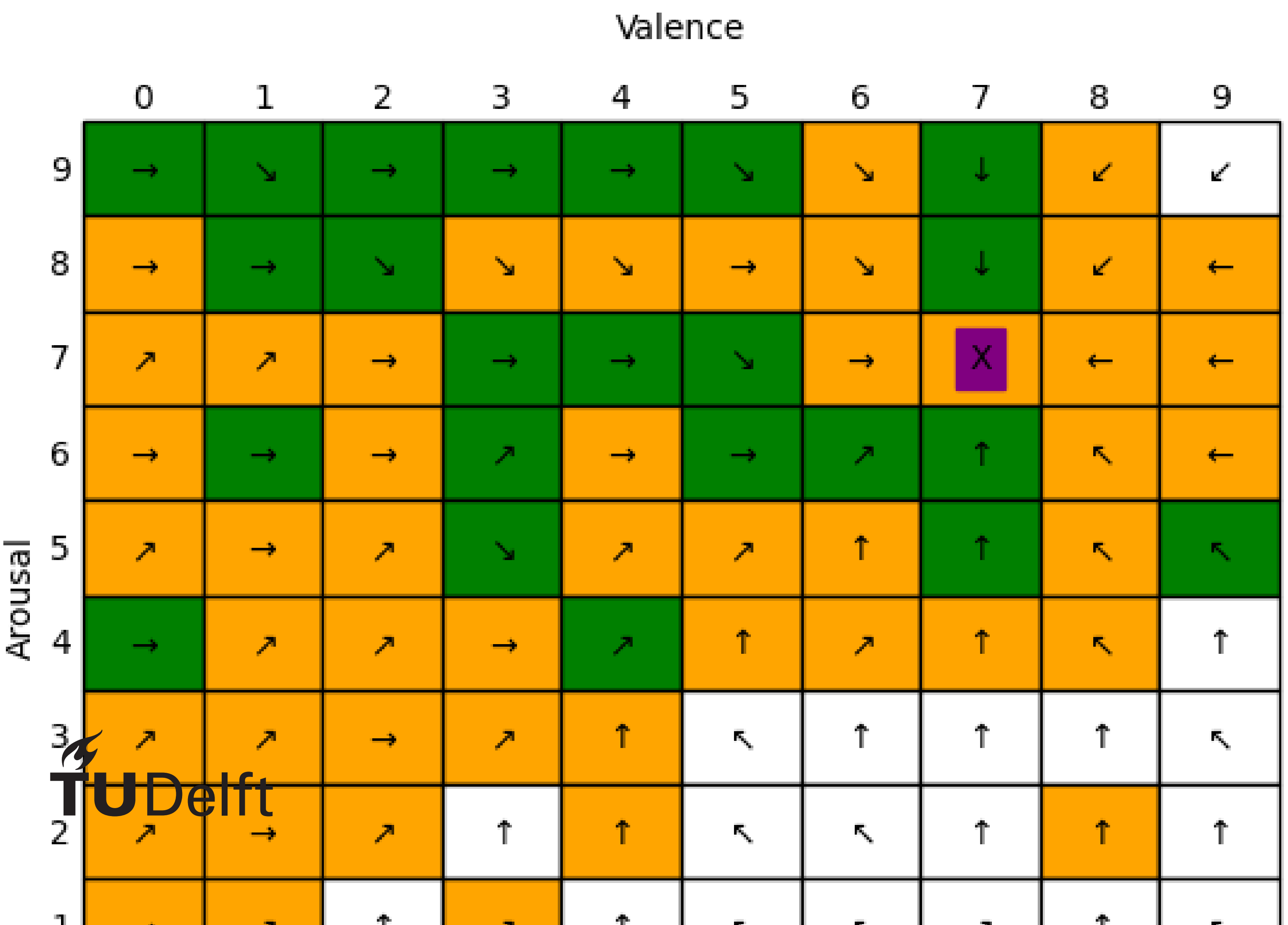
Music Playlist Generation for Emotion Regulation

A Functional Component in the Individualized Music Intervention for Persons with Dementia

Bernd Kreynen



Music Playlist Generation for Emotion Regulation

A Functional Component in the Individualized Music Intervention for Persons with Dementia

by

Bernd Kreynen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday July 5, 2021 at 14:00.

Student number:	4331842	
Project duration:	March, 2020 – July 5, 2021	
Thesis committee:	Prof. dr. Mark Neerincx,	TU Delft, Interactive Intelligence, supervisor
	Dr. Willem-Paul Brinkman,	TU Delft, Interactive Intelligence
	Dr. Matthijs Spaan,	TU Delft, Algorithmics

TU Delft

Preface

Dementia and the care of persons with dementia (PwDs) is an increasing problem due to a combination in the rise of dementia cases but also due to a shortage in the healthcare workforce. PwDs often deal with agitation. Non-pharmacological treatments for these symptoms have been studied in Psychology. One of these is the individualized music intervention as proposed by Gerdner [36]. In this thesis we defined two main research questions relating to this IMI and successfully answered these.

This report was made in the context of a master thesis for the master *Computer Science* at *Delft University of Technology*. It was supervised by Prof. Dr. Mark Neerincx from the *Interactive Intelligence Group* and the thesis committee members were Dr. Willem-Paul Brinkman from the *Interactive Intelligence Group* and Dr. Matthijs Spaan from the *Algorithmics Group*.

I would like to thank my supervisor, who supported me throughout the whole process of this thesis with regular meetings and feedback on my work. I am also grateful for the participants who completed the experiment and everyone who helped share the link to the online experiment. I was especially surprised by the warm welcome the Dutch music therapy community provided to research from a different discipline. In particular I would like to thank the different educational institutions who helped share my experiment amongst their students. These were different Dutch Universities of Applied Sciences (HAN University, Leiden, NHL Stenden, Zuyd, Utrecht) and two Universities of Arts (ArtEZ, Codart Rotterdam). The *Nederlandse Vereniging voor Muziektherapie* (Dutch Association for Music Therapy) was also very helpful, allowing me to give a presentation for their members to gain some feedback and bringing me into contact with several music therapists who shared their knowledge and experience with me.

Bernd Kreynen
Delft, June 2021

Abstract

Dementia care is a growing problem, both due to a rising number of cases and due to a shortage in healthcare workers. Aside from cognitive symptoms persons with dementia (PwDs) often deal with psychological symptoms such as agitation. The individualized music intervention (IMI) by Linda Gerdner has been proposed to reduce these. This is the basis for our first research question "How can we incorporate the IMI into technology to care for PwDs?". We answer this by dividing the guidelines for IMI into functional components with their own responsibilities. This allows for more independent development of these components. We also research one of the components, a playlist generator. This leads us to our second research question "Assuming knowledge of a list of preferred music for a user, can we generate playlists which better regulate emotions than a random shuffle?". Our generator is based on a Markov Decision Process (MDP) and allows for a target to be set in the 2D valence-arousal space. This was tested in an online experiment with 22 participants. The control group listened to a random shuffle of their preferred music. The experimental group listened to a playlist generated by the MDP. The target was set to a valence > 0.75 (valence ranging from 0 to 1). A significantly higher final valence for the experimental group was observed, showing a clear improvement over the control group when the goal is to positively influence the valence. The dataset obtained through this experiment is available for future research..

Contents

1. Introduction

By 2030 the WHO predicts a global healthcare workforce shortage of 14.5 million [103]. Part of the problem is an ageing population with special needs [102], often in combination with an ageing workforce [103]. In particular there is an expectation of having more persons with dementia (PwDs). Both in Europe and worldwide the number of cases could more than double by 2050 [61, 68].

PwDs often deal with agitation and depression which cause feelings of helplessness and distress in families and caregivers of the PwDs [3]. Due to the adverse side effects of pharmacological treatments, non-pharmacological treatments for these symptoms can be preferable [3]. One of these treatments that has received attention, and has been found to be effective in reducing agitation, is music [3, 10]. It has also been found that personalization of this music for the PwD plays an important role in the effectiveness of the treatment [10]. Guidelines for implementing an individualized music intervention (IMI) for persons with dementia have been described and implemented by researchers and caregivers with success [36]. Clearly an important part of this treatment is creating individual playlists for PwDs. However there is little research which looks into how we can combine this treatment with other technological developments such as music recommender systems and affective computing.

Researching the automation of (parts of) this intervention could allow it to be incorporated into different technologies to help support PwDs and their caregivers. Examples could be socially assistive robots, support tools for music therapist or other healthcare professionals, personalized music players for PwDs, ... This makes it an interesting topic to look into further.

From a research perspective it is also interesting since, as we will show, implementing the IMI requires implementing different components which show a lot of overlap with current emerging fields such as affective computing and contextual music recommendations. To implement automated IMI we have to answer similar questions to the ones asked in those fields such as: "How do we sense the emotional state of a user?" "How do we model the emotional state and influences on it?". It also has overlap with already more established fields such as machine learning, which could offer solutions to the automation of some of the components, and traditional music recommender systems, which could be used to automate the process of retrieving music a PwD might prefer.

In this thesis we define two main research questions. The first research question is "How can we incorporate the IMI into technology to care for PwDs?". We will attempt to do this by dividing the IMI into functional components with their own responsibilities. This forms a framework of components which can then be fulfilled in various ways depending on the IMI implementation. This could be a tool to support music therapists with IMI or in the future perhaps an assistive robots that applies the IMI. We will also show that many of these functional components have overlap with currently emerging research fields such as affective computing. By dividing the intervention in these components research from these fields can later be used to implement them.

As mentioned earlier generating individual playlists is an important part of IMI therefore one of the functional components is responsible for generating playlists. In the current IMI this is often determined by the device used to play the music. If an iPod Shuffle is used to store the preferred songs of a PwD for example then the playlist is simply a random shuffle of the preferred list. If the aim is to reduce agitation then perhaps it is possible to improve upon this random shuffle when this affective aspect is kept in mind. Therefore our second research question focuses on playlist generation for emotion regulation. The question is "Assuming knowledge of a list of preferred music for a user, can we generate playlists which better regulate emotions than a random shuffle?".

To answer this question we will develop an algorithm based on Markov Decision Processes (MDP). Modelling it as such allows us to use research from the field of Reinforcement Learning to help us solve this problem. To test this algorithm we will perform an online experiment with real participants (without dementia).The main goal of our experiment will be to answer our second research question. Answering this question allows us to implement this improved playlist generation algorithm in current tools to for example support music therapists during IMI, or other interventions where emotion regulation is desired.

Next to answering the second research question the experiment has two subgoals. The first subgoal of the experiment is to generate more data which can be used to study music related induced emotions. Much of the current research on emotion-based music recommendations focus on using emotion as a contextual factor and generating music which is more in accordance to the users preference with this information. Lacking the focus on induced emotions. The datasets relating to emotion and music often also focus on perceived emotions. Many of the datasets that do focus on perceived emotions use short fragments of songs [19, 50, 99]. The only one we found using full songs uses an emotion scale specific to music [5], making it more difficult to integrate with more general research from the field of affective computing. This lack of data limited us in the design of our playlist generation algorithm. In our experiment we used a fairly simple heuristic MDP model which did not require this data. More available data on this topic would be a great benefit to future research into playlist generation for emotion regulation.

The second subgoal is to check if we can find a noticeable difference in emotional reactions to songs that have memories attached to them. This second subgoal occurred more naturally due to the need to ask some questions during the experiment. Making one of these about memory was motivated by previous research showing that memories intensify induced emotions when watching music videos [29]. If we observe a difference it could show potential for incorporation into future algorithm. It also adds some additional personal information about the songs to the dataset. This aspect will be further explained in Chapter 5.

In this thesis we used Socio-Cognitive Engineering (SCE) [1, 63] to guide our design. By using this method we arrived at our current research goals. In the appendices some more artefacts produced by using SCE can be found for interested readers. Appendix A contains the ontology. Appendices B to D contains various artefacts produced due to using SCE from earlier iterations of the thesis. One of these artefacts is the foundation, it contains the relevant domain, human factors and technological knowledge [1]. Chapter 2, the background, incorporates the parts of the foundation that are most relevant to our two main research questions.

The structure of the rest of the report will be as follows; Chapter 3 attempts to answer the first main research question by giving an overview of the functional components of IMI. After this Chapter 3 we move on to the second main research question; Chapter 4 serves as an introduction to our experiment and explains how MDPs could be applied to the problem of playlist generation in IMI. Chapter 5 explains the goals and hypothesis of the experiment. It also explains the design and implementation of the experiment. Chapter 6 gives an overview of the results of our experiment and Chapter 7 has a discussion related to the experiment. Finally we wrap up the report with a general conclusion relating to both research questions in Chapter 8.

2. Background

In this background section we start by defining the terms emotion, mood and affect. We then move on to giving some background on dementia and musical interventions applied in the care of PwDs. We cover relevant music recommendations research, in particular emotion-based music recommendations. We give an overview of the available emotion based music datasets and give a very brief introduction on reinforcement learning and its application in music recommendations.

2.1. Emotion, Mood and Affect

As mentioned in our introduction part of our thesis focuses on a playlist generation algorithm which regulates emotions. Therefore it is important to start by defining what we mean by different terms relating to emotions. Affective computing, an emerging research field, also uses the terms emotion, mood and affect. Sometimes clearly defined but sometimes also vaguely defined or even used interchangeably. This is in part due to them not having clearly unique definitions [83]. This can be noticed in for example reviews that use both the terms mood and emotion without any further clarification as to their meaning [56, 84, 92, 106]. For clear communication it is important that we do define these. To do so we look at a thesis by Sternheim [90]. In this thesis an ontology for affective reasoning was developed for the PAL (Personal Assistant for a healthy Lifelstyle) project. A project which tries to improve self-management of diabetes in children through a social robot. The ontology was developed to be able to add affective abilities to the robot [90]. The following definitions are based on this thesis [90].

Emotion Sternheim looks at the definition given by Shiota and Kalat [82]: Emotions are a universal, functional reaction to an external stimulus event, temporally integrating physiological, cognitive, phenomenological and behavioural channels to facilitate a fitness-enhancing environment-shaping response to the current situation'. This tells us that emotions are temporary (more short lived), they are a reaction to a certain stimulus and cause changes in the person, physiologically or otherwise [90].

Mood The most important difference between mood and emotions is the longevity. Moods are more long-term and are not a direct result of a reaction to an external stimulus [90]. Due to this there is also less research available on how to for example represent or induce moods since this is more difficult in an experimental setting [90].

Affect Affect is a term used in research that can refer to either mood, emotion or both [90].

Affective (emotional) State Sternheim further defines the emotional (affective) state in their ontology. It encompasses both mood and emotions. A user has an emotional state and in turn an emotional state has both a mood and an emotion [90]. To avoid confusion later on we will refer to this as the affective state. The term emotional state will be used differently in this thesis later on.

2.1.1. Emotion and Mood Interaction

In the thesis Sternheim also proposes a model for the affective state, the *Emotional State Model for Affective Semantics* (ESMAS) [90]. It is shown in Figure 2.1. The details of this model will not be explained in this thesis. However important to note for our research is that emotions and mood have an influence on each other. Sternheim notes for example that techniques such as forcing a smile have been shown to work to improve the mood of someone [90]. In a similar fashion we expect that by inducing emotions in our users we will also influence their mood.

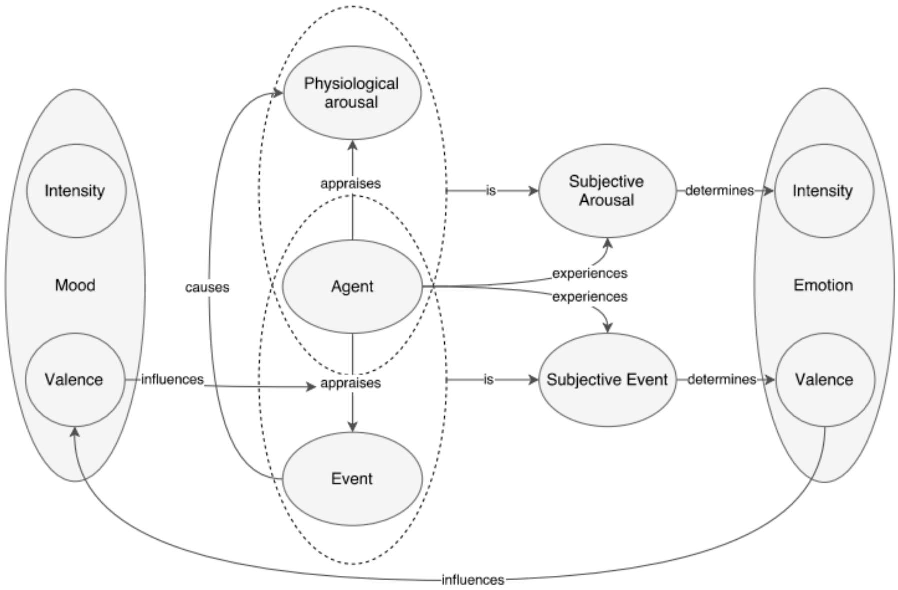


Figure 2.1: The visual representation of Sternheim's ESMAS. Image from [90]

2.1.2. Emotion representation

In this thesis to represent emotions we will mostly use the VA-scale. A two dimensional scale. In this scale valence is the first dimension and defines the pleasure/displeasure of an emotion or how positive/negative the emotion is [70]. The second dimension, arousal, is a sort of intensity and describes activation/deactivation [31]. Calm and relaxed are for example positive valence and negative arousal, while excited and happy are positive valence and positive arousal [70]. Occasionally we will also discuss a third dimension, dominance, from the PAD-scale, discussed later in this section. For completeness we give an overview of some emotional representations commonly used in emotion recognition research.

Dimensional

Two Dimensional The VA-scale is based on the valence arousal model of Russel [70] described in the previous paragraph. It has been widely used in research. It it is for example used in datasets related to music and emotion [6, 85, 86, 111], music emotion recognition [39, 43, 66, 100, 107] and emotion recognition from different physiological signals [12, 44, 88, 95].

Three Dimensional Sometimes a three dimensional scale is used. These use the valence and arousal scale and add an additional third dimension. One example of this is the PAD scale. This scale has the dimensions of pleasure (valence), arousal and dominance [38]. The scale has been used in for example automated emotion recognition [25].

Other third dimensions have also been proposed. Deng et al. for example use a third dimension of Resonance in their emotion-based recommendations [26]. In music related research the scales have also been called valence, energy and tension [99].

GEMS The Geneva Emotional Music Scale (GEMS) has been made specific to music and contains nine scales [99, 110]. The scales are: wonder, transcendence, tenderness, nostalgia, peacefulness, energy, joyful activation, tension and sadness [110].

Categorical

Categorical approaches have also been used widely, using many different categories. The categorical approach can be seen in for example datasets related to music and emotion [5, 15, 65, 99], emotion-aware music recommendation systems [2, 37, 45] and automated emotion recognition [4, 75].

Discussion on approaches
Vuoskoski et al. researched different representations (categorical, 3-dimensional and GEMS) in the context of induced emotions from music [99]. They found that the overall consistency of the dimensional model was the highest. Additionally they also found that the three dimensional model had some redundancy in the third scale and suggest it may be removed [99]. These findings, combined with its widespread use in other research makes the VA-scale seem like a good choice to represent emotions in this thesis.

2.2. Dementia

In the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) the previous diagnosis of dementia is now placed under major or mild cognitive disorders. However for continuity it retains the use of the term dementia which can for example be used in settings where patients and physicians are used to this term. In research we can also find both terms being used [7]. There are many different forms of major cognitive disorders with the most common and well studied one being Alzheimer's Disease [47, 105]. It is estimated that 60 to 90% of dementia cases are attributable to Alzheimer's disease [7]. Other forms include vascular dementia, Lewy bodies, frontotemporal dementia and cognitive disorders due to Parkinson's, HIV, traumatic brain injury or other underlying causes [7]. However distinguishing these forms in a patient can be difficult [47].

2.2.1. Demographic

According to the World Health Organization in 2018 approximately 50 million people worldwide were affected. Age is the strongest known risk factor and therefore the overall cases rise quickly with age, but dementia does not exclusively affect older people [7, 104]. Dementia before the age of 65 is called early onset dementia (EOD), but the prevalence of this has been poorly studied [97]. In this section we will give some background on dementia, first we will cover both the cognitive symptoms and the behavioural and psychological symptoms afterwards we will look into the effect of dementia on caregivers.

2.2.2. Symptoms

Cognitive
The primary deficit in dementia, as the name major cognitive disorder suggests, is in cognitive function [7]. The DSM-5 describes 6 cognitive domains in which decline can occur [7]:

- Complex attention (sustained attention, divided attention, selective attention, processing speed)
  - Example symptoms include increased difficulty with multiple stimuli, difficulty with holding new information such as reporting what was just said and the inability to perform mental calculations.
- Executive function (planning, decision making, working memory, responding to feedback, inhibition, mental flexibility)
  - Example symptoms include abandoning complex projects, difficulty with multitasking and difficulty with resuming tasks after an interruption.
- Learning and memory (immediate memory, recent memory, very-long-term memory)
  - Example symptoms include repeating themselves in conversation, not being able to keep track of short lists such as shopping lists, requiring frequent reminders to perform tasks and difficulty recalling events.
  - The DSM-5 notes that except in severe forms of major neurocognitive disorders semantic, autobiographical and implicit memory are relatively preserved in comparison with recent memory.
- Language (expressive language and receptive language)
  - Example symptoms include preferring general pronouns over names, difficulty finding words, peculiar word usage and grammatical errors.
- Perceptual-motor (integrating perception with purposeful movement)
  - Example symptoms include difficulties with previously familiar activities such as using tools and difficulty navigating environments.

- Social cognition (recognition of emotions, theory of mind)
  - Example symptoms include behaviour and attitude changes, decreased empathy, decreased inhibitions, increased extraversion or introversion and insensitivity to social standards.

The diagnosis of major neurocognitive disorder requires significant cognitive decline in one of the previous domains from a previous level of performance, interference with independence in everyday activities from this cognitive decline and that the deficits are not better explained by another mental disorder such as major depressive disorder or schizophrenia [7].

Behavioural and psychological

Behavioural and psychological symptoms (also called neuropsychiatric symptoms [79]) are also common in people with neurocognitive disorders. Symptoms include psychotic features, hallucinations, paranoia and other delusions, sleep disturbances, wandering, dis-inhibition, apathy, depression, anxiety, aggression and elation [7, 46]. The most prevalent symptoms are apathy, depression, irritability, agitation and anxiety [17]. These symptoms do not only present in later stages but also in the early stages of cognitive decline [17, 57]. They occur more frequently in hospitals or long-term care facilities [17].

Behavioural and psychological symptoms are seen almost universally in all types of dementia but certain behaviours can be more common in certain types or in certain stages of the disease [17, 46]. Progressively worsening depression and anxiety are for example common in Alzheimer's disease [46]. Elation on the other hand occurs more commonly in frontotemporal dementia [7]. These behavioural and psychological symptoms are episodic and have a tendency to fluctuate, contributing to the difficulty of managing these symptoms [46]. Dementia can causes these symptoms directly by disrupting brain circuitry involved in behaviour and emotion [46]. Dementia can also can also cause these symptoms indirectly by increasing the vulnerability to stressors, on which other factors, such as environmental triggers, can also have an effect [46].

2.2.3. Effect on caregivers

A review by Cooper et al. studied the prevalence of anxiety in caregivers for PwD and found that a quarter of them suffer from anxiety and that the rates of anxiety are significantly higher than in control groups [24].

In a review Cheng found that the symptoms described in Section 2.2.2 and in particular disruptive behaviours, such as agitation, aggression and disinhibition, are most predictive for caregiver burden and depression [20]. This is both due to the impact on the emotional connection with the PwD and due to them causing difficulties in care [20].

High costs are also associated with managing these behavioural and psychological symptoms. In a study from 2002 for example Beeri et al. found that approximately 30% of the total costs of caring for persons with Alzheimer's disease is invested in the direct management of these symptoms [77]. In a 2006 study Herrmann et al. also found that these symptoms contribute significantly to the overall cost of dementia care [41].

2.3. Musical Interventions

Studies show that PwDs respond positively to music and that this is preserved into late stages of dementia [10, 73]. In the following subsections we will describe a few different musical interventions for PwD that have been researched.

2.3.1. Musical Therapy

Musical therapy is performed in many different ways, both in individual and group setting. Sessions are lead by a therapist [60]. It has been found to reduce agitation [10, 55, 60] and improve mood, alleviating symptoms such as depression and anxiety [10, 60]. However when comparing the group therapy to other group therapies not involving music, no significant difference was observed between the two, suggesting that in group activities perhaps the social component plays a bigger role [10, 62, 98].

2.3.2. Interactive Interventions

Some research is reporting better results on improving the mood and agitation of PwD when interactive interventions are used in combination with music [10]. These can include things such as singing, clapping or dancing [71]. Sánchez et al. also compared multi-sensory stimulation and found that it might have better effects on anxiety and cognitive impairment than individualized (receptive) music on persons with severe dementia [72]. However a recent meta-analysis concluded that receptive music therapy could reduce agitation, behavioral problems and anxiety compared to usual care, while no significant difference was found between interactive music therapy and usual care [94].

2.3.3. Individualized Music Intervention

Individualized music interventions have also been studied. In these interventions a PwD listens to a personalized playlist. Gerdner et al. first studied the effects of individualized music for PwDs in 2000 and found a significant difference in the reduction of agitation between individualized and classical music [33, 35]. Following research has also found the use of individualized music to be able to reduce agitation and the emotional well-being in people with dementia [23, 32, 35, 67, 72, 93]. Since then Gerdner has made evidence based guidelines for individualized musical interventions [36]. These guidelines have been shown to be effective both when implemented by research staff and by trained staff and family members [36]. We will give an overview of these guidelines in Section 3.1. The functional components in Chapter 3 are also based on these guidelines. Currently in individualized musical interventions playlists are manually created specifically for the PwD, in co-operation with for example family to find music that was relevant to the person's life and in line with their personal preferences before the on-set of dementia [71, 72, 93]. From the guidelines by Gerdner we also have a simple method of for assessing preference though:

2.3.4. Assessment of Personal Music Preference Questionnaire

In the evidence based guidelines by Gerdner to help with the creation of a playlist she suggests using the Assessment of Personal Music Preference Questionnaire (APMPQ) [36]. This APMPQ has been used in research and positive effects on agitation were found when utilizing it [34, 67]. Sung et al. used a modified version of the questionnaire in Taiwan and found significant differences in anxiety when implementing the intervention [91].

The questionnaire includes questions about their history with music (did music play an important role before their illness?, Do/did you play an instrument?, ...), their preferences (selecting favourite music types) and finally asks for specific songs/albums/artists/... [36]. If the individual cannot fill this in themselves then a different version is also available for family or caretakers to fill [36].

2.4. Music recommendations

In this section we briefly cover some approaches that have been used in general music recommendations. With general we mean music recommendation not specifically for PwDs or specifically relating to emotions.

2.4.1. Collaborative Music Recommendation

In collaborative music recommendation systems songs are recommended by looking at the choices of other, similar users [11, 87]. It is one of the most successful approaches in recommendation systems [87]. Predictions for scores on unseen songs are based on a combination of songs of users whose past ratings have strong correlation [11, 87]. The assumption is that users that behave similarly have similar tastes, but this assumption has not been widely studied [87].

2.4.2. Content-based Music Recommendation

In content-based music recommendation predictions are made based on data from the songs themselves [11, 74, 87]. The data used can be metadata (tags from social networks, data from web documents, manual annotations, ...), Audio content (acoustic and musical features) [74]. Songs that are similar to songs the user has listened to or rated positively are recommended. Some limitations are that generating features to base the predictions on can be difficult and that it has a tendency to overspecialize due to recommending similar types of items [11].

2.4.3. Context-aware Music Recommendation

These methods incorporate contextual information to make their recommendations. This can be for example the time of day, location or information from wireless network sensors at the time of recommendation [11, 16]. This makes them aware of certain contexts in which the recommendation has to be made [11]. The context can be environmental-related (location, weather, ...) or user-related (user's activity, emotional state, ...) [16]. A specific subset of Context-aware systems are emotion-aware or emotion-based systems. We will cover emotion-based music recommendation in Section 2.5.

2.4.4. Hybrid Models

It is also possible to create hybrid models by combining models in various ways. Proper hybrid models can outperform single approaches due to incorporating advantages of different methods, while having less of the disadvantages [74, 87]. Some studies have explored the combination of collaborative and content-based methods, and context-aware and collaborative methods, however relatively few have combine context-aware with content-based [74].

2.5. Emotion and Music Recommendation

Affective computing is becoming a more popular research field. Related to context-aware music recommender systems, some music recommender research has also focused on the aspect of emotions. This affective factor can come into play in different ways. Many use emotions as an extra context to improve recommendations, but not necessarily to induce, regulate and/or improve emotions [9, 27, 81]. Therefore this research is less interesting to our goal of regulating or inducing emotions. Deng et al. try to get information about the emotional state of a listener through social media, they mention the difficulty with the acquisition of emotions as a reason for researching this option [27].

Han et al. [40] use a SVM to learn emotion state transitions, but they use a categorical approach to representing emotions. They also propose a context-based music recommendation system based on the data on emotion transitions from this SVM. While certainly interesting this data does not seem to be available for further research. They also tested the system on 30 participants; they only report that the participants found the results satisfactory and do not report results on the influence on the emotional state of the participant. So while they do estimate emotion transitions with their SVM and use this in their recommender system, in their evaluation they also focus on recommending music that is satisfactory to the listener in a given context instead of focusing on regulating emotions.

2.5.1. Physiological signals

Some work has also been done in incorporating physiological signals into music recommendations. Ayata et al. for example use physiological sensors (plethysmography and galvanic skin response) in wearables to estimate the listeners emotions [9]. Shin et al., Guo et al. and Morita et al. all propose systems incorporating EEG signals for mood-based music recommendation systems.

Some research in this area also attempts to influence certain factors related to affect. Liu et al. designed and tested a music recommendation system that attempts to make music recommendations to reduce stress on long haul flights based on heart rate monitoring. They found that incorporating the heart rate monitoring in the music recommendation can reduce stress. [52, 53]. Shin et al. also propose a similar music recommendation system [80]. Nirjon et al. developed a system using earphones that monitor the heart-rate and combine it with a music recommendation system that attempts to keep the listeners heart rate at a target level [64]. Liu et al. describe a system for improving driver alertness through music recommendation that incorporates EEG signals [54].

2.6. Datasets on music and emotion

This section gives an overview of the available datasets on music and emotions. Table 2.1 gives an overview of all the datasets. Important distinctions between the datasets are in the sensors and feedback they use. The items of the dataset (e.g. a whole song, excerpts from songs, music videos, ...). Whether they are static or continuous, meaning whether the dataset has one label for an entire item in the dataset or whether a song has labels throughout the duration of the item. Whether they try to annotate the emotions induced in the participant or the emotions expressed in the item. The type of labels that are given (e.g. valence-arousal scale, categorical, ...). And finally they vary a lot in the number of participants and number of items evaluated. For our research datasets about perceived emotions are not that interesting. Therefore in this section we will only give a short overview of the datasets containing induced emotions:

2.6.1. AMG1608

This dataset contains induced emotions and was created with the aim of enabling the testing of personalized music emotion recognition (MER) systems [19]. They used the valence-arousal (VA) scale. With 665 subjects it is quite a large dataset and 46 participants rated more than 150 songs. Participants provided their emotional state explicitly. They also used this dataset to train both a general and personalized MER and note that the personalized MER performs better [19]. It only covers short excerpts of 30 seconds however. Furthermore it also does not contain information about the sequence in which which participants listened to the excerpts.

2.6.2. DEAP

A Database for Emotion Analysis using Physiological Signals (DEAP) is a dataset containing multimodal data on induced emotions [50]. DEAP contains data from 32 participants who all rated 40 one minute excerpts from music videos. Participants were asked to self-asses their arousal, valence, liking for the item and dominance once per item. Additionally the following were collected during the viewing of the item; electroencephalography (EEG), electromyography (EMG), electrooculography (EOG), heart rate, respiration, respiration pattern, skin temperature, blood volume pressure, galvanic skin response (GSR). 22 participants also had frontal face videos recorded [50]. It is also limited by the shorter time of excerpts (one minute).

2.6.3. SoundTracks - Induced

There are two variants of the SoundTracks dataset. One was created with the aim of comparing categorical, discrete and dimensional models for self-rating perceived emotions, the other was made to compare the same for self-ratings of induced emotions [31, 99]. The items were soundtracks from films, which varied in length but were on average about 57 seconds long. Different groups of participants were asked to use different scales to rate their emotions. In total 102 participants listened to 16 different excerpts [31], the participants were divided in groups according to the scales they used. The dataset seems less interesting due to its shorter excerpts and focus on single excerpts and not sequences.

2.6.4. Emotify

This dataset was created to study induced emotions using the Geneva Emotional Music Scales (GEMS) [5]. Annotations were collected in the form of a game where participants could select a genre to play music from, like a song and self-rate a song according to the GEMS scale. [5]. They intentionally take lesser known music to reduce the influence of familiarity and preferences [5]. The choice to use GEMS however is in a way limiting since it was specifically made with music in mind and it is not used in the research of affective systems outside of this niche.

2.6.5. PMEmo

Created for MER this dataset has participants evaluate the chorus part of a song. It contains self-ratings about perceived emotions and data on induced emotions in the form of electrodemal activity (EDA) [111]. In total there are 457 participants who rated 794 different items. Each item is rated by at least 10 participants. In between listening to different songs participants listened to a part of *In My Life* by *Kevin Kern* [111]. It seems less interesting for our research due to its big focus on perceived emotions and due to listening to *In My Life* in between making evaluating emotion transitions more difficult.

In general we can conclude that the datasets either focus on perceived emotions, have short excerpts or use the newer and less used GEMS scale.

Name	Feedback	Label types (self-ratings)	Items	Participants	Perceived or Induced	Static or Continuous
AMG 1608 [19]	Explicit self-ratings	VA-scale	1608 excerpts of 30 seconds	665 total	Induced	Static
DEAP [50]	EEG, EMG, EOG, heart rate, respiration, respiration pattern, blood volume pressure, GSR and explicit self-ratings	VA, dominance, liking, familiarity	40 1-minute excerpts from music videos	32 total	Induced	Sensors are continuous, self-ratings static
SoundTracks, induced [99]	Explicit self-ratings	categorical, discrete and dimensional	16 excerpts from film soundtracks (+- 57 seconds)	102 total	Induced	Static
Emotify [5]	Explicit self-ratings	GEMS	400 songs	1595 total, not all rated all songs	Induced	Static
PMEmo [111]	EDA and explicit self-ratings	VA-scale	794 chorus parts of a song	457 total, 10+ per item	Perceived self-rating, induced EDA	Continuous and static
DEAM [6, 86]	Explicit self-ratings	VA-scale	1802 total, some are 45 sec excerpts some are full songs	Total number unclear	Perceived or expressed (contradiction in [6] and [86]	Continuous
1000 songs [85]	Explicit self-ratings	VA-scale	1000 full songs	100 total, 10+ per item	Perceived	Static and continuous
SoundTracks, perceived [31]	Explicit self-ratings	Categorical, discrete and dimensional	110 excerpts from film soundtracks (+- 15 seconds)	116 total	Perceived	Static

Table 2.1: Table containing basic information on musical emotion based datasets

2.7. Affect Recognition

The field of affect recognition is getting more attention [30, 76, 109]. Many different methods to automatically get information about the affective state of a user have been proposed [109]. As mentioned in Section 2.5.1 some have tried use physiological signals of various forms in the context of music recommendation. This approach has become more popular with the advent of wearables [76], however many challenges still remain with these approaches. Among them are the lack of available datasets. Current approaches also reach high accuracy in terms of arousal but still get lower accuracy on valence recognition [76].

In the field of computer vision facial emotion recognition (FER) has also been studied [48], this has also been applied in the context of music and multimedia recommendations [37, 58]. Many different machine learning approaches to recognize emotions from facial expressions through a camera have been proposed [48]. In the wild performance is still a challenge for these algorithms currently [30, 59]. Dupre et al. only report an accuracy ranging from 48% to 62% on currently available systems on spontaneous expressions [30]. Especially on PwD this could currently cause problems. They might not be aware of having to look straight at the camera. Additionally when discussing this option with music therapists they mentioned that something to keep in mind is that as dementia progresses the expressions of the PwD fade away and become more subtle. Therefore even if these improve on the general public it might not be a good option for PwDs.

In the future physiological sensors seem the most promising to provide an automatic method for measuring affective states in PwDs but for some PwDs these wearables could also be a source of agitation. Furthermore research on how to detect emotional states from these signals would have to be extended to PwDs after being improved on people without dementia first.

2.8. Reinforcement Learning

Reinforcement learning has been applied in research to a wide variety of problems [8, 14, 49], including music recommendation [18, 21, 42, 51]. These approaches generally look at a sequence of songs, and have some sort of memory [42]. Hu et al.l models the state as a windowed history of played songs and actions as picking a song. To deal with the curse of dimensionality they use a clustering algorithm to compress the state space.

The clustering is based on a collaborative filter [42]. They report good performance on datasets but do not perform experiments with real participants [42]. Liebman et al. take a similar approach with regard to the state and action representation they then use Monte Carlo Tree Search to approximate solutions [51]. Chi et al. also take a very similar approach but deal with the curse of dimensionality by classifying songs into four emotion classes. They use these classes to represent the songs. An action is also choosing an emotion class instead of picking a song [21]. Wang et al. take a different approach and model the problem as a multi-armed bandit [101]. In the model of utility a listener gets from a song they add a novelty factor. [101]. This has to be incorporated in the utility since the multi-armed bandit approach does not have a state to save history in unlike the previous three approaches. They also use approximations to deal with the large action space [101]. Mostly our research will differ from this existing research by explicitly targeting the regulation of emotions and not just using reinforcement learning to recommend the "most preferred" song for a listener.

In Chapter 4 and appendix E we take a closer look at MDPs and POMDPs (used in reinforcement learning) and how they can be applied to the topic of this thesis.

2.9. Conclusion

From this chapter we can conclude a few things; In the treatment of PwDs musical interventions have been applied in various ways. One of these is the IMI, focusing on using individualized music to reduce agitation in the PwD. How this can be integrated into tools or perhaps even automated has seemingly not been studied. This leads us to our first research question "How can we incorporate the IMI into technology to care for PwDs?". Since the focus is on reducing agitation, combining it with current research on affective music recommendation seems like a logical step. However current research has a lack of focusing on induced emotions for emotion based music recommendation systems. This observation lead us to our second research question: "Assuming knowledge of a list of preferred music for a user, can we generate playlists which better regulate emotions than a random shuffle?". This question will be further explained in Chapter 3. In datasets we also notice either a focus on perceived emotions or the usage of short excerpts. The only exception to this is Emotify but one of the goals of their research was to also test a music specific scale (GEMS). Therefore this dataset is not as easy to combine with existing research on affective computing.

3. Individualized Music Intervention: A Functional View

In this chapter we give an answer to our first research question: "How can we incorporate the IMI into technology to care for PwDs?". This is done by dividing it into different functional components which have to be implemented to implement the IMI. This allows for flexibility to adapt to different contexts in which the system operates. In an app for current use in a healthcare facility some of the components might still be fulfilled by a healthcare worker. While in the far-future some of these components might be replaced by research from for example affective computing to automate the process in an assistive robot for PwDs. We start by making an overview of the prescribed guidelines of the IMI. We then generalize this to the functional components. Each component has its own responsibilities and interfaces to connect to other components. For these components we will explain how they are currently implemented in IMI. Finally we will also show how these components might be implemented and integrated in future systems.

3.1. The Guidelines

In Chapter 2 we already discussed the IMI as described by Gerdner [36]. This section provides an summary of the guidelines by Gerdner. The guidelines first have an assessment period followed by applying the intervention itself, we will cover them in separate sections here as well.

3.1.1. Assessment

The assessment prepares for the intervention. During the assessment the PwD should be monitored. This monitoring lasts a certain period of time, the guidelines give an example of one week [36], but do not prescribe an exact amount of time. The observations are based around observing agitation and determining if there are certain identifiable temporal patterns or observable causes for agitation. Any identified causes that can be removed should be removed first. Identified patterns can be used later to determine when the intervention should be applied. Gerdner notes that it is also important to be able to identify the PwDs music preference during the assessment. Otherwise the intervention will not be effective.

3.1.2. Intervention

We start by noting that the intervention can be carried out by a multitude of people such as different healthcare professionals and in a variety of scenarios such as, day care, long-term care, home care, ... The intervention itself is described as having the following steps by Gerdner [36]:

- First the preferences of the PwD are determined
  - If the PwD is able to, this can be done by asking the PwD for their preferences.
  - If the PwD is not able to do this, a family member or friend is asked for this information.
  - These preferences have to be made available somehow, e.g. by putting them on an mp3-player.
- The intervention has to be initiated at some point, the guidelines suggest it is most effective when it is started 30 minutes before the PwD usually is most agitated.
- The intervention itself:
  - It should last approximately 30 minutes.
  - It should be in a location where the PwD spends a lot of time.

- The volume must be set to a comfortable level.
- Speakers are preferred, but if it causes annoyances to other people (e.g. other residents at a facility) headphones can be used.
- If the PwD shows increased agitation or confusion the intervention should be stopped. And the playlist reassessed.

3.2. Components

Based on the guidelines described in Section 3.1.2 we made an overview of the functional components that have to be implemented to be able to fulfill the intervention as described in the guidelines. This section serves as an explanation for the component diagram, Figure 3.1. We will explain the responsibilities of each component. We will also discuss the different interfaces that exist between components. These interfaces can be machine readable interfaces (when both components are on a media player for example) or they can be human readable interfaces when one component is fulfilled by a machine and the other by a human. We will also explain how these component are implemented in the current implementation. This is shown Figure 3.2. With current implementation we mean the implementation as it is being applied already today by healthcare professionals. In the documentary "Alive Inside" [69] iPod Shuffles are used in the current implementation as media players, for the rest of the current implementation sections we will assume the use of an iPod Shuffle to fulfil some of the components. The main difference between Figure 3.2 and the general diagram Figure 3.1 is that components are assigned to either being fulfilled by the iPod Shuffle or by a trained worker/music therapist.

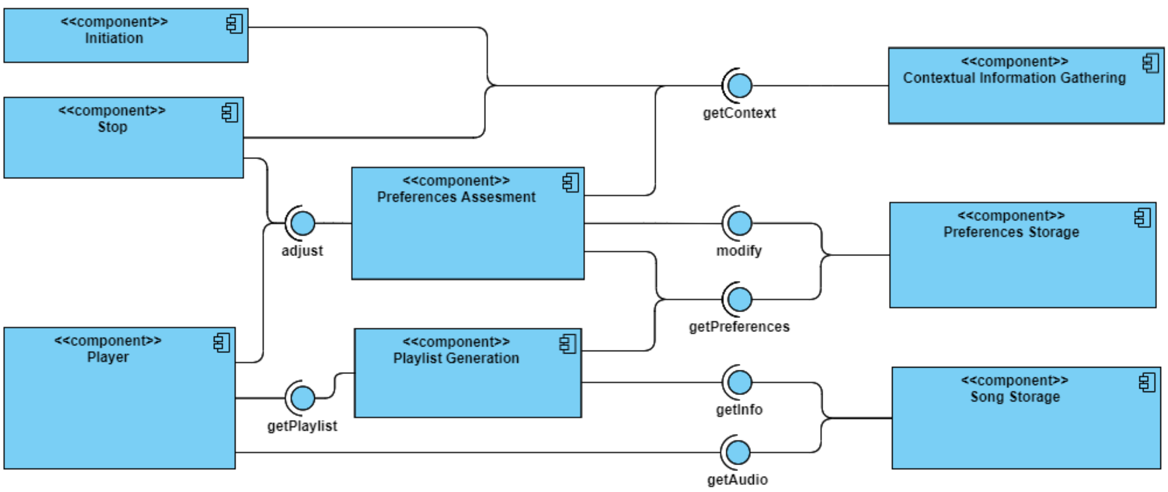


Figure 3.1: UML component diagram showing the components and interfaces

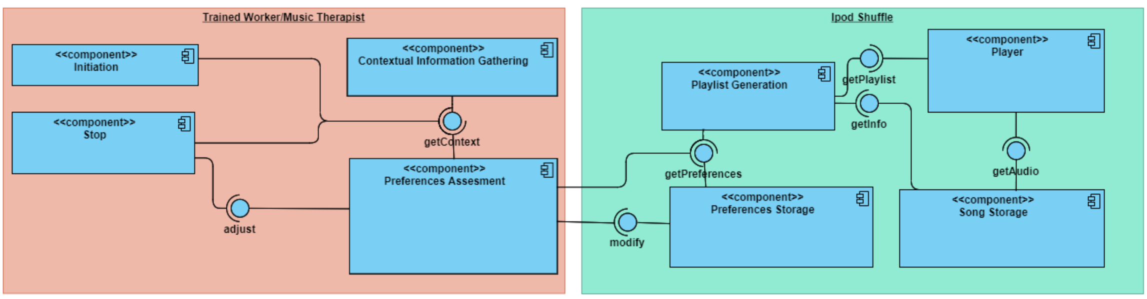


Figure 3.2: UML component diagram showing the division of components in the current IMI

We give an overview of each component and their interfaces.

3.2.1. Initiation

The responsibility of the initiation component is to initiate the intervention when it would help the PwD. It initiates the Player component. It gets information for its decisions from the Contextual Information Gathering component.

Current Implementation Currently this component is fulfilled by a trained worker, it can be the caregiver of the PwD or a healthcare professional at a facility for PwDs. They decide when the intervention should be started based on their own observation and the information from the assessment.

3.2.2. Stop

The responsibility of the stop component is to stop the intervention when the PwD has an adverse reaction. If necessary it should reassess the preferences of the PwD to prevent a future adverse reaction. It stops the player component, and should be able to reassess the preferences via the Preferences Assessment component. It gets information for its decisions from the Contextual Information Gathering component.

Current Implementation As with the initiation this component is fulfilled by a trained worker. If they observe adverse effects they have to stop the intervention and reassess the song selection.

3.2.3. Contextual Information Gathering

This component has the responsibility of gathering contextual information required by the other components. E.g. the emotional state of a PwD, or other contextual information such as the setting of the room, or other people present.

Interfaces

getContext: Allows other components to get contextual information which has been gathered by this component.

Current Implementation Again this is done by the trained worker. They observe the PwD and then fulfil the responsibility of the start and stop component themselves based on the observed information.

3.2.4. Preferences Assessment

The responsibility of this component is to assess the preferences of the PwD, in a format suitable for the Playlist Generation component. It should also be able to adjust the preferences if required by the Stop or Player component. In some implementations it could get information from the Contextual Information Gathering component for deciding on adjustments.

Interfaces

adjust: Allows other components to ask the Preferences Assessment component to adjust the preferences. (e.g. when adverse affects are noticed and the intervention is stopped)

Current Implementation Again fulfilled by the trained worker. During the assessment period they will talk to the PwD, family, friends, ... to determine what music the PwD prefers.

3.2.5. Preferences Storage

This component stores the preferences gathered by the Preferences Assessment component so that it is available to the Playlist Generation component.

Interfaces

modify: Used by Preferences Assessment to change the preferences.
getPreferences: Allows other components to retrieve the stored preferences.

Current Implementation The preferences in the current implementation is a list of songs that came out of the assessment. This list is stored on the iPod Shuffle in the current intervention.

3.2.6. Song Storage

Stores the audio files of songs that can be played by the player. It can also contain additional information about the song such as the artist, title or features used by the playlist generator.

Interfaces

getInfo: Retrieves info about stored song which can be used by the Playlist Generation component.
getAudio: Gets the audio of a certain song. Required for playback by the player.

Current Implementation The iPod Shuffle also implements the song storage component as it will have the audio files on the songs that the PwD preferred.

3.2.7. Player

This component, once initiated, gets a playlist from the Playlist Generation component and plays the audio files of the songs from the Song Storage. Depending on the implementation we could imagine the Player component adjusting the preferences via the Preferences Assessment component continuously. This is different than if we strictly look at the guidelines since they describe adjusting it on adverse reactions only.

Current Implementation The iPod Shuffle also servers as the player as it will playback the stored songs. It can be paired with headphones or speakers to play the audio.

3.2.8. Playlist Generation

This component generates the playlist that should be played once requested to by the player. This could be a continuous process; after each song the player asks for a new song so that new contextual information can be taken into account. Or it could be a one time event where a full playlist is generated once when an intervention starts. This component is not really covered in the guidelines by Gerdner [36], but inherently appears in the intervention, in the medium on which songs are saved (e.g. an mp3 player might randomize, while a cd-player might play sequentially). More complex playlist generation techniques could also be interesting to investigate.

Current Implementation When using an iPod Shuffle as the player then this component is also implemented in the iPod. Since these will randomly shuffle the stored songs.

3.3. Near-future: Implementation in support app

In the near future we could envision an application to support workers in applying the IMI. The main idea behind this app would be to support the workers by taking care of the preference storage, song storage and playlist generation aspects. It would mostly be a replacement for the mp3-player/iPod Shuffle in the current implementation, while allowing for more specific playlist generation aglorithms than a random shuffle.

3.3.1. Initiation And Stop

Similarly to the guidelines a trained worker or family/friends would initiate the intervention. The same goes for stopping the intervention.

3.3.2. Contextual Information Gathering

This could be provided to the application by trained workers or by family or friends. Additional methods of gaining feedback through for example the camera could be used as well. Emotion recognition algorithms could be one of them. Therefore the implementation of the responsibilities of this component could be spread across both solutions involving workers and automated technological solutions

3.3.3. Preferences Assessment

Preferences Assessment is the responsibility of the app. However the contextual information from the worker or other people provided to the application could be of use.

3.3.4. Preferences Storage

Preferences of a user a stored in a database. These preferences are tags on songs (e.g. liking a song).

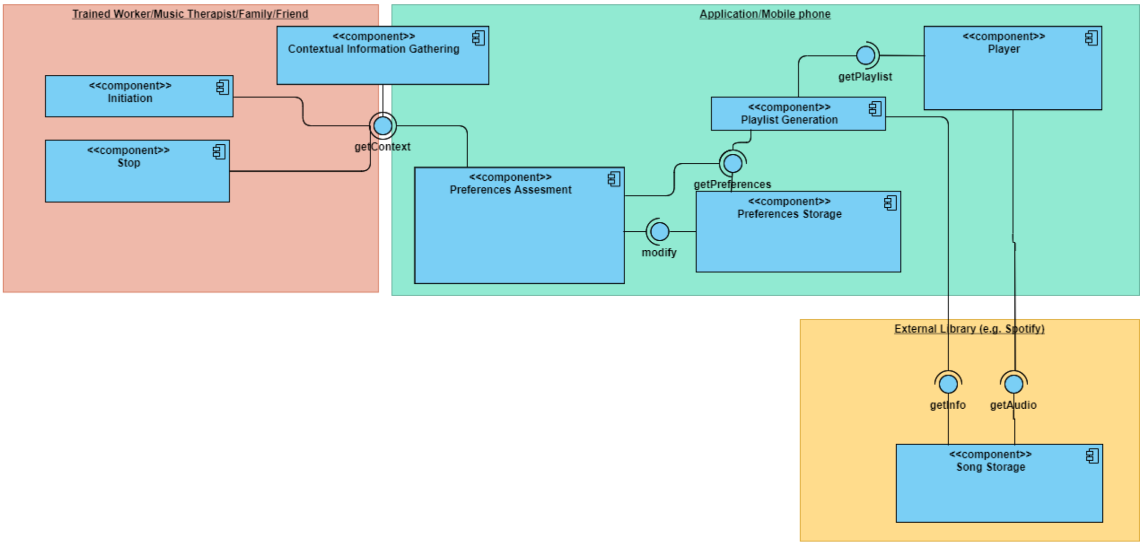


Figure 3.3: UML component diagram showing the division of components in the current IMI

3.3.5. Song Storage

This can be done in various ways. An application could have their own library of songs in their own database. But if a wide variety of music is to be provided, more realistically an external library is used through publicly available APIs or through partnerships with other services such as Spotify.

3.3.6. Player

In the case of the application, the application itself (and in part the mobile phone with its speakers, or headphones) serve as the player.

3.3.7. Playlist Generation

Playlist generation in such an application could have simple solutions, we could keep the random shuffle from the current implementation. However an app specifically geared towards IMI would allow the implementation of a Playlist Generator which is more complex and attempts to take contextual information, such as the affective state, in mind in the playlist generation. This more complex case is interesting to study and is the focus of the second research goal of this thesis. The chapters after this one will focus on this second goal and the more complex Playlist Generation component.

3.4. Far-future: Automated IMI

In the far future we could perhaps imagine the whole process of the intervention being automated. Such an automated approach could then be implemented in for example an assistive robot. To reach this goal a lot of research is still necessary however. In this section we want to highlight for each component some active areas of research that are interesting and what still would require work.

3.4.1. Initiation and Stop

With an advanced enough Contextual Information Gathering component it could be possible to develop an algorithm which predicts when the intervention should be applied. This would require research into which contextual information is important for this decision and how it can be used to make this decision. The same goes for stopping the intervention in case of adverse effects.

3.4.2. Contextual Information Gathering

Research into affective computing will offer a lot of new avenues in this regard. We can apply systems discussed in Section 2.7. We can think of systems such as facial expression recognition or those that use physiological signals from wearables. Wearables seem like a promising source of additional contextual information for PwDs. However the challenges in Section 2.7 will have to be kept in mind. For a more detailed look into affect recognition using wearable and the current challenges we also refer readers to a review by Schmidt et al. [76].

3.4.3. Preferences Assessment

For preferences assessment advances in more "traditional" (non-context aware) recommendation systems could be of use. These continuously get better at finding music to users according to different criteria such as preference, recommending more new music, ... These could also be used to for example try to translate a limited list of preferences to a larger list of songs a PwD might like or to use other information to infer preferences of PwD. As with the Contextual Information Gathering component care has to be taken that research from these other fields applies to our target group of PwD and if not they might have to be adjusted or redesigned for this target group.

3.4.4. Storage

In terms of storage (both for preferences and songs) we do not expect current advances in research to offer a lot of new avenues. Important in the future success of technologies incorporating this automatic intervention will be the access to a music library. Therefore we do see co-operation with current services that provide a large music library (such as Spotify) as vital.

3.4.5. Player

What the player (physically) looks like will depend on the application. We could envision it being incorporated into an assistive robot for example in which case the robot would fulfil the responsibilities of the player component.

3.4.6. Playlist Generation

As mentioned in the Near-future section there are a lot of possibilities for Playlist Generation. Later on in this thesis we propose one based on Markov Decision Processes. Simpler techniques such as a random shuffle could also still be viable. Future research, especially in contextual music recommendation could also offer new insights in how these playlist could be generated. Important for this component is that research is done into what works to reduce agitation for PwDs.

3.5. Playlist Generation

As mentioned in the introduction we will have an experiment based on our own playlist generation component. Currently the playlist generation in IMI is often randomized by the media player. However it seems like emerging fields such as affective computing and already more established fields such as music search and recommendation could improve on basic techniques such as a random shuffle. Stakeholders in the music therapy field also expressed a wish for a look into how playlists can be generated, in particular interest was expressed in playlists with certain effects on the mood (e.g. calming, activating, ...). The focus on affective states also fits well in the context of IMI where the aim is to use the music to reduce agitation.

We therefore considers playlist generation in IMI as the following problem:

Definition 3.5.1 *Given a start emotional state ES_s, consisting of a valence and arousal value, and a target emotional state ES_t, consisting of a valence and arousal range, generate a playlist which moves the listener's emotional state from ES_s to a value in ES_t, while keeping in mind the preferences of the listener.*

The major difference with this approach to emotion based music recommendation approaches mentioned in Chapter 2 is the focus on inducing/regulating emotions. As summarized in Chapter 2 most current research focuses more on finding music that fits the current affective state or on regulating specific physiological signals such as the heart-rate.

If we can devise a simple algorithm which performs better then this randomization it could be used in many contexts. It could be used in a support app in the near future as described above. In the far-future a similar algorithm could also be implemented in the described assistive robot.

This combination of novel research and wide applicability of the lessons learned make it an interesting component to focus on. Components such as the song storage and preference storage are already solved problems since they can be implemented with a database or with music streaming services such as Spotify. Preference assessment is already being studied heavily in the form of general music recommendation, with general meaning not emotion related. Players can already be implemented in a wide variety of form factors (e.g. phone, computer, mp3-player, ...). Contextual Information Gathering is already being studied in a myriad of ways for context-based systems. Finally the initiation and stop component will be heavily dependent on what kind of contextual information is available and can currently already be solved adequately by workers at a facility for example.

To conclude research into this problem seems interesting and promising. It can combine many emerging fields such as affective and contextual computing but does pose a novel research avenue. Furthermore a simple algorithm which already improves over randomization could be applied both in the very near-future, integrated in for example a support app, or in the far-future integrated in for example an assistive robot.

4. Markov Decision Processes

In this chapter we start by explaining Markov Decision Processes (MDPs). Beginning with a definition of MDPs and some methods that can be used to solve MDPs. Finally, we move on to different ways in which this can be applied to the problem of playlist generation in the IMI. While there are many possibilities we could consider to generate playlists, we think MDPs offer an interesting option. Modelling the problem of playlist generation in IMI allows use to use existing research into reinforcement learning to solve the problem. Furthermore (PO)MDP formulations have been used before in (contextual) music recommendation as explained in Section 2.8.

4.1. Markov Decision Process Definition

In this section we give a short, formal definition of MDPs. This section is heavily based on the chapter by Van Otterlo and Wiering [96], interested readers can find a more in depth explanation there. They give the following formal definition [96]:

Definition 4.1.1 *A Markov decision process is a tuple* $< S, A, T, R >$ *in which* S *is a finite set of states,* A *a finite set of actions,* T *a transition function defined as* $T : S \times A \times S \rightarrow [0, 1]$ *and* R *a reward function defined as* $R : S \times A \times S \rightarrow \mathbb{R}$

State

A state is a unique characterization of what is important to the problem at hand [96]. Van Otterlo and Wiering give the example of a chessboard where the state would be a complete configuration of board pieces [96].

Action

Actions can be used by an agent to control the system state [96].

Transition Function

When an action a is taken in state s, a state transition occurs from s to a new state s', this is based on a probability distribution over the set of possible transitions [96].

Markov Property

Important for the transition function is that it should satisfy the *Markov Property* [96].

Definition 4.1.2 *A system is Markovian if the result of an action only depends on the current state and the chosen action, not on previous actions and states.*

However a system where the last k states are sufficient to determine the result of an action, can always be transformed to a problem that is *Markovian* [96].

Reward Function

The reward function gives a reward for being in a state or performing an action in a state [96]. The reward function can be defined in three interchangeable ways [96]:

$$\begin{aligned} R &: S \rightarrow \mathbb{R} \\ R &: S \times A \rightarrow \mathbb{R} \\ R &: S \times A \times S \rightarrow \mathbb{R} \end{aligned} \tag{4.1}$$

Policies

A policy is a function which outputs an action a for each state s. A policy can be either deterministic or stochastic [96]. They are defined as:

$$\pi : S \rightarrow A \qquad \text{Deterministic} \tag{4.2a}$$

$$\pi : S \times A \rightarrow [0,1] \qquad \text{Stochastic} \tag{4.2b}$$

Optimality Criteria

To define how good a certain policy is, we have to define an optimality criteria. There are three basic models of optimality. The finite horizon; discounted, infinite horizon and average reward [96]. When a task is either continuing or when the length of the task is unknown the discounted, infinite horizon model is best suited [96]. This is the case for our task since we do not know beforehand how many songs the user will listen to. Therefore we take a closer look at the discounted infinite horizon model. Its formula is given by:

$$E\left[\sum_{t=0}^{\infty} \gamma^t r_t\right] \tag{4.3}$$

Where r_t is the reward at time-step t and γ, $0 \leq \gamma < 1$ is the discount factor [96].

Value Functions

Value functions are used to link the optimality criteria to the policy. There are two types of value functions. Firstly there are V-functions, $V^\pi(s)$, it is the expected value of starting in state s and following policy π from that state. Secondly there are Q-functions, $Q^\pi(s,a)$, it is the expected value of taking action a in state s and following policy π afterwards. The V-function represents how good it is for an agent to be in state s, the Q function represents how good it is for an agent to take action a in state s [96]. Using the optimality criterion given in Equation 4.3 these are given by:

$$V^\pi(s) = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s\right\} \tag{4.4}$$

$$Q^\pi(s,a) = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s, a_t = a\right\} \tag{4.5}$$

These can be rewritten in terms of a Bellman Equation [96]:

$$V^\pi(s) = \sum_{s'} T(s,\pi(s),s')\Big(R(s,\pi(s),s') + \gamma V^\pi(s')\Big) \tag{4.6}$$

$$Q^\pi(s,a) = \sum_{s'} T(s,a,s')\Big(R(s,a,s') + \gamma Q^\pi(s',\pi(s'))\Big) \tag{4.7}$$

Greedy Policy

Having defined the above we now want to find the *optimal policy*, π^*. This is the policy for which $V^* = V^{\pi^*}$ and $\forall s \in S : V^*(s) \geq V^\pi(s)$ [96]. This is given by the following equation [96]:

$$V^*(s) = \max_{a \in A} \sum_{s' \in S} T(s,a,s')\Big(R(s,a,s') + \gamma V^*(s')\Big) \tag{4.8}$$

This can be defined analogously for the Q-function:

$$Q^*(s,a) = \sum_{s' \in S} T(s,a,s')\Big(R(s,a,s') + \gamma \max_{a' \in A} Q^*(s',a')\Big) \tag{4.9}$$

These optimal V- and Q-functions can be used to define a policy that selects the optimal action given the function. This is called the *greedy policy* and can be defined both in terms of the V- and Q-function:

$$\pi^*(s) = arg\max_{a} \sum_{s' \in S} T(s,a,s')\Big(R(s,a,s') + \gamma V^*(s')\Big) \tag{4.10}$$

$$\pi^*(s) = arg\max_a Q^*(s,a) \tag{4.11}$$

If we have a full model of the problem (reward function and state transitions) we can use dynamic programming to compute the optimal policy from the V- or Q-function, this is the *model-based* approach [96]. In a *model-free* in this case we need to make a trade-off between exploration and exploitation. This is done by utilising different policies such as ϵ-greedy and softmax which will be explained in the next sections. Combinations between the *model-free* and *model-based* also exist [96].

ϵ-greedy Policy

The ϵ-greedy policy is a fairly simple trade-off policy. It selects the best action according to the greedy policy with probability (1-ϵ) and a random action with probability ϵ [96].

Softmax Policy

The softmax policy takes an action with a probability that is weighted by its Q-value:

$$P(a) = \frac{e^{\frac{Q(s,a)}{T}}}{\sum_i e^{\frac{Q(s,a_i)}{T}}} \tag{4.12}$$

T is a *temperature* parameter, the higher T is the more random the strategy will be [96].

4.2. Solving MDPs

A myriad of methods to solve MDPs have been devised. They can be categorized into two major categories, namely model-based and model-free solutions [96].

4.2.1. Model-based

As the name suggest these assume a model of the MDP is available. This model can then be used to calculate policies and value functions using the Bellman Equations (Equations (4.6) and (4.7)) through dynamic programming algorithms such as Policy Iteration [96].

4.2.2. Model-free

Again as the name suggests these methods do not assume the availability of a model of the MDP. Instead they interact with the environment and through this estimate the value functions [96]. This introduces an exploration-exploitation trade-off where a trade-off has to be made between selecting optimal actions based on current information or obtaining more information to improve the model on which decisions are made [96]. Once an agent has a sufficient model, similar techniques as in the model-based approaches can be used to determine the policy [96].

4.3. Applied to IMI

If we look at the setting of IMI and in particular the problem of generating playlists within IMI then model-based approaches seem more suitable. In the model-free approaches we have to explore the environment to find the best actions, this could mean first making a lot of "wrong" choices, which could cause distress to the PwD. Therefore we will look more into model-based approaches. One question then becomes which model do we use?

4.3.1. Songs as Actions

One method, also used in contextual music recommendation as mentioned in Section 2.8, is that the action an agent takes is picking a song. Inherently however this means that the state-action spaces grows with the number of songs that can be picked. With this the computation time of solutions to the MDP formulation will also grow. Approximation algorithms might therefore be necessary with this formulation if a big library of songs is used.

States

We also have to define the states. The state of the agent could be the current emotional state of the listener. However if we do this then we are left with defining T which would be a probability distribution from the current emotional state to new emotional states after listening to a song. This would be a complex function to learn for which we at the moment do not seem to have enough data for. This method also lack memory. This transition function most likely also depends on other factors such as the song that has just been listened to. The emotional reaction of a listener is most likely not the same when a song is picked twice in a row. Incorporating such additional factors only makes it more difficult to approximate it with the data we have available right now. The question then also becomes how much memory do we have to add? Do we only keep that last played song, the last ten songs, ... The more history we add the more our state space grows.

Rewards
if we define the reward function as $r(s_t, p_t, s_{t+1}) = \mathbb{R}$ then since our goal is to reach the target emotional state it seems we include a reward regarding this. We could give a reward based on how close s_{t+1} is to the target emotional state or the difference in distance to the target emotional state between s_t and s_{t+1}. Additionally we could use the reward to model other aspects that we might find important in the selection of a next music piece. For example we could incorporate a similarity measure to previously picked songs so that picking similar songs incurs a higher reward.

Conclusion
If we want to use songs as actions we have multiple options to model both the states and the rewards. However to deal with the large continuous state pace and large action space we will need to develop approximation algorithms as in the other recommendation systems that use this approach (Section 2.8), for which we currently have little data. Therefore it seems infeasible to build these currently. More data for research would have to be gathered first.

4.3.2. GridWorld, A heuristic approach

This model is inspired by a common abstraction used in Reinforcement Learning, that of a gridworld in which an agent can move one grid at a time. Some grids in the world contain rewards and others might incur punishments. Many variations on this problem have been thought of. GridWorld inspired algorithms have also been proposed in relation to recommendation systems [22, 108] We can apply a similar idea to the problem of playlist generation. To model the problem as a GridWorld we start by dividing the valence-arousal space in an nxn grid. The valence and arousal values are converted to this nxn grid. The agent starts in the grid corresponding to the start emotional state of the listener. The actions the agent takes are moving one place on the grid (up/down, left/right or diagonally). To experiment with this idea we made a prototype which is discussed in the next chapter. This prototype will also make the idea behind this method of modeling clearer

4.4. GridWorld, Prototype

4.4.1. Assumptions

To develop the algorithm we made some key assumptions:

- We have knowledge about the user preferences.
  - For example a list of rated songs, or an algorithm which computes ratings for users on songs.
- We have knowledge about the emotion a song induces for a user.
  - We also assume that the transitions are deterministic, in other words the emotion of a listener always follows the emotion of the songs played.
    - This choice is made as a simplification since we have too little data to model how the transition function would work.
- We want subsequent recommendations to not be too far away from each other in the valence-arousal dimensions. In other words we assume that a song will only move the emotional state of a user by a certain amount.

4.4.2. Implementation

We use reinforcement learning to solve this problem. To do so we first model the valence-arousal scale as a grid world. The songs we have knowledge about are also mapped to this same grid based on their expected induced emotion. This allows us to model the problem as a MDP process. With the following:

- State: The state is the square on the grid our agent (recommender system) is in.
- Actions: each time the agent can move one square up, down, left right or diagonally
- Reward: For each square the agent lands in we give a reward based on the preference of a user for songs in that square. When we reach the square corresponding to the target state we also give a reward.

A first prototype was developed based on data from Spotify:

- Spotify provides a list of all your saved songs and a list of 50 top songs.
  - For this experiment this was retrieved for my own account.
  - A square with one or more songs in my top list received a reward, R_t.
  - A square with one or more songs in my saved list received a reward ,R_s.
  - We also give a reward, R_g, for reaching a goal state
  - Finally we define, R_e, to allow us to define a reward (or punishment) for a square that is empty. If it is not specified it is 0.
- Per song Spotify gives a "valence" and "energy" value between 0-1.
  - This was used as the expected induced emotion of a song and as explained used to associate a song with their square on the grid.

For the first prototype we divide the valence-arousal scale in a n by n grid, the initial examples show $n = 10$. This gives the gridworld from Figure 4.1. Green squares are squares with top songs, orange squares with saved songs and the purple X denotes a possible target state.

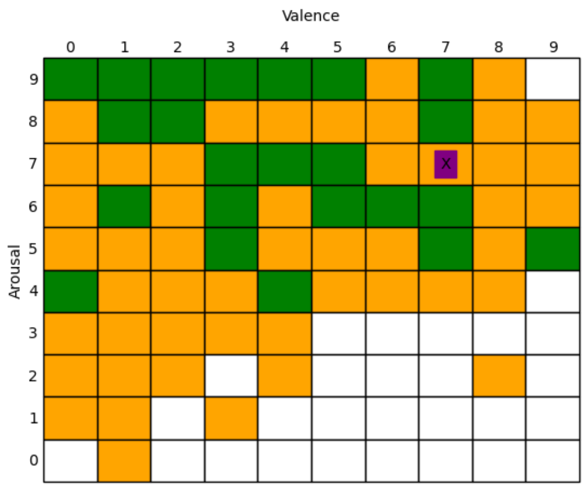


Figure 4.1: Gridworld based on Spotify saved and top songs

We then apply value iteration and compute the greedy policy once the values have converged. We use a discount of 0.9 and the reward for the goal state is 10. Figure 4.2 shows the actions the greedy policy would take based on these values for the gridworld of Figure 4.1. The action our recommender system would take in a certain state is denoted by the direction the arrow points in.

If we know the start state of a user and the target state the path from one to the other can be determined trivially from figure 4.2. To get the recommendations in our prototype for each square the agent visits (including start and end state) we randomly picked a song from the list of saved or top songs (depending on whether the square is a green or orange square) associated with that square. An example of this is shown in Figure 4.3. The blue square is the start state. The squares with red in them show the path the agent takes and the list on the right shows the associated recommendations made.

4.4.3. Results

In this subsection we document some of our observation when testing the prototype.

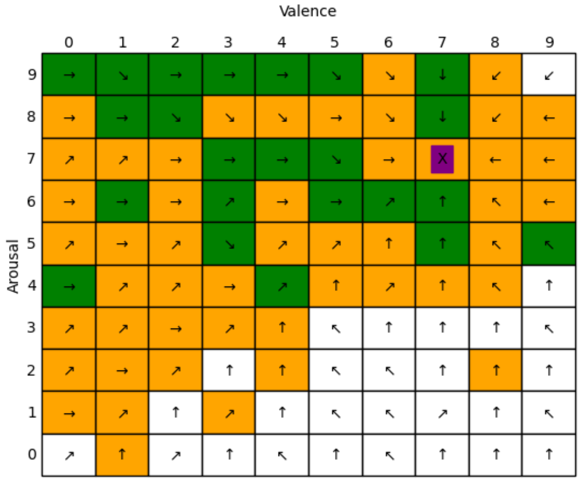


Figure 4.2: Actions of a greedy policy after policy iteration with a discount of 0.9, $R_t = 2$, $R_s = 1$, $R_g = 10$

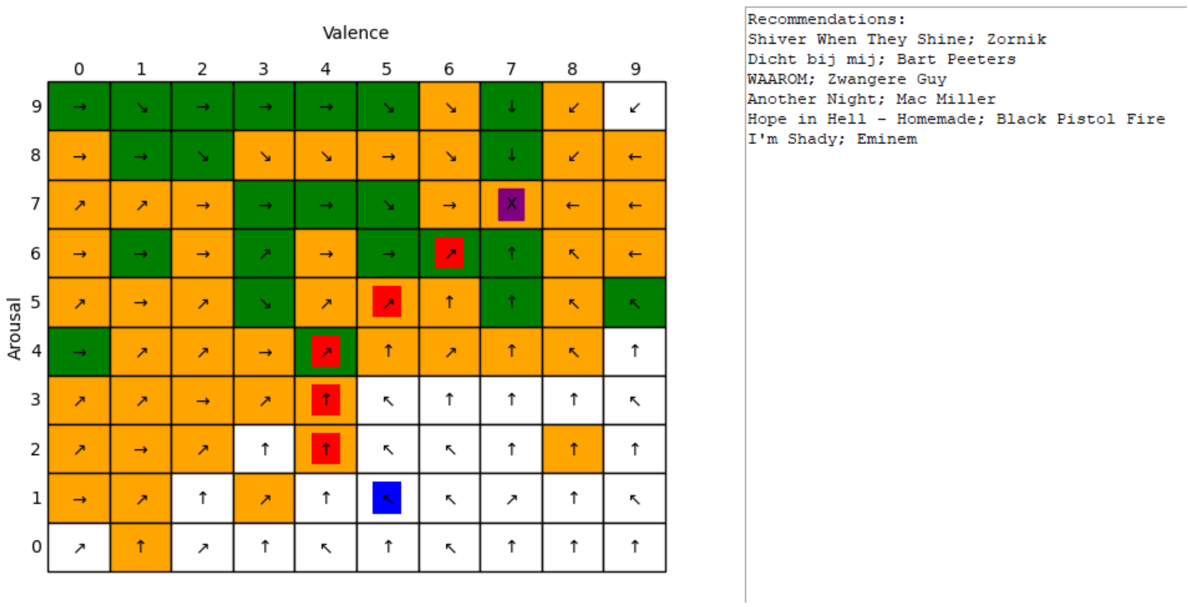


Figure 4.3: Example path of the recommender system with its recommendations

Running Times

The running time of our algorithm will depend heavily on the grid size, since the size of our state space is n^2. To get an idea about which kind of grid sizes are feasible to solve on current hardware with simple value iteration we performed a test checking the running times with varying grid size of 5, 10, 20, 30, 40 and 50. This test was performed on a Intel Core i7-9750H running at 2.60 GHz. The results shown are averages of 20 runs. Table 4.1 shows the average run time and the average number of iterations for each grid size. We notice that the running time indeed goes up quickly with the grid size. The number of iterations only increases slightly. Figure 4.4 shows the results in a graph and confirms this that it seems like the running time increases with n^2. If we want to generate playlists quickly with current hardware we will need to stick to smaller grid sizes. However an easy optimization would be to use policy iteration instead of value iteration. In policy iteration the search is stopped when the policy converges instead of the values. These tend to converge more quickly and therefore would provide results faster.

n	Average run time (s)	Average iterations
5	0,17	178
10	0,81	181
20	3,65	190
30	8,57	197
40	16.1	202
50	26.0	210

Table 4.1: Table comparing performance with increasing grid size. With $\gamma = 0,9, R_g = 10, R_t = 2, R_s = 1$. Goal states are uniformly randomly generated. Average of 20 runs.

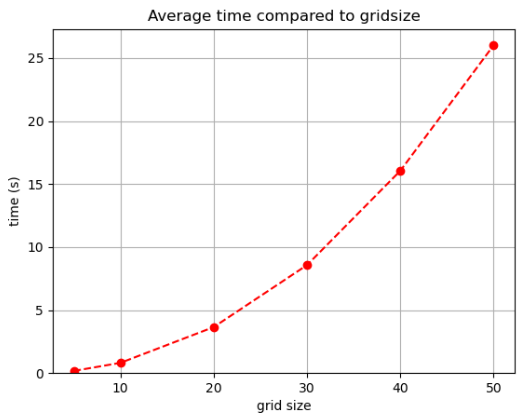


Figure 4.4: Table plotting average run times averaged over 20 runs compared to increasing grid size. With $\gamma = 0,9, R_g = 10, R_t = 2, R_s = 1$. Goal states are uniformly randomly generated.

Parameters

When building the prototype we noticed that a few things are important to keep in mind when setting parameters. First of relative sizes of the rewards are very important. Giving a too small reward to the goal state for example makes it so that the recommender system might not reach the goal state. On the other hand, giving too little reward for grid squares with preferred music makes it so the recommender system does not really take into account the preferences of a user and just goes to the goal state as quickly as possible. The discount factor also has a similar trade-off. Putting it too close to 1 means it will go straight to the goal state, putting it too close too low will make it so the agent only considers immediate rewards and then the agent never reaches the goal state.

4.4.4. Discussion

In terms of running times with current hardware it seems it would be possible to use it in real time for a recommender system. However the running time will heavily depend on the grid-size since this determines the size of the state space.

Furthermore we also find that the recommendations we receive depend heavily on some parameters chosen. Namely on the rewards and the discount factor. When implementing the system, we will have to look into how to determine these parameters. It is also possible that these parameters have to differ for different users. Important to note is as well that choosing the parameters wrong could make it so the recommender system never reaches the goal state.

We also note that one of the limitations is that this MDP assumes we know the current state perfectly. In practice this is doubtful, POMDPs can deal with uncertainty in observing the state and could be interesting to look into.

4.5. Conclusion

In this chapter we gave a short overview and definition of MDPs. While there are undoubtedly a lot more methods one can think of to model the problem of playlist generation in IMI as an MDP we presented two in this chapter, songs as actions and GridWorld. We also discussed that some methods require data which is currently not available and therefore might be more suitable for future research. With GridWorld we showed a simple prototype based on simple heuristics to generate playlists. While there are a lot of design considerations with this method, such as the grid size, how to determine preferences of a listener, and how to pick songs when in a certain grid we showed that a simple prototype of this algorithm is currently possible with data from Spotify. Such a prototype could therefore be interesting for some initial research. In the next chapters we will discuss the design and results of an experiment we conducted based around this prototype. However we also noted that the MDP formulation inherently has a limitation in that it assumes a fully observable environment, while in real life applications our method to determine the emotional state for example might have noise. One way to deal with this would be to extend our approach to a Partially Observable Markov Decision (POMDP). Partially Observable Markov Decision Processes are defined similarly to Markov Decision processes with the key difference that we do not assume to fully know the state the agent is in. Instead we have observations and associated with these observations are probabilities for a certain state. Due to this POMPDs are able to deal with decision making with uncertain sensing [89]. However additional decisions as to how to model this uncertainty have to be made. With limited data and no certainty about which emotion recognition system might be hooked up to our playlist generator component we decide to opt for sticking with a more simple MDP formulation for our experiment. For interested readers Appendix E contains a chapter from a previous version of this report with a bit more details on POMPDs and how they might be applied. For more information on POMDPs in general we refer interested readers to *Chapter 12: Partially Observable Markov Decision Processes* by Matthijs Spaan from the book *Adaptation, Learning and Optimization: Reinforcement Learning* [89].

5. Experiment

In the previous chapter we explained how MDPs could be applied to the problem of playlist generation in IMI. In this chapter we explain an experiment relating to this topic. We start by introducing the experiment and explaining the goals of this experiment. We then move on to the method of the experiment, this includes the participants, design, measures and procedure of the experiment. We briefly discuss how we implemented this experiment and finally we briefly discuss some issues we encountered when we launched the experiment for the first time.

5.1. Introduction

As explained in our introduction, Chapter 1, this thesis contains two main research questions and this experiment deals with the second of our research questions. This second research question is "Assuming knowledge of a list of preferred music for a participant, can we generate playlists which better regulate emotions than a random shuffle?". In the previous chapter we have showed how we can apply MDPs to the problem of playlist generation with a target affective state. This experiment will mainly aim to test how this differs from a random shuffle in terms of emotion regulation with real participants. Our MDP solution will be further explained in Section 5.2.2. This approach differs from most current research into emotion-based music recommendation by focusing on induced emotions. As also mentioned in Chapter 1 this experiment has two subgoals. The first is that we want to generate more data which can be used to study induced emotions and in particular playlist generation for emotion regulation. With this subgoal we hope to enable and stimulate more research into this topic. Current datasets are not really suited for the type of research we want to do. As mentioned in Section 2.6 often they suffer from focusing on perceived or expressed emotions such as PMEmo [111] or the 100 songs dataset [85]. Other datasets such as AMG1608 [19] or DEAP [50] only had participants listen to short fragments. We hope to build an initial dataset which can be used for some initial research into this topic. However we expect it to be somewhat limited in the number of participants. As also mentioned in the introduction, Chapter 1, the second subgoal is to check if we can find a noticeable difference in emotional reaction to songs that have memories attached to it. This second subgoal occurred more naturally due to the need to ask some questions during the experiment. Picking the topic of memory was motivated by by previous research showing that memories intensify induced emotions when watching music videos [29]. It therefore might offer additional avenues for future research, by for example integrating information about associated memories in the playlist generation component. In total we can say that the experiment has three goals, two of them also have an associated hypothesis:

- Test a simple MDP model and see if we can perceive a noticeable difference when we try to improve the valence of a listener compared to a randomly shuffled playlist.
  - ***H*1**: By using the MDP formulation the emotional state of a participant after listening to a playlist is more positive than the state of someone listening to a randomly shuffled playlist.
- Generate more data which can be used for future research into the topic of music recommendations to regulate emotions.
- Check if songs with memories attached to them have larger/different effects on the emotional state.
  - ***H*2**: Songs that have a strong memory attached to them will have a larger effect on the emotional state of a listener.

The experiment will be conducted completely online. Participants will visit a website on which they can partake in the experiment. As mentioned the rationale behind conducting a completely online experiment are the restrictions concerning COVID-19 at the time of the experiment. In the experiment we have a control group who listens to a randomly shuffled playlist of songs that they have saved on Spotify. With saved songs in this thesis we mean songs that the participant has "liked" on Spotify (for those familiar with Spotify: given a heart). The experimental group listens to a playlist generated by a deterministic MDP using the saved songs from Spotify. More info on the details of the playlist generation for both groups is given in Section 5.2.5. For both groups the size of the playlist is four songs. Each participant will be asked some background questions before listening to the playlist. A few questions will be asked during the playback of songs. Finally some questions will be asked after they have finished listening to the playlist. The next section described the method of the experiment.

5.2. Method

In this section we cover the method of the experiment this included the participants, design, measures and procedure. We end on a subsection which explains how the playlists for this experiment are generated.

5.2.1. Participants

For the experiment we want "regular" participants, that is people without dementia. We choose to initially focus on this group because they can provide more feedback. They also can do the experiment online by themselves which was advantageous with the restrictions around COVID-19 at the time of performing the research. Even though this means the data generated is not from PwDs we think that some lessons that are learned from this could be applied to playlist generation for PwDs. Due to this the research and dataset also becomes more generally applicable to other applications where affective playlist generation could be interesting. We do have some limitations on who can participate. Participants need to be 18+ and due to technical limitations a Spotify Premium account is required. The reasons behind this are described in Section 5.3. The experiment will be in English and therefore participants have to speak English as well. When designing the experiment we hoped to get around 60 participants, giving us about 30 participants for both the control and experimental group. However after running the experiment we had 34 people sign in with their Spotify account, 22 of these fully completed the experiment. Twelve of those participants were in the control group and ten in the experimental. To recruit participants the experiment was shared through posts on the author's social media pages (LinkedIn and Facebook). The experiment was also shared with friends and family of the author. At the TU Delft it was shared amongst computer science students and faculty staff. It has been shared with the member of the "Nederlandse Verenining voor Muziektherapie" (Dutch Association for Music Therapy). Dutch educational institutions who have music therapy related studies also shared the experiment with their students. This included several Universities of Applied Sciences (HAN University, Leiden, NHL Stenden, Zuyd, Utrecht) and two Universities of Arts (ArtEZ, Codart Rotterdam). The experiment was also shared through Reddit, on a forum dedicated to music therapy.

5.2.2. Design

This section briefly discusses the overall design of the experiment and explains how this design is used to test the hypotheses we introduced in the introduction. As mentioned in our experiment participants are assigned to either the control or experimental group, this happens randomly. Participants in the control group listen to a random shuffle of their liked songs and participants in the experimental group to a playlist generated by a MDP formulation. In our introduction we mentioned our two hypotheses. This design section will also be divided into those two hypotheses. To get information about the emotional state of a participant we use the AffectButton [13]. We also ask the participant to rate their change in emotional state after each song and after the whole playlist.

Hypothesis 1

- ***H*1**: By using the deterministic MDP formulation the emotional state of a participant after listening to a playlist is more positive than that of someone listening to a randomly shuffled playlist.
  - We first test this by using the answers from the AffectButton.
    - **H1.1:** The valence of the final emotional state of the participants in the experimental group is higher than that of the participants in the control group.
    - μ_c and μ_e are the means of the valence of the final emotional state of the control and experimental group respectively.

- ◇ $\boldsymbol{H1.1}$: $\mu_c < \mu_e$
- ◇ $\boldsymbol{H1.1_0}$: $\mu_c \geq \mu_e$
- ◇ We will use a t-test to test the hypothesis.

– We also use the data from the self reported emotion improvement.

- ◇ $\boldsymbol{H1.2}$: The self-reported improvement in emotional state will be greater in the experimental group than in the control group.
- ◇ e_i and c_i represent the answers on the Likert scale from a subject from the experimental and control group respectively.
- ◇ $\boldsymbol{H1.2}$: $P(e_i > c_i) > \frac{1}{2}$
- ◇ $\boldsymbol{H1.2_0}$: $P(c_i > e_i) \leq \frac{1}{2}$
- ◇ We will use a t-test to test the hypothesis.

To test this hypothesis we have two dependent variables: the final valence as reported by the AffectButton of a participant and the self reported change in emotion. We have one independent variable: the group of a participant (experimental or control).

Hypothesis 2

- $\boldsymbol{H2}$: Songs that have a strong memory attached to them will have a larger effect on the emotional state of a listener.
  - $\boldsymbol{H2.1}$: The distance in emotional state before and after listening to a song is larger when there is a memory attached to the song.
    - ◇ μ_m is the mean distance of the emotional state in the VA domain before and after a song to which a strong memory is attached. μ_n is the same for a normal song, without a strong memory attached to it.
    - ◇ $\boldsymbol{H2.1}$: $\mu_n < \mu_m$
    - ◇ $\boldsymbol{H2.1_0}$: $\mu_n \geq \mu_m$
    - ◇ We will use a t-test to test the hypothesis.

To test this second hypothesis we have one dependent variable: the delta in valence as reported by the AffectButton of a participant when listening to a song. We also have one independent variable: the memory tag a participant gave to the song (positive memory, neutral memory, negative memory, no memory, other memory).

These will be the most important analyses we do after carrying out our research since these very directly aim to help in attaining the goals we set for this experiment.

5.2.3. Measures

During the experiment quite a lot of data will be collected. In the previous section we already mentioned what some of our dependent and independent variables are. In this section we summarize all the data gathered during the experiment This is divided in two tables; Table 5.1 contains the experiment level data. This is the data we have once for each participant, think of for example their age, music listening frequency or responses to the end questions. Table 5.2 contains the song level data, for each experiment we have four of these and it contains information about the song itself such as the artist or title, but also contains the response the participant gave to questions relating to this song. The song level data and experiment level data are linked by the experiment id. This is also the format in which the data eventually will be published for future research. There will be two comma separate value files, one containing the experiment level data and the other the song level data. When reporting our results in the next chapter we will also use the names as they are given in these tables.

5.2.4. Procedure

This sections describes the procedure of the experiment and the rational behind it. We start by devoting a section to why we chose to use Spotify for our source of saved songs. After that we cover each major section of the experiment that a participant goes through chronologically. We include some images to show what the experiment looked like in this chapter, interested readers can find a full step by step walkthrough in Appendix G. The chronological order of the procedure is also captured in the chart in Section 5.2.4. While explaining the procedure we will refer back to this chart.

Name	Values	Translates to	Notes
experimentId	integer	Id of the experiment	This is the id that was given to the experiment in the database, this allows us to link song-level data to experiment-level data.
age	integer: [0-9]	Age range: [0: 18-20, 1: 21-25, 2: 26-30,3-35, 3: 36-40, 4: 41-45, 5: 46-50, 6: 51-55,7: 56-60, 8: 61-65,9: 65+]	Background question about age ("What is your age")
listeningFrequency	integer: [0-4]	Selected option: [less than once a month, once a motnh, once a week, once a day, more than once a day]	Background question question about listening frequency ("How often do you usually listen to music?")
performsMusic	integer: [0-1]	Selected option: 0 is no, 1 is yes	Background question about performing music ("Do you actively perform music?")
performsMethod	List of strings	List of selected option + item filled in other box	Background questions about which method of performing, only if performs music question is yes ("How do you perform music")
performsFrequency	integer: [0-4]	Selected option: [less than once a month, once a motnh, once a week, once a day, more than once a day]	Background question question about performing frequency ("How often do you usually perform music?")
nSaved	integer	Number of saved songs we were able to retrieve for this participant	
valence0...4	float: [0-1]	Valence in PAD space	Response to the AffectButton. Valence0 is before listening, valence1 after the first song and so on.
arousal0...4	float: [0-1]	Arousal in PAD space	Response to the AffectButton. Arousal0 is before listening, arousal1 after the first song and so on.
dominance0...4	float: [0-1]	Dominance in PAD space	Response to the AffectButton. Dominance0 is before listening, dominance1 after the first song and so on.
selfImprovementPlaylist	integer: [0-4]	5 point likert: [large decline, smalldecline, no improvement or de-cline, small improvement, largeimprovement]	Asked after listening to the full playlist, ("How would you rate your current emotional state compared to the emotional state before you listened to the playlist?")
playlistRating	integer: [0-4]	5 point likert (very poor, poor, neutral, good, very good)	Asked after listening to the full playlist, ("How would you rate the sequence of song you have just listened to?")
notes	string	Open question, can be any text	

Table 5.1: Table summarizing the experiment-level data generated during the experiment

Name	Values	Translates to	Notes
experimentId	integer	Id of the experiment	This is the id that was given to the experiment in the database, this allows us to link song-level data to experiment-level data.
spotifyId	string	Spotify id of the song	Identifies songs on Spotify
Artist	string	Name of the artist	Retrieved from Spotify
Album	string	Name of the album	Retreived from Spotify
Title	string	Name of the song	Retrieved from Spotify
timestep	integer: [1-4]	Index of the song in playlist	Time at which the song was played in the experiment (1 is the first song, 4 is the last song).
spotifyValence	float: [0-1]	Valence in VA space	The valence Spotify has stored for this song in their database
spotifyArousal	float: [0-1]	Arousal in VA space	The arousal/energy Spotify has stored for this song in their database
valenceBefore	float: [0-1]	Valence in PAD space	AffectButton response. Valence before listening to the song.
valenceAfter	float: [0-1]	Valence in PAD space	AffectButton response. Valence after listening to the song.
arousalBefore	float: [0-1]	Arousal in PAD space	AffectButton response. Arousal before listening to the song.
arousalAfter	float: [0-1]	Arousal in PAD space	AffectButton response. Arousal after listening to the song.
dominanceBefore	float: [0-1]	Dominance in PAD space	AffectButton response. Dominance before listening to the song.
dominanceAfter	float: [0-1]	Dominance in PAD space	AffectButton response. Dominance after listening to the song.

Table 5.2: Table summarizing the song-level data generated during the experiment

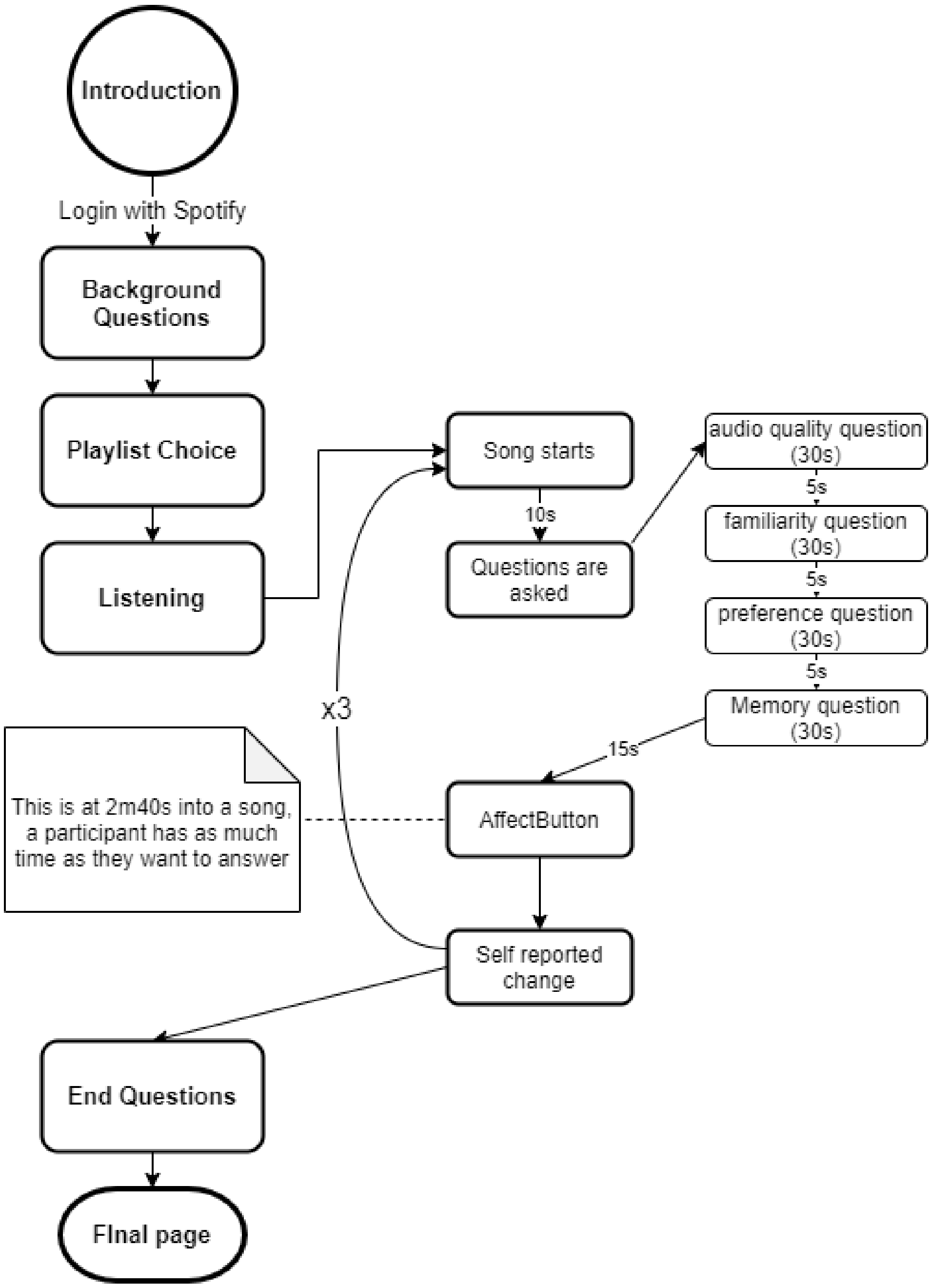


Figure 5.1: Graph showing each part the participant goes through in chronological order.

Spotify

Spotify is used to get songs that a participant has saved on Spotify. For these we assume that the participant "likes" them or has a preference for them. This can be used to generate a random list of liked songs for the control group. Spotify also provides a *valence* and *energy* value (in music related affect energy is sometimes used to denote arousal) for every song. We can use this to map the songs to the griwdworld from the deterministic MDP. Our focus will be on regulating the valence, this decision is further explained in Section 5.2.5. A Spotify premium account also allows us to play the full length songs in the website that implements the experiment.

The Introduction

The next sections will chronologically cover the parts of the experiment participants go through.

First the participants are introduced to the experiment. They receive a short explanation about the goal of the experiment, the requirements, how the collected data will be used and what will be expected of them. They can also test their audio set-up during this introduction to ensure that before they advance everything is working properly to partake.

The introduction is the entry point of the chart in Section 5.2.4.

Background Questions

Before listening to music the participants are asked some background questions. We mostly use these to get a better idea about who is participating. The following questions are asked:

- What is your age?
  - **Options:** Select one: 18-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 65+
  - **Reason:** This gives us some information about our demographics. We chose brackets since this makes it harder to identify a participant than when their specific age or birth date is asked.
- How often do you usually listen to music?
  - **Options:** Select one: less than once a month, once a month, once a week, once a day, more than once a day.
  - **Reason:** Gives us an idea of the listening habits of our participants.
- Do you actively perform music?
  - **Options:** Select one: yes, no
  - **Reason:** Gives further info on the musical background of our participants.
  - The following additional questions appear if yes is selected:
    - ◇ How often do you usually perform music?
      - · **Options:** Select one: less than once a month, once a month, once a week, once a day, more than once a day.
      - · **Reason:** More info on the habits of our participants.
    - ◇ How do you perform music?
      - · **Options:** Select multiple: Vocals, Woodwinds (flute, recorder, clarinet, ...), Brass instruments (trumpet, trombone, ...), Bowed string instruments (violin, ...), Guitar, Keyboard/Piano, Percussion instruments (xylophone, tambourine, ...). Additionally participants can fill in any instrument in an "other options"
      - · **Reason:** Get an idea of which instruments the participant plays (if they play any).

As shown in Section 5.2.4 after the introduction from the previous section the participant first logs in via Spotify before being shown these background questions.

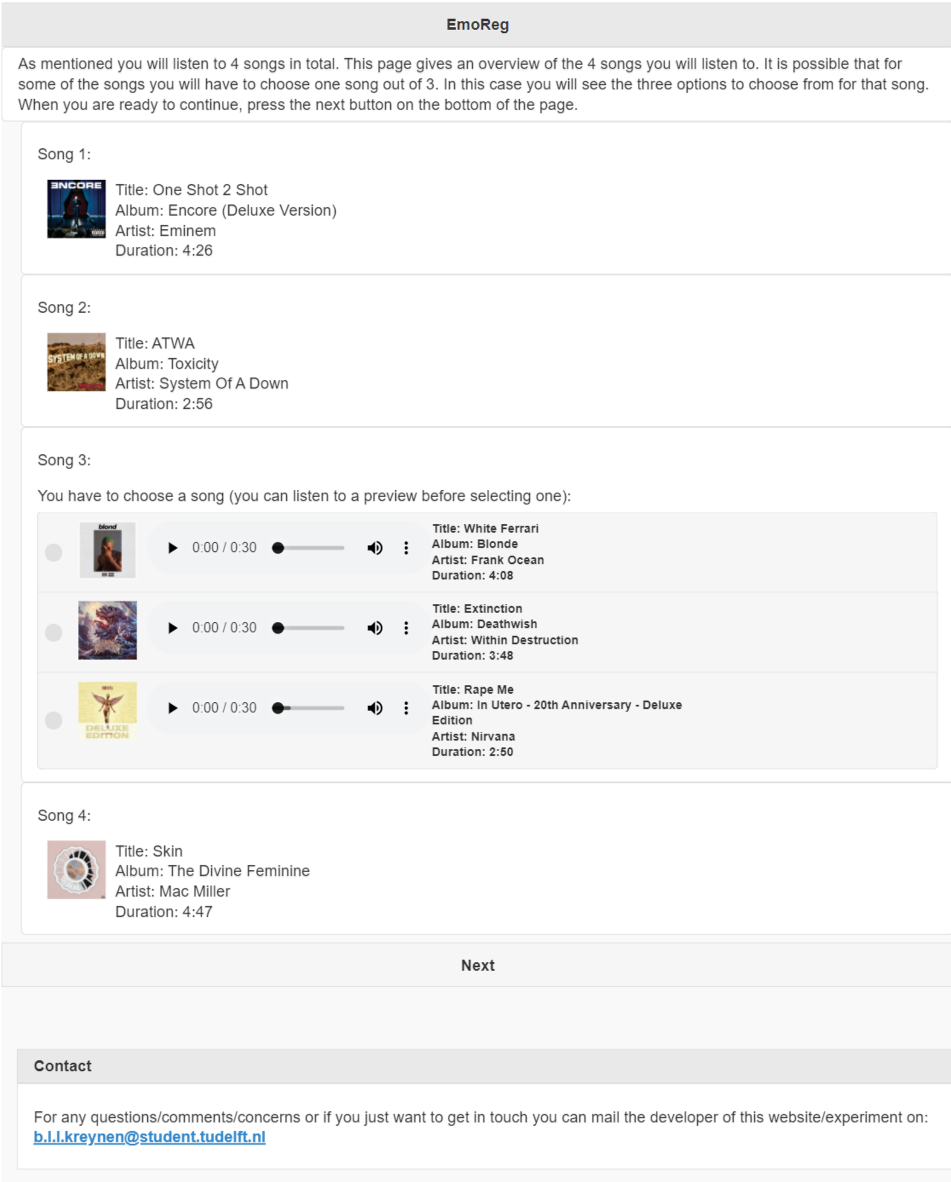


Figure 5.2: Presenting the playlist and their song choices to the participant

Listening to the Playlist
After the participants have answered the background questions they are almost ready to start listening to the playlist. First they are asked their emotional state via the AffectButton [13]. This is needed for our playlist generation and is explained in Section 5.2.5. After this participants are presented with a generated playlist of four songs. It is possible that for some of these four songs they have to make a choice between three songs before continuing. The reason for this choice is explained in Section 5.2.5 as well. An example of this screen with a choice for the third song is shown in Figure 5.2. This aspect of choosing songs is listed as "Playlist Choice" in Section 5.2.4. After submitting the choice the participant starts listening to the first song in the playlist. During the playback of a song the participant will be asked four questions. The details of these questions will be explained later in this section. After these questions the participant is asked to give their emotional state via the AffectButton again. Finally we will ask the question "How would you rate your current emotional state compared to the emotional state before you listened to this song?" with the options: large decline, small decline, no improvement or decline, small improvement, large improvement). Upon answering this last question the participant continues to the next song in the playlist. When the last song in the playlist has been played we continue to our outro; the end questions. This is the "Listening" box in Section 5.2.4, the steps that happen while a participant is listening to a song are also detailed to the right of this box.

AffectButton The AffectButton [13] was chosen as a tool to ask participants about their emotional state. We chose this tool because it has been verified in previous research and works in the PAD-scale [13], we can use the pleasure and arousal dimensions as valence and arousal in our VA-scale and the third dimension could be interesting to analyze or for future research that might want to use the three dimensional scale instead. The first time the participant sees this button they are also given instructions on how to use it. This is shown in Figure 5.3. As a second method of getting information about the effect on the emotional state we ask the participants to self rate their emotional improvement after every song as well. The following question is asked:

- How would you rate your current emotional state compared to the emotional state before you listened to the song?
  - **Options:** 5 point likert: large decline, small decline, no improvement or decline, small improvement, large improvement
  - **Reason:** This is added to have a second way to check the influence of the playlist on the emotion. This simple self-reported improvement can be compared to the changes in the VA-scale.

Questions during playback During the playback of a song we ask four questions. The first question appears after the participant has listened to the song for 10 seconds. Each question then appears for a maximum of 30 seconds and the interval between the start of two subsequent questions is 35 seconds (i.e. if the previous question was open for the maximum of 30 seconds a new question will appear after 5 seconds). There are multiple reasons for posing these questions:

- It keeps the attention of the participant on the site, reducing the variation in factors outside of the experiment influencing the emotions of a participant.
- The questions are timed. If a participant does not answer it in time it tells us they are not paying attention to the page of the experiment.
- Answers from participants give us more information on the song and on how the participant experiences the song.

The order of the questions is always the same. These are the questions, in the correct order:

- How would you rate the audio quality?
  - **Options:** 5 point Likert scale (1: Very poor, 5: Very good)
  - **Reason:** Used to to check if people have issues with audio playback during particular moments
- How familiar is this song?

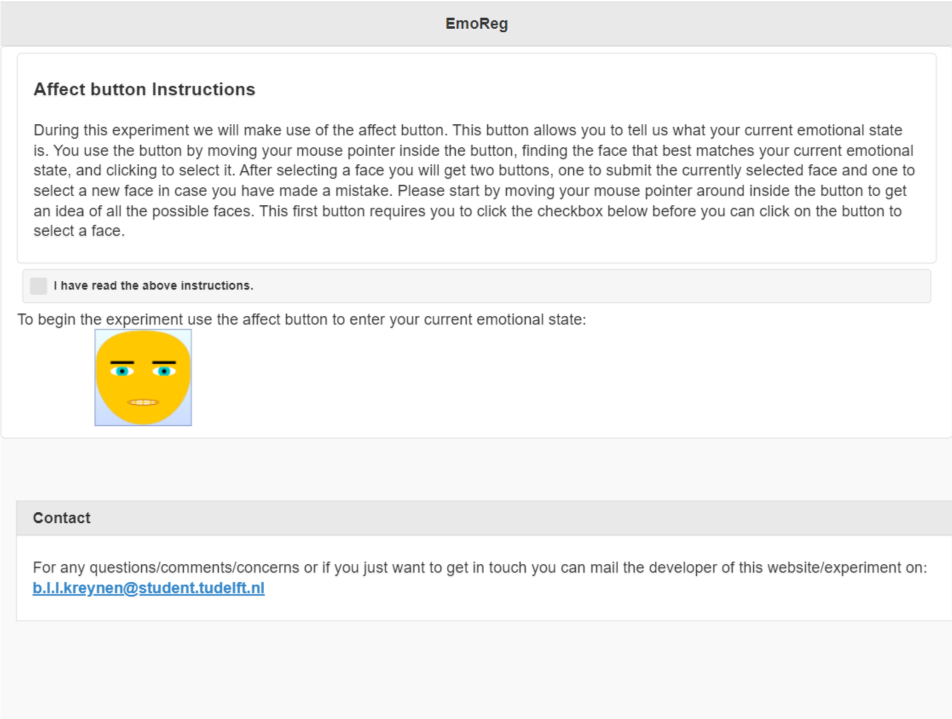


Figure 5.3: The affect button introduction.

  - **Options:** 5 point Likert scale (1: Not at all familiar, 5: Very familiar)
  - **Reason:** Gives information about the relationship between a participant and the song.
- How would you rate your preference for this song?
  - **Options:** 5 point Likert scale (1: Dislike a lot, 5: Like a lot)
  - **Reason:** Allows us to get an idea of whether the songs we played were indeed songs the participant likes.
- Do you recall a memory when listening to this song?
  - **Options:** Select one: Yes, a positive memory; Yes, a neutral memory; Yes, a negative memory; No; Yes, other
  - **Reason:** Allows us to check whether memories associated with a song have a noticeable difference in their effect on the affective state.

The questions appear in the box below the album art in Figure 5.4.

Song duration To allow enough time for all questions we filter all Spotify songs to be at least 2 minutes and 40 seconds. To prevent a participant from spending too much time on one song we also filter out songs that take longer than 8 minutes and 30 seconds. The questions mentioned in the previous sections end after approximately 2 minutes and 30 seconds. At 2 minutes and 40 seconds we present the participant with the AffectButton and immediately after with the self reported emotion improvement question. This allows the participant to continue to the next songs after around 2 minutes and 40 seconds on any song in case they think the song drags on too long.

Playlist duration The playlist consists of four songs. As mentioned in the previous section songs are at least 2 minutes and 40 seconds long, therefore the minimum duration of the songs played back to back is 10 minutes and 40 seconds. With a maximum song length of 8 minutes and 30 seconds in principle the longest playlist can be 34 minutes, a similar length to one intervention as described by Gerdner [36]. However since participants can proceed to the next song more quickly if they want we expect most participants to end up somewhere between 11 and 15 minutes of listening time.

Simple visualization During playback the participants also is shown a simple visualization of the music. This visualization shows a participant that the page is active. It also is another way of keeping the attention of a participant on the page. The visualization is updated every half beat of the song and is based on the spectrum analysis of the song. Data required for this is all provided by the Spotify API. This can be seen in Figure 5.4 together with the whole page a participant sees when listening to a song.

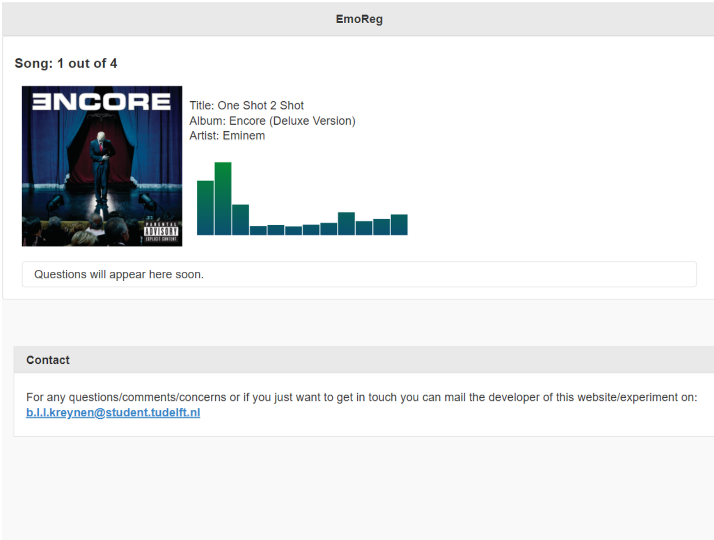


Figure 5.4: The music player before the participant receives a question.

End Questions
After listening to the playlist we ask the following questions:

- How would you rate your current emotional state compared to the emotional state before you listened to the playlist?
  - **Options:** 5 point Likert: large decline, small decline, no improvement or decline, small improvement, large improvement
  - **Reason:** Similar to the self rated improvement asked after every song. This simple self-reported improvement can be compared to the changes in the VA-scale.
- How would you rate the sequence of songs you have just listened to?
  - **Options:** 5 point Likert: Very poor, poor, neutral, good, very good
  - **Reason:** Feedback on the quality of the playlist.
- Is there anything else you would like to mention (about the experiment, sequence of songs or anything else)?
  - Open question, can enter any text
  - **Reason:** Get more feedback which can be used in a qualitative analysis.

Final Page
After the participant has submitted the end questions they are shown a final page thanking them for participation and mentioning that they can contact the author of this thesis in case of any concerns/questions/comments/.... This is the exit point of the chart in Section 5.2.4

Duration of entire experiment
Not including reading the introduction, so after logging in with Spotify, most of the experiment will be spent on listening to the playlist. Which, as mentioned in the *Playlist duration* paragraph above, is expected to be between 11-15 minutes. Additionally participants also have to fill in the background questions and end questions. However since most information is gathered during playback we do not expect this to take a lot of time therefore we expect the duration of the entire experiment to be 15 to 20 minutes.

5.2.5. Playlist Generation

In this section we describe the playlist generation algorithm for both the control group and the experimental group, including the design rational behind it. We start by explaining the experimental algorithm. The method for generating the experimental playlist is also summarized in Figure 5.5. The "current state" is grid corresponding to the current state of the MDP formulation. When starting the playlist generation this is the grid that corresponds to the first values that the participant entered via the AffectButton.

Experimental
The experimental group uses a 7x7 GridWorld MDP formulation, as explained in Chapter 4, to generate the playlist. In this first experiment we will only focus on changing the valence axis of the VA scale, hence the target states of the MDP are all grids with a valence of 0.75 or higher. The agent gets a reward of 1 for playing a saved song and a reward of 5 for reaching the target, going to state for which we know no song that the participant has saved induces a punishment of -1. We use value iteration and a greedy policy with a discount of 0.8. If it takes more than four songs to reach the target state from the start state then we cut off the playlist after four songs. If we reach the target state in less than four songs then we continue playing songs from the target states, moving a maximum of one grid at a time and giving priority to states for which we know a saved song for the participant. Finally if the path goes over a state without a saved song we will give the participant a choice between 3 songs. This is shown in Figure 5.5 after the decision "current state has saved song?". The "no" path shows how these choices are generated which is in the following way:

- If the playlist already contains songs:
  - We ask Spotify for three recommendations with the correct valence/arousal values based on a seed consisting of the songs that are already in the playlist
    - If we get these then these are used. Otherwise we use songs from the "defaultlist"
- Else we use songs from the "defaultlist"

The "defaultlist" is constructed in the following way:

- First we added all songs from the top 2000 to the default list.
- Then we iterated over all VA grids which still had less than 3 songs and kept asking Spotify for recommendations for this grid based on a random seed of 5 songs from the top2000 until we had at least 3 songs per VA grid.

We chose the top 2000 because it is constructed to somewhat reflect the preferences of the dutch audience. With it containing 2000 songs it is also contains quite a lot of songs. Additionally the playlist can be found on Spotify. With many of our participants likely coming from the Netherlands it seemed like a good source for songs that had a bigger chance of being familiar to our participants. After testing we found that most grids in the VA-space were covered with at least 3 songs from the top2000.

Valence Target We focus on the valence dimension first. In the MDP formulation we use we can set a target in both the valence and arousal domain, but it is also possible to set a range of targets. We chose to have all grids with a valence above 0.75 as our target states. With this first experiment we want to start by focusing on this dimension to limit the scope a bit. Furthermore if we look at the model described in Section 2.1.1 then if we want to influence the more longer term mood instead of only the emotion then the valence in emotions has a direct influence on the valence of someones mood. While there is not such a strong influence between the arousal. We pick 0.75 because with 0.5 as neutral and 1 it is in the middle between the neutral value and the most positive value, making it a clearly positive valence value.

Grid size The number of grids is kept small for two reasons. To begin with we do not know how many songs a participant will have saved on average. With a lower grid size we increase the chance that a participant will have a saved song for a certain grid. With a 7x7 grid we have 49 grids. With a uniform distribution we hence have an expectancy of having filled each grid with at least one song with 49 saved songs. It seems reasonable that many people will have more than 49 songs saved on Spotify. However we have to keep in mind that the distribution of saved songs is more than likely not uniform. Therefore with such a gridsize we can assume that if a participant has 50 saved songs a large part of the VA space will most likely be covered, with a few gaps

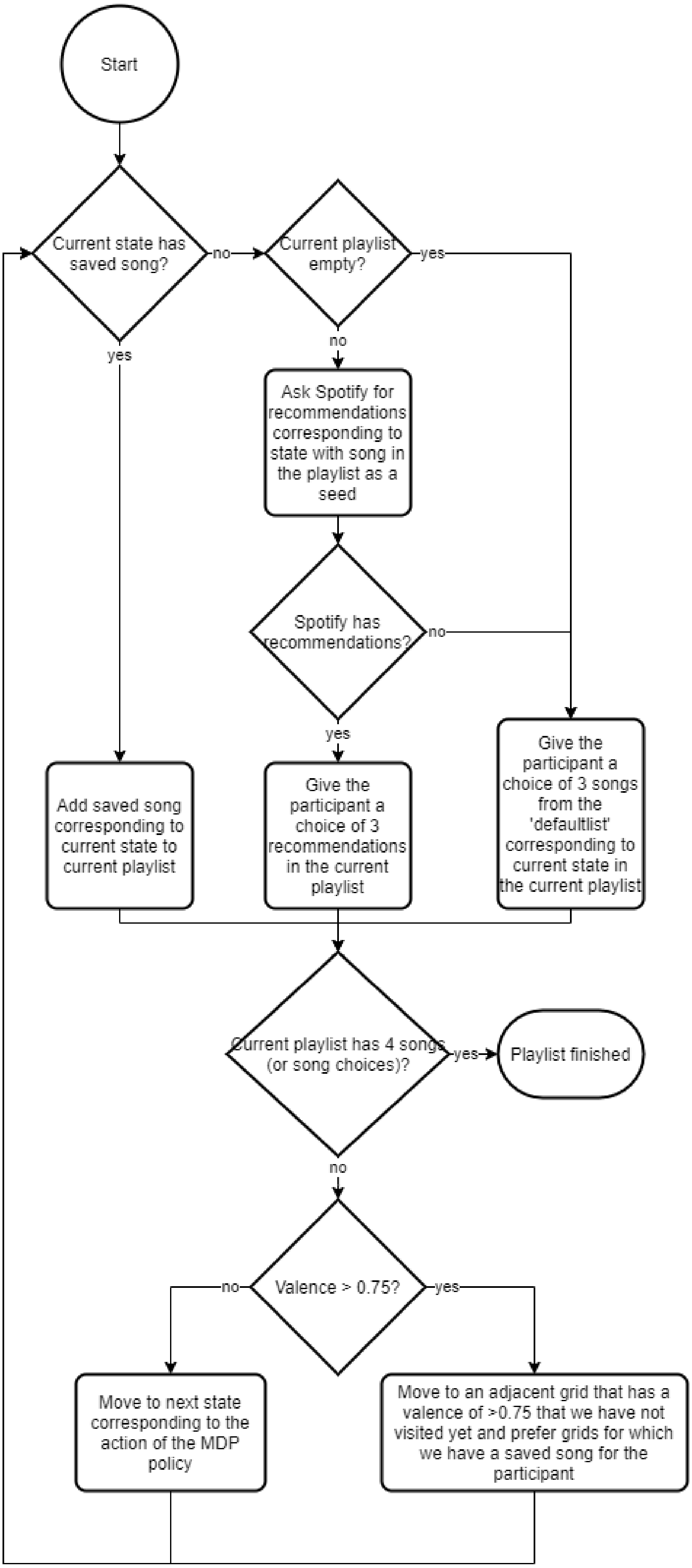


Figure 5.5: Flowchart showing how the playlist generation works for the experimental group.

here and there. 50 songs also seems like a reasonable number of preferred songs to get from family or friends of a PwD when they for example know of a few albums that they listened to. Secondly the size of the steps we can take in the VA-space is determined by the size of the grid as well. We only have four songs in a playlist with a gridsize of 7, without detours we can reach from the most negative valence of 0 to at least a positive valance of over 0.5, since if we move 4 grids to the right we have traveled over half the width of the gridworld. With this we will hopefully also see that most of the participants in the experimental group have a positive valence after listening to the four songs during the experiment. Finally a smaller grid size also makes sure our value iteration solution is still fast enough.

Rewards & discount Determining which rewards we should give to a grid with a song, a grid without a song and a target grid is a difficult choice. We started by defining what we find important in our solutions:

- We want to reach our target on average in 4 steps
- We want most of the selected songs to be saved songs
- We want a policy to reach the target state, and not get stuck. This can happen by creating an infinite loop between two grids for example.

To determine our rewards and discount factor with these in mind we made simulations based on the following assumptions:

- The saved songs of a participant are uniformly distributed in the VA space.
  - For each participant this distribution is generated once and saved to be used for future tests.
  - We assume a participant has 50 saved songs.
- For each participant we test a starting position of both 0.25, 0.25 and 0.25, 0.75 in the VA space.
  - The reason behind it is that in the final application to PwD we would assume that when the target emotional state is a positive valence it will be used when the valence of a listener is negative.

We tested various combinations of rewards and discounts on 1000 simulated participants. With the rewards described earlier and a discount factor of 0.8 for all of the participants we eventually reached the target state, no loops were encountered. On average the target states were reached in 4 steps. And 96.8 percent of the grids contained a saved song. In conclusion with these settings we expect everyone to reach the target state in a reasonable time while mostly encountering saved songs, hence fulfilling our requirements for the algorithm for the experiment. We do not in any way think that these choices are necessarily very well optimized or the best choice but it is a choice that is good enough for the purpose of this experiment. However since saved songs are most likely not uniformly distributed we expect a worse result on the percentage of saved songs in our actual experiment. To test this assumption we reran our simulations but we changed the assumption that saved songs are uniformly distributed to that they are distributed in the VA space according to a truncated normal distribution with a mean of 0.25 and a standard deviation of 0.25. This distribution is used for both the valence and arousal dimension. If we change to these assumptions then still all the participants reach the target state and in an average of four steps. However the number of grids with a saved song goes down to 83.8 percent. As expected we more often encounter a grid without saved songs, however the majority of used grids still had a saved song.

After reaching target In our experiment, since we always want to have a playlist of size four, we have to decide what we do if we reach a target state. We cannot necessarily keep following the MDP policy because often it moves between two states after reaching the target, and we want to prevent picking the same song multiple times in this short playlist. Therefore we devised a simple solution when we reach the target state based on the goal of having a valence higher tan 0.75 and preferring grids with a saved songs:

- We first select the adjacent target grids (grid with a valence above 0.75) that have a saved song which we have not yet visited
  - If there is one or more then we pick one at random and move to this grid.

- If there are no adjacent target grids with a saved song we pick the adjacent target grids which have not yet been visited.
  - We pick one at random and move to this target grid.
- We repeat the above until the playlist is of length four.

Picking a song associated with a square when we move to it is done in the same way as for the MDP formulation. This simple algorithm keeps with the main idea of the grid world agent moving one grid at a time. At the same time it also prefers grids with a saved song over those without one. This is shown in Figure 5.5 in the decision "Valence > 0.75?". If the answer is no we keep following the MDP formulation, if it is yes we follow the above method to move to the next state.

Control

For the control group we simply select four random songs from the participants saved songs. To replicate the choice aspect of the experimental playlist one of those four songs will be replaced by a choice between 3 songs that we get through Spotify recommendations with the other songs in the playlist as a seed. With this control group we want to come close to replicating the traditional IMI where for example an iPod is put on shuffle to randomly select from a list of known preferred songs.

Now that we have explained the method of the experiment and the rational behind the important design decisions in detail we will move on to the implementation of the experiment.

5.3. Implementation

In this section we give an overview of how the experiment has been implemented. As mentioned the experiment is conducted completely online. We built a website to allow people to access the experiment. In this chapter we first describe how we implemented the backend of the website. We then movfe on to the frontend. Finally we talk about how this experiment was brought online. For some parts of the implementation we will show pictures of what the frontend looks like. A complete walkthrough of the frontend can be found in Appendix G.

5.3.1. The backend

The backend was implemented in Django. Django is a web framework for Python. According to its site it is high level and "encourages rapid development" [28]. This is exactly what we needed to quickly develop a website to host our experiment on. Additionally Django has widespread usage and with that also come a lot of packages that work with it. This includes the Python Social Auth package which allows us to easily implement a login through Spotify.

Spotify

As mentioned in the experiment you login through Spotify. This in turn allows us to access the saved songs of a participant. It also allows us to play full songs on the website if they have a premium account. Additionally we can use it to track if the same participant participates more than once. The Spotify API can also be used to access info about track, including the title, artist, album, album cover image, and valence and energy values.

MDP

The prototype we described in Chapter 4 based on value iteration was written in Python. Since our backend was also written in Python it was easy to implement a slightly adjusted version of this prototype in our backend.

5.3.2. The frontend

The front-end is a combination of HTML served through the Django backend and Javascript. Javascript is mostly used to create the music player and serve the relevant questions during playback of songs. Some of the pages as viewed by a participant have already been shown in the Design section. Interested readers can find a full walkthrough of the frontend in Appendix G.

5.3.3. Accessing the experiment

A server running the Django backend was brought online. The domain name www.emoreg.nl was reserved and redirected to the server. Therefore the experiment was publicly accessible to anyone by visiting the link. The first page shown to someone when visiting the link is the introduction page. After the experiment is over www.emoreg.nl will be updated to include a short summary of the results, a link to the produced dataset and a link to this report.

5.4. Conducting the experiment

After implementing the experiment and bringing the website online we started sharing the experiment with potential participants. The experiment was first shared in February 2021 in a Facebook post on the Facebook and Linkedin page of the author of this thesis. Since this was the first time letting people access the experiment we wanted to limit it to a first small pool to check if everything worked well. During this initial phase we found two issues with the website. Linking from social media sites did not properly redirect to the https version of the site, this lead to it being flagged by some antivirus systems. This was quickly remedied but might have caused some participants to not continue to the site when it was first shared. Secondly due to a bug in the selection form for some participants nothing happened when clicking next in the playlist screen (Figure 5.2). From our logs we confirmed that four participants had this issue.

Having ironed out the last issues we started spreading the experiment further. It was shared with friends and family of the researcher but also amongst computer science students and faculty staff from the TU Delft. Around the same time it was also shared with the member of the "Nederlandse Verenining voor Muziektherapie" (Dutch Association for Music Therapy). We then spread it to Dutch educational institutions who have music therapy related studies. We had many positive responses from these institutions and the experiment was shared with students from different Dutch Universities of Applied Sciences (HAN University, Leiden, NHL Stenden, Zuyd, Utrecht) and two Universities of Arts (ArtEZ, Codart Rotterdam).

Finally we also shared the experiment on Reddit, on a forum dedicated to music therapy.

The experiment was open for almost four months (until the end of May 2021). In this time span in total 34 participants logged in with their Spotify account on the website.

Now that we have introduced the experiment, its hypotheses, method and implementation we can move on to the results of our experiment. These are presented in the next chapter.

6. Experiment: Results

In this chapter we present the results of the experiment we carried out to answer our second main research question. The experiment was implemented as described in the previous chapter. We start by taking a closer look at our participants and their background information. We then move on to the emotional states as given via the AffectButton. We also report the answers given to the end questions. Finally we report on the answers given to the questions that are asked while the songs are playing.

6.1. Number of participants

In this section we want to give an overview of the number of participants we have and how far they got in the experiment. We start by looking at the total sign ins and then continue through each step until the final step of the end questions. This is summarized in Figure 6.1. In this chart "begin playlist" means submitting the choices for the playlist (and thus progressing to the first song). As can be seen in the Figure in total 34 people signed in and 22 people finished the experiment. Out of the participants who finished the experiment 12 belonged to the control group and 10 to the experimental group.

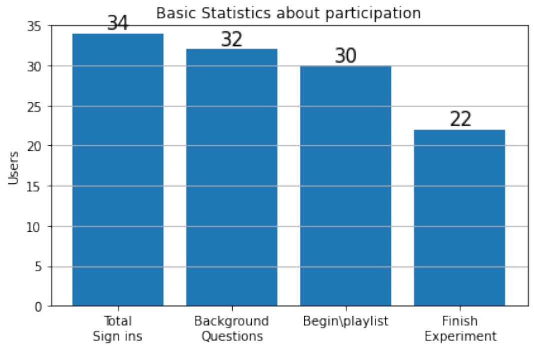


Figure 6.1: Bar chart showing how far participants came in the experiment.

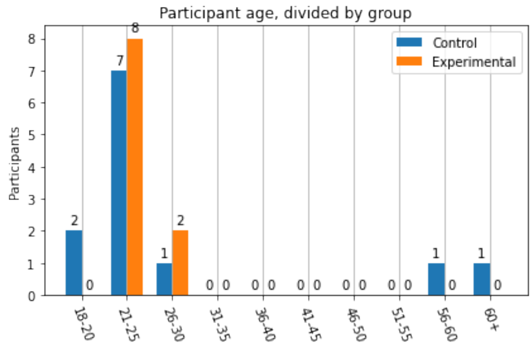


Figure 6.2: Bar chart showing Age of participants, divided by group.

6.2. Background Questions

In this sections we give an overview to the answers given to the background questions. We only show the participants that finished an experiment since we also want to show the distinction between the control and experimental groups. Furthermore it is also these participants that we are more interested in for our future analyses since we have complete data on them.

6.2.1. Age

We show the age bracket participants selected and whether these participants belonged to the control or experimental group in Figure 6.2. It shows that the majority of our participants are in the 21-25 age bracket. Only two participants were above 30, one in the 56-60 bracket and the other in the 60+ bracket. This seems to correspond to age range we would expect from students participants, amongst who we indeed heavily spread the experiment.

6.2.2. Listening frequency

The same has been done for the question of how often they listen to music. This is shown in Figure 6.3. In this Figure we see that a majority of people selected the more than once a day option and no one went for any of the options that are less than once a day.This seems to suggest that our participants are quite actively involved with music.

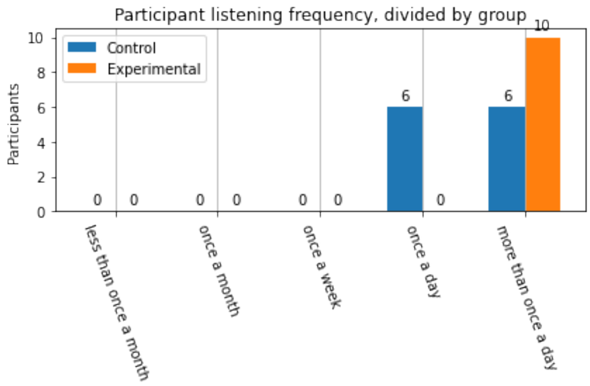


Figure 6.3: Bar chart showing listening frequency of participants, divided by group.

6.2.3. Performs Music

Finally we had the questions relating to performing music. Whether participants perform music is shown in Figure 6.4, we notice that the division between people who do and do not play music is almost equal. The fact that almost half of our participants perform music again seems to suggest that they are quite actively involved with music. How frequently those that do perform music perform is shown in Figure 6.5 and which instruments they perform is shown in Figure 6.6. Vocals, guitar and keyboard seem more popular than others. These three also seem about equally popular compared to each other.

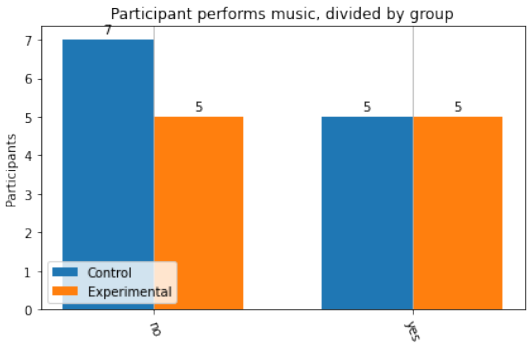


Figure 6.4: Bar chart showing whether participants perform music, divided by group.

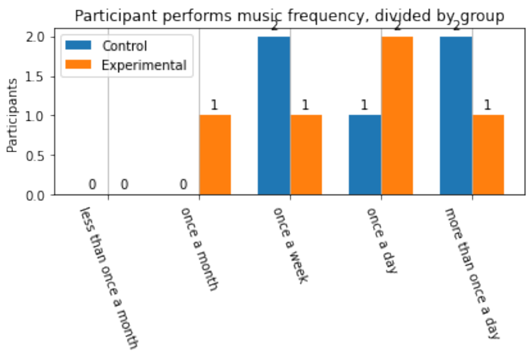


Figure 6.5: Bar chart showing performance frequency of participants, divided by group.

6.3. Experiment Duration

We measured the duration of the experiment from the moment the participant entered the first emotional state via the AffectButton to the submission of the end questions. The shortest experiment was 739 seconds long and the longest 1934 seconds. The mean is 1022.8 seconds with a standard deviation of 293.1. This mean is nicely within the 15 to 20 minutes we estimated the experiments to be.

6.4. Saved Songs

On average we were able to obtain 724.8 songs per participant, with a standard deviation of 661.5. The participant with the least amount of saved songs had 23 saved songs and the participant with the most had 2364

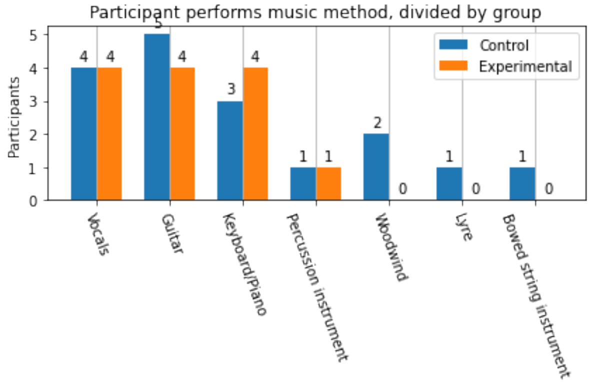
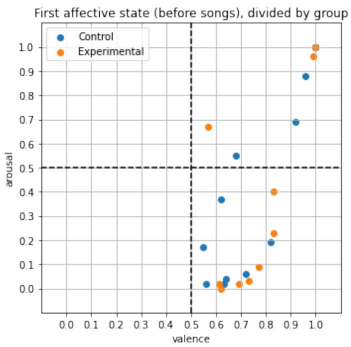


Figure 6.6: Bar chart which instrument participants perform, divided by group.

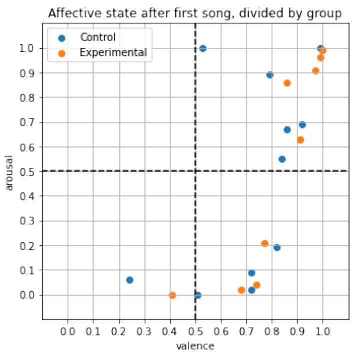
saved songs.

6.5. Emotional States

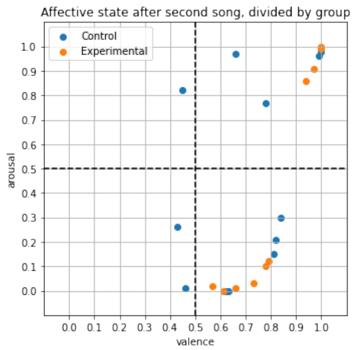
During the experiment the emotional states of the participants were asked via the AffectButton. We can plot this in the VA space. For each time we asked the participants their emotional state this has been plotted in a scatter plot in Figure 6.7. participants in the experimental and control group are plotted in separate colours. On initial observation it seems that the begin state is much more towards a low arousal and somewhat positive valence, while at the end participants are more spread out on the arousal dimension. The experimental participants also seem to move more toward a positive valence in the last state.



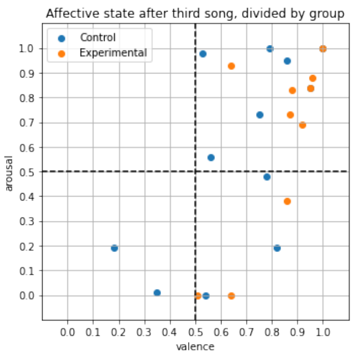
(a) At the beginning, before playlist



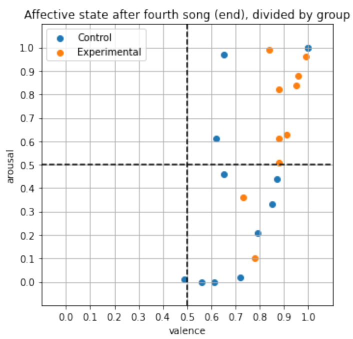
(b) After 1st song



(c) After 2nd song



(d) After 3rd song



(e) After 4th song

Figure 6.7: Emotional states of participants throughout experiment

We can look closer at the valence by only plotting this dimensions. To compare the experimental to the control group we made a boxplot of the valence at each timestep, see Figure 6.8. In this boxplot we can see that the mean of the final valence of the experimental group seems larger than that of the control group. The deviation in valence in the control group is also larger for each timestep, except before listening to the experiment where they are almost equal. This seems to suggest that the valence before listening to the playlist of both groups were similar but that they started diverging when listening to the playlist. Similar graphs

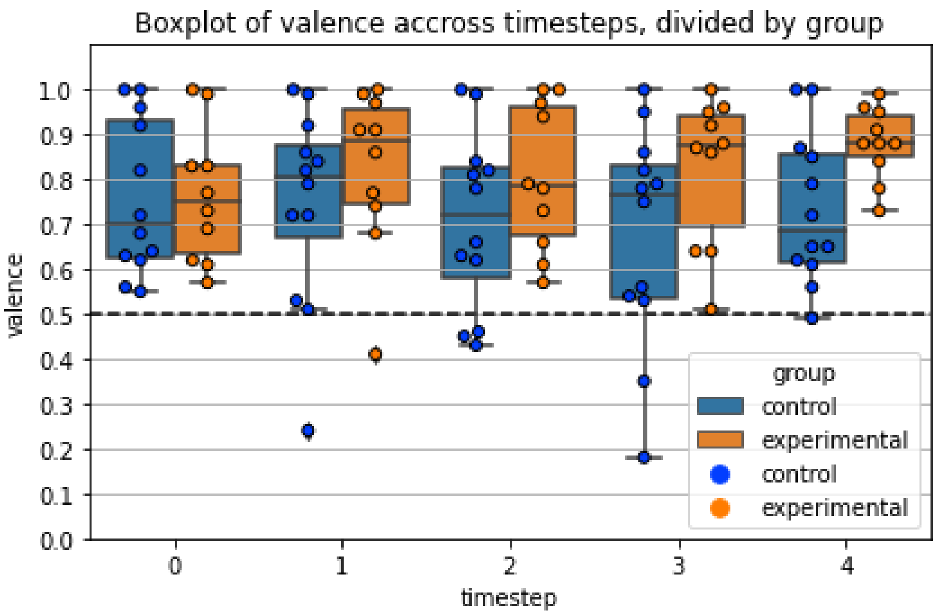


Figure 6.8: Box plot of the valence at each timestep, divided by group. Timestep 0 is before listening to the playlist, 4 is after listening to the final song. The dotted line shows a neutral valence and the line in a box is the median. The boundaries of the box are the medians of the upper half (Q3) and the lower half (Q1) of the data. The whiskers show the minimum and maxium values without outliers. Outliers are calculated at 1.5IQR ($IQR = Q3 - Q1$). The dots are the actual values for each participant.

are available for the arousal and dominance, however since this is was not our focus these are included in Appendix F for interested readers.

To see the change in a participant from listening to the playlist the delta in valence is also interesting to look at. Figure 6.9 shows the delta between the last and first valence. In the plot a histogram with 20 bins is drawn, a kernel density estimation is drawn on top on top. In the valence dimension the control group has a delta with a mean around 0. In the experimental group we see this mean shift towards the positive side slightly, but the difference is not very pronounced.

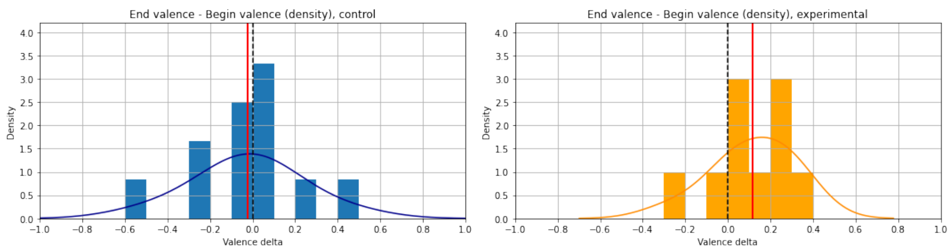


Figure 6.9: Kernel density estimation (Gaussian kernel), bandwidth=0.8 of begin and final valence, divided by group. The KDE is plotted on top of a histogram with 20 bins with the same data. The sample mean is shown in red. It shows the valence in delta over the entire experiment (before listening to the playlist until after the last song). The dots are the actual values for each participant.

Additionally we also plot the delta with the previous answer for each timestep in Figure 6.10. The deltas are shown in a boxplot. Timestep 1 means substracting the valence after listening to the first song from the valence before listening to the playlist, timestep 2 means substracting the valence after the first song from the valence after the second song and so on. Similarly to Figure 6.9 we notice that in general the delta seems slightly more positive for the control group. The deviation also seems larger in the control group for most timesteps. However just as in Figure 6.9 this difference is not very pronounced. This small difference is in contrast with the larger difference in Figure 6.8 where the final valence of the experimental group seems to be higher. However we note that the begin valence of participants in general is also positive, and some are even very positive. Some participants even started with a maximum valence of 1. With a smaller negative delta these participants still end up with a relatively high valence.

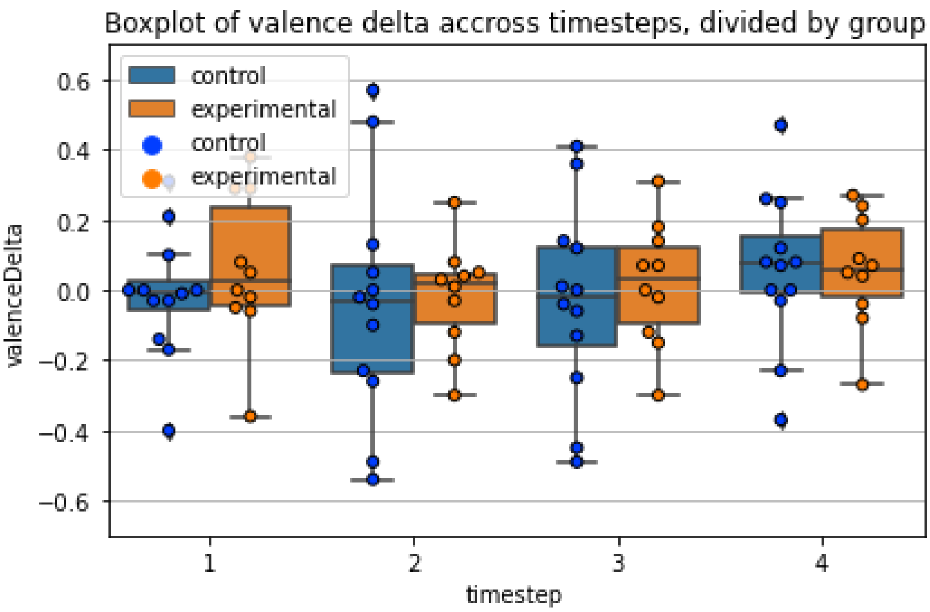


Figure 6.10: Boxplot of the delta in valence. The line in a box is the median.The boundaries of the box are the medians of the upper half (Q3) and the lower half (Q1) of the data. The whiskers show the minimum and maxium values without outliers. Outliers are calculated at 1.5 Inter Quartile Range (IQR) ($IQR = Q3 - Q1$).

To study the differences between the control more in depth we perform a Welch's t-test in SPSS. The results of this can be seen in Figure 6.11. The results of this correspond pretty well with what we saw in Figure 6.9 and Figure 6.8. Namely that there is a significant difference between the final valence (valence4) of the control (M=0.73, SD=0.17) and the experimental (M=0.88, SD=0.08) group, t(16.4),p=0.017. But as we already noted in the graph the difference in delta is not pronounced and there is no significant difference there. We also do not yet notice a significant difference in any of the states before the final state. Interested readers can find t-tests performed on the other dimensions of the AffectButton answer in Appendix F. The full descriptive statistics (mean, standard deviation, standard error, confidence intervals for the mean and minimum and maximum values) from SPSS of all dimensions are also given in Appendix F. The correlations between the different emotional dimensions are also given there.

Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
valence0	Welch	.007	1	19.938	.935
valence1	Welch	.840	1	19.999	.370
valence2	Welch	1.594	1	20.000	.221
valence3	Welch	2.769	1	19.245	.112
valence4	Welch	7.112	1	16.402	.017
valencedelta	Welch	2.367	1	19.811	.140

a. Asymptotically F distributed.

Figure 6.11: Output of Welch's t-test in SPSS 26. Performed for the valence at each timestep. valenceDelta is the delta between the final and begin valence.

6.6. End Questions

6.6.1. Self reported emotion change

In Figure 6.12 we show a bar chart of the answer to the self reported change in emotion. In line with our results of the delta in valence we do not see a pronounced difference between both. It is interesting to note that only one participant rated a decline (a small decline) and that this participant is from the control group. We performed a Welch's t-test on the answer to this question but found no significant difference t(19.9), p=0.85 between the answer of the control group (M=2.83, SD=0.84) and the experimental group (M=2.90, SD=0.74)

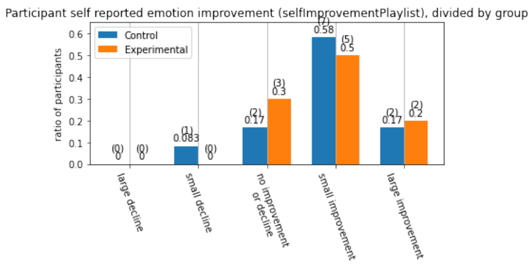


Figure 6.12: Bar chart showing results of the self reported emotion change question, divided by group and normalized on group size. In brackets we show the number of participants that gave the answer.

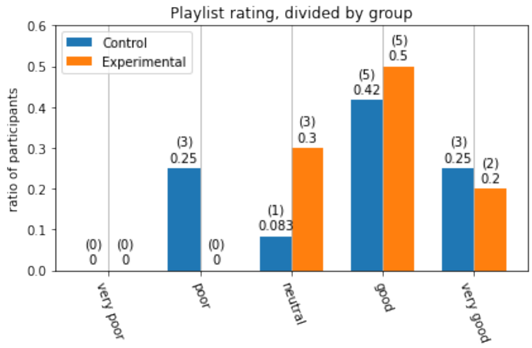


Figure 6.13: Bar chart showing results of the playlist change, divided by group and normalized on group size. In brackets we show the number of participants that gave the answer.

6.6.2. Playlist Rating

A bar chart of the playlist ratings is shown in Figure 6.13. We note that only the control group had a few participants rate the playlist as poor, while more participants rated it neutral in the experimental group than the control. We also performed a Welch t-test on this answer but also here we did not find a significant difference t(18.9),p=0.57 between the answer of the control group (M=2.67, SD=1.15) and the experimental group (M=2.90,SD=0.74)

6.6.3. Open Question

Twelve Participants left an answer on the open question. One mentions the audio quality question being asked too quickly. One participant from the experimental group mentions that the sequence of songs feels random. One experimental participant reports feeling calmer. A participant in the experimental group reports a lot of Dutch music while they do not listen to this often themselves. A participant from the control group compliments the taste in music. A participant from the experimental group notes that their mood was very good at the start so that a small decline was expected. This participant started with a VA value of (1,1) and ended with (0.99, 0.96). On their self reported change they noted no improvement or decline. In general they felt that their mood improved when a song had a "driving rhythmic nature". A participant from the control group noted that they liked rediscovering music they hadn't listened to in a while in the experiment. One participant notes that the AffectButton makes sense for a questionnaire but "comical and unnatural" from a user's perspective. A control group participant reports having not heard two of the songs before and that the choice of song should be personal. A control group participant notes that there is a mix of different genres and that they did not recognize any of the songs. An experimental participant reports having a lot of Spanish songs as options.

6.7. Songs

This section will deal with data on a song basis (e.g. answers to questions during playback).

6.7.1. Answered Questions

Looking at the completed experiments we only observe one instance where a question asked during playback was not answered within the given 30 seconds. In total 353 questions were asked, 352 received a response.

With 22 participants listening to four songs having four questions each we would expect a total of 352 questions. We therefore can immediately write off the single unanswered question as a technical issue which got solved and to which we received an answer to nonetheless later.

6.7.2. Audio Quality

Figure 6.14 shows a bar plot of the ratio of songs that received a certain answer on the audio quality question, divided between the control group and experimental group. On a 5 point Likert scale 3 is neutral so only 3 song in total received a negative rating for the audio quality. This seems to indicate that there were no major issues with playback via Spotify.

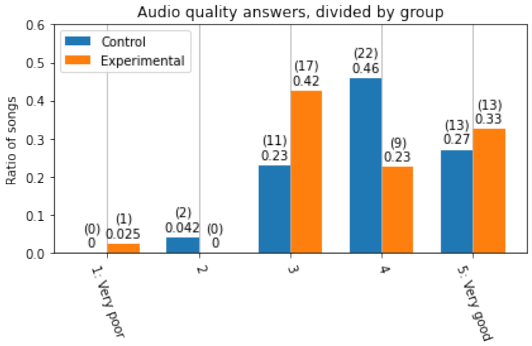


Figure 6.14: Bar chart showing ratio of songs which received a certain audio quality rating. In brackets we show the number of songs this was.

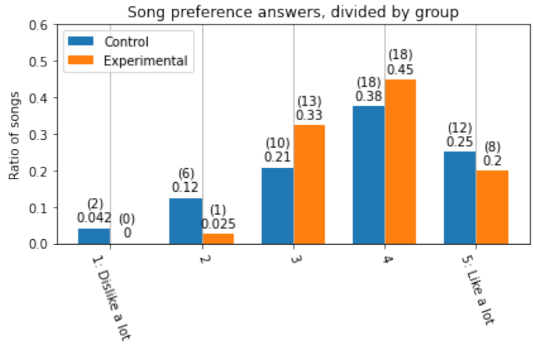


Figure 6.15: Bar chart showing ratio of songs which received a certain preference rating. In brackets we show the number of songs this was.

6.7.3. Preference

The preference rating of participants for songs is shown in Figure 6.15. Also here most songs receive at least a neutral rating. A majority of all songs receive a positive rating of 4 or more, in both groups. We notice that most of the negative ratings occur in the control group. Only one song was rated negatively by a participant in the experimental group. Only 2 songs that received negative ratings were not from a participants saved songs. This shows that playing a saved song is not a guarantee for getting a song a participant actually likes.

Familiarity

Figure 6.16 shows a bar chart of the answers songs received to the familiarity questions. The control and experimental group seemed to have given very similar ratings for this question.

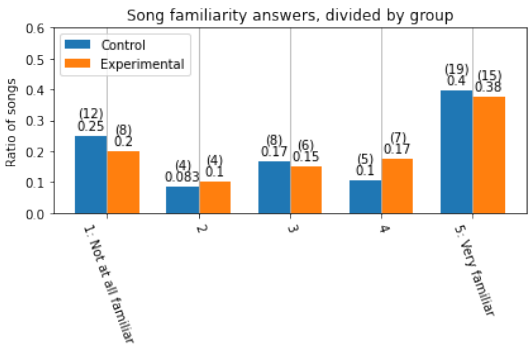


Figure 6.16: Bar chart showing ratio of songs which received a certain familiarity rating. In brackets we show the number of songs this was.

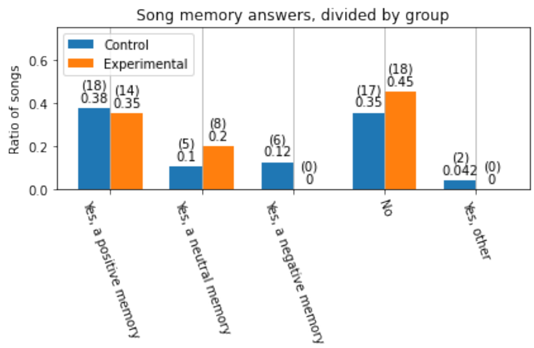


Figure 6.17: Bar chart showing ratio of songs which received a certain memory answer. In brackets we show the number of songs this was.

6.7.4. Memory

Figure 6.17 Shows the answer to the memory question for each songs. It shows that there is an almost equal divide between a memory (of any type) or no memory, with somewhat more songs having some memory attached to it. Interesting to note is that most reported memories are positive, with only six songs receiving the answer of a negative memory and 32 songs of a positive memory. Additionally in the experimental group no songs were tagged with a negative memory. We also performed t-tests, dividing the songs into songs with or without a memory and looking at the delta in emotional state (valence, arousal, dominance) when a participant listened to it. We did not find significant differences here though. Interested readers can find the full results of these t-tests in Appendix F.

Since there is only a small number of negatively tagged songs compared to ones with positive memories it is difficult to perform more rigorous statistical analyses on them. However since we do think it is interesting to take a closer look we perform a qualitative analysis. First we zoom in on the songs with a negative memory. We note that two participants tagged two songs with a negative memory. These two participants therefore already make up for four out of the six negatively tagged songs. For four out of the six songs the Spotify valence for the song is also negative. None of the songs have a very high valence on Spotify (the highest having a value of 0.628). This seems to suggest there at least might me some link between the valence of a song and the memory attached. These songs did generally receive high preference ratings. Four out of six were given a rating of 4 (out of the 5 maximum), one was given a rating of 3 and one a rating of 2. Meaning only one of the participants rated their preference for a song lower than the neutral answer of 3. The delta in valence that participants gave via the AffectButton for all but two of these songs is negative. For one of the other the delta is 0 and thee last one is 0.12. While we only have a small sample size to base this on there does seem to be some correlation there. This is also reflected in the self reported emotional change that the participants answered via the Likert scale question. Here all songs receive a score lower than the neutral value of 3, except the aforementioned two songs. For the neutral memories less clear lines can be drawn. The Spotify valence of these songs is more all over the place (the lowest for example being 0.21 and the highest 0.97). The same goes for the delta in valence, varying from 0.27 to -0.37. Finally we also take a look at songs with a positive memory attached to them. Since this is a larger group we plot some of the values. We start by plotting the valence provided by Spotify for these songs in Figure 6.18 we note that there seems to be a bias towards positive songs. We also plot the delta of valence when a participant listens to this song as reported by

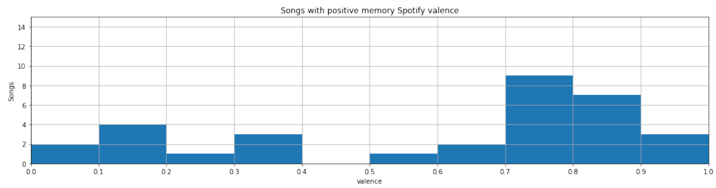


Figure 6.18: histogram with 20 bins of the valence value provided by Spotify for songs that were tagged as having a positive memory.

the AffectButton. This is plotted in a histogram in Figure 6.19. While not very pronounced we note that there seems to be a small trend towards a positive delta. This seems to be confirmed by the self reported change in

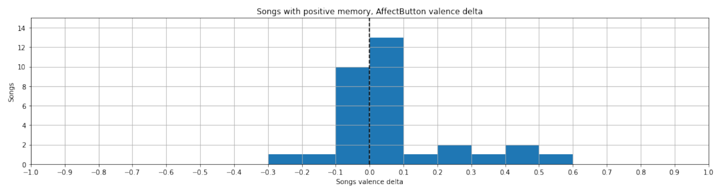


Figure 6.19: histogram with 10 bins of the delta in valence reported by participants via the AffectButton when listening to the song. Only for songs that were tagged as having a positive memory (control and experimental data are merged).

emotion by participants on the Likert scale. This is plotted for the positive memory songs in Figure 6.20. We notice the same trend towards a positive answer with most participants noting at least a small improvement. In conclusion it seems that there might be a relationship between the Spotify valence and the likelihood of a memory being positive or negative. As might be expected songs with a positive valence generally seem to have more positive memories. The effect on the emotional change also seems in line with whether a memory is positive or negative. We notice this effect slightly in the answers given via the AffectButton and in the answers given on the self reported emotional change.

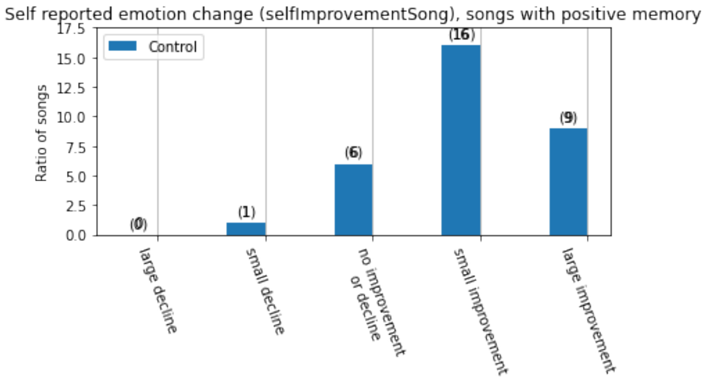


Figure 6.20: Bar chart showing the answer to the self reported emotion change. Only for songs that were tagged as having a positive memory (control and experimental data are merged).

6.7.5. Emotion change

Figure 6.21 shows the answer to the emotion change question per song divided per group. We do not notice a large difference between both groups except that there are a bit more songs in the control group that received a large decline as an answer. A large part of the songs in both groups received the no improvement or decline answer. To see how this self reported change relates to the dimensions gathered through the AffectButton

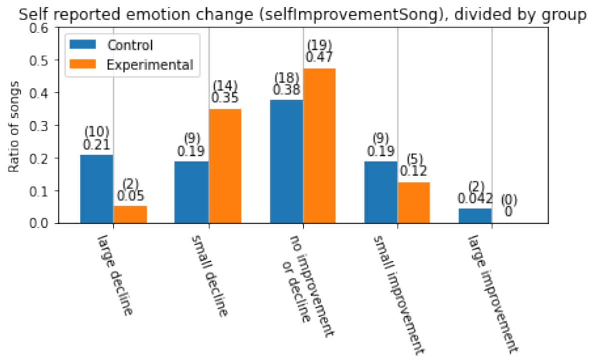


Figure 6.21: Bar chart on emotion change question for each song, divided by group.

we can look at the correlations. Full correlation matrices are given in Appendix F we note that for each of the dimensions (valence, arousal, dominance) there are significant weak to moderate correlations ($p < 0.01$) to the answer on this question. This shows that as expected there is a correlation between this self reported improvement and the answers given through the AffectButton. This could be an indication that participants used the AffectButton correctly.

6.7.6. Songs in Saved

We first take a look at how many songs were from a participants saved songs and how many were not for each groups. This is shown in Figure 6.22. We can see that for both groups the majority of songs came from a participants saved songs, as hoped. The control group had a higher ratio of saved songs. For the control group we expected to have around 75% saved songs since we offered a choice not from their saved songs for one in the four songs. The actual 79% for the control group comes close. For the experimental group 60% was in the participants saved songs. This indicates that for around 40% of the steps taken we encountered an empty square for which we did not have a saved song in the experimental group.

6.7.7. Spotify VA values

Figure 6.23 shows the VA values provided by Spotify for the songs played during the experiment, divided by group. There is a clear difference between the experimental group and control. The control group seems well spread out, which is to be expected since it selects songs randomly without regard to the VA space. The songs from the experimental group are clearly all positive in valence. With all participants also having at least

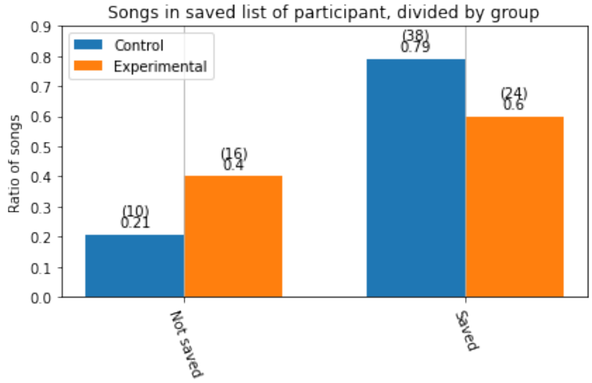


Figure 6.22: Bar chart showing ratio of saved vs not saved songs, divided by group.

a positive begin valence this is the outcome we expected on correct functioning of the algorithm since the algorithm should try to increase in valence to reach the target of 0.75

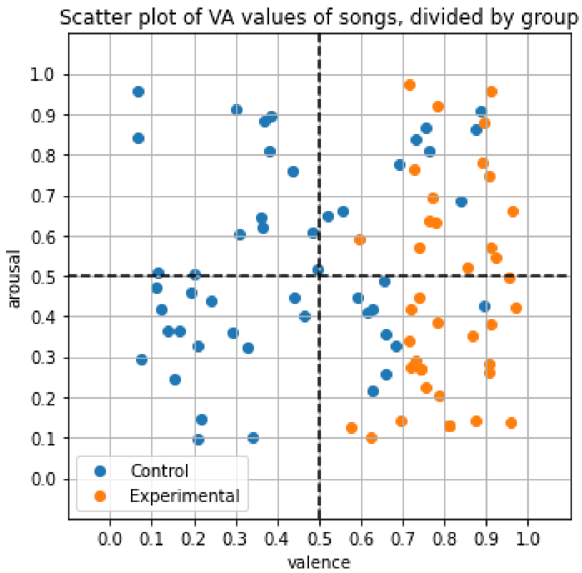


Figure 6.23: Scatter plot of the VA values provided by Spotify for each song, divided by group.

Having presented the results in this chapter we will move on to the discussion of our results in the next chapter.

7. Experiment: Discussion

7.1. Discussion

In this section we will discuss the results from the previous chapter. We discuss the three goals we had for this experiment one by one. Starting with the first goal of testing our MDP formulation, followed by the second goal of generating data and finally we discuss the third goals of checking the effect of memories with a song attached to them. After this we discuss some limiting factors of our experiment and recommendations for future work.

7.1.1. Goal 1: MDP

Recall that for this first goal we set up an hypothesis with two subhypotheses, from Section 5.2.2:

- ***H*1**: By using the deterministic MDP formulation the emotional state of a user after listening to a playlist is more positive than that of someone listening to a randomly shuffled playlist.
    - **H1.1:** The valence of the final emotional state of the participants in the experimental group is higher than that of the participants in the control group.
    - ***H*1.2**: The self-reported improvement in emotional state will be greater in the experimental group than in the control group.

For **H1.1** in Figure 6.11 we showed that the final valence of the experimental group (M=0.88, SD=0.08) is significantly higher than the final valence of the control group (M=0.73, SD=0.17), t(16.4),p=0.017. We can therefore reject the null hypothesis that was associated with **H1.1**. For **H1.2** we did not find a significant difference between the self reported change. However we note that there is an important difference between **H1.1** and **H1.2**, the first looks at the final emotional state and the second asks after the change in emotional state. Since many participants already had a high begin valence it seems logical that the difference in reported change might be less pronounced. However in general from the data in this experiment we can conclude that our MDP formulation did succeed in getting participants to a certain target valence after listening to a playlist of four songs. We therefore conclude that we succeeded in our first goal.

7.1.2. Goal 2: Dataset

Our second goal was to generate more data on which future research into music recommendation for emotion regulation/induction can build. A factor which limits us here is the amount of participants we had in the experiment. This means that the dataset is smaller than hoped. However with 22 participants all listening to four songs we do still have data on 88 songs in total. When analyzing the results we also found interesting observations such as the above described difference in valence between the control and experimental group but also observations relating to memories described in the next section. Expanding upon this first analysis could reveal more interesting observations. The data will still be made available but if we want to look into more complex techniques than our MDP formulation a first step will still be to gather more data.

7.1.3. Goal 3: Memories

The third goal was to check if songs with memories attached to them have a different effect on the affective state. We had the following hypotheses:

- ***H*2**: Songs that have a strong memory attached to them will have a larger effect on the emotional state of a listener.

– ***H*2.1**: The distance in emotional state before and after listening to a song is larger when there is a strong memory attached to the song.

We could not make strong claims relating directly to this hypothesis since we did not observe an immediate significant difference. A qualitative analysis into the memory question leads to some interesting insights. First it seems that there are very little negative memories associated with songs in general. This observation in itself is interesting and worth confirming in future research. At the same time the data also suggests that there might be a link between whether a memory is positive or negative and its influence on the change in valence, however making conclusions for this concrete is difficult due to the low number of songs that were tagged as having a negative memory. To conclude for this goal; while we could not make hard conclusions based on the data of this experiment we do make some interesting observations relating to memory. This is in line with earlier research by for example Schulkind et al. [78] who found a correlation between memories and 'emotionality' (induced emotions) when listening to music. We therefore expect that more in depth research would obtain more interesting results.

7.1.4. Considerations

In this section we describe a few aspects about the experiment and the results we have to consider when interpreting the results. This will also lead us to the recommendations for future work in the next section.

Participants

We have multiple factors related to our participants that we have to consider. First there is the limited sample size of 22 people. As mentioned above, this limits us in the goal of building a dataset for future research. For our own results this also means the results might not generalize as well as we hoped. This also leads us to our second consideration relating to the participants, in this small sample size our background questions also reveals little variation in age and musical background. Most of our users are quite young with the largest group being between 21-25. Quite a lot of the participants perform music and all of them listen to music at least once a day. This means the results might not translate well to people who are less actively involved with music or of a different age range. Especially for our original target audience of PwDs the young age is quite a difference in demographics. Additionally on average we have 724 (presumably) preferred songs for a user this is quite a bit more than we expect to have for many PwDs, our results might suffer when we have smaller lists of preferred songs.

Questions during music playback

Everyone answered all questions they were asked during music playback. This shows that our participants in general seemed to stay attentive to the website. For the smaller sample size we expected this is good since we reduce the influence of outside factors. However we have to keep in mind that in real situation these outside influences might have a great effect on the affective state which we did not encounter during this experiment.

Begin Emotional State

Something to keep in mind is that the mean of the start valence of all participants is already positive. With a target of a valence of higher than 0.75 this means a portion of our participants was already above the target before the experiment. Some participants even have a valence of 1 so even if their valence reduces by quite a bit it will still be relatively positive. It also means that participants who improved did not necessarily make very large improvements. This could explain the difference we are seeing between the significant result for the final valence and the not significant result for the delta in valence. At the same time we have to keep in mind that due to this our results might not be representative of a scenario in which someone with a very negative valence wants to invoke a positive valence, which could be a realistic scenario when considering the context of music therapy.

7.2. Future Work

In this final section of the discussion we go over some suggestions for future work based on our results and above discussion.

7.2.1. Gathering More Data

In the end only 22 participants completed the experiment. Even though the experiment has been spread around quite widely. It was shared amongst many students. One indication of how many people came across it comes from Linkedin which shows that the Linkedin posts have been viewed by 779 people. A limiting factor for some participants will of course be the Spotify premium account since not everyone will have one. To open up similar experiments to more users instead a free platform to play the music could be used such a Youtube. For this research we stuck with Spotify since they provided easy access to the valence and arousal values and access to a list of songs the users presumably likes. For future research music streaming services such as Spotify could also be contacted to see if there are possibilities for co-operation (a premium account could be given as a reward for participation for example, allowing the music to be used during the experiment).

For many people simply the time needed to take the experiment might also be a hurdle, to gather more data in the future systems which offer a reward such as mechanical turk could also be considered. This reward could influence the response however. Another option would be implementing a system using gamification to stimulate tagging songs, a similar idea was used successfully for the dataset Emotify [5]. Although gamification could once again affect the tags by affecting the induced emotions.

In the far future perhaps research into affective computing can help with generating a lot more data. Research in for example facial expression recognition [48] could offer avenues to gather data on people listening to music more passively. Other methods being researched include different physiological signals such as heartbeat, EMG, ... Section 2.5.1 These in the future could offer an unobtrusive way to measure the emotional state when people are listening to music in various natural contexts. Some sensors used for this are already being implemented more and more into consumer technology such as smartwatches. Studying how these wearables can be used to measure affect has become more popular [76]. We do have to keep in mind that these systems initially might not work for our target group of PwDs and that they might have to be adapted for them or that some might not even work well when applied to PwDs at all. As dementia advances facial expressions for example often become less clear which could pose a problem with a solution like facial expression recognition.

7.2.2. Algorithm

While in the far-future with more data available we could envision a system which implements a POMDP using information from an automatic emotion recognition system, for immediate follow up research we think it is interesting to first study other simpler methods. For example what if we just randomly sample music the user likes from a certain target VA range? If these also perform well these have much potential to be implemented already in current technology. These simpler algorithms have the benefit of not needing the current VA affect as an input and therefore could already be implemented with current technology without having to worry about how we can sense the affective state of a PwD.

7.2.3. PwD

Since this research has been performed on regular people we do not yet know how well it translates to PwD. A possible next step for the research could be to implement a similar experiment with PwD. Here perhaps a caretaker would have to fill in the AffectButton or another method of determining the affect should be used.

7.2.4. Music Emotion Recognition Algorithm

In our research we used the Spotify values for valence and arousal because they were easily accessible to us. However it is not clear based on what data Spotify constructs these. Implementing a different emotion recognition algorithm based on data specific for this problem could offer better results. Especially when the data used for this is based specifically on induced emotions.

7.2.5. Begin Emotional State

As mentioned we did not have participants who started with a negative valence. In real situations we might want to use the playlist generation to bring someone from a negative to a positive affective state. Therefore for future research it would be interesting to see how participants who start with a lower valence respond. We might encounter those participants automatically when more participants are used for a study or participants could specifically be asked to use it at a later time when they have a negative valence.

7.2.6. Memories

For us the research into the memory aspect occurred naturally due to the need of having questions which kept the attention of the participant. However we did find some interesting observations. It would be interesting to study these further in a next experiment with a better way to tag memories than our simple question with five options. This could confirm whether for example mostly positive memories are associated with songs and how the type of memory influences the change in affective state. If this pans out it could also be interesting to look into how this can be incorporated further into an algorithm to regulate emotions. Data on memories for songs could be gathered explicitly by asking about them or perhaps in the far-future a recognition system could be developed which attempts to automatically detect when a user recalls a memory for a song.

In this chapter we have discusses the results of the experiment that were presented in the previous chapter. The next chapter will provide a conclusion to this report.

8. Conclusion

In our introduction we started by highlighting the rising number of people with dementia combined with a shortage in the healthcare workforce. This combination makes looking into methods to support the care of PwD interesting. One of these being the individualized music intervention (IMI) as described by Gerdner [36]. In this intervention music, which is known to be preferred by the PwD, is played to reduce agitation.

We defined two main research questions relating to this topic. The first was: "How can we incorporate the IMI into technology to care for PwDs?". We answered this research question by showing the different functional components, along with their responsibilities, which have to be fulfilled to implement the intervention. We showed how these components work in the current implementation in which a trained worker fulfills many of the components. We also showed how they might work in a music therapy support app in the near future and even in a fully automated intervention in an assistive robot in the far future. We also discussed how many of these functional components, which will have to be developed to reach a fully automated system, have a lot of overlap with currently emerging research in for example the field of affective computing. With this we offer a functional view on the currently available evidence based guidelines by Gerdner [36]. By defining these functional components with their own responsibilities it allows us and other researchers to develop them individually using research from various other fields.

Our second research question related to a specific functional component, the playlist generator. The question was: "Assuming knowledge of a list of preferred music for a user, can we generate playlists which better regulate emotions than a random shuffle?". To generate playlists we implemented an algorithm based on Markov Decision Processes. We then compared this algorithm to a random shuffle via an online experiment. In this experiment we showed that our implementation performs better than a random shuffle of participants preferred music when the goal is to regulate the valence of a participants affective state. In the experiment our playlist generator had a target valence of ≥ 0.75. The final valence of the experimental group (M=0.88, SD=0.08) was significantly higher than the final valence of the control group (M=0.73, SD=0.17), $t(16.4)$,$p=0.017$. Additionally in this experiment we also observed that participants generally did not seem to have negative memories associated with songs. We however could not investigate the aspect of memory more in depth due to the our limited sample size. Previous research into emotion based music recommendations most of the time did not focus on inducing or regulating emotions [9, 27, 81, 111] or did not focus on evaluating the induction of emotions on real participants [40]. Others tried to directly influence physiological factors such as the heart-rate or EEG signals [52–54, 64, 80], this makes it more difficult to incorporate it with currently available data such as the affect data available through the Spotify API. MDPs have also been used in music recommendation research before [18, 21, 42, 51] but these also did not focus on the induction of emotions and/or did not perform experiment with real participants [42]. This makes the research we performed novel.

We also discussed some of the limitations of our experiment. A limited sample size, combined with little variation in the backgrounds of participants means we are unsure of how well the results generalize, especially to our eventual target group of PwD. The sample size also limits the usefulness of our dataset for future research. Our participants also all already started with a positive valence, a more diverse set of affective states before listening to the playlists might also offer different results.

Based on these limitations we made some recommendations for future work. This includes gathering more data to build more complex methods, testing even simpler methods that do keep in mind affect in some way (e.g. randomly sampling songs the user likes from a certain region in the valence-arousal-space) and testing similar algorithms with PwDs.

While there is still work left in general we can conclude that we answered both the main research questions set out in this thesis. To answer the first we showed which functional components are required to implement IMI into various technologies, we also provided a roadmap from the current implementation towards the far-future of a fully automated IMI. To answer the second we implemented a playlist generation algorithm based on MDPs which, comparing this to a random shuffle in an online experiment showed promising results. The results to this last question show that a lot could be gained by generating playlist in IMI with the specific goal of regulating emotions.

A. Ontology

Affect: Term that can refer to mood and/or emotion

Affective State: State containing both emotion and mood

Emotion: Short-lived reaction to stimulus as defined by Shiota and Kalat

Emotional state: Emotion expressed in either the VA or PAD-scale. The emotional state can be that of a user. But also of a song (the Spotify VA values).

Family or Friend (FoF): Family or Friend, person close to the PwD who might for example have knowledge about their music preferences.

Individualized Music Intervention (IMI): With this we mean the intervention as described in the guidelines by Gerdner [36].

Markov Decision Process (MDP): Markov Decision Process, framework to model decision making, defined in Chapter 4.

Mood: More long-term than emotions, not a direct reaction to a stimulus.

Partially Observable Markov Decision Process (POMDP): Partially Observable Markov Decision Process, framework to model decision making under uncertainty, defined in Appendix E.

Person with Dementia (PwD): Person with dementia, in our thesis often seen as the person for whom an intervention is done or a playlist is generated.

Pleasure Arousal Dominance Scale (PAD scale): Similar to the VA-scale, but Valence is pleasure and there is an extra third dimension of dominance.

User: User can mean the PwD/FoF or worker operating the system.

Valence-Arousal Scale (VA-scale): Valence-arousal scale, 2D representation for emotions, most often used to represent emotions in this thesis.

Worker: Worker will often be used in short for all kind of healthcare workers ranging from workers at a facility to people supporting PwDs at home.

B. Prototype 1: Data Gathering Application

In this chapter we discuss the first prototype that was developed for this thesis. We start by discussing the motivation for this prototype. We move on to the details that were worked out for the prototype. And finally we discuss

B.1. Motivation

The lack of data makes research into music recommendations for PwD difficult. Therefore the first proposal concerned an application which could be used to gather data about music preferences for PwD. The application would be made so that family or friends (FoF) of the PwD could listen to music together with them and provide feedback on their responses. The dataset resulting from this could then be used for preliminary research into music preferences of PwD and be published for further research. Two ideas included to motivate people to use the system was a content-based recommendation system and using data from Wikipedia to provide facts about the songs as talking-points.

B.2. Details

For this proposal we worked out a first version of the specifications, a questionnaire and a non-functional visual prototype. We briefly discuss each.

B.2.1. Specification

Through applying the socio-cognitive engineering method a first version of the specifications of such an application was made. The full specification can be found in Appendix C

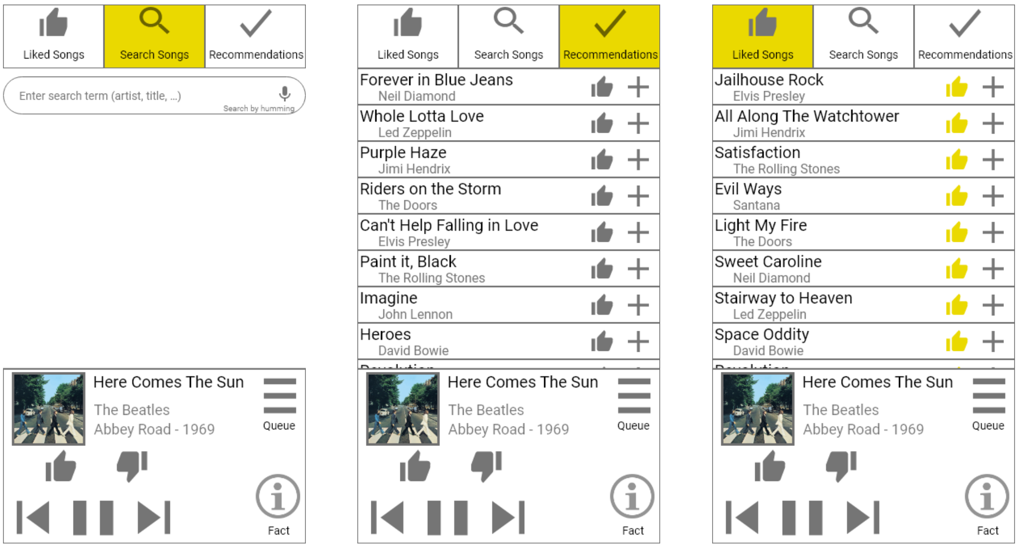
B.2.2. Visual prototype

Based on these specifications a first prototype was made. This prototype had working buttons to move between tabs but no other functionality. The three main tabs of this prototype can be seen in Figure B.1. Figure B.1a shows the screen where the FoF could search for songs, Figure B.1b where they could get recommendations and Figure B.1c shows where they could view previously liked songs. Figure B.2a shows how a fact could be and Figure B.2b shows what happens when you click on a link in the fact.

B.2.3. Questionnaire

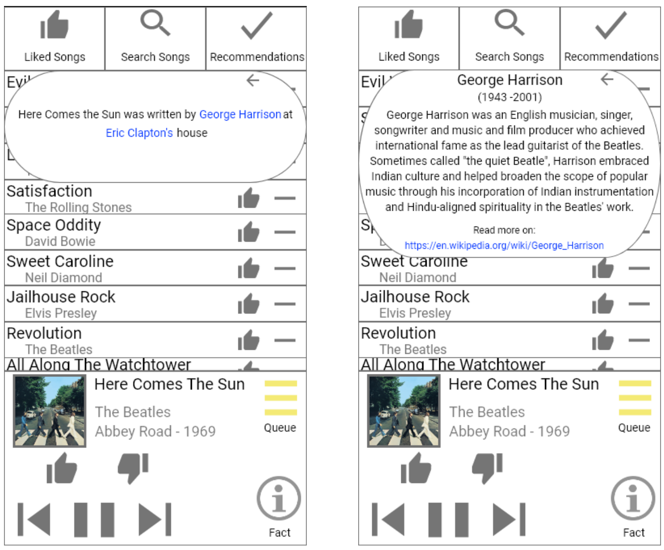
To further develop the design of this prototype a questionnaire was also developed which was to be carried out by the FoF of PwDs. The idea was that the questionnaire would be spread through various organizations to the FoF of PwD. The questionnaire was developed both in Dutch and English with the hope that this would provide both local and international responses. The aim of the questionnaire was mostly to gain insight into:

- Demographic information on potential users.
- Exploring design space and stakeholders
  - What do FoF already use music for (if anything?)
  - What devices do they have available?
  - Do they use streaming services?



(a) Search screen (b) Recommendation screen (c) Liked songs screen

Figure B.1: The main screens of the prototype



(a) Showing a fact (b) Clicking through on a fact screen

Figure B.2: Showing facts about a song in the prototype

  - What sort of values do they find important in such a system?

- Gauging interest and getting feedback on the system by presenting the visual prototype.
- Gauging knowledge of FoF about music preferences of PwD.
- Gain initial small dataset on music preferences of PwD.

The full questionnaire, both in English and Dutch can be found in Appendix D

B.3. Reason for change

Due to Covid-19 the possibilities of people using this app together in real-life declines. It was therefore decided to look into a different direction and not focus on only the preferences of the PwD but also the affective state.

C. Prototype 1: SCE Specification

C.1. Design Scenarios

Scenario 1 - Visit - First use

Jan is visiting Bart. Jan has heard about a new app which he can use to listen to music together with Bart. Jan decides to try it out. He opens the app on his phone and has to fill out a short questionnaire about the music preferences of Bart. From this questionnaire Jan gets some music recommendation to listen to with Bart. Jan tries some of them out and marks the songs that Bart seemed to react well to.

Scenario 2 - Visit - Repeated Use

Jan visits Bart again after first using the system. Jan notices that Bart is anxious and conversation is difficult. Jan remembers the reactions from Bart to the music of last time and decides to open the app again. He can then listen together with Bart to the songs that have been marked before or he can get new recommendations from the app.

Scenario 3 - Video Call

Jan is unable to visit Bart, instead he decides to video call. Again Bart is anxious and conversation is difficult. Jan can listen to music again through the video call together with Bart. To work out further: How would work? (e.g. incorporated in video call app? Or perhaps a phone that links to Jan's phone that the workers can give to Bart?)

C.2. Use Cases

Use Case 1 - Set up

Actors: PwD, FoF, system
Circumstance: FoF has heard about the usage of the system and wants to set it up.
Precondition: FoF has system available. PwD and FoF can communicate during use case.
Postcondition: System knows initial questionnaire answers, FoF understands how to use the system, FoF understands benefits of the system
Method: Inform FoF, take questionnaire, tutorial on functionality
Steps:

1. FoF starts the system.
2. The system explains the potential benefits of listening to music together with PwD.
3. The system explains how it will make music recommendations based on an initial questionnaire and further feedback from the FoF.
4. The system presents the FoF with the initial questionnaire.
5. FoF fills in the initial questionnaire together with the PwD.
6. The system gives a tutorial to the FoF it explains:
   (a) How they can see music recommendations.
   (b) How they can search for music themselves.

   (c) How they can tag music as giving a positive response.
   (d) How they can view previously tagged music.

7. The system presents the user with the first recommendations and incites the user to start listening to music together.

Use Case 2 - FoF listens to recommendations with PwD

Actors: PwD, FoF, system
Circumstance: FoF has set up the system and wants to listen to music together with the PwD.
Precondition: Set up use case has been completed. FoF and PwD are together.
Postcondition: FoF have listened to music recommendations together. Hopefully PwD mood is improved and PwD and FoF had meaningful interactions. Possibly new songs have been tagged.
Method: Recommend music, play music
Steps:

1. FoF starts the system.
2. The system presents the FoF with music recommendations.
3. The FoF selects one of them to listen to.
4. The system starts playing the song and replaces it with a new music recommendation.
5. The PwD responds positively to the song.
6. The FoF tags the song as having a positive reaction.
7. Once the song has ended the system plays the next recommendation.

Alternative Steps:

1. FoF starts the system.
2. The system presents the FoF with music recommendations.
3. The FoF selects on of them to listen to.
4. The system starts playing the song and replaces it with a new music recommendation.
5. The PwD does not respond to the song.
6. The FoF asks the system to skip the song.
7. The system starts playing the next song in the recommendation.

Alternative Steps:

1. FoF starts the system.
2. The system presents the FoF with music recommendations.
3. The FoF selects one of them to listen to.
4. The system starts playing the song and replaces it with a new music recommendation.
5. The PwD responds negatively to the song.
6. The FoF tags the song as having a negative reaction.
7. The system stops playing the song and presents the FoF with new recommendations to pick from.
8. Repeat from step 3

Use Case 3 - FoF listens to tagged playlist with PwD

Actors: PwD, FoF, system
Circumstance:
Precondition: Set up use case has been completed. FoF and PwD can see each other. Some songs have been tagged as having a positive reaction
Postcondition: FoF have listened to the tagged playlist together. Hopefully PwD mood is improved and PwD and FoF had meaningful interactions.
Method: Show tagged playlist, play music
Steps:

1. FoF starts the system.
2. The system presents the FoF with music recommendations.
3. The FoF wants to listen to previously tagged music and asks the system to show this instead.
4. The system shows previously tagged music.
5. The FoF selects one of the songs.
6. The system starts playing the song.
7. When the song ends the system starts playing a random song from the tagged playlist.

Use Case 4 - Skipping song

Actors: FoF, system
Circumstance: FoF are using the system and want to skip a song due to a multitude of reasons. It could be that the PwD is currently not responding well to the song.
Precondition: FoF or worker are using the system to play music
Postcondition: Next song is being played
Method:
Steps:

1. FoF asks the system to skip the current song.
2. The system plays the next song in the current playlist.

Use Case 5 - Pause/play song

Actors: FoF, system
Circumstance: FoF are using the system and want to interrupt a song to for example talk to the PwD.
Precondition: FoF are using the system to play a playlist
Postcondition: Music stops/starts playing.
Method:
Steps:

1. FoF asks the system to pause/play the current song.
2. The system stops/starts playing the current song.

Use Case 6 - Changing volume

Actors: FoF, system
Circumstance: The FoF are using the system and notice that the music is either too loud or too silent.
Precondition: FoF are using the system to play music
Postcondition: Volume has been raised/lowered
Method:
Steps:

1. FoF asks the system to raise/lower volume.
2. The system raises/lowers the volume according to what was asked.

C.3. Requirements

Requirement 1 - Educate on benefits

Requirement: The system should be able to inform the FoF about the benefits of using the system (both for themselves and for the workers).
Claim: Knowing the benefits will make FoF more likely to try it.

Requirement 2 - Educate on working of the system

Requirement: The system should be able to inform the FoF about how the system works.
Claim: This will help the FoF understand the importance of certain actions they take and will improve confidence in using the system.

Requirement 3 - Take initial questionnaire

Requirement: The system should be able to take a initial questionnaire from the FoF about the music preferences of the PwD.
Claim: This will provide the system with enough information to make initial recommendations.

Requirement 4 - Tutorial FoF

Requirement: The system should be able to give the FoF a tutorial explaining the interface of the system.
Claim: After this the FoF knows how to use the system and gains further confidence in using the system.

Requirement 5 - Present music recommendations

Requirement: The system should be able to present the FoF with music recommendations, based on the initial questionnaire and previously tagged music.
Claim: The music recommendations will help the FoF build a tagged playlist with songs the PwD responds positively to.

Requirement 6 - Music playback

Requirement: The system allow the FoF to play music (from music recommendations, searched music or playlists).
Claim: Playing the music allows FoF and PwD to have interactions about the music being played. It also allows FoF to gauge the reaction to make tagged playlists.

Requirement 7 - Volume control

Requirement: System should allow FoF and workers to raise volume and lower volume while listening to music
Claim: FoF or worker is able to adjust the volume to a correct level.

Requirement 8 - Playback control

Requirement: System should allow FoFto skip, play and pause songs.
Claim: Gives autonomy to FoF.

Requirement 9 - Tagging songs

Requirement: During playback FoF should be able to tag a song as having a positive or negative response. The system should save this information.
Claim: FoF will tag music based on reactions of the PwD providing information for recommendations and creating an tagged playlist in the process.

Requirement 10 - Searching songs FoF

Requirement: The system should allow FoF to search for music to play themselves.
Claim: FoF will use this to look for music they know the PwD likes and help with creating tagged playlists.

D. Prototype 1: Initial FoF Survey

Research Questions/Topics

- Insight into potential user demographics
  - Country of residence of potential users
  - Age of potential users
- Insight into design space
  - What systems do they typically have available when visiting PwD?
  - What do they maybe already do with music?
  - Do they visit often?
- Gauging interest in a system to facilitate listening to/making playlists
  - Additionally determine if there are factors of influence here (such as age, previous experience with music, ...)
- Are FoF able to give examples of music the PwD likes when prompted for it?
- Gain data on music preferences PwD.

D.1. Survey (Dutch)

Intro

U bent uitgenodigd om deel te nemen aan een studie met de titel Situated Music Recommendation for People with Dementia. Deze studie wordt uitgevoerd door Bernd Kreynen van de TU Delft. Deelname aan de studie is bedoeld voor mensen die een familielid of kennis met dementie hebben.

Het hoofddoel van deze studie is om meer te weten te komen over de muziekvoorkeuren van mensen met dementie. Daarnaast willen we ook kijken naar de interesse in muziek van vrienden/familie van mensen met dementie. Het invullen van de vragenlijst zal ongeveer 15 minuten duren. De verkregen data zal gebruikt worden in de ontwikkeling van een applicatie die u in staat stelt om samen met uw kennis met dementie muziek te luisteren en voor onderzoek naar muziekaanbevelingen voor mensen met dementie.

Uw deelname in deze studie is volledige vrijwillig en u mag op ieder moment stoppen met de vragenlijst. U bent ook vrij om vragen op te laten als u deze niet wilt beantwoorden.

We verwachten niet dat er risico's verbonden zijn aan het deelnemen van deze studie, maar zoals met elke onlineactiviteit is een schending van de gegevens altijd een mogelijkheid. Dit risico zullen we beperken door de antwoorden op te slaan op beveiligde project storage waar enkel de onderzoekers toegang tot hebben. Er zal geen enkele data buiten uw antwoorden op de vragen bewaard worden.

De data zal later wel publiek beschikbaar gesteld worden voor onderzoek via het 4TU.Centre for Research Data. Indien antwoorden Identificeerbare of gevoelige data bevatten zullen deze echter niet mee gepubliceerd worden.

U kan de onderzoeker altijd contacteren op b.l.l.kreynen@student.tudelft.nl

Verder kan u op het einde van de survey ook optioneel uw e-mailadres invullen. Als u dit doet zullen we u informeren over:

- De resultaten van de studie

  - Inzichten in muziekvoorkeuren van mensen met dementie
  - Tips voor het opstellen van afspeellijsten voor mensen met dementie gebaseerd op de resultaten.

- U zal uitgenodigd worden om de applicatie, waarmee u samen met uw kennis naar muziek kan luisteren, uit te proberen.

Persoonlijke gegevens

In dit deel wordt om uw eigen persoonlijke gegevens gevraagd.

Wat is uw leeftijd?

____

Wat is uw geslacht?
(u mag 1 antwoord aanduiden)

- ○ M
- ○ V
- ○ andere: ____________

In welk land verblijft u?
(u mag 1 antwoord aanduiden)

- ○ Nederland
- ○ België
- ○ andere: ____________

Persoonlijke gegevens kennis met dementie

In dit deel wordt om de persoonlijke gegevens van uw kennis met dementie gevraagd.

Wat is hun leeftijd?

____

Wat is hun geslacht?
(u mag 1 antwoord aanduiden)

- ○ M
- ○ V
- ○ andere: ____________

Wat is uw relatie met hun?
(u mag 1 antwoord aanduiden)

- ○ Ik ben hun kind
- ○ Ik ben hun broer/zus
- ○ Ik ben hun echtgeno(o)t(e)/partner
- ○ andere: ____________

Eigen ervaring met muziek

In dit deel wordt om informatie gevraagd over uw eigen ervaringen met muziek.

Speelt u zelf een instrument?
(u mag 1 antwoord aanduiden)

- ○ Ja
- ○ Nee
- ○ andere: ____________

	Minder dan eens per maand	Eens per maand	Eens per week	Eens per dag	Meer dan eens per dag
Hoe vaak luisterd u naar muziek?	1 ○	2 ○	3 ○	4 ○	5 ○

Hoe luistert u naar muziek? (bv. Spotify op mobiel, radio, ...)

____

Heeft u een account op een muziek streaming service?
(u mag meerdere antwoorden aanduiden)

- ☐ Spotify
- ☐ Apple Music
- ☐ Google Play Music
- ☐ Tidal
- ☐ Neen
- ☐ andere: ______________

Ervaring met muziek kennis met dementie

In dit deel wordt om informatie gevraagd over de ervaring met muziek van uw kennis met dementie.

Spelen of Speelde ze vroeger een instrument?
(u mag 1 antwoord aanduiden)

- ○ Ja
- ○ Nee
- ○ andere: ______________

	Minder dan eens per maand	Eens per maand	Eens per week	Eens per dag	Meer dan eens per dag
Hoe vaak luisterde ze vroeger naar muziek?	1 ○	2 ○	3 ○	4 ○	5 ○
Hoe vaak luistert hij/zij nu naar muziek?	1 ○	2 ○	3 ○	4 ○	5 ○

Hoe eens bent u het met de volgende statement(s)?

	Helemaal oneens			niet eens en niet oneens			Helemaal eens
Ik ben bekend met zijn/haar muziekvoorkeuren	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○

Hoe luisterde uw kennis vroeger naar muziek
(u mag meerdere antwoorden aanduiden)

- ☐ Radio
- ☐ Cd-speler
- ☐ Platenspeler
- ☐ Bandrecorder
- ☐ Cassettedeck
- ☐ Streaming service
- ☐ MP3-speler/IPod
- ☐ andere: ____________

Hoe luistert uw kennis nu naar muziek
(u mag meerdere antwoorden aanduiden)

- ☐ Radio
- ☐ Cd-speler
- ☐ Platenspeler
- ☐ Bandrecorder
- ☐ Cassettedeck
- ☐ Streaming service
- ☐ MP3-speler/IPod
- ☐ andere: ____________

Kan u voorbeelden geven van muziek waar uw kennis graag naar luisterde, speciale herinneringen aan heeft en/of sterke emoties mee verbonden zijn? U kan optioneel aan het nummer ook een herinnering en/of een emotie toevoegen die verbonden is aan het nummer voor uw kennis (enkel een nummer, een nummer + emotie of een nummer + herinnering is dus ook prima). (Ook niet alle 20 rijen moeten ingevuld worden, vul zo veel of weinig in als u zelf weet).

	Nummer (Titel - Artiest)	Herinnering	Emotie
Nummer 1			
Nummer 2			
Nummer 3			
Nummer 4			
Nummer 5			
Nummer 6			
Nummer 7			
Nummer 8			
Nummer 9			
Nummer 10			
Nummer 11			
Nummer 12			
Nummer 13			
Nummer 14			
Nummer 15			
Nummer 16			
Nummer 17			
Nummer 18			
Nummer 19			
Nummer 20			

Bezoek/contact met kennis

In dit deel wordt om informatie gevraagd over bezoeken/contact met uw kennis.

	Minder dan eens per maand	Eens per maand	Eens per week	Eens per dag	Meer dan eens per dag
	1	2	3	4	5
Hoe vaak gaat u op bezoek bij uw kennis? (In normale omstandigheden)	◦	◦	◦	◦	◦

Wanneer u op bezoek gaat bij uw kennis, tot welke van de volgende heeft u dan toegang?
(u mag meerdere antwoorden aanduiden)
☐ Een computer
☐ Een tablet
☐ Een smartphone
☐ andere: ____________

Heeft u soms op andere manieren contact met uw kennis? (bv. telefonisch)
(u mag 1 antwoord aanduiden)
◦ Ja
◦ Nee

Alternatieve manieren contact met kennis

In dit onderdeel wordt om meer informatie gevraagd over andere manieren waarop u contact hebt met uw kennis. **Note:** Deze sectie is alleen zichtbaar in de online vragenlijst als er op de vorige vraag ja geantwoord wordt.
Op welke manieren heeft u nog contact met uw kennis, buiten normale bezoeken?
(u mag meerdere antwoorden aanduiden)
☐ Bellen
☐ Videobellen
☐ Email
☐ Sms
☐ Whatsapp
☐ andere: ____________

	Minder dan eens per maand	Eens per maand	Eens per week	Eens per dag	Meer dan eens per dag
	1	2	3	4	5
Hoe vaak heeft u contact op een van deze manieren met uw kennis? (In normale omstandigheden)	◦	◦	◦	◦	◦

Ervaringen met samen naar muziek luisteren

In dit onderdeel wordt gevraagd naar uw ervaring met samen naar muziek luisteren.
Hoe eens bent u het met de volgende statement(s)?

	Helemaal oneens			niet eens en niet oneens			Helemaal eens
	1	2	3	4	5	6	7
Ik heb interesse in samen met mijn kennis naar muziek luisteren.	◦	◦	◦	◦	◦	◦	◦

Heeft u wel eens samen met uw kennis naar muziek geluisterd?
(u mag 1 antwoord aanduiden)

- ◦ Ja
- ◦ Neen

Samen muziek luisteren

Dit onderdeel gaat verder in op samen naar muziek luisteren. **Note:** Dit onderdeel wordt in de online vragenlijst enkel zichtbaar als op de vorige vraag ja ingevuld is.
Hoe eens bent u het met de volgende statement(s)?

	Helemaal oneens			niet eens en niet oneens			Helemaal eens
	1	2	3	4	5	6	7
Hoe vaak luisterd u samen met uw kennis naar muziek?	◦	◦	◦	◦	◦	◦	◦

Wie bepaalt er dan welke muziek er geluisterd wordt?

Naar wat voor muziek luisteren jullie dan?

Hoe luisteren jullie naar de muziek? (bv. met spotify op mobiel, radio, ...)

Waar luisteren jullie samen naar muziek?

Wanneer luisteren jullie samen naar muziek? (Zijn er bv. dingen die daar aanleiding toe geven?)

Systeem

Voor het afstudeerwerk zal er een product ontwikkeld worden (naam X). Het doel van X is om u in staat te stellen samen met uw kennis met dementie naar muziek te luisteren die ze leuk vinden wanneer u bijvoorbeeld op bezoek bent bij uw kennis. In de applicatie kan u naar muziek zoeken en deze beluisteren. Door tijdens het luisteren feedback te geven over hoe uw kennis reageert zal het systeem meer te weten komen over de voorkeuren van uw kennis en aan de hand hiervan muziek aanbevelen die uw kennis mogelijk ook leuk vindt. Dit zal er uiteindelijk voor zorgen dat u een uitgebreide afspeellijst krijgt van muziek die uw kennis leuk vindt. Deze kan u dan gebruiken bij bezoeken om samen als activiteit naar muziek te luisteren, maar u kan hem ook delen met bijvoorbeeld zorgverleners zodat zij de muziek kunnen afspelen wanneer u er niet bent en uw kennis bijvoorbeeld geïrriteerd raakt. Een voorbeeld van hoe dit systeem er uitkomt te zien kan u op de volgende link bekijken (sommige navigatie knoppen werken wel, maar verder is het voorbeeld niet functioneel): https://xd.adobe.com/view/7bb7310b-c962-49fc-9ec6-45191062318a-2ee4/?fullscreen

Hoe eens bent u het met de volgende statement(s)?

	Helemaal oneens			niet eens en niet oneens			Helemaal eens
Ik heb interesse in X uit te proberen	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○
Ik heb interesse in een systeem dat mij helpt muziek te vinden voor mijn kennis (niet per se X)	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○
Ik heb interesse in een systeem waarmee ik samen naar muziek kan luisteren met mijn kennis (niet per se X)	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○
Ik vind het kunnen afspelen van gepersonaliseerde muziek voor mijn kennis belangrijk	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○

Indien u het kunnen afspelen van gepersonaliseerde muziek belangrijk vindt; Waarom vind u dit?

Is er iets specifiek dat je graag zou willen dat het systeem kan doen?

Heeft u nog commentaar/opmerkingen op het voorbeeld dat hierboven vermeld werd? (u kan het weer bekijken op https://xd.adobe.com/view/7bb7310b-c962-49fc-9ec6-45191062318a-2ee4/?fullscreen) .

Afsluiting

Is er nog iets wat u kwijt wilt? (Over de survey, X, of iets helemaal anders)

Einde

Bedankt voor uw deelname aan de survey! Als u nog meer mensen kent die hierinh geïnteresseerd zouden zijn zou het leuk zijn als u ze ook uitnodigt met de volgende link: **TODO:** voeg link toe.

Als u geïnteresseerd bent om de resultaten van de studie te krijgen en om een uitnodiging te krijgen om X uit te proberen wanneer het verder ontwikkeld is kan u (optioneel) hier uw email adres nog opgeven. Dit email adres zal los van de vorige antwoorden opgeslagen worden en zal nooit met een derde partij gedeeld worden.: Wat is uw email adres?

Als er nog vragen/opmerkingen/... zijn mag u mij altijd contacteren op b.l.l.kreynen@student.tudelft.nl .

D.2. Survey (English)

Intro

You are being invited to participate in a research study titled *Situated Music Recommendation for People with Dementia.* This study is being done by *Bernd Kreynen* from the TU Delft. **Participation in the survey is meant for family members or friends of a person with dementia.**

The purpose of this research study is to explore **music preferences of people with dementia** and to gauge interest in systems that utilize music recommendations to personalize music for people with dementia. It will take you approximately **15 minutes** to complete. The data will be used to shape the development of an application that allows family and friends to listen to music together with a person with dementia.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any question.

We believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach is always possible. We will minimize any risks by storing the responses on secure project storage which only the researchers will have access to. No additional data outside of your responses will be stored.

The data of this research will be made publicly available after the research as well, through the 4TU.Centre for Research Data, but answers of open questions will only be published after being checked for sensitive and identifiable data.

You can contact the researcher, *Bernd Kreynen*, of this study via: *b.l.l.kreynen@student.tudelft.nl*

Furthermore at the end of the survey you can optionally leave your e-mail address. If you do so we will use this to inform you about:

- The results of the study
  - Insights into music preferences of people with dementia
  - Tips for the creation of playlists for people with dementia based on the results
- You will be invited to try out the application that will be developed that lets you listen to music together with your friend/family member with dementia.

Personal Details

This part concerns your personal details.

What is your age?

____

What is your gender?
(you may check one answer)

○ M
○ F
○ other: __________

In which country do you reside?
(you may check one answer)

○ Australia
○ England
○ North-America
○ Northern Ireland
○ Republic of Ireland
○ Scotland
○ Wales
○ other: __________

Personal Information Acquaintance with Dementia

This part concerns the personal details of your acquaintance with dementia.

What is their age?

____

What is their gender?
(you may check one answer)

○ M
○ F
○ other: __________

What is your relation to them?
(you may check one answer)

○ I am their child
○ I am their sibling
○ I am their spouse
○ other: __________

Own experience with music

This part concerns your own experiences with music.

Do you play an instrument?
(you may check one answer)

- ○ Yes
- ○ No
- ○ other: ____________

	Less than once per month	Once per month	Once per week	Once per day	More than once per day
How often do you listen to music?	1 ○	2 ○	3 ○	4 ○	5 ○

How do you listen to music? (e.g. Spotify on mobile, radio, ...)

____

Do you have an account on a music streaming service?
(you may check multiple answers)

- ☐ Spotify
- ☐ Apple Music
- ☐ Google Play Music
- ☐ Tidal
- ☐ No
- ☐ other: ____________

Experience with music acquaintance

This part asks about the experience with music of your acquaintance with dementia.

Do/did they play an instrument?
(you may check one answer)

- ○ Yes
- ○ No
- ○ other: ____________

	Less than once per month	Once per month	Once per week	Once per day	More than once per day
How often did they use to listen to music?	1 ○	2 ○	3 ○	4 ○	5 ○
How often do they listen to music now?	1 ○	2 ○	3 ○	4 ○	5 ○

In how far do you agree with the following statements?

	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
I am familiar with their music preferences	○	○	○	○	○	○	○

How did they use to listen to music?
(you may check multiple answers)

- ☐ Radio
- ☐ Cd-player
- ☐ LP-player
- ☐ Band-recorder
- ☐ Cassette-deck
- ☐ Streaming service
- ☐ MP3-player/IPod
- ☐ other: ____________

How do they listen to music now?
(you may check multiple answers)

- ☐ Radio
- ☐ Cd-player
- ☐ LP-player
- ☐ Band-recorder
- ☐ Cassette-deck
- ☐ Streaming service
- ☐ MP3-player/IPod
- ☐ other: ____________

Can you give examples of songs that your acquaintance used to listen or to which special emotions/memories are attached?
Not every column of the table needs to be filled in. A song without an emotion or memory is also fine for example.
Not all 20 rows needs to be filled either fill in as many or as few songs as you know.

	Song (Title - Artist)	Memory	Emotion
Song 1			
Song 2			
Song 3			
Song 4			
Song 5			
Song 6			
Song 7			
Song 8			
Song 9			
Song 10			
Song 11			
Song 12			
Song 13			
Song 14			
Song 15			
Song 16			
Song 17			
Song 18			
Song 19			
Song 20			

Visiting/contact with acquaintance

This part concerns your contact/visits with your acquaintance.

	Less than once per month	Once per month	Once per week	Once per day	More than once per day
	1	2	3	4	5
How often do you visit your acquaintance? (In normal circumstances)	○	○	○	○	○

When you're visiting your acquaintance to which of the following do you have access?
(you may check multiple answers)

- ☐ A computer
- ☐ A tablet
- ☐ A smartphone
- ☐ other: ____________

Do you have contact with your acquaintance in other manner (e.g. by phone)
(you may check one answer)

- ○ Yes
- ○ No

Alternative methods of contact

This part concern other manner which you use to contact your acquaintance. **Note:** This section is only visible when yes has been answered to the previous question.
In which ways do you have contact with your acquaintance, outside of normal, physical visits?
(you may check multiple answers)

- ☐ Calling

☐ Video-calling
☐ Email
☐ text message
☐ Whatsapp
☐ other: ____________

	Less than once per month	Once per month	Once per week	Once per day	More than once per day
	1	2	3	4	5
How often do you contact your acquaintance in one of these manners? (In normal circumstances)	○	○	○	○	○

Experiences with listening to music together

This part concerns your experiences with listening to music together.
In how far do you agree with the following statements?

	Strongly disagree			Neither agree nor disagree			Strongly agree
	1	2	3	4	5	6	7
I am interested in listening to music together with my acquaintance.	○	○	○	○	○	○	○

Have you listened to music together with your acquaintance?
(you may check one answer)

○ Yes
○ No

Listening to music together

This part asks for more information about listening to music together with your acquaintance. **Note:** This part only becomes visible if they answered yes on the previous question.

	Less than once per month	Once per month	Once per week	Once per day	More than once per day
	1	2	3	4	5
How often do you listen to music together with your acquaintance?	○	○	○	○	○

Who decides which music is played?

____

Which music do you listen together?

How do you listen to the music together? (e.g. Spotify on mobile, radio, ...)

Where do you listen to music together?

When do you listen to music together? (Are there for example certain things that give rise to it?)

System

As part of this research a product will be developed (named X). The goal of x is to allow you to listen to music together with your acquaintance. While listening to the music you give feedback to the system about how they respond to it. Based on this the system will recommend more music that your acquaintance might like. Over time this will build up a playlist of songs your acquaintance likes. This can be used when you're visiting to listen to it together but could also be shared to for example healthcare staff who could play the music when your acquaintance becomes agitated. A demo of the application can be viewed with the following link (some of the navigation buttons do work, but otherwise the demo is not yet functional): https://xd.adobe.com/view/7bb7310b-c962-49fc-9ec6-45191062318a-2ee4/?fullscreen

In how far do you agree with the following statements?

	Strongly disagree			Neither agree nor disagree			Strongly agree
I am interested in trying X	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○
I am interested in a system that helps me find music for my acquaintance (not necessarily X)	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○
I am interested in a system that allows me to listen to music together with my acquaintance (not necessarily X)	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○
I think being able to play personalized music for my acquaintance is importnat.	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○

If you think listening to music together is important, why do you think this??

Is there something specific you would like X to be able to do?

Do you have any comments/suggestion/,,. on the demo shown above? (you can revisit it on https://xd.adobe.com/view/7bb7310b-c962-49fc-9ec6-45191062318a-2ee4/?fullscreen)

Outro

Is there anything else you would like to let us know? (About the survey, X, or something entirely different?)

End

Thanks for your participation in the survey. If you know any more people who would be interested in this then it would be great if you'd invite them to the survey via the following link:
If you're interested in getting the results of the study and to get and invite to try out the application once it is online then you can leave your e-mail address through the following link:
What is your e-mail address?

____

If there are any remaining questions/comments/… then you can always contact me through b.l.l.kreynen@student.tudelft.nl
.

Als u geïnteresseerd bent om de resultaten van de studie te krijgen en om een uitnodiging te krijgen om X uit te proberen wanneer het verder ontwikkeld is kan u (optioneel) hier uw email adres nog opgeven. Dit email adres zal los van de vorige antwoorden opgeslagen worden en zal nooit met een derde partij gedeeld worden.:

E. Partially Observable Markov Decision Processes

In this chapter we will take a closer look at Partially Observable Markov Decision Processes (POMDPs). This chapter will have a similar structure to the previous chapter on MDPs. We will start by giving a formal definition, move on to some methods that can be used to solve POMDPs and finally look into how this could be applied to playlist generation in IMI.

E.1. Partially Observable Markov Decision Process Definition

This section will give a formal definition of POMDPs. It will be heavily based on *Chapter 12: Partially Observable Markov Decision Processes* by van Matthijs Spaan from the book *Adaptation, Learning and Optimization: Reinforcement Learning* [89]. Partially Observable Markov Decision Processes are defined similarly to Markov Decision processes with the key difference that we do not assume to fully know the state the agent is in. Instead we have observations and associated with these observations are probabilities for a certain state. Due to this POMPDs are able to deal with decision making with uncertain sensing [89], such as when the sensing of the emotion is uncertain for example. Spaan gives the following definition:

Definition E.1.1 *A partially Observable Markov Decision Process is a tuple* $< S, A, \omega, T, O, R >$*. In which S is a finite set of states, A is a finite set of actions, ω is a finite set of observations, T is a transition function defined as* $T : S \times A \times S \rightarrow [0,1]$*, O is an observation function defined as* $O : S \times A \times \omega \rightarrow [0,1]$ *and R is a reward function defined as* $R : S \times A \times S \rightarrow \mathbb{R}$

Observations

As mentioned in the definition there is a finite set of observations, ω. In MDP the agent receives a state, in a POMDP the agent instead receives an observation $o \in \omega$. The observation an agent receives is conditional on the next state and can also be conditional on the action taken. As shown in the definition this gives $O : S \times A \times \omega \rightarrow [0,1]$ as the observation function. Additionally if the action has no influence on the observations it can also be noted as $O : S \times \omega \rightarrow [0,1]$ [89].

Belief Vector

To make optimal decisions in a POMDP we require memory [89]. One possible solution is to store all past actions and observations, but this memory would grow indefinitely [89]. Instead we can define the POMDP as a belief-state MDP. These use a belief vector b(s) to contain all information about the past. b is a probability distribution which is then used as the state in a MDP [89]. This belief vector is then updated by Bayes' rule every time the agent makes an observation. However for this we need to know the transition and observation function. This is only available in model-based approaches to solving the MDP [89].

Value Functions

The value function is defined similarly as in an MDP. However instead it is defined over the belief vector b [89].

Policies

Like value functions policies are defined similarly as in an MDP. With the difference that a policy now maps beliefs to actions [89].

E.2. Solving POMDPs

E.2.1. Model-based

There are several model-based method for computing policies [89]. This section gives a brief overview. For more detailed explanations on them we refer to [89].

Heuristics based on MDP
These still track the beliefs but instead of solving a POMDP they solve an MDP which is of lower complexity [89]. A simple example of these is the most likely state (MLS). In MLS we take the state from the belief vector with the highest probability and use this as the state of an MDP [89]. An advantage of these approaches is that they are less complex to solve, however they likely fail when the belief has a more complex shape [89].

Exact Value Iteration
Similarly to value iteration in an MDP we can apply value iteration to POMDPs. But due to POMDPs being intractable in general [89] these solutions might not be that interesting and will not be looked into further in this thesis.

Point-based Value Iteration
This is an approximation technique which is based on the idea of computing solution only for the believes that are actually reachable in the environment [89]. Instead of using the whole belief space we only use a smaller set of prototype believes that the agent sampled [89]. Important to the solutions in this approach is how the belief state is sampled and many different methods for this have been proposed [89].

Grid-based approximation
These use fixed or variable grids on the belief space. value backups are performed only for the grid points [89]. Different approaches for the selection of grid points and interpolation rules have been proposed [89]. Spaan mentions that regular grids do not scale well with high dimensionality and that non-regular grids have the downside of expensive interpolation rules [89]

Policy Search
These search through the policy space, multiple methods have been proposed. Examples are gradient ascent and bounded policy iteration [89]. It can also include heuristic search methods [89]. This approach can be prone to local optima [89].

E.2.2. Model-free

These can be divided into direct and indirect methods. Direct methods do not reconstruct the underlying POMDP model, instead they for example map observation histories to actions. Indirect methods try to estimate the POMDP model when interacting with it [89].

E.3. Applied to IMI

Model-free methods with the POMPD have the same issues in the application to PwD as model-free MDP methods, these have been explained in the previous chapter, Section 4.3.

E.3.1. Songs as Actions

As in Section 4.3 we can define actions as picking a song. We incur similar problems with defining and computing the transition function. Namely that we have too little data. Additionally we also now have to model an observation function. This observation function will depend on what method we use to observer the emotional state. For example do we use the AffectButton [13] or do we automated affect recognition (Section 2.7) And once again we will need previous data or research to base our model on.

E.3.2. GridWorld

The approach described in Section 4.3 for a GridWorld model will no longer work with most methods of solving POMDPS. In Section 4.3 we defined the actions as relative movements (e.g. up/down, left/right) and select a song based on the current grid of the agent. This relative movement does not work if we for example utilise a belief vector. This approach would only work when we apply MLS, from the Heuristics based on MDP as described in Appendix E.2.1. In this case we would also have to model the observation function since MLS uses the most likely state from the belief vector, for which we need to know both the transition function and observation function. In our prototype in Section 4.4 we simplified our transition function to be deterministic, since we have little data to make other assumptions on. However this is unlikely to be correct. Not being able to properly model the transition function yet could greatly impact the performance of the policy from the solved POMDP, even if we are able to model the observation function.

E.4. Conclusion

HoIn this chapter we gave an overview and definition of POMDPs. Because POMDPs can deal with environments that are not fully observable. Such as an emotional state from a PwD which might not be observed completely correctly. However we also showed that to correctly model POMDPs for this problem we lack data to make decisions on. Therefore in this thesis we will focus on initial research with the simple MDP formulation and through this also generating data which future research relating to the problem of playlist generation to invoke emotions can build. Our design for an experiment with this focus will be explained in the next chapter.

F. Additional Statistics

In this appendix we show some additional graphs and tables that we generated but which were not used in our own analysis. Interested readers can take a look at these and make their own interpretations. The names used in this chapter correspond to the names in Table 5.1 and Table 5.2.

F.1. Descriptive Statistics

F.1.1. Emotional States

Appendix F.1.1 shows the descriptive statistics about emotional states during the experiment. Figures F.3 and F.4 show density plots for the arousal and dominance. We also show the graphs of the deltas in arousal and dominance. These are shown in Figures F.5, F.7 and F.8

F.1.2. End Questions

Appendix F.1.2 shows the descriptive statistics about answers of the end questions.

F.2. Welcth test

We also performed Welch tests on the other PAD dimensions. The full tables for all dimensions is shown in Figure F.10

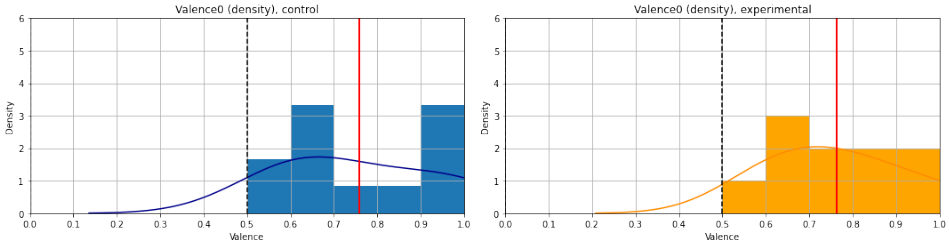
F.3. Correlations

Finally we made a few correlation matrices. These are shown in Figures F.11 to F.14.

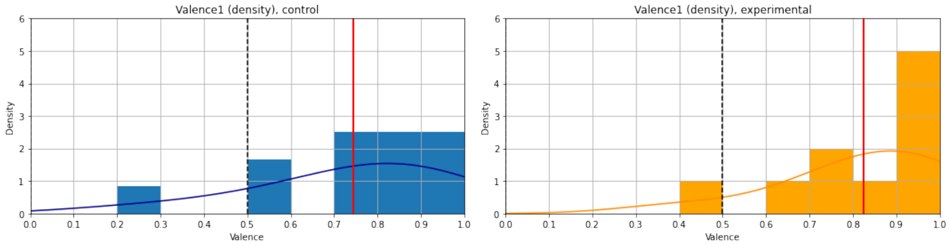
Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
valence0	0	12	.7583	.17251	.04980	.6487	.8679	.55	1.00
	1	10	.7640	.15064	.04764	.6562	.8718	.57	1.00
	Total	22	.7609	.15913	.03393	.6904	.8315	.55	1.00
arousal0	0	12	.4158	.39149	.11301	.1671	.6646	.02	1.00
	1	10	.3420	.39771	.12577	.0575	.6265	.00	1.00
	Total	22	.3823	.38663	.08243	.2108	.5537	.00	1.00
dominance0	0	12	.6858	.21589	.06232	.5487	.8230	.14	.98
	1	10	.7350	.13534	.04280	.6382	.8318	.56	.95
	Total	22	.7082	.18136	.03867	.6278	.7886	.14	.98
valence1	0	12	.7450	.22224	.06416	.6038	.8862	.24	1.00
	1	10	.8240	.18198	.05755	.6938	.9542	.41	1.00
	Total	22	.7809	.20417	.04353	.6904	.8714	.24	1.00
arousal1	0	12	.5125	.41517	.11985	.2487	.7763	.00	1.00
	1	10	.5250	.41540	.13136	.2278	.8222	.00	.99
	Total	22	.5182	.40532	.08641	.3385	.6979	.00	1.00
dominance1	0	12	.7942	.22857	.06598	.6489	.9394	.35	1.00
	1	10	.7290	.17854	.05646	.6013	.8567	.44	.99
	Total	22	.7645	.20526	.04376	.6735	.8556	.35	1.00
valence2	0	12	.7075	.19813	.05720	.5816	.8334	.43	1.00
	1	10	.8050	.16406	.05188	.6876	.9224	.57	1.00
	Total	22	.7518	.18592	.03964	.6694	.8343	.43	1.00
arousal2	0	12	.4525	.41070	.11856	.1916	.7134	.00	.98
	1	10	.4040	.46445	.14687	.0718	.7362	.00	1.00
	Total	22	.4305	.42593	.09081	.2416	.6193	.00	1.00
dominance2	0	12	.6567	.30285	.08742	.4642	.8491	.06	1.00
	1	10	.6890	.13755	.04350	.5906	.7874	.47	.96
	Total	22	.6714	.23753	.05064	.5660	.7767	.06	1.00
valence3	0	12	.6758	.24670	.07122	.5191	.8326	.18	1.00
	1	10	.8230	.16580	.05243	.7044	.9416	.51	1.00
	Total	22	.7427	.22201	.04733	.6443	.8412	.18	1.00
arousal3	0	12	.5775	.39493	.11401	.3266	.8284	.00	1.00
	1	10	.6280	.37169	.11754	.3621	.8939	.00	1.00
	Total	22	.6005	.37625	.08022	.4336	.7673	.00	1.00
dominance3	0	12	.7042	.31466	.09083	.5042	.9041	.11	1.00
	1	10	.6880	.23275	.07360	.5215	.8545	.24	.98
	Total	22	.6968	.27413	.05844	.5753	.8184	.11	1.00
valence4	0	12	.7342	.16763	.04839	.6277	.8407	.49	1.00
	1	10	.8800	.08055	.02547	.8224	.9376	.73	.99
	Total	22	.8005	.15174	.03235	.7332	.8677	.49	1.00
arousal4	0	12	.4208	.39846	.11503	.1677	.6740	.00	1.00
	1	10	.6700	.28554	.09030	.4657	.8743	.10	.99
	Total	22	.5341	.36638	.07811	.3716	.6965	.00	1.00
dominance4	0	12	.7117	.26013	.07509	.5464	.8769	.35	1.00
	1	10	.8490	.09085	.02873	.7840	.9140	.69	1.00
	Total	22	.7741	.20948	.04466	.6812	.8670	.35	1.00
valencedelta	0	12	-.024166666666667	.244148179049800	.070479508448280	-.179291018851883	.130957685518550	-.510000000000000	.450000000000000
	1	10	.116000000000000	.182525492892241	.057719628858436	-.014570871856164	.246570871856164	-.260000000000000	.340000000000000
	Total	22	.039545454545455	.224954540862249	.047960469259279	-.060193801325774	.139284710416683	-.510000000000000	.450000000000000
arousaldelta	0	12	.005000000000000	.532003075862605	.153576059529492	-.333018627969762	.343018627969762	-.990000000000000	.950000000000000
	1	10	.328000000000000	.464585837063508	.146914941377656	-.004344686959542	.660344686959542	-.600000000000000	.850000000000000
	Total	22	.151818181818182	.517545823723542	.110341140364714	-.077648781327306	.381285144963670	-.990000000000000	.950000000000000
dominancedelta	0	12	.025833333333333	.276979104342592	.079956980226048	-.150150793589940	.201817460256606	-.630000000000000	.310000000000050
	1	10	.114000000000000	.170632809141605	.053958832044027	-.008063358404622	.236063358404622	-.260000000000000	.289999999999900
	Total	22	.065909090909091	.233842826557642	.049855458136710	-.037771010067248	.169589191885429	-.630000000000000	.310000000000050

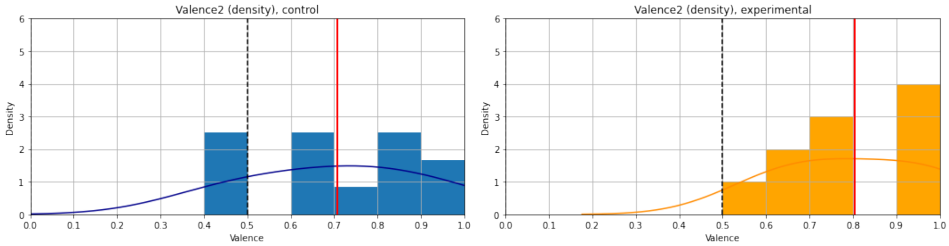
Figure F.1: Descriptive statistics table from SPSS 26 showing the statistics for the emotional states. 0 is used to denote the control group and 1 for the experimental group. 0 on the emotional state dimensions is before listening to the playlist, 4 is after the last song. The deltas are the difference between the last and first value for a participant.



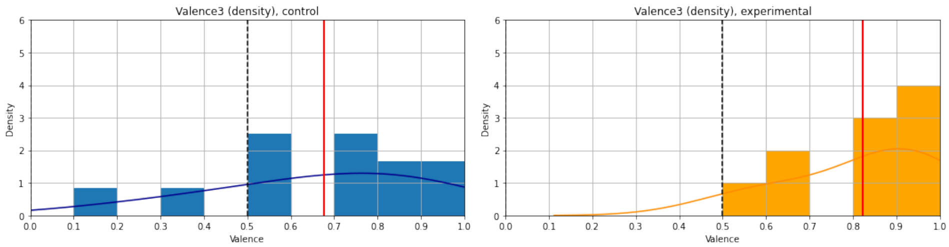
(a) begin valence, before playlist



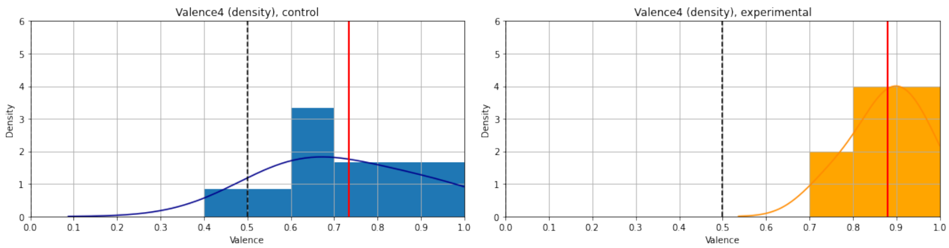
(b) valence after 1st song



(c) valence after 2nd song

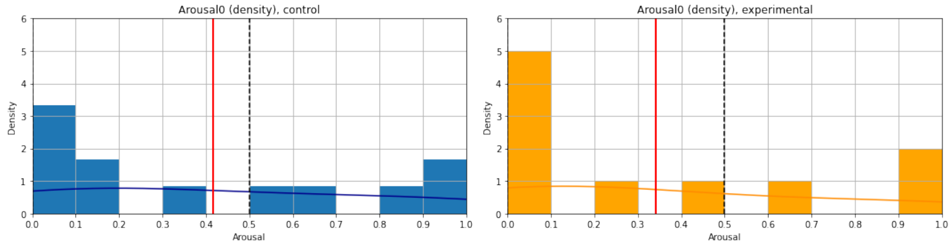


(d) valence after 3rd song

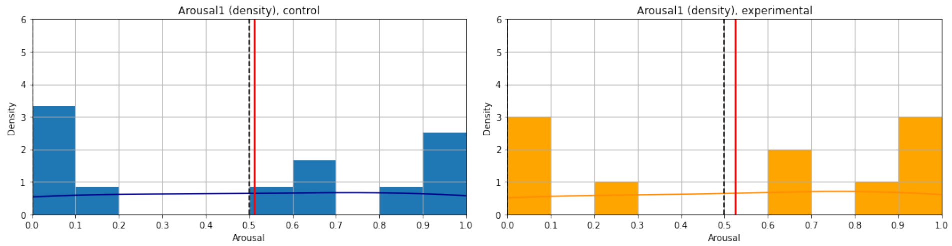


(e) Final valence, after 4th song

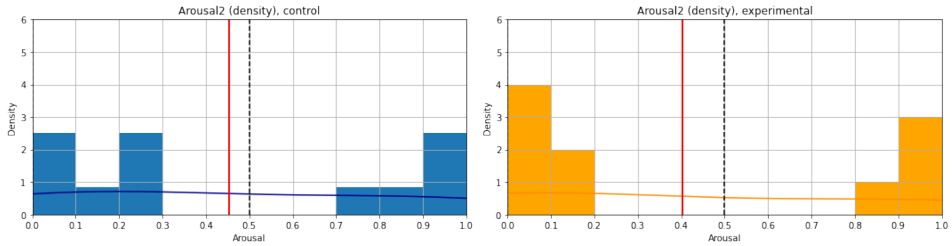
Figure F.2: Kernel density estimations (Gaussian kernel), bandwidth=0.8 of each timestep, divided by group. The KDE is plotted on top of a histogram with 10 bins with the same data. The sample mean is shown in red



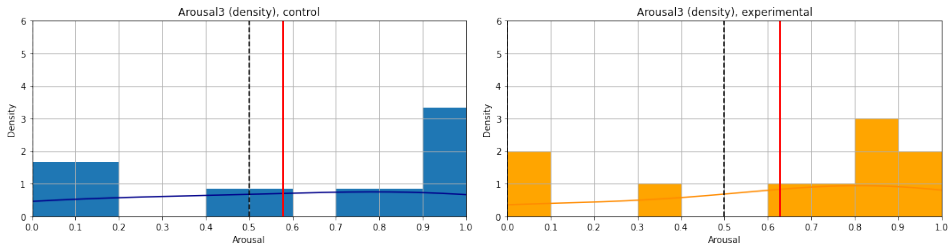
(a) begin arousal, before playlist



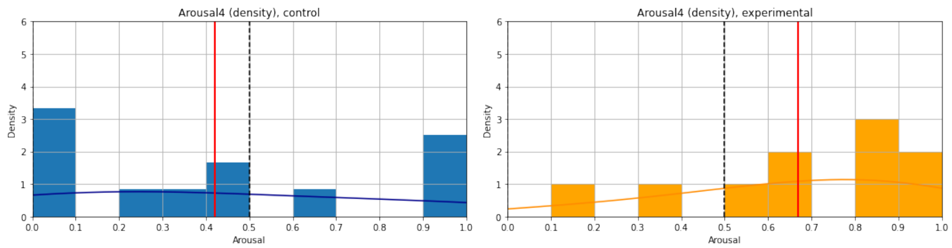
(b) arousal after 1st song



(c) arousal after 2nd song

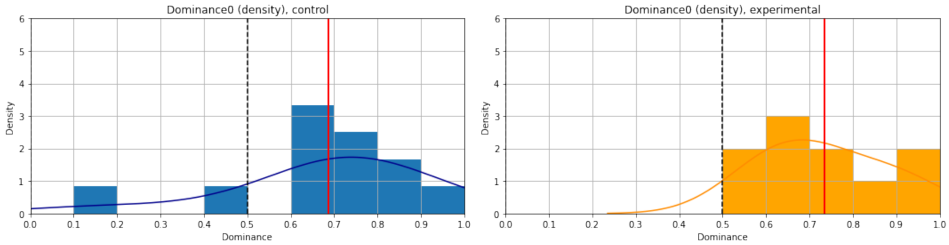


(d) arousal after 3rd song

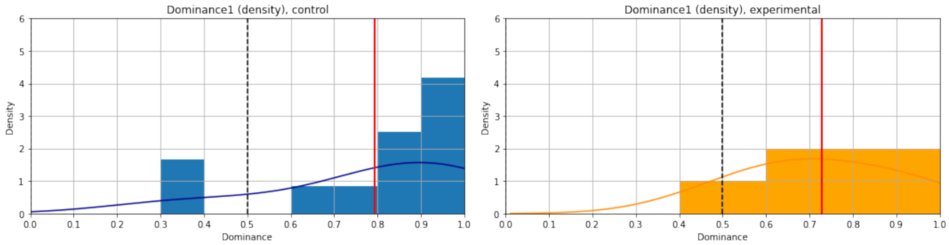


(e) Final arousal, after 4th song

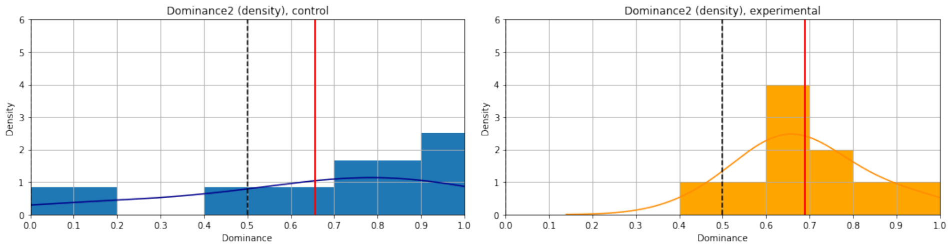
Figure F.3: Kernel density estimations (guassian kernel, bandwidth=0.8 of arousal of each timestep, divided by group. The KDE is plotted on top of a histogram with 20 bins with the same data. The sample mean is shown in red.



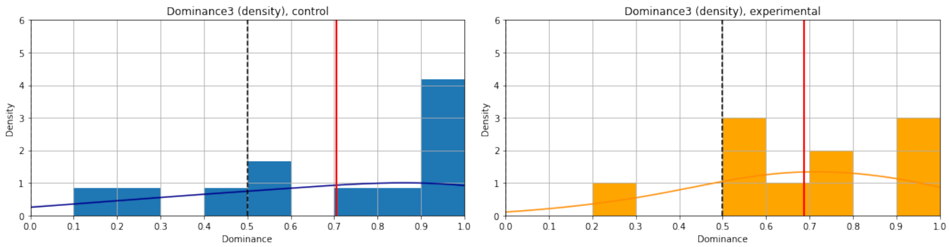
(a) begin dominance, before playlist



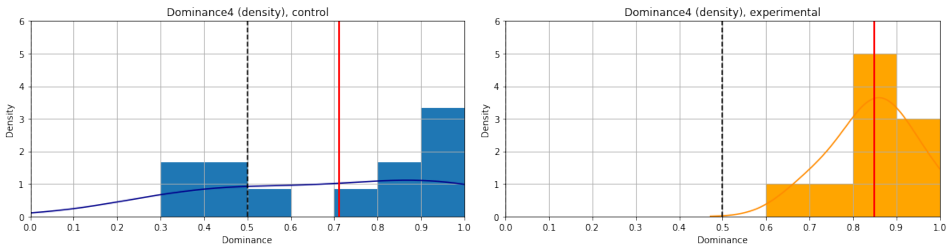
(b) Dominance after 1st song



(c) Dominance after 2nd song

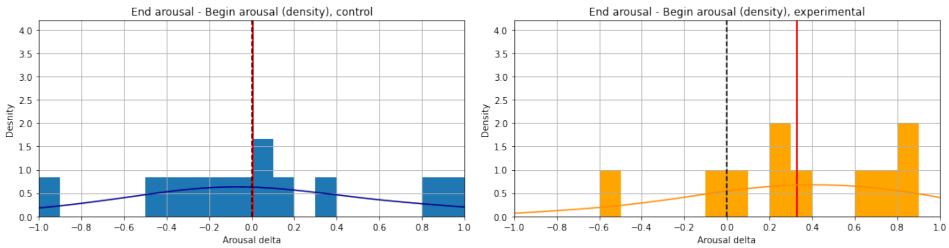


(d) Dominance after 3rd song

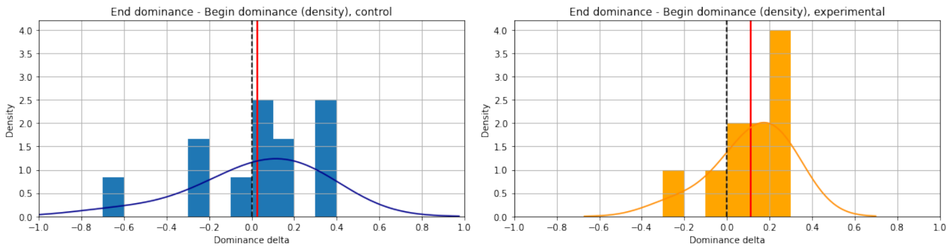


(e) Final dominance, after 4th song

Figure F.4: Kernel density estimations (guassian kernel, bandwidth=0.8 of dominance at each timestep, divided by group. The KDE is plotted on top of a histogram with 20 bins with the same data. The sample mean is shown in red.

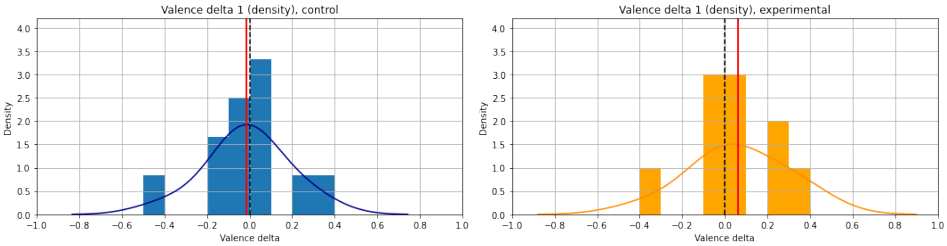


(a) Arousal delta

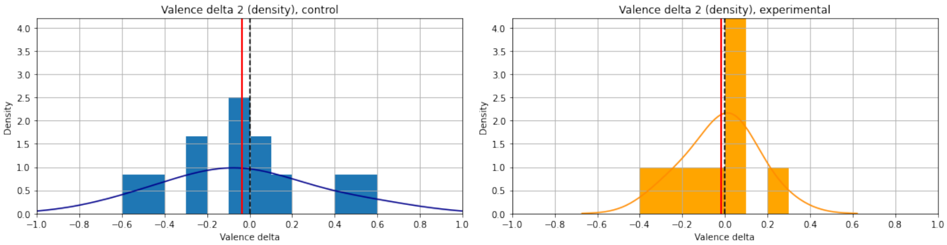


(b) Dominance delta

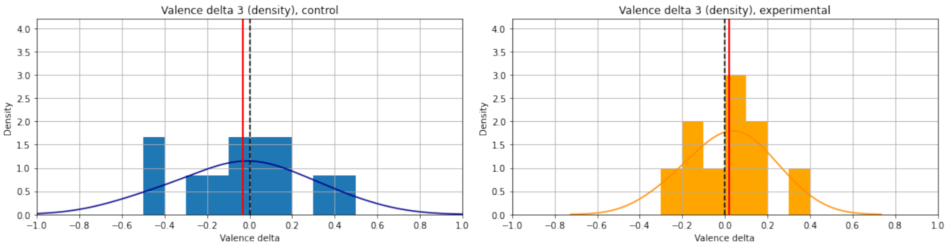
Figure F.5: Kernel density estimations (guassian kernel, bandwidth=0.8 of begin and final dominance, divided by group. The KDE is plotted on top of a histogram with 20 bins with the same data. The sample mean is shown in red.



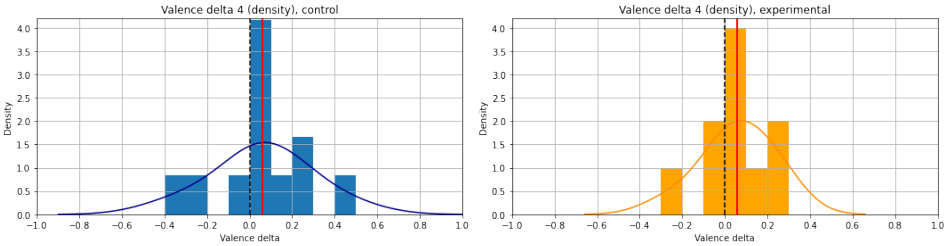
(a) Delta valence after 1st song



(b) Delta valence after 2nd song

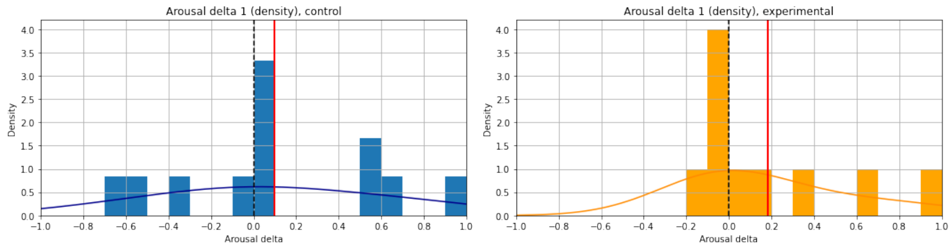


(c) Delta valence after 3rd song

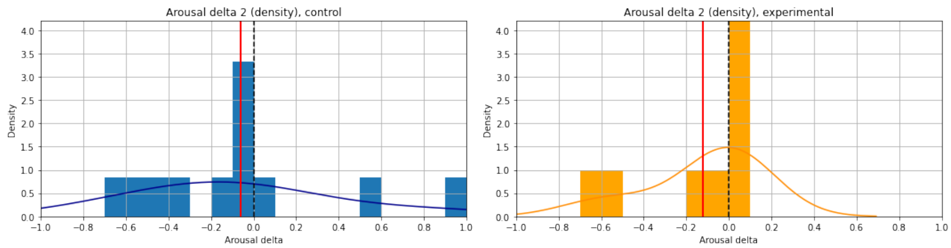


(d) Delta valence after 4th song

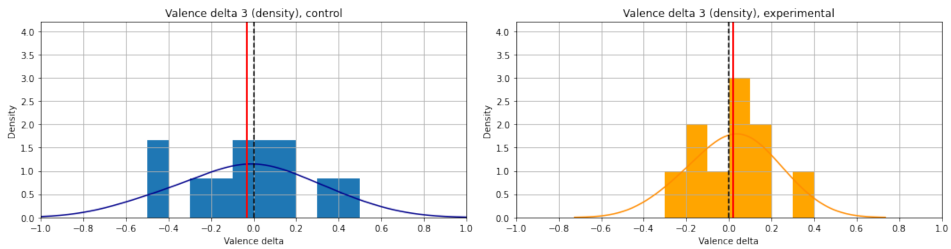
Figure F.6: Boxplot of the delta in valence.



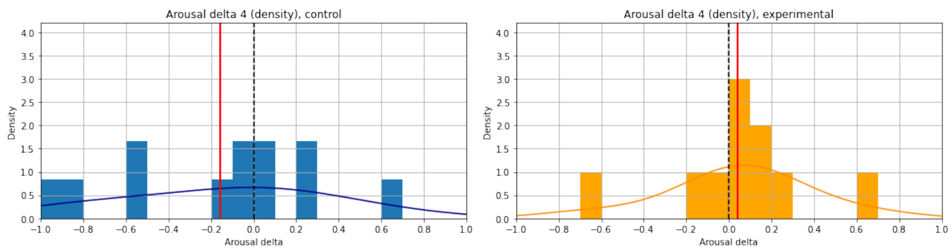
(a) Delta arousal after 1st song



(b) Delta arousal after 2nd song

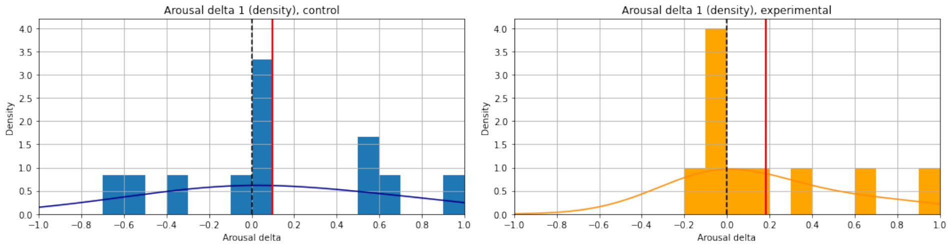


(c) Delta arousal after 3rd song

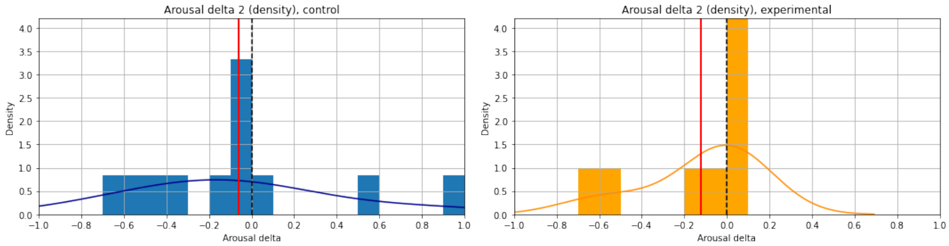


(d) Delta arousal after 4th song

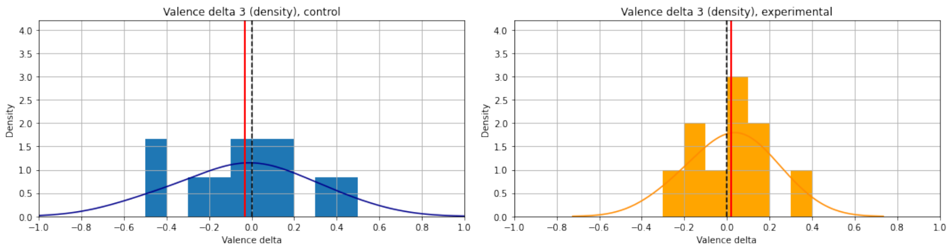
Figure F.7: Kernel density estimations (guassian kernel, bandwidth=0.8 of each timestep, divided by group. The KDE is plotted on top of a histogram with 20 bins with the same data. The sample mean is shown in red. We look at the delta between the AffectButton arousal before and after listening to the song at timestep i.



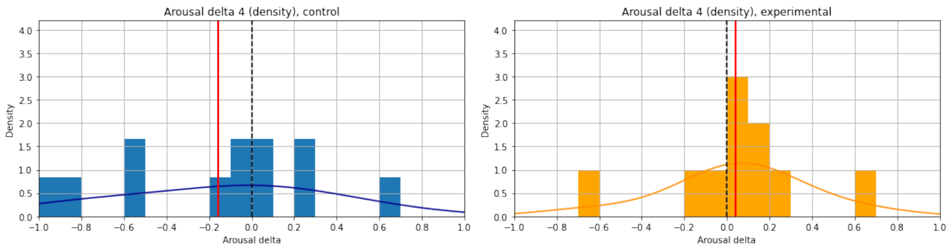
(a) Delta dominance after 1st song



(b) Delta dominance after 2nd song



(c) Delta dominance after 3rd song



(d) Delta dominance after 4th song

Figure F.8: Kernel density estimations (guassian kernel, bandwidth=0.8 of each timestep, divided by group. The KDE is plotted on top of a histogram with 20 bins with the same data. The sample mean is shown in red. We look at the delta between the AffectButton dominance before and after listening to the song at timestep i.

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
selfImprovementPlaylist	0	12	2.83	.835	.241	2.30	3.36	1	4
	1	10	2.90	.738	.233	2.37	3.43	2	4
	Total	22	2.86	.774	.165	2.52	3.21	1	4
playlistRating	0	12	2.67	1.155	.333	1.93	3.40	1	4
	1	10	2.90	.738	.233	2.37	3.43	2	4
	Total	22	2.77	.973	.207	2.34	3.20	1	4

Figure F.9: Descriptive statistics table from SPSS 26 showing the statistics for the final questions. 0 is used to denote the control group and 1 for the experimental group. It contains the playlist rating and the self reported emotional change.

Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
valence0	Welch	.007	1	19.938	.935
arousal0	Welch	.191	1	19.174	.667
dominance0	Welch	.423	1	18.730	.523
valence1	Welch	.840	1	19.999	.370
arousal1	Welch	.005	1	19.286	.945
dominance1	Welch	.563	1	19.940	.462
valence2	Welch	1.594	1	20.000	.221
arousal2	Welch	.066	1	18.221	.800
dominance2	Welch	.110	1	15.927	.745
valence3	Welch	2.769	1	19.245	.112
arousal3	Welch	.095	1	19.662	.761
dominance3	Welch	.019	1	19.770	.891
valence4	Welch	7.112	1	16.402	.017
arousal4	Welch	2.903	1	19.625	.104
dominance4	Welch	2.918	1	14.087	.110
valencedelta	Welch	2.367	1	19.811	.140
arousaldelta	Welch	2.310	1	19.938	.144
dominancedelta	Welch	.835	1	18.589	.372

a. Asymptotically F distributed.

Figure F.10: Output of Welch t-test in SPSS 26. 0 is used to denote the control group and 1 for the experimental group. 0 on the emotional state dimensions is before listening to the playlist, 4 is after the last song. The deltas are the difference between the last and first value for a participant.

	valence0	arousal0	dominance0	valence1	arousal1	dominance1	valence2	arousal2	dominance2	valence3	arousal3	dominance3	valence4	arousal4	dominance4	valencedelta	arousaldelta	dominancedelta
valence0	1.000																	
arousal0	.713**	1.000																
dominance0	.045	.309	1.000															
valence1	.418	.465*	.075	1.000														
arousal1	.208	.316	.204	.738**	1.000													
dominance1	.201	.337	.367	.465*	.752**	1.000												
valence2	-.036	.044	-.045	.227	.195	-.055	1.000											
arousal2	.211	.322	.189	.410	.607**	.329	.682**	1.000										
dominance2	-.284	-.155	.061	-.245	.074	.231	.533*	.460*	1.000									
valence3	.306	.071	-.140	.409	.184	-.124	.391	.207	-.153	1.000								
arousal3	.161	.280	-.126	.389	.521*	.276	.319	.515*	.117	.590**	1.000							
dominance3	-.003	.145	.201	.317	.526*	.518*	.232	.487*	.461*	.224	.622**	1.000						
valence4	-.008	-.136	-.062	.373	.086	.077	.259	-.080	.072	.520*	.182	.338	1.000					
arousal4	-.227	-.035	.169	.299	.312	.338	.298	.260	.278	.241	.335	.485*	.724**	1.000				
dominance4	-.324	-.106	.153	.022	.107	.147	.114	.118	.096	.012	.097	.154	.379	.783**	1.000			
valencedelta	-.750**	-.727**	-.021	-.139	-.113	-.103	.131	-.242	.257	.159	-.012	.234	.596**	.548**	.428*	1.000		
arousaldelta	-.687**	-.728**	-.017	-.222	-.082	.048	.038	-.179	.301	.003	-.094	.196	.515*	.645**	.606**	.892**	1.000	
dominancedelta	-.332	-.307	-.580**	.079	.016	-.162	-.016	-.132	-.187	.172	.198	-.009	.306	.429*	.571**	.368	.449*	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure F.11: Spearman correlations for the emotional states. 0 on the emotional state dimensions is before listening to the playlist, 4 is after the last song. The deltas are the difference between the last and first value for a participant.

Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
beginValence	Welch	1.360	1	73.996	.247
beginArousal	Welch	.935	1	66.079	.337
beginDominance	Welch	2.033	1	85.076	.158
endValence	Welch	.560	1	70.257	.457
endArousal	Welch	.344	1	66.511	.559
endDominance	Welch	.000	1	83.468	.996
valenceDelta	Welch	.116	1	67.881	.735
arousalDelta	Welch	.153	1	77.470	.697
dominanceDelta	Welch	1.401	1	85.552	.240

a. Asymptotically F distributed.

Figure F.12: Welch t test on songs, divided by whether or not they have a memory associated to them. It tests the value of each dimension before and after the song. As well as the delta between these two for each dimension.

	selfImprovement	qualityAnswer	preferenceAnswer	familiarityAnswer	valence	arousal	beginValence	beginArousal	beginDominar	endValence	endArousal	endDominance	valenceDelta	arousalDelta	dominanceDelt
selfImprovementSong	1.000														
qualityAnswer	-.058	1.000													
preferenceAnswer	.552**	-.018	1.000												
familiarityAnswer	.314**	.005	.547**	1.000											
valence	.025	-.121	-.067	-.059	1.000										
arousal	.024	-.074	.139	.385**	.036	1.000									
beginValence	.014	-.318**	.263*	.296**	.277**	.321**	1.000								
beginArousal	.047	-.283**	.273**	.296**	.101	.454**	.708**	1.000							
beginDominance	-.119	-.078	.046	.179	-.065	.219*	.370**	.553**	1.000						
endValence	.512**	-.092	.472**	.349**	.296**	.359**	.412**	.312**	.045	1.000					
endArousal	.522**	-.159	.439**	.391**	.192	.414**	.293**	.457**	.278**	.690**	1.000				
endDominance	.623**	.037	.448**	.393**	.017	.295**	.069	.276**	.288**	.398**	.651**	1.000			
valenceDelta	.407**	.248*	.171	.024	.011	-.012	-.535**	-.329**	-.238*	.481**	.301**	.286**	1.000		
arousalDelta	.475**	.145	.137	.073	.111	-.040	-.398**	-.510**	-.249*	.338**	.464**	.349**	.630**	1.000	
dominanceDelta	.553**	.155	.319**	.150	-.004	.085	-.207	-.145	-.545**	.245*	.246*	.561**	.375**	.376**	1.00

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure F.13: Spearman correlations for the emotional states and questions during songs. 0 on the emotional state dimensions is before listening to the playlist, 4 is after the last song. The deltas are the difference between the last and first value for a participant. emotion-Likert is the self reported emotion change.

	listeningFreqency	emotionEndForm	playlistRating
listeningFreqency	1.000	-.044	.068
emotionEndForm	-.044	1.000	.170
playlistRating	.068	.170	1.000
valence0	.371	-.354	.130
arousal0	.323	-.442*	.141
dominance0	-.177	-.049	-.102
valence1	.242	-.089	.420
arousal1	-.040	.064	.385
dominance1	-.145	.061	.035
valence2	.330	.231	-.011
arousal2	.169	.186	.042
dominance2	-.024	.331	-.303
valence3	.234	.326	.574**
arousal3	.081	.391	.505*
dominance3	-.073	.479*	.482*
valence4	.121	.327	.373
arousal4	.024	.445*	.218
dominance4	-.153	.329	.027
valencedelta	-.225	.540**	.126
arousaldelta	-.209	.582**	-.028
dominancedelta	-.048	.328	.291
*. Correlation is significant at the 0.05 level (2-tailed).			
*. Correlation is significant at the 0.05 level (2-tailed).			

Figure F.14: Spearman correlations for the questions and emotional states. 0 on the emotional state dimensions is before listening to the playlist, 4 is after the last song. The deltas are the difference between the last and first value for a participant.

G. Frontend Walkthrough

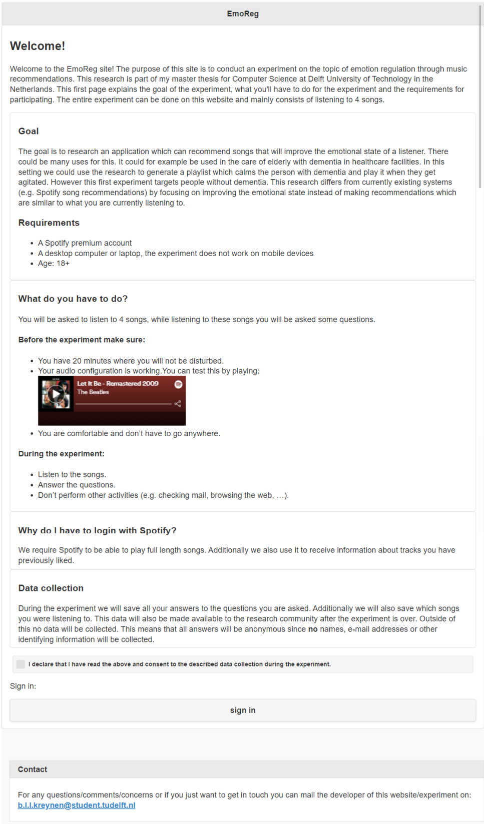


Figure G.1: The Introduction

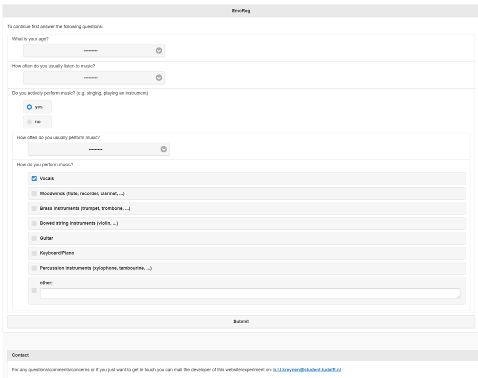


Figure G.2: The background questions

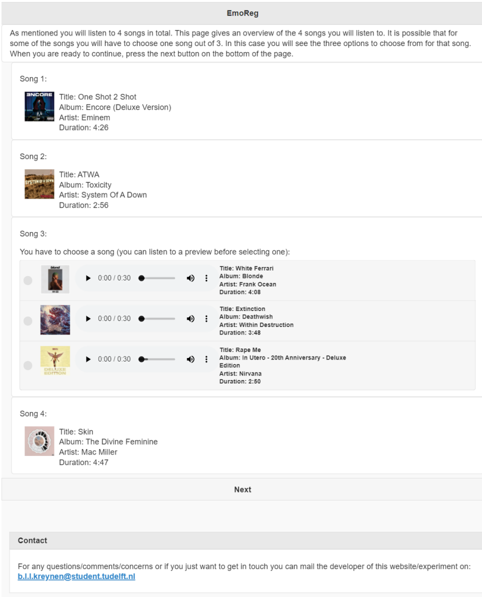


Figure G.3: Presenting the playlist and their song choices to the participant

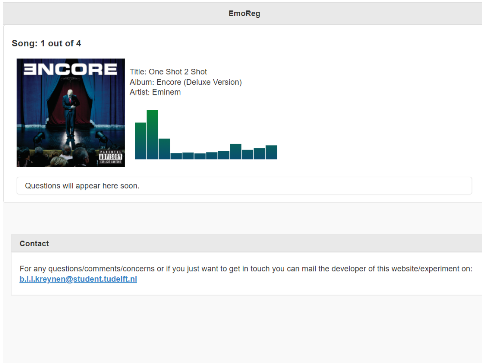
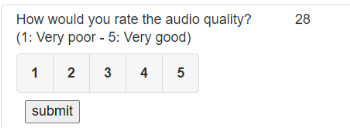
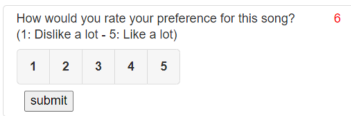


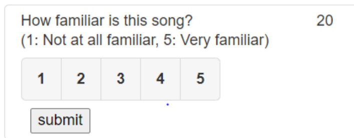
Figure G.4: The music player before the participant receives a questions



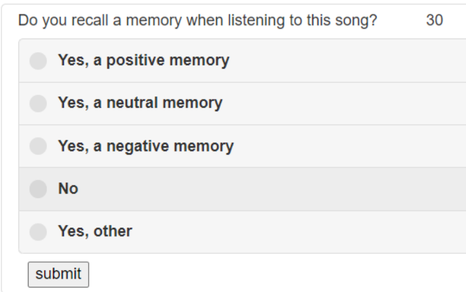
(a) First question, audio quality



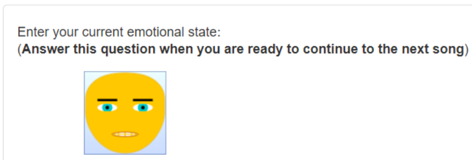
(b) Second question, preference screen



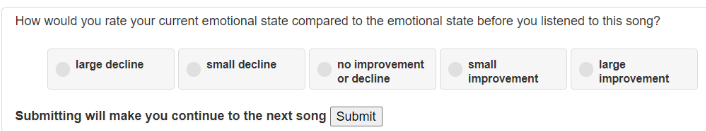
(c) Third question, familiarity screen



(d) Fourth question, memory screen



(e) Fifth question, affect button screen



(f) Sixth question, emotion improvement screen

Figure G.5: The questions posed in the player.

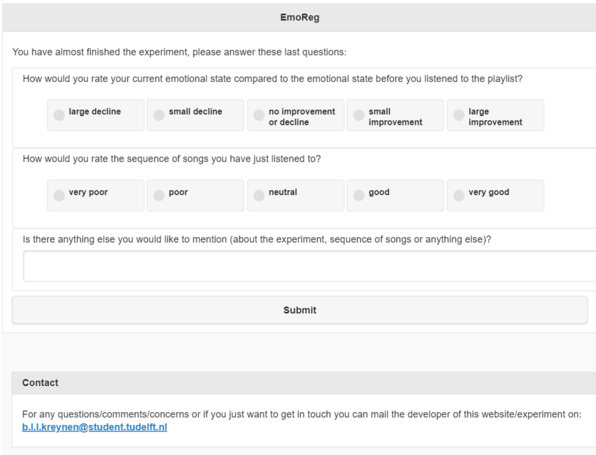


Figure G.6: The final questions asked after listening to the full playlist

Bibliography

[1] Socio-Cognitive Engineering | Socio-Cognitive Engineering. URL `https://scetool.ewi.tudelft.nl/`.

[2] Ashu Abdul, Jenhui Chen, Hua-Yuan Liao, and Shun-Hao Chang. An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach. *Applied Sciences*, 8(7): 1103, 7 2018. ISSN 2076-3417. doi: 10.3390/app8071103. URL `http://www.mdpi.com/2076-3417/8/7/1103`.

[3] Iosief Abraha, Joseph M. Rimland, Fabiana Mirella Trotta, Giuseppina Dell'Aquila, Alfonso Cruz-Jentoft, Mirko Petrovic, Adalsteinn Gudmundsson, Roy Soiza, Denis O'Mahony, Antonio Guaita, and Antonio Cherubini. Systematic review of systematic reviews of non-pharmacological interventions to treat behavioural disturbances in older patients with dementia. the SENATOR-OnTop series. *BMJ Open*, 7(3): e012759, 3 2017. ISSN 20446055. doi: 10.1136/bmjopen-2016-012759.

[4] Andra Adams and Peter Robinson. Automated recognition of complex categorical emotions from facial expressions and head motions. In *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, pages 355–361. Institute of Electrical and Electronics Engineers Inc., 12 2015. ISBN 9781479999538. doi: 10.1109/ACII.2015.7344595.

[5] Anna Aljanaki, Frans Wiering, and Remco C. Veltkamp. Studying emotion induced by music through a crowdsourcing game. *Information Processing and Management*, 52(1):115–128, 1 2016. ISSN 03064573. doi: 10.1016/j.ipm.2015.03.004.

[6] Anna Aljanaki, Yi-Hsuan Yang, and Mohammad Soleymani. Developing a benchmark for emotional analysis of music. *PLOS ONE*, 12(3):e0173392, 3 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0173392. URL `https://dx.plos.org/10.1371/journal.pone.0173392`.

[7] American Psychiatric Association and others. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[8] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey, 11 2017. ISSN 10535888.

[9] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak. Emotion Based Music Recommendation System Using Wearable Physiological Sensors. *IEEE Transactions on Consumer Electronics*, 64(2):196–203, 5 2018. ISSN 00983063. doi: 10.1109/TCE.2018.2844736.

[10] Amee Baird and Séverine Samson. Music and dementia. In *Progress in Brain Research*, volume 217, pages 207–235. Elsevier, 2015. doi: 10.1016/bs.pbr.2014.11.028.

[11] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 7 2013. ISSN 09507051. doi: 10.1016/j.knosys.2013.03.012.

[12] Danny Oude Bos. EEG-based Emotion Recognition The Influence of Visual and Auditory Stimuli. Technical report, University of Twente, 2006.

[13] Joost Broekens and Willem Paul Brinkman. AffectButton: A method for reliable and valid affective self-report. *International Journal of Human Computer Studies*, 71(6):641–667, 6 2013. ISSN 10715819. doi: 10.1016/j.ijhcs.2013.02.003.

[14] Lucian Buşoniu, Robert Babuška, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning, 3 2008. ISSN 10946977.

[15] Erion Çano and Maurizio Morisio. MUSIC MOOD DATASET CREATION BASED ON LAST.FM TAGS. pages 15–26, 2017. doi: 10.5121/csit.2017.70603. URL http://www.music-ir.org/mirex/wiki/2007:Audio_Music_Mood_Classification.

[16] Òscar Celma and Xavier Serra. FOAFing the music: Bridging the semantic gap in music recommendation. *Web Semantics*, 6(4):250–256, 11 2008. ISSN 15708268. doi: 10.1016/j.websem.2008.09.004.

[17] J. Cerejeira, L. Lagarto, and E. B. Mukaetova-Ladinska. Behavioral and psychological symptoms of dementia. *Frontiers in Neurology*, MAY:73, 5 2012. ISSN 16642295. doi: 10.3389/fneur.2012.00073.

[18] Jia-Wei Chang, Ching-Yi Chiou, Jia-Yi Liao, Ying-Kai Hung, Chien-Che Huang, Kuan-Cheng Lin, and Ying-Hung Pu. Music recommender using deep embedding-based features and behavior-based reinforcement learning. 2019. doi: 10.1007/s11042-019-08356-9. URL https://doi.org/10.1007/s11042-019-08356-9.

[19] Yu An Chen, Yi Hsuan Yang, Ju Chiang Wang, and Homer Chen. The AMG1608 dataset for music emotion recognition. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 2015-Augus, pages 693–697. Institute of Electrical and Electronics Engineers Inc., 8 2015. ISBN 9781467369978. doi: 10.1109/ICASSP.2015.7178058.

[20] Sheung Tak Cheng. Dementia Caregiver Burden: a Research Update and Critical Analysis. *Current Psychiatry Reports*, 19(9):1–8, 9 2017. ISSN 15351645. doi: 10.1007/s11920-017-0818-2.

[21] Chung Yi Chi, Richard Tzong Han Tsai, Jeng You Lai, and Jane Yung Jen Hsu. A reinforcement learning approach to emotion-based automatic playlist generation. In *Proceedings - International Conference on Technologies and Applications of Artificial Intelligence, TAAI 2010*, pages 60–65, 2010. ISBN 9780769542539. doi: 10.1109/TAAI.2010.21.

[22] Sungwoon Choi, Heonseok Ha, Uiwon Hwang, Chanju Kim, Jung-Woo Ha, and Sungroh Yoon. Reinforcement Learning based Recommender System using Biclustering Technique. *arXiv*, 1 2018. URL http://arxiv.org/abs/1801.05532.

[23] Jiska Cohen-Mansfield, Marcia S Marx, Maha Dakheel-Ali, and Khin Thein. The use and utility of specific nonpharmacological interventions for behavioral symptoms in dementia: An exploratory study. *American Journal of Geriatric Psychiatry*, 23(2):160–170, 2015. ISSN 15457214. doi: 10.1016/j.jagp.2014.06.006.

[24] Claudia Cooper, T. B.S. Balamurali, and Gill Livingston. A systematic review of the prevalence and covariates of anxiety in caregivers of people with dementia. *International Psychogeriatrics*, 19(2):175–195, 4 2007. ISSN 10416102. doi: 10.1017/S1041610206004297.

[25] Weihui Dai, Dongmei Han, Yonghui Dai, and Dongrong Xu. Emotion recognition and affective computing on vocal social media. *Information and Management*, 52(7):777–788, 11 2015. ISSN 03787206. doi: 10.1016/j.im.2015.02.003.

[26] James J. Deng and Clement Leung. Emotion-based music recommendation using audio features and user playlist - IEEE Conference Publication. In *2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012)*, Taipei, Taiwan, 2012. IEEE. URL https://ieeexplore.ieee.org/abstract/document/6528741.

[27] Shuiguang Deng, Dongjing Wang, Xitong Li, and Guandong Xu. Exploring user emotion in microblogs for music recommendation. *Expert Systems with Applications*, 42(23):9284–9293, 12 2015. ISSN 09574174. doi: 10.1016/j.eswa.2015.08.029.

[28] Django Software Foundation. Django: The web framework for perfectionists with deadlines. URL https://www.djangoproject.com/.

[29] Bernd Dudzik, Hayley Hung, Mark Neerincx, and Joost Broekens. Investigating the Influence of Personal Memories on Video-Induced Emotions. In *UMAP 2020 - Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 53–61. Association for Computing Machinery, Inc, 7 2020. ISBN 9781450368612. doi: 10.1145/3340631.3394842.

[30] Damien Dupré, Eva G. Krumhuber, Dennis Küster, and Gary J. McKeown. A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLoS ONE*, 15(4): e0231968, 4 2020. ISSN 19326203. doi: 10.1371/journal.pone.0231968. URL `https://doi.org/10.1371/journal.pone.0231968`.

[31] Tuomas Eerola and Jonna K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39(1):18–49, 1 2011. ISSN 0305-7356. doi: 10.1177/0305735610362821. URL `http://journals.sagepub.com/doi/10.1177/0305735610362821`.

[32] Minah Amor Gaviola, Kerry J. Inder, Sophie Dilworth, Elizabeth G. Holliday, and Isabel Higgins. Impact of individualised music listening intervention on persons with dementia: A systematic review of randomised controlled trials. *Australasian Journal on Ageing*, 39(1):10–20, 3 2019. ISSN 17416612. doi: 10.1111/ajag.12642. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ajag.12642`.

[33] Linda A Gerdner. Effects of individualized versus classical 'relaxation' music on the frequency of agitation in elderly persons with Alzheimer's disease and related disorders. *International Psychogeriatrics*, 12(1):49–65, 2000. ISSN 10416102. doi: 10.1017/S1041610200006190.

[34] Linda A Gerdner. Use of individualized music by trained staff and family: translating research into practice. *Journal of gerontological nursing*, 31(6):22–30, 2005. ISSN 00989134. doi: 10.3928/0098-9134-20050601-08.

[35] Linda A Gerdner. Individualized music for dementia: Evolution and application of evidence-based protocol. *World Journal of Psychiatry*, 2(2):26, 2012. ISSN 2220-3206. doi: 10.5498/wjp.v2.i2.26.

[36] Linda A Gerdner. Evidence-Based Guideline: Individualized Music for Persons with Dementia. *Journal of gerontological nursing*, 36(6):7–15, 2013.

[37] Shlok Gilda, Husain Zafar, Chintan Soni, and Kshitija Waghurdekar. Smart music player integrating facial emotion recognition and music mood recommendation. pages 154–158. Institute of Electrical and Electronics Engineers (IEEE), 2 2018. doi: 10.1109/wispnet.2017.8299738.

[38] Hatice Gunes, Bjöorn Schuller, Maja Pantic, and Roddy Cowie. Emotion representation, analysis and synthesis in continuous space: A survey. In *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pages 827–834, 2011. ISBN 9781424491407. doi: 10.1109/FG.2011.5771357.

[39] Byeong-Jun Han, Seungmin Rho, Roger B Dannenberg, and Eenjun Hwang. SMERS: MUSIC EMOTION RECOGNITION USING SUPPORT VECTOR REGRESSION. Technical report, 2009.

[40] Byeong-Jun Han, Seungmin Rho, Eenjun Hwang, Sanghoon Jun, B.-J Han, S Jun, E Hwang, and S Rho. Music emotion classification and context-based music recommendation Music Emotion Recognition View project Environmental Sound Classification View project Music emotion classification and context-based music recommendation. *Multimed Tools and Applications*, 47:433–460, 2010. doi: 10.1007/s11042-009-0332-6. URL `https://www.researchgate.net/publication/229031003`.

[41] Nathan Herrmann, Krista L. Lanctôt, Robert Sambrook, Nadia Lesnikova, Réjean Hébert, Peter McCracken, Alain Robillard, and Elizabeth Nguyen. The contribution of neuropsychiatric symptoms to the cost of dementia care. *International Journal of Geriatric Psychiatry*, 21(10):972–976, 10 2006. ISSN 0885-6230. doi: 10.1002/gps.1594. URL `http://doi.wiley.com/10.1002/gps.1594`.

[42] Binbin Hu, Chuan Shi, and Jian Liu. Playlist Recommendation Based on Reinforcement Learning. 2017. doi: 10.1007/978-3-319-68121-4{_}18. URL `https://www.researchgate.net/publication/320053683`.

[43] Arefin Huq, Juan Pablo Bello, and Robert Rowe. Automated Music Emotion Recognition: A Systematic Evaluation. *Journal of New Music Research*, 39(3):227–244, 9 2010. ISSN 0929-8215. doi: 10.1080/09298215.2010.513733. URL `http://www.tandfonline.com/doi/abs/10.1080/09298215.2010.513733`.

[44] Christian Martyn Jones and Tommy Troen. Biometric valence and arousal recognition. In *Australasian Computer-Human Interaction Conference, OZCHI'07*, pages 191–194, New York, New York, USA, 2007. Association for Computing Machinery (ACM). ISBN 9781595938725. doi: 10.1145/1324892.1324929. URL http://portal.acm.org/citation.cfm?doid=1324892.1324929.

[45] Hafeez Kabani, Sharik Khan, Omar Khan, Shabana Tadvi, and Asst Professor. Emotion Based Music Player. *International Journal of Engineering Research and General Science*, 3(1), 2015. ISSN 2091-2730. URL www.ijergs.org.

[46] Helen C. Kales, Laura N. Gitlin, and Constantine G. Lyketsos. Assessment and management of behavioral and psychological symptoms of dementia. *BMJ (Online)*, 350, 3 2015. ISSN 17561833. doi: 10.1136/bmj.h369.

[47] Stella Karantzoulis and James E Galvin. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics*, 11(11):1579–1591, 11 2011. ISSN 1473-7175. doi: 10.1586/ern.11.155. URL http://www.tandfonline.com/doi/full/10.1586/ern.11.155.

[48] Byoung Chul Ko. A brief review of facial emotion recognition based on visual information. *Sensors (Switzerland)*, 18(2):401, 2 2018. ISSN 14248220. doi: 10.3390/s18020401. URL www.mdpi.com/journal/sensors.

[49] Jens Kober, J. Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *International Journal of Robotics Research*, 32(11):1238–1274, 9 2013. ISSN 02783649. doi: 10.1177/0278364913495721.

[50] Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. DEAP: A database for emotion analysis; Using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 1 2012. ISSN 19493045. doi: 10.1109/T-AFFC.2011.15.

[51] Elad Liebman, Maytal Saar-Tsechansky, and Peter Stone. DJ-MC: A Reinforcement-Learning Agent for Music Playlist Recommendation. Technical report, 2014.

[52] Hao Liu, Jun Hu, and Matthias Rauterberg. Music playlist recommendation based on user heartbeat and music preference. In *ICCTD 2009 - 2009 International Conference on Computer Technology and Development*, volume 1, pages 545–549, 2009. ISBN 9780769538921. doi: 10.1109/ICCTD.2009.246.

[53] Hao Liu, Jun Hu, and Matthias Rauterberg. Follow your heart: Heart rate controlled music recommendation for low stress air travel. *Interaction Studies*, 16(2):303–339, 1 2015. ISSN 1572-0373. doi: 10.1075/is.16.2.12liu.

[54] Ning-Han Liu, Cheng-Yu Chiang, and Hsiang-Ming Hsu. Improving Driver Alertness through Music Selection Using a Mobile EEG to Detect Brainwaves. *Sensors*, 13(7):8199–8221, 6 2013. ISSN 1424-8220. doi: 10.3390/s130708199. URL http://www.mdpi.com/1424-8220/13/7/8199.

[55] Gill Livingston, Lynsey Kelly, Elanor Lewis-Holmes, Gianluca Baio, Stephen Morris, Nishma Patel, Rumana Z Omar, Cornelius Katona, and Claudia Cooper. A systematic review of the clinical effectiveness and cost-effectiveness of sensory, psychological and behavioural interventions for managing agitation in older adults with dementia. *Health Technology Assessment*, 18(39):1, 2014. ISSN 20464924. doi: 10.3310/hta18390.

[56] A. Luneski, E. Konstantinidis, and P. D. Bamidis. Affective medicine: A review of affective computing efforts in medical informatics, 2010. ISSN 00261270.

[57] Constantine G. Lyketsos, Maria C. Carrillo, J. Michael Ryan, Ara S. Khachaturian, Paula Trzepacz, Joan Amatniek, Jesse Cedarbaum, Robert Brashear, and David S. Miller. Neuropsychiatric symptoms in Alzheimer's disease. *Alzheimer's and Dementia*, 7(5):532–539, 9 2011. ISSN 15525279. doi: 10.1016/j.jalz.2011.05.2410.

[58] Mahesh Babu Mariappan, Myunghoon Suk, and Balakrishnan Prabhakaran. FaceFetch: A user emotion driven multimedia content recommendation system based on facial expression recognition. In *Proceedings - 2012 IEEE International Symposium on Multimedia, ISM 2012*, pages 84–87, 2012. ISBN 9780769548753. doi: 10.1109/ISM.2012.24.

[59] Brais Martinez and Michel F. Valstar. Advances, challenges, and opportunities in automatic facial expression recognition. In *Advances in Face Detection and Facial Image Analysis*, pages 63–100. Springer International Publishing, 1 2016. ISBN 9783319259581. doi: 10.1007/978-3-319-25958-1{_}4. URL `https://link.springer.com/chapter/10.1007/978-3-319-25958-1_4`.

[60] Orii McDermott, Nadia Crellin, Hanne Mette Ridder, and Martin Orrell. Music therapy in dementia: A narrative synthesis systematic review, 2013. ISSN 08856230.

[61] Thibault Mura, J. F. Dartigues, and Claudine Berr. How many dementia cases in France and Europe? Alternative projections and scenarios 2010-2050. *European Journal of Neurology*, 17(2):252–259, 2010. ISSN 13515101. doi: 10.1111/j.1468-1331.2009.02783.x.

[62] Pauline Narme, Sylvain Clément, Nathalie Ehrlé, Loris Schiaratura, Sylvie Vachez, Bruno Courtaigne, Frédéric Munsch, and Séverine Samson. Efficacy of musical interventions in dementia: Evidence from a randomized controlled trial. *Journal of Alzheimer's Disease*, 38(2):359–369, 2014. ISSN 18758908. doi: 10.3233/JAD-130893.

[63] Mark A. Neerincx, Willeke van Vught, Olivier Blanson Henkemans, Elettra Oleari, Joost Broekens, Rifca Peters, Frank Kaptein, Yiannis Demiris, Bernd Kiefer, Diego Fumagalli, and Bert Bierman. Socio-Cognitive Engineering of a Robotic Partner for Child's Diabetes Self-Management. *Frontiers in Robotics and AI*, 6:118, 11 2019. ISSN 2296-9144. doi: 10.3389/frobt.2019.00118. URL `https://www.frontiersin.org/article/10.3389/frobt.2019.00118/full`.

[64] Shahriar Nirjon, Robert F. Dickerson, Qiang Li, Philip Asare, John A. Stankovic, Dezhi Hong, Ben Zhang, Xiaofan Jiang, Guobin Shen, and Feng Zhao. MusicalHeart: A hearty way of listening to music. In *SenSys 2012 - Proceedings of the 10th ACM Conference on Embedded Networked Sensor Systems*, pages 43–56, New York, New York, USA, 2012. ACM Press. ISBN 9781450311694. doi: 10.1145/2426656.2426662. URL `http://dl.acm.org/citation.cfm?doid=2426656.2426662`.

[65] Renato Panda and Rui Pedro Paiva. MUSIC EMOTION CLASSIFICATION: DATASET ACQUISITION AND COMPARATIVE ANALYSIS. In *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, York, UK, 2012. URL `http://www.music-ir.org/mirex/wiki/MIREX_HOME`.

[66] Renato Panda, Ricardo Manuel Malheiro, and Rui Pedro Paiva. Novel audio features for music emotion recognition. *IEEE Transactions on Affective Computing*, 3 2018. ISSN 19493045. doi: 10.1109/TAFFC.2018.2820691.

[67] Heeok Park. The effect of individualized music on agitation for home-dwelling persons with dementia. *Open Journal of Nursing*, 03(06):453–459, 10 2013. ISSN 2162-5336. doi: 10.4236/ojn.2013.36061. URL `http://www.scirp.org/journal/doi.aspx?DOI=10.4236/ojn.2013.36061`.

[68] Martin Prince, Renata Bryce, Emiliano Albanese, Anders Wimo, Wagner Ribeiro, and Cleusa P. Ferri. The global prevalence of dementia: A systematic review and metaanalysis, 1 2013. ISSN 15525279.

[69] Michael Rossato-Bennett (writer/director), Alexandra McDougald (producer), and Regina Scully (producer). Alive inside: A Story of Music and Memory, 2014.

[70] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6): 1161–1178, 12 1980. ISSN 00223514. doi: 10.1037/h0077714. URL `/record/1981-25062-001`.

[71] Mayumi Sakamoto, Hiroshi Ando, and Akimitsu Tsutou. Comparing the effects of different individualized music interventions for elderly individuals with severe dementia. *International Psychogeriatrics*, 25(5):775–784, 2013. ISSN 10416102. doi: 10.1017/S1041610212002256.

[72] Alba Sánchez, Ana Maseda, M Pilar Marante-Moar, Carmen De Labra, Laura Lorenzo-López, and José Carlos Millán-Calenti. Comparing the Effects of Multisensory Stimulation and Individualized Music Sessions on Elderly People with Severe Dementia: A Randomized Controlled Trial. *Journal of Alzheimer's Disease*, 52(1):303–315, 2016. ISSN 18758908. doi: 10.3233/JAD-151150.

[73] Teppo Särkämö, Sari Laitinen, Mari Tervaniemi, Ava Numminen, Merja Kurki, and Pekka Rantanen. Music, Emotion, and Dementia: Insight From Neuroscientific and Clinical Research. *Music & Medicine*, 4(3):153–162, 2012. ISSN 1943863X. doi: 10.1177/1943862112445323.

[74] Markus Schedl, Peter Knees, Brian McFee, Dmitry Bogdanov, and Marius Kaminskas. Music Recommender Systems. In *Recommender Systems Handbook*, pages 453–492. Springer US, 2015.

[75] Stefan Scherer, Sayan Ghosh, Eugene Laksana, and Louis-Philippe Morency. Representation Learning for Speech Emotion Recognition The Multi modal Assessment of Team Cohesion View project Representation Learning for Speech Emotion Recognition. 2016. doi: 10.21437/Interspeech.2016-692. URL `http://dx.doi.org/10.21437/Interspeech.2016-692`.

[76] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. Wearable-based affect recognition—a review, 10 2019. ISSN 14248220. URL `www.mdpi.com/journal/sensors`.

[77] Michal Schnaider Beeri, Perla Werner, Michael Davidson, and Shlomo Noy. The cost of behavioral and psychological symptoms of dementia (BPSD) in community dwelling Alzheimer's disease patients. *International Journal of Geriatric Psychiatry*, 17(5):403–408, 5 2002. ISSN 0885-6230. doi: 10.1002/gps.490. URL `http://doi.wiley.com/10.1002/gps.490`.

[78] Matthew D. Schulkind, Laura Kate Hennis, and David C. Rubin. Music, emotion, and autobiographical memory: They're playing your song. *Memory and Cognition*, 27(6):948–955, 3 1999. ISSN 0090502X. doi: 10.3758/BF03201225. URL `https://link.springer.com/article/10.3758/BF03201225`.

[79] Davide Seripa, Grazia D'Onofrio, Frisardi Vincenza, Solfrizzi Vincenzo, Mecocci Patrizia, and Pilotto Alberto. Neuropsychiatric Symptoms, Endophenotypes, and Syndromes in Late-Onset Alzheimer's Disease: Focus on APOE Gene. *International Journal of Alzheimer's Disease*, 3 2011. doi: 10.4061/2011/721457.

[80] Il Hyung Shin, Jaepyeong Cha, Gyeong Woo Cheon, Choonghee Lee, Seung Yup Lee, Hyung Jin Yoon, and Hee Chan Kim. Automatic stress-relieving music recommendation system based on photoplethysmography-derived heart rate variability analysis. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, pages 6402–6405. Institute of Electrical and Electronics Engineers Inc., 11 2014. ISBN 9781424479290. doi: 10.1109/EMBC.2014.6945093.

[81] Saim Shin, Dalwon Jang, Jongseol J. Lee, Sei Jin Jang, and Ji Hwan Kim. MyMusicShuffler: Mood-based music recommendation with the practical usage of brainwave signals. In *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, pages 355–356. Institute of Electrical and Electronics Engineers Inc., 2014. ISBN 9781479912919. doi: 10.1109/ICCE.2014.6776039.

[82] M.N. Shiota and J.W. Kalat. *Emotion*. Wadsworth, USA, 2 edition, 2012.

[83] Aaron Sloman. Review of Affective Computing. Technical Report 1, 3 1999. URL `https://www.aaai.org/ojs/index.php/aimagazine/article/view/1448`.

[84] Marina Sokolova and Antonio Fernández-Caballero. A Review on the Role of Color and Light in Affective Computing. *Applied Sciences*, 5(3):275–293, 8 2015. ISSN 2076-3417. doi: 10.3390/app5030275. URL `http://www.mdpi.com/2076-3417/5/3/275`.

[85] Mohammad Soleymani, Michael N. Caro, Erik M. Schmidt, Cheng Ya Sha, and Yi Hsuan Yang. 1000 songs for emotional analysis of music. In *CrowdMM 2013 - Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia*, pages 1–6, New York, New York, USA, 2013. Association for Computing Machinery. ISBN 9781450323963. doi: 10.1145/2506364.2506365. URL `http://dl.acm.org/citation.cfm?doid=2506364.2506365`.

[86] Mohammad Soleymani, Anna Aljanaki, and Yi-Hsuan Yang. DEAM: MediaEval Database for Emotional Analysis in Music. 2018. URL http://doi.acm.org/10.1145/2502081.2502224.

[87] Yading Song, Simon Dixon, and Marcus Pearce. A Survey of Music Recommendation Systems and Future Perspectives. In *9th International Symposium on Computer Music Modeling and Retrieval*, pages 395–410. Centre for Digital Music Queen Mary University of London, 2012. URL https://www.researchgate.net/publication/277714802.

[88] Olga Sourina and Yisi Liu. A FRACTAL-BASED ALGORITHM OF EMOTION RECOGNITION FROM EEG USING AROUSAL-VALENCE MODEL. 2011. doi: 10.5220/0003151802090214.

[89] Matthijs T.J. Spaan. Partially Observable Markov Decision Processes. In *Adaptation, Learning, and Optimization: Reinforcement Learning*, volume 12, chapter 12, pages 387–414. 2012. ISBN 978-3-642-27644-6. doi: 10.1007/978-3-642-27645-3{_}12.

[90] Anne Merel Sternheim. *Affective Semantics: An Ontology for Affective Reasoning in the PAL Project.* PhD thesis, Utrecht University, 10 2017.

[91] Huei Chuan Sung, Anne M Chang, and Wen Li Lee. A preferred music listening intervention to reduce anxiety in older adults with dementia in nursing homes. *Journal of Clinical Nursing*, 19(7-8):1056–1064, 2010. ISSN 09621067. doi: 10.1111/j.1365-2702.2009.03016.x.

[92] Jianhua Tao and Tieniu Tan. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*, volume 3784 LNCS, pages 981–995. Springer, Berlin, Heidelberg, 10 2005. ISBN 3540296212. doi: 10.1007/11573548{_}125. URL https://link.springer.com/chapter/10.1007/11573548_125.

[93] Kali S Thomas, Rosa Baier, Cyrus Kosar, Jessica Ogarek, Alissa Trepman, and Vincent Mor. Individualized Music Program is Associated with Improved Outcomes for U.S. Nursing Home Residents with Dementia. *American Journal of Geriatric Psychiatry*, 25(9):931–938, 2017. ISSN 15457214. doi: 10.1016/j.jagp.2017.04.008.

[94] Kelvin K.F. Tsoi, Joyce Y.C. Chan, Yiu Ming Ng, Mia M.Y. Lee, Timothy C.Y. Kwok, and Samuel Y.S. Wong. Receptive Music Therapy Is More Effective than Interactive Music Therapy to Relieve Behavioral and Psychological Symptoms of Dementia: A Systematic Review and Meta-Analysis. *Journal of the American Medical Directors Association*, 19(7):568–576, 2018. ISSN 15389375. doi: 10.1016/j.jamda.2017.12.009. URL https://doi.org/10.1016/j.jamda.2017.12.009.

[95] Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Transactions on Affective Computing*, 3(2):237–249, 2012. ISSN 19493045. doi: 10.1109/T-AFFC.2011.30.

[96] Martijn Van Otterlo and Marco Wiering. Reinforcement learning and markov decision processes. In *Adaptation, Learning, and Optimization: Reinforcement Learning*, volume 12, chapter 1, pages 3–42. Springer Verlag, 2012. ISBN 978-3-642-27644-6. doi: 10.1007/978-3-642-27645-3{_}1. URL https://link.springer.com/chapter/10.1007/978-3-642-27645-3_1.

[97] Renata Teles Vieira, Leonardo Caixeta, Sergio Machado, Adriana Cardoso Silva, Antonio Egidio Nardi, Oscar Arias-Carrion, and Mauro Giovanni Carta. Epidemiology of early-onset dementia: a review of the literature. *Clinical Practice & Epidemiology in Mental Health*, 9(1):88–95, 7 2013. doi: 10.2174/1745017901309010088.

[98] Annemieke C Vink, Marij Zuidersma, Froukje Boersma, P De Jonge, Sytse U Zuidema, and J. P.J. Slaets. The effect of music therapy compared with general recreational activities in reducing agitation in people with dementia: A randomised controlled trial. *International Journal of Geriatric Psychiatry*, 28(10): 1031–1038, 2013. ISSN 08856230. doi: 10.1002/gps.3924.

[99] Jonna K Vuoskoski and Tuomas Eerola. DOMAIN-SPECIFIC OR NOT? THE APPLICABILITY OF DIFFERENT EMOTION MODELS IN THE ASSESSMENT OF MUSIC-INDUCED EMOTIONS. Technical report, 2010. URL www.jyu.fi/music/coe/materials/emotion/soundtracks/.

[100] Muyuan Wang, Naiyao Hang, and Hancheng Zhu. User-adaptive music emotion recognition. In *International Conference on Signal Processing Proceedings, ICSP*, volume 2, pages 1352–1355, 2004. doi: 10.1109/icosp.2004.1441576.

[101] X Wang, Y Wang, and D Hsu. Exploration in Interactive Personalized Music Recommendation: A Reinforcement Learning Approach. *ACM Trans. Multimedia Comput. Commun. Appl*, 2(3), 2013. doi: 10.1145/0000000.0000000. URL http://doi.acm.org/10.1145/0000000.0000000.

[102] Gregg A Warshaw and Elizabeth J Bragg. Preparing the health care workforce to care for adults with Alzheimer's disease and related dementias. *Health Affairs*, 33(4):633–641, 2014. ISSN 15445208. doi: 10.1377/hlthaff.2013.1232.

[103] World Health Organization. Global strategy on human resources for health: Workforce 2030. *Who*, page 64, 2016. URL http://www.who.int/workforcealliance/media/news/2014/consultation_globstrat_hrh/en/%0Ahttp://apps.who.int/iris/bitstream/10665/250368/1/9789241511131-eng.pdf?ua=1%5Cnhttp://www.who.int/hrh/resources/pub_globstrathrh-2030/en/.

[104] World Health Organization. Towards a dementia plan: a WHO guide. Technical report, 2018. URL https://www.who.int/mental_health/neurology/dementia/policy_guidance/en/.

[105] World Health Organization. WHO: Dementia, 9 2019. URL https://www.who.int/en/news-room/fact-sheets/detail/dementia.

[106] Chih-Hung Wu, Yueh-Min Huang, and Jan-Pan Hwang. Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology*, 47(6):1304–1323, 11 2016. ISSN 00071013. doi: 10.1111/bjet.12324. URL http://doi.wiley.com/10.1111/bjet.12324.

[107] Yi Hsuan Yang, Yu Ching Lin, Ya Fan Su, and Homer H. Chen. A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2):448–457, 2 2008. ISSN 15587916. doi: 10.1109/TASL.2007.911513.

[108] Yang Yu, Zhenhao Gu, Rong Tao, Jingtian Ge, and Kenglun Chang. Interactive Search Based on Deep Reinforcement Learning. 12 2020. URL http://arxiv.org/abs/2012.06052.

[109] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009. ISSN 01628828. doi: 10.1109/TPAMI.2008.52.

[110] Marcel Zentner, Didier Grandjean, and Klaus R. Scherer. Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion*, 8(4):494–521, 8 2008. ISSN 15283542. doi: 10.1037/1528-3542.8.4.494.

[111] Kejun Zhang, Hui Zhang, Simeng Li, Changyuan Yang, and Lingyun Sun. The PMEmo dataset for music emotion recognition. In *ICMR 2018 - Proceedings of the 2018 ACM International Conference on Multimedia Retrieval*, volume 18, pages 135–142, New York, NY, USA, 6 2018. Association for Computing Machinery, Inc. ISBN 9781450350464. doi: 10.1145/3206025.3206037. URL https://dl.acm.org/doi/10.1145/3206025.3206037.