LISTEN OR INTERACT? A LARGE-SCALE SURVEY ON MUSIC LISTENING AND MANAGEMENT BEHAVIOURS

by

Mohsen Kamalzadeh
B.Sc. (Computer Engineering), Sharif University of Technology, 2010

a THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in the
School of Computing Science
Faculty of Applied Sciences

© Mohsen Kamalzadeh 2013
SIMON FRASER UNIVERSITY
Spring 2013

All rights reserved.
However, in accordance with the Copyright Act of Canada, this work may be reproduced without authorization under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
APPROVAL

Name: Mohsen Kamalzadeh

Degree: Master of Science

Title of Thesis: Listen or Interact? A Large-Scale Survey on Music Listening and Management Behaviours

Exchanging Committee:
Hao (Richard) Zhang
Associate Professor, Computing Science
Simon Fraser University
Chair

Dr. Torsten Möller
Professor, Computing Science
Simon Fraser University
Senior Supervisor

Dr. Dominikus Baur
Postdoc, Computer Science
University of Calgary
Supervisor

Dr. Joanna McGrenere
Associate Professor, Computer Science
University of British Columbia
Examiner

Date Approved: March 25, 2013
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011
Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010
Abstract

The results of an online survey on music listening and management are presented and analyzed. With 590 participants, the main goal was to understand how much control the respondents desired on their music listening experience and how much interaction with their music source they would be willing to have in order to exert such control. A need for interaction techniques which take minimal effort and let users steer the listening experience by controlling key attributes of songs was observed. The time required for this interaction should be similar to how long it takes to skip one song. Examples of attributes that needed to be controlled were found to be mood, familiarity, tempo, and how distracting the songs are. Some other notable findings were that our participants had a median of 4600 songs in their music collections, that portable devices are the most popular music source, that commuting and work are the top activities accompanying music, and that online music services have not gained much traction with users.

**Keywords:** Music listening, interaction, behaviours, habits, online survey, design implications
Acknowledgments

A few words of acknowledgement are certainly not sufficient when it comes to thanking people who have made it possible for me to be who I am. They may however bring a smile to their faces if one day they decide to have a look at this thesis, and that is what matters.

It is of utmost importance to me to thank my family, especially my parents, who have backed me in every possible way throughout all these years and missed me when I was away. I can not fathom having been able to reach this stage and write this page, had it not been for their ceaseless care and support.

It is an honour for me to have worked my senior supervisor, Dr. Torsten Möller. It was with his invaluable guidance and never-ending support that I learned how to organize my ideas, how to make them heard, how to be critical of myself, and how to overcome obstacles in research. I would like to thank my supervisor Dr. Dominikus Baur for always being there for me and helping me in every stage of my work. I would also like to seize this opportunity to thank Dr. Arthur Kirkpatrick and Dr. Derek Bingham for their extensive cooperation in my research project, and Dr. Joanna McGrenere for her insightful review of my thesis.

Finally, I would like to thank my friends Shayan, Arash, Amir Hossein, Zia, Naghmeh, Pariya, Nima, Iman, Marziyeh, Yasaman, Parnian and many others who will hopefully forgive me for not having mentioned them, for making it easier to fit in, to not miss home too much, and to be with people who cared about me and whom I cared about.
Contents

Approval ii
Partial Copyright License iii
Abstract iv
Acknowledgments v
Contents vi
List of Tables viii
List of Figures ix
1 Motivation 1
2 Related Work 5
3 Methods 11
   3.1 Demographic group comparisons and sampling bias 14
   3.2 Listener types 15
4 Results 17
   4.1 Demographics 17
   4.2 Category 1: music listening 19
   4.3 Category 2: music collection management 32
   4.4 Open ended question on most important features missing from current music listening tools 35
List of Tables

3.1  Description of listener types ................................................................ 16

4.1  Demographic characteristics of listener types. ............................... 18

4.2  Popularity of various online music services ...................................... 32

4.3  Popularity of various music sources, e.g. physical media, portable players, etc. ................................................................................. 33

A.1  Demographic characteristics of samples ........................................... 47

A.2  Difference between attention and non attention scenarios for interaction tolerance, importance, familiarity, and mood variety ..................... 48
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Distribution of age</td>
<td>18</td>
</tr>
<tr>
<td>4.2</td>
<td>Academic background of samples</td>
<td>19</td>
</tr>
<tr>
<td>4.3</td>
<td>Activities during which music is listened to</td>
<td>20</td>
</tr>
<tr>
<td>4.4</td>
<td>Methods of playback during each activity</td>
<td>23</td>
</tr>
<tr>
<td>4.5</td>
<td>Interaction tolerance</td>
<td>27</td>
</tr>
<tr>
<td>4.6</td>
<td>Importance of music (precision in control on music)</td>
<td>29</td>
</tr>
<tr>
<td>4.7</td>
<td>Difficulties faced using Media Player and iTunes</td>
<td>35</td>
</tr>
<tr>
<td>A.1</td>
<td>Histogram of number of non-answered questions</td>
<td>49</td>
</tr>
<tr>
<td>B.1</td>
<td>Survey welcome page - part 1</td>
<td>51</td>
</tr>
<tr>
<td>B.2</td>
<td>Survey welcome page - part 2</td>
<td>52</td>
</tr>
<tr>
<td>B.3</td>
<td>Survey page 1</td>
<td>53</td>
</tr>
<tr>
<td>B.4</td>
<td>Survey page 2 - part 1</td>
<td>54</td>
</tr>
<tr>
<td>B.5</td>
<td>Survey page 2 - part 2</td>
<td>55</td>
</tr>
<tr>
<td>B.6</td>
<td>Survey page 3 - part 1</td>
<td>56</td>
</tr>
<tr>
<td>B.7</td>
<td>Survey page 3 - part 2</td>
<td>57</td>
</tr>
<tr>
<td>B.8</td>
<td>Survey page 4</td>
<td>58</td>
</tr>
<tr>
<td>B.9</td>
<td>Survey page 5 - part 1</td>
<td>59</td>
</tr>
<tr>
<td>B.10</td>
<td>Survey page 5 - part 2</td>
<td>60</td>
</tr>
<tr>
<td>B.11</td>
<td>Survey page 6 - part 1</td>
<td>61</td>
</tr>
<tr>
<td>B.12</td>
<td>Survey page 6 - part 2</td>
<td>62</td>
</tr>
<tr>
<td>B.13</td>
<td>Survey page 7</td>
<td>63</td>
</tr>
<tr>
<td>B.14</td>
<td>Survey page 8</td>
<td>64</td>
</tr>
<tr>
<td>B.15</td>
<td>Survey page 9</td>
<td>65</td>
</tr>
</tbody>
</table>
Chapter 1

Motivation

With the immense growth of technology and the Internet, the landscape of music consumption has been rapidly evolving. Having access to libraries of millions of songs with just a few mouse clicks is an example of this change. With such a wide array of choices, selecting what to listen to can become a burden. A library of songs can be looked at as a multidimensional space where every song can have tens, even hundreds of features attached to it. These features range from low-level acoustic features like timbre, tempo, etc. to meta-data and more abstract characteristics like artist, album, or even how dark or gloomy the song is. Having this space, one major question emerges: How do people like to explore this space and how can technology help it? The goal of simplifying a music listener’s journey in this space is what gave birth to music recommendation services like Pandora and Last.fm. But are current recommendation services the right vessel for this journey?

We divide Music consumption into two general categories: (a) music listening, and (b) music management. Listening encompasses the reasons why music is listened to, types of music, source of music, methods of playback, social listening, sharing, and so on. For instance, the reason for listening can be to get in or out of a mood, the method of playback can be a pre-compiled playlist, and the source can be one’s music collection on a portable device. Management consists of obtaining music, maintaining tags and playlists, managing files and folders, etc. The boundary between listening and management can be fuzzy, but this categorization can provide a clear enough basis on which this thesis is organized.

Throughout the past decade, we have seen major transitions in how music is consumed. Physical media is becoming all but extinct for mainstream listeners, mobile media players are transforming from low capacity and limited functionality mp3 players into advanced
connected devices with tens of gigabytes of space, and more recently, there has been an upsurge in the availability and popularity of online music recommendation and subscription services. In various stages of these transitions, empirical studies have striven to understand how the consumers’ interaction with music and technology has been evolving. Even so, one area that has garnered less attention from researchers is popularity of various methods of music playback and interaction with music sources.

Examples of playback methods are pre-compiled playlists, shuffling on a library and skipping songs, filtering for specific attributes like artist or genre, choosing songs one after another, and so on. Every playback method is essentially a compromise between precision in control on what is being played and the amount of interaction required to exert such control. We classify a listener’s control on a music source into two types: (a) initiating and (b) adjusting. Initiating control happens when one starts listening. Adjusting control happens when the experience is deviated from what initiating control establishes. In that sense, initiating control can define an experience in the form of a sequence of songs, and adjusting control can alter it. Both these control types can have a level of detail ranging from coarse to precise. For example, with a pre-compiled playlist, the listener exerts precise initiating control by specifying both the songs and their order in the playlist. If the listener does not interact with the music source anymore, the sequence of songs played forms a musical experience. Adjusting control can now happen in the form of skipping songs, jumping around in the playlist, or adding some song on the fly. At this point, if the listener switches to another playlist or another playback method, she crosses the boundaries of adjusting and initiates a new experience.

With most current recommender services, the listener can provide seed songs or other seed attributes like artist name (relatively coarse initial control), but has minimal control on how the songs get chosen later on (almost non-existent adjusting control). Current playback interfaces mostly rely on notions of library management and the hierarchical structuring of songs based on their attributes. That is why currently, adjusting control is limited to concepts like skipping tracks or occasionally finding specific songs that come to mind during listening. Both of these have coarse and minimal impacts on the experience, and can only lead to a somewhat desirable outcome if they happen frequently. This poses the question; is there a need for new forms of adjusting control which take less effort and bring about more precise control? And how much effort is too much? We tried to answer these with our survey.
CHAPTER 1. MOTIVATION

Knowing the consumers’ current choice of which method to use sheds light on the kind of compromise between control and interaction that they find appropriate, and thus, provides us with an understanding of what needs improvement. To have an idea of what types of interaction with the music source are usually plausible for music listeners, we need to know which devices are most frequently used and which activities accompany music listening most often. Finding out what characteristics of songs influence their inclusion or placement in a playlist helps us figure out the characteristics that listeners may want to be able to control in a novel music interface which does not rely on the old concept of a rigid pre-compiled playlist and embraces dynamic tunable experiences. More importantly, is such an interface even needed? As will be discussed in the next chapter (Related Work), previous studies have investigated some of these issues, however, most of them suffer from having a very small number of participants.

This study, reports and analyses the results of an online survey with 590 respondents, covering both listening and management categories, including the issues raised above, with a focus on playback methods, control and interaction. Music listening can be both a personal or a social activity. Here, all topics are discussed in the scope of personal listening. In that sense, party music, restaurant music, and the likes, are not discussed. The contributions here are twofold:

- insight into how our participants listen to music and manage their music collections
- implications for design of music listening and discovery tools

To be able to reach out to as many participants as we could, we performed convenience sampling. In the end, 744 people from 7 populations responded to our call. These populations included a mailing list for research in auditory perception (sample 1), an opt in news mailing list from Ludwig Maximilians University in Germany (sample 2), the researchers’ friends in various social networks (sample 3), a mailing list including all Computing Science and Engineering faculty and students of Simon Fraser University in Canada (sample 4), a mailing list consisting of professional sound and music engineers (sample 5), radio professionals (sample 6), and a mailing list for research in music information retrieval (sample 7). Later on, when there is a need to refer to participants from any of these populations, the sample numbers will be used.

We put all the participants in one pool and compared groups based on age and gender. Statistically significant differences between these groups are reported in the Results chapter.
We also tried to identify various types of listeners based on three basic characteristics: (a) one’s average number of active music listening hours per day, (b) one’s average number of passive music listening hours per day, and (c) the size of one’s personal music collection. Active listening is when one listens to music just for the sake of listening and not during other activities. Passive listening is listening to music during an accompanying activity, like commuting, work, etc.

In general, a desire for precise control coupled with low interaction was observed. The survey results indicated that there is a need for novel interaction techniques which facilitate manipulating various attributes of songs on the fly, with minimal required interaction. It was found that some of these attributes are mood, familiarity of songs, how distracting the songs are, and tempo.

The rest of this thesis is organized as follows: In the Related Works chapter, a number of previous studies relevant to music collection management and music listening will be discussed. The Methods chapter explains how the online survey was conducted, what topics the questions covered, how the data was pre-processed for non-response treatment and outlier removal, and the methods used for statistical analysis. The responses to the survey will be presented in the Results chapter. These will be structured into questions on music listening and music collection management. In the Discussion chapter, a deeper interpretation of the results will be provided, along with design implications.
Chapter 2

Related Work

Music consumption research includes studies on information seeking behaviours [1, 9, 15, 15], browsing and accessing music [6, 2], playlist generation [27, 12], finding new music [4], contexts and purposes of listening [7, 17, 27, 23], social listening [27, 7, 23], and library management [29, 5]. In this chapter, a selection of these studies will be introduced and discussed in conjunction with the topics covered in our survey. A full list of these topics is provided in the next chapter (Methods).

Looking at studies published in the field of MIR\(^a\), there is a body of papers investigating how users seek music information. This encompasses both seeking music itself using a set of input characteristics, and seeking information surrounding a piece of music. We will refer to these as tasks 1 and 2. While music information retrieval is usually used as an umbrella term for both of these tasks, it is important to distinguish between them. Task 1 often involves searching for a song or a set of songs that satisfy certain conditions, while task 2 entails obtaining information on a specific song, album, artist, etc. The conditions used in task 1 and the information retrieved in task 2 can both be combinations of the seven facets of music information described by Downie [8]. These are pitch, temporal, harmonic, timbral, editorial, textual, and bibliographic facets.

As Randel [25] puts it, pitch is “the perceived quality of a sound that is chiefly a function of its fundamental frequency – in the number of oscillations per second.” To put it simply, pitch is what decides the vertical placement of notes in musical notation. The temporal facet contains information on the duration of various musical events. This includes tempo

\(^a\)Music Information Retrieval
indicators, meter, pitch duration, harmonic duration, and accents. The harmonic facet consists of information on the harmonic sequences and events of a musical work. Harmony, or polyphony, happens when two or more pitches occur at the same time. Timbre, color, or tone color is what causes the distinction between the sounds made with different instruments even if the same pitches are played. The timbral facet thus consists of all the information regarding orchestration, and performance methods. The editorial facet is comprised mainly of performance instructions, including fingerings, ornamentation, and so on. Thus, the boundary between editorial and timbral facets is often considered blurry. That discussion is however not within the scope of this thesis. The textual facet encompasses a song’s lyrics or any other text read during the performance, e.g. chorales and hymns. Finally, the bibliographic facet is made up of the meta-data associated with a musical work. These include artist, album, composer, year of release, and so on.

As Downie [8] discusses, an MIR system geared towards music experts needs to have precise facet information for at least the first six facets mentioned above. A consumer oriented system on the other hand needs to have more breadth and cover a broader selection of musical pieces, at the expense of fine-grained facet information. It is not realistic to expect average consumers to be able to express their music information needs in terms of facets other than textual and bibliographic ones. Returning to our earlier distinction between tasks 1 and 2, while consumers may be content with retrieving bibliographic and textual information in a task 2 scenario, these facets are not enough when it comes to task 1. That is, finding and listening to unfamiliar music, or even songs within one’s own library that may satisfy the user’s aspiration for music, but do not come to his mind immediately. This is why Kim and Belkin [15] set out to understand how non-experts describe music.

In 2002, Kim and Belkin [15] tried to discover categories of terms describing affective and functional aspects of music as means for listeners to express their information needs. They intended to understand (a) How users who did not have musical backgrounds perceived and described music that they heard, and (b) How such users thought they would go about searching for music that they had heard in an ideal MIR system, particularly, the descriptive terms they would have used. Both of these, especially (b), revolve around what we called task 1. They divided their 22 participants into two groups, one performing the description task and one performing the search task, and both groups were asked to listen to the same musical pieces. It was found that in both cases, “Emotions” and “Occasions or filmed events” were the most popular categories, ending up first (31%) and second (23%) in (a),
and second (24%) and first (29%) in (b). “Emotions” were defined as words that explicitly indicate emotional status like happy, joyful, sad, etc. “Occasions or filmed events” were words describing specific occasions or events, for example, “celebration,” “grand arrival or entry,” “exploring forests,” etc. In 2006, Bentley et al. [2] reported that a search for music resembles a “satisficing” behaviour more often than not. What that means is when it comes to selecting or finding music for listening, users tend to search for songs that are good enough more often than searching for specific songs.

In 2004, Lee and Downie [18] found that the purposes for seeking information surrounding musical pieces (task 2) are mainly building musical collections and verifying the identity of musical works. The results of their 427 participant web survey showed that the most sought music information were “title of work(s),” “lyrics,” and “artist information”. In 2003, Bainbridge et al. [1] analysed questions and answers from 502 postings to the music category of Google Answers. They reported that in over 80% of the cases, the users were able to provide some form of bibliographic meta-data when describing their information request. The most common ones of these meta-data were reported to be “performer(s)” and “title of a work”. Other popular attributes were “date,” “orchestration,” “collection title (album, LP, etc.),” and “composer”. Also in 2004, Vignoli [29] asked participants about the attributes they used for structuring their libraries, and found the order of importance to be “artist name,” “album name,” and “song name”. Taking into account the similar observations in these studies, one can conclude that the most sought information bits about music are also the ones that are most popular in managing libraries.

Few papers specifically investigate playback methods, one of which is the study by Vignoli [29]. He reports that the most used methods of playback chosen by interview participants, were “I choose one or more albums,” “I search for a single song,” and “I choose one or more artists”. Participants had between 1200 and 3500 songs (average of 2300) in their collections and the majority of them managed their music collections based on artist name, album name, and song title hierarchically with folders and sub-folders. In contrast, they preferred attributes like mood and activity for browsing and selecting songs. The participants also said they preferred to create playlists each time they listened to music, instead of using pre-compiled ones. From a control perspective, this can indicate a desire for high precision. These results were reported for only 7 participants. The author does mention that a web questionnaire with 130 respondents was held to confirm results. However, not all results from this questionnaire are discussed in the paper and regarding the above issues, it
is only said that artist name was at the top of the hierarchy of 60% of user libraries. Album name and genre were found to be the next most popular choices, each with around 15%.

Also relevant to playback methods is a 2006 study by Leong et al. [19] that investigates the merits of shuffling. They report that out of 113 users, 91 chose shuffling over sequential playback as their favourite method of playback. 69 of these 91 users said they would rather shuffle their whole collection than constrain it first. That said, the most popular constraining method was found to be a playlist.

In 2004, Cunningham et al. [5] examined their participants’ interaction with music collections. The participants were students of a Human Computer Interaction course who performed “personal ethnographies” to examine their own interaction with their music collections. In a time when music collections were only beginning to move from physical media to digital files, Cunningham et al. [5] mainly focused on how music CDs are accessed and organized. They also point out that there is a need for more robust searching functionalities in popular music organization software like iTunes. Other needed features included the ability to personalize the appearance of individual songs, compilations and playlists in a way that is more profound than having different skins for the player interface. Also, a need for supporting multiple users on a single library was mentioned by some participants. The fact that at the time most manufacturers strived to design smaller and smaller mp3 players was said to lead to missed opportunities. None of the participants found the small size of the mp3 players to be beneficial. They expressed a desire for portable devices to have music retrieval features that were only possible with larger screen sizes; as large as a PDA at the time.

In their paper on recommending collections of items, published in 2009, Hansen et al. [12] study playlist creation habits of 52 subjects. They report that the song appearing first in a playlist is very likely to be among a playlist creator’s favourite songs. Songs unknown to the creator were less likely to appear on the playlist, and less so in the beginning of the list. In case of creating playlists for people other than the creators themselves, participants said that the list should generally consist of songs that are liked by the target audience. 70% of subjects mentioned that songs should have a common theme relating to a particular mood, activity, time period, a certain “sound,” genre, beat, place, person, or story. One third of the participants expressed a need for variety in the list and two thirds of them said the order of songs and how they flow and blend was important. Some strategies proposed by participants to achieve this flow included “spreading the best songs in the mix throughout,”
CHAPTER 2. RELATED WORK

and building up to a “climax” and then “taking you back down,” and “not putting too many slow/fast songs together.”

In 2011, Stumpf and Muscroft [27] studied playlist creation behaviours of 7 participants, for 3 different purposes (large party, small gathering, private travel). They found out that tempo, mood, and genre were the most important factors when creating playlists for private travel, which is the only personal context studied and thus of interest to us. Also, participants expressed high importance for being able to control the playlist and said that it needed to flow. Unfortunately, the small number of participants makes it hard to generalize these results to any population.

As for where and when music is listened to, in 2001, in a study with 36 participants who were reported to be music enthusiasts, Brown et al. [3] found the top places for music listening to be car (82%), living room (61%), and work (38%). The study focuses on CDs as digital songs were not prominent at the time. It is also reported that the choice of music was in part reliant on what technologies were available at the place where music was listened to, and also the collection of music available. Nowadays, these limits are slowly fading away with the music collections shifting to portable devices and the cloud.

In 2007, Salvucci et al. [26] studied the effect of interacting with an iPod while driving and observed that selecting music can significantly affect driving performance. In 2009, Lamont and Webb [17] found the five most frequent activities accompanying music listening to be, in decreasing order, working, socializing, driving, and public transport. In 2010, Liikkanen et al. [20] found that on average, 1 out of 9 commuters have headphones on and that most of them do not interact much with their device while on transit. Interestingly, they report that the number of portable music devices sold is much higher than what this 1 out of 9 fraction suggests, indicating that most people would rather not use these devices while commuting.

In 2012, we [14] conducted an earlier stage of our online survey, in which we investigated various aspects of music consumption, with a focus on playback methods, control, and interaction. With 222 participants, we found that elements such as (a) familiarity of songs, (b) how distracting they are, (c) the mood of songs and if it needs to vary during one listening session, and (d) how much interaction should be needed for this purpose are all dependant on the amount of attention required by the activity during which music is listened to. We observed that listeners may want to alter the flow of music even if they are paying attention to an accompanying activity. To achieve that, they did not mind interacting with their music.
source as long as it was not much more than the effort it takes to skip a song. We also reported that mood, genre, and artist were the most important factors in playlist creation. A caveat however was that most of our participants had a technical background with the majority being Computing Science and Engineering students and faculty from Simon Fraser University. This time, we add responses gathered from 522 more respondents, which come from more diverse backgrounds than before, and provide a deeper analysis.

\[b\] with a total of 744 respondents. Out of these, 154 were removed due to high non-response.
Chapter 3

Methods

Initially, the survey had a total of 744 respondents from 7 populations. Six of these populations were invited with mailing lists and one by spreading a link through our personal social networks. Since one of our largest samples came from Germany, for the convenience of respondents, the survey was available in both English and German. After pre-processing, 590 respondents were chosen for data analysis.

The questionnaire went through 2 pilot tests with a total of 10 respondents. In the first pilot stage, 3 respondents filled the questionnaire while explaining their thoughts on each question in detail. This method is often referred to as a think-aloud interview. After revisions, a second pilot stage was carried out with retrospective interviews with 7 other respondents. In this method, the participant fills the questionnaire while taking note of any issues she has with the questions. The researcher may also observe the respondent’s behaviour and keep track of any notable behaviour, like certain questions taking longer than expected to answer.

The final version of the questionnaire consisted of 33 questions spread across 11 pages. It took about 10 minutes to fill out, and participants were free to not answer any of the questions if they desired. The full questionnaire can be found in section B.1 of Appendix B.

Four of the questions were removed from analysis as we later decided they were either not sound enough or too tangential. For listening, the following topics were included in analysis:

- average hours of active listening per day (ordinal ranges)
- average hours of passive listening per day (ordinal ranges)
- activities during which music is listened to (ranking of 4 activities)
• methods of music playback for each of the 4 activities ranked earlier (choice of 1 out of 6 methods for each activity)
• important attributes of songs for playlist creation (Likert scale for each attribute)
• interaction tolerance and control (explained later)
• popularity of online music services (one multiple choice and one single choice question)

Regarding management of music collections, the following topics were included:

• music sources, e. g. physical media, radio, personal digital collections, etc. (Likert scale for each source)
• size of personal music collections (response in numbers: number of physical media, and gigabytes/number of digital songs)
• correctness of meta-data (tags) in personal collections (categorical)
• number of playlists maintained (ordinal ranges)
• use of manual folder structures versus applications, for collection management (categorical)
• primary music management/playback software and issues faced using them (categorical then multiple choice)

All the above topics, except for size of music collections, were asked using closed-ended questions with categorical or ordinal response choices.

The following demographic questions were asked at the end of the survey:

• age
• gender
• country of residence
• level of education
• academic discipline (asked only if level of education was college or higher)

In addition to the above, an open ended question was presented to the participants late in the questionnaire which asked them what they thought the most important features missing from current music listening tools were.

The first public stage of the study had 79 participants. After this stage the results were examined for non response. A new version of the questionnaire with 4 slightly revised questions was used for the remaining 665 participants. In case of these revised questions, only the responses from the newer version are reported here.
As participants were allowed to not answer any question if they desired so, we needed a strategy to handle non-response. Looking at a histogram of number of non answered questions (Figure A.1 in Appendix A), going from 5 to 6 non answered questions, the number of respondents fell almost flat with no further major change. We observed that 590 (79%) respondents had skipped less than 6 questions. This was chosen as a cut-point and the remaining 154 participants were removed from further analysis. From this point on, whenever we refer to “participants” or “respondents”, we mean these 590. Note that demographic questions were not included in this pre-processing step to acknowledge the relative importance of other questions and the fact that the question on academic discipline got only 226 responses. To deal with missing data in this 590 participant pool, in each analysis, we removed the participants who had skipped any question involved in the analysis. The only place where we deviated from this method was when clustering participants into listener types. This will be explained shortly.

Most of the responses to questions were in the form of ordinal variables. For these, the Mann-Whitney U test [22] was used to check the statistical significance of differences between two groups. For comparisons among more than two groups, the Kruskal-Wallis one way ANOVA test [16] was used. In case of significant omnibus effects with the Kruskal-Wallis test, we used Dunn’s [10] method for pairwise comparisons, which performs the necessary corrections required for post-hoc pairwise tests. We had categorical variables as well, for which Pearson’s Chi Squared statistic was used. For interval variables, like size of music collections, the t-test was performed. Finally, to compare one group’s answers to two related questions, the Marginal Homogeneity test [13] was used for ordinal and dichotomous variables and Pearson’s Chi Squared test was performed for categorical variables. An alpha level of 5% was used in all the tests.

In the end, we found 21 statistically significant differences between males and females, 21 statistically significant omnibus differences among age groups, and 11 statistically significant omnibus differences among listener types, adding up to a total of 53. With an alpha level of 5%, we need to keep in mind that 5% (about 3 out of 53) of these observations may have happened due to chance. It is worth noting however that considering the conservative nature of the non-parametric tests we used, it is likely that our results will not be prone to

\[\text{The word “group” here and later on in the paragraph is not referring to any of the initial 7 samples. It can refer to any group of participants that may be included in a statistical test, e. g. a listener type or a demographic sub-population.}\]
such type I errors. Besides, the sheer number of effects observed increases our confidence in not having made type II errors.

3.1 Demographic group comparisons and sampling bias

As noted earlier, comparisons of different groups of participants will be presented in the results section. We took two approaches, one based on demographic characteristics (age and gender), and another based on a “listener type” characteristic that we introduce later.

When investigating effects of demographic aspects, it is ideal to have all demographic domains represented equally in the data. By a “domain” we mean each area of a high dimensional contingency table made up of all demographic variables. Since we have performed convenience sampling, there is sampling bias in our data, and these domains do not have equal presence in our samples. For instance, in our gathered responses, the females are generally younger than males and this correlation is statistically significant. As a result, no matter how we define our age strata boundaries, our sample bias will cause the domains consisting of older female participants to be under-sampled compared to males in the same age ranges. The undesirable consequence of this phenomenon is that if we compare music consumption behaviours of these age strata, our observations can be caused by both gender and age differences, and will thus be inconclusive.

To achieve the most meaningful results with what data we have, it is ideal to control for all other demographic variables when investigating the effects of one. Our data allowed us to control for gender when investigating age and vice versa. For controlling, we used random over-sampling to increase the size of under-sampled domains. This is preferable to random under-sampling of larger domains as it will not result in information loss. We should however be careful with the extent to which we over-sample. It is generally believed that a minimum sample size of 30 is required for drawing any meaningful conclusion regarding a population, and that is when random sampling has been performed. Hence, we tried to keep our domains as large as possible. This is only achievable if we try to have equally sized age strata. These were chosen to be:

- age $\leq 21$, with 177 participants
- $21 < \text{age} \leq 27$, with 211 participants
- $27 < \text{age}$, with 185 participants
Our data were not diverse enough to let us control for country of residence, when investigating the effects of age and gender. Most of our respondents resided in either Canada (160) or Germany (282), making the rest of the domains too small for over-sampling. Regarding the effects of country of residence on music consumption behaviour, we did not observe a large number of significant differences, and thus grouping based on country is not reported here. The data was also too sparse to allow controlling for academic discipline or drawing meaningful conclusions on how it can affect music consumption. Finally, we observed a large correlation between age and level of education, which made it pointless to investigate the effects of both. Hence, we will only focus on age.

3.2 Listener types

As a second approach, we used three basic characteristics of our participants’ music consumption to define listener types. These three were average number of active listening hours per day, average number of passive listening hours per day, and size of personal music collection. Our goal was to find the most dense areas of the three dimensional space formed by these three attributes and see if the respondents fell into clusters with clear boundaries.

We first removed any of the 590 participants who had not answered more than one of these three questions from further analysis (5 cases). Then we used median replacement to fill the missing data for those who had one missing response (16 cases). To diminish the effect of each dimension having a different scale than others, we calculated Z scores and used them instead of raw responses for further analysis. In the final pre-processing step, we removed the multivariate outliers (13 cases) of this three dimensional response space using the Mahalanobis distance \[21\]. After these modifications, we were left with 572 respondents.

No obvious clusters were distinguishable in our data. The cases seemed to spread out from one dense central point in all directions. We used the DBSCAN [11] density based clustering algorithm coupled with manual parameter optimization to carve out a core area which consisted of the majority of respondents. We then clustered the participants outside of this core using the K-Means algorithm into four clusters. In the end, we got five listener types.

\[\text{Refer to section A.2 of Appendix A for an explanation of this method.}\]
CHAPTER 3. METHODS

<table>
<thead>
<tr>
<th>Type Nickname</th>
<th>Number of Participants</th>
<th>Description</th>
<th>Active Hours Median</th>
<th>Passive Hours Median</th>
<th>Collection Size Median (songs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>472</td>
<td>The most dense area of the active/passive/collection size space. These participants’ responses are closer to the medians of all participants, and thus, they reflect the whole population.</td>
<td>1</td>
<td>3</td>
<td>3,600</td>
</tr>
<tr>
<td>Actives</td>
<td>23</td>
<td>Mildly higher than Majority active hours and collection size, but very low passive hours.</td>
<td>2</td>
<td>1</td>
<td>15,300</td>
</tr>
<tr>
<td>Collectors</td>
<td>26</td>
<td>Similar to Majority active hours, with mildly higher passive hours; but much larger collections.</td>
<td>1</td>
<td>4</td>
<td>55,100</td>
</tr>
<tr>
<td>Passives</td>
<td>20</td>
<td>Similar to Majority active hours and collection sizes, but very high passive hours.</td>
<td>1</td>
<td>7</td>
<td>4,200</td>
</tr>
<tr>
<td>Musicals</td>
<td>31</td>
<td>These participants have very high active hours along with relatively high passive hours, but similar to Majority collection sizes.</td>
<td>3</td>
<td>5</td>
<td>6,200</td>
</tr>
</tbody>
</table>

Table 3.1: Description of listener types defined from three basic characteristics of music consumption: average active listening hours per day, average passive listening hours per day, and collection size. A total of 572 respondents were included in this analysis. The values shown in the table for active and passive hours are on the following scales. Active hours scale: (1) Less than 1 hour, (2) 1-2, (3) 2-4, (4) 4-8, and (5) More than 8 hours. Passive hours scale: (1) Less than 1 hour, (2) 1-2, (3) 2-4, (4) 4-6, (5) 6-8, (6) 8-10, (7) 10-12, (8) and More than 12 hours.

types, which are described in Table 3.1. We will discuss the significant behavioural differences found between the four outer clusters and the core majority. We should note however that due to the relatively small size of these outer clusters, the statistical tests used may not be thoroughly reliable. Nevertheless, there are some major differences among listener types that can be insightful towards understanding our participants’ behaviours. For each question where a significant overall effect among types is seen, the groups that caused it will be mentioned.

*Differences in responses given to questions other than the three used for clustering.*
Chapter 4

Results

In this chapter, we will first have a look at demographic characteristics of the samples. Then, survey results will be presented structured into the two music consumption categories introduced earlier, namely, music listening, and music collection management. Differences between age and gender groups will be presented here. The last topic discussed will be the clustering of participants based on their listening hours and size of music collection.

4.1 Demographics

Table A.1 in Appendix A provides a brief description for each population and summarizes age and gender characteristics of samples from each of them. Figure 4.1 shows the distribution of age for all participants. The median, mean, and standard deviation of age were respectively 24, 26.9, and 9. Of all respondents, 47% were females, 94% had at least a college degree, and 42% had a Bachelor’s. The demographic characteristics of our 5 listener types are summarized in Table 4.1. Members of Majority were 50.5% males and 49.5% females. Passives and Musicals also had roughly uniform gender distributions. There was however a male bias in Actives and Collectors. These clusters were 64% and 86% male respectively (Pearson’s Chi Squared, $\chi^2(4, N = 561) = 12.360, p = 0.015$). Musicals were the youngest cluster with a median age of 22 and Collectors were the oldest one with a median of 32. It is important to note that the reason behind the Collectors’ older age might be the actual time needed to gather collections as massive as theirs.

Only 226 participants answered the question on academic discipline. Hence, we cannot have a complete understanding of the level of academic diversity of our participant pool.
What we can say for sure is that unlike Sample 4 which had a heavy bias towards Engineering or Computer Science backgrounds and formed the main part of our previously published results [13], the rest of the samples had more diversity. What information we have on academic discipline for each sample is shown in Figure 4.2 classified into three categories: (a) applied sciences, (b) formal and natural sciences, and (c) humanities and social sciences.

The canonical members of the applied sciences group were computer science and engineering, making up for 69% and 16% of the group respectively. In group (b), physics (42%), mathematics (15%), life sciences (15%), and earth sciences (12%) were the more prominent fields. Lastly, arts (38%), psychology (15%), history (13%), and sociology (11%) had the largest shares in group (c).

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of Participants</th>
<th>Male</th>
<th>Female</th>
<th>Min Age</th>
<th>Median Age</th>
<th>Max Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>472</td>
<td>231</td>
<td>226</td>
<td>14</td>
<td>24</td>
<td>78</td>
</tr>
<tr>
<td>Actives</td>
<td>23</td>
<td>14</td>
<td>8</td>
<td>19</td>
<td>26.5</td>
<td>65</td>
</tr>
<tr>
<td>Collectors</td>
<td>26</td>
<td>19</td>
<td>3</td>
<td>21</td>
<td>29</td>
<td>56</td>
</tr>
<tr>
<td>Passives</td>
<td>20</td>
<td>9</td>
<td>11</td>
<td>16</td>
<td>26.5</td>
<td>50</td>
</tr>
<tr>
<td>Musicals</td>
<td>31</td>
<td>15</td>
<td>14</td>
<td>18</td>
<td>22</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 4.1: Demographic characteristics of listener types. Participants were allowed to not answer any of the questions, including demographics. That is why the sum of the numbers of males and females may not be equal to the size of the group.
CHAPTER 4. RESULTS

Figure 4.2: Academic background of samples for participants who provided a response (N = 226). The vertical axis depicts the percentage of each academic discipline’s share in each sample. Only the participants who replied to the question were included in calculating the percentages. Due to the particularly low response rate for this question, the percentage of each sample’s members that replied is also shown below the chart as “response rate”.

4.2 Category 1: music listening

Listening to music can happen both actively and passively. Active listening is when one listens to music for the sake of listening, not accompanying other activities. Passive listening is when music is listened to during other activities for purposes like getting in or out of moods, overcoming ambient noise, helping concentration, and so on [7].

Average hours of music listening per day (both active and passive): Participants were asked to choose one of 5 choices for active listening hours. These were: “Less than 1 hour,” “1-2,” “2-4,” “4-8,” and “More than 8 hours.” Since passive listening generally happens more than active listening, the choices were changed to cover a larger range: “Less than 1 hour,” “1-2,” “2-4,” “4-6,” “6-8,” “8-10,” “10-12,” and “More than 12 hours.” In order not to impose too much mental strain on the respondents, we decided not to have actual number inputs or finer grained response scales. Furthermore, we found the level of accuracy provided by these scales sufficient for the purpose of extracting listener types, without failing to include higher ends of the spectrum.

The medians for active and passive listening hours for all participants were “Less than 1 hour” and “2-4 hours” respectively. No significant overall difference between age groups or listener types was observed in regards with either active or passive listening hours. Male
Figure 4.3: Activities during which music is listened to, ranked from 1 to 4. Part (a) includes non-controlled data from all participants (N = 578). Parts (b) and (c) show gender sub-populations, controlled for age (Female: N = 301. Male: N = 338). Parts (d), (e), and (f) show age sub-populations, controlled for gender (Part (d): N = 190. Part (e): N = 214. Part (f): N = 232). Please note that all participants with an answer for rank 1 were included in these analyses. Therefore, the percentages in ranks 2 to 4 may not add up to 100.

Participants however had a significantly larger number of active listening hours (Mann-Whitney U, \(U(N = 64^{(f)}) = 43258, Z = -4.318, p = 0.000\)), but with the same median as females (“Less than 1 hour”). There was no significant difference for passive listening hours.

**Activities during which music is listened to:** Participants were asked to rank 4 activities: commuting, exercising, work, and housework. An option to provide other activities was also offered. The survey asked about the methods of playback that participants used for the same activities as well, which will be presented later on. The choice of these four was partly based on findings by Lamont and Webb [17]. However, to keep the number of choices small, some modifications were made. Since driving is by nature an activity which

\[\text{The number exceeds 590 due to over-sampling}\]
does not allow high interaction and because it was previously found that most commuters do not have much interaction with their music sources either [20], the two activities of driving and commuting were merged under “commuting”. Also, “socializing” was removed as the focus in this study is only on personal listening. Finally, exercising was included in order to see if its reliance on specific tempo and rhythm would show a surge in usage of pre-compiled playlists.

Figure 4.3(a) shows the results for all participants. Commuting and work make up for most of the first rank and exercising seems to only hold its ground in the third and fourth ranks. To further investigate activities, the Ranked Pairs method was performed. This method revealed commuting as the top activity and housework as second most popular, with work and exercising at third and fourth places.

Parts (b) and (c) of Figure 4.3 show the results for female and male sub-populations. Here, a considerable difference in rank 1 activities can be seen. Work (40%) was much more popular with male participants, going neck and neck with commuting (41%) for rank 1. For females however, commuting dominated rank 1 with 54%. Both exercising and housework were more prevalent with females, at the expense of work getting only 20% of the rank 1 votes. The difference in rank 1 activities between males and females was found to be statistically significant (Pearson’s Chi Squared, $\chi^2(3, N = 639) = 32.067, p = 0.000$). This was also reflected in Ranked Pairs results, were females’ top activities were found to be commuting and housework, whereas for males, this changed to commuting and work.

Our age groups also had different choices for their top activities accompanying music listening. Parts (d), (e), and (f) of Figure 4.3 demonstrate these groups’ results. The largest difference was the much higher popularity of work as rank 1 in the third age group (age > 27). This was found to be statistically significant (Pearson’s Chi Squared, $\chi^2(6, N = 636) = 20.883, p = 0.002$). Again, this was apparent in Ranked Pairs results as well, with the third age group’s second most prevalent activity after commuting being work, as opposed to housework for age groups 1 and 2.

Majority’s responses closely resembled those of the whole population, as seen in Figure 4.3(a). Work, while being the second top activity overall, could only get one first rank vote from Actives. The votes were instead divided between commuting and housework. Musicals also did not listen to music during work that much. In their case, commuting

^[Refer to section A.1 of Appendix A for an explanation of this method.]
gathered most of the votes as first rank. On the contrary, Collectors and Passives rated work overwhelmingly higher than Majority, and made it their top activity (Pearson’s Chi Squared, $\chi^2(12, N = 561) = 37.828, p = 0.000$).

In our previous results [14], housework was a clear third after commuting and work. This time around, the distance between housework and work seems much smaller. Brown et al. [3] actually found work to be in the third place (38%) after car (82%) and living room (61%). Those results were however published back in 2001, did not discuss differences between demographic groups with regard to these activities, and had only 36 participants. Thus, a direct comparison with our results does not seem appropriate.

The top “other” activity was leisure time on the computer, which was mentioned a total of 57 times. Other frequently mentioned activities included shower/bathroom (32), gaming (27), before/after sleep (24), and reading (20).

Methods of playback for various activities: Respondents were asked to choose between one of 6 playback methods for each of the four activities we had before plus active listening. These methods included:

(1) choosing songs one after another
(2) picking an album, artist, or genre
(3) a prepared playlist or folder of songs
(4) a shuffle on your whole collection
(5) online recommendation services
(6) radio, including online stations

Besides understanding the participants’ playback habits, this question was designed to provide insight into how precise of a control the participants desire on their listening experience. Figure 4.4(a) visualizes the responses to our playback methods question for all participants. It is interesting how choosing songs one after another is only popular for active listening, suggesting that the kind of very precise control facilitated by this method is not needed during passive listening. In case of commuting, methods 2, 3, and 4 have pretty similar shares with “playlist or folder” having the highest percentage (27%). As expected, prepared playlists rose to their highest percentage (40%) with exercising, marking a need for very precise control there. That said, exercising was not among the top activities, finishing last in our Ranked Pair analysis for all demographic sub-populations. This confirms the findings by Lamont and Webb [17] who reported exercising to be the 11th popular activity.
accompanying music. As for work, methods 2 and 3 dominate the scene, with shuffle scoring lower than with other activities, indicating that shuffle cannot provide sufficiently precise control when it comes to listening during work. Finally, radio gets a large bump during housework, which can signal less desire for precise control there. This surge may also be due to the fact that too much interaction is usually not possible during housework as the source may be away from the listener, making radio the best choice.

The most notable contrasts between female and male groups relate to the use of radio. It is apparent in parts (b) and (c) of Figure 4.4 that for both commuting and housework, females showed a much higher preference for radio than males. In fact, all of the passive
listening methods were found to be significantly different between males and females, with the weakest difference being in methods during work (Pearson’s Chi Squared, commuting: \(\chi^2(6, N = 648) = 43.899, p = 0.000\), exercising: \(\chi^2(6, N = 648) = 20.661, p = 0.002\), work: \(\chi^2(6, N = 648) = 13.594, p = 0.035\), housework: \(\chi^2(6, N = 648) = 40.376, p = 0.000\).

It is interesting to review the patterns emerging in parts (d), (e), and (f) of Figure 4.4. The youngest and oldest groups have the largest number of significant differences (active listening, commuting, and work). This is while groups 1 and 2 are only different in work, and groups 2 and 3 are different in active listening. So, it seems like as we enter the second period \((21 < \text{age} \leq 27)\), the use of playlists decreases significantly for both commuting and work, giving its place to more shuffling and higher usage of radio. As we go from age period 2 to 3 \((\text{age} > 27)\), the tendency towards using an album, artist, or genre as a filter for song selection increases. In general, our results can suggest that desire for control decreases as one grows older. These pairwise differences between age groups caused the overall effects to be significant for all activities (Pearson’s Chi Squared, active listening: \(\chi^2(12, N = 646) = 30.242, p = 0.003\), commuting: \(\chi^2(12, N = 646) = 27.913, p = 0.006\), work: \(\chi^2(12, N = 646) = 32.469, p = 0.001\), housework: \(\chi^2(12, N = 646) = 23.377, p = 0.025\).

The only activity for which a significant difference was found among listener types was work. Like other questions, members of Majority had a distribution similar to what is shown in Figure 4.4a. As this was a categorical response with 6 choices, it is hard to pin point the exact reason for a significant overall difference (Pearson’s Chi Squared, \(\chi^2(24, N = 570) = 49.046, p = 0.002\)). One obvious heavy weight difference involved Collectors, who deviated from Majority by a large margin. They chose “choosing an album, artist, or genre” considerably more than any other type. They also had much lower preference for playlists and radio than Majority.

**Important attributes of songs for playlist creation:** We asked participants about how important the following attributes were to them when creating playlists: artist, album, tempo, genre, mood, instruments used. Overall, mood and genre were the first and second most important attributes. Artist and tempo shared the third place with very similar ratings. These top four are the same as what we had previously found [14]. Our top attribute is also in agreement with findings by Hansen and Golbeck [12]. However, we see an inconsistency with what Stump and Muscroft observed [27] as they found the order of importance to be tempo, mood, genre. The most popular attribute from the “other” field
was lyrics/content, which was mentioned 10 times.

Age groups did not show significant differences for any of the attributes. The third and fourth most important attributes saw some fluctuations, but the four top ones stayed the same. Likewise, the overall order of attributes was identical for both gender groups. That said, females rated four attributes significantly different from males. They rated mood higher (Mann-Whitney U, $U(N = 442) = 29650$, $Z = 5.443, p = 0.000$), genre lower (Mann-Whitney U, $U(N = 442) = 20735$, $Z = -3.273, p = 0.001$), tempo higher (Mann-Whitney U, $U(N = 442) = 27168$, $Z = 2.205, p = 0.027$), and album lower (Mann-Whitney U, $U(N = 442) = 20635$, $Z = -3.199, p = 0.001$).

Musicals rated instruments used in the song higher than other listener types, creating a significant overall difference (Kruskal Wallis, $\chi^2(4, N = 391) = 24.171, p = 0.001$).

**Interaction tolerance and control:** As discussed in the Motivation chapter, one of our goals was to find out how much interaction with a music source would be too much, and how much precision in control is required. In this regard, we intended to study the following 4 issues:

(a) How much interaction is acceptable in case of a desire for changing the mood.
(b) How much precision in control is desired (asked from participants in terms of “importance of music”.)
(c) How familiar the songs should be.
(d) If the listener prefers a constant or varying mood throughout a listening session.

The first of the above topics is geared towards understanding the maximum amount of acceptable interaction. The second one attempts to quantify the overall precision of control the listeners desire on their experience. The third question explores an attribute that has been mostly neglected in music listening tools, familiarity of songs, and tells us if this element forms part of the listeners’ desired control. The reason for choosing familiarity over other under explored attributes was that it can be easily derived from a user’s listening history. The fourth question investigates the mood attribute that is known to be important from previous studies. This time however, we wanted to know if the mood of one listening session should be monotonous or varying. Simply put, this last topic would tell us if the first topic regarding interaction tolerance was even relevant.

It would have been naive to assume that these do not change with the context of listening. It was also not reasonable to try and investigate these issues for every one of the four
activities we had before, as that would lengthen the questionnaire too much. Moreover, if
it did turn out that the above issues are dependant on the activity during which music is
listened to, we would still not know what characteristic of the activities caused this effect.
Thus, we decided to choose a factor that we thought would be influential. We picked the
amount of attention an accompanying activity needs and formed the following hypothesis
around it:

- The level of desired precision in control, the maximum amount of interaction tolerated
  in order to exert control, the desire for having songs with constant or various moods,
  and if the songs should be familiar or not, are all dependent on the amount of attention
  the activity accompanying music listening needs.

Participants were first asked to imagine an activity that needs their attention and answer
4 questions, one for each of the topics discussed above. Then, the same four were asked
regarding activities that do not need much attention. From this point on, the former will
be called attention activities, and the latter will be called non-attention activities.

The participants also had the option to choose if they did not listen to music during
either of the activity types. A total of 13.7% of participants said they did not listen to music
during attention activities. This translated to 14.1% of females and 11.2% of males. As for
age groups, 9.3% of group 1, 17.9% of group 2, and 12.2% of group 3 did not listen during
attention activities. This difference turned out to be statistically significant (Pearson’s Chi
Squared, $\chi^2(2, N = 629) = 6.666, p = 0.036$). A mere 0.8% of all participants said they did
not listen during non-attention activities. No significant difference was observed among any
of the age or gender sub-populations.

**How much interaction is acceptable:** The respondents were asked about the maxi-
mum amount of interaction they would want with their music source in case they wanted to
change the mood of their listening experience. They were given four choices ranging from
very low interaction to high interaction which were based on the time needed to complete
one interaction instance. For example, skipping a track needs very low interaction, but a
long sequence of consecutive skips can take a long time. One interaction instance is just
one skip. To make sure the choices were clear enough, participants were given examples for
each of them. The four choices were:

1. very low interaction: e.g. skipping tracks
Figure 4.5: The maximum interaction the participants were willing to have with their music source, in case they wanted to change the mood. Part (a) includes non-controlled data from all participants (Attention: N = 458. Non-attention: N = 517). Parts (b) and (c) show female and male sub-populations, controlled for age (Attention, female: N = 243, male: N = 269. Non-attention, female: N = 282, male: N = 294). Parts (d), (e), and (f) show age sub-populations, controlled for gender (Attention, part (d): N = 169, part (e): N = 154, part (f): N = 191. Non-attention: part (d): N = 184, part (e): N = 183, part (f): N = 209).

(2) low: e.g. specifying your desired change in mood but not having to find any particular song

(3) medium: e.g. switching to another playlist

(4) high: e.g. finding specific songs one after another

The dominant trend was lower interaction tolerance for attention activities and higher interaction tolerance for non-attention ones. Overall results for both activity types and a break-down for gender and age groups are shown in Figure 4.5. A tendency for lower interaction tolerance was apparent for females compared to males. In the case of attention activities, this difference was found to be statistically significant (Mann-Whitney U, U(N =
The middle age group also had a significantly lower interaction tolerance than both the youngest and oldest groups. This resulted in a significant overall effect for age groups (Kruskal Wallis, $\chi^2(2, N = 587) = 13.141, p = 0.001$).

During attention activities, Collectors were more open to interaction than other types, with a median equal to 2. This caused a significant overall difference (Kruskal Wallis, $\chi^2(4, N = 512) = 10.616, p = 0.031$). Collectors had higher interaction during non-attention activities too, but were not different enough for statistical significance. Majority’s responses resembled Figure 4.5(a) closely, with a median of 1.

**How important the music is:** The choices provided covered a range from precise to coarse control. They were:

1. I choose each song carefully.
2. I don’t care about individual songs but they should match my mood.
3. I don’t care about individual songs or their moods, but the music shouldn’t distract me from my other activity.
4. I just want to hear some music, I don’t care what’s being played.

An “other” option was also included for this question as a need for one was expressed by our pilot participants. In the end however, so few respondents chose “other” that the overall picture was not affected. Figure 4.6 summarizes the responses to this question for both activity types. In general, for attention activities, most of the respondents expressed a need for more precision than what choice (4) offered, and for non-attention activities, this shifted up to choice (2). It is interesting to see that even during attention activities, one fourth of the participants wanted ultimate control (choice 1) and another third desired music that matched their mood. What catches our attention most though is the difference between gender groups (Figures 4.6(b) and (c)), where, a much higher need for control with males and attention activities results in a significant effect (Mann-Whitney $U$, $U(N = 585) = 41421, Z = 2.733, p = 0.006$). A similar pattern was also visible for non-attention activities, which was not strong enough for statistical significance. In addition, the third age group expressed a considerably lower need for control than groups one and two during non-attention activities (Figures 4.6(d), (e), and (f)), causing a significant overall effect for age groups (Kruskal Wallis, $\chi^2(2, N = 626) = 8.843, p = 0.012$).
Figure 4.6: The importance of the music listened during both activity types. Part (a) includes non-controlled data from all participants (Attention: N = 505. Non-attention: N = 582). Parts (b) and (c) show female and male sub-populations, controlled for age (Attention, female: N = 258, male: N = 304. Non-attention, female: N = 303, male: N = 342). Parts (d), (e), and (f) show age sub-populations, controlled for gender (Attention, part (d): N = 177, part (e): N = 179, part (f): N = 201. Non-attention: part (d): N = 194, part (e): N = 214, part (f): N = 233).

No significant difference was found among listener types for the attention scenario. For non-attention activities, Collectors voted for much more control than the Majority type participants, and had a median of 1. This is what caused a significant overall difference (Kruskal Wallis, \(\chi^2(4, N = 550) = 19.171, p = 0.001\)). The Majority cluster had a response distribution similar to what is seen in Figure 4.6(a), with a median of 2.

**How familiar the songs should be:** Three choices were provided for familiarity of songs:

1. I mostly listen to music I’ve heard before.
(2) I mostly listen to new songs (for example on the radio).

(3) I don’t care if the songs are familiar or not.

An overall tendency towards familiar songs was observed. In all demographic sub-populations but age group 3, choice 1 gathered more than half of the votes for both activity types. Preference for new songs (choice 2) stayed below a mere 8% for attention activities in all sub-populations, and below 20% for the non-attention scenario. Choice 3 stayed between 20 and 30 percent, and was second to choice 1 in all demographic sub-populations. Lastly, in all groups, going from attention activities to non-attention ones resulted in a roughly 12% drop in choice 1, which was almost equally shared between choices 2 and 3.

Although the overarching pattern was the same for all demographic sub-populations, there were notable differences when we looked deeper. Females chose the second option more often than males for both scenarios. This led to a significant difference in the case of attention activities (Pearson’s Chi Squared, \( \chi^2(2, N = 550) = 7.851, p = 0.020 \)). Regarding age groups, as we go from younger age to older age, choice 1 loses popularity. This loss results in higher popularity for choice 2 as we transition from group 1 to group 2, and in a surge of popularity for choice 3, as we move further to group 3. This phenomenon exists for both types of activity (attention: \( \chi^2(4, N = 546) = 31.384, p = 0.000 \), non-attention: \( \chi^2(4, N = 629) = 24.002, p = 0.000 \)).

This was the question where the outer clusters had the largest number of deviations from Majority. For attention activities, the Majority group had high preference for choice 1 (67%). Choice 3 got 27%, and choice 2 was last with only 6%. Collectors had much lower preference for familiar songs than Majority. This came in exchange for a higher popularity for choice 3. Besides, none of them voted for new songs. Passives showed a preference similar to Majority for previously heard songs. However, they voted for new songs much more than other types, in exchange for a low choice 3 rating. Musicals preferred familiar songs more heavily (86%). These differences resulted in a significant overall effect (Pearson’s Chi Squared, \( \chi^2(8, N = 477) = 36.669, p = 0.000 \)).

As for non-attention activities, the Majority group still had high preference for choice 1 (57%). Next were choice 3 with 30%, and choice 2 with 13% (almost double that of attention activities). Collectors again had much lower preference for familiar songs than the Majority cluster, in expense of a higher popularity for choice 3. Passives kept their high preference for new songs, but this time, it was in exchange for choice 1. These differences resulted in a significant overall effect (Pearson’s Chi Squared, \( \chi^2(8, N = 555) = 23.059, p = 0.003 \)).
If the listener prefers a constant or varying mood throughout a listening session: A preference for constant mood during attention activities and varying mood during non-attention ones was observed. The percentages for the whole participant pool were 67% constant mood to 33% varying moods during attention activities and an almost mirror 34.5% to 65.5% during non-attention ones. All demographic sub-populations had a similar distribution of responses with no significant difference across the board. Collectors were the only group that gave the edge to various moods even during attention activities. In a direct comparison with Majority participants, this effect was statistically significant (Pearson’s Chi Squared, $\chi^2(1, N = 371) = 7.430, p = 0.006$), but since Passives and Musicals also showed similar but less strong deviations from Majority, the overall effect came just short of being significant (Pearson’s Chi Squared, $\chi^2(4, N = 431) = 9.373, p = 0.052$).

Our hypothesis concerning control, interaction, and attention was confirmed for all demographic sub-populations, and the Majority listener type. Table A.2 in Appendix A contains the corresponding test results. In some cases, certain listener types did not show significantly different behaviours between attention and non-attention activities. Regarding importance of songs, Passives and Musicals had an already higher desire for precise control than other types during attention activities, so the move to non-attention activities did not result in that much of a difference. As for familiarity of songs, as already discussed, listener types other than Majority had vastly different responses for both types of activities. These responses were also not significantly different between attention and non-attention activities for any of these outer listener types. Lastly, Collectors, Passives, and Musicals had high preference for various moods even during attention activities, so a move to non-attention ones did not produce a significant difference.

Popularity of online music services: Participants were asked about both their favourite online service and the ones they had ever tried. They were instructed to check any of 8 available choices (last.fm, iTunes Genius, Grooveshark, Zune Smart DJ, Pandora, Spotify, iLike, and Musicover), and provide any other services they may have had used in an “other” field.

Out of all participants, 21.4% had never tried any service, 36.8% had tried only one, 23.1% had tried two, and 12.2% had tried 3. Despite having the same median of one service tried, males had tried significantly more services (Mann-Whitney U, $U(N = 652) = 46984.5$, $Z = -2.572, p = 0.010$). Older participants had also tried a larger number of services, such that the median for age group 3 rises to two services. An overall significant
effect, caused mainly by the difference between groups 1 and 3, was observed (Kruskal Wallis, $\chi^2(2, N = 649) = 14.587, p = 0.001$). No significant difference was found among listener types.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Tried</th>
<th>Favourite</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Last.fm (43%)</td>
<td>Last.fm (10%)</td>
</tr>
<tr>
<td>2</td>
<td>iTunes Genius (27%)</td>
<td>Grooveshark (8%)</td>
</tr>
<tr>
<td>3</td>
<td>Grooveshark (25%)</td>
<td>iTunes Genius (5%)</td>
</tr>
<tr>
<td>4</td>
<td>Pandora (17%)</td>
<td>Pandora (5%)</td>
</tr>
<tr>
<td>5</td>
<td>Spotify (16%)</td>
<td>Spotify (4%)</td>
</tr>
</tbody>
</table>

Table 4.2: Ranking of online music services. Tried: Most tried services. Favourite: Top services chosen as favourites.

When asked about their favourite service (with the same choices as above), 59.5% of all participants said they didn’t normally use these services. This low usage is also apparent in Figure 4.4 where online recommendation services have low popularity. Besides having tried more services, the older age group also used these services more. The percentage of non-users dropped to 50% for this group as opposed to 64% for younger groups, causing a significant effect (Pearson’s Chi Squared, $\chi^2(2, N = 519) = 9.508, p = 0.009$). A similar effect was visible with Actives and Musicals, but it did not amount to a significant difference.

Table 4.2 shows the top tried and favourite services for all respondents, for the provided choices. Considering the similarity of these two rankings (save for a switch between ranks 2 and 3), and the generally low number of tried services, it is possible for the choice of favourite to be highly dependant on the services one has tried rather than actual preference for a service over others. The top “other” tried service was Youtube, mentioned a total of 13 times. Simfy was also popular, but only in Sample 2 with 7 mentions.

4.3 Category 2: music collection management

Music sources: The sources under question were physical media, radio (including online stations), one’s collection on computers, one’s collection on portable devices, and interactive online services. For each of these, respondents had to choose from a five point Likert

\[^{d}\]The mailing list at Ludwig Maximilians University in Munich, Germany
scale from “never” to “very often” regarding how often they used them. Looking at all participants, computers and portable devices end up as the most popular sources. Table 4.3 describes the distributions of the scores for each of these sources for all participants.

<table>
<thead>
<tr>
<th>Source</th>
<th>Median</th>
<th>Mean</th>
<th>Std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>One’s music collection on computers</td>
<td>4</td>
<td>3.96</td>
<td>1.12</td>
</tr>
<tr>
<td>One’s music collection on portable devices</td>
<td>4</td>
<td>3.92</td>
<td>1.29</td>
</tr>
<tr>
<td>Radio</td>
<td>3</td>
<td>3.16</td>
<td>1.3</td>
</tr>
<tr>
<td>Physical media</td>
<td>3</td>
<td>2.68</td>
<td>1.22</td>
</tr>
<tr>
<td>Online services</td>
<td>2</td>
<td>2.33</td>
<td>1.44</td>
</tr>
</tbody>
</table>

Table 4.3: Median, mean, and standard deviation of popularity of music sources. Scale is from 1 (“never”) to 5 (“very often”).

The ranking of sources was the same in all demographic sub-populations, however, there were numerous significant differences between the popularity of various sources in these groups. Physical media and radio were both rated much higher by females than by males. On the other hand, males rated computers and online services significantly higher (Mann-Whitney U, physical media: $U(N = 652) = 63818, Z = 4.682, p = 0.000$, radio: $U(N = 652) = 70992, Z = 7.719, p = 0.000$, computers: $U(N = 652) = 45238, Z = -3.377, p = 0.001$, online services: $U(N = 652) = 48877, Z = -2.595, p = 0.009$). Several differences were observed between age groups as well. Use of radio increased with age, such that group 3 scored radio considerably higher than both groups 1 and 2. A similar effect was observed for usage of online services. On the contrary, the popularity of computers and portable devices declined with older age. (Kruskal Wallis, radio: $\chi^2(2, N = 649) = 28.691, p = 0.000$, online services: $\chi^2(2, N = 649) = 8.514, p = 0.014$, computers: $\chi^2(2, N = 649) = 9.222, p = 0.010$, portable devices: $\chi^2(2, N = 649) = 9.257, p = 0.010$).

There were significant overall differences among listener types regarding computers and portable devices (Kruskal Wallis, portable devices: $\chi^2(4, N = 572) = 14.208, p = 0.007$, computers: $\chi^2(4, N = 572) = 17.154, p = 0.002$). Both were caused by Musicals rating these sources higher than Majority. We also noticed a couple of other peculiarities in the responses, which we thought would be interesting to point out, despite not being deemed significant enough by the Kruskal Wallis test. Nearly 40% of Actives rated their physical media use as “very often”. This is in stark contrast with Majority, from which only 9% had the same answer. Also, 40% of Passives gave the “very often” score to radio, whereas only
18% of Majority did the same.

**Size of personal music collections:** Collection size was asked in number of songs, number of physical media, or number of gigabytes (respondent’s choice). Number of gigabytes and physical media were then converted to number of songs assuming an average 4MB size for each song and 12 songs per album. The median number of songs for all participants was 4650. Females had a median collection size of 3180, whereas for males, this was 6000. Difference between gender groups was found to be significant (Mann-Whitney U, \(U(N = 646) = 40971.5, Z = -4.620, p = 0.000\)). The size of music collections had a positive correlation with age, resulting in a significant effect (Kruskal Wallis, \(\chi^2(2, N = 644) = 43.132, p = 0.000\)). Age groups 1 to 3 had medians of 2800, 3913, and 7906 songs respectively.

**Correctness of meta-data (tags) in personal collections:** Participants were asked about how likely they were to correct wrong or inaccurate tags in their collection. The choices were: (1) Probably not, (2) Only my favourite songs, (3) Only if it makes a song hard to find, (4) I always correct/My tags are already correct. Overall, the most popular answer was 4, which was also the median. Our data shows that males were more keen on having correct tags in their collections (Mann-Whitney U, \(U(N = 623) = 43963, Z = -2.276, p = 0.023\)). The youngest age group also showed a stronger desire for correct tags than both other age strata (Kruskal Wallis, \(\chi^2(2, N = 619) = 17.283, p = 0.000\)).

**Number of playlists maintained:** The median response for number of playlists maintained was “2 - 4 playlists”. A significant overall effect was found between age strata (Kruskal Wallis, \(\chi^2(2, N = 635) = 8.843, p = 0.012\)). Pairwise comparisons showed the difference to be caused by group 3 having a higher number of playlists with a median of “5 - 10”.

**Use of manual folder structures versus applications, for collection management:** Participants were provided with three choices: (a) Prefer manual folder structure, (b) Prefer applications like iTunes and Media Player, and (c) Both. These got 49%, 29%, and 22% of the responses respectively. The middle age group was found to prefer manual folder management more than others, in exchange for lower popularity of choice (b) (Pearson’s Chi Squared, \(\chi^2(4, N = 641) = 11.542, p = 0.021\)). Interestingly, manual folders were more popular with all outer clusters than with Majority, by at least 12%. Passives and Collectors had the largest deviations, choosing manual folders 25% and 20% more than Majority respectively (Pearson’s Chi Squared, \(\chi^2(24, N = 570) = 49.046, p = 0.002\)).

**Primary music management/playback software and issues faced using them:**
Participants were asked to choose one of iTunes, Windows Media Player, or “other” as their primary listening/management application and were given a multiple choice question on difficulties they had faced with it. iTunes came out on top with 49%. Next were “other” with 27.5% and Media Player with 23.5%. We found significant differences among both our age and gender groups. The “other” choice was much more popular with males than with females, who also chose Media Player more than males (Pearson’s Chi Squared, $\chi^2(2, N = 605) = 31.442, p = 0.000$). The popularity of Media Player declined with older age too (Pearson’s Chi Squared, $\chi^2(4, N = 597) = 17.527, p = 0.002$). We provided eight multiple choice options for difficulties faced using the chosen primary software. The results for all participants are shown in Figure 4.7. Winamp was the most popular “other” software with 34 mentions. Next were Foobar2000 with 18, and VLC Player with 9 mentions.

4.4 Open ended question on most important features missing from current music listening tools

There were a total of 230 responses to this question, out of which, 14 said they had no opinion or that current tools were good enough. We took a coding approach to identify the most salient concepts from the open ended responses. In the process, some concepts that we thought were similar enough or could have risen from similar viewpoints were grouped
CHAPTER 4. RESULTS

into a more general category.

The leading category, with 33 supporters, was improvement in recommendation systems. Participants touched on various aspects of this issue including recommendation of unpopular music that one may like (the well known “long tail” issue), accuracy of recommendations, and covering non-western music more robustly. One participant explained that “the programs suggesting music won’t know that something unheard-of suits your tastes if it doesn’t find enough people who match your taste.” Another complained how recommender systems do not seem to take into account music that one actively dislikes, saying: “They do not infer which artists or instruments I don’t like.”

For 29 respondents, searching, sorting, and meta-data management were the most lacking departments. A particularly prominent theme was a need for more accurate tags, automatic tag correction and music identification, multi tag searching, and more efficient ways to add or edit meta-data. Six participants expressed a desire for adding arbitrary keywords to files and being able to search and sort their libraries based on them.

The next top category was what we named “audiophile features”, which included 25 responses. Support for customizable equalizer and popular high quality formats, for instance FLAC, were the top missing features according to these participants.

Smart playlists were the fourth most popular concept with 15 participants. Intelligent playlists based on listening history, user created playlists, actively listened music, and context were among the points made by these participants. One participant remarked that “I want to make a playlist by adding a huge pile of songs, then the program should order the songs automatically so the pacing between songs of up and down tempo is pleasing (there are various pleasing patterns to this).”, as opposed to having a playlist created from scratch.

Other Concepts with at least 5 supporters were more convenient multi-device library management (9), mood meta-data (8), clear uncluttered interfaces (7), emphasis on listening history and listen count (7), additional info like bio, context of lyrics, etc. (7), customizable interfaces (6), and query by melody (5).
Chapter 5

Discussion

Some believe that music listening should come with the least amount of interaction possible. Efforts have been made to predict listeners’ preferences based on the context of listening, so as to relieve them from the burden of choice. While these systems may be successful to some extent, the argument here is that this may not be the optimal solution. As studied by DeNora [7], the choice of music and the reasons that lie behind it can be unique to each listener. Thus, it is hard to believe that one can rely on context or inferred user conditions to tell the complete story. Moreover, the responses to the question about varying or constant moods confirm that even one session of listening is often not supposed to be a monotonous experience. The results of this study indicate that the choice of music should be ultimately left to the user, but the process of selecting what music to play needs to be simplified.

Of all the respondents, 86% said they did listen to music during activities that needed their attention, and seeing as we found work to be the second top activity during which music is listened to, we can conclude that attention needing activities account for a large chunk of the respondents’ listening hours. Out of this 86%, 33% expressed a need for varying moods even during attention needing activities. Changing moods, which was the task targeted by the interaction tolerance questions, is one example of adjusting control. As seen in Figure 4.5(a), 55% of participants were only willing to have “very low” interaction to achieve this goal. Our example for this level of interaction was skipping a song, but skipping can not be a proper solution for adjusting control. That is because skipping is not an experience defining act. With skipping, the listener has no control over what plays after the current song is done. Even worse, if one can not find a desirable song right away, consecutive skipping can become tedious and time consuming and break out of the limits
of “very low” interaction. Another conventional adjusting control method, choosing songs manually, is equal to our “high” interaction, which is out of the limits of interaction tolerance for 94% of the respondents during attention activities. Likewise, the “medium” level was not acceptable for 80% of the respondents. Therefore, we are left with “low” interaction as the sweet spot for adjusting control during attention activities.

The picture changes when we look at non-attention activities. Here, 63% of the participants were willing to have at least “medium” interaction. That said, the example we used for defining this level of interaction (switching to another playlist) is only practical if we assume the listener has the suitable playlist ready every time a change of mood is needed. This assumption is probably not true however, as our median respondent had only between 2 and 4 playlists. In a more generalized sense, we can take the “medium” level to be any form of interaction that is not as extreme as having to choose songs one after another, but needs more effort than specifying what needs to change about the currently playing song (low interaction). Our results show that for non-attention activities, this level of interaction is also acceptable.

It is worth noting that requiring attention is just one factor that can limit the possibility of interaction, but not the only one. Driving, walking, and public transport are the main forms of commuting, which was the top activity accompanying music. While driving is an activity that needs a fair amount of attention, public transport and walking can also impose limits on one’s interaction with a music source. Liikkanen et al. [20] confirm this for walking commuters in public transport hubs, reporting that a very small percentage of them were seen interacting with their devices.

Putting it all into perspective, it appears as though the largest part of our respondents’ listening hours happen in situations were high interaction is not plausible. And still, in most of these situations, the need for adjusting control is present. The least common denominator between attention and non-attention activities was found to be “low” interaction. This suggests that there is a need for new interaction techniques that fall into the “low” category and provide enough adjusting control. The starting point can be based on some seed attributes like how current recommendation services work. As the listening experience moves forward however, the users needs to be able to steer and tune their path in the musical space to a desired direction. This interaction should not force the user to find and choose songs, but rather facilitate real-time control over attributes that need to be controlled.

These results also give us ideas on what these attributes are. Other than mood which
we discussed above, we found that participants listened to familiar songs much more than new songs, so familiarity is one other important attribute that needs to be controlled, but is mostly neglected in current music services. Tempo, genre, and how distracting the songs are were also found to be important. While things like familiarity and tempo may be easier to measure and control, we need to come up with ideas on how to let users manipulate the likes of mood and genre easily and efficiently, within the limits of “low” interaction. We also need to keep on looking for other attributes that are both important and understandable for users.

It can be argued that smart playlist creation interfaces that allow manipulating the above attributes but in form of initial control can also be the solution, and thus, they nullify the need for new adjusting techniques. The issue is that with such an interface the users need to specify any alterations they might desire in the course of one listening session before starting it, along with the time each variation should occur. Although this survey did not ask about this specific issue, judging by the responses to the questions on interaction tolerance, it is likely that such a task would be too taxing and time consuming.

For the listener type clustering, three basic characteristics of listeners were used which a tool can easily figure out after a few days of use. Yet, there were several differences in behaviour between the clusters. For example, instruments used in the songs were of much higher importance to Musicals than other listener types for playlist creation, and Collectors were open to a higher level of interaction than others. We also found several differences between demographic sub-populations. For instance, males’ interaction tolerance and desire for control were higher than females, especially during attention activities. These indicate a potential for designing adaptive music listening interfaces matching each user’s characteristics.

Better and more accurate recommendations were the most prominent concept mentioned by our participants in response to our open-ended question on important missing features of current music listening tools. Current music services are quite rigid when it comes to recommendation. Most services operate based on an unexplained “similarity” of artist or genre. More importantly, there is often no way for adjusting the experience. These could be some of the reasons behind the limited popularity of these services, along with regional or device specific access restrictions. In this regard, explanations that Pandora provides for recommending each song and the ability to alter the experience through mixing seed attributes are steps in the right direction. The importance of these explanations has also
been emphasized by previous studies [21].

With portable devices and computers being nearly equally popular sources of music, we have to focus on interaction techniques that are scalable between small and large screens. The portable front is arguably more challenging as we need to figure out how a touch enabled screen that is around 4 inches in diameter can provide support for manipulating all the above attributes without becoming too cluttered. One potential solution might be to present the musical space to the user in different separate views instead of trying to present it all at once. The touch-screen is however not always accessible. Situations that do not allow high interaction are often the same ones where taking the device out of a pocket is not desirable, for instance, while exercising or walking in a crowded area. Thus, we may need to consider techniques that let users apply adjusting control while the device remains in their pockets.

In summary, the results suggest that one needs to have the following guidelines in mind when designing a modern music listening tool:

• Real-time adjusting control should be supported. Attributes that need control include mood, tempo, genre, familiarity, and how distracting the songs are.
• The interface should be scalable to both portable devices and computers.
• Interaction should not take much more time than skipping songs. In many situations, finding and selecting songs, albums, or playlists can be too much.
• It is possible for the interface for controllable attributes to adapt to the users based on the size of their personal music collection, number of listening hours, age, and gender.

It is important to note that the respondents’ answers for how important the songs are, that is, the amount of desired control could have been influenced by the type of control that current interaction techniques can offer them. In other words, our questions might not have been able to completely isolate desired control from what is currently possible. Nevertheless, we still see a higher desire for control than what current techniques can afford, especially in the case of attention activities. Therefore, it is reasonable to believe that had our questions done a better job of making the issue clear, we would have seen even stronger disparity between desired control and current available interfaces.

Another possible issue with our data can be the lack of information on the respondents’ occupation. One’s work circumstances can greatly dictate various aspects of music listening, specially the amount and type of interaction possible. Looking back at the participants, we also need to acknowledge the fact that the majority of them (94%) had a college or higher
degree. Our participant pool was also relatively young with a median age of 24. Also, there was a high concentration of participants with computer science background in Samples 3\textsuperscript{a} and 4\textsuperscript{b}. These issues can bias the results in favour of state of the art technology. That said, the same generation of people are often the target of novel interfaces, so the bias may not be that harmful. Moreover, we believe that our relatively large sample size places our results among the most reliable that the academic community has been able to gather in this field. A next step would be to strive for truly representative samples from more diverse populations. This may however require the cooperation of relevant corporations with access to such data and participants, as without the aid, it may be very difficult to reach out to populations not confined within the limits of academic societies.

\textsuperscript{a}The researchers’ friends in various social networks

\textsuperscript{b}Mailing lists including Computing Science and Engineering students and faculty of Simon Fraser University
Chapter 6

Conclusion

We reported and analysed the results of an online survey on music listening and management behaviours and presented implications for design of modern music listening tools. The survey had a total 744 participants across 7 populations. Out of these 590 were chosen for analysis and the rest were removed due to high non-response. We specially focused on the methods of playback currently used, how precise of a control participants desired on their music listening experience, and how much interaction they would be willing to have with their music source in order to exert that control. We also compared various groups of participants with each other and found several significant differences between them, some of which we point out in this section. The groups included demographic sub-populations of age (age ≤ 21, 21 < age ≤ 27, and 27 < age), gender, and a “listener type” concept that we defined based on three basic characteristics of participants, namely, active[a] and passive hours[b] of music listening per day and size of music collection.

Most participants had an average of less than 1 hour of active listening per day and 2-4 hours of passive listening. While male participants had significantly more active listening hours, their median was identical to females. The most popular activities accompanying music were found to be commuting and work. For females however, work was much less popular and was overshadowed by commuting. Work also gained popularity as we went from younger to older age. The top methods of playback for listening during these activities were found to be “a prepared playlist or folder of songs” and “picking an artist, album, or

[a]Listening to music while being focused on the music, and not doing another activity.
[b]Listening to music during other activities, e.g., work.
genre”. This indicates a desire for a certain level of precision in control on the music which is more than what methods like shuffle can provide, especially during work.

Studying the participants’ interaction tolerance in conjunction with their desired amount of control, we concluded that there is a need for novel interaction techniques which facilitate manipulating various attributes of songs on the fly, with minimal required interaction. We found that some of these attributes are mood, familiarity of songs, how distracting the songs are, and tempo. Demographic groups showed different levels of interaction tolerance and control desire. Males were more open to interaction than females. The same was true for the middle age group compared to the younger and older groups. Males also had a higher desire for control than females, which is in line with their higher interaction tolerance. Also, the need for control declined with older age.

The median music collection size was found to be 4,650 songs. Males had significantly larger collections with a median size of 6,000 songs, as opposed to 3,180 for females. Older participants also had larger collections. Most respondents maintained 2 - 4 playlists, except for the oldest age group which had a median of 5 - 10 playlists. The majority of participants did not like to have any incorrect tags in their libraries and this was stronger in males. Manual folder structures were more than twice as popular as applications when it came to managing libraries. In general, computers and portable devices were the top sources for music playback. We did however find that females used physical media and radio much more than males. Males on the other hand rated computers and interactive online music services significantly higher than females. In addition, use of radio and online services increased with age, whilst computers and portable devices declined in popularity.

We had five listener types, one of which was made up of the majority of respondents. We named the other four types Actives, Collectors, Passives, and Musicals (Table 3.1). Actives were participants with relatively high active listening hours but low passive hours. Collectors were participants with mildly high passive hours and much larger than majority collection size. Passives, as the name suggests, where participants with very high passive hours. Finally, Musicals were those with high active and passive hours.

Musicals were the youngest listener type with a median age of 22 (Table 4.1). Next were Majority with a median of 24, Active and Passives with medians of 26.5, and Collectors with a median of 29 years. As for gender distributions, there was a high concentration of males in Collectors (19 to 3) and Actives (14 to 8). The rest of the types however consisted of rather similar numbers of males and females.
With the majority type being the largest of all groups, their responses were always similar to the whole participant pool. Comparing the other four types to the majority, we found that work as an activity accompanying music was not as popular with Actives and Musicals as with the majority. This is while Collectors and Passives rated work significantly higher than the majority. As for methods of playback, Collectors preferred “choosing an album, artist, or genre” significantly more than other listener types and had much lower preference for playlists and radio. Looking at interaction tolerance and control, once again, Collectors were different, leaning towards more interaction and more precise control. They also had higher preference for songs with varying moods during one session of listening. In the end, Collectors were the group with the largest number of deviations from the majority. These differences suggest that, by having even basic knowledge about the users, we may be able to design adaptive music listening interfaces tailored to their needs.
Appendix A

A.1 The Ranked Pairs method

Ranked Pairs is a method for summarizing ranked choices. In our case, participants were asked to rank four activities that could accompany music listening, based on how large of a share of their music listening hours the activities accounted for. In Ranked Pairs, choices are scored against each other according to their placements in the rankings provided by respondents. For instance, each time commuting appears before work, one point is added to the tally “commuting vs. work,” and each time work appears before commuting, one point is added to the opposite tally, “work vs. commuting.” Having an n*n matrix of tallies where n is the number of choices, each tally can be treated as a graph edge from the winning choice to the losing choice. That way, “commuting vs. work” would be an edge from commuting to work. Edges are then sorted in descending order and added to a directed graph beginning from the largest tally. Edges are added until adding more would create a cycle in the graph. Now each node’s final score is its number of outgoing edges minus incoming edges.
A.2 Multivariate outlier removal with Mahalanobis distance

In a univariate space, the observations with the largest Z scores can be counted as outliers. The Z score\(^a\) is defined as the Euclidean distance between one observation and the sample mean, divided by the sample standard deviation:

\[
    z_i = \frac{x_i - \mu}{s}, \tag{A.1}
\]

In a similar fashion, the Mahalanobis distance of a multivariate observation from the multivariate sample mean is equal to:

\[
    MD_i = \sqrt{(x_i - \mu)^T C^{-1} (x_i - \mu)} \tag{A.2}
\]

Where \(C\) is the covariance matrix of all variables and \(\mu\) is a vector consisting of means all variables. The observations with larger than threshold distance are then regarded as multivariate outliers. This threshold is often arbitrarily chosen. In our case, we used the chi-squared distribution.

Assuming multivariate normality, the squares of Mahalanobis distances follow a chi-squared distribution with degrees of freedom equal to the number of variables. For each observation, we can then find out the percentage of observations that would have less than or equal values, using the CDF\(^b\) of the corresponding chi-squared distribution. The cases with a higher than threshold percentage would then be our multivariate outliers. With visual aid from the three dimensional scatter plot of Z scores for active listening hours, passive listening hours, and collection size, we found a threshold value of 99% to be the most appropriate for our analysis.

\(^a\)Also known as the standard score

\(^b\)Cumulative Distribution Function
A.3 Tables

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Population Description</th>
<th>Sample Size</th>
<th>Male</th>
<th>Female</th>
<th>Min Age</th>
<th>Median Age</th>
<th>Max Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A mailing list for research in auditory perception</td>
<td>38</td>
<td>19</td>
<td>16</td>
<td>21</td>
<td>31.5</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>An opt-in university wide mailing list for general announcements at Ludwig Maximilians University in Germany</td>
<td>282</td>
<td>97</td>
<td>178</td>
<td>18</td>
<td>22</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>Our social networks (Facebook, Twitter, a mailing list of friends)</td>
<td>79</td>
<td>52</td>
<td>23</td>
<td>14</td>
<td>31</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>A mailing list including all Computing Science and Engineering students and faculty of Simon Fraser University</td>
<td>143</td>
<td>104</td>
<td>34</td>
<td>17</td>
<td>23</td>
<td>67</td>
</tr>
<tr>
<td>5</td>
<td>Professional sound and music engineers</td>
<td>14</td>
<td>11</td>
<td>2</td>
<td>21</td>
<td>33</td>
<td>52</td>
</tr>
<tr>
<td>6</td>
<td>A mailing list for research in music information retrieval</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>21</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>Radio professionals</td>
<td>24</td>
<td>11</td>
<td>11</td>
<td>23</td>
<td>40</td>
<td>54</td>
</tr>
<tr>
<td>All</td>
<td><strong>participants</strong></td>
<td><strong>590</strong></td>
<td><strong>302</strong></td>
<td><strong>266</strong></td>
<td><strong>14</strong></td>
<td><strong>24</strong></td>
<td><strong>78</strong></td>
</tr>
</tbody>
</table>

Table A.1: Demographic characteristics of samples, after pre-processing. Participants were allowed to not answer any of the questions, including demographics. That is why the sum of the numbers of males and females may not be equal to the number of participants.
### Table A.2

Difference between attention and non-attention scenarios for interaction tolerance, importance, familiarity, and mood variety. For interaction tolerance, importance, and mood variety, the marginal homogeneity test is used. For familiarity, Pearson’s chi squared test is used. Non-significant p-values are highlighted with grey background.

<table>
<thead>
<tr>
<th>Sub-population</th>
<th>Interaction Tolerance</th>
<th>Importance</th>
<th>Familiarity</th>
<th>Mood Variety</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std. MH</td>
<td>p</td>
<td>Std. MH</td>
<td>p</td>
</tr>
<tr>
<td>all participants</td>
<td>-14.21</td>
<td>0.000</td>
<td>8.06</td>
<td>0.000</td>
</tr>
<tr>
<td>females</td>
<td>-10.87</td>
<td>0.000</td>
<td>7.14</td>
<td>0.000</td>
</tr>
<tr>
<td>males</td>
<td>-10.66</td>
<td>0.000</td>
<td>5.20</td>
<td>0.000</td>
</tr>
<tr>
<td>age ≤ 21</td>
<td>-9.55</td>
<td>0.000</td>
<td>6.83</td>
<td>0.000</td>
</tr>
<tr>
<td>21 &lt; age ≤ 27</td>
<td>-8.76</td>
<td>0.000</td>
<td>4.43</td>
<td>0.000</td>
</tr>
<tr>
<td>age &gt; 27</td>
<td>-7.46</td>
<td>0.000</td>
<td>3.55</td>
<td>0.000</td>
</tr>
<tr>
<td>Majority</td>
<td>-12.55</td>
<td>0.000</td>
<td>6.68</td>
<td>0.000</td>
</tr>
<tr>
<td>Actives</td>
<td>-3.051</td>
<td>0.002</td>
<td>2.10</td>
<td>0.036</td>
</tr>
<tr>
<td>Collectors</td>
<td>-2.86</td>
<td>0.004</td>
<td>2.61</td>
<td>0.009</td>
</tr>
<tr>
<td>Passives</td>
<td>-3.02</td>
<td>0.003</td>
<td>1.92</td>
<td>0.055</td>
</tr>
<tr>
<td>Musicals</td>
<td>-3.32</td>
<td>0.001</td>
<td>1.64</td>
<td>0.101</td>
</tr>
</tbody>
</table>
A.4 Figures

Figure A.1: Histogram of number of non-answered questions out of 24. Demographic questions were excluded in this analysis. The histogram shows all the initial 744 participants, that is, before removal of participants with high non-response. The cut-off point was chosen to be 5 questions and participants who had not answered more than 5 questions (N = 154) were removed from further analysis. This left us with 590 participants.
Appendix B

B.1 The survey

This section contains the actual online survey used in our study. The survey consisted of 11 pages of questions, 1 welcome page, and 1 submission confirmation page. The figures that follow depict the survey pages in the same format presented to respondents. In cases were a survey page is too long for one thesis page, it is broken into two consecutive figures.
Music Listening Survey

Thank you for coming!

We are researchers at the School of Computing Science at Simon Fraser University (SFU) in Canada, and Department of Informatics at Ludwig Maximilians University Munich (LMU) in Germany.

We are conducting this survey about music listening habits which consists of questions on listening hours, preferred devices, use of playlists, active or passive listening, and so on. Our ultimate goal is to create a new tool to make music listening more fun, simple, and intuitive.

Your participation will be anonymous, as we will not record a name, email address, IP address, or anything else that can identify you from your answers. Your connection to this website is secured and encrypted. For maximum confidentiality, we will also remove any potential identifiers from responses (like in the comments you make) and store the raw data in a locked cabinet. The anonymous data will be made available online for other researchers interested in the topic.

You can cancel your participation in the survey at any time by clicking "Exit and Clear Survey" at the bottom right of the page. This will also clear your responses and no record of your participation will be saved. If you wish to keep record of this page, please save or print it. Participation in this survey is entirely voluntary. Your decision to participate or not, or your decision to cancel this survey, will have no effect on your grades (if you are a student) or your job (if you are employed). This is a minimal risk study, which means the risks involved with participation, if any, are not greater than those ordinarily encountered in daily life.

If you wish a copy of any reports or papers from this survey, they will be available at: http://gruvi.cs.sfu.ca/projects/dynamic-musical-experience.

By clicking "Next" and filling out this survey, you are agreeing with the above and consenting to participate.

In case of any questions, comments, or complaints please contact my research supervisor:

Dr. Torsten Möller
office: TASC 8021, Burnaby campus
Simon Fraser University
Burnaby, B.C. Canada
+1-778-782-2215
http://www.cs.sfu.ca/~torsten/
torsten@sfu.ca

Principal Investigator:
Mehsan Kamalzadeh
M.Sc. student of Computing Science
Simon Fraser University, B.C. Canada
mkamalza@sfu.ca

LMU collaborator:
Dominikus Baur
Diplom Medieninformatiker
Department of Informatics
Ludwig Maximilians University Munich, Germany
http://www.medien.ifi.lmu.de/team/dominikus.baur
donnikus.baur@ifi.lmu.de

Figure B.1: Survey welcome page - part 1
Figure B.2: Survey welcome page - part 2
Figure B.3: Survey page 1
**Music Listening Survey**

How often do you listen to collections taken from friends or downloaded online? (including playlists, mixtapes, mixCDs, folders, etc)

- Never
- Rarely
- Sometimes
- Often
- Always
- No answer

Do you ever create playlists or selections of songs (like folders)?

- Yes
- No
- No answer

How important are each of the following for you, when creating a playlist?

<table>
<thead>
<tr>
<th></th>
<th>Not Important</th>
<th>Somewhat Important</th>
<th>Very Important</th>
<th>No answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tempo</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Album</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Genre</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Mood</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Instruments used in the song</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Artist</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
</tbody>
</table>

Other:

Figure B.4: Survey page 2 - part 1
APPENDIX B.

Figure B.5: Survey page 2 - part 2
Figure B.6: Survey page 3 - part 1
Figure B.7: Survey page 3 - part 2
Figure B.8: Survey page 4
Music Listening Survey

Important:
For all the questions in this page, please imagine you want to listen to music during an activity that needs your attention like work or studying.

During activities that need attention, which of the following best describes the music you listen to in terms of song familiarity?
- I mostly listen to music I’ve heard before.
- I mostly listen to new songs (for example on radio).
- I don't care if the songs are familiar or not.
- I don't listen to music while doing such activities.
- Other: [space]
- No answer

During activities that need attention, which of the following best describes how important the music you listen to is?
- I choose what to listen to, carefully
- I don't care about individual songs but they should match my mood (sad, happy, dark, etc)
- I don't care about individual songs or their moods, but the music shouldn't distract me from my other activity
- I just want to hear some music, I don't care what's being played
- Other (please provide a short comment)
- No answer

Please enter your comment here:

Figure B.9: Survey page 5 - part 1
During activities that need attention, do you:

- prefer songs with the same mood
- like to have mood changes once in a while
- No answer

During activities that need attention, if you want to change the mood, what is the maximum amount of interaction you would want to have with your music player?

- very low interaction: for example, skipping tracks
- low: for example, specifying your desired change in mood but not having to find any particular song
- medium: for example, switching to another playlist
- high: for example, finding specific songs one after another
- No answer

Figure B.10: Survey page 5 - part 2
APPENDIX B.

Figure B.11: Survey page 6 - part 1
During activities that **don't need much attention**, do you:
- prefer songs with the same mood
- like to have mood changes once in a while
- No answer

During activities that **don't need much attention**, if you want to change the mood, what is the maximum amount of interaction you would want to have with your music player?
- very low interaction: for example, skipping tracks
- low: for example, specifying your desired change in mood but not having to find any particular song
- medium: for example, switching to another playlist
- high: for example, finding specific songs one after another
- No answer

Figure B.12: Survey page 6 - part 2
Figure B.13: Survey page 7
Music Listening Survey

If you see a song with wrong or inaccurate information tags (like artist, album, song title, etc), would you correct them?
- I correct most of the inaccurate or wrong tags that I see in my collection.
- I correct only the tags that can make a song hard to find for me.
- I correct only the tags of my favorite songs.
- As far as I know, most of the tags in my collection are already correct.
- Probably not
- Other: [Text Box]
- No answer

How important are each of the following for you in organizing your music collection or choosing what to listen to?

<table>
<thead>
<tr>
<th></th>
<th>Very Important</th>
<th>Important</th>
<th>Somewhat Important</th>
<th>Not Important</th>
<th>No answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artist</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Album</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Genre</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Mood</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Tempo</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Instruments used in the song</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

Other aspects: (along with their importance)

[Text Box]

Figure B.14: Survey page 8
Figure B.15: Survey page 9
Music Listening Survey

Please estimate the size of your music collection.

Number of physical media you own (Audio CDs, LPs, cassettes, etc):

Size of digital collection: (Please answer at least one of the choices)

- Number of songs:
- Number of gigabytes:

In your opinion, what are the most important features missing from today’s music listening tools?

Figure B.16: Survey page 10
Figure B.17: Survey page 11
Your opinions have been recorded. Thank you very much for your cooperation!
If you would like to help us gather more responses, please help spread the link to this survey:

http://sample-link

Figure B.18: Submission confirmation page
B.2 Invitation e-mail

As pointed out in the Methods chapter, 6 of the samples were recruited through sending invitation e-mails to 6 mailing lists. What follow are the contents of the invitation e-mail.

Subject: Requesting your help with research on music listening

We are researchers at the School of Computing Science at Simon Fraser University (SFU) in Canada, and Department of Informatics at Ludwig Maximilians University Munich (LMU) in Germany. 

We are conducting a study on improving personalized music listening tools and we would really appreciate if you can help us with this survey on music listening habits. The survey consists of questions on listening hours, preferred devices, use of playlists, active or passive listening, and so on. Our ultimate goal is to create a new tool to make music listening more fun, simple, and intuitive.

To take the survey, or to find out more about it, click on the following link: http://sample-link

We assure you that we won’t keep any record of your name, email address, IP address, or anything else that can identify you from your answers.

Thank you,

Mohsen Kamalzadeh, Dominikus Baur, Torsten Möller
Computing Science
Simon Fraser University, Ludwig Maximilians University Munich
Bibliography


