on the device running the code.

3.2 Process
During the development of both the concept and the prototype, a number of significant adjustments have been made. Below, the most impactful changes have been described:

Generated music to Spotify Web API
In the beginning of the conceptualization, the project's aim was to generate audio based on spoken text. Various experiments have been taken place to look at different manners to go from this speech-to-text recorded text to generated audio. To achieve this, models like Meta's MusicGen (audiocraft)[20] and Google's MusicLM [21] have been tested. While the result of these generated music files was interesting to listen to, the time it took to generate was quite long compared to the length of the music file, which introduced the necessity for high compute costs or looping or slowing of the generated music fragments. Besides this, discussion was raised whether generated audio was even necessary in this stage or whether utilizing and existing audio database/library was more useful. Utilizing an existing database offers several advantages, including higher audio quality, broader search capabilities, and a significantly reduced time to begin playback. Based on these considerations, the decision was made to integrate the Spotify API for audio selection, using the extensive library of existing playlists already available on the platform.

Experimentations with Spotify
Two methods were experimented with to play a playlist on Spotify. The first method involved locally launching Spotify on a laptop, entering a playlist name, and playing it using PyAutoGUI [22]. This approach did not require the Spotify Web API. However, a major limitation of this method was that the exact playlist name needed to be known and generated by the LLM in advance. To address this, the Spotify Web API was adopted, as it allows for searching existing playlists without requiring the exact name beforehand.

During iterations on the code, it was discovered that the playlist search is sensitive and only functions correctly when the input uses capitalized first letters for each word and avoids special characters. This resulted in adjustments to the LLM prompt to ensure the playlist search was successful in most cases.

Experimentations with Large Language Models
Our concept uses an LLM to transform spoken text into a suitable input for the Spotify web API. When testing the ability to create prompts for Musicgen with Ollama it became clear that a 'middle-man' might add a lot of compute and was not very accurate. Testing various web-based models created insight into the capabilities of different models. At first the use of Llama 3 8B seemed a logical choice, but this model was difficult to use for our use-case without cloud API. Finally, Gemini (1.5 Flash) was able to be tuned to send the desired output with high accuracy. This helped tremendously when hooking into the Spotify web API.

Gradio to Figma
To run the model and conduct an experiment in the form of a user test, various approaches were considered for the front-end development of a user interface. Initially, as can be seen in Figure 3. It was created using Gradio [23], as it allows for running the model in the background of the application. When considering the feasibility of integration of the model into a UI, the decision was made to develop a prototype in Figma, figure 4, while the

model is functioning in the background on a laptop controlled by the researcher. In this way more attention could be given to a useful, more intuitive and aesthetically pleasing UI while the functionality of this interface and the connection to the model was simulated using the Wizard-of-Oz method [24] on a laptop by the researcher. The attached video demonstrates how the interface works in combination with the model.

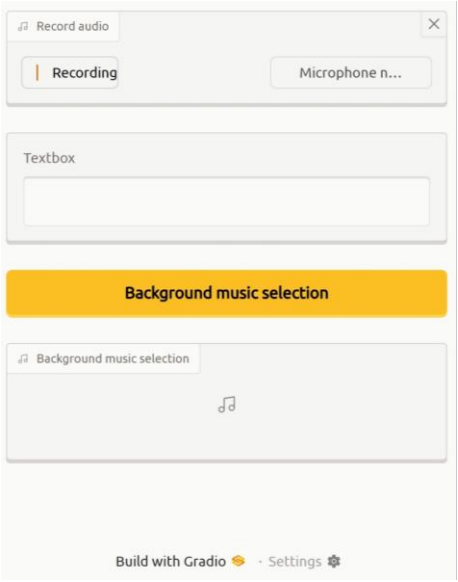


Fig. 3. First Application created using Gradio

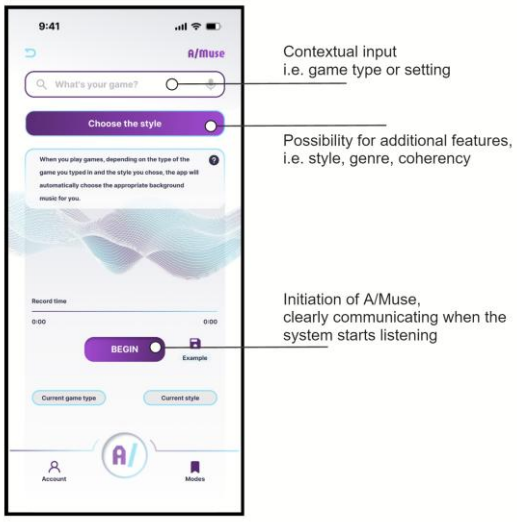


Fig. 4. Core touchpoints of the A/Muse interface. After inputting the context and rules and the system is initiated, the system will run in the periphery of the social setting. See appendix E.

4 Methods

To assess the final developed model in combination with the user interface prototype, a user test was conducted. The main goal was to evaluate the user experience and identify areas for improvement. The prototype, developed using Figma, served as the front-end interface through which participants interacted. Simultaneously, a program operated by the researcher ran in the background to manage music playback and simulate real-world functionality. Apart from the user interface, the model is fully functional, listening and the user and selecting appropriate music for the context.

Participants
A total of five participants (3 Female, 2 Male) took part in the study. Participants were selected based on the criteria for being students at TU/e and were recruited through personal networks. Additionally, participants were required to have proficiency in English to ensure the model works appropriately.

Procedure
The experiment consists of three different phases, as can be found in the full protocol in appendix F. After the informed consent form was signed, appendix G, the game of Exploding Kittens was explained to new players. This game was chosen as it is quick to learn, and suitable for the relevant time of the experiment.

The first phase of the study involved participants interacting with the user interface prototype, filling in the name of the game. Afterward, they play a game of Exploding Kittens [25] and engage in conversation with the other players, who are team members acting as conversational partners. After interacting with the prototype, participants played a game with a team member for approximately five minutes, during which they were instructed to converse naturally. Throughout this interaction, the model operated in the background, listening to the conversation and selecting a playlist based on the dialogue. This playlist provided background music during the session. Due to the limitations of the code at the time, the microphone is too sensitive when the music is playing, so after a song is played it will start up again and look for relevant music based on the conversation. In this time creating a short pause in the music.

After five minutes the participant is asked to fill in the User Experience Questionnaire [26], which rates the user experience of interactive products on 26 user experience dimensions, using a Likert scale from a quantitative data perspective.

The final stage of the study consisted of semi-structured interviews. Participants were asked to elaborate on their responses to the User Experience Questionnaire. Additionally, they were invited to rate their overall experience, suggest opportunities for improvement, and provide ideas for applying the system in other scenarios.

Data collection and analysis
During each session, one team member was responsible for documentation. The data collected through the User Experience Questionnaire (UEQ) was analyzed using the UEQ data analysis tool. By combining these quantitative results with the session notes from the interview, preliminary findings were drawn about user to provide insights on user experience and future improvements.

5 Results

The test collected scores from five participants on 26 UEQ dimensions and their opinions. The mean score is shown in Table 1. The high mean score in the positive dimension and the low mean score in the negative dimension together indicate a positive disposition of the results.

Table 1. Mean scores of dimensions

Dimensions	Mean score (1-7)	Dimensions	Mean score (1-7)
annoying/enjoyable	6.2	unlikable/ pleasing	5.9
not understandable/ understandable	6.2	usual/ leading edge	4.8
creative/ dull	2.2	unpleasant/ pleasant	5.8
easy to learn/ difficult to learn	1.6	secure/ not secure	3.6
valuable/ inferior	2.6	motivating/ demotivating	2.9
boring/ exciting	5.5	meets expectations/ does not meet expectations	3.3
not interesting/ interesting	5.8	inefficient/ efficient	4.3
unpredictable/ predictable	2.5	clear/ confusing	2.4
fast/ slow	2.4	impractical/ practical	5.1
inventive/ conventional	2.4	organized/ cluttered	3.8
obstructive/ supportive	4.5	attractive/ unattractive	2.1
good/ bad	2.1	friendly/ unfriendly	2.1
complicated/ easy	5.9	conservative/ innovative	6

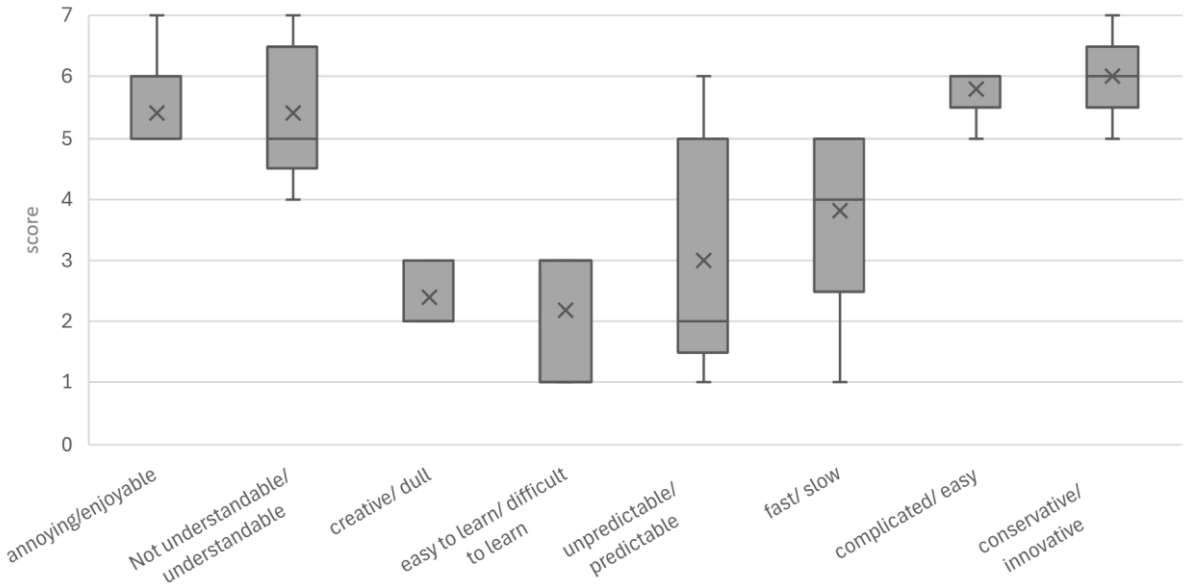


Fig. 5. Boxplot of some interesting mean scores of eight UEQ dimensions

Based on the results of the participants' questionnaires, the mean score for each dimension of this experiment can be obtained, as shown in appendix G. Select the highest (mean score ≥ 5.4) and lowest (mean score ≤ 2.4) mean scores' dimensions, as shown in Figure 5.

From this figure, it becomes clear that participants generally found that A/muse was enjoyable (“enjoyable” has a mean score of 6.2), and other dimensions about motion feeling show the same result. Besides, being understandable (“understandable” has a mean score of 6.2) also can be seen, and other dimensions about usability show the same result. What’s more, participants agree that the design is creative (“dull” has a mean score of 2.2).

In the interviews, the following issues were identified: the desire for music to be mixed in with each other and not to have periods between music (P1), the desire for the length of music to be controllable (P1), and the desire for greater precision (P3).

It is worth noting that there were divergent opinions on the question of randomness. Some participants found the music unpredictable (P1, P2, P5), with some of them finding this unpredictability to be a bad experience (P1, P2) and others finding it interesting (P5). There was also a participant who found it predictable (P3).

With regard to the music and mood, some participants felt that adjusting the music to the atmosphere could sometimes be counterproductive (P3, P4), and that adjusting it to one's own wishes would be more comfortable (P1, P3). In addition, in terms of usage scenarios, most participants expressed a preference for the period used in social scenarios (such as parties) (P1, P2, P3), while it was considered less suitable for work scenarios (P3).

6 Discussion

In this study, our main goal was to explore the effect of context-based, automatic music selection on the perceived game experience. We designed a system that selects background music based on speech content and sentiment analysis and an interface that allows the specification of music styles. The results show that the whole process of designing the A/Muse music system can enhance the game atmosphere and is easy to use (the mean UEQ scores for enjoyable and ease of use are 6.2 and 5.9 respectively). This provides indication that the theoretical basis in the literature that providing dynamic music based on events enhances player immersion.

6.1 Limitations

There are also some shortcomings and limitations in this design. Some music is not suitable for the scene, and there is a lack of continuity between pieces of music. The current function relies on predefined logic (static playlists from the Spotify API) rather than dynamic generation (such as the music generation model Meta MusicGen). Although the problem of long latency times has been effectively solved, it has brought about a new problem for participants considering that some music is not suitable, and there is a lack of continuity between pieces of music. Perhaps a mixing mode would be better, which would combine a music library with lightweight music generation to make up the transitional parts between pieces of music.

Ethical considerations should be made with regards to the use of speech data and what that might mean if used by ill-intended people. In our case the source data is only passed through as initiator, but if taken into production it should be considered how this personal data is handled to ensure the user has control over their data.

Unpredictability in the music output brings about differences in the need for a sense of control. As mentioned in the results, different testers have different understandings and attitudes towards the unpredictability of music selection and even show completely opposite attitudes (such as considering it a source of surprise and considering it a bad experience). This difference reflects the varying needs of users for a sense of control: in social games, moderate randomness can enhance immersion (such as D&D background music adapting to changes in the situation), but the lack of transparent rules can undermine trust (P2: "The AI decision-making logic needs to be clear"). How to achieve controllable randomness will be a newly emerging question, which is also consistent with the ISO-principle of guiding user emotions through automation [5].

Technical limitations in execution and validating the concept arose as well. While testing speech recognition across different devices it became clear that the quality and internal noise reduction affected the quality of speech recognition. The breath of speech recognition options also showed difficulty translating to real-life settings when finding a balance between speed, continuous streaming and accuracy. Furthermore, when API requests were sent to Spotify, the connection to the server sometimes failed, resulting in no music being played. Limitations of the microphone and room audio resulted in unclear audio input, which sometimes caused a delay in sound pickup. These testing details may also have affected the test results. In general, this test mainly focused on gaming scenarios and only had 5 testers. Therefore, the results are limited and more comprehensive testing with more people and scenarios is needed in the future. Furthermore,

Other limitations regarding the Spotify Web API are that the selected playlists are created by Spotify's users. Which means that they might not accurately represent the intended result from the LLM. Besides this the output of the API is influenced by the listening behavior of the host of the API, which also has interfered with the playlist selection. This could for example be seen by the amount of Dutch music recommended.

Lastly, the model and the user interface were not connected but simulated by the researcher. Through connecting this together the user experience might have been different.

6.2 Future work

The concept and purpose of A/Muse should be further developed in several areas. Firstly, facilitating a more robust control over the music selection through UI and back-end adaptation are important to increase the effectiveness of the concept. Specifically, specified genres and multi-modal restrictions for music recommendation from Spotify's end will greatly enhance the robustness of the system.

A/muse's degree of user autonomy could be improved as well. Some participants expressed a desire to be able to control the length of music (P1) and to be able to adjust the music style according to their personal preferences rather than the mood (P1, P3). Therefore, it may be possible to increase A/muse's degree of user autonomy, for example, by providing a music instant switch button and more music style options. But care should be taken with keeping the user's shared control over the system present.

Additional context adaptation and further testing should be done to define the concept's impact. Currently, A/muse is focused on facilitating atmosphere based on leisure conversation and context. If work mode is considered, is there a need to change the functions? Furthermore, other modes and contexts might be considered, i.e. improv, theater, story-building or storytelling.

Interface and program connection and platform expansion. Limited by the API, the interface design and the program are not actually connected on the prototype now, which will be a research direction in the future. In addition, the current A/Muse is running on a mobile phone, and whether it needs to be used on other platforms and how to achieve this is also a point of future exploration.

Lastly, it might be important to consider our design to have a certain emotional guiding function, that is, to perceive anger and transform it into calmth or happiness. This could create an initiation towards a mental health service [5].

7 Conclusion

This design of A/Muse aims to enhance the immersion and user experience through context-based, automatic music selection. The preliminary results gathered in a table-top game setting demonstrate A/Muse's potential as an interactive music selection system providing shared autonomy to the users. Participants mentioned the designed system to be attractive, interesting and creative, reinforcing the existing role of dynamic music in player immersion.

Overall, A/muse creates an immersive and enjoyable experience. However, there is room for further adaptability and enhancement. Some limitations were identified, such as occasional mismatches between music and scenario transitions, as well as the unpredictability of the selected music. Addressing these issues could improve user experience and reliability of the incorporated AI. Future work could explore refining the algorithms that select music, improving the continuity in between transitions and expanding the application of A/muse to contexts beyond gaming. Additionally, by integrating user-controllable features in the app, a more personalized experience can be offered.

The design of A/muse provides a contribution to the development of AI in intelligent ambient music systems, enhancing immersion and atmosphere in a social environment.

Acknowledgements

We acknowledge the course coordinators and TA's for knowledge, guidance, and feedback during this project. Additionally, we would like to thank the participants that were willing to perform in the usertest. We acknowledge the use of generative AI such as chatGPT and Claude.ai for troubleshooting during the coding and integration phase. Additionally, we acknowledge the use of DeepL for translating some of the professional terms from Chinese to English. No text or code was generated by AI.

References

[1] Kalkumbe K. HARMONIOUS RHYTHMS: EXPLORING THE IMPACT OF MUSIC ON PRODUCTIVITY IN EDUCATIONAL AND WORKPLACE SETTINGS. *Proceeding of International Conference on Social Science and Humanity*. 2024;1(1):116-122. doi:https://doi.org/10.61796/icossh.v1i1.10

[2] Wilson S. The Effect of Music on Perceived Atmosphere and Purchase Intentions in a Restaurant. *Psychology of Music*. 2003;31(1):93-112. doi:https://doi.org/10.1177/0305735603031001327

[3] Mycka J, Jacek Mańdziuk. Artificial intelligence in music: recent trends and challenges. *Neural Computing and Applications*. Published online November 16, 2024. doi:https://doi.org/10.1007/s00521-024-10555-x

[4] Makkaoui K. The Algorithmic Influence: How Digital Platforms Shape Music Culture | Outlook. Aub.edu.lb. Published 2024. https://sites.aub.edu.lb/outlook/2024/02/26/the-algorithmic-influence-how-digital-platforms-shape-music-culture/

[5] Hoffer M. How Music Affects Your Mind, Mood and Body. www.tmh.org. Published December 2, 2022. https://www.tmh.org/healthy-living/blogs/healthy-living/how-music-affects-your-mind-mood-and-body

[6] Cowen AS, Fang X, Sauter D, Keltner D. What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*. 2020;117(4):1924-1934. doi:https://doi.org/10.1073/pnas.1910704117

[7] Berkeley.edu. Published 2020. https://www.ocf.berkeley.edu/~acowen/music.html#

[8] Tarr B, Launay J, Dunbar RIM. Music and social bonding: “self-other” merging and neurohormonal mechanisms. *Frontiers in Psychology*. 2014;5(1096). doi:https://doi.org/10.3389/fpsyg.2014.01096

[9] Boer D, Fischer R, Strack M, Bond MH, Lo E, Lam J. How Shared Preferences in Music Create Bonds Between People. *Personality and Social Psychology Bulletin*. 2011;37(9):1159-1171. doi:https://doi.org/10.1177/0146167211407521

[10] Rentfrow PJ. The Role of Music in Everyday Life: Current Directions in the Social Psychology of Music. *Social and Personality Psychology Compass*. 2012;6(5):402-416. doi:https://doi.org/10.1111/j.1751-9004.2012.00434.x
[11] Shih YN, Huang RH, Chiang HY. Background music: Effects on attention performance. *Work*. 2012;42(4):573-578. doi:https://doi.org/10.3233/wor-2012-1410
[12] Farkas T, Denisova A, Wiseman S, Fiebrink R. The Effects of a Soundtrack on Board Game Player Experience. Published online April 29, 2022. doi:https://doi.org/10.1145/3491102.3502110
[13] Cakmak D, Maghsudi S, Perez Liebana D, et al. Computational Creativity for Game Development 1 Executive Summary. *Report from Dagstuhl Seminar*.:24261. doi:https://doi.org/10.4230/DagRep.14.6.130
[14] Ferreira LN, Lelis LHS, Whitehead J. Computer-Generated Music for Tabletop Role-Playing Games. arXiv.org. doi:https://doi.org/10.48550/arXiv.2008.07009
[15] Scharenborg O, Larson M. The Conversation Continues: the Effect of Lyrics and Music Complexity of Background Music on Spoken-Word Recognition. *Research Repository (Delft University of Technology)*. Published online September 2, 2018. doi:https://doi.org/10.21437/interspeech.2018-1088
[16] Zhang (Uberi) A. SpeechRecognition: Library for Performing Speech recognition, with Support for Several Engines and APIs, Online and offline. PyPI. Published 2023. https://pypi.org/project/SpeechRecognition/
[17] IANCU B. Evaluating Google Speech-to-Text API's Performance for Romanian e-Learning Resources. *Informatica Economica*. 2019;23(1):17-25. Accessed January 30, 2025. https://ideas.repec.org/a/aes/infoec/v23y2019i1p17-25.html
[18] https://drive.google.com/viewerng/viewer?url=https://storage.googleapis.com/deepmind-media/gemini/gemini_v1_5_report.pdf.
[19] Content Restricted. Spotify.com. Published 2025. Accessed January 30, 2025. https://developer.spotify.com/documentation/web-api.
[20] Copet J, Kreuk F, Gat I, et al. Simple and Controllable Music Generation. arXiv.org. doi:https://doi.org/10.48550/arXiv.2306.05284
[21] Kreuk F, Synnaeve G, Polyak A, et al. AudioGen: Textually Guided Audio Generation. *arXiv (Cornell University)*. Published online September 30, 2022. doi:https://doi.org/10.48550/arxiv.2209.15352
[22] Welcome to PyAutoGUI's documentation! — PyAutoGUI documentation. pyautogui.readthedocs.io. https://pyautogui.readthedocs.io/
[23] Meena Devii Muralikumar, McDonald DW. An Emerging Design Space of How Tools Support Collaborations in AI Design and Development. *Proceedings of the ACM on Human-Computer Interaction*. 2025;9(1):1-28. doi:https://doi.org/10.1145/3701181
[24] Dahlbäck N, Jönsson A, Ahrenberg L. Wizard of Oz studies. *Proceedings of the 1st international conference on Intelligent user interfaces - IUI '93*. Published online 1993. doi:https://doi.org/10.1145/169891.169968
[25] https://www.explodingkittens.com/products/exploding-kittens-original-edition.
[26] Schrepp M, Hinderks A, Thomaschewski J. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. *Design, User Experience, and Usability Theories, Methods, and Tools for Designing the User Experience*. Published online 2014:383-392. doi:https://doi.org/10.1007/978-3-319-07668-3_37

Appendix

Appendix A: Contributions

Within the course of the project, a shared responsibility was placed within the group regarding the contribution to brainstorming sessions, and research on useful AI models for the concept. Additionally, the report writing and reviewing was equally divided among team members. The following tasks were contributed to individually:

Joep van den Berg:
- Speach-to-text programming
- Background research
- Preforming user study and setup
- Organizing sources

Olivier Blom:
- Testing AudioGen models
- Performing user study
- Context & purpose definition
- Integration of Gemini and system integration

Febe Meijer:
- Approaching and creating process outcome LLM towards the playing of music (PyAutoGui & Spotify Web API)
- Scripting and editing of demonstrator video
- User test preparation including method, ERB approval & consent form
- Technical support during system integration

Xinyue Liu:
- Testing MMAudio model in early explorations
- Front-end interactive interface by Gradio
- From low to high fidelity of final interface design by Figma
- User test analysis

Appendix B: Individual reflections

Joep van den Berg:

My M11 courses were all chosen broadening my horizons in mind, including this course. With the goal to develop myself using AI in design projects, and allow for better communication with people well versed in AI. I did not have any experience with AI, LLM, or machine learning before this, so most of the information covered in this course was new to me. What I expected to learn was gaining a basic understanding of AI models, which to use in what context, and how to apply them to a project.

The course lived up to this expectation, and now nearing the end of the course I understand not only how to apply AI to a project, but also if it is necessary to do so. Some technological contexts just don't need AI applied to them, but knowing where it can be, is very useful.

A great source of information was provided during the lectures. Through this I learned about different kinds of LLMs, what their possible applications were, and how they were trained. This gave me insight into which of these to use in my project and how to do so. AI is currently applied to many situates, but many of these are not appropriate, so thorough research is required before coming to an AI approach in a project. Using hugging faces and taking their models, these could be combined with a bit of programming. I did have experience in programming, so applying these during the lectures where usefull exercises. Intimidation of getting in there and working with the LMM models and combining the data together with other models at first. The importance of the data flow, input, processing, and output becomes apparent in how they function, and this aids me in applying them in our current and future projects.

In our project, we created a model that would automatically play appropriate music for tabletop games based on current spoken word sentiment. It would listen until a certain amount of words were input and give an appropriate prompt of the situation. This prompt would be interpreted to a Spotify API playlist and play the song. I had a hand in the direction of applying background music to a game setting since I was interested in doing something music-related and making it function in real to create immersion. I also contributed to the user test, by participating and conducting interviews. Most of our participants liked the concept but thought the songs could be quite unpredictable. Furthermore, I incorporated the speech-to-text element in our project, but I do regret not putting in helping more to combine all the elements together. This could have given me a better technical insight into the more complex elements of applying these models and combining them. Something I must do myself now in the future.

One of the areas that require improvement for my M12 research project is my analytical reading and writing capabilities, therefore I took up the task of writing the related works and managing the references. I was in a good position to do this since I worked closely with my teammates on the parts they designed. Reading papers about new possibilities when it came to music and combining it with AI was an added benefit to start thinking about my own M12 project. If and how I could apply it to the context of musical instruments. I explored this subject in my poster regarding the pickup winder sound analysis. This is still something I'm ambitious about and for my FMP I want to create a musical instrument that is intuitive to use. I am now considering using AI to shape sound based on user input, allowing dynamic adjustments for greater personalization. This would simplify sound design, making it more intuitive and accessible, even for those without technical expertise.

Overall, this course gave me a better insight into new technological possibilities and AI integration, where and when to apply them, and getting some hands-on experience in applying them. And it also sparked new ideas and directions for future projects.

Olivier Blom:

I chose this course as I wanted to keep up with the state-of-the-art of combining UX with AI. I had used conversational models, sentiment analysis and object recognition in research projects before and even researched human-agent interaction four years ago. Immediately it became clear however, that I only knew the top-level elements of designing with AI. Throughout the course I learnt and experienced how I as a designer can utilize and challenge the boundaries of existing models and data(sets) to use the appropriate approach and tools for design purposes. An eye-opening moment was when testing CLIP classification capabilities of combining image and text modalities across multiple domains at a time.

Going further, it was a new experience for me to determine the need for AI in concepts and choose which data, modalities and model fit to reach the goal of the concept. This made me look different at the term modalities, helping me understand the importance of using the right data type to optimize your system.

After having established the need, product purpose and data/modalities it proved to be key to understand the capabilities of the chosen model. The hands-on approach of this course was very valuable, but it became even more fun when we got to think of novel implementations. It was interesting to see a group come up with the concept that I had researched during my M1.2. With this new knowledge new opportunities arise in the use and value of available data.

A key takeaway of this course is the difference in design process that designing with AI asks. It is different than the design process of interaction design in the sense that you need to take a much more data-centred approach. I learnt the importance of validating bias, data storage and considerations of in and output. All these things will influence the final user experience, but the difficulty is that a user might not notice whether the output is biased, whereas with tangible interactions faults will be much more evident.

Working with top-level machine learning and artificial intelligence made me realize the importance of changing perspectives often. The way data, input and output are handled determines the effectiveness, speed, and capabilities of your product. When developing our concept, we first opted to precisely choose which song to play, fitting the mood. However, we quickly pivoted towards a more generalistic approach that utilized the full capabilities of Spotify's search engine for example. This eliminated a step that could confuse the conversational model.

It was a personal goal to bridge the gap between AI and UX. I assumed that AI, being software, might be difficult to implement in practical day-to-day use-cases outside conversational models, or research purposes. And while I was tinkering with the LLM to create input for the Spotify API I learnt how important it is to understand the capabilities and limitations of the technology you are working with. In this case we tried to use a generalistic model for one specific task. Diving into and testing the model's capabilities made me realize the importance of being critical of the purpose and need for AI.
In the end I focused on critically reflecting on our concept's narrative and steered it towards a system that tries to break free from the black-box or all-knowing algorithms. By experiencing the development process of this concept, I see how important, but difficult it is to keep sight of the user and their goal. With the same system architecture for example, we could have also made a product that actively extracts user data to further wall-off their music explorations. Instead, we opted to open up the ability to provide contextual input and let the system do what it is good at, in a human centered way.

All in all I gained much experience which helps me see the possibilities and hazards in a design-with-AI process. But in future projects I will need to stay critical and curious to successfully improve concepts by using AI.

Febe Meijer:

My initial goal for this course was to develop a broader and deeper understanding of AI, specifically in the context of design. Given the likelihood of working in a multidisciplinary team in the future, where others may have more experience in AI, I wanted to gain foundational knowledge about its concepts and capabilities. This understanding would allow me to effectively communicate and bridge the gap between individuals with technical expertise in AI and those without.

A key takeaway from this course has been learning how AI can be integrated into design projects and, more importantly, identifying when it is appropriate or necessary to do so. The term 'AI' seems to be incorporated into every innovatively developed project, while it might not even be necessary. Therefore, I do acknowledge that it is important to first research what elements of a design might truly benefit from adding AI solutions. This conveyed the idea that AI should not be implemented for the sake of novelty but rather as a tool to address specific design challenges. Understanding this helped me critically assess the relevance of AI in various contexts and make informed decisions on its application.

What really contributes to the above is now having a broader understanding of the possibilities and capabilities of various machine learning models and the data that it is trained on. By learning not only how these models are currently applied across different functionalities and settings but also how I can implement or utilize them in my own work, I feel better equipped to make informed decisions Gaining more technical knowledge about the relationship between input data, output data, and the processes in between has strengthened my ability to select and apply the most suitable models. This deeper technical insight has also influenced my ability to evaluate the strengths and limitations of different models and approaches, allowing me to tailor solutions that best fit the needs of a given design project.

This course taught me the practical skill of implementing an already trained machine learning model in the specific context it is designed for. While this resulted in integrating only a minor aspect of the broader availability of models in the final assignment, it provided valuable hands-on experience. It also taught me how to use Hugging Face as an exploration platform for possibilities. Understanding the capabilities, limitations, and expected outcomes of a model serves as a crucial guide when searching for suitable models. Besides this, it is essential to consider the data on which a model has been trained and how this influences the goals that need to be achieved.

Additionally, individually I played a significant role in the preparation of the user test, which involved a critical focus on ethical considerations regarding data handling when working with machine learning models. Because adherence to the GDPR is key in design projects, I recognize the importance of transparency on what happens to the user's data (both collected in the user test and used within the model), how it is collected, stored and processed.

In past projects, I have often faced challenges during the system integration phase, where various individual elements are combined into a cohesive working model. Within this project, I served in a supportive role during that phase. I was able to successfully contribute to making the entire process, from speech-to-text to audio playback on Spotify, fully functional.

What I personally experienced as enjoyable in this course is combining the data-centered approach of the design with a more user-centered approach of development and testing. While other electives merely focus on one of these aspects, a more holistic approach is provided. This helps me better understand how to integrate both technical and user-oriented considerations, which better prepares for future scenarios and projects.

Xinyue Liu:

My knowledge of programming was very lacking, so I intentionally chose this course with the expectation that I would be able to advance in the field of Math, Data & Computing. Actually, this course did achieve my goal, and I gained a lot from it. I learnt about various AI types and their differences, I mastered the ability to use various models in Colab, and I learnt about using Gradio to create front-end interactive interfaces.

In this course, I was responsible for some parts. At the very beginning, we discussed what model to use for our idea (generating music from audio) and each of us tested different models to compare their advantages and disadvantages. I tested the MMAudio model. It was tested with a version that someone else had packaged online, but it was enough for testing. Unfortunately, all the models we could find had the same problem in that it would take a long time to generate music of sufficient quality, which did not satisfy our needs all over. However, after a discussion with our instructors, we changed the need for generating music to selecting music from a music library, which resulted in a very fast response time. Sometimes the direction of the problem that can't be solved, maybe a different way of thinking will lead to a more ingenious solution.

We then re-divided the work and divided the whole algorithm writing into four parts. I was responsible for the front-end interface. Firstly, it was written in Gradio, which was not too complicated and fun, and I learnt how to automatically transfer outputs from one model to inputs of another one. However, radio is limited in that it's essentially an interactive interface to the content of the code, rather than a prototyping tool for the app. Therefore, when I needed a more highly finished interface, i.e., one with different, non-coded modules, Gradio was either unable to meet the requirements or was too cumbersome to do so. In addition, Gradio 's interface has very basic and cumbersome color and module matching. When testing the Gradio application, I initially used the basic speech-to-text model whisper for the first half of the front-end interface because I had to wait for the rest of the team to finish their codes before I could put in the model part of Gradio, but the text-to-music selection in the second half could not be tested at that time. It didn't matter, because we changed the approach later.

Because of the various inconveniences of Gradio, as well as other group members' code writing needs more time, so in the front-end part, I changed to the model of simulating the interface + background output. In other words, I used Figma, a more widely used interactive interface design tool, to simulate the interactive interface, while the other devices did not do the front-end and directly called the code to achieve the purpose. This approach was very successful, and there were no major errors in the interactive interface part of the test session. In this Figma interface design, I learnt how to interact with components. In the past, I used to design different interfaces and then use the interactive component → response page approach. However, this time, I learnt to design by interacting with the component → responsive version of the component itself. In addition, I also learnt to simulate typing interaction design in Figma. This part helped me a lot because I want to be a human-vehicle interaction designer in the future, which covers the interface design part of HMI, also working with Figma. The interaction interface design skills in this course have given me more specialized design habits and knowledge in HMI.

I was also responsible for the analysis of the test results, and it was in this section that I became more proficient in data visualization. How to generate a boxplot directly from a large amount of data in excel was the main thing I learnt. This was helpful because I used to use Echarts for data visualization, so once I encountered a large amount of data I wasn't as proficient, whereas generating it directly using excel was quick and convenient.

Overall, the course was very helpful to me. In addition to the AI-related knowledge and skills I learnt as mentioned in the first paragraph, I learnt to think in a way that incorporates ML and AI into my design approach. In the future learning process, if I encounter an event that cannot be simply solved, I will now take the initiative to find out if ML and AI have a solution, and I have the ability to master this solution.

Appendix C: Code

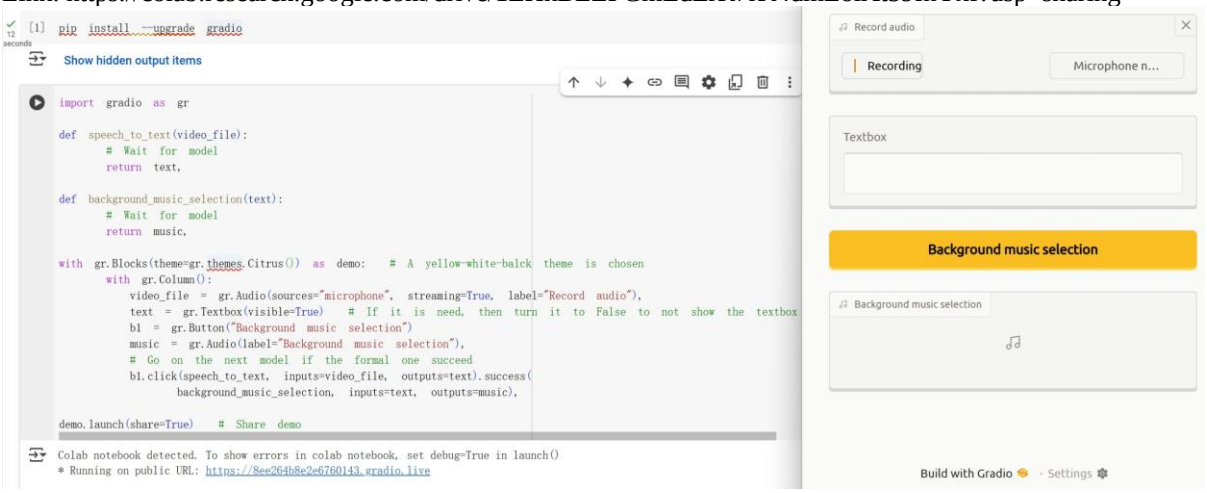
Github link to code: https://github.com/OlivierBlom/Designing-with-advanced-AI

Appendix D: Gemini system prompt

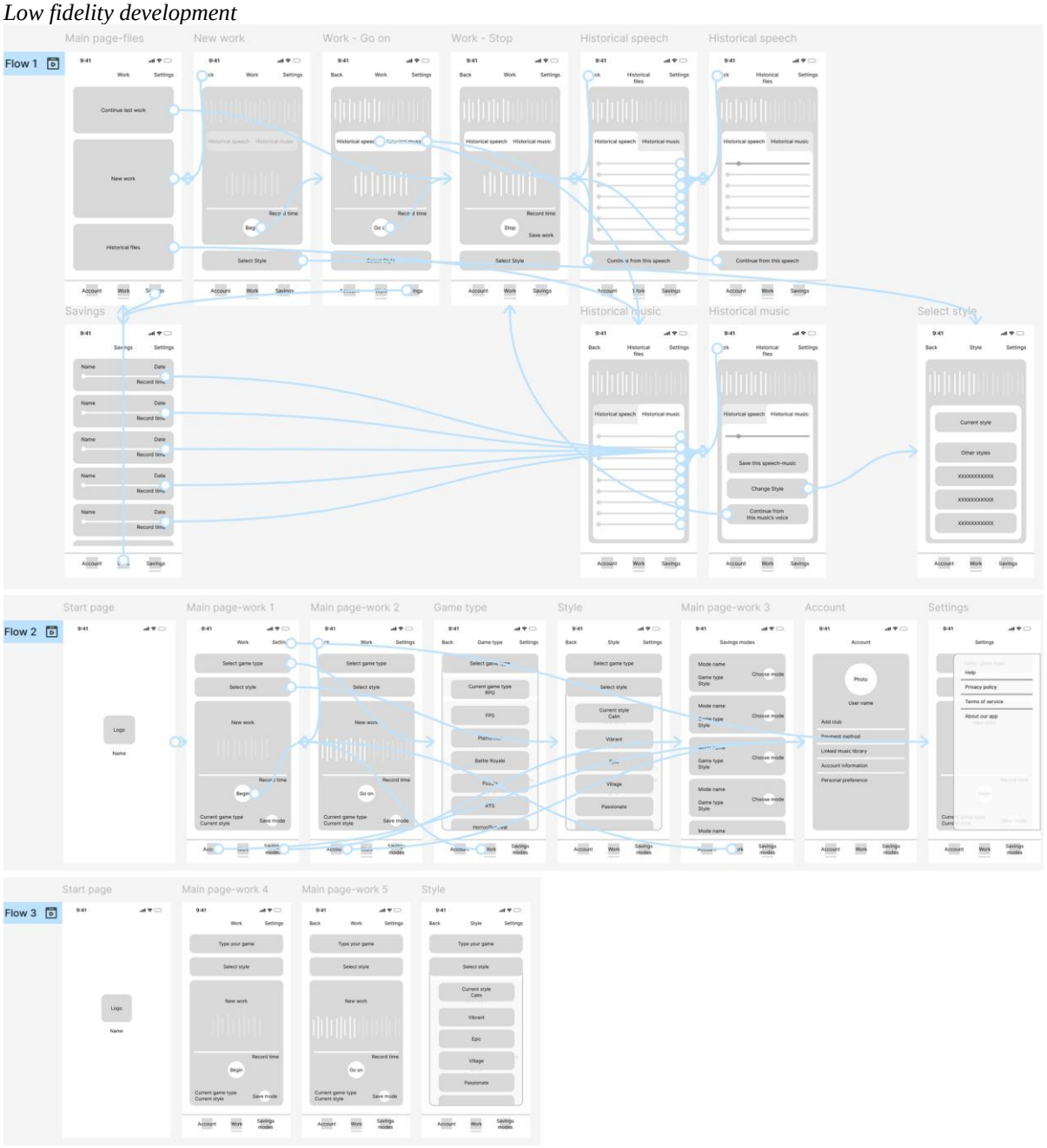
```
  system_message = f'''INSTRUCTIONS: Do not respond to the system message. After the system message respond with a max. 5 word playlist title of a vibe based on a conversation snippet SYSTEM MESSAGE: You are a spotify music recommendation agent. You will provide a vibe description based on one contextual word: {game-Name} and a conversation snippet. You will analyse the conversations sentiment and provide only a maximum of 5 words describing a playlist title or vibe in a manner that is logical for a playlist title. Don't use special characters in the output. If the criteria are not met, dont respond with anything '''
```

Appendix E: User interface

Front-end interactive interface by Gradio
Link: https://colab.research.google.com/drive/1EHkBLLFGmEdErR7rK4dmLoiAt59fnYxf?usp=sharing

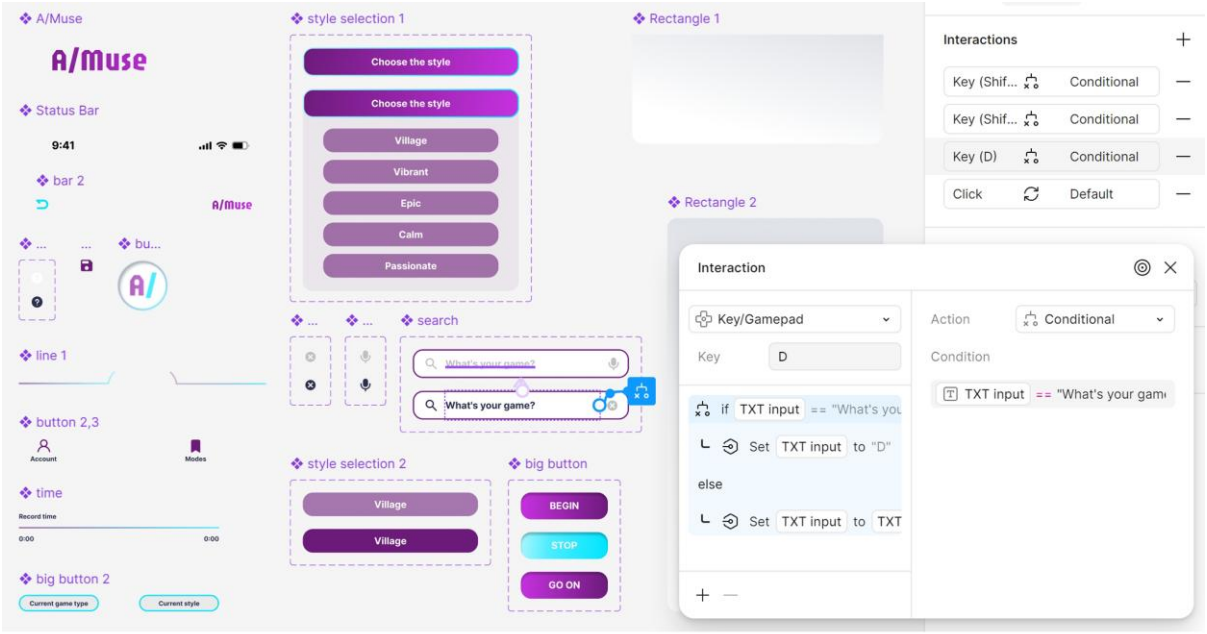


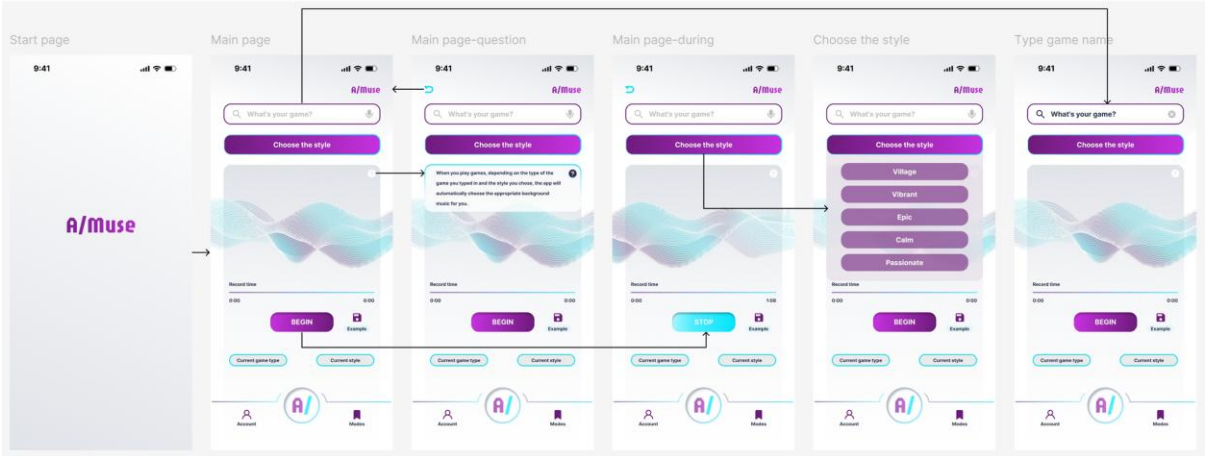
Low fidelity development



High fidelity
Figma link: https://www.figma.com/design/mFrlGLGQLj1TRtwsOm54F0/Advanced-AI-sharing-file?node-id=0-1&t=ZNoB7aDRMj0xvanQ-1 (with password: AI)

High fidelity components and interaction:





Appendix F: Study protocol

Part 1: User test

Participants are explained the game of Exploding Kittens. After this they are asked to play the game while interacting with the UI and model for about 15 minutes.

Part 2: User Experience Questionnaire

Please make your evaluation now.

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	○	⊗	○	○	○	○	○ unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. Please remember: there is no wrong or right answer!

Please assess the product now by ticking one circle per line.

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

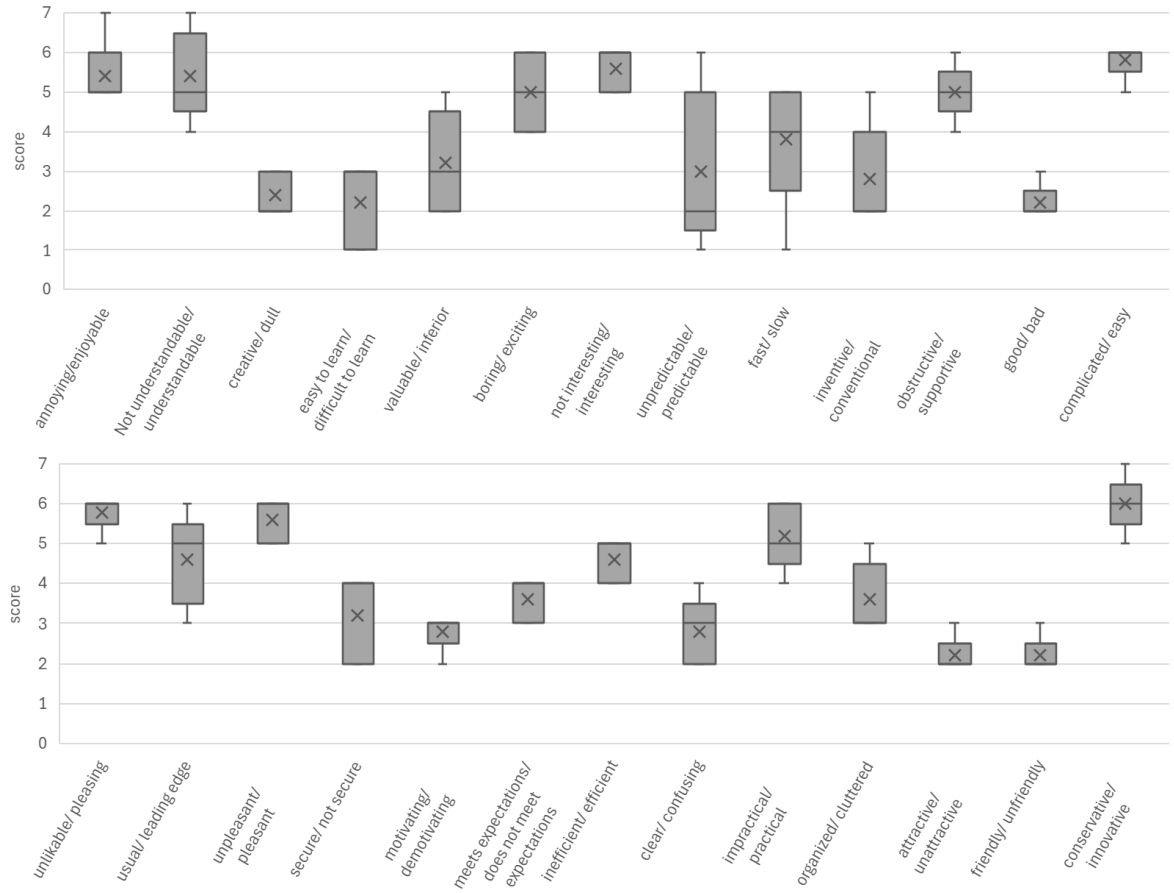
Part 3: Semi-structured interview:
Based on User Experience Questionnaire:

- Which elements did you rate very high/low? Why?
- Which of the elements in the UEQ stood out to you? Why?
- Which elements do you regard as important in the setting of this application? Why?

Based on the experience:

- What did you like/dislike with regards to the experience with the application?
- What are possible improvements?
- In which USE case do you believe this application might be interesting?

Appendix G: Boxplots of all UEQ dimensions



Appendix H: Consent form

Information sheet for research project "Speech to music in home and entertainment settings"

1. Introduction

You have been invited to take part in a research project Speech to music in home and entertainment settings, because you have been contacted by the researcher to participate.

Participation in this research project is voluntary: the decision to take part is up to you. Before you decide to participate, we would like to ask you to read the following information, so that you know what the research project is about, what we expect from you and how we deal with processing your personal data. Based on this information you can indicate via the consent declaration whether you consent to take part in this research project and the processing of your personal data.

You may of course always contact the researcher via f.a.meijer@student.tue.nl, if you have any questions, or you can discuss this information with people you know.

2. Purpose of the research

This research project will be managed by Febe Meijer.

The purpose of this research project is to research the user experience of a designed app with an integrated speech-to-music model.

3. What will taking part in the research project involve?

You will be taking part in a research project in which we will gather information by:

- Observation
- Presenting you a questionnaire about user experience which you can fill in in writing
- Interviewing you about the given answers in the questionnaire and the overall experience and write down your answers.

This study will be completely anonymous, and the data obtained from the study will not be traceable to you.

For your participation in this research project you will not be compensated.

4. Potential risks and inconveniences

Your participation in this research project does not involve any physical, legal or economic risks. You do not have to answer questions which you do not wish to answer. Your participation is voluntary. This means that you may end your participation at any moment you choose by letting the researcher know this. You do not have to explain why you decided to end your participation in the research project. Ending your participation will have no disadvantageous consequences for you.

If you decide to end your participation during the research, the data which you already provided up to the moment of withdrawal of your consent will be used in the research. Do you wish to end the research, or do you have any questions and/or complaints? Then please contact the researcher via f.a.meijer@student.tue.nl.

5. Confidentiality of data

The raw and processed research data will be retained for a period of 2 months. Ultimately after expiration of this time period the data will be either deleted or anonymized so that it can no longer be connected to an individual person. The research data will, if necessary (e.g. for a check on scientific integrity) and only in anonymous form be made available to persons outside the research group.

This research project was assessed and approved on 21-01-2025 by the ethical review committee of Eindhoven University of Technology.

Consent form for participation by an adult

By signing this consent form I acknowledge the following:

1. I am sufficiently informed about the research project through a separate information sheet. I have read the information sheet and have had the opportunity to ask questions. These questions have been answered satisfactorily.

2. I take part in this research project voluntarily. There is no explicit or implicit pressure for me to take part in this research project. It is clear to me that I can end participation in this research project at any moment, without giving any reason. I do not have to answer a question if I do not wish to do so.

Name of Participant:
Signature:
Date:

Name of researcher:
Signature:
Date: