Algorithmic music generation using quantification of cognitive properties, and utilisation in the DAW environment

Alexandros Stamatiadis

Faculty of Arts and Humanities

Department of Music

The University of Sheffield

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

June 2022

Table of contents

Acknowledgments

I would like to begin by thanking my brother, my parents and my extended family for the support and the encouragement. Also for inspiring me, each one in their own special way. Further, I want to thank everyone who filled these 4 years with great memories, and everyone whose presence afforded me the opportunity to grow.

I would of course like to thank my supervisors, Renee Timmers and Guy Brown. Even though it is typical for PhD students to thank their supervisors, I honestly felt lucky to have Renee and Guy who vastly improved me as a researcher. Besides, with their expertise, experience and genuine interest, they managed to help me unify and develop this project into something that immersed my curiosity and will to work, aligning it perfectly to my personal interests and future goals. It was not the most typical project, and therefore, not the easiest task. They also managed to make me conform to academic writing somewhat, which was not the easiest task either.

[Redacted text]

Abstract

This thesis is concerned with cognitive concepts related to melodic preferences, and attempts to capture quantifiable elements of what makes melodies sound pleasing to listeners, through the use of statistical properties and analysis of listener preferences. By using both empirical findings and drawing from literature, these elements were subsequently used as a basis for the development of software that generates musical sequences. The thesis included two studies investigating the perception of melodies, where the concept of a Uniformity Principle (a preference of listeners for melodies with distributionally uniform pitches), and the relation between working memory and liking were examined. Further, a study was conducted with the aim of understanding the usefulness and quality of the software we developed, showing positive results in both of those dimensions.

1. Introduction

Music is an art form that can be remarkably hard to define. If we use the notion of sound, a composition such as 4’33” by John Cage, which is a piece consisting of 4’33” minutes of silence, shows that we would still miss something important. This important element is the view of music as a sequence of time spaces or “time buckets” (Taruskin, 2009) that could contain any amount of sound and any amount of silence in it. Therefore, in the most minimal sense, music is a temporal structure. And being an art form, it is a structure that conveys aesthetic appeal.

This PhD project is an exploration of this aesthetic appeal, through the lens of analysing musical structures with the use of quantitative music cognition. In simpler words, it is an attempt to capture quantifiable elements of what makes melodies sound pleasing to listeners, through the use of statistical concepts and analysis of listener preferences. By using both new insights and drawing from literature, these elements are subsequently used as a basis for the development of a software that generates MIDI musical segments. MIDI stands for “Musical Instrument Digital Interface” protocol (Smith & Wood, 1981). In the modern music making environment of Digital Audio Workstations (DAWs), music composition practices involve the writing of music in a virtual piano roll in the form of MIDI information. The software that was developed as part of this project is a Virtual Studio Technology (VST) plugin that can be embedded into DAWs and act as a compositional aid tool for music makers.

My motivation for starting this project dates back to the days of high school. Between the daily high traffic on my way to school, and my love for the excellent songwriting of bands like *System Of A Down*, I had ample time to contemplate the beauty of very simple yet emotionally powerful repeating melodies that drove many of these songs. Particularly, I was interested in the idea that some melodies were “better” than others, in some kind of objective way. The simplicity of melodies found in popular music arrangements offered a framework with few variables, which could therefore be studied extensively and systematically, offering a possibility to examine this idea. The project is a direct continuation of my curiosity on the topic, supplied with my knowledge of statistics through my BSc studies, and of music cognition in postgraduate studies. Furthermore, as a university student I was driven to work on projects that could have a direct impact on “real life” environments. I was particularly fascinated by the thought of helping people of all skill levels to create music, by using music cognition research to develop an automated music generation tool.

1.1 State of the VST market and our contribution

In the current VST market there are a number of different approaches to software development that lead to distinct categories of products. Specifically, there is a sector concerned with the digital recreation of analog audio equipment, forefronted by companies such as Acustica Audio, Arturia and Brainworx. These companies use digital signal processing techniques to recreate the circuits and sound manipulation behaviour of equipment such as equalisers, compressors, guitar amplifiers and analog synthesisers. Another direction in music software development has to do with the recreation of real life, non-electronic instruments and sounds in the DAW environment, primarily led by Native Instruments, Spectrasonics and Output. The main technique used in this approach is sampling, which refers to recording sounds and arranging the recordings in a way that a musician could play the recorded instrument “inside the box”. Finally, a third section in the plugin market corresponds to software that uses the digital domain to make tools that could not be possibly implemented in the analog domain, and ones that are not recreating any particular hardware device, usually with functions not frequently seen in hardware equivalents. Examples of this approach include vocal pitch correction tools (e.g. Antares

Auto-Tune and Celemony Melodyne), machine learning-based audio extraction software from iZotope and Steinberg, music generation software by companies such as Mixed In Key, and tools oriented towards mixing, such as VolumeShaper by Cableguys, and Pro-Q 3 equaliser by FabFilter.

The importance of pursuing this project, in addition to personal motivation, is established in the untapped potential that is to be found at the intersection of music software development and academic research on the aesthetic quality of sound. Especially in the category of novel VST plugins which do not aim to replicate a hardware module, programmers could potentially gain a significant benefit from the application of knowledge that exists in the academic realm about how listeners perceive musicality and emotion in musical material. Our project demonstrates the development of this type of study and the use of existing ones, for the purpose of creating a novel VST plugin that is further examined for usability and usefulness by a group of testers. Being a music generator, the plugin belongs to a subcategory of products that is underdeveloped and not widely used in the music production process yet. With this work we aim to contribute in this subcategory and highlight the benefits of our approach in music software creation. At the time of writing, this software is under further development for a release as a product with the VST company

[Redacted text]

The deep and decisive interest of this company in our software underscores the contribution that our project can potentially have in the connection between music cognition research and music software development.

1.2 Research aims

This concept of using music cognition research to develop novel music production software requires a diverse context of investigation. Our overarching question is to what degree plugin creation can be assisted by the quantification of aesthetic properties and utilisation of the potential contained in academic music cognition literature. We are particularly interested in areas that can yield quantifiable results when it comes to how listeners experience responses such as aesthetic pleasure, complexity and interestingness, while listening to musical excerpts. Also, we wanted to evaluate methodically the quality and usability of the software that was a result of this research.

Study 1: The effect of pitch distributions in melody liking ratings

In the first study of this project we explored the effect of pitch distributions on liking ratings of melodies, along with perceived complexity and how interesting the melodies sounded. The central concept under investigation was the possible positive effect in liking ratings, of a pitch distribution where the mean pitch is always at the centre of the range of the pitches in the melodies. This idea was called the Uniformity Principle, as it corresponds to a distribution that exemplifies the characteristics of the discrete uniform statistical distribution. The listener's responses to melodies that had pitches conforming to this distribution were compared to responses based on progressively more skewed distributions, in order to test the hypothesis of the Uniformity Principle. During this study we further explored responses on complexity, perceived interestingness and how this data could be influenced by demographic factors of the listeners.

Study 2: The role of working memory in the aesthetic appreciation of melodies

The second study is an expansion to the topic of melodic complexity. Our aim with this study was to examine the effect of complexity on liking, testing the theory of an aesthetically optimal amount of complexity. An inverted U relationship between pleasantness and stimulus complexity has been supported by several studies (Vitz, 1966; Berlyne, 1971; Saklofske, 1975; Farley and Weinstock, 1980; Imamoglu, 2000), and this study is an attempt to model and test the function of complexity in the context of melodic expectation. Additionally, we assessed listeners’ ability to understand the relationship between melodies that have undergone inversion transformations and their original versions. By doing so we aimed at understanding the relationship between cognitive load and liking, as well as the importance of fulfilling melodic expectations in liking, and the interplay between the two pairs of factors. Our results could help materialise these variables in a way that would allow their implementation in the MIDI music generator software.

Development of the MMM Generator and evaluation study

The third phase of the project concentrated on implementing the findings, insights, and concepts from the first two studies into a software plug-in, which we term the MMM Generator, as a reference to the Music Mind Machine research centre of the University of Sheffield. This software was intended to produce quality musical content for the modern DAW-based music making process. After developing an initial version of the MMM Generator, we used it to collect user experience data, and then proceeded with analysing it. The purpose of the software evaluation was to see how effective we were in producing a helpful and high-quality music composition tool by utilising the studies of this project, which focused on quantifying aesthetic and cognitive properties of music, along with pre-existing academic literature on the same topic, and popular music composition practices. The focus of our investigation is on the users' perceptions of the software's usability and their general impressions, both favourable and negative. We also gathered suggestions for improvement, which were taken into account and resulted in an updated version of the MMM Generator that was tailored to the demands based on users' feedback.

1.3 General methodology

The studies that were carried out in this project shared common methods of data collection and sample selection. A percentage of the participants of these studies was recruited from the University of Sheffield via the email recruitment system, and University of Piraeus at which I completed my BSc in Statistics, via Facebook. The rest of the participants were recruited through contact via the social media platform Instagram. The majority of the data collection happened whilst the Covid-19 pandemic had forced governments to enforce national lockdowns, which made online questionnaires the only viable choice of data collection. Google Forms questionnaires provided an optimal solution for this type of task, and allowed the use of Likert scales and video embedding. Studies 1 and 2 utilised these scales for the collection of responses on the perception of melodies. They were used as a rating tool for the assessment of liking and perceived complexity of melodies, the degree of melodic inversion, and so on for the variables under consideration for the two studies. It further allowed testers of the MMM Generator plugin to work in their own music making environment and use the software as part of their normal workflow, which was an important benefit of collecting the data through online questionnaires. The task of the participants in this study was to fill scale-based and free text questions on the usability and design of the MMM Generator, the quality of the musical output, and point out general positive and negative remarks.

For the creation of the melodic stimuli in the perception questionnaires, generation algorithms were developed according to the variables under consideration in each study. For the preservation of a variability in the stimuli, the generators created melodies with shapes that confined to a code-based implementation of archetypal contours, deriving from Meyer's work (Meyer, 1973). In the case of the third study of the project, the main material of consideration was the MMM Generator plugin. The user interface (UI) design of this software was done in a way that gave focus on the direct control of the variables that were under consideration in studies 1 & 2. It also included a variety of features that aided in the usability and musical quality of the MIDI output. In terms of coding, it used the algorithms that were utilised for the creation of the stimuli of the previous studies as an initial basis, and expanded on them. The coding of all generation algorithms in the project, as well as part of the data analysis throughout the studies was done by using the Python 3 programming language. The majority of data analysis on the stimuli and participant’s responses was carried out on SPSS.

1.4 Layout of the thesis

The thesis is organised in three main parts. It contains the current introductory chapter as an overview of the project, highlighting the motivation for its conception, the contribution of it in the intersection of music software and cognition research, and summarising the nature and methods of the studies. The next section consists of three chapters, each of them reporting the three studies of the project respectively. Chapter 2 analyses the investigation on the Uniformity Principle idea, comparing statistical distributions of pitches in melodies, and looking into the variables of complexity and interestingness. In Chapter 3, the focus of the exploration is on perceived complexity, and the associated notions of cognitive load and optimally complex stimuli. Chapter 4 is devoted to the MMM Generator software and it is split into two sections. The first part of the chapter explains the background of algorithmic generation of music in the academic and industry fields, and describes the link between the research under consideration in the current project with the development of the MMM Generator plugin. The remainder of the chapter is devoted to the study of how music producers used the MMM Generator as a part of their creative workflow, allowing us to evaluate its quality, usefulness and points for improvement. Lastly, Chapter 5 summarises the key findings of the project, connecting the main contributions with existing literature on music cognition. It also presents the discussion about limitations and suggestions for future work on the link between cognition research and application development in the music software industry. As a visual map, the structure of the project is represented as follows:

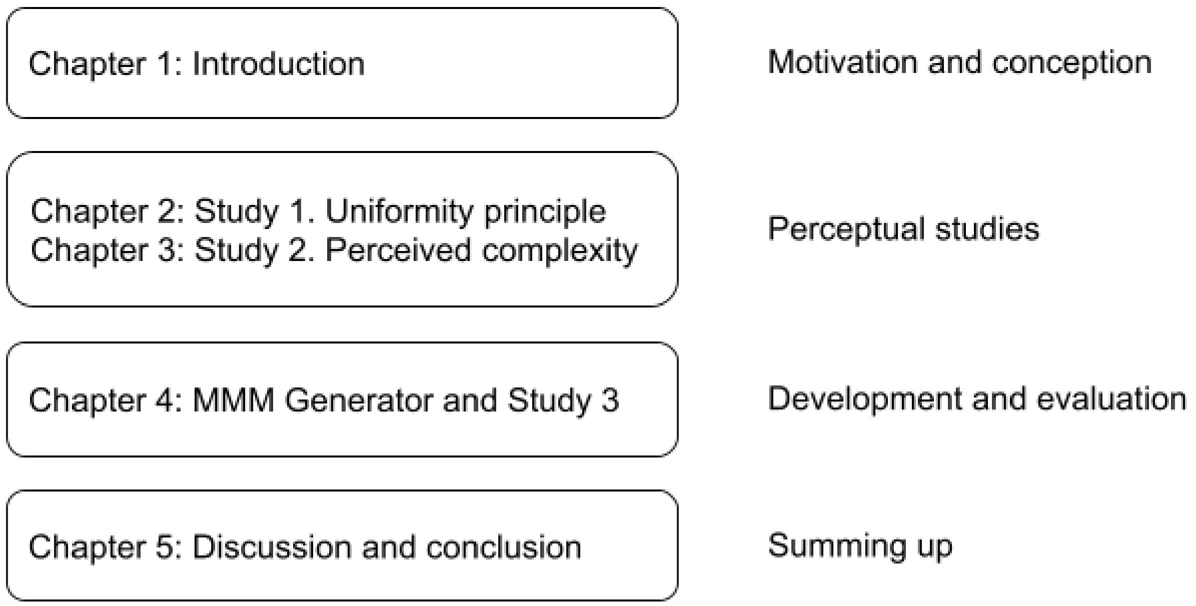


Figure 1: Visual map of the thesis' structure.

2. Study 1: The effect of pitch distributions in melody liking ratings

2.1 Introduction

Our interest in properties of melodies that can present quantifiable results, in relation to how listeners experience aesthetic responses, constituted the motivation for this study. Such properties include psychological principles, like Gestalt laws (Kubovy & van Valkenburg, 2001), and statistical universals (Brown & Jordania, 2013), forming a context of research that could allow the systematic analysis of liking responses. An analysis of this type could be subsequently used for the development of melody generation algorithms, with the purpose of achieving positive responses by listeners. Particularly, the current study focuses on the distributional properties of pitches, and their contribution to the appreciation of melodies.

In music cognition literature it has been shown that listeners are sensitive to the distribution of pitches in music, and a “distributional view” regarding schematic expectations, such as inference about the key of musical pieces, has been given academic support in the context of tonality perception (Levitin, 2002). This idea holds that there exists a cognitive process of comparing template distributions with the ones of a new stimulus, in order to find which archetypal distribution matches the one of the stimulus. This procedure has been modelled in various ways, such as the Krumhansl-Schmuckler algorithm that uses key-profiles derived through probe-tone experiments, in order to determine the key of a piece by examining the correlation of the piece’s profile with the template profiles (Krumhansl & Kessler, 1982; Krumhansl, 1990). Similarly, Chew’s (2002) model uses a 3-dimensional space where each key has a characteristic point, and the finding of the key of a passage is based on the proximity of the average position of the passage’s events to the nearest characteristic key point. Other distributional models are based on the frequency of pitch occurrences and their fit to specific keys, either through notation (Vos & Van Geenen, 1996; Yoshino & Abe, 2004) or audio signal (Martens et. al, 2005). Finally, Temperley (2007) used a Bayesian approach to construct a distributional key-finding model that calculates the probability of occurrence of a melody, given that the melody was generated by using a specific key. Overall, listeners seem to use pitch distributions for key detection (Temperley, 2008), and given that key is an obvious influence in deriving aesthetic value in tonal music, there is evidence that the distribution of pitches can potentially be connected to and influence the liking and aesthetic responses of listeners regarding a particular musical stimulus.

Another area in the literature that focuses on statistical properties of music is the perception of well-formedness of melodies. In this topic, a number of factors have been found to influence or comply with the expectations of listeners and to make melodies sound "satisfactory". One such factor is post-skip reversal, which refers to the observation that a melody tends to change in pitch direction after a big interval occurs. Some proposed reasons for this phenomenon are the psychological need for "filling gaps" (Meyer, 1973; Narmour 1992), and more simply the existence of melodic range constraints. Such constraints are, for example, the possible range of pitches that the guitar can produce due to its length, or the range that the human voice can reach which is limited by the physiology of the vocal cords. This explanation was tested by Von Hippel & Huron (2000) by conducting statistical analysis of a large music dataset. They found that randomly ordered notes still show the effect of post-skip reversal given such range constraints. Other principles include a tendency for small intervals, declination of pitch when melodic intervals are small and rise when they are big, and the continuation of the direction of small intervals. Their role in expectations, either through statistical regularities (Pierce & Wiggins, 2004) or psychological properties, gives rise to the perception of well formedness in melodies that comply to them, and changes in these properties influence the melodies' induced aesthetic experience, as argued by Meyer and Narmour's work and confirmed in the aforementioned empirical studies.

By combining the idea that listeners could be sensitive to pitch distributions and form schematic expectations based on them (Temperley & Marvin, 2008), and evidence that statistical regularities interact with expectations and ease of processing of melodies (Schellenberg et. al, 2002), we can raise the question of how these two kinds of properties could be related. In past studies, researchers have used Zipf's 1/f distribution for note pitches and duration in order to generate melodies, as an attempt to understand the listeners' preferences. In Zipf's 1/f function, the second most frequent observation occurs half as often as the first, the third one occurs at 1/3 frequency, and the nth at 1/n. These experiments on note series that are shaped according to Zipf's distribution showed that listeners had a preference for series that follow a distribution at 1/f, as opposed to 1/f0 and 1/f2 (Voss & Clarke, 1978). Further, 1/f power law is documented to apply in harmony (Wu et. al, 2015), rhythm (Levitin, Chordia & Menon, 2012) and has been used to develop pleasantness metrics (Manaris et. al, 2003).

In view of these studies we asked: Could a deliberate fine-tuning of the parameters of pitch distributions improve ratings of liking in melodies? Do specific pitch distributions sound better than others? Also, how do listeners link liking with other parameters such as complexity and interestingness, and what is the role of a musical background in how individuals perceive pitch distributions? Our study can be seen as a continuation of the research direction that we presented, as it tests the distributional properties of melodies as well, by generating and comparing melodies that comply with different statistical distributions. Namely, our stimuli consisted of melodies that followed the uniform distribution, and 3 other distributions of progressively more skewed shapes. As we are focusing on distributions in the sense of common probability density functions, the rationale for using uniform is that it is the only such common distribution that can be implemented in a range of 12 discrete data points and can provide meaningful deviation of values, due to its property of giving all values the same probability of occurrence. The skewed shape distributions only differ from the uniform in their mean values and, as such, are the best fit to enable direct comparisons between uniform and alternative stimuli.

The melodies of the study use a number of different contours, in order to control for the possible influence of melodic shapes in the results. This is a design aspect of the study that allows us to avoid a scenario where using only one shape for the melodies would influence the results according to how the chosen shape would interact with the rest of the variables. More broadly, the ability of perceiving contour is acquired in early childhood, and it is an important factor in the experience of musicality (Trehub, Bull & Thorpe, 1984; Patel, 2010; Trehub, Becker & Iain Morley, 2015). In the literature of music cognition and ethnomusicology, researchers have observed that it is possible to group melodic contours that are frequently used in compositions into archetypes (Seeger, 1960; Meyer, 1973; Adams, 1976; Hood, 1982). For the stimuli of this study the archetypes that were used are gap fill (post-skip reversal), linear ascending, linear descending, arpeggio (constant up and down motions), and random shapes.

The participants of the study had to listen to these melodies, and rate them in the dimensions of liking, perceived complexity and interestingness. The two extra variables of complexity and interestingness are relevant for their potential use of revealing information about the underlying reasoning behind liking ratings. In past music research, it has been found that interestingness may be positively correlated with pleasantness and positively correlated with complexity when pleasantness is partialled out (Russell, 1982). Since further research has shown a high correlation between pleasantness and liking (Ritossa & Rickard, 2004), the variables of complexity and interestingness are good candidates for showing potential association with liking ratings.

We asked potential participants to take part in our study in order to help us with the overall aim of it, which is the development of a better understanding of how quantitative properties of melodies can influence people's perception of these melodies. The main hypothesis under exploration is concerned with the idea of how a uniform distribution - based generation of melodies could influence liking responses by participants. What we expected to find is that the mean liking ratings of uniform melodies would be higher than those of progressively more skewed melodies. The task of the participants involved listening to melodies through questionnaires, and rating them according to three Likert scales on how much they like each melody, how complex and how interesting the melody is (a sample of the questionnaire is presented in Appendix). We informed them that the results of the study could help in the formation of theory about statistical regularities and their effect on the appreciation of melodies. The study was approved by the ethics committee of the Department of Music.

2.2 Method

2.2.1 Generation of the stimuli

Regarding the generation procedure of the stimuli used in the study, which are the melodies of varying pitch distributions, we present and explain pseudocode in order to illustrate the key points of the algorithm. We used 16 melodies, which consisted of 4 distributions (uniform, and 3 progressively more skewed distributions) x 4 shapes (gap-fill, linear, arpeggio, random). The actual programming was done in Python 3 and the integrated development environment (IDE) that was used is Spyder (Raybaut, 2009). The generation is split in two scripts: The first one contains the functions that do the generation and contour shaping, and the second one is used to export the data to a MIDI file. For ease of programming an initial script was also written that named the MIDI numbers that correspond to notes, as the note names. Next, scales were created through exclusions of off-scale notes, and subsequent mapping of the scales to cover the whole MIDI note range. The remainder of this section explains the melody generation process for the stimuli of the study.

In Box 1 we show a pseudocode version of the Generate function that was used in the creation of the stimuli. We present the case of uniformly distributed pitches, while the remaining types of melodies use the same algorithm, just with a progressively more off-centre target mean than the UniformMean. The LowestNote and HighestNote are variables that take as input the lowest and the highest note that we want our melody to have, in terms of pitch. By keeping these constant for all melodies, we make them all comparable for the experiment as we maintain a constant pitch range. The Scale variable is a list that contains all banned notes, so in other words, it denotes the scale in which the notes are generated from. Keeping the scale constant for all melodies ensures that this aspect is controlled as well. NotesList is the list of final / approved notes of the melody (which initially always contains two notes, i.e. LowestNote and HighestNote). Similarly, the CandidateNotesList is the list of notes that are candidates for inclusion in the NotesList, and it gets updated with every iteration. The LoopCounter variable inputs a note in NotesList if after 20 repetitions, an error reducing note has still not been found in an iteration and the necessary total number of notes has not been reached. This is done in order to avoid an infinite loop, which can happen because sometimes, the event of getting the same mean as the uniform one with a very small number of notes can occur. In such an occasion there can be no optimisation, making the algorithm stuck and unable to fill the remaining slots and reach the number of notes that are required. The number of repetitions (20) was chosen through experimentation, to give enough tries without making the generation too slow.

FUNCTION Generate

INPUT: LowestNote, HighestNote, Scale

OUTPUT: NotesList

STEPS:

1. Initialise the following variables:

UniformMean <- MEAN([LowestNote, HighestNote])
NotesList <- [LowestNote, HighestNote]
CandidateNotesList <- [LowestNote, HighestNote]
LoopCounter <- 0

2. Set CandidateNote as a random note between LowestNote and HighestNote.

3. If CandidateNote is one of the off-scale notes, generate a new CandidateNote. Else add it to the CandidateNotesList.

4. Measure the difference between the mean of the CandidateNotesList list and the

UniformMean.
5. If CandidateNote reduces the difference, add it in the NotesList and go to step 7. Else go to step 6.
6. If LoopCounter < 20, then remove CandidateNote from CandidateNotesList and repeat from step 2 and add 1 to the LoopCounter. Else add CandidateNote in the NotesList.
7. If the mean note of the NotesList is equal to the UniformMean and the NotesList length is greater than 3 or equal to 8, continue to step 8. Else repeat from step 2.
8. Return NotesList

Box 1: Pseudocode of the generation process of uniform melodies.

After the generation process, the Generate function returns as output the final list of notes. This list can now be shaped in various contours in order to fit in the melodic archetypes that will be used in the study. As an example, the gap fill contour function works as follows: It takes as input the output of the generate function (NotesList). It then finds the lowest and highest notes, putting the lowest first and the highest second. Finally, it sorts the remaining notes in a descending fashion from note #3 to the end of the melody. This explanation is further documented through the pseudocode presented in Box 2.

FUNCTION Gap fill

INPUT: NotesList

OUTPUT: NotesList

STEPS:

1. Initialise the following variables:
MinNote <- MIN(NotesList)
MaxNote <- MAX(NotesList)
MinPosition <- NotesList.INDEX(MinNote)

MaxPosition <- NotesList.INDEX(MaxNote)
2. Switch the position of MinNote (note with the lowest pitch), with the note at position 0 of NotesList.
3. Switch the position of MaxNote (note with the highest pitch), with the note at position 1 of NotesList.
4. Sort the notes of NotesList at positions 2 to -1 in descending order.
5. Return NotesList

Box 2: Pseudocode describing the contour shaping procedure for the gap fill shape.

The melody is then exported to MIDI through another script, where the parameters of note durations, tempo, note velocities and file name are being set. For the design of the stimuli we used durations of quarter notes, a moderate tempo at 120 BPM and no dynamics in note velocities, by keeping a constant maximum MIDI value of 127. We then fed the MIDI file into the Kontakt VST plugin by Native Instruments, and used the “Alicia’s Keys” virtual piano library. The rendering of the final audio files was done through the Cockos’ Reaper digital audio workstation. We extracted them to mp3 format in order to keep low file sizes and embedded them in the questionnaire.

2.2.2 Statistical properties of generated melodies

Statistics of the generation algorithms

In order to get insights on the statistical properties of the stimuli, we conducted an analysis on a dataset of melodies generated by the algorithms that were developed for the study. For this reason, the analysis shares similarities to a simulation process, and it aimed to show how the algorithm outputs stimuli that are proper for the objectives of the study. The dataset consisted of 30 melodies generated under the uniformity rule, and 90 melodies generated by using progressively more skewed distributions in two steps. The 60 melodies of the most skewed distribution were split into 30 positively and 30 negatively skewed melodies. The melodies had their lowest and highest notes fixed and equal, in order to render comparisons regarding the behaviour of their means possible.

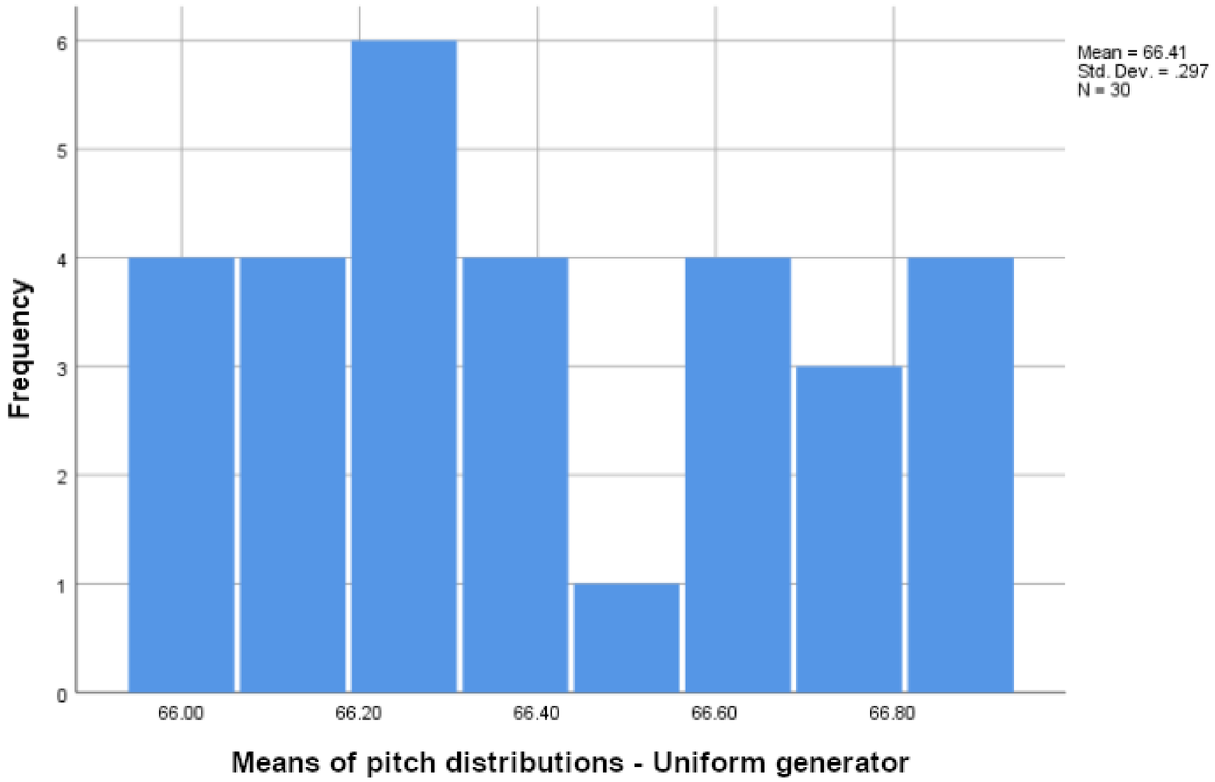
Regarding the distributional properties of the mean pitches of the stimuli, by constituting the "average pitch" of each melody they are values of continuous nature, rather than discrete (they are not MIDI pitches, but an average value of multiple MIDI pitches). In Figure 2 it is shown that the means of the melodies that were generated by the uniformity algorithm are at the 66 MIDI pitch which would correspond to the pitch F#4 (60 = C4, 72 = C5). What is important here is the variation of the means of the melodies. The uniform melodies show an absence of variation of their means, which are always at 66. The figure shows that melodies had means ranging from 66, to 66.80. This displays the effect of the algorithm that replaces generated notes until the melody has the particular ideal mean. At this point it is important to ask, what difference does the "uniform generation" have to that of a "random generation"? After all, a generator of random numbers works via a uniform distribution, since every number has an equal probability of occurrence. We emphasise that with the term "uniform generation" of "uniform melodies" in this project we refer to a procedure that forces the output data to always have a mean value that is equal to the theoretical mean value of the discrete uniform distribution, as shown in (1).

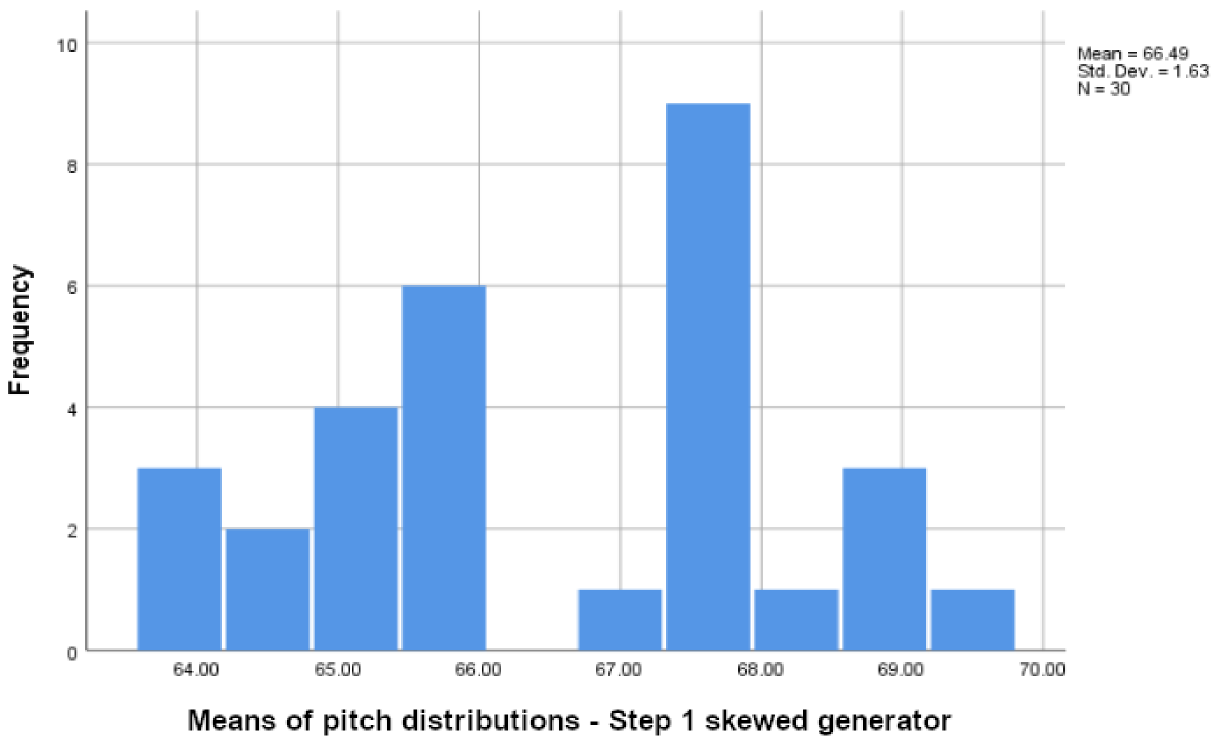
$$\frac{lowest\ pitch + highest\ pitch}{2} \tag{1}$$

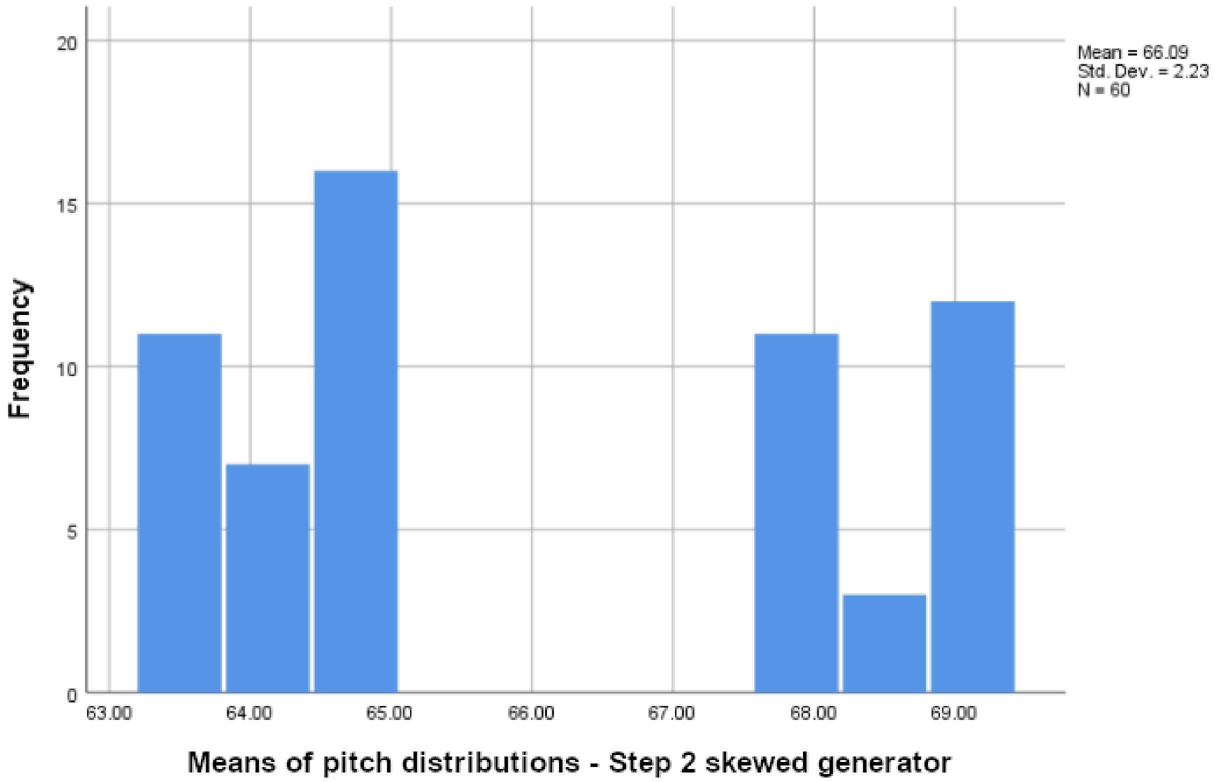
This is different from implementing a generation procedure that uses the uniform distribution without a forced mean, as this would simply constitute a random generation process. Similarly, the skewed generation uses a forced mean that is at distance of discrete values from the theoretical uniform mean. For example, in the case of step 1 skewed generator the mean is described by (2).

$$\frac{lowest\ pitch + highest\ pitch}{2} \pm 1 \qquad (2)$$

The step 1 skewed generator (Figure 3) shows a slight peak of means near the centre of the range, which is what we should expect because of the Law of Large Numbers (Fischer, 2011). Finally, the step 2 skewed distribution generator outputs two peaks of means that are above and below the value of the uniform means (Figure 4). This is the effect of the generation algorithm which ensures the avoidance of a skewed melody having a mean near the ideal uniform one. It is interesting to note that the mean of the means of the skewed melodies is almost equal to the one of the uniform and random melodies, even though there is zero data in that range. If we wanted to find the central tendency of such a distribution we should consider using the median, instead of the mean. Generally, all of the generation procedures have the same mean of means, which is expected due to the range constraints that we set in order to have an even comparison. In conclusion, the shapes of the distributions of the means showed what we would expect from the algorithms, and the figures confirmed that the algorithms are working properly.

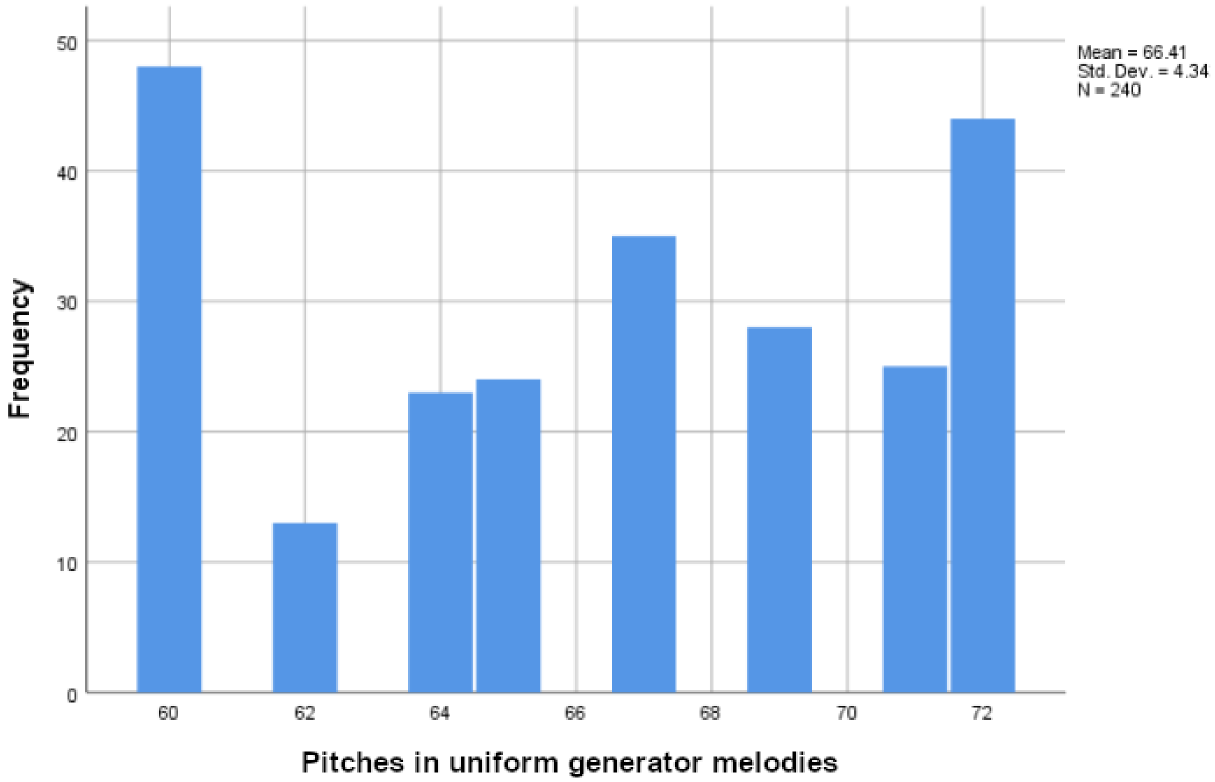

Mean = 66.41
Std. Dev. = .297
N = 30
Frequency
0
1
2
3
4
5
6
66.00
66.20
66.40
66.60
66.80
Means of pitch distributions - Uniform generator

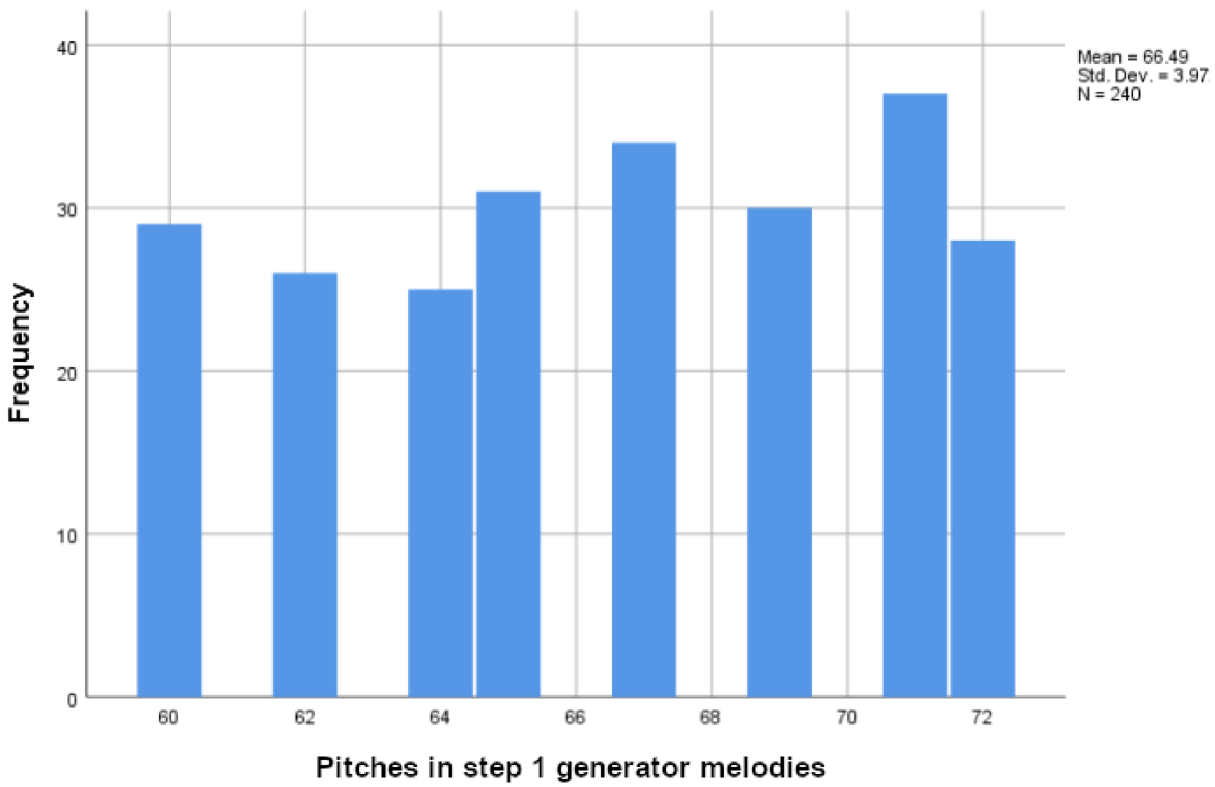

Mean = 66.49
Std. Dev. = 1.63
N = 30
Frequency
0
2
4
6
8
10
64.00
65.00
66.00
67.00
68.00
69.00
70.00
Means of pitch distributions - Step 1 skewed generator

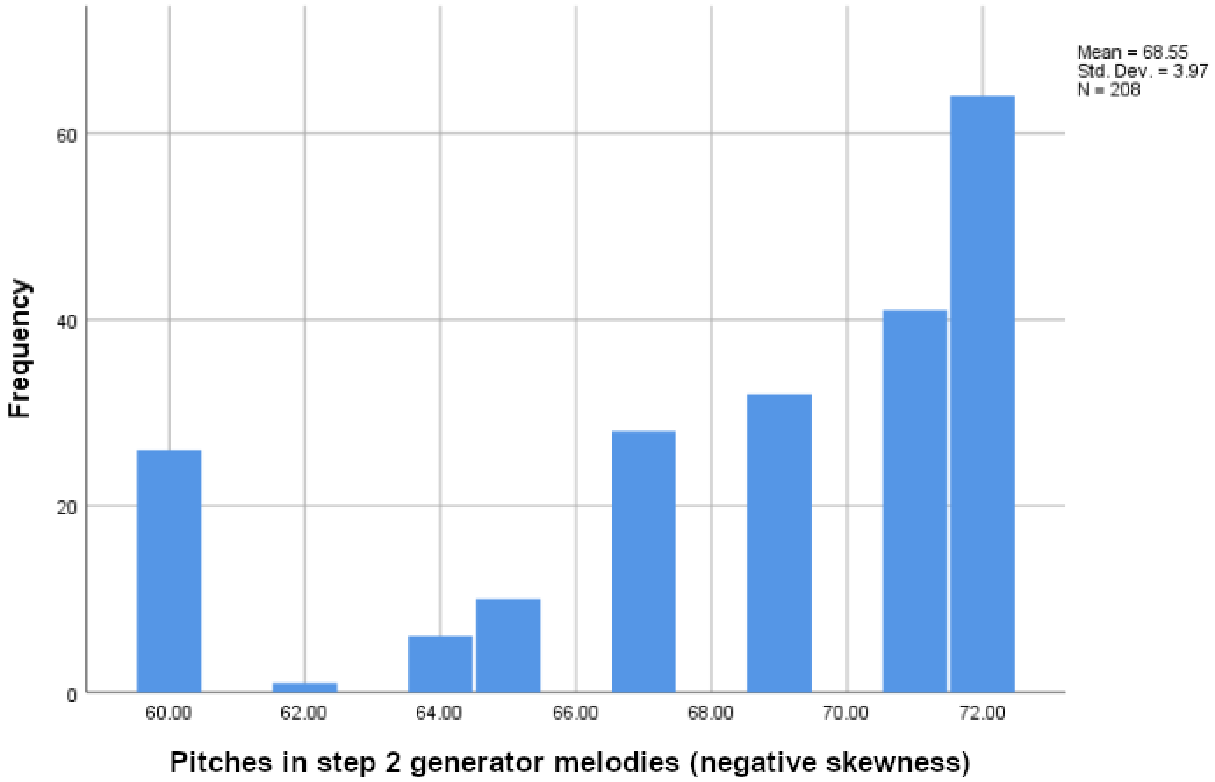


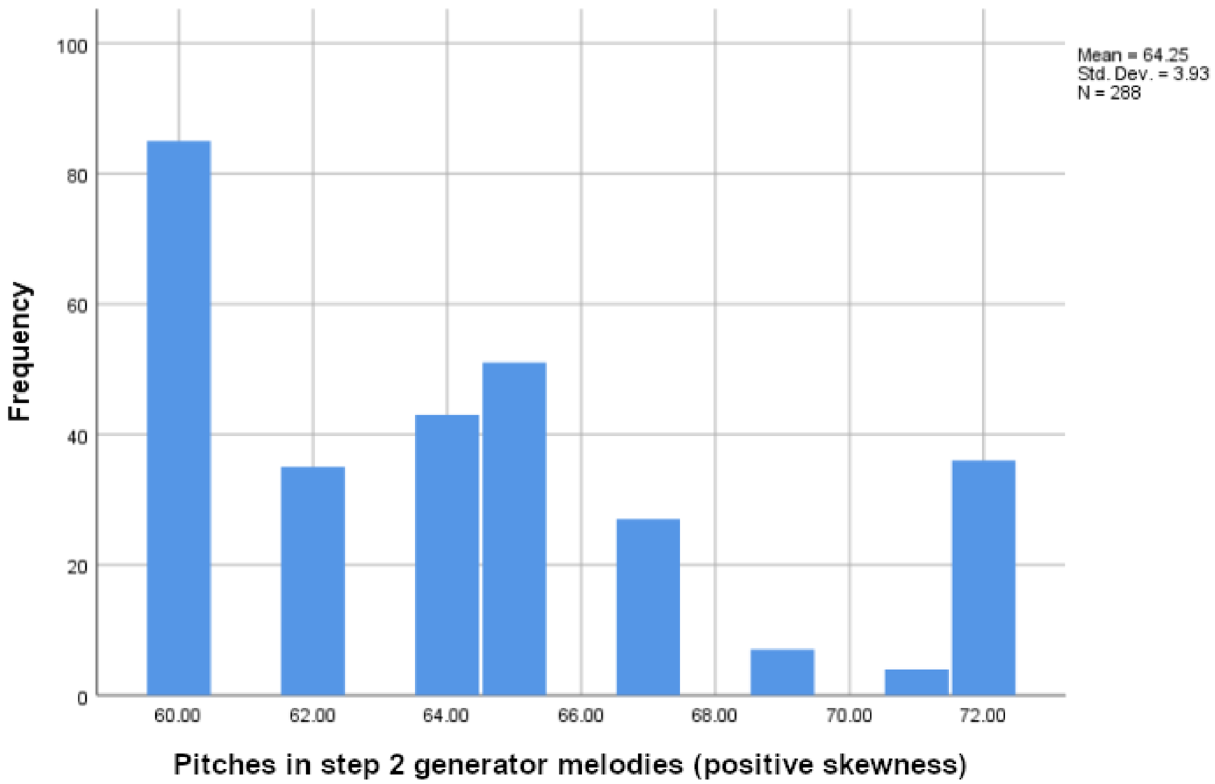
Figures 2-4: Frequencies of means of each generator rule.

For a further representation of the distributional properties of the melodies, in Figures 5-8 we show the distribution of pitches for each generation rule, summing all melodies generated by each rule. In all cases, the possible pitches were chosen to be from C4 to C5 in the C Major Scale, in order to acquire melodies that could be compared in the exact same pitch range. The high frequency of extreme pitches in the case of highly skewed, step 2 distribution, is the direct result of the algorithm trying to create melodies that have a mean that is not in the centre of the range. In the case of step 1 skewed melodies, there is a more even distribution of pitches across the specified range.


Frequency
50
40
30
20
10
0
60
62
64
66
68
70
72
Mean = 66.41
Std. Dev. = 4.34
N = 240
Pitches in uniform generator melodies


Frequency
40
30
20
10
0
60
62
64
66
68
70
72
Mean = 66.49
Std. Dev. = 3.97
N = 240
Pitches in step 1 generator melodies



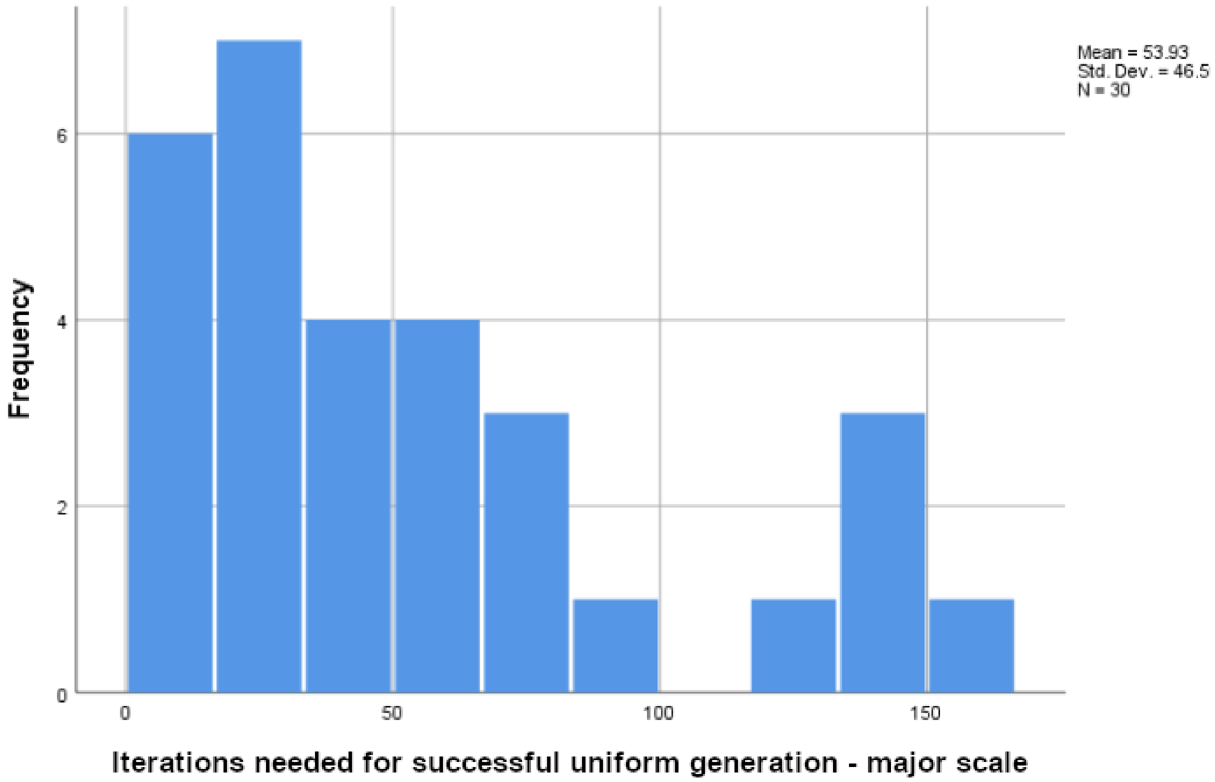


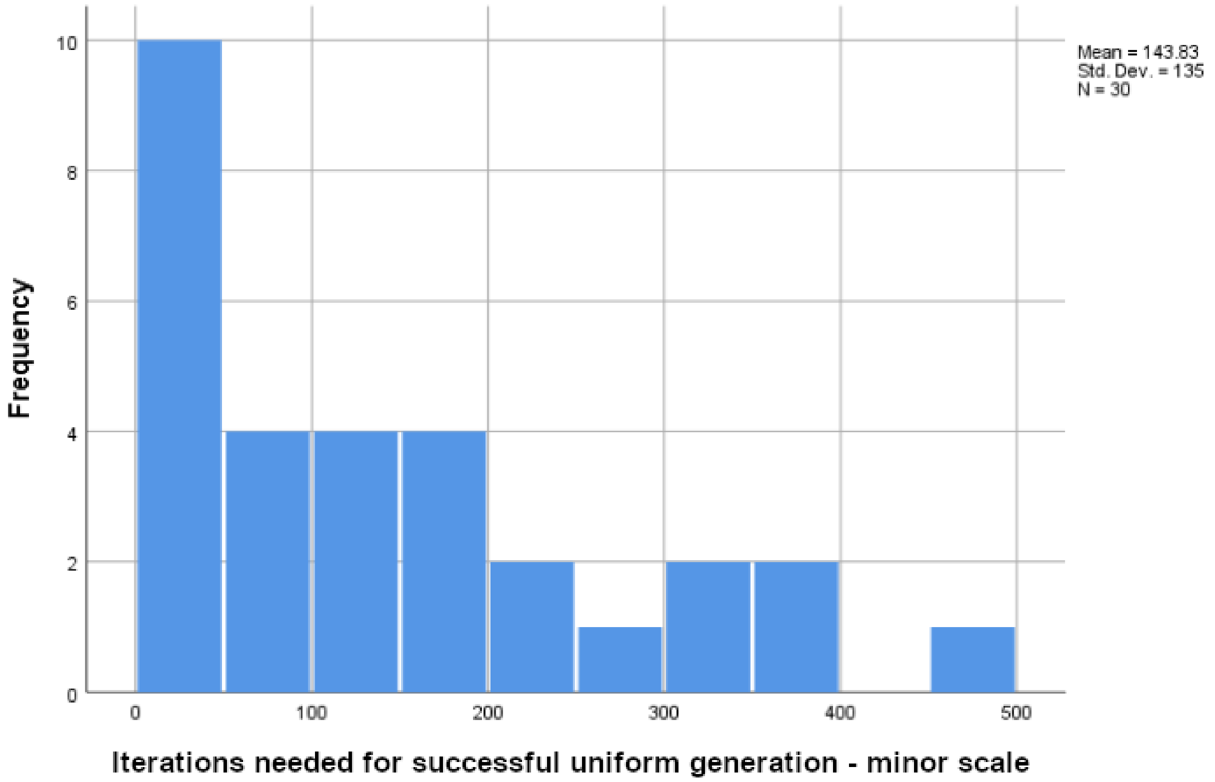
Figures 5-8: Distribution of pitches of each generation rule.

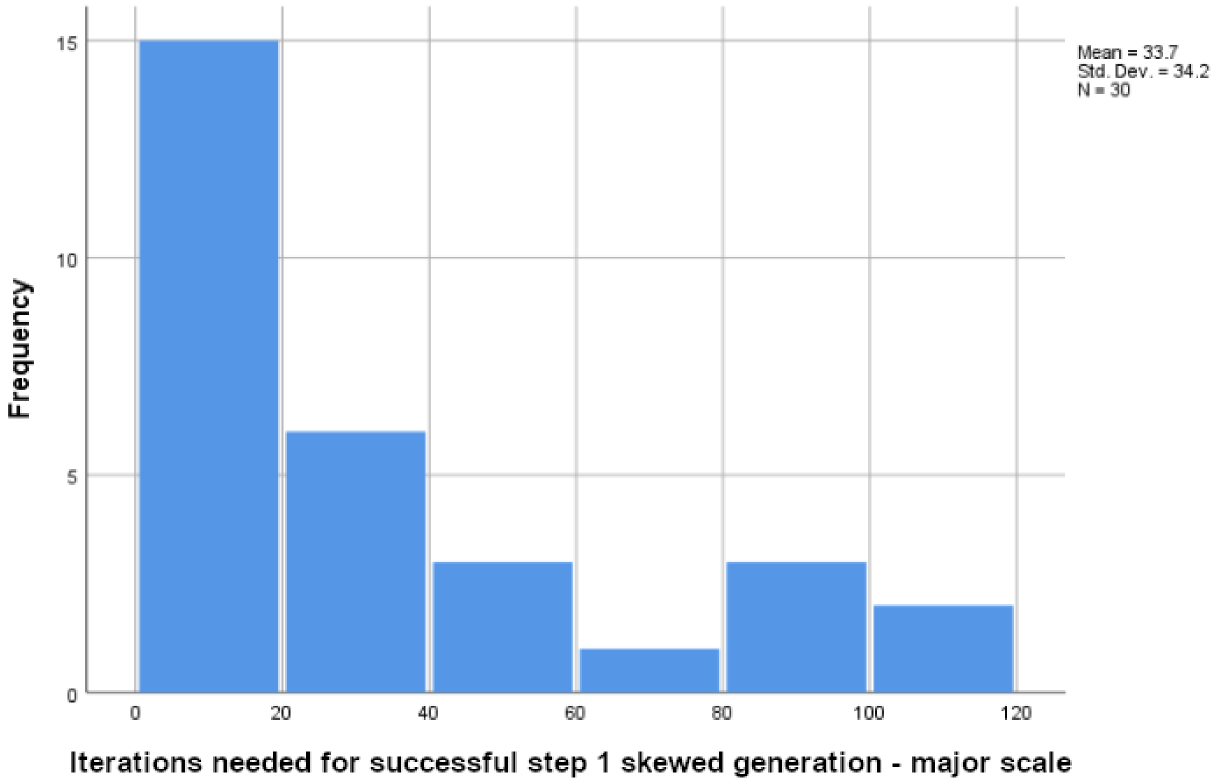
Iteration statistics

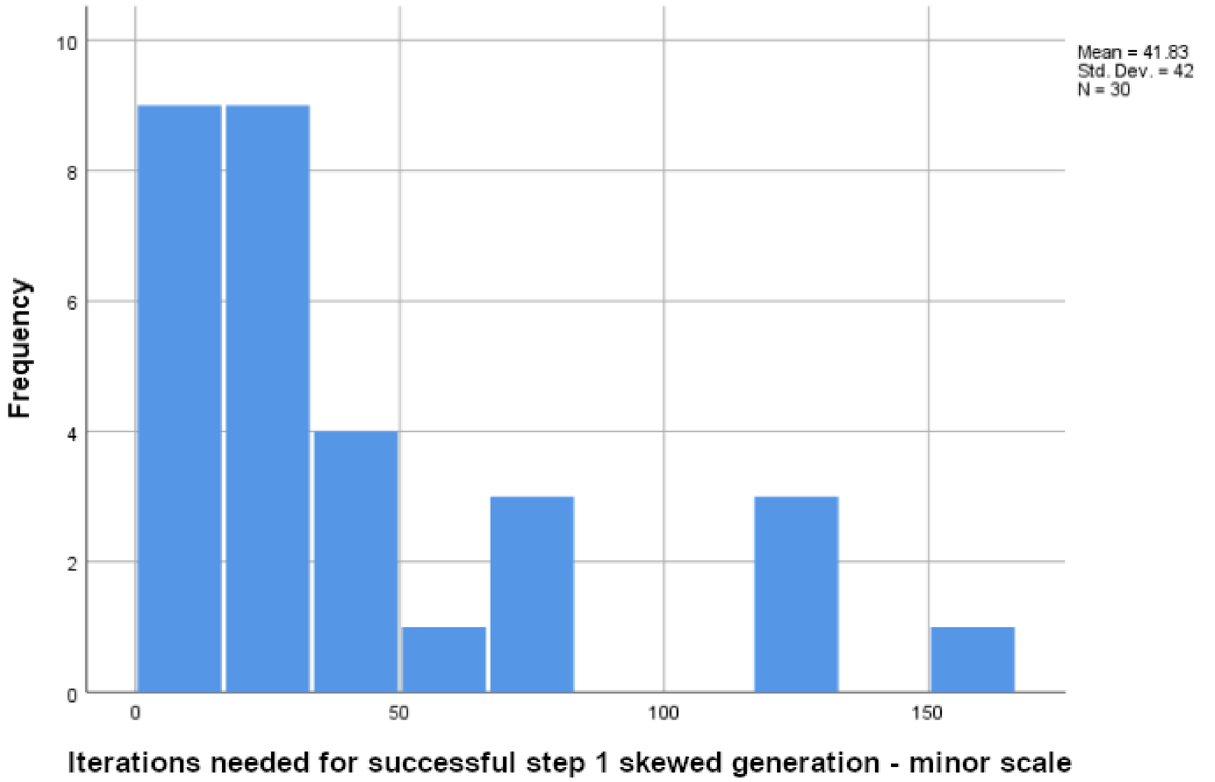
An important element of the algorithms was the ability to choose how many notes the generated melodies would have. In a 4/4 rhythmic context, which is the most familiar one in western music and thus, chosen for the experiment, having 8 notes of quarter duration creates 2-bar melodies. This allows for a bigger variation between the stimuli compared to single bar, 4-note melodies. Also, it maintains a balance of length without sacrificing listenability, which could happen if we were to make them longer than that; for symmetry we would need to have 4 bars, which, at our chosen 120 beats per minute (BPM), have a duration of 8 seconds.

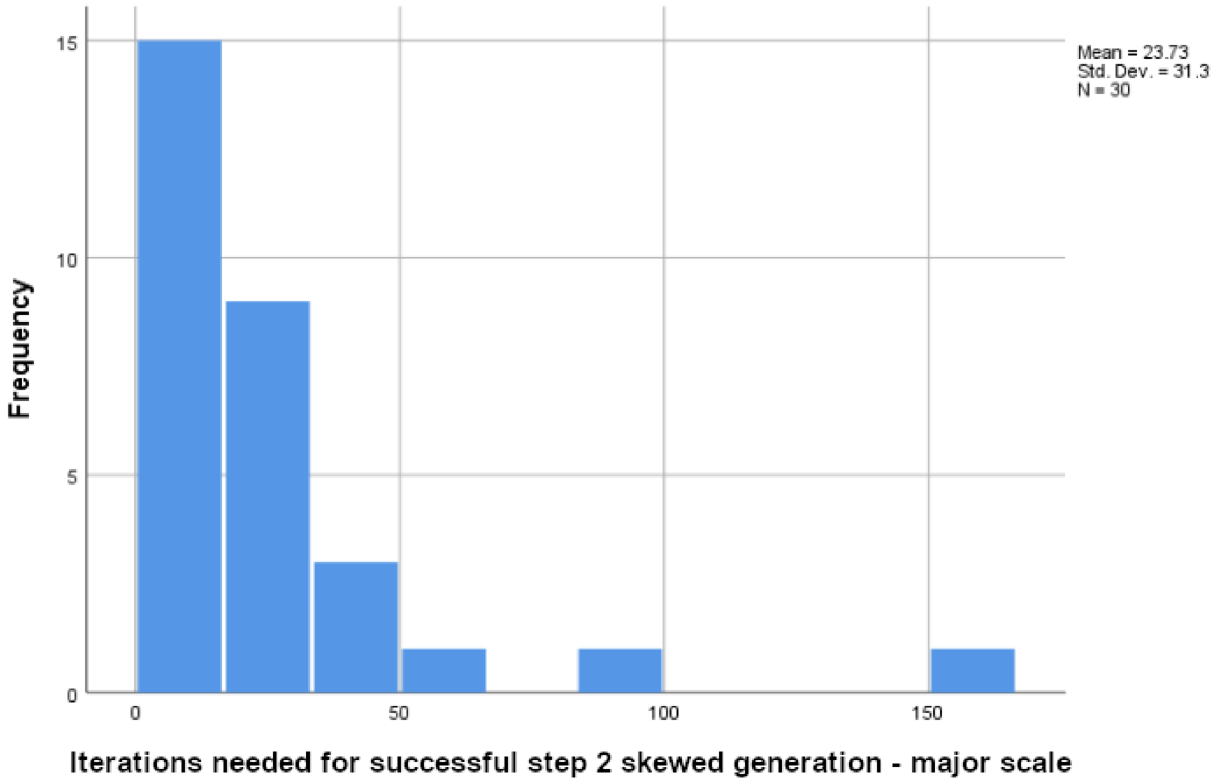
In order to observe the degree of difficulty in the formation of melodies that follow each generation rule we analysed a dataset of 180 melodies. This dataset consisted of 30 melodies for each rule-based generation procedure (uniform, skewed with a distance of means from the uniform mean equal to 2 notes, and skewed with a distance of means from the uniform mean equal to 3 notes) and for each tonal mode (major and minor). Below we present the histograms of each generation rule and each tonal mode, regarding the number of iterations which were needed for the algorithm to create a melody that follows each rule.

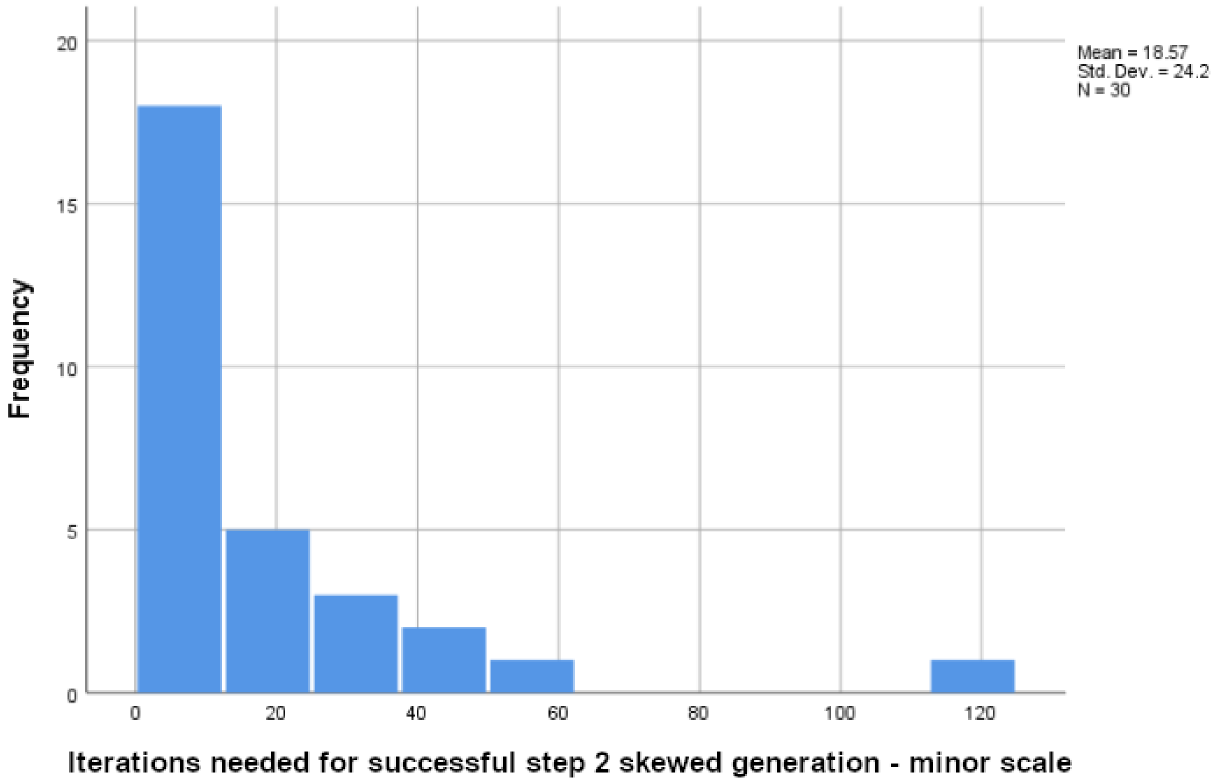

Mean = 53.93
Std. Dev. = 46.5
N = 30
Frequency
0
2
4
6
0
50
100
150
Iterations needed for successful uniform generation - major scale


Mean = 143.83
Std. Dev. = 135
N = 30
Frequency
0
2
4
6
8
10
0
100
200
300
400
500
Iterations needed for successful uniform generation - minor scale


Mean = 33.7
Std. Dev. = 34.2
N = 30
Frequency
0
5
10
15
0
20
40
60
80
100
120
Iterations needed for successful step 1 skewed generation - major scale


Mean = 41.83
Std. Dev. = 42
N = 30
Frequency
0
2
4
6
8
10
0
50
100
150
Iterations needed for successful step 1 skewed generation - minor scale


Mean = 23.73
Std. Dev. = 31.3
N = 30
Frequency
Iterations needed for successful step 2 skewed generation - major scale


Mean = 18.57
Std. Dev. = 24.2
N = 30
Frequency
Iterations needed for successful step 2 skewed generation - minor scale

Figures 9-14: Number of iterations required for each generation rule in order to reach a pre-specified mean target.

Through the histograms and in all cases, we can note that the more a generator tries to approach the uniform mean, the more iterations it requires in order to produce a valid melody. This observation indicates that while the uniform mean is at the centre of the melodic range, it does not get realised naturally (e.g. as it is argued by Von Hippel & Huron, 2000 to be the case for the post skip reversal rule).

Also, in the case of uniform minor melodies we found a mean number of required iterations at 143.83. This is a large outlier compared to the rest of generators, and it was caused by the fact that the minor mode uses notes that frequently produce means at precisely 1 semitone distance from the uniform mean. When this happens, the generator can not find any combination of notes that could make the melody arrive to the uniform mean, which means that it has to restart the procedure from the beginning, so it can find a melody (of any length) that matches the desired mean. Therefore, the difficulty of this generator to produce uniform means explains the difficulty of finding a valid melody of a specific length (i.e. 8 notes).

Statistics of melodic archetypes

An important aspect of the experiment is the addition of specific contours in the melodies, in order to test how different shapes interact with the generation rules and their ratings. An intuitive thought regarding this interaction could be that contours consisting of a bigger number of required rules for shaping depend less on the generation rules, because they allow less variation in their shape, making the melodies of their type to sound more similar regardless of their note distribution. For a methodologically valid approach, a theoretical argument is not enough, which means that we had to develop a quantitative measure to express this difference of contours in their degrees of freedom.

For this reason we used a procedure of pairwise comparison for all contours and all generation rules, to measure the Mean Absolute Difference (MAD) for each pair. The generation was split into 4 progressively more skewed distributions, which is the segmentation that we used in the experimentation. Specifically, the procedure was as follows:

1. The melodies of the dataset were all restructured according to the archetypal shape of gap fill. Then, a duplicate of the dataset was left unchanged, so it maintained the random contour structure.
2. For the gap fill archetype, we did a note to note comparison between the melodies of all generation algorithms. For example, we compared the first note of the first melody in the uniform subset, to the first note of the first melody in the step 1 skewed (onwards: “skewed 1”) subset. Then, the 2nd note for these melodies. The comparison yields the absolute difference of the pitch values of the notes. After the first pair of melodies had been compared, we calculated the sum of absolute pitch differences for the specific pair and divided it by the number of melodies. We then proceeded to melody #2 for the uniform - skewed 1 comparison. This procedure was done repeatedly for every contour and every archetype in the same manner. The procedure required *4 (generation rules) x 5 (contours) x 8 (notes per melody) x 30 (number of melodies) = 4800* comparisons and was implemented in Python.
3. The procedure essentially created variables that contained the absolute pitch differences for each melody pair, for each contour. As an example, the variable “gap fill uniform - skewed 1 difference” contains the absolute differences for each pair of melodies between the uniform and skewed 1 generation algorithms. Similarly, we created the variables “gap fill uniform - skewed 2 difference”, “arpeggio skewed 1 - skewed 3 difference”, and so on.

4. We calculated the sum of each of these variables. Then, we divided the sum by the number of melody comparisons (N) that each variable contained, in order to derive the Mean Absolute Difference (MAD) of each variable.

The results are presented in Table 1. It is worth reminding that the MAD in our case is essentially a measure of freedom, because the more two melodies vary between different rules while maintaining the same contour, the more this contour allows variance in its shape, making the melody less dependent on the contour. As Table 2 shows, gap fill allows the least freedom in shape, followed by arpeggio and by double linear contours. As expected, the random contour shows the biggest MAD allowing the least influence of melodic shape on the properties of the melodies.

Table 1: Mean Absolute Differences (MAD) for pairwise pitch comparisons of contour archetypes.

Rule 1	Rule 2	Contour	MAD
Uniform	Skewed 1	Random	40.70
Uniform	Skewed 2	Random	40.47
Uniform	Skewed 3	Random	41.13
Skewed 1	Skewed 2	Random	42.30
Skewed 1	Skewed 3	Random	44.57
Skewed 2	Skewed 3	Random	43.47
Skewed 1	Skewed 2	Random	42.30
Skewed 1	Skewed 3	Random	44.57
Skewed 2	Skewed 3	Random	43.47
Uniform	Skewed 1	Double Linear Ascending	20.70

Uniform	Skewed 2	Double Linear Ascending	22.60
Uniform	Skewed 3	Double Linear Ascending	26.53
Skewed 1	Skewed 2	Double Linear Ascending	21.43
Skewed 1	Skewed 3	Double Linear Ascending	26.57
Skewed 2	Skewed 3	Double Linear Ascending	28.73
Uniform	Skewed 1	Double Linear Descending	20.70
Uniform	Skewed 2	Double Linear Descending	22.60
Uniform	Skewed 3	Double Linear Descending	26.53
Skewed 1	Skewed 2	Double Linear Descending	21.43
Skewed 1	Skewed 3	Double Linear Descending	26.57
Skewed 2	Skewed 3	Double Linear Descending	28.73
Uniform	Skewed 1	Gap Fill	12.63
Uniform	Skewed 2	Gap Fill	14.87
Uniform	Skewed 3	Gap Fill	22.00
Skewed 1	Skewed 2	Gap Fill	14.50
Skewed 1	Skewed 3	Gap Fill	22.97
Skewed 2	Skewed 3	Gap Fill	23.53
Uniform	Skewed 1	Arpeggio	12.63
Uniform	Skewed 2	Arpeggio	15.00
Uniform	Skewed 3	Arpeggio	26.07
Skewed 1	Skewed 2	Arpeggio	14.63

Skewed 1	Skewed 3	Arpeggio	26.23
Skewed 2	Skewed 3	Arpeggio	26.67

Table 2: Sums of Mean Absolute Differences per contour archetype.

Random	252.63
Double Linear Ascending	146.57
Double Linear Descending	146.57
Gap Fill	110.5
Arpeggio	121.23

In summary, the data generation procedures showed that the MIDI generation algorithms operate as expected, by creating data of mean values that are aligned with the dispositions which must be present in the stimuli. Also, we found that different melodic shapes lead to varying freedom of note placements. This has the possible implication of a direct impact of the pitch shapes on the strength of the effect on ratings caused by the generation algorithms.

2.2.3 Participants

Our sample consisted of people that we approached online, through the email system of University of Sheffield, and social media. Regarding the demographic characteristics of our sample, we had 15 female and 13 male participants, split into 16 participants between the ages of 18-24, 10 between 25-44 and 2 at or above 45 years. 9 of our respondents stated that they did not have knowledge of music, while the rest had some experience either by being self taught musicians (7) or having formal music education (12).

2.2.4 Design

Having examined the properties of the stimuli and the operation of the generation algorithms, we present the study design that uses the aforementioned materials with the purpose of understanding the influence of pitch distributions in liking ratings of melodies. In summary, the ratings could be described approximately through an equation of the following form:

$$Y = a + b\, X_1^{X2} + c\, X_3 + d\, X_4 \qquad (3)$$

Where Y is the rating of liking, X_1 is the generation rule used and X_2 is the shape of the melody, which could be influencing the effect of X_1. Further, X_3 is the perceived complexity of the melody and X_4 is the perceived interestingness.

The independent variables of the study are the degree to which the pitch distribution conforms to uniform (four levels: uniform, skewed #1, skewed #2, skewed #3) and the contour archetype (gap fill, linear ascending, linear descending, arpeggio, random sequence - no fixed contour style). Dependent variables are the ratings of liking, complexity and interestingness of the melodies. The ratings used a Likert scale with 0 as the lowest point and 10 as the highest point. Likert was preferred to other scales for its ease of design, ease of understanding and completion by subjects, and due to its wide use in unsupervised settings (Stathakopoulos, 2005). Lastly, the control variables consist of the musical background of participants (no music education, high school, graduate, postgraduate, self taught), their age group (18-24, 25-64, 65+) and gender (male, female).

Regarding the statistical tests to be used for the study, the following design was followed: An initial view of the data was presented through descriptive statistics, with the use of frequencies and histogram views of the variables, in order to acquire knowledge on the distributional properties of the data. The Lilliefors corrected Kolmogorov - Smirnov and Shapiro Wilk tests were then used in order to examine average ratings data for normality, and 2 sample Kolmogorov - Smirnov tests were performed in this first section of the analysis in order to examine the influence of demographic factors in melody ratings. Additionally, the associations between average melody ratings were tested with pairwise comparisons using the Spearman’s Rho test. Finally, the analysis of the melody generation rules was conducted with signed ranked tests that reveal statistically significant differences in distributions of related samples.

2.2.5 Procedure

The study was run by using an online interactive questionnaire that is able to reproduce sound files, and show images. We presented various music clips of approximately 8 seconds of duration that contained a melody played with a piano timbre. A part of the stimuli was generated by using the uniform distribution (optimisation of the mean note value), and equal parts of them used progressively more distant means from the uniform mean. As noted in 2.2.2, 16 melodies were used, to achieve having 1 for every shape archetype and level of skewness. The order of the melodies in the questionnaires was random, and further, 5 different questionnaires were created to be used for each session in order to avoid possible order bias. Further, we used 2 additional test melodies at the start of each questionnaire in order to make this kind of stimuli sound more familiar, in order to prevent a bias of either exaggerated bad or positive ratings towards the melodies that are presented first. For each melody, the responders had to give these ratings after listening to it:

- Rating of the melody regarding how much the participant liked it. This is the measurement of aesthetic appreciation.
- Rating of the melody in terms of the perceived complexity of it.
- Rating of the melody in terms of how interesting it sounded.

At the end of the questionnaire, the participants were asked if they had music education before, their age group and their gender.

2.3 Results

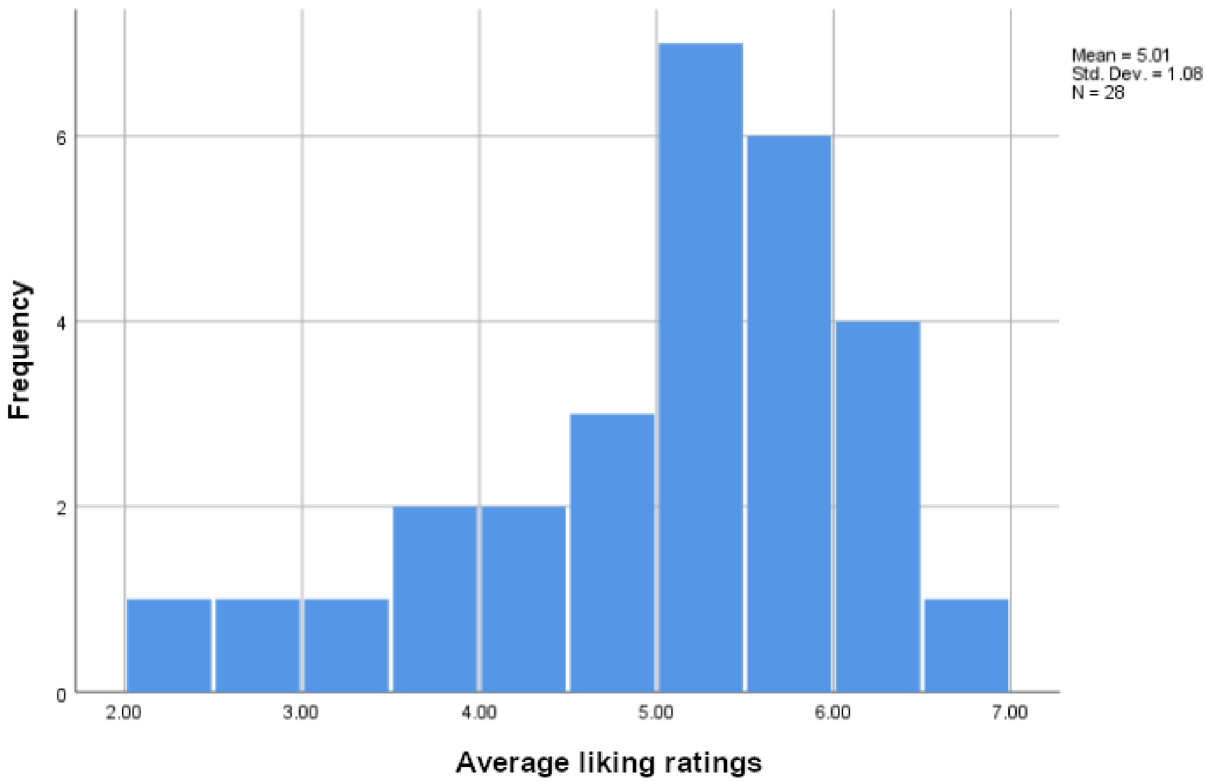
The results section is split into four parts. At first, we present the descriptive statistics of the variables of the experiment and discuss the possibilities and limitations of the acquired data when it comes to the degree to which they can be used to properly address the questions of the study. We proceed with addressing the study's aims that are related to the effect of complexity, interestingness and demographic elements on liking ratings. Even though this part of the analysis was complementary to the main hypothesis testing of the generation rules, we addressed it first because it considered our dataset on a more general level. After that, we focus on the main research question of what are the differences in liking ratings between the uniform rule in comparison to the others. We describe the procedure of re-coding the raw liking data to a point-system format that is suitable for the nature of the data, and proceed with showing the results of the statistical tests that we used to address these differences.

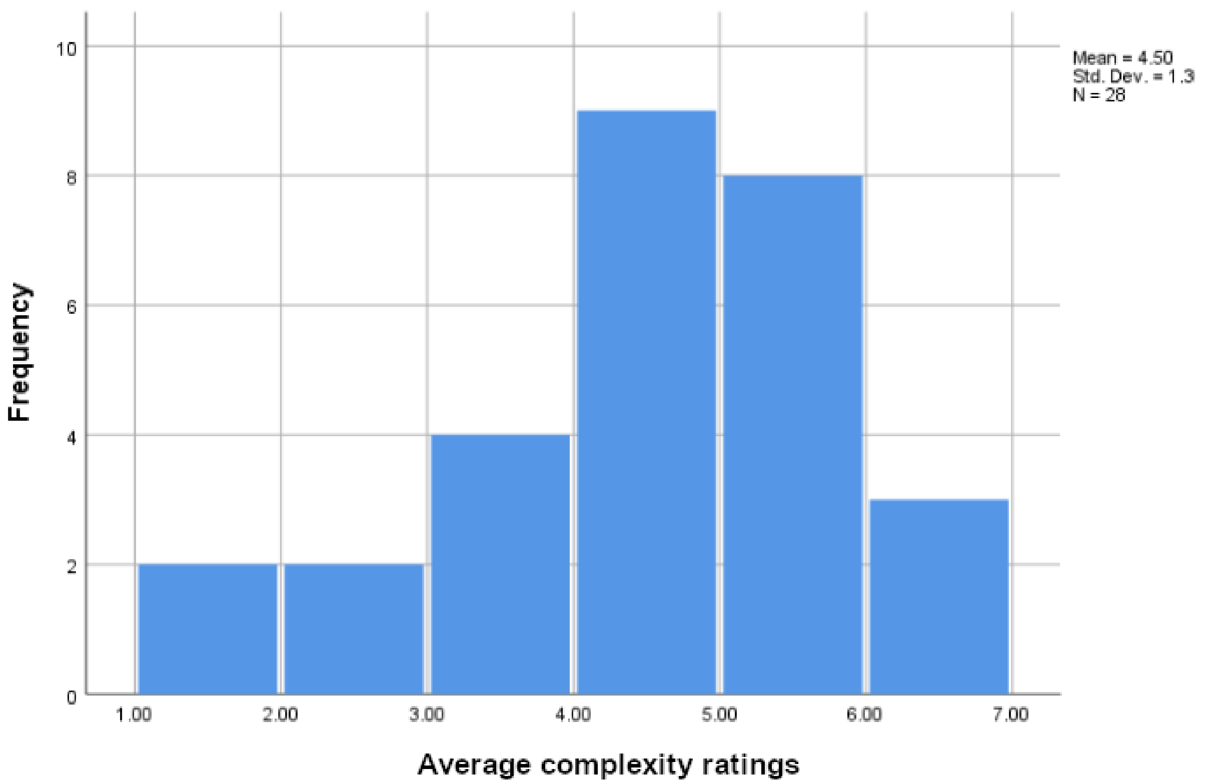
2.3.1 Descriptive statistics

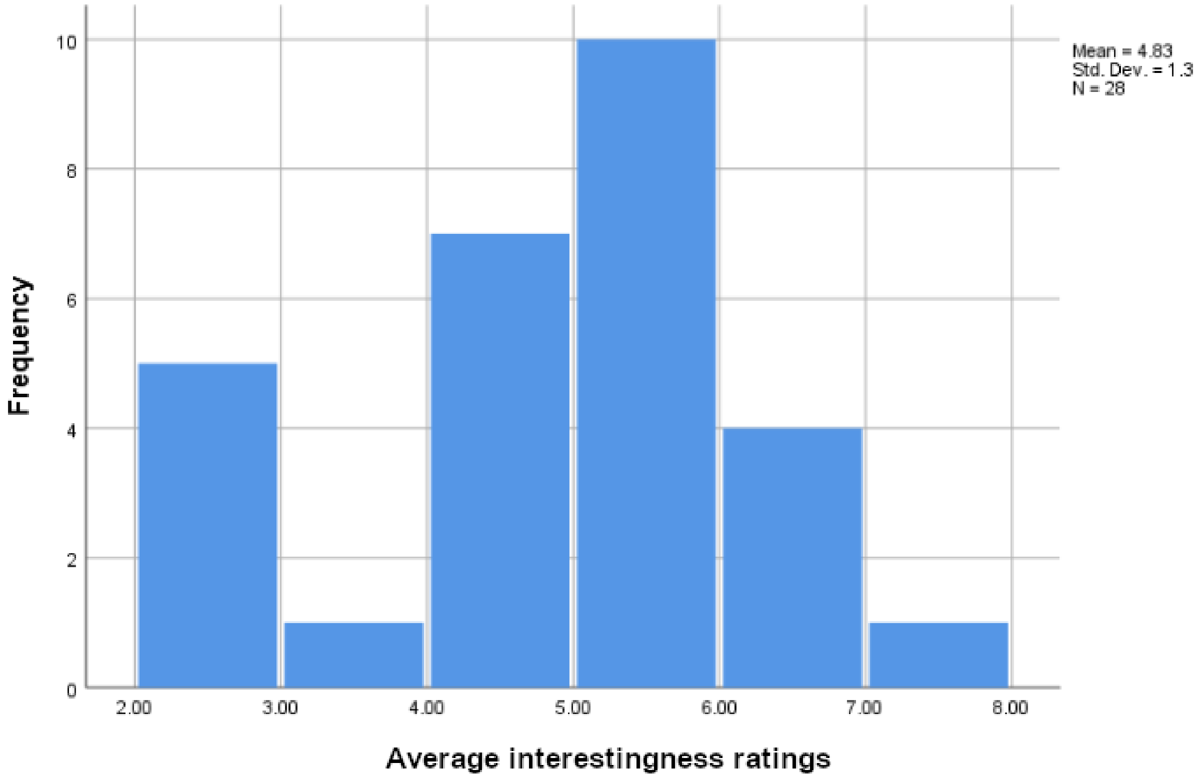
Our sample consists of people from varying age groups, music education levels, musical preferences and genders, and it has a size of N = 28. Initially it was 30, but the data of two responders was excluded from the analysis because of random completion of the questionnaire after a specific point. From the first excluded participant, there was more than one melodic shape that had precisely equal ratings across all types of uniformity levels, and by the second excluded participant, a variability of less than 3 points for all questions along with a lack of ratings above 3/10 in any question.

The sample of our study, as described in 2.2.1, contains a sizable demographic variability, which can benefit the study in the sense of discovering factors that influence the main variables of interest, which are the melody ratings. For conducting the analysis, given the relatively small sample size, we need to have an approximately equal split of binary demographic pairs that would allow the appearance of statistical significance. Therefore, for the requirements of the study we aimed for a split of about 14 participants in each binary variable. In the 4 variables of the dataset (age group, music education, preferred genre and gender) we can see that there was generally an unequal split between groups, except for the case of gender where we had 15 female and 13 male participants. However, by examining the data we found possible recodings that were implemented, in order to aid the inference potential of our statistical analysis. For example, in the age group variable we have 16 participants in the 18-24 group, 10 in 25-44 and just 2 in 45+. From this data we created a new variable which differentiated between young adults (18-24) and adults (25+), and that it has a split of 16 - 12. Similarly, with regards to music education, there were 9 responders who stated that they do not have any knowledge of music. The other 19 had a form of music education, either self taught (7) or formal (12), thus forming this binary distinction of musicians and non-musicians. Finally, in the case of preferred music genres we had a big variability and a lack of usable recoding of the categories, since no combination could provide a meaningful separation of the dataset. Having created these new categories, we could proceed with examining the descriptive statistics of our main points of interest: the variables of liking, complexity and interestingness.

For these ratings we started the analysis by looking at their distributional properties on a broad level, and specifically at the fitness of their overall average for each participant to the normal distribution. To calculate those, we created new variables that consisted of the average values of each variable, for each participant. For example, liking averages has an N = 28 and every data point is the average liking rating of a participant across all the melodies of the questionnaire. The histograms of liking, complexity and interestingness are shown in Figures 15-17.


Mean = 5.01
Std. Dev. = 1.08
N = 28
Frequency
Average liking ratings


Mean = 4.50
Std. Dev. = 1.3
N = 28
Frequency
Average complexity ratings



Figures 15-17: Distributions of average ratings of liking, complexity and interestingness for each participant.

By looking at the figures above we get a sense of the distributional properties of this data. They all have a similar mean and variance. While the number of melodies is sufficient for data analysis, it can quickly get smaller when we analyse the ratings in subgroups of age groups, genders and music education. As a result, we need to test the data for normality so we can decide what kinds of statistical tests we will use, that is either parametric or non parametric tests. We ran the Lilliefors Corrected Kolmogorov Smirnov and the Shapiro Wilk normality tests, and the results are presented in Table 3.

Table 3: Tests of normality for the distributions of liking, complexity and interestingness.

	Kolmogorov - Smirnov			Shapiro - Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Liking Average	.156	28	.078	.934	28	.077
Complexity Average	.103	28	.200	.974	28	.690
Interestingness Average	.187	28	.013	.898	28	.010

As we can see from the Significance rows, we could not reject the null hypothesis for the variables liking average and complexity average, at a 5% significance level. The H_0 hypothesis for both of the tests is that the distribution is normal, which meant that in our case, we could not reject the hypothesis that the variables liking average and complexity average follow the normal distribution. Regarding interestingness average ratings, both of the normality tests rejected the null hypothesis at a 5% level. Because at least one of our variables did not follow the normal distribution, we did not use a parametric approach of analysis. Importantly, since we had to analyse subgroups of this relatively small sample, non parametric tests would be strongly recommended in the subsequent analyses of this data.

2.3.2 Non-parametric tests selection for the analyses of the average liking, complexity and interestingness variables

Regarding data characteristics, since both the ratings and the demographic questions are in discrete scales, the nature of the data is not continuous. Also, not all ratings follow a distribution that could approximate normal, as the tests revealed. These observations are sufficient to minimise the amount of possible tests that we can run in the analysis. Specifically, for the reasons mentioned above we cannot execute a Multivariate Analysis of Variance (MANOVA) test that could be a starting point. However, a test of this type would only be helpful as an initial look into the possibility of possible significance in the mean differences of the data. Since we wanted to look at and compare various demographic subgroups as a part of the exploratory analysis of the dataset, we chose to use a series of 2-samples means comparison type of test, like the t test. Other than the comparisons of these pairs, we also performed a correlation analysis between the average rating variables of liking, complexity and interestingness.

Having a dataset that contains non-normal distributions, especially given the small number of data points, called for a non parametric approach. In order to work with the statistical tests of this kind, we examined our dataset, and to what degree it conforms to the assumptions of the non parametric equivalents of the tests that are relevant to our aims. Specifically, we had to check the test assumptions of the independent samples Mann - Whitney U test as the most common alternative to the independent samples t test, and of a correlation test, such as Spearman’s Rho.

Analysis of the effect of demographic characteristics and correlations of average ratings

When it comes to the Mann - Whitney U test, since we were interested in comparing the means in pairs of two independent samples (e.g. liking average comparison between younger and older participants), we wanted to ensure that the distributions of the compared variables are of the same shape. This would allow us to interpret the outcome of the test as a result that considers the difference between the means of the compared variables. However, an analysis of the histograms of our variables as split by age, musical expertise and gender, did not ensure such similarity of shapes. Under this scenario, we chose an alternative route to the analysis on possible demographic subgroup variations of the means of average ratings. We tested the differences between the empirical distributions of the compared pairs, and thus, tested the probability of the compared subgroups having data that is drawn from the same probability distribution. This, in essence, gave us the information of whether or not the two compared demographic subgroups differ in average ratings of liking, complexity and interestingness. The implementation of the comparisons between empirical distributions was done with the 2 samples Kolmogorov - Smirnov test. We compared the distributions of liking average, complexity average and interestingness average ratings between musicians and non-musicians, females and males, and participants that are younger than 24 or older.

Tables 4-6: Comparisons of empirical distributions between demographic groups.

Two-Sample Kolmogorov-Smirnov Test Statistics Between Females And Males				
		Liking Average	Complexity Average	Interestingness Average
Most Extreme Differences	Absolute	.323	.303	.359
	Positive	.046	.200	.077
	Negative	-.323	-.303	-.359
Kolmogorov-Smirnov Z		.852	.798	.947
Asymp. Sig. (2-tailed)		.461	.547	.331

Two-Sample Kolmogorov-Smirnov Test Statistics Between Musicians And Non-musicians				
		Liking Average	Complexity Average	Interestingness Average
Most Extreme Differences	Absolute	.246	.211	.181
	Positive	.246	.140	.181
	Negative	-.170	-.211	-.170
Kolmogorov-Smirnov Z		.607	.520	.448
Asymp. Sig. (2-tailed)		.855	.949	.988

Two-Sample Kolmogorov-Smirnov Test Statistics Between Under And Over 24 Years				
		Liking Average	Complexity Average	Interestingness Average
Most Extreme Differences	Absolute	.417	.229	.250
	Positive	.417	.229	.104

	Negative	-.250	-.125	-.250
Kolmogorov-Smirnov Z		1.091	.600	.655
Asymp. Sig. (2-tailed)		.185	.864	.785

The Two - Sample Kolmogorov - Smirnov Test comparisons of the scale average variables work by calculating the supremum value of the differences between the two empirical distribution functions. The Tables 4-6 presented above show that there were no statistically significant differences between different age groups, genders or musical expertise level. These results held true for the average ratings of all variables (liking, complexity and interestingness).

Correlation statistics of average liking, complexity and interestingness values

So far we have presented our analysis on the control variables and showed that we could not reject the hypothesis that there was no significant difference between the empirical cumulative distributions of the sub samples. These results mean that we did not find evidence to support that age, gender or musical expertise influence the values of the average liking, complexity and interestingness scores. Addressing factors that may be affecting liking ratings, the next step is to approach the related research question: What is the role of complexity and interestingness considering liking ratings? We looked at the correlation between the three variables, with the aforementioned Spearman's Rho test.

Regarding Spearman's Rho test assumptions, we needed to examine whether the pairs of variables have a monotonic relationship, in order to be able to perform the test. This relationship can be verified through a matrix of scatter plots, as shown in Figure 18.

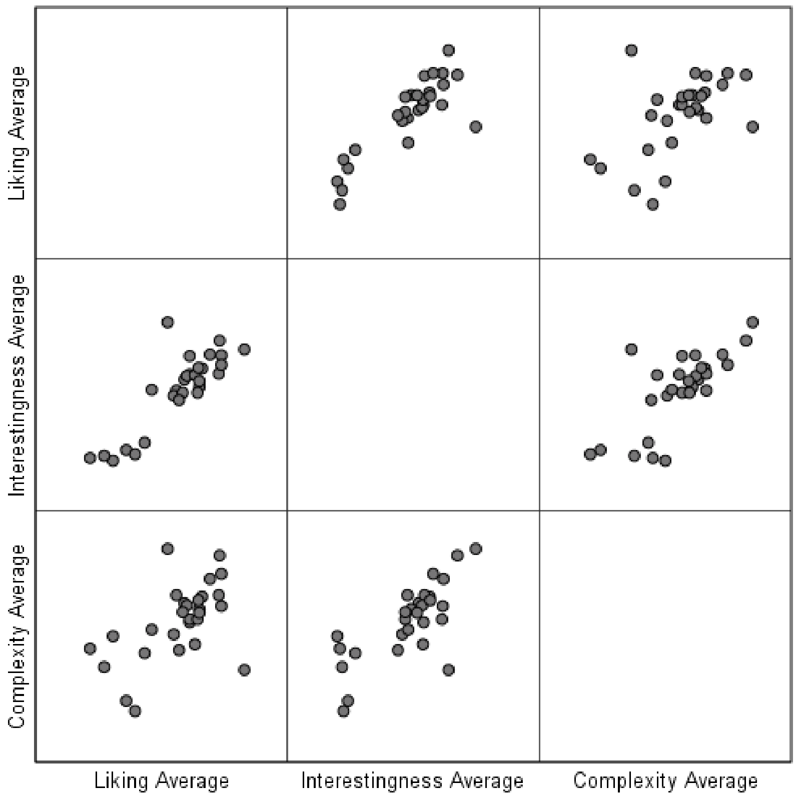


Figure 18: Matrix of scatterplots between scale average pairs.

Considering that the scale averages dataset conformed to the prerequisites of our chosen correlation test, we could now address the exploration of how complexity and interestingness relate to liking. The Spearman's Rho tests results are shown below in Table 7.

Table 7: Correlation analysis for liking, complexity and interestingness average ratings. Values with * indicate 1% significance level.

<table>
<tr><th colspan="5">Spearman's Rho Correlations (N=28)</th></tr>
<tr><td colspan="2"></td><td>Liking Average</td><td>Interestingness Average</td><td>Complexity Average</td></tr>
<tr><td rowspan="2">Liking Average</td><td>Correlation Coefficient</td><td>1</td><td>.789</td><td>.540</td></tr>
<tr><td>Sig. (2-tailed)</td><td>.</td><td>.000*</td><td>.003*</td></tr>
</table>

Interestingness Average	Correlation Coefficient	.789	1	.673
	Sig. (2-tailed)	.000*	.	.000*
Complexity Average	Correlation Coefficient	.540	.673	1
	Sig. (2-tailed)	.003*	.000*	.

Table 7 contains pairs of comparisons between the average ratings for each participant, that show the correlation coefficient of each pair, along with the statistical significance of the coefficient. Liking average ratings showed a high correlation with interestingness average (.789) and a moderate correlation with complexity average (.540). The third pair, between interestingness and complexity average values shows a moderate to high correlation (.673). The results of all comparisons were statistically significant at a 1% level. We could initially conclude that participants tied how interesting the melodies sounded to them to how much they liked them. Furthermore, we saw a degree of connection between all values, as participants did not exhibit big variability in the three ratings within each melody item.

2.3.3 Analysis of melodic preferences between pitch distributions

In the next section of this chapter, the central focus is to present the analysis of preferences between the generation rules that produced melodies of different pitch distributions, which was the main subject of the study. The dataset of melodic preferences between these rules passed through transformations, since we needed to rearrange the raw ratings of each participant, in a way that would present each participant as a single data point. For this data point, we would have each generation rule in the form of a variable, so we could compare these rules through related - samples statistical testing procedures. In order to reframe the dataset in this way, a ranking procedure was developed that compared the generation rules according to the preferences of each participant. This procedure consisted of looking at each melodic shape group rated by the participant, and distributing points to the generation rules according to their relative score. For example, suppose that regarding the arpeggio shaped melodies the skewed #1 distribution had the highest liking ratings, followed by the uniform, and skewed #3 and #2 had the 3rd and 4th best liking ratings. In that case, for the particular shape skewed #1 generator would get 3 points, uniform would get 2 points, and skewed #3 would get 1 point. This process was repeated for all shapes, and the summary of points for each generation rule resulted in the total points of the rule for the particular participant.

This dataset could now allow the comparison of the generation rules by also having every participant act as a data point. The way to perform this analysis is to use multiple related samples comparisons and compare all possible pairs of rules, for every participant. Similarly to the steps taken for the demographic analysis, we also had to make sure that this data is in accordance with the assumptions of the statistical tests. We used Wilcoxon Signed Rank, as the most usual non parametric test for related samples testing. Therefore, we needed to examine whether the variables follow probability distributions that are symmetrical in nature. In particular, we were interested in the whole sample, as well as the subsamples of musicians and non-musicians which could provide a meaningful comparison due to possible differences in melodic perception. We tested the symmetry of distributions through boxplots, as shown in Figures 19-20.

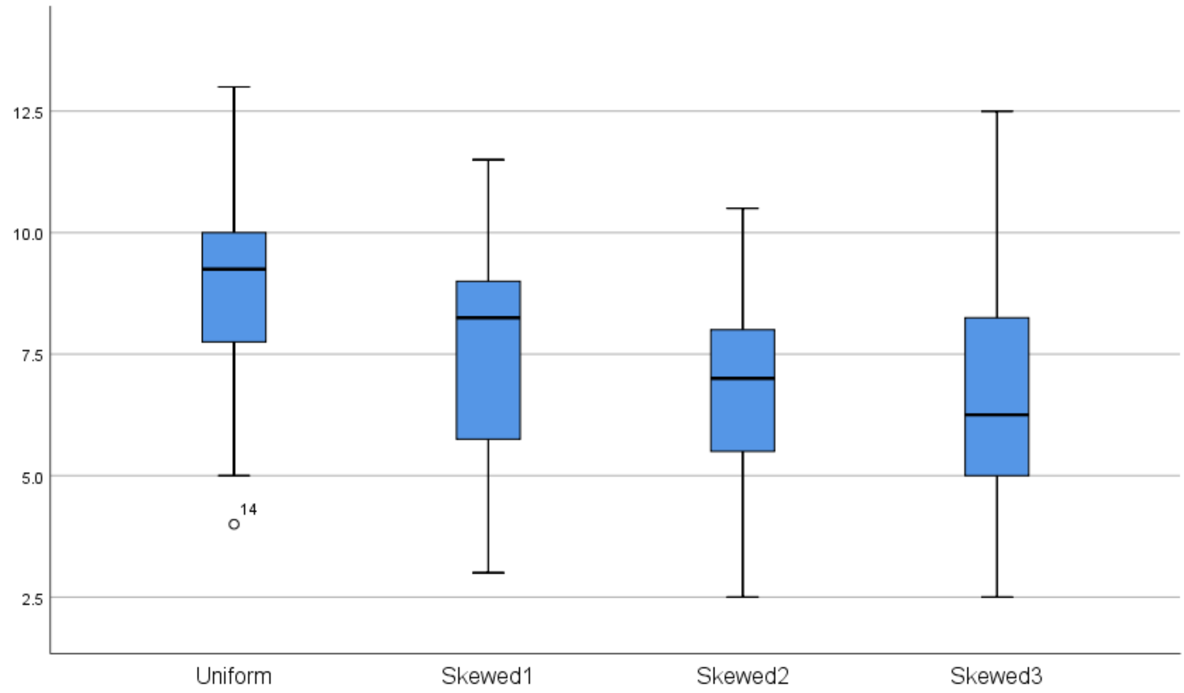


Figure 19: Boxplots of all generation rules, for the whole sample of the study.

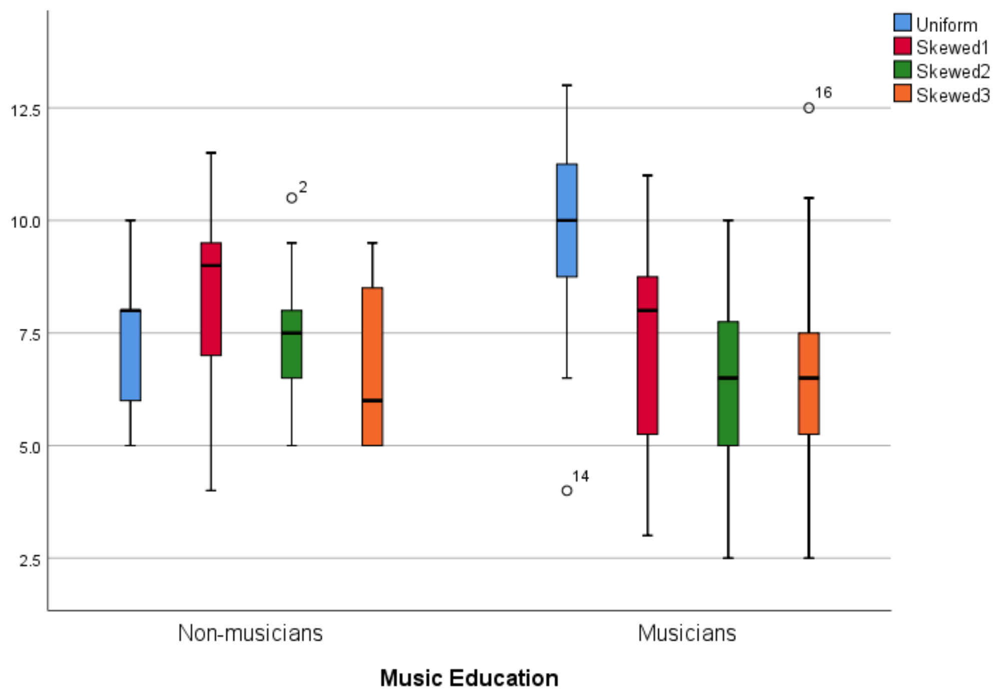


Figure 20: Boxplots of all generation rules, for the distributions of non-musicians and musicians subsamples.

The horizontal axis of Figure 19 represents the generation rules, and the vertical axis shows the value of each variable represented by the boxplots. The questionnaire contained 5 melodic shapes, and for every shape, the most highly rated generation rule gained 3 points. Therefore, the maximum possible value for a generation rule was 15. The theoretical minimum was 0 (the rule should always be the least liked in all melodic shape scenarios). In Figure 20, the horizontal axis represents the two subsamples of non-musicians and musicians, while the colour of each boxplot stands for a corresponding generation rule. The values of generation rule variables are shown on the vertical axis, like in Figure 19.

While looking for symmetry through boxplots, we must observe a median that is in the middle of quartiles 1 and 3, and a plot of approximately equal lengths of whiskers. We found that in Figure 19 there is no symmetry in the distributions of Skewed #1 and Skewed #3 variables. Also, in the subsamples graphs of music education, there was no symmetrical distribution in non-musicians. By performing a Wilcoxon Signed Rank test, we would have difficulty interpreting the results as a difference in means, because of the asymmetrical shapes of the distributions in comparison. Therefore, we preferred to use the paired samples Sign test, as an alternative to Wilcoxon Signed Rank, since it does not assume symmetrically distributed variables. This test was performed for the sample as a whole, but also ran independently for the subsamples of musicians and non-musicians. The results of the Sign tests are presented in Tables 8-10.

Tables 8-10: Sign comparisons between liking rating rankings for all combinations of melody generation rules, separated by samples under consideration.

Paired Samples Sign Test (Full Sample)						
	Skewed1 - Uniform	Skewed2 - Uniform	Skewed3 - Uniform	Skewed2 - Skewed1	Skewed2 - Skewed3	Skewed3 - Skewed1
Exact Sig. (2-tailed)	.078	.021	.005	.170	.124	.441

Paired Samples Sign Test (Non-musicians)						
	Skewed1 - Uniform	Skewed2 - Uniform	Skewed3 - Uniform	Skewed2 - Skewed1	Skewed2 - Skewed3	Skewed3 - Skewed1
Exact Sig. (2-tailed)	.727	1	.508	.508	.508	.180

Paired Samples Sign Test (Musicians)						
	Skewed1 - Uniform	Skewed2 - Uniform	Skewed3 - Uniform	Skewed2 - Skewed1	Skewed2 - Skewed3	Skewed3 - Skewed1
Exact Sig. (2-tailed)	.008	.004	.004	.332	.238	1

The significance of the comparisons between melody generation rules considers the degree to which the differences in liking ratings exist due to chance. In the full sample test, there was a 5% significant difference between liking ratings of uniform and skewed 2 and 3. However, when we split the sample into musicians and non-musicians, we found a notable difference in the significance of the comparisons between the two subsamples. Musicians showed a 1% statistically significant difference in all comparisons of the uniform melodies with the skewed ones. On the contrary, there were no significant differences in any comparison between the generation rules for the non-musicians subsample. This contrast echoes the results of the boxplot graphs (Figure 20), where uniform liking ratings seemed to be at higher values than the rest of the rules in the musicians group.

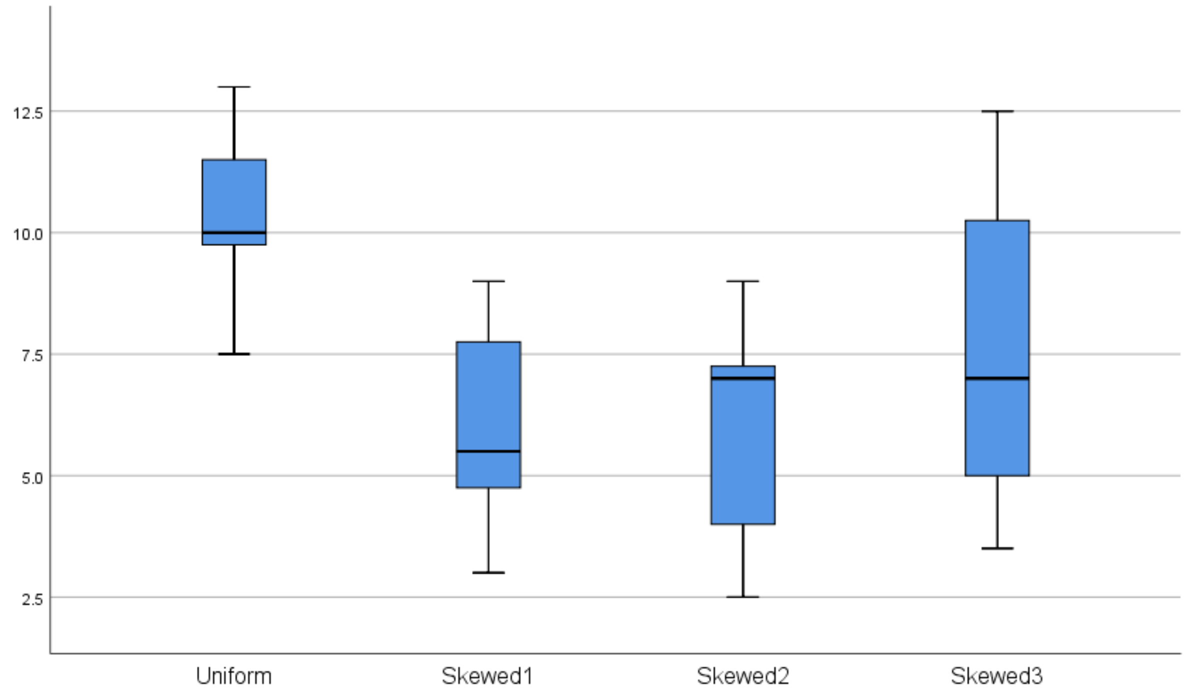
The multiple comparisons that we performed with the Sign tests yielded significant results for musicians, which seemed to be in accordance with the idea of Uniform distribution’s effect on melody liking ratings. We further used a multiple testing correction test, to account for the fact that an increasing number of tests could end up with some falsely positive results. We used Benjamini - Hochberg, a modern test that uses False Detection Rate (FDR), which is a statistic that plays the role of p-value in the decision of how significant the original multiple testing p-values are (Benjamini & Hochberg, 1995). The results are presented in Table 11.

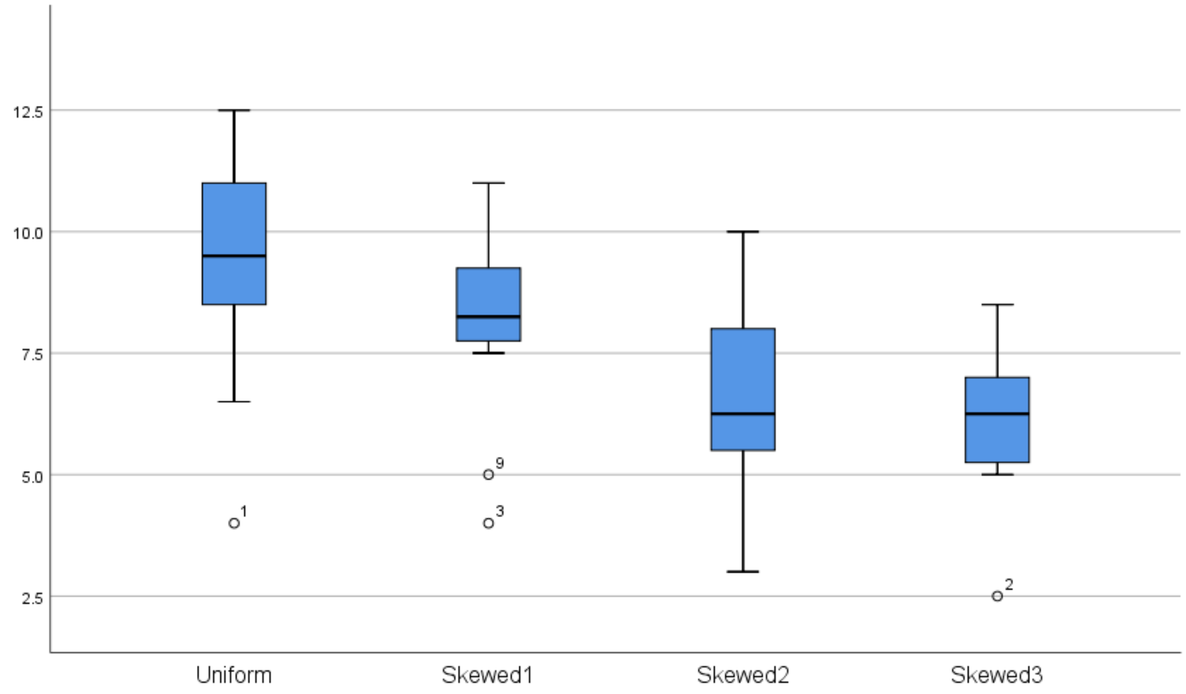
Table 11: Multiple testing correction test for the Sign comparisons of generation rules in musicians.

Benjamini - Hochberg Multiple Testing Correction for the Musicians Subgroup.						
	Skewed1 - Uniform	Skewed2 - Uniform	Skewed3 - Uniform	Skewed2 - Skewed1	Skewed2 - Skewed3	Skewed3 - Skewed1
False Discovery Rate	.012	.012	.016	.286	.286	1

After performing the Benjamini Hochberg test, the differences between uniform and all skewed rules were retained, at a 5% False Discovery Rate. This outcome provides us with evidence that our main hypothesis for this study holds true, that is, the mean liking ratings of uniform melodies appeared to be higher than those of progressively more skewed melodies. The effect was strongly present in the subsample of musicians, but not in non-musicians.

Finally, it is interesting to note that a part of the musicians (7) was made of self-taught musicians. While the dataset is not big enough to allow us meaningful statistical analysis, a simple view on the ratings of self-taught and formally educated musicians through boxplots (Figures 21-22) reveals that the self-taught subgroup played at least an equal role on the significance of our findings, compared to the formally educated one.





Figures 21-22: Liking ratings of the generation rules by self-taught (21), and formally educated (22) musicians.

2.4 Discussion

In this chapter we presented a study on the effect of pitch distributions to the aesthetic appreciation of melodies. Specifically, we tested the idea that uniform pitch distributions could result in melodies that would be rated as more aesthetically pleasing than melodies using more skewed distributions. Also, we examined the association of complexity and interestingness with liking ratings, both in an overall sense and by the use of subsamples of demographic elements.

Regarding the latter, a strong correlation between the overall ratings of interestingness and liking was found, as well as a moderate correlation between complexity and the other two variables. We can conclude that interestingness could be a concept that is closely associated with liking, and that we would need further examination about how the notion of complexity is related to liking in melodies. The association between interestingness and liking is in accordance with scales-based studies about this topic in music (Russell, 1994) and art (Aitken, 1974). Considering the demographic subgroups analysis, we did not find evidence of statistically significant differences between any of the subsamples in the average ratings of liking, complexity and interestingness, as separated by age groups, genders and level of music education. These comparisons were tests on probability distributions, meaning that we did not find evidence against the assumption that the subgroups rate the melodies in the same way and that any distributional differences between the groups may have occurred by chance.

The results of the analysis regarding melodic preferences between generation rules showed a strong statistically significant difference in liking ratings between melodies of uniformly distributed pitches, and pitches of progressively more skewed distributions. We call this positive effect of uniformly distributed pitches on liking ratings as the Uniformity Principle. Notably, this was true for self taught and formally educated musician participants, but not for non-musicians, indicating a difference in how the two groups perceive melodic information. Overall, we can conclude that this experiment indicates a strong likelihood that musician participants show preference for the melodies originating from uniformly distributed pitches, and that interestingness may be a notion that is closely associated with liking.

This outcome can be regarded as a contribution to the literature of aesthetic preferences quantification, providing an objective rule that supports the generation of melodies using statistical target values. As such, it can be useful in the creation of compositional assistance tools for music making. Further, it is an addition to the studies that support the idea that musician listeners are sensitive to and influenced by the statistical distributions and properties of pitch. However, we could not confirm a stable link between the complexity of melodic stimuli with liking, either linearly or as an inverted U type of relationship. It should also be noted that the study was done through an online questionnaire, meaning that the participants worked on an unsupervised setting and on varying listening systems. We can not rule out a possible effect of this fact on the ratings, and especially on the ratings and results that were non significant.

3. Study 2: The role of working memory in the aesthetic appreciation of melodies

3.1 Introduction

As a next step to Study 1, we focused on the notion of complexity in the context of aesthetic appreciation of melodies, and in what ways we could quantify any possible effects of complexity as a means to use this idea for the generation of melodic output. This step would complete our overarching goal of creating a framework of quantified aesthetic properties. This framework would be used as a basis, in order to create a music generating software solution for music makers in the form of a Digital Audio Workstation plugin.

When we examine the properties of music in regard to their relation to responses of aesthetic preference, we can find rich literature on the attribute of complexity. Specifically, it has been observed that certain levels of it in musical stimuli can be optimal for the aesthetic experience of listeners (Beauvois, 2007), even though complexity can be perceived differently among individuals (Eerola et. al, 2006). An important related contribution in musical aesthetics is that of Berlyne (1974), who examined the relationship between aesthetic response and what he called “collative properties”, which are elements of the stimuli such as complexity, novelty and familiarity. Subsequent research has supported Berlyne’s theory of an “inverted U” relationship between aesthetic preference and complexity (Gordon & Gridley, 2013; Delplanque et. al, 2019), and has further looked on areas such as ways of modelling complexity (Marin & Leder, 2013; Eerola, 2016) and its relationship to variables such as creative potential (positive correlation between potential and preferred complexity level - Ziv & Keydar, 2009) and rhythmic properties (positive emotional valence effect of syncopation induced complexity - Keller & Schubert, 2011). Elements of novelty in music can also be seen as aiding preference in a general sense (Berlyne, 1970) but at the same time, multiple exposures to the same stimulus can increase enjoyable-ness according to the mere exposure effect (Zajonc, 1968).

However, the main points of interest between the effects of novelty and repetition overlap only on a surface level: The first notion relates to the properties which are in the compositional structure of a musical piece or the construct of a melody, and the related literature refers to the importance of complexity levels in the achievement of aesthetic pleasure. Repetition of an auditory stimulus (e.g. a repeating sound), on the other hand, is seen as an influencing variable on the level of its experienced musicality (Margulis & Simchy-Gross, 2016; 2018). In other words, of the degree to which that sound or set of sounds is perceived as music. Thus repetition, in the sense that it is discussed in this chapter, is not an influencing variable of a primarily structural nature.

A concept related to complexity is that of working memory, as it is a system of limited capacity and accountable for the processing of temporary information (Miyake & Shah, 1999). Musical complexity is specifically related to working memory both in terms of capacity limitations (melodic memory capacity is approximately 7-12 notes - Pembrook, 1987) and of influence in performance (Silverman, 2012). The working memory system, as modelled by Baddeley and Hitch (Baddeley and Hitch, 1974; Baddeley, 2003), is composed of a central executive component that supervises and controls the flow of information between itself and the subsystem components, the phonological loop and the visuo-spatial sketchpad. The first one is responsible for storing and manipulating verbal content, while the second element stores visual information. In relation to music, there is debate on whether the classic Baddeley model is accurate in depicting the processing of musical information as a procedure of the phonological loop subsystem, or whether an extra system specialised in music is required. Studies on the interference effect of language in tone recall, with a sample of musicians and non-musicians, showed significant differences between interference from language compared to music for musicians (non-significant for non-musicians), which is an indication that there exist differences between language related and musical related memory (Deutsch, 1970; Pechmann & Mohr, 1990; Jones & Macken, 1993; Williamson, Mitchell, Hitch, & Baddeley, 2010). Further experiments by Thompson & Yankeelov (2012) on musical and verbal memory with the use of irrelevant sounds, in order to compare the performances of musicians (at least seven years of musical training) and non musicians, reveal that the phonological loop system could be altered to include a temporary storage element for music.

This type of memory has potential for exploration within the framework of aesthetic preferences, as it is thought to be responsible for constantly updating harmonic expectancies and to raise activity in Brodmann Area 44, which is implicated in music-syntactic processing, and ventral premotor cortex (important for the processing of musical structure) brain regions when irregular chords are perceived (Koelsch, 2006). Similarly, it can be thought of as responsible for updating note to note expectations within a phrase (Rohrmeier & Koelsch, 2012). Therefore, by being linked to musical expectations and complexity, working memory and its capacity limits can potentially be a factor of influence on musical aesthetic preferences.

To the author's knowledge, this possible connection has not previously been researched in music cognition literature yet. More broadly in art, Sherman, Grabowecky & Suzuki (2015) have found that appreciation of visual artworks increases when the visual complexity level of the stimuli matches the working memory capacity of the subjects. In music, the impact of working memory on appreciation has only been approached through the tangent of studies on complexity and aesthetic preferences. However, a combination of working memory with the aforementioned factors can provide a base on which the influence of capacity-complexity relations on ratings can be analysed, informing the current literature of quantitative studies on aesthetic preferences in music.

Modelling complexity and expectations in working memory

In order for this analysis to happen, the ideas of complexity and expectations must be clarified and defined in the context of working memory. Under the time-based resource-sharing (TBRS) model of working memory span, as presented by Barrouillet et. al (2004), it is assumed that the processing and maintenance of information require attention as a common functional resource. Through a series of 7 experiments involving resource-sharing tasks of elementary operations, in their article they presented conclusive evidence regarding the above assumption, as well as they showed that working memory span can be modelled using estimates of cognitive load.

Within the TBRS theory, complexity is seen as a direct result of the concatenation of the elementary processes of a task (Barrouillet et. al, 2008). Also, any score differences in working memory related tasks should be interpreted as differences in the cognitive load resulting from the atomic constituents of the stimuli, if all other aspects of cognitive load are constant. Therefore, by having a quantitative measure of cognitive load, we can control its individual aspects and create a framework in which we can create a task that could explore complexity (in the context of working memory capacity) in relation to aesthetic preferences. Such a measure is given by Barrouillet et. al (2004), with the equation (4).

$$CL = \frac{\sum_{i=1}^{n} a_i n}{T} \tag{4}$$

In which:

- CL stands for Cognitive Load
- a_i is the difficulty / attentional demands of retrieval i
- n is the number of retrievals
- T is the total duration of the activity

For the creation of such a task, with which we aim to effectively address a link to aesthetic preferences, having a quantitative measurement of cognitive load is not enough, because in our context of music listening it involves simply the storing of contour and information about pitch relations. As a result, the attentional requirements it has are used for accurate representation of the incoming musical stimuli. However, this study considers the role of working memory (and not of passive short term memory) in the link between the complexity of musical stimuli and the aesthetic appreciation of them. On top of making participants use their attentional resources for the purpose of storing the aforementioned information, we must, therefore, force a simultaneous task that requires the sharing of attentional resources and that depends on active processing of the musical stimulus. By doing this we can explore the impact of varying the storage related cognitive load on the performance of the active processing task, and of the solving strategy that is being used by the subjects when attention resources for the active processing task are scarce. By varying storage cognitive load systematically in relation to the active processing task, we achieve the important aim of making sure that working memory - specific parameters are utilised in the analysis. Finally, in order to connect how the manipulation of storage related cognitive load and the limitations of available attention influence aesthetic appreciation, the active processing task should be based on a mechanism that is known to be connected to musical appreciation. Such a task can be based on real time (active) formation of melodic expectations, a mechanism that is well documented to play an important role in music related aesthetic responses (Meyer, 1956; Narmour, 1989; Schellenberg, 1997; Huron, 2006; Brattico & Pearce, 2013). In order to model this task, we can use the variables of pitch direction and pitch intervals, which listeners would evaluate using probabilities that are conditional on heard pitches. Overall, the components that we manipulate are cognitive load, and the formation of melodic expectations. The ways this is achieved are described in more detail in 3.2. Further, we will take into account the effect of pitch grouping, that is the perception of melodic parts as segmented into groups that are coherent in terms of their notes being in similar pitches (Bregman, 1994). To summarise as a diagram, we can visualise the connection between complexity, working memory and aesthetic appreciation as follows:

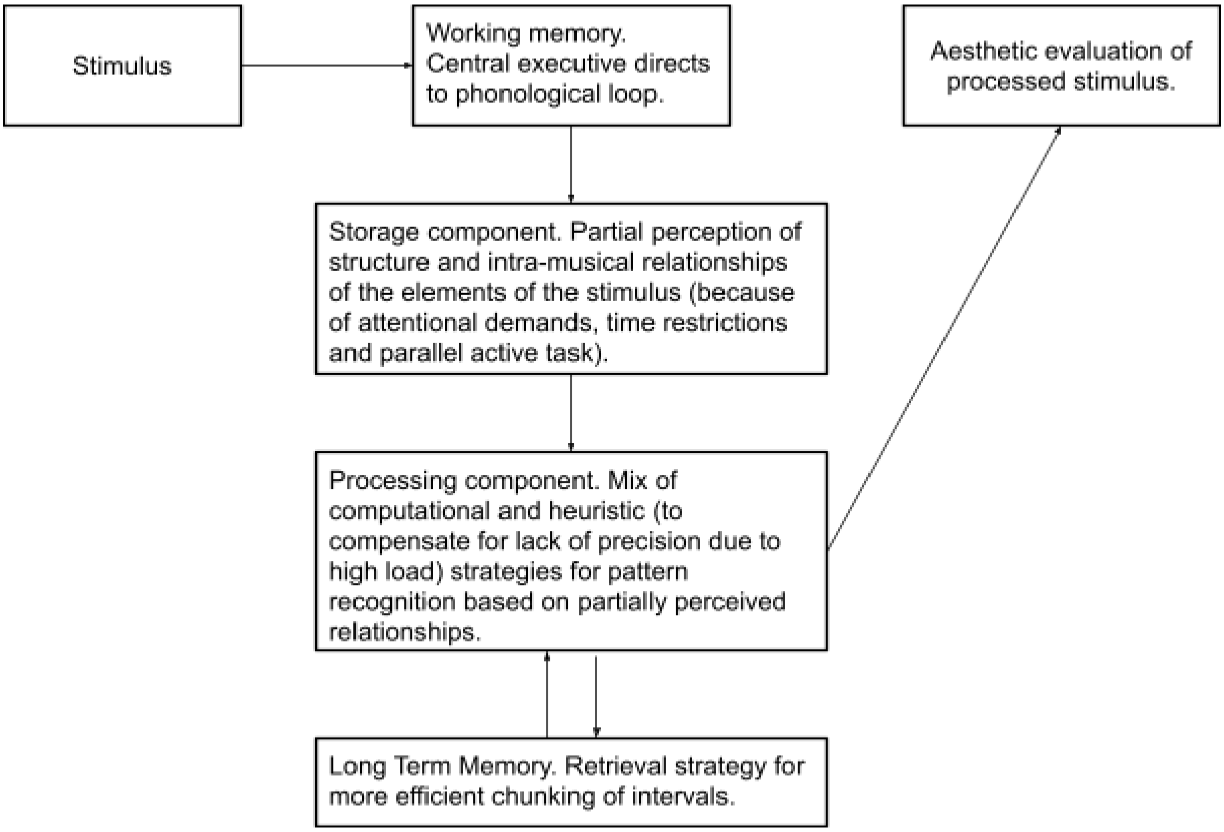


Figure 23: Connection between complexity, working memory and aesthetic appreciation.

3.2 Method

3.2.1 Design

In order to test how working memory capacity limitations may affect the perception of aesthetic quality in melodies, in this study we used music generation algorithms that output diatonic melodies consisting of two parts. An initial first part and a second part that is an inversion of the first part (same contour and scale, but opposite note to note direction). This design was incorporated in order to test if and to what degree the part will be perceived as an inversion, and how the perception is altered by different levels of induced storage cognitive load from this part. The cognitive load (CL) level was manipulated by creating second parts which varied in intervallic distances. They were either keeping the intervallic distances of the first part (medium CL melodies, exact inversion parts), keeping approximately only the pitch direction removing the information of note to note distances (simpler melodies, contour-only inversions), or expanding the intervallic distances of the initial part (higher CL melodies). What changes in the formation of expectations when storage load is varied is that the melodies with exact inversion parts are in theory (mathematically) complying more to the expectations created by the first melodic segment (maintaining intervallic distances as opposed to less complex melodies). This implies a possible preference of listeners towards the medium cognitive load melodies instead of simpler ones, because of better fulfilment of expectations. However, in this study the listeners report the extent of fulfilment of expectations they experience by addressing the perceived relationship between the two segments of each melody. Thus, the preference contrast between better fulfilment of expectations / more demanding stimulus does not clash with our focus of study which is to analyse the extent to which cognitive load will hinder the perception of the relationships of the two melodic segments, and what strategy, if any, do the listeners use in absence of available attention for the active task. Further, the ratings of aesthetic preference will be compared to the perceived fulfilment of expectations, which is not necessarily in line with the mathematically correct realisation.

As suggested above, the removal of intervallic information can be interpreted as creating a less complex stimulus which is less demanding for processing in working memory (Sweller, 1994). Also, various studies by Dowling have shown that the tracking of intervals is a difficult task for the memory to retain, since listeners are generally not able to maintain the exact intervallic distances of melodies that have undergone transformations such as inversion, retrogration and retrograde inversion (Dowling, 1971; 1972; Berz, 1995).

Other possible ways of modulating working memory load could be the number of notes, or rhythm complexity. However, working with intervals makes it possible to keep only this specific parameter free to vary. If rhythm was preferred, the prediction mechanisms in terms of melodic contour would not differentiate between stimuli, and in the case of number of notes there would also be unavoidable confounding changes in intervals. In relation to the cognitive load equation described above, the stimuli have identical number of retrievals (n), and duration of activity (T). Therefore, their only difference is in the extra load that comes from computational demands in the case of non single step, actual inversion parts (a_i, i being the sequential note number). Importantly, a characteristic of inverted parts is that the inversion of contour creates difficulties in the memory representation of the first part as related to the inverted one (Dowling, 1971). As a consequence, when we vary intervallic distances in the inverted part, we vary the available attention of the subjects which they can use in order to find the relation between the initial and the inverted contour. On the other hand, manipulation of the duration of activity (T) can present a different insight to our analysis by changing the cognitive load without changing note relations. In a part of the study, the BPM of the stimuli are raised to a faster pace, in order to explore possible differences in the results of expectancy fulfilment when comparing slower and faster variations of the same melodies (keeping everything in the equation constant, except T).

3.2.2 Stimuli

For the stimuli creation, music generation algorithms were written in Python 3. The generators produce randomly distributed notes in a given scale and pitch range, forming a melody which is then shaped according to a quantified implementation of contour archetypes as described by Meyer (1973). This melody forms the first part of the stimulus, as discussed in 3.1. Another melody is then generated, and each note of the new melody follows a rule that makes it an inverted (complementary) melody of the initial one. Suppose that C_i is the i-th in order note of the complementary melody C, P_i the i-th note of the first part (primary) melody P and the notes are described in terms of MIDI pitch values. The complementary melody is, then, created according to the equation (5).

$$C_i = P_1 - (P_1 - P_i)^b \tag{5}$$

Where:

- $i = 2, \dots, n$
- $n = \#P$
- $C_i \in S$
- $P_i \in S$
- $S = \{0, 1, \dots, 127\} \setminus \{\text{off-scale MIDI pitch values}\}$
- b is a scaling factor of the intervallic information.

An example of this kind of melody is shown in Figure 24 (b=1) and Figure 25 (b=⅔).

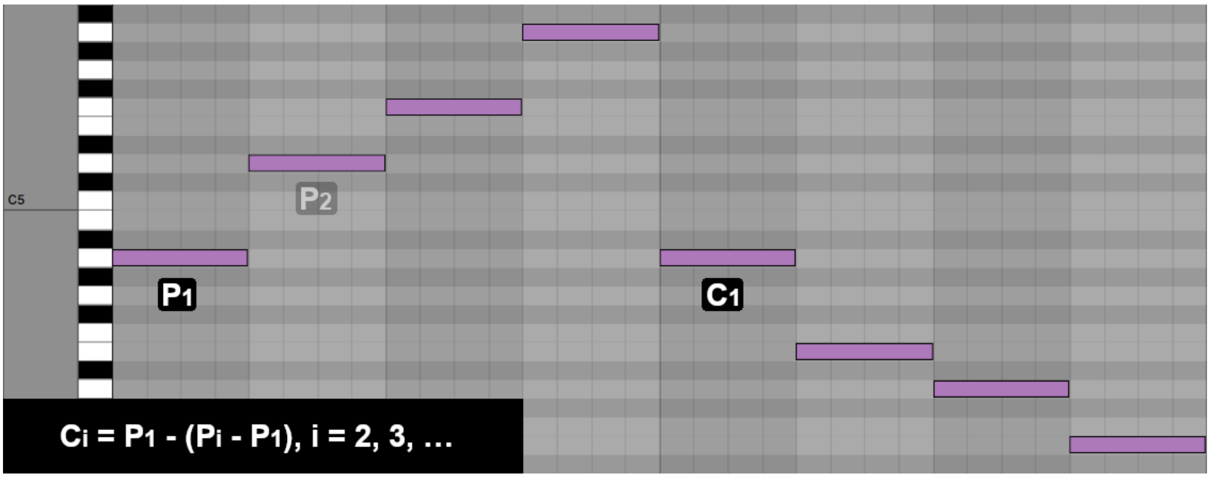


Figure 24: Last four notes are the complementary part without reduction of intervallic information (b=1).

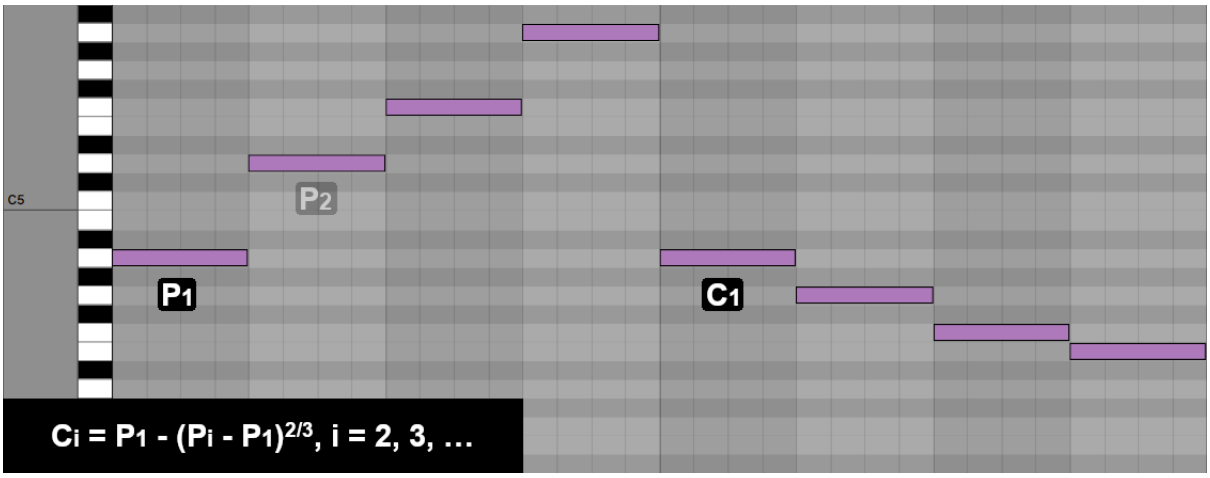


Figure 25: Last four notes are the complementary part with elimination of intervallic information (b=⅔).

By using this procedure we generated stimuli that were given the contours of linear ascending, linear descending and gap fill melodic archetypes, plus random shaped contours. This was done in order to control for the factor of melodic shape. Examples of these shapes are given in Figures 26-29.

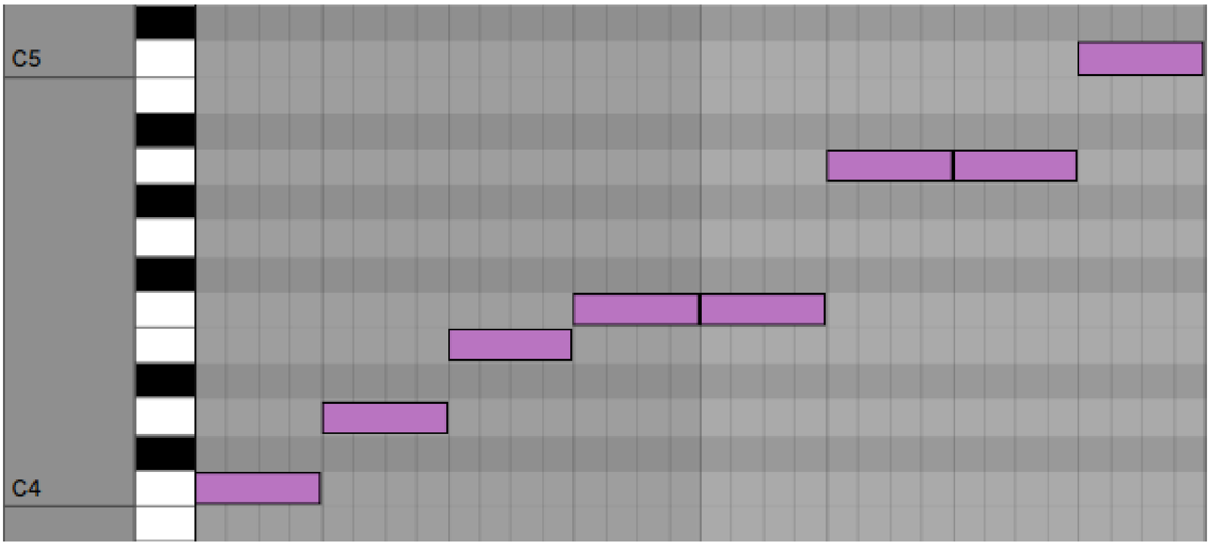


Figure 26: A linear ascending melody. Each note is always at an equal or higher pitch than all previous ones.

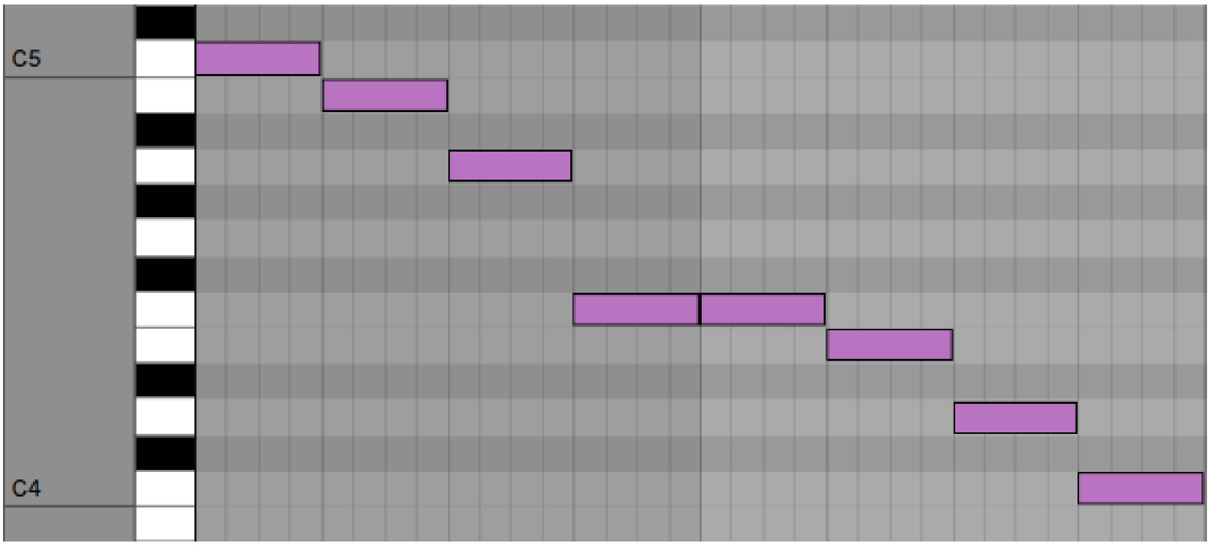


Figure 27: A linear descending melody. Each note is always at an equal or lower pitch than all previous ones.

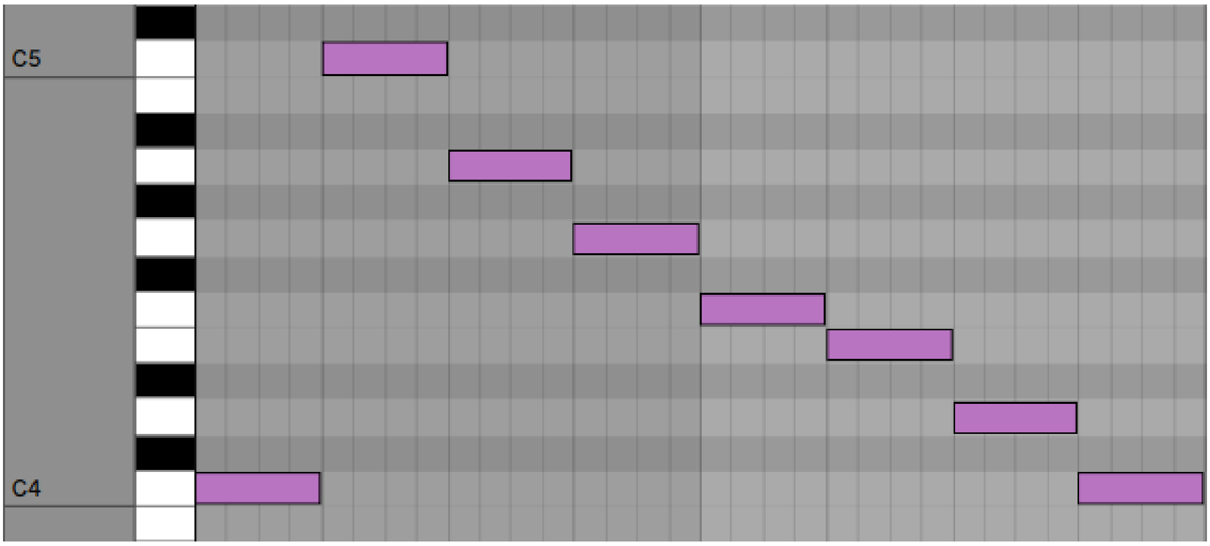


Figure 28: A gap-fill melody. The first note is the lowest in pitch, the second one is the highest, and the following notes form a descending linear pattern.

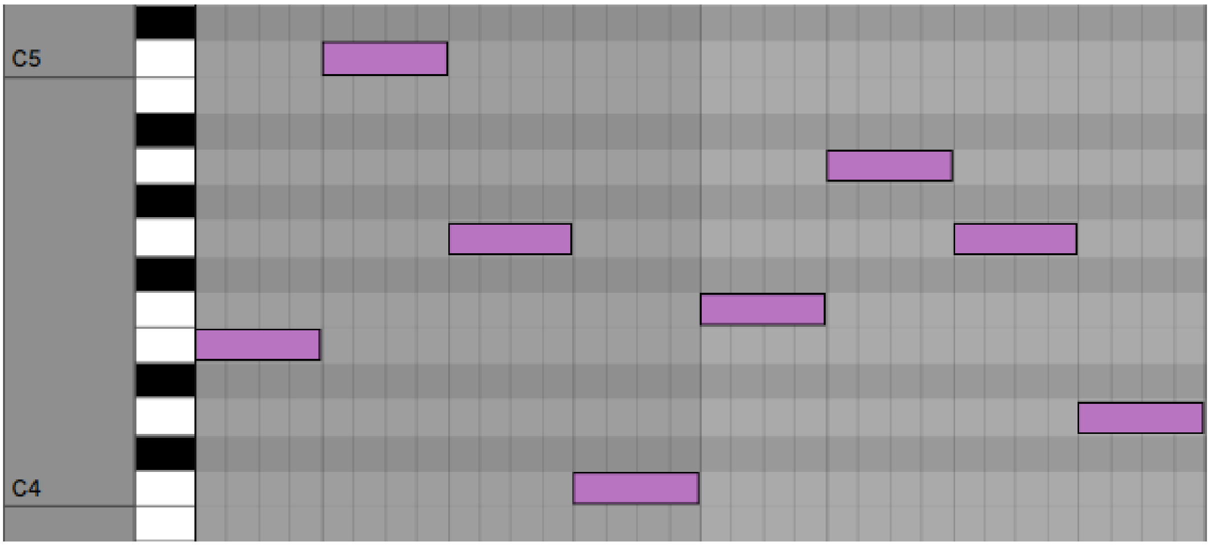


Figure 29: A melody with random contour. The notes do not follow any specific order of sequence.

In terms of code implementation, the note generation and contour shaping procedures were identical with the ones used in Study 1. The inverted part algorithm is presented in Box 3 with pseudocode. When off-scale notes were created by the inversion, their repitching was handled by following the idea of auditory scene analysis regarding how melodic grouping occurs via pitch proximity (Bregman, 1994). In particular, we reduced separation when the note had a pitch distance of 5 or less with the previous note, and increased the separation when the distance was higher than 6. At a (middle-of-octave) distance equal to 6, the direction of repitching was chosen at random.

FUNCTION Inversion

INPUT: NotesList, b, Scale

OUTPUT: InvertedPart

STEPS:

1. Initialise the following variables:

InvertedPart <- []

BottomNote <- MIN(NotesList)

2. For the whole length of NotesList, for every note in position i of NotesList create a new note in position i of InvertedPart. The new note InvertedPart[i] is created by using the equation: InvertedPart[i] = INTEGER(BottomNote - (BottomNote - NotesList[i]))^{b}

3. While InvertedPart[i] is one of the off-scale notes:

3.1 If InvertedPart[i] - InvertedPart[i-1] <= 5, reduce InvertedPart[i] value by 1.

3.2 Else if InvertedPart[i] - InvertedPart[i-1] >= 7, then increase InvertedPart[i] value by 1.

3.3 Else either increase or decrease InvertedPart[i] value by 1 via random choice.

4. Return InvertedPart

Box 3: Pseudocode describing the creation of inverted parts for the stimuli of Study 2.

In relation to the study's aims, the participants were requested to assess the amount of exactness of inversion found in the stimuli. For the purpose of varying the difficulty of this task, we generated stimuli on three levels of the scaling factor b. The levels were at b = 0.67 (reduced intervals), b = 1 (exact inversion), and b = 1.15 (expanded intervals). These levels correspond to stimuli of low, medium and high level of informational complexity, respectively. Further, in order to vary complexity without also varying the intervals, we created another set of questionnaires. For those, the tempo of the melody was varied instead, on two levels (slow was at 85 BPM, and fast at 120 BPM), providing the listeners with a narrow and a wide time frame to process the melodic information. When it comes to the chosen scale, we preferred C major as a comfortable to listen to and widely used scale. There was no change of scales between the stimuli, since Study 1 showed no influence of the scale on any observed statistical relationships. Finally, the melodies used hard quantised quarter notes and had a piano timbre (Native Instrument's "Alicia's Keys" with the velocity of all notes set at the maximum MIDI value of 127).

3.2.3 Participants

A sample of 20 university music students and musicians (N = 9 females) participated in this study, giving their informed consent to participate. The sample consisted of listeners with exposure to either musical instrument playing experience, or music production / composition experience. Drawing from the experience of Study 1, we recruited participants with a musical background, because they are more likely to notice nuanced differences in the kind of stimuli that we are using in this project. Like Study 1, this study was approved by the ethics committee of the Department of Music.

3.2.4 Hypotheses

Participants listened to these melodies through an online questionnaire (partially presented in Appendix). They rated how much they liked each melody, to what extent they think the second part is an exact inversion of the first, and how complex the melody sounded. We hypothesised that listeners would prefer the melodies of which the inverted part omits intervallic jumps, because it will be easier to process the part and therefore infer that it is related to the first part as the inversion of it, making the melody more sensical / less random. Indeed, such longer than note-to-note distance relationships are thought to be a working memory task (Rohrmeier & Koelsch, 2012), raising the question of whether or not listeners have enough capacity to successfully process both the intervallic distances of the second part and its relationship to the first part, and how does their capacity affect the link between expectations and liking. Importantly, musical expectations theory would predict that an exact inversion of the initial melodic segment would be more pleasant, by presenting a true realisation of veridical expectations. The term “veridical” in psychology refers to the degree to which an experience, perception, or interpretation accurately represents reality (APA Dictionary of Psychology, n.d.). Referring to expectations and perception, it concerns the agreement between how a particular listener perceives these factors, and how they might theoretically be perceived by an “ideal listener”, enculturated in a particular music tradition (in our case, western tonal music). We are therefore presented with the questions: 1) Does veridical perception of melodies rely on a mechanism of a probabilistic implication-realisation procedure, or on less cognitively demanding heuristics (visible through the realisation of a relationship only after simplification of the structure while CL is high)? 2) Does aesthetic appreciation rely on the fulfilment of expectations as a basis, or could this be a side effect of a reward mechanism that relates to the sense of successful completion of computational tasks in working memory?

In relation to the first question, based on the current literature we would expect to replicate findings that have shown the use of heuristics and encoding strategies that are less demanding than complete reliance on analysis of statistical properties of the stimuli. For example, Krumhansl et. al (1999) has found that familiarity with stylistic features of the stimuli influences the veridical effects of the knowledge of that style in terms of statistical regularities present in it. Also, when presented with familiar tunes, listeners show a good long-term memory (LTM) retrieval of exact intervallic distances (Dowling, 1970), meaning that they can use a strategy of encoding intervals based on LTM information. Finally, with regard to the perception of short melodies, the contour and intervallic distances vary in their influence on recognition, depending on the amount of delay between the presentation of stimuli; contour was used for a short delay time and intervals for a long delay (Dowling & Bartlett, 1981; Dewitt & Crowder, 1986). This finding suggests a procedure of choice as to what recognition strategy should be employed for better results. In accordance to this literature, we predict to find that differentiations in cognitive load of the stimuli will lead to different perception of the relations between the presented melodic segments, because the minimisation of intervallic distances of the inverted part will lead to the availability of a contour-only analysis of the statistical similarities presented in the segments. Further, the ease of processing induced by longer available time frames will lead to an alteration in the perception of a similarity between segments due to more available time for analysis through the strategy of using intervals.

As far as the second question is concerned, we would predict, based on the literature of musical expectations, that stimuli possessing statistical regularities and patterns of an “exact inversion” relationship between the two melodic parts, would be more aesthetically preferable due to better fulfilment of expectations. However, it is worth noting that the present study intends to challenge this idea and test the aesthetic appreciation of stimuli under several cognitive load stressing factors that intend to force listeners into using a variety of strategies for the analysis of the intra-musical properties of the melodies. If they tend to prefer less expectation-fulfilling stimuli and at the same time they misinterpret the degree of present regularities in them, a possible answer to our question would be that aesthetic appreciation depends on rewarding that derives from a perceived sense of successful analysis of the stimulus and not simply from the stimulus providing a resolution that was probabilistically likely to occur. To summarise the main elements of the study, we present the hypotheses and variables below. Note that similarity of the two segments is not a variable - the segments are considered identical when the interval size variable is set in “exact intervals”, and the precise deviation from this when it is set in “contour only” can be calculated with statistical parameters.

Independent variables:

- Interval size of second part (reduced, exact, expanded intervals)
- Tempo (slow, fast)
- Melodic shape (linear ascending, linear descending, gap fill, random)

Dependent variables:

- Ratings of liking.
- Ratings of perceived exactness of inversion.
- Rating of perceived complexity.

Hypotheses:

- **H_0 about veridical perception and intervallic distances:** Interval size of the second part does not affect ratings of exactness of inversion for any melodic shape and tempo (mean rating of small interval melodies is equal to the mean rating of melodies with exact and expanded intervals).
- **H_0 about veridical perception and time frame:** Tempo does not affect ratings of exactness of inversion for any melodic shape and tempo (mean rating of fast melodies is equal to the mean rating of slow melodies).
- **H_0 about aesthetic appreciation and veridical perception:** Aesthetic appreciation is in accordance with probabilistic expectations formed by similarity between the melodic segments, for any melodic shape and tempo (ratings of liking are not higher in small intervallic distances or slower tempo).

3.2.4 Data collection

The study was implemented with the use of online questionnaires that were sent to students of the University of Sheffield via email, and other university students via social media distribution. Structurally, the questionnaires consisted of a series of melodic stimuli. After the presentation of each melody, the responders were asked to rate how much they liked the melody, to what degree they think that the second part of the melody is the exact inversion of the first part, and how complex they perceived it to be. Each one of the series presented various melodies of the same contour archetype. The stimuli were split into questionnaires that varied the b factor (intervallic jumps of the inverted part), and those that varied the tempo of the melodies. Finally, the participants were asked about their age and their gender.

Conclusively, the variables under consideration are those of liking, perceived exactness of inversion, and perceived complexity. Their role in the study is documented in the Hypotheses section of 3.2.1. The variable of gender was also collected, in order to control for possible effects (Christenson & Peterson, 1988; Colley, 2008). The analysis considering these variables could possibly be implemented with a Related Samples and an

Independent Samples way. For each of them, the dataset would have to be structured in different ways.

3.2.5 Datasets

Related samples dataset

The first step of the related samples analysis is the reconstruction of the raw dataset of the questionnaire, in a way that makes participants act as data points. To achieve this we created new variables based on averaged data from the questionnaire's Likert scales. These variables are divided into scores of melodies with low, medium and high complexity, mirroring the "reduced intervals", "exact intervals", and "expanded intervals" types of melodies.

In the cases of questionnaires where tempo was the factor that affected complexity, only the low and medium complexity variables were given values. Tempo is a binary factor (slow and fast), and the choice of linking each value with the low and medium complexity variables was decided based on the mean perceived complexity values from the raw dataset. Slow melodies had a mean perceived complexity of 3.04, and "reduced intervals" melodies had a mean of 3.15, making them their nearest neighbours. Similarly, fast melodies had a mean rating of 3.56, which made "exact intervals" melodies their closest pair with a mean of 3.53.

Table 12: Mean values of perceived complexity scales among all melodies, split by complexity levels.

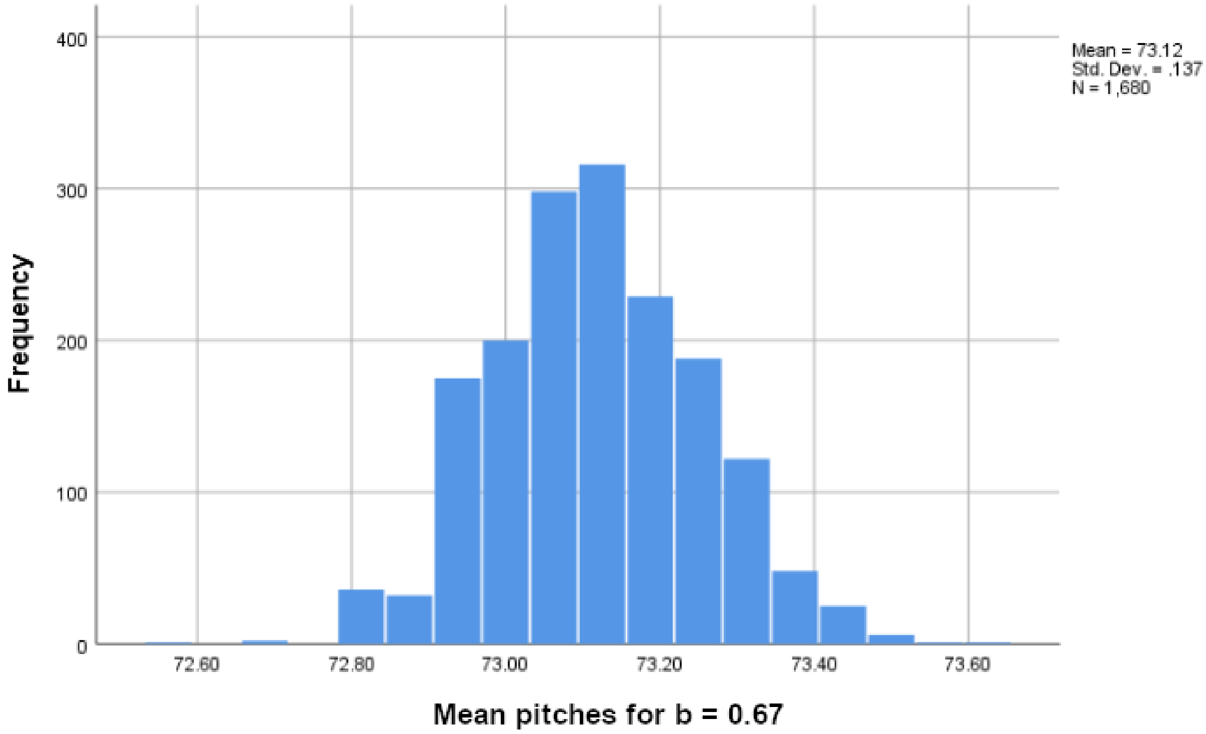
	Slow tempo	**Fast tempo**	**Reduced intervals**	**Exact intervals**	**Expanded intervals**
N of melodies	9	9	20	20	11
Mean	3.037	3.556	3.146	3.533	3.621

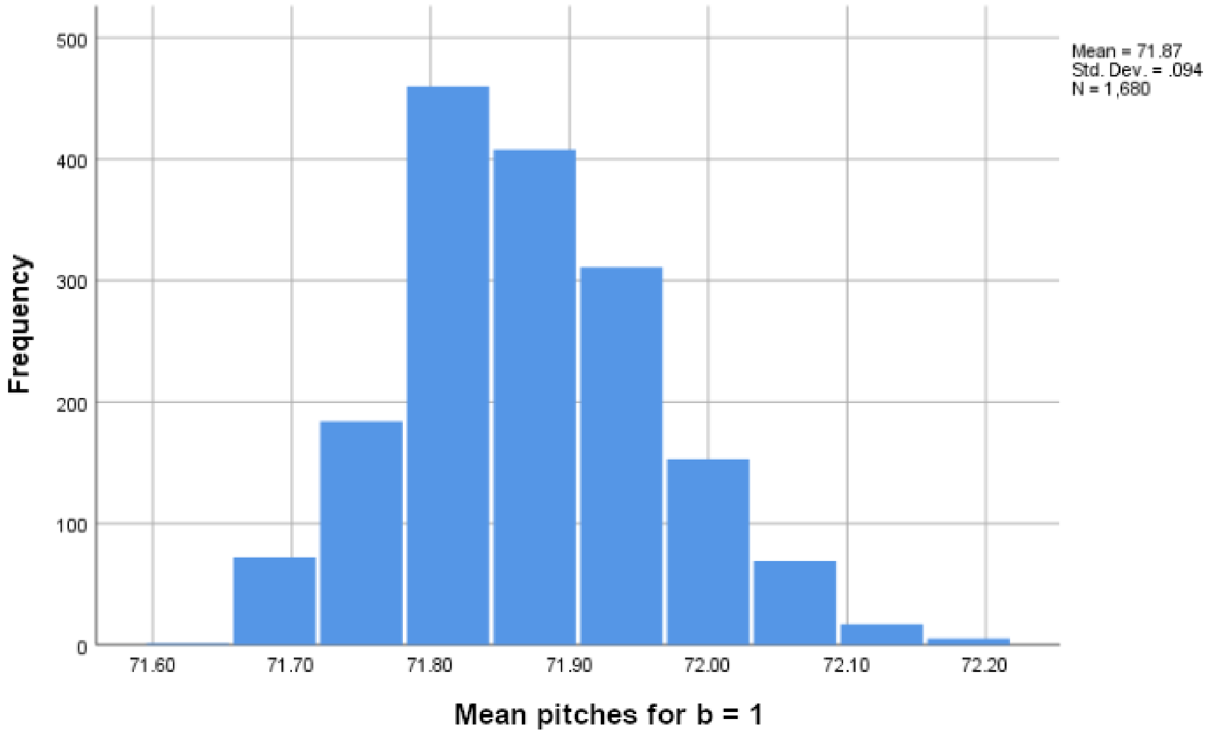
Independent samples dataset

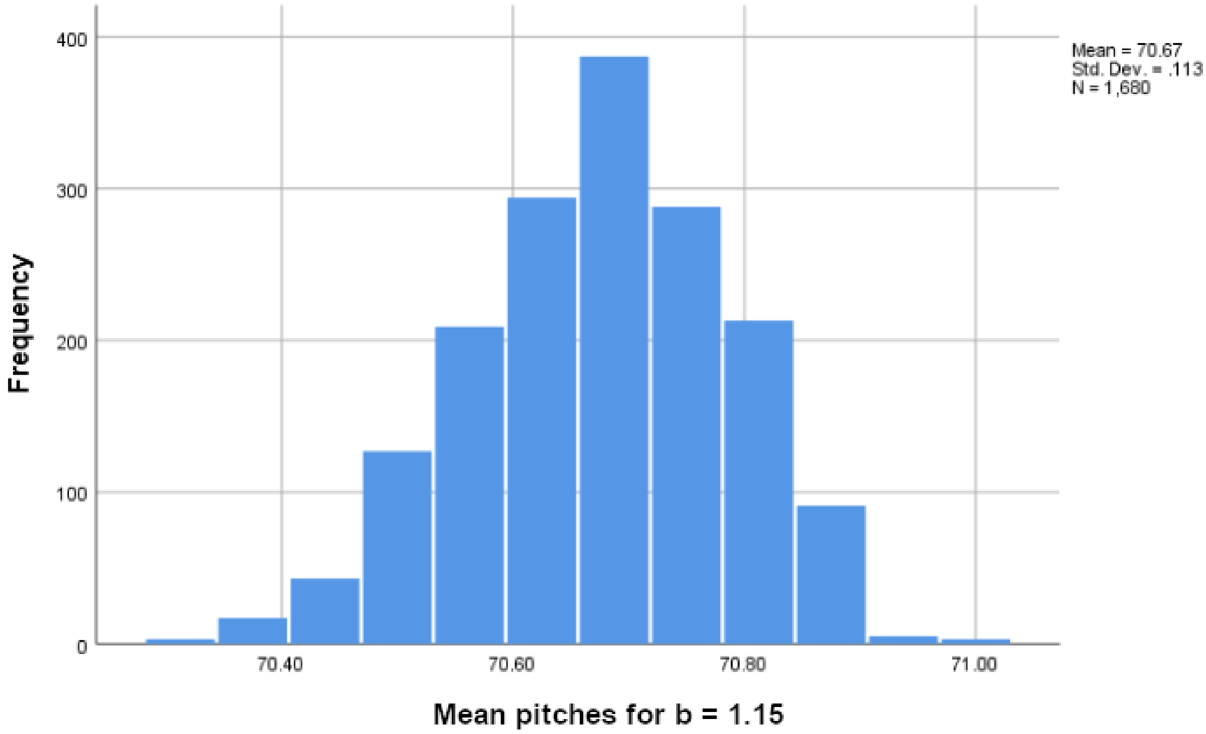
Another approach for this analysis would be to treat the raw questionnaire responses of all respondents as one variable, for each of the liking, exactness of inversion & perceived complexity scales. A prerequisite for doing this is to ensure that all participants had a similar behaviour in how they rated the scales. Specifically, it is to check that the distributions of responses are homoscedastic and with similar central tendencies. The methods we used for this are intraclass correlation coefficient (ICC) for testing the agreement between raters, and homogeneity of variance examination with Levene's Test. For the implementation of ICC, the dataset consists of variables that each of them represent one rater's responses for a particular scale. For the purpose of testing the equality of variance, the dataset consists of liking, exactness of inversion and perceived complexity variables that include all responses from all participants. The responses are tied to a nominal Participant ID variable, in order to allow analysis of variance testing.

Statistics of materials

Before proceeding with the results of the study, we present an analysis regarding the distributional characteristics of the stimuli. We will look at the parameters of mean and variance for each type of melody, as a way to assess the effect of interval sizes and tempo, which are related to the complexity of the stimuli. This will be compared to the categorisation of the melodies into complexity tiers based on perceived complexity scores, that was shown in Table 12. For this analysis we generated melodies of all pitch distance (b) levels (0.67, 1, 1.15). The total number of melodies is 5040, split equally to 1680 for each b level. An initial view of the frequency distributions of notes is shown below, for each complexity tier. The plots consider the mean and the variance of each of the generated melodies. The numbers shown in the histograms refer to MIDI note numbers.

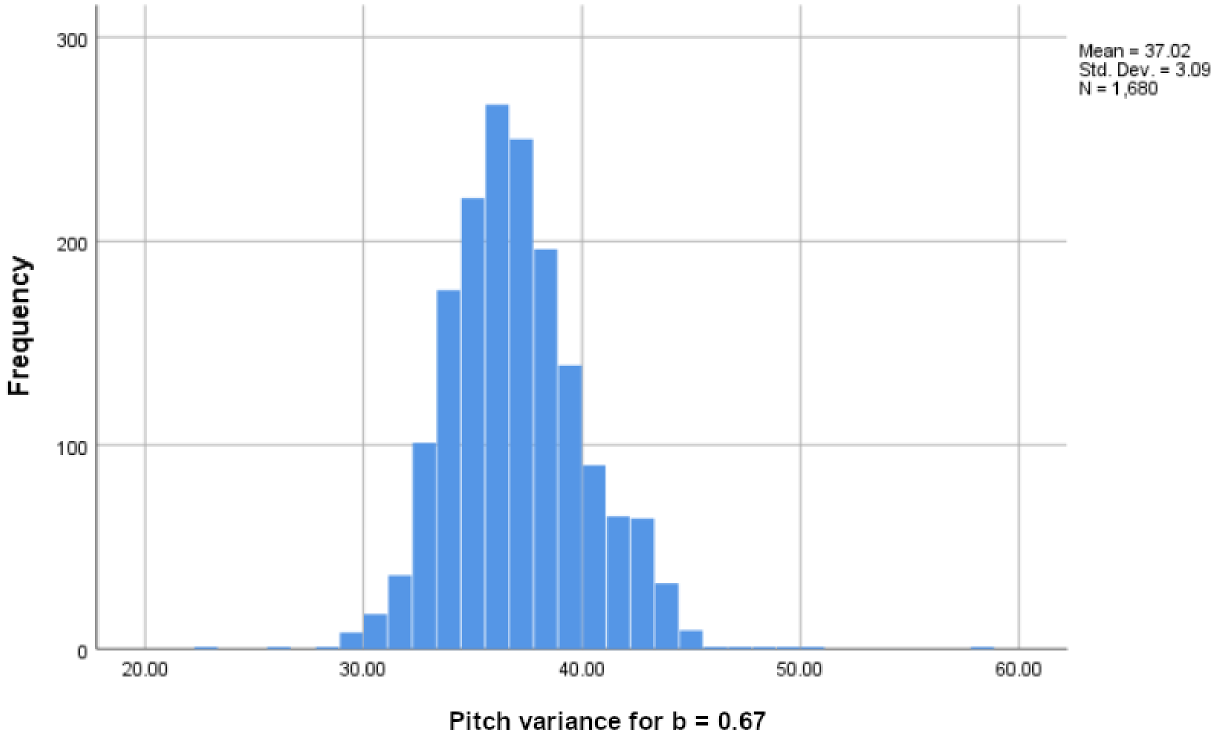

Mean = 73.12
Std. Dev. = .137
N = 1,680
Frequency
400
300
200
100
0
72.60
72.80
73.00
73.20
73.40
73.60
Mean pitches for b = 0.67

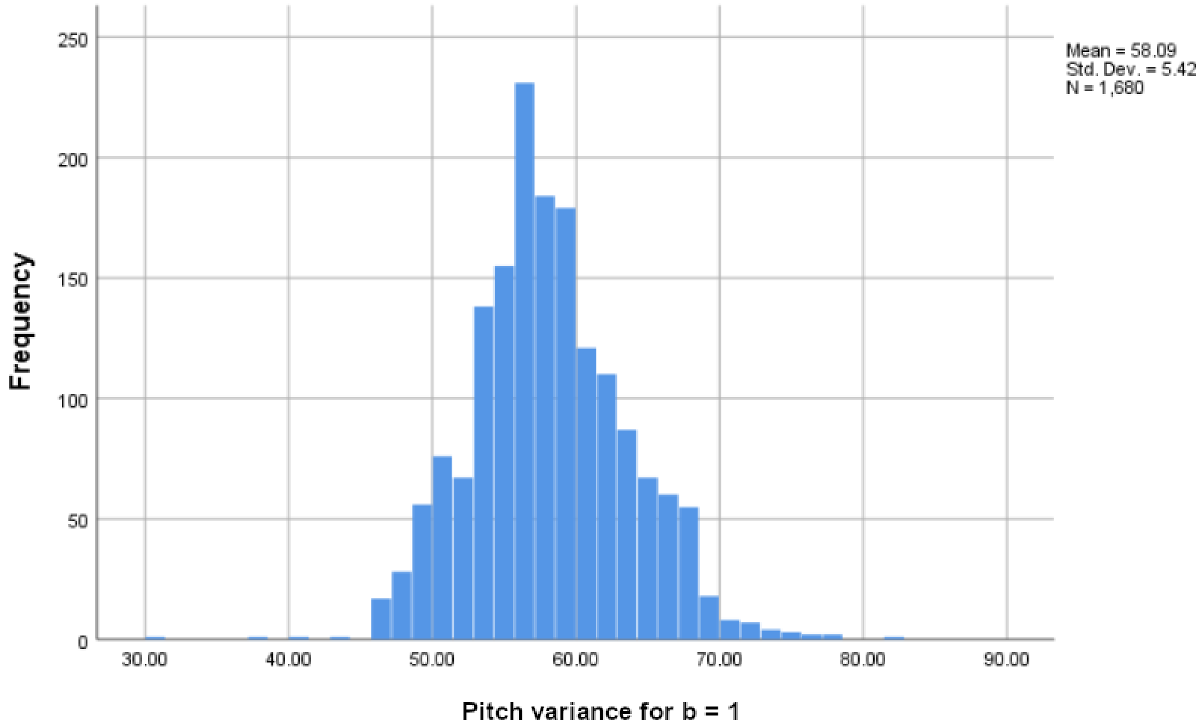

Mean = 71.87
Std. Dev. = .094
N = 1,680
Frequency
500
400
300
200
100
0
71.60
71.70
71.80
71.90
72.00
72.10
72.20
Mean pitches for b = 1

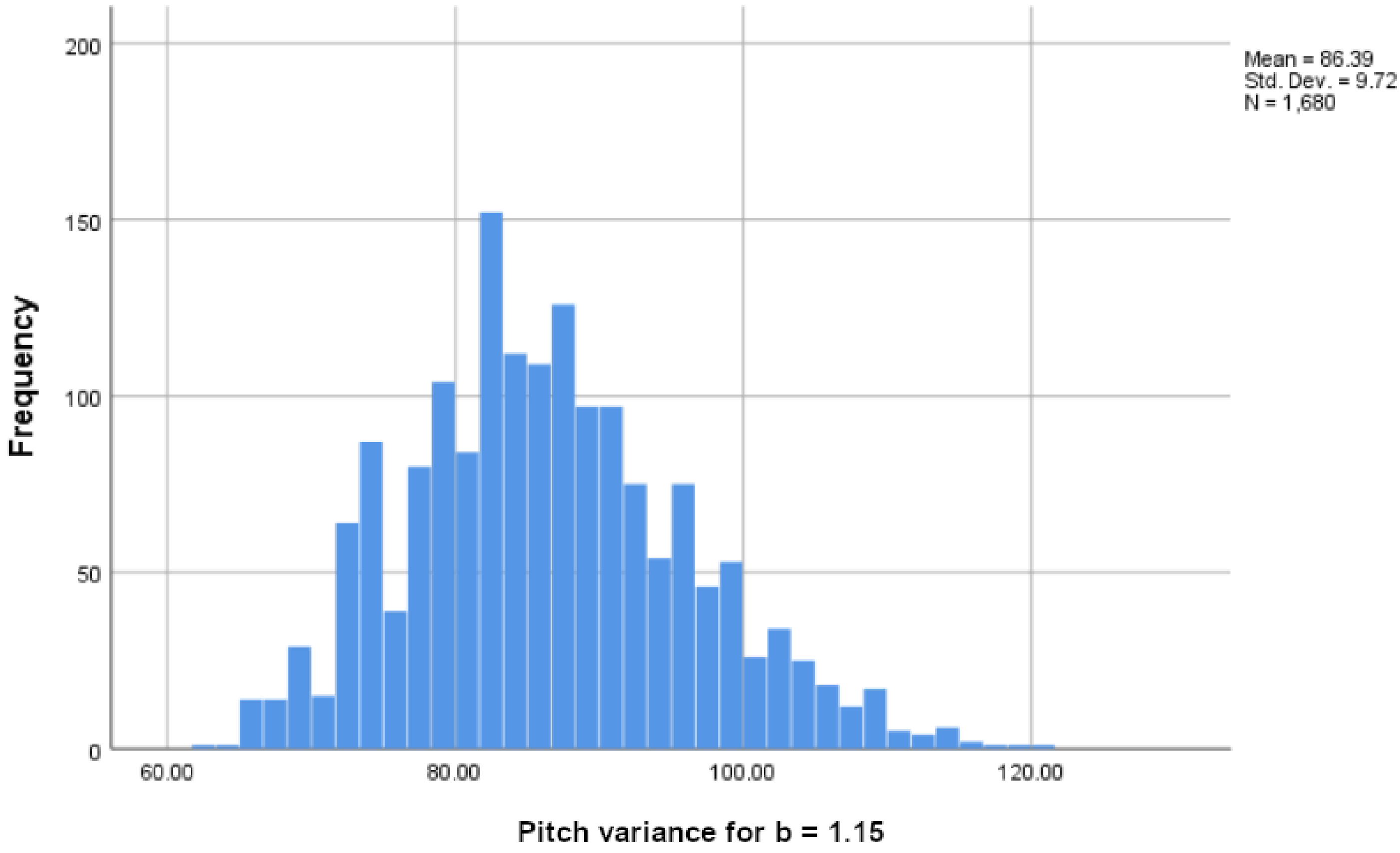


Figures 30-32: Mean MIDI note values from a pool of melodies created with the generation script, with varying interval sizes.

It can be noted that as the b parameter increases, the mean pitch value decreases. This happens because melodies reach to further (lower) notes as the intervals of the inverted part get bigger. We also looked at the variances of the pitches for each b tier.

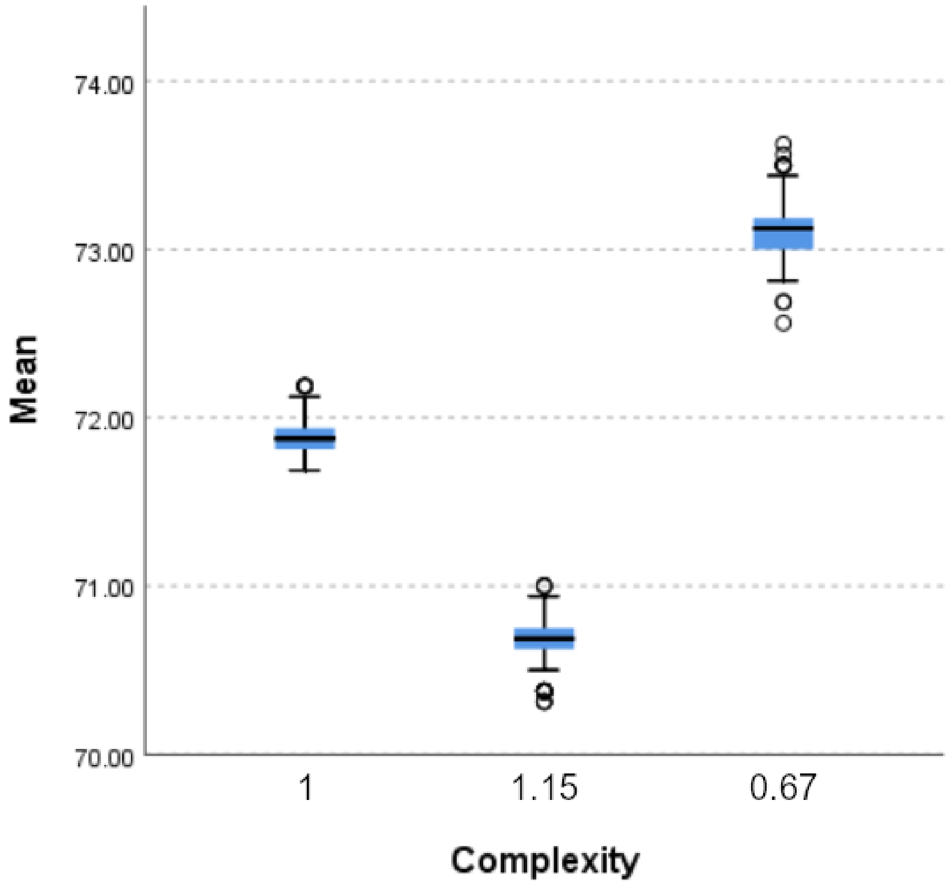

Mean = 37.02
Std. Dev. = 3.09
N = 1,680
Frequency
300
200
100
0
20.00
30.00
40.00
50.00
60.00
Pitch variance for b = 0.67

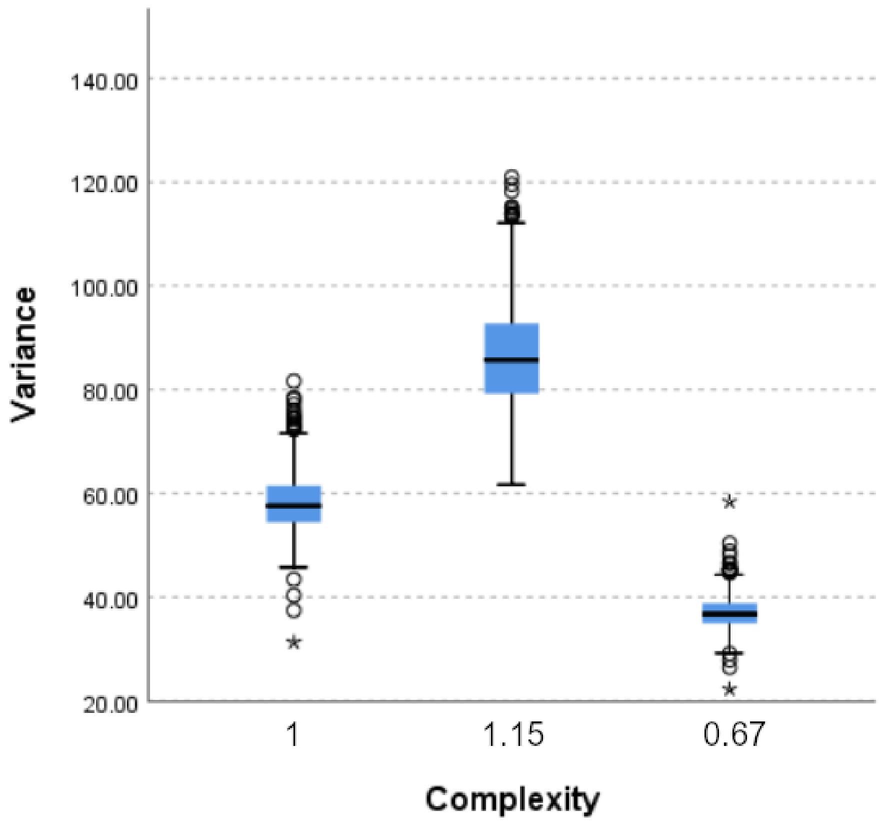

Mean = 58.09
Std. Dev. = 5.42
N = 1,680
Frequency
250
200
150
100
50
0
30.00
40.00
50.00
60.00
70.00
80.00
90.00
Pitch variance for b = 1



Figures 33-35: Histograms of the variance of pitch values from a pool of melodies with varying levels of interval sizes.

Similarly, Figures 33-35 show that with higher values of b, the pitches attain greater variability. This is an observation that we expected due to the rise of pitch gaps in the inverted part, and it is the reason why b is considered a parameter that is linked with informational complexity. What is shown with these histograms can be further verified with statistical testing. None of the distributions in this analysis is normal, as shown by an assessment of normality with Kolmogorov-Smirnov tests (0% significance). Furthermore, Kruskal-Wallis tests showed differences between the distributions at 0% significance level. These differences can also be observed in the respective K-W boxplots, showing that the chosen levels of complexity result in melodies that are different enough to allow meaningful comparisons in the study.


74.00
73.00
72.00
71.00
70.00
Mean
1
1.15
0.67
Complexity


140.00
120.00
100.00
80.00
60.00
40.00
20.00
Variance
1
1.15
0.67
Complexity

Figures 36-37: Kruskal-Wallis boxplots that show differences between statistical distributions of melodic characteristics (here, mean and variance of pitch distributions) in melodies of different interval size levels.

Finally, we can look at the pairwise comparisons tables of these distributions (13-14), to locate precisely where the K-W differences are found.

Table 13: Pairwise comparisons of the mean distributions of each b level.

Sample 1 - Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig. (Bonferroni corrected)
1.15 - 1	1680	50.121	33.519	.000
1.15 - 0.67	-3360	50.121	-67.038	.000
1 - 0.67	-1680	50.121	-33.519	.000

Table 14: Pairwise comparisons of the variance distributions of each b level.

Sample 1 - Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig. (Bonferroni corrected)
0.67 - 1	1679.863	50.204	33.461	.000
0.67 - 1.15	-3353.433	50.204	-66.796	.000
1 - 1.15	-1673.571	50.204	-33.335	.000

As Tables 13 and 14 show, all distributions are different from each other. This points to the mean values of mean and variance distributions being statistically significantly different between all interval size levels. The order of the mean values of variances shows that higher intervallic distances equaled higher variance. This effect mirrors the perceived complexity ratings as shown in Table 12.

In conclusion, the way we structured the dataset is in agreement with both the complexity introduced by variance of b, and the perceived complexity of the same types of stimuli as rated by participants. It is important to note that complexity is not a universally identical term with variability, but in the context of this research variability is linked with complexity, in the sense of informational complexity - richness of information (bigger number of pitch distance values). We can now proceed to the main analysis of the responses that were obtained from the questionnaires. The analysis is split into two ways, namely Related Samples and Independent Samples. Each of them draws from the two relevant datasets as discussed in the beginning of 3.2.5.

3.3 Results

3.3.1 Related samples

Correlation analysis

At the beginning of this analysis correlations between the liking, complexity and exactness of inversion variables are explored for all melodic complexity levels. In order to decide what test to use we inspected for outliers and the linearity of our values, as they are important assumptions for using Pearson's r, the most common test for assessing correlations. There were no values that surpass a distance limit of more than 3.29 sd from the mean for any variable, so there were no outliers found.

Table 15: Descriptive statistics of Likert scales.

	N	Minimum	Maximum	Mean	Std. Deviation
Simple_Liking	20	1.67	5.5	3.39	.9
Simple_ExactInv	20	1.17	6.33	4.22	1.45

Simple_Compl	20	1.67	5	3.15	.89
Normal_Liking	20	1.83	6.34	3.58	1.11
Normal_ExactInv	20	1.67	6.34	4.33	1.35
Normal_Compl	20	2.17	6.17	3.53	1
Expand_Liking	11	2.33	5.67	3.42	.96
Expand_ExactIn v	11	2	6	3.98	1.24
Expand_Compl	11	2	6	3.62	1.03

When testing for linearity, there appeared to be cases of nonlinear relationships between our variables.

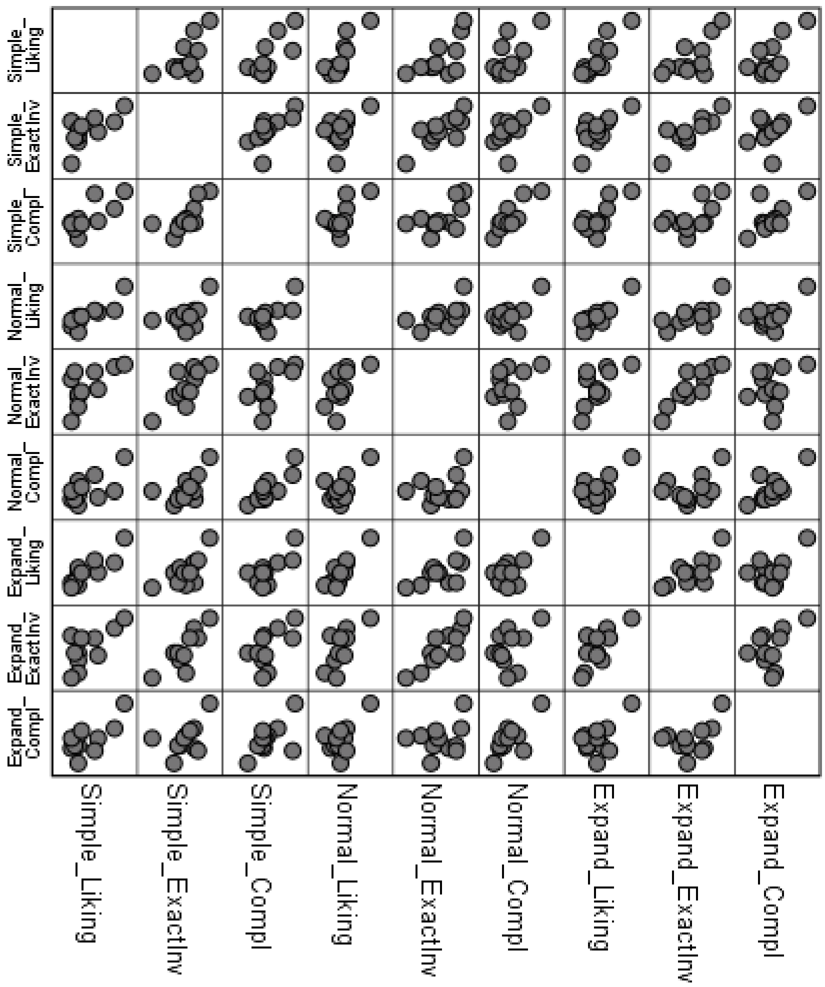


Figure 38: Scatterplot matrix of all scales.

Therefore, we used the non parametric correlation test of Kendall's Tau, instead of Pearson's r. After adjusting the p-values of the bivariate correlations by using the Benjamini-Hochberg method of multiple testing correction, we found significant correlations with a False Detection Rate of 5%, which are presented in Table 16.

Table 16: Statistically significant correlations between complexity levels for different rating variables.

Correlation	**Significance**	**Rating Variable**	**Melodic Variables**
Moderate (0.564)	5%	Liking ratings	Low and Medium complexity
Moderate (0.692)	5%	Liking ratings	Medium and High complexity
Moderate (0.66)	5%	Liking ratings	Low and High complexity
Moderate (0.599)	1%	Exactness of Inversion ratings	Low and Medium complexity
High (0.748)	5%	Exactness of Inversion ratings	Medium and High complexity
Moderate (0.667)	1%	Complexity ratings	Low and Medium complexity

Table 17: Non-parametric bivariate correlations of all scales. The shown significance is before correcting for multiple testing.

		Simple_ Liking	**Simple_ ExactInv**	**Simple_ Compl**	**Normal_ Liking**	**Normal_ ExactInv**	**Normal_ Compl**	**Expand_ Liking**	**Expand_ ExactInv**	**Expand_ Compl**
Simple_Liking	Correlation Coefficient	1	.126	**.33***	**.564****	**.361***	.226	**.66****	.419	.457
	Sig. (2-tailed)	.	.452	.049	.001	.031	.179	.007	.081	.057
Simple_ExactInv	Correlation Coefficient	.126	1	.247	-.038	**.599****	.181	**.510***	**.574***	.315
	Sig. (2-tailed)	.452	.	.141	.819	.000	.280	.033	.015	.183
Simple_Compl	Correlation Coefficient	**.33***	.247	1	.236	**.328***	**.667****	.427	.343	.419
	Sig. (2-tailed)	.049	.141	.	.159	.050	.0	.079	.154	.081
Normal_Liking	Correlation Coefficient	**.564****	-.038	.236	1	**.327***	.264	**.692****	**.509***	.245
	Sig. (2-tailed)	.001	.819	.159	.	.050	.116	.004	.033	.306
Normal_ExactInv	Correlation Coefficient	**.361***	**.599****	**.328***	**.327***	1	.219	**.534***	**.748****	.187
	Sig. (2-tailed)	.031	.000	.050	.050	.	.191	.027	.002	.432
Normal_Compl	Correlation Coefficient	.226	.181	**.667****	.264	.219	1	.346	.094	**.509***
	Sig. (2-tailed)	.179	.280	.0	.116	.191	.	.153	.694	.033
Expand_Liking	Correlation Coefficient	**.66****	**.510***	.427	**.692****	**.534***	.346	1	**.519***	.135
	Sig. (2-tailed)	.007	.033	.079	.004	.027	.153	.	.032	.578
Expand_ExactInv	Correlation Coefficient	.419	**.574***	.343	**.509***	**.748****	.094	**.519***	1	.170
	Sig. (2-tailed)	.081	.015	.154	.033	.002	.694	.032	.	.478
Expand_Compl	Correlation Coefficient	.457	.315	.419	.245	.187	**.509***	.135	.170	1
	Sig. (2-tailed)	.057	.183	.081	.306	.432	.033	.578	.478	.

Comparison of distributions

Before proceeding with comparing the distributions of the questionnaire data, the shape of the distributions was tested for normality. A Shapiro-Wilk test did not produce results that could allow the rejection of the null hypothesis after correcting for multiple testing, for any variable. Therefore, the comparisons of distributions were done with related samples T-tests. With an FDR of 5%, there was a 5% significant difference in mean values between the perceived complexity ratings in low and medium complexity melodies (0.388 higher in medium complexity). No other statistically significant differences were found after correcting p-values for multiple comparisons.

3.3.2 Independent samples

Comparison of distributions

ICC tells us if raters gave similar ratings for each of the scale items in the questionnaire. In the SPSS implementation that was used in this analysis, the ICC assessment includes a statistic that shows whether the correlation coefficient is statistically significantly different from 0. In our case, the ICC test did not provide statistically significant results using a consistency definition.

Table 18: Intraclass correlation coefficient of two way random effects and consistency.

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.045	-.076	.324	1.284	11	55	.258
Average Measures	.221	-.743	.742	1.284	11	55	.258

A more liberal approach would be to only require similar statistical parameters of the distributions of raters' responses. For this reason we tested the equality of variances between participants for all scales, by using Levene's test. The Levene Statistic was found statistically significant at a 5% level for all three liking, exactness of inversion and perceived complexity variables, meaning that the hypothesis of equal variances was rejected. With these findings, it is shown that the ratings of participants do not follow the same distributions, and putting them under the same variable would change the original distributional properties of each independent participant's responses.

Table 19: Levene's test for the homogeneity of variances of responses among participants for all Likert scales.

		Levene Statistic	**df1**	**df2**	**Sig.**
Liking	Based on Mean	5.210	19	286	.000
	Based on Median	3.078	19	286	.000
	Based on Median and with adjusted df	3.078	19	192.339	.000
	Based on trimmed mean	5.008	19	286	.000
ExactInv	Based on Mean	1.979	19	286	.010
	Based on Median	1.589	19	286	.058
	Based on Median and with adjusted df	1.589	19	229.441	.060
	Based on trimmed mean	2.006	19	286	.008
Complexity	Based on Mean	1.926	19	286	.012
	Based on Median	1.038	19	286	.417

	Based on Median and with adjusted df	1.038	19	205.966	.419
	Based on trimmed mean	1.826	19	286	.020

3.3.3 Summary

To summarise, the results show a statistically significant difference between the perceived complexity ratings in low and medium complexity melodies, meaning that the participants linked the increase in intervallic jumps with the perception of a more complex melody. This is also visible in the lack of correlation between low and high complexity melodies when it comes to ratings of perceived complexity. In relation to the hypotheses of 3.2.1, the current data does not suffice to support the rejection of any null hypothesis.

3.4 Discussion

3.4.1 Research questions

In 3.2.1 we focused on two main research questions. Namely: 1) What kind of mechanism do listeners use in order to recognise relationships between melodic segments? 2) How do listeners receive aesthetic pleasure from successfully recognising such relationships?

About the first question, current literature (as discussed in 3.1), allows us to expect that the complexity of the stimuli would influence how accurately listeners recognise the relationship between the inverted part and the initial part of the melodies. In terms of variables, this translates to an expectation of getting results that would show the mean exactness of inversion rating of small interval melodies to be different than the mean rating of melodies with exact and expanded intervals, for each participant. For the tempo-altering stimuli, the corresponding results should show that the mean rating of fast melodies is not equal to the mean rating of slow melodies. The results of this experiment did not show statistically significant differences in the aforementioned pairs of means, and contrary to what we could expect from other studies on this topic, the null hypotheses of 3.2.1 that are related to exactness of inversion ratings cannot be rejected.

Concerning the second question, an ease of processing caused by smaller intervals and slower tempo allowed us to anticipate that these stimuli would be perceived as more aesthetically pleasing, which could be interpreted as an effect of permitting easier recognition of existing patterns between the two distinct parts of the melodies. Accordingly, the ratings of liking would be higher for melodies with small intervallic distances and slow tempo. However, like the results for the first question, the results of related samples comparison of the means for the liking ratings did not show a statistically significant difference between the lower and higher complexity melodies.

Technically, when interpreting these results under the perspective of how complexity influences ratings, we make the assumption that bigger intervallic gaps and faster tempos lead to more complex melodies. This can be supported by the increase in data variability and richness of information, as shown in 3.2.5. Importantly, in the present study it was found that participants did rate the melodies with bigger intervallic gaps and faster tempos as more complex than the simpler ones, at a 5% significance level (that is, without faster melodies having more notes). Therefore we find that this connection can be claimed not only through distributional properties, but also by the listener’s subjective judgements as revealed by their responses.

3.4.2 Conclusion and further research

In summary, the findings of this study did not show that differences in the complexity of musical material changed the recognition of patterns between musical segments. Further, liking ratings did not differ significantly under varying complexity levels. The role of complexity could have been different if the range of it was brought to more extreme low and high points. Naturally, one-note melodies (lowest complexity) and chaotically complicated note-heavy melodies would produce, at minimum, changes in exactness of inversion ratings. That is to say, the present study is not directly comparable to the ones concerning the theory of an inverted U relationship between complexity and aesthetic responses. However, while Dowling & Bartlett (1981) and Dewitt & Crowder (1986) found through a series of experiments that intervallic information and contour influences the recognition of similarity between melodies, this was not the case in the present study. An explanation for this contrast is the complicated nature of the task of recognising the amount of how inverted a melody is, since the two previous studies used exact copies of the melodies.

Overall, the findings of the current study support that intervallic information and tempo affect the perceived complexity of melodies, with bigger intervals and higher tempos resulting in higher perceived complexity. Further, complexity did not influence the perception of patterns or liking ratings among listeners. More steps should be taken towards the exploration of how cognitive load and working memory affect the aesthetic appreciation of melodies, possibly by adding a non music related task that uses the same two mechanisms and using a neuroimaging approach. This is a suggestion based on the idea of analysing brain activity of an unrelated cognitive load - working memory task and comparing it with a musical task that will use the same mechanisms. Such a design could bring insight into how music appreciation relates to these mechanisms by means of correlating brain activity, rather than by varying the levels of parameters during a sole musical task.

4. Development of the MMM Generator and evaluation study

4.1 Introduction

For the third part of the project we focused on the challenge of developing a music generator software with the use of a music cognition based framework. This approach could allow us to tackle possible issues found in machine learning based generators, and offer an alternative solution to gaps presented in other commercial generator plugins when it comes to user interfaces and controls. The MMM Generator (the plugin of this project) was made possible by using as a basis the outcomes, insights and ideas that resulted from the first two studies, in algorithms that aim to generate useful musical material for the modern music production industry. The algorithms were developed in the setting of an application for digital audio workstations, in the form of a VST plugin. In this chapter we focused on presenting the design of the generation functions, the use of academic methodology, and the utilisation of user feedback into the process of making this software. Prior to analysing the process of creating the software, we present an overview of the academic literature that is related to the most common techniques and procedures in the area of music generation algorithms. Also, we give a number of examples of commercially available VST plugins that are designed to generate MIDI music and discuss their design and their relation to our own software.

The value of a musical tool such as this rests in whether musicians find it practically useful. So we undertook a thorough user evaluation of it, which is introduced in the remainder of this chapter, following the presentation of the MMM Generator. The aim of this study was to understand the degree to which we managed to reach the aim of developing a useful music making product that is built based on music cognition theory, and to determine if it provides an output of high quality and usability for DAW-based producers. Doing so, by using a combination of qualitative and quantitative assessment tools to examine the usability and quality of the MMM Generator, we are addressing the broader idea of how academic literature can aid in the development of DAW plugins, either for the purposes of improving currently available tools, or by helping with the creation of new and unique ones.

4.1.1 Music generation algorithms - literature and existing programs

With regard to music generation literature, efforts have been made with the goal of producing aesthetically pleasing melodies, chords, rhythms and their combinations. The principal methods that academics have used for the creation of music generators are found in the broad framework of artificial intelligence (see for a summary Nierhaus, 2009). The most common way is that of deep neural networks (Briot, Hadjeres & Pachet, 2017), which researchers have incorporated in various forms. Deep networks are systems that operate in two distinct phases. Initially, the system is trained with the use of a dataset that is responsible for providing examples to it. The parameters of the network are tuned with a procedure that tests its performance, as it observes statistical regularities that occur in the raw data, and uses layers that extract progressively more abstract features based on these regularities. After this training phase, the next main phase is the generation of output. It involves input data that is used to initiate the generation process, and the output data which is the outcome of the deep network's operation.

A main approach is that of recurrent neural networks (Eck & Schmidhuber, 2002; Lyu, Wu, Zhu & Meng, 2015; Chu, Urtasun & Fidler, 2016; Jaques et. al, 2017), which specialise in processing long sequences of values, usually of variable length. Specifically, Long Short-Term Memory (LSTM) RNN networks are considered the most preferred option in music generation (Kulkarni et. al, 2019), as they take the output of the previous time-step in a generation process and use information from it in order to create its next output, imitating the interdependence of notes in human made musical phrases. However, a new architecture termed as transformer has appeared, making use of the idea of an attention mechanism. This mechanism focuses on some specific elements of the input sequence at each time step, and has provided even better long-term structure results than RNN in music generation (Huang et. al, 2018). Alternatively, convolutional neural networks, which are successful in many practical applications and specialise in processing grids of values, like time series and pixels in images (Goodfellow, Bengio & Courville, 2016), are also used in music generation (Yang, Chou & Yang, 2017; Lattner, Grachten & Widmer, 2018; Huang et. al, 2019), but due to lacking the temporal depth of LSTM or transformer they are not considered as efficient. Another important approach is autoencoders (Bretan, Weinberg & Heck, 2016; Roberts, Engel, & Eck, 2017; Roberts et. al, 2018), a type of neural network that is used to learn features in an unsupervised setting with unlabelled data, and is particularly efficient in dimensionality reduction and information retrieval tasks. In music generation, they are used for the benefit of effective high level feature extraction, and in many cases they are combined in architectures that include LSTM networks (Fabius & van Amersfoort, 2015; Hadjeres, Nielsen & Pachet, 2017).

Similarly, genetic algorithms, which are models designed for optimisation problems and are inspired by the biological processes of natural evolution, have been used for algorithmic composition, especially in earlier years of music generation (Matić, 2010; Eigenfeldt & Pasquier, 2013; Ponce de León, Inesta, Calvo-Zaragoza & Rizo, 2016; Sulyok, McPherson & Harte, 2016). Finally, some academics have used alternative and more strictly mathematical procedures such as Markov models (Van Der Merwe & Schulze, 2011; Hadjeres, Pachet & Nielsen, 2017). These models are stochastic processes that describe a sequence of possible events, where the probability of occurrence of each event depends solely on the state of the previous event. Case-specific algebraic and statistical models (Drewes & Högberg, 2007; Whorley & Conklin, 2016), as well as pattern detection based methods (Collins, Laney, Willis & Garthwaite, 2016; Conklin, 2016) have also been used as generation approaches. The literature of music generation algorithms is, in general, based on the training of algorithms with musical corpuses of human made compositions, and on the utilisation of observed patterns in such human-made music.

A domain that is close to the one of music generation algorithms, in the sense that it is also a quantification and / or coded implementation of musical elements, is that of computational and statistical music cognition. In that field of study the quantification process is about the perception of musical elements, as well as the modelling of patterns formed by such elements. The main literature of this nature is concerned with topics such as the modelling of what melodic continuations do listeners prefer (Narmour, 1992; Krumhansl, 1997; Schellenberg, 1997), what instrumentations and frequency contents have the best chance of becoming popular in various years (Ni, Santos-Rodriguez, Mcvicar & De Bie, 2011; Herremans, Martens & Sörensen 2014), in what ways listeners perceive musical elements (Lerdahl & Jackendoff, 1996; Eichert, Schmidt & Seifert, 1996; Cambouropoulos, 1997; Tillmann, Bharucha & Bigand 2000; Eerola, Himberg, Toiviainen & Louhivuori, 2006; Fritz et. al, 2013; van der Weij, Pearce & Honing 2017), and the statistical modelling of musical properties and patterns (Werts, 1997; Margulis, 2005; Temperley, 2008; Pearce, 2018; Harrison & Pearce, 2020).

Reflecting on these trends, the framework of deep learning represents the main approaches to music generation, with the use of neural network architectures. Advances of recent years, including the availability of massive amounts of data and the increase of strong and affordable computing power, have resulted in the development of projects that can learn rules and create musical material in training settings of non “hand-picked” corpuses of music. The big majority of projects output symbolic representations of music, but a few audio producing systems exist too, like WaveNet (Van Den Oord et. al, 2016). Such systems have the benefit of modelling audio signals, resulting in projects like human speech mimicking, and offer an alternative way in music generation, e.g. by synthesising audio that is based on a training dataset of classical piano music. However, symbolic representations have the benefit of focusing on compositional aspects, which may be considered more essential in music than sound. Generally, the compositional aspects of music (melody, harmony and rhythm) have been captured in the generation procedure in a broad range of symbolic deep learning systems, with focus that varies between systems.

The main achievement of trending music generation research is the ability to create output for a variety of tasks even if the stylistic rules are too complicated for a design of manually coded rules, with the ease of a relatively generalisable creation of models if the initial training dataset is changed. Regarding its limitations, the biggest issue is a lack of detailed user control in the generation of output material. Since neural network models operate in a way of feature learning, they can be guided towards specific behaviour but cannot be controlled with perfect accuracy, which could be desired in cases where the users know beforehand what type of musical material they want. Direction attempts with neural networks can be made either through strategies that are based on the control of the input data in training and generation phase, or of the outcome via altering it after the fact. Alternatively, the generation can be controlled by actualising only outputs that conform to the users' requests. However, these strategies can lack in terms of the amount of data needed for training and input, as well as in the speed of the music generation process.

4.1.2 Music generation as compositional aid

For the purpose of aiding music software users, i.e. allow them to alter core parts of the generation process, the existing generation algorithms usually require expert knowledge of the systems in which they are implemented (Hunt, Nash & Mitchell, 2017). Programming environments like Impromptu (Sorensen & Gardner, 2010), Sonic Pi, Overtone (Aaron & Blackwell, 2013) and Supercollider (Wilson, Cottle & Collins, 2011) allow the use of generators and at the same time the change of parameters of the generation procedure, but they are directed towards music makers who are also programmers. Likewise, frameworks related to neural network algorithms, such as the Magenta project (Jaques et. al, 2017), WaveNet & MuseNet (Payne, 2019), require knowledge of mathematics and advanced computer science if one wishes to alter the generation code. Even then, these algorithms operate in a way of training through large datasets and the parameterisation is limited to changes in the way that the networks are trained. Moreover, the systems mentioned in this paragraph can be used only in their respective environments, and in most cases, they are not implemented in the form of plugins that could be inserted in Digital Audio Workstation tracks. This renders the tools not usable for the purposes of integration in the workflow of a typical music production setting (meaning, music making by producers who are not also programers, and who do not use coding in the creative process).

4.1.3 Music generation algorithms in music production

The idea of melody and music generators in music production has attracted software developers and music makers since the early days of computers, with the first programmed melody inside a computer being played in 1950 on the CSIR Mark 1 (Doornbusch, 2004), but it is only in recent years that several significant developments have been made in the commercial music software space. These products are integrated into the DAW-based music production workflow, and don't require computer science knowledge from the users. This section presents some of the most notable contributions in music generation VST plugins. The amount of projects is relatively limited, and the inclusion criteria for this list of examples are popularity and quality. We address the products Generate (Ableton plugin rendition of Google's Magenta), Melody Sauce (2020), Captain Melody (2020) and Riff Generation (2020). The projects are initially described in terms of the generation processes they use, and their approach to user interface design. Their functionality is then analysed in order to present possible opportunities for improvement, and our attempt at addressing these improvements with the MMM Generator is described.

Generation mechanisms

Starting with Generate, this plugin is Google's approach to put their music generating machine learning algorithms in the form of a DAW-based tool. In the first few years their results with using recurrent neural networks for melody generation sounded musical, but by the nature of their RNN it was not possible to produce long-duration output, as the independence from specific rules in the generation process made the output progressively more random sounding. However, their latest progress with using a novel, hierarchical structure in the model (Roberts et. al, 2019) produces better results for long duration musical output. Another product with a similar approach to generation, in the sense that it uses a sophisticated algorithm, is Melody Sauce. A product of Evabeats, it is a melody generator that, according to the developer, aims to provide a starting point for a melody that the user will complete via the software's simple controls. Its AI uses an algorithm that is informed by mathematical analyses of common features in modern pop and EDM hit songs.

Leaning to a user control - centred approach, the company Mixed In Key developed Captain Melody, a plugin that was released in 2018 and that appears to be one of the most popular solutions in the new melody generation market. Since their code is not publicly available it is not easy to understand how the note generating part of the plugin works precisely. Behind this feature there could be a random notes generator, or an AI assisted process. Nevertheless, the focus of Captain Melody is on the user's options to control and morph the melodic ideas. These controls adjust the shape of the melody, either leaving it as is, or giving it a linear shape. Features include the choice between different types of linear shapes (ascending / descending / ascending + descending / descending + ascending) and a curve-like contour, a variety of strength regarding note leaps, and the fitting of the melody in relation to chord changes. All these parameters are configurable in detail by the user, along with further parameterisation that includes the frequency of stepwise and leap motions, amount of produced notes, and a control of the frequency of triplets. Finally, the user can alter the probabilities of occurrence for each of the notes in the scale. Similarly to Captain Melody, In Session Audio developed the Riff Generation, a plugin that uses a random notes generator and that is parameterisable by the user through a big assortment of options. These controls are about the boundaries of the generated melody, and the user's preferences on scale, rhythms, sound selection, sound effects, and various other in-detail customisations of the output.

User interface

As seen in the previous section, some of the generators are centred around the idea of giving users a relatively simple interface and a melody that requires relatively few manual changes, while others support the approach of letting users shape the melody significantly after the generation process, with a multitude of controls. Belonging in the first category, Google has incorporporated an approach in Generate that is of a minimalistic design with very few options. As noted, with these algorithms there is difficulty in dictating the musical direction of the output material with great precision because of the “black box” nature of neural networks. Concerning the shaping of the produced output, Generate provides the user with one option, called temperature. This is a slider that essentially randomises the produced notes by a certain amount (DuBreuil, 2020). The user can also select how many variations the plugin will produce, with a second slider. In a similar approach, Melody Sauce focuses on a simple, single-page user interface, even though the number of controls is significantly higher than Generate. Central to the interface there is a section of 3x3 square “drum pad” buttons, that when any of them is pressed, a melody is generated. Each pad works with different algorithmic settings, but it is not clear from a UI perspective what the changes are. Left from the pads there is a section with scale and rhythm settings, while on the right side the user is presented with a view of the previously generated melodies along with editing options.

While that software aims to provide a simple interface design and melodies that require relatively few manual alterations, other manufacturers support the idea of allowing users to influence the melody extensively after it has been generated, by using a wider variety of controls. This is the viewpoint behind products such as Captain Melody. The user interface for that plugin is decidedly more complex than the previous examples, featuring 4 tabs meant for various song sections, 4 further tabs in each section that are about the shaping of the notes, and 2 tabs about technical settings. Below the sections of tabs, it includes a piano roll editor where the MIDI melody is shown and manual edits can be performed. At the left side of the screen the user has options that relate to various tools that can shape the melody, and sound options. Further, Captain Melody can be interconnected with other generator plugins of the Captain series, allowing a fuller range of possibilities. In like manner, Riff Generation makes considerable use of submenus and tabs as part of its layout design, as it aims to supply music producers with an ample amount of choices regarding the notes and sound of the generated output. Its UI features two main areas of focus, which resemble the layout of two step sequencers. These areas change their content depending on the tab that is activated by the user. For the upper side area there is an assortment of 4 tabs about velocity, note duration, volume and panning for each step of the sequencer. The lower section of the UI shows a grid of 32 x 5, that alters 5 types of sound settings for each of the 32 time points of the sequencer. The same area becomes a menu of 5 sections of generation settings when the user presses a “generation options” button.

Analysis

The examination of the generation processes and user interfaces of that software can reveal possible points for improvement that the current project’s MMM Generator plugin could address. The generation mechanism of Google’s plugins is considered to be useful for starting a melodic idea that the music producer will use as a basis when building the actual melody of their track, and in general, as a tool to spark initial inspiration (SadowickProduction, 2019; Kozmik Kandi, 2021). Melody Sauce, developed with a direction reminiscent of Google’s plugins, received a mixed response from the users and it seems to be considered as a tool that generates ideas of limited variation and quality, that serve as an initial material to be further developed by the user (Cowby, 2018; Weaver Beats, 2019). The same sentiment is shared about Captain Melody, in which some parts of the generated melodies are useful for further tweaking (Qpbeats, 2019). On the other hand, Riff Generation is thought by some users to be a powerful arpeggiator or sequencer, but to not classify as a melody generation tool (Tatanka, 2018). In summary, the generation algorithms of current plugins seem appealing to users who like to experiment with alternative ways of music creation, but they are not considered to be sufficient in providing fully ready melodies and replace the composition process of users.

In terms of user interfaces, Google's Generate features a minimal design with few options, which is a good design practice that helps with its usability (Nielsen, 1994). However, the lack of more controls means that the user cannot influence how the music generation operates in any way other than the amount of produced melodies and the amount of note randomisation that occurs after the fact. Also, due to the fact that the plugin is coded as a Max/MSP package for Ableton Live, the workflow differs from how typical VST plugins integrate with the DAW. This makes Generate less easy to use, since it outputs the MIDI to clip slots (Hughes et. al, 2022) instead of the timeline of tracks, and it does not support operation within the typical arrangement view of the project in Live. Likewise, Melody Sauce is designed with simplicity as a priority, without multiple screens and with a surface level control of the generation, i.e mainly basic rhythm and scale settings. By having the generation buttons at the centre, the plugin achieves a good visual hierarchy that helps the users understand the relative importance of the interface's elements (Gordon, 2020). The straightforward design of the aforementioned products is in contrast with the perspective used for Captain Melody. With a reliance on heavy editing of the melody by the user, the centrepiece of the interface is a piano roll. The various controls and tabs are around the piano roll, without visual clues about what are the most important elements or how the melody generation is initiated. This leaves the user with no obvious entry point to act on, which is a bad practice as it provides no instant gratification experience for using the software (Tidwell, Brewer & Valencia-Brooks, 2020). Finally, like Captain Melody, the Riff Generation plugin imposes a large cognitive cost on the users due to its complicated interface, and it does not present an hierarchical ordering of the elements. These types of user interface designs may lead the users to refrain from using the particular plugins, in favour of simpler ones, even if they are not equally good in terms of the quality of generated melodies. The potential cost of time and effort when learning a complicated system can lead music producers to accept this difference in quality if the simpler plugins are "good enough" for their required task, which is a behaviour described as “satisficing” (Herbert, 1956). Therefore, the examples we examined show us that in order to provide a good user experience, we need to address the issue of balancing clear action points and simplicity with a versatile, musically meaningful control of the generation process.

4.1.4 Interim conclusion

In relation to the aforementioned background literature and software developments, this part of the project utilises the approach of statistical and code based modelling of cognitive elements, by using the results of the research that was carried out in the first and second studies of the thesis. These elements were used in order to create a MIDI based music generation application, and by strictly using models based on music cognition it is an alternative method to the one used in most related research, i.e. the training of the algorithm from a human made music corpus. The use of these models taps into the opportunity that the rest of the presented VST plugin projects leave, by not utilising music cognitive processes in the generation algorithms. Furthermore, this different approach is an attempt to address the creation of a plugin that fulfils the role of a tool that outputs ready to use melodies, instead of producing output that mostly serves the role of an initial spark of inspiration for a human composed final melody, which is what currently published generators are mainly used for. Concerning the user interface, the background research revealed that current plugins present a gap between using highly sophisticated generators over which the user has no musical control, and seemingly random notes generation combined with complicated interfaces and a big variety of options, often not related to music generation.

The main point in the creation of this software as part of the thesis relies on the aim of offering a solution to the task of starting a new music track according to broad artistic directions by the user, e.g. the choice of a melodic archetype and setting of atmosphere through choices of rhythm and chord progression types. It was implemented by using a generator that allows the user to need relatively little control and a simple, intuitive interface.

Building on the other VST plugins in the marketplace, we wanted to achieve simplicity that is reminiscent of Google's Magenta, while also allowing a fuller control of the possible outcome as observed in projects like Mixed In Key's Captain Melody. We attempted to accomplish this by using results from the previous parts of this project, i.e. by embedding information about the aesthetic impact of specific quantitative elements found in melodies. The interface was also informed by researching user needs in the form of interviewing music producers, and by implementing UI design literature concepts. Finally, a differentiating factor from a part of the currently available software is that it has a direct aim to be used as a writing aid tool for producers and does not try to blend tasks not related to efficient MIDI music generation, such as sound designing and mixing, or the use of plugin combinations and low level control of probabilities.

4.2 Creation of the generator

During the development of the MMM Generator we created a system of different sections that worked together to create the musical output. Each section serves a different purpose in the chain of the generation process and was developed according to a corresponding set of rules, drawing from music cognition literature, the studies of the current project, and analysis of real-life professional music production techniques. A view of the initial user interface is presented in Figure 39.

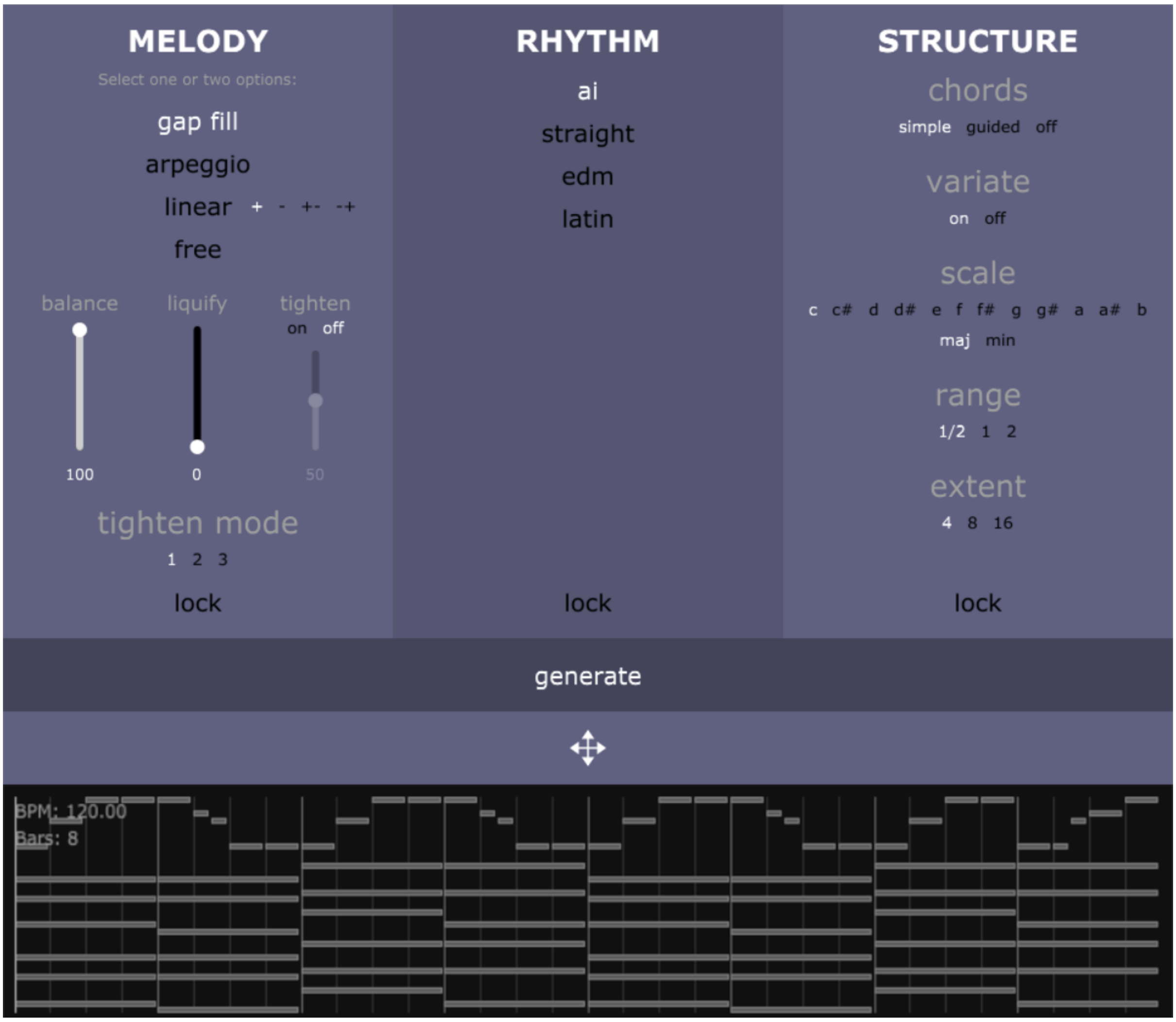


Figure 39: The first rendition of the user interface of the Generator. It consists of three sections for the controls of the melody, rhythm, and structure of the output. It also features a drag-and-drop function that puts the output into a DAW track as a MIDI item, and a visual representation of the output in the form of a piano roll.

The first of these sections consists of a note generator that creates an initial MIDI melody according to prespecified broad directions by the user. The next one is an interconnected section that is responsible for the rhythmic elements of the musical output. Finally, the last section is related to the structure of the output, and handles the chords and parameters related to the extension of the duration of the music segment. The sections do not work in a serial “forward only” way, but are connected and work in parallel throughout the generation process, transferring information between them. In 4.2.1-3 we explain at a high level how these parts work, while we connect their development with the background research behind them.

4.2.1 Initial notes generation

[Redacted text]

[Redacted text]

[Redacted text]

4.2.2 Rhythm generation

[Redacted text]

[Redacted text]

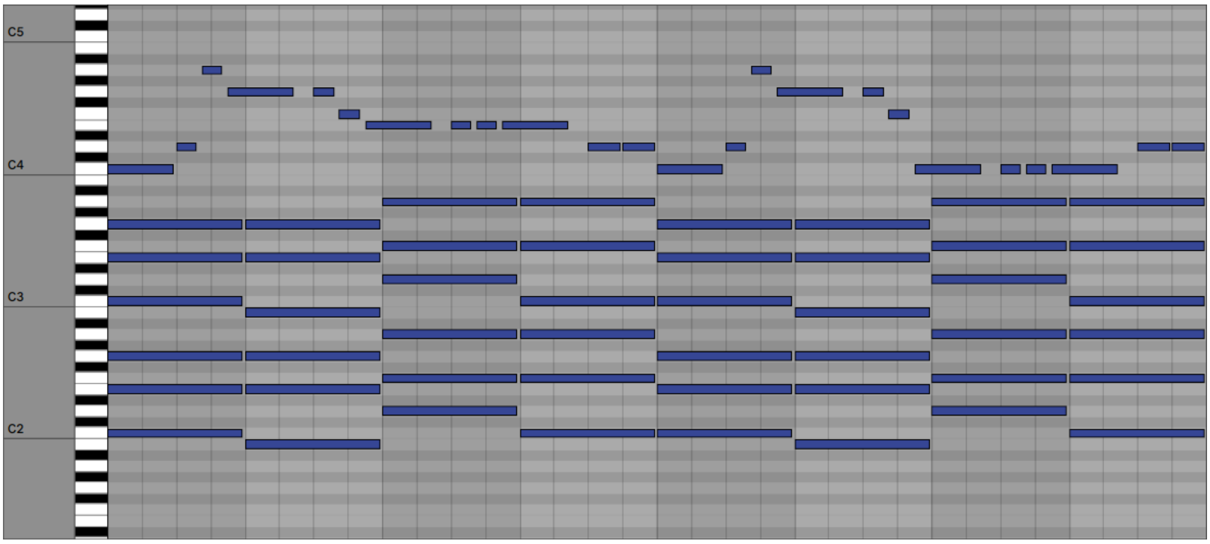
[Redacted text]

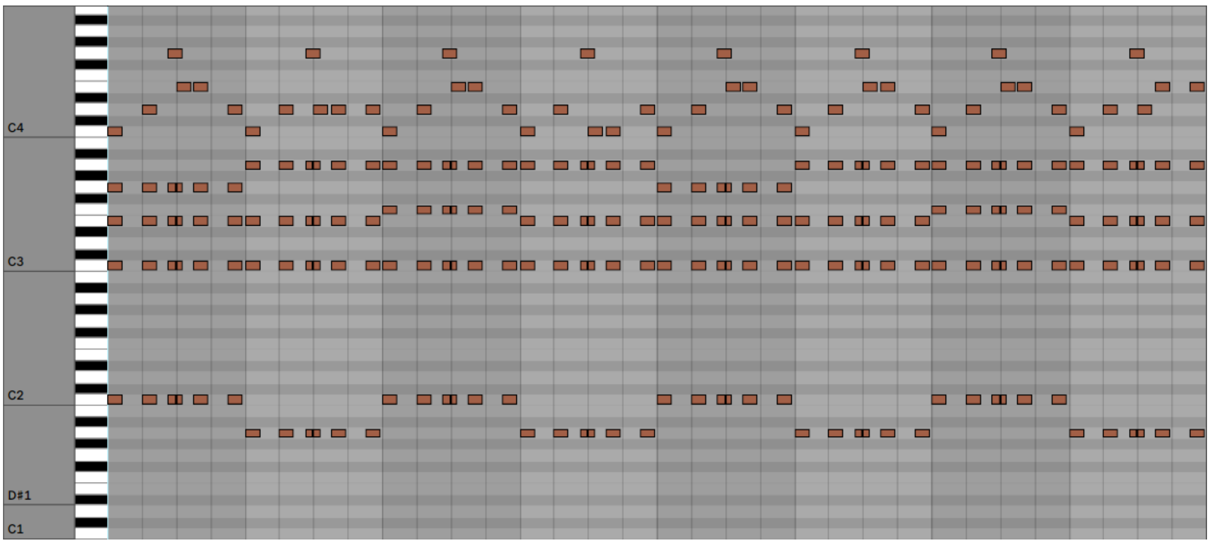
Box 4: [Redacted text]

[Redacted text]

4.2.3 Generation and choices of structural elements

An important aspect of the MIDI generation that surrounds melody is structure, in terms of harmony and of the development of a long duration output. We created features of harmony by employing chord progressions. Specifically, the user has the option to add progressions that use chords built directly below the melody on the strong beats (also called “block chords”), or progressions that use notes that follow the exact rhythm pattern of the melody. Figures 42-43 show MIDI piano roll examples of block chords and rhythmical chords respectively, below a melody. The examples show music that has been created by the MMM Generator.





Figures 42-43: A musical output of the MMM Generator that uses block chords (42), which play for the duration of a whole bar. In (43) there is another example that incorporates rhythmical chords, playing with the same rhythm as the melody. The chord changes happen at 1-bar intervals, like in the case of block chords.

[Redacted text]

The rest of the development concerned functions about secondary structural elements of the musical output. The function of “Range” allows users to control the size of the possible pitch range of the main melody, by setting the variable HighestNote as shown in Box 1 at either 6, 12 or 24 semitones above LowestNote. The “Extent” function alters the amount of notes that the generator uses in order to create the melody. It is a feature that controls how repetitive or non repetitive the output is going to be, and it can be set to three levels of repetitiveness. In particular, at step 7 of Box 1, the algorithm stops adding notes to the melody if the mean pitch is at the target value, and the number of notes is greater than 3 or equal to 8. With the “Extend” function, the user can explicitly stop this initial pitch generation process at specifically 4, 8 or 16 notes. Further, the MMM Generator also features “Lock” buttons in each of the three sections of the user interface, as shown in Figure 39, where the discussed functions can be identified. The “Locks” allow the user to keep either the contour and initial notes, or the rhythm, or the chords and structure of the musical output that was generated, to be used in further generations. By locking a section, any new generated music will vary all the other aspects of the generation, but it will "remember" from the previous generation and not change what aspects exist in the section that is locked by the user. This allows the users to build a musical output that they like in a step by step fashion, and it is performed by storing the generation settings in temporary log files. Finally, the "Variate" function works with an on / off switch that controls the presence or absence of variations in the generated melody, as the structure develops. Those variations are created by triggering via random choice one element in a list of deterministic manipulation series, to allow a local re-shaping of parts within the melody, by using the controls of the "initial notes generation" section (whose functionality is detailed in 4.2.1). Finally, users can choose a major or minor scale in the key of their liking.

4.2.4 Bridging the MMM Generator to the Digital Audio Workstation environment

During the previous sections we discussed the creation and features of the MMM Generator, as a software that creates MIDI data according to user instructions. The context in which this software operates is within digital audio workstations, in the form of an external sub-software that the DAWs "see" and call it to operate inside their environment. These sub-software are what is known as virtual studio technology plugins. For the operation of the MMM Generator as a plugin, we used a hybrid system that incorporated a web server, where all generation and functional algorithms are stored, and a file that is stored in the user's computer and acts as the VST plugin. The plugin, when loaded into the DAW, shows the interface depicted in Figure 39 on which the user chooses their required settings. Once the user presses "Generate", the plugin sends a request to the server, and once the server processes the input settings and creates the musical output, it sends it back to the plugin. This hybrid approach was required, because the generation algorithms are coded in Python, while VST plugins use the C++ programming language. Python was necessary due to a dependence on [Redacted text]. Therefore, a server that supported the execution of Python code was required in the process.

4.2.5 Summary of the interface design

In terms of design, the sections of the interface were structured in a way that directly addressed the development as described in 4.2.1-4. The “Melody” section begins with multiple choice options that allow the users to choose between melodic shape archetypes. The Balance slider inputs the value of the “target mean” variable that was described in 4.2.1. Next to it, Liquify is a slider that controls the effect of a randomiser algorithm that changes the melodic shape structure of the output. Finally, Tighten controls the complexity of the melody when it is activated, and it can further apply this change on inverse or duplicate continuations of the melodies, depending on Tighten Mode.

The next section on the interface of the plugin is “Rhythm”, which contains a multiple choice section relating to different rhythm types. Each choice corresponds to a different pool of rhythm guides that suits the stated style of music. Finally, “Structure” is a section that is responsible for harmonic and structural controls of the output. In “chords” the user can choose between block, rhythmic and no chords. Variate, Range and Extent options control the corresponding variables as discussed in 4.2.3, and “scale” allows users to choose a major or minor scale in the key of their preference.

After the sections that contain controls about the parameterisation of the generation process, the “generate” button can be used to initialise the MIDI creation. The MMM Generator plugin achieves this by sending the user’s chosen values to the server. It then receives MIDI data that can be played back when the user initiates playback in the DAW. The arrows symbol below “generate” is a click-and-drag function that can place the MIDI data as a MIDI item in the DAW tracks. Lastly, the generated MIDI music can be seen in visual form at the bottom section of the plugin, in a piano roll of relative distances between the notes.

The plugin file and the interface code were written in C++ by [Redacted text], founder of the music software company [Redacted text]. His contribution included the idea and implementation of the visual representation of the generated MIDI. With his assistance, we were able to connect the MMM Generator algorithms that are stored in the server to the plugin interface, enabling us to use the MMM Generator in digital audio workstations. This allowed us to perform the study concerned with the evaluation of the MMM Generator.

4.3 Evaluating the MMM Generator: Aims

Having built a prototype version of the MMM Generator plugin, we distributed it to music producers, acquired user feedback about it, and proceeded with evaluating it. The aims regarding plugin evaluation were to discover the degree to which we succeeded in creating a useful and high quality music production tool, by using a music cognition literature - informed approach. The analysis of our study focuses on user sentiment in terms of the software's usability, and on user's positive and negative impressions. Further, thoughts for improvement were collected, in order to be used for a possible updated version of the MMM Generator that would be adapted to users' needs.

The plugin evaluation took place with the combination of an online questionnaire and text-interviews with music producers. In order to achieve an assessment that would answer our aims, questions on a series of user experience and usability features were employed. Such features are related to the ease of use, the workflow, the aid on inspiration, and fitting the users' purposes. Overall, we wanted to understand to what extent it is possible to approach music production plugin development through quantifying aesthetic properties and tapping into the potential that is found in academic music cognition literature. More specifically, we drew insights from the statistical properties of melodies, rhythms and harmonies, in combination with evaluating listener's reactions and expectations. By doing so, we attempted to aid in the improvement of current, and the creation of new and unique music production plugins. This study of plugin evaluation addresses our larger aim by focusing on unsupervised remote testing of usability (Barnum & Dragga, 2001).

4.4 Method

4.4.1 Design

A mixture of quantitative and qualitative questions were used for the purpose of deriving concrete quantifiable evaluations of the plugin, as well as qualitative insights into the experience of it (Creswell & Plano Clark, 2011). The quantitative questions used rating scales consisting of 10 points, evaluating statements related to the user's experience of the MMM Generator. By comparing to the mean point, we can evaluate whether satisfaction was reliably positive with respect to a number of experience factors.

The qualitative questions were open-ended prompts that were non-directive, apart from asking to evaluate positive and negative aspects of the MMM Generator, and aspects for future development. The qualitative data analysis consisted of defining trends within key themes that were extracted from the raw responses of participants. The focal themes included pre-determined and emerging ones, to permit a balance between gaining insight that is related to the research questions, and having space for unexpected responses, or new ideas. The questionnaire is shown in Appendix.

4.4.2 Participants

Music producers over the age of 18 were contacted via the social media website Instagram, using a combination of drawing from personal contacts and a virtual snowball sampling technique (Coleman 1958; Baltar & Brunet, 2012). This led to a total sample of twenty male music producer participants. To confirm participation, they were asked to download the MMM Generator software, read the manual that was included in the downloaded package, and spend at least thirty minutes using it in their digital audio workstation. After this session, 16 randomly selected participants filled out a questionnaire, while the remaining 4 took part in a free text interview via direct messaging on the Instagram platform. The study was approved by the ethics committee of the Department of Music.

4.4.3 Materials

The online questionnaire used for the study was based on the post-study system usability questionnaire / PSSUQ (Lewis, 1992), and was modified to suit the assessment of usability in the context of a VST plugin. Further, free text questions were included, with the purpose of evaluating any specific positive and negative impressions, as well as to receive ideas for implementation in future development. These questions were adapted to be used as the basis for the text-based interviews that were conducted with the second group of participants, which were carried out in order to enrich the dataset. Both the questionnaire and the free text based assessments took place after the participants used the VST plugin version of the MMM Generator in their DAW of choice.

4.5 Results

4.5.1 Quantitative analysis

The distributions of the quantitative questionnaire data were tested against a theoretical median equal to 5.5, in order to examine the degree to which the participants had a positive or negative experience of the MMM Generator. These concerned ratings of usefulness, interestingness, ease of use, uniqueness, cleanliness, speed, how creative it feels, how inspiring it is, to what degree it has the features the user needs, and to what degree it fits the workflow and the purposes of the user.

The data acquired from these scales called for a non parametric approach to the hypothesis testing (distributions of the answers being tested against the theoretical median value), since the responses were not distributed according to the normal distribution, as it was found with the use of a Shapiro-Wilk test for normality. Instead, the Wilcoxon Signed Rank test was used, for the purpose of comparing the median of the distributions of our variables with 5.5. This value was chosen as the middle value between 1 ("Very Unuseful") and 10 ("Very Useful"), that correspond to the two extremes of the Likert scales. Table 20 shows the multiple Wilcoxon Signed Rank tests that were performed.

Table 20: Hypothesis testing for the usability questionnaire scale items.

One-Sample Wilcoxon Signed Rank Tests						
Variable	**Median**	**Min**	**Max**	**Comparison Value**	**Wilcoxon Sig.**	**Adjusted Sig.**
Useful	8	3	10	5.5	.002	.007
Interesting	9	5	10	5.5	.001	.006
Easy To Use	8	2	10	5.5	.011	.016
Unique	8	2	10	5.5	.008	.015
Creative	8	3	10	5.5	.007	.015
Inspiring	9	4	10	5.5	.001	.006
Clean	8	2	10	5.5	.007	.015
Quick	8	1	10	5.5	.012	.017
Features	6	1	10	5.5	.323	.323

Fits Workflow	8	2	10	5.5	.027	.033
Fits My Purpose	6	2	10	5.5	.130	.143

Through these tests, all variables except “Fits My Purpose” and “Has the Features I Need” were found to have a median value that was statistically significantly above the test value. These findings held the same outcome after performing a Benjamini-Hotchberg test for multiple testing, at a significance threshold of $q < .05$. In Table 21 (see Appendix), we show the distribution of the ratings for the scale variables.

It is interesting to note that the variables had a median score of 8 or 9, showing that the MMM Generator was perceived as a useful and high quality tool. However, ratings of “Fits My Purpose” and “Has the Features I Need” did not give significance, which means that not all producers needed the software. Further, by not fitting one’s purpose, it is also logical to have a similar score in the “Features I Need” scale. A correlation analysis using Kendall’s Tau-b was conducted in order to test this assumption, which indeed indicated a highly significant, strong correlation between the two variables (Kendall’s Tau = .792, $p < .001$). This points to the fact that even though the plugin was evaluated positively, the participants did not specifically request it, or all of them necessarily need it. This is different from assuming that the producers, in general, did not need it, as the contrary is highlighted through the significantly high ratings in the various aspects of usability and output quality. As the text data shows in the upcoming qualitative analysis section, a percentage of musicians generally prefered to not use a "helper software" for music composition. Therefore, the existence of these two sides did not allow significance. As the correlation reveals, if a music producer felt that this software is not for them, they (logically) also said that it does not have the features they need.

4.5.2 Qualitative analysis

These qualitative responses were helpful to cross-validate the findings of the scale items analysis, and allowed for a more detailed view behind the reasoning with which the participants completed the questionnaire scales. The data under examination was concerned with the general impressions of the users. The interviews aimed to gather information regarding the general thoughts gained from the MMM Generator usage, as well as the positives, negatives and possible improvement ideas about the MMM Generator. Similarly, the free text data deriving from the usability questionnaire included responses from four text-based questions: Their general first impression, what feature they found to be the most useful, what was the least useful feature, and what would they do to improve about the Generator given the chance.

The free text data that was gathered from the participants was examined in order to categorise the feedback into key themes, based on the content of the responses. We used the themes of Output Quality and User Experience that fit the objectives of our study, along with ones that emerged during the categorisation of the responses. These themes are described in Table 22, alongside the number of occurrences in the dataset.

Table 22: Thematic grouping and interpretation of the free text data, along with the frequency of occurrence for each key theme.

Category of feedback	Types of responses	Example response	Number of references
Good output quality	Good melodies and chords. Useful. Helps with inspiration.	“I do use plugins like Scaler, Riffer, Captain Chords when i get writers block but these at times just generate random notes which sound like a mess. Your plugin actually sounds good and the	14

		melody works with the chords which is awesome.”	
Good user experience	Unique. User friendly. Nice. Fast.	“Useful and fun to work with. Relatively intuitive, with fast results.”	10
Helps with music knowledge	Helps with music knowledge.	“I personally think that it would be quite useful for those who have difficulties to make their own melodies and chord progressions.”	5
Good market potential	Good market potential.	“It makes music making possible for almost anyone. It can prove to be a revolutionary plugin.”	5
Ideas / requests	Humanisation. More styles.	“I would have liked more options for rhythms and structure such as time signature, syncopation, polyrhythms etc."	13
Negative feedback (all)	Not intuitive. I would not use it.	“I wouldn’t use it for my own songs though. I feel like the progressions and the melodies are a bit generic.”	8

Overall, we got 34 positive impressions in the topics of music quality, helpfulness, user experience and market potential. Also, there were 8 negative feedback points, of which 4 were about the labelling of the parameters and user interface, 1 about a lack of features, and 3 about the actual output. In relation to the research questions, the impressions of the participants suggest a positive sentiment towards the usability of the plugin. This can be observed from the frequent mention of the plugin's high market potential, as well as the responses regarding the quality of the musical output. These results are in line with the quantitative analysis, where the medians of the distributions of the scale variables of usability were higher than the theoretical median.

4.6 Discussion

With this study we sought to find to what degree we succeeded in creating a usable music generator by using an academic approach. Namely, by using functions that derived from music cognition studies and literature, and by analysing the features of related music and software. Also, we wanted to understand what is the overall quality of the generator in terms of user experience.

The data that we acquired was split into two approaches, that is quantitative and qualitative data collection. The results of the quantitative approach showed that the ratings for most usability variables were statistically significantly higher than the middle value. In particular, this outcome suggests that the generator achieves positive standards in the areas of usefulness, interestingness, ease of use, uniqueness, cleanliness, speed, creativity and inspiration. Similarly, the qualitative answers showed a favourable stance on the quality of the musical output, usability and market potential. Therefore, we can conclude that the MMM Generator satisfies the goal of being a usable plugin with a good quality standard.

The qualitative remarks of the users showed us the reasoning behind the ratings that were gathered as quantitative data. Uniqueness, as a highly rated aspect of the software, seemed to have been achieved by the fact that this kind of software is new in the plugin industry, and further, the currently available products have not managed to appeal to a considerably large number of users. This led to feedback related to the fact that some producers had "never seen programs like this". Another factor that we infer it added to the ratings of uniqueness is that the implementation was not based on AI trained models, which would be a “black box” approach in the creation of the output, and thus it presented controls and functionality that have not been implemented in other music generation plugins. Being able to provide detailed music-focused controls is an added benefit of using music cognition concepts such as archetypal contours, instead of relying on training a model to learn melodic contours from data.

This design, by achieving a standard of music output that was received positively by music producers, led to high usefulness ratings as well. One of them commented: “I have noticed that the generated melodies and the matching chords are very appealing to myself and about ~90% of the time I go inspired instantly and had the urge to create a track out of the generated music coming from the plug-in”. This sentiment, in various degrees, was repeated through the qualitative feedback and through the ratings of the Likert scales. The usability of the MMM Generator is what allows it to have a good market potential, because in the modern music business where producers undertake the role of the composer, they would often need an aid for this role, if they find a good one. According to a participant in our study who has produced a big number of hit pop songs in Greek music for the duration of the past decade: “In modern music production, I think it is the future. Because the new guys are also a bit lacking [music theory skills]. And it is what they are searching for”.

As discussed in 4.1.3, the market of music generation plugins is at a place where the users still hold a generally negative view when it comes to their quality of output and usability, when it comes to scenarios of wanting to use the output in a project as it is, keeping it relatively unedited. The most positive users hold the point of view that music generation plugins should be better seen as a secondary tool that can perhaps initiate the creation of real musical material later on by the songwriter. However, our work can be interpreted as a contribution to music generation software that can provide insights about how to approach the creation of such generators. Specifically, we showed that it is possible to utilise the existing potential found in music psychology literature, as well as the creation of new studies that quantify aesthetic responses to music. This utilisation can lead to the development of software that has the potential to surpass the weak points of non-cognition theory - informed works, as it was shown by the high ratings and feedback on usefulness and quality in the study. However, academic literature was not the only source of ideas for the development of this software. A use of modern music production techniques and songwriting theory was also implemented. Also, the current study was not designed to fully differentiate between academic literature and music making techniques / theory regarding the degree to which the two different domains contributed to the ratings and qualitative feedback.

Further, our study showed that the test value was not significant in the scales “Fits My Purpose” and “Has the Features I Need”. Similarly, the critique coming from qualitative feedback considered primarily how the parameters were named, and the user interface. Qualitative feedback also showed us that a percentage of the producers just felt that they do not want to use a compositional aid of this kind, because it does not fit with how they like to create music, which is a possible interpretation of the non significance of “Fits My Purpose” and “Has the Features I Need”. We can conclude that the MMM Generator is not for every musician who works in a digital audio workstation environment. Rather, it is a better fit for users who make music that uses electronic elements, such as Electronic Dance Music, Trap, Pop, and so on, as we recognised from the responses of the producers who were known to have worked on these musical genres. Also, it appeared useful to people who want to incorporate less human-like melodies into their works, either by considering the output as generic, or as artificial, like in the case of one participant: “I love this thing already. I've been looking for an affordable melody generation tool - primarily to help in creating electronic tracks. Sometimes you want to create something that sounds really artificial - and this can help with that. Though, that isn't to discount the variety this tool possesses. I prefer to use it as a melody generation tool - but the harmony generation was good too”. A possible explanation about this view of some responders is the absence of timing and velocity humanisation controls, and that it does not incorporate instrument-specific motifs. Ultimately, given that it is not possible to create a tool that will appeal to every possible workflow and music production style, we could only act on the user interface and parameter-labelling critique, that was general and not dependent on specific workflow or musical genres.

Another limitation of our work relates to the fact that we used online questionnaires, instead of an offline examination of how the participants used and evaluated the plugin. While allowing the participants to work in their own environment certainly has benefits related to the natural implementation of the MMM Generator to their personal DAW and system choices, it could be interesting to have some cases of in-person, richer evaluations of their work with it. This idea had become impossible to implement, since the study ran through late 2020 and early 2021, in a period when full lockdown measures had taken effect due to the Covid-19 pandemic.

As next steps to progress towards fulfilling the needs of users, we rethought the user interface with a focus on simplicity and easy to use functionality. Specifically, the UI will now consist only of the Scale option, and buttons that trigger pre-configured combinations of settings, also known as presets. The buttons will be categorised in the sections Full Part (includes extended Melody and Chords), Lead (main melody only), Bass (presets that create bass parts), Keys (oriented towards chords), and FX (presets that output notes that are meant to be used as starting points for the sound design of effects). Further, the UI will include an “advanced settings” button that leads to a new screen. This option would allow users to access the settings that were in the interface of the version of the study, giving them the opportunity to create musical output through all possible combinations of settings, instead of relying on the pre-configured combinations of the initial screen. With this approach, we greatly simplify the UI and the process of generating musical output, while keeping the quality of it at an optimal level through the use of preset configurations. At the same time, we allow the more advanced or curious producers to fine-tune their output and delve deeper into the possibilities of the MMM Generator.

5. Discussion and conclusion

5.1 Introduction

5.1.1 Main research questions

This thesis was concerned with cognitive concepts related to melodic preferences, and attempted to capture quantifiable elements of what makes melodies sound pleasing to listeners, through the use of statistical concepts and analysis of listener preferences. By using both new insights and drawing from literature, these elements were subsequently used as a basis for the development of software that generates MIDI musical segments. The thesis included two studies concerning the perception of melodies, and a study about the user satisfaction in regards to the software. In particular, the main research questions were as follows:

1) What is the role of statistical distributions of pitches in melodies when it comes to liking, and how is perceived complexity and interestingness associated with liking?
2) What is the relationship between working memory load and liking ratings and how does the fulfilment of melodic expectations affect the ratings?
3) To what degree did we succeed in creating a high-quality product that addresses the needs of music producers? This question is asked in the context of using a methodology of quantifying aesthetic and cognitive properties of melodies, found in the studies of this project and past music cognition literature, and using them in order to produce a music generator software.

In this final chapter, we summarise how these questions have been answered through the research we conducted in the project. Also, we present the methodology that was used and the limitations of our work, along with ideas regarding future steps.

5.1.2 Rationale for doing this research

This research is based on the overarching aim of creating the MMM Generator, a music production software that performs algorithmic music generation and that was developed using insights from academic research in the area of music cognition. In order to achieve this, we developed research questions as presented in 5.1.1, which have individual aims focusing on the theme of each question. Specifically, Question 1 explored the perception of pitch distributions and aimed to contribute to the topic of how they affect aesthetic and cognitive responses to music. This question was answered through Study 1, and resulted in main insights regarding what distributions participants preferred, which were subsequently coded and used in the creation of music generating algorithms.

Likewise, Question 2 was concerned with the effect of working memory limitations and fulfilment of expectations on melody liking. The aim of this question was to provide new insights in that topic with the formation of Study 2, which examined the connection between liking and the other variables by using pitch manipulation of the stimuli, in a similar way to Study 1. This manipulation allowed a fine-tuned control of pitch-related cognitive load and it constituted an approach that offered a new way of examining this topic. By doing so, it also helped in the development of a feature that controls in detail the aspect of complexity in the output of the MMM Generator, contributing to the overall goal of the project.

Finally, the third main research question examined the idea of combining the creation of a commercial VST plugin with using a music - cognitive academic approach as the basis for it. With the implementation of this idea as described in Chapter 4, we wanted to address a considerable gap that exists in the utilisation of this non computer science based field for plugin creation. By doing so, our aim was to address the weaknesses of current state of the art neural network approaches, and offer interesting new opportunities and insights in music generation. Further, we wanted to evaluate the benefits of using this approach of development, by analysing user feedback about the plugin.

5.2 Novel contributions

5.2.1 Addressing the research questions

The contribution of this thesis in academic literature is found in the context of music cognition research. The aim was to find and express factors that influence the melody liking preferences of listeners in quantitative ways. As such, it presented studies about how these possible factors could work. Also, it contributed to the DAW plugin industry with the development of the MMM Generator. A summary of the research questions along with the related contributions of this project is presented in the following subsections.

Question 1: What is the role of statistical distributions of pitches in melodies when it comes to liking, and how is perceived complexity and interestingness associated with liking?

The role of statistical distributions of pitches in liking was explored via the comparison of multiple pitch distributions, that were related to what was termed in Study 1 as the Uniformity Principle: A hypothesis stating that a completely even distribution of pitches within the range of a melody results in liking ratings that are higher than those of less even pitch distributions. The results of the particular analysis did show strong statistical significance in the liking differences between distributions that were created with the uniform generation rules, as described in Chapter 2, and those that were created by progressively more skewed ones. In particular, melodies generated from the uniform algorithm had more positive ratings than the rest, supporting our hypothesis. Furthermore, the significance of these differences held for the group of musician listeners, as defined by attendance of formal music education or self-taught music skills, and not for non-musicians.

In Study 1, the secondary idea of exploring the performance of variables that are related to liking, as perceptual properties, was also assessed. Specifically, we tested for a possible association of complexity and interestingness ratings with the ones of liking. This was performed both for the whole sample, and for demographic subsamples. What was shown in the whole sample analysis is a strong association between interestingness and liking, and moderate associations in the complexity-interestingness and complexity-liking pairs of comparisons. Regarding the demographic subsamples analyses, no statistically significant associations were found within the binary split groups of age, gender or musical experience, in any pairwise comparison of liking, interestingness and complexity. Overall, the exploration of these variables resulted in the implication that interestingness is an attribute of melodies that is associated with liking.

We consider the results regarding the Uniformity Principle to imply that a sense of balance, when it comes to the distribution of pitches within the range of a melody, has a positive impact in liking. This contribution offers a quantifiable and directly implementable way to improve the perceived quality of melodic output in music generation algorithms in general. Also, it offers a rule that is reminiscent of Gestalt principles, such as symmetry and good figure, adding to the understanding of how universal psychological principles may apply to the aesthetic appreciation of music.

Regarding its place in literature, this finding is an addition to the examination of the sensitivity of music listeners in statistical distributions, where it has been shown that distributions influence listener's perception of musical features, such as the key of a composition (Vos & Van Geenen, 1996; Yoshino & Abe, 2004), and aesthetic preferences (Voss & Clarke, 1978; Manaris et. al, 2003). Further, it indicates that as in other areas of music cognition (Kishon-Rabin et. al, 2001; Liu et. al, 2018; Medina & Barraza, 2019), there are differences in how musicians and non-musicians perceive musical properties, since the effect of Uniformity Principle was only found in the group of musician participants. Regarding the association between interestingness and liking, we obtained a result that shows a positive effect of curiosity in participants' liking responses, making this another perceptual variable that affects liking, in addition to the perception of pitch distributions. It is an outcome that confirms the past, yet few, scales-based studies about the particular link of interestingness and preferability in music (Russell, 1994) and art (Aitken, 1974).

Question 2: What is the relationship between working memory load and liking ratings and how does the fulfilment of melodic expectations affect the ratings?

In the second study of the thesis, the role of working memory regarding aesthetic responses in melodies was explored, as a possible factor of influence in liking ratings that can be controlled quantitatively. In particular, we investigated how accurately listeners recognise relationships between melodic segments under varying stimuli complexity. Also, we explored the degree to which the perception of a successful recognition of such relationships is linked to aesthetic pleasure, by testing for differences in liking ratings between stimuli of lower and higher perceived complexity. As pointed out in Margulis (2016), systematically studying perceived complexity could allow us to measure how listeners process musical structure, and how it is involved in aesthetic preferences. Study 2 is a contribution in this area of research. Its results comprise implications about the perception of complexity as a musical feature, and as a factor of aesthetic judgement of melodies.

Concerning the link between aesthetic judgement and perceived complexity, there was no association found, since the comparisons of the means of liking ratings did not show statistically significant differences between lower and higher complexity stimuli. While it is a study not directly comparable to the literature of the inverted U relationship between complexity and liking, due to the absence of a wide enough range of complexity levels, it examined the same link and found no significance in the specific range. Further, the study showed that listeners perceived changes in complexity in accordance with the design of the study, by rating melodies with bigger intervals as more complex than the ones with smaller intervals. This was also the case with faster melodies, which were perceived as more complex than slower ones. We concluded that listeners were able to correctly identify changes in complexity between the stimuli, and that complexity did not affect their preferences. In connection with musical structure processing, the participants' performance in the pattern recognition task was not affected by changes in the complexity of melodic stimuli. This is a result that does not replicate the findings of past studies with the same hypothesis by Dowling & Bartlett (1981) and Dewitt & Crowder (1986), and that could be explained by the difficulty of the task of how exactly inverted a melody is (contrary to the task in the other studies, that used exact copies of the melodies). As such, it showed that in non extreme low and high points of complexity, there were no statistically significant differences in liking ratings and in changes of melodic expectations.

Question 3: To what degree did we succeed in creating a high-quality product that addresses the needs of music producers?

Developing the generator software in the context of the PhD project provided a novel framework for plugin making, that showed the possibility and displayed an approach of utilising music cognitive literature in the processes of creation and evaluation. It addressed the sub questions that contribute to the answer of Question 3: 1) How can we use this literature for the creation of a DAW plugin? 2) How can we increase the usefulness and quality of the product by analysing the marketplace and utilising user feedback?

Our contribution regarding the first question has been a central point in the project, and a thread that connected the perceptual studies with the creation of a music generator plugin for the DAW environment. As it was shown in Chapter 4, the generation algorithms that were used for the creation of the stimuli in the studies formed a basis, on which further development took place and resulted in the finalised features of the plugin. In order for this to happen, the design of the studies and the levels of manipulation of the cognitive variables were created in a way that allowed the transition from the studies to a generator software. In particular, the studies explored musical features (such as pitch distributions and complexity) which were tested in order to find their optimal values, and results regarding hypotheses of music-cognitive concepts. The melodies used for the scale rating tests were created with generation algorithms that varied the values of the musical features. As a result, the MMM Generator got its initial values and ranges of values from the data acquired by the listener's responses, and precisely setting the values of the variables was made possible by creating corresponding sliders on the plugin's user interface. This way of development showed how it is possible to explore variables of interest, study their significance in aesthetic perception of musical material, and use the design and results of the studies in order to develop functions of a plugin based on them. This approach of analysis by synthesis has been used in music software development for research purposes in the past (Friberg et. al, 2000; Friberg, Bresin & Sundberg, 2006; Livingstone et. al, 2010; Williams et. al, 2017; Micallef Grimaud & Eerola, 2021), but not in the context of music generation, or of the plugin market.

Concerning the second question, about how to increase usefulness and quality of the plugin, the thesis contributed in two ways. Firstly, it did so by providing ways to incorporate theoretical knowledge and concepts in music cognition, along with modern music production best practices, into the process of development. This approach was coupled with an analysis of the marketplace, in order to identify how other generators functioned and what their strengths and weaknesses are, according to user opinions and interface design literature. The complexity of the interfaces and workflow, along with the music output quality, were reviewed and used in order to inform the development of the MMM Generator in terms of avoiding bad practises found in other products, and identifying gaps in the marketplace where there was room for improvement. Secondly, as discussed in the upcoming 5.2.2 section, it provided a study that allowed the analysis of user's thoughts about usefulness and quality, with the use of scales, open text questions and interviews. In summary, the processes of collection and analysis of the data in Study 3 can be considered as a contribution to the plugin industry when it comes to insights about how to approach product development and evaluation. The rationale behind the consideration of this procedure as contribution is based on the fact that it can be used as a generalisable framework, which plugin companies can use for market research purposes during the developmental phase of software creation.

Finally, an important contribution of the project in the industry is the MMM Generator plugin, which achieved statistically significantly high ratings during its evaluation study in the areas of usefulness, interestingness, ease of use, uniqueness, cleanliness, speed, creativity and inspiration. Further, the responses of the open text questions showed an equally positive evaluation regarding the quality of the musical output, as well as usability and potential in the

DAW plugin marketplace. These results were achieved despite the fact that, as documented via online review sources discussed in Chapter 4, users do not yet hold a positive view about the majority of music generation products in terms of quality and usefulness in their music production process, for the role of plugins that create ready-to-use melodies. They suggest that our approach of using a quantitative music cognition based design along songwriting and production techniques was successful, exceeding some of the limitations in the development of music generation plugins regarding usability and output quality, as evidenced by the interviews and open ended answers during Study 3. In particular, users of all levels reported that the plugin generated melodies that sounded well-formed, in contrast to a frequent “random note” - like output they get from other generators. The MMM generator was especially useful in creating ideas to start a new song, by providing a good melodic and chord progression basis, and further, users felt that it was useful due to its features that allowed precise control of the output material.

5.2.2 Methodological contributions

With this thesis we aimed to contribute not only in cognition literature and the industry, but also in ways that extend to broader music research methodology. Specifically, academic concepts were operationalised and implemented in a music technology product. The steps with which this goal was accomplished constitute a generalisable framework that can be used for other projects that aim to perform a transition of theoretical research into its use in industry applications. In the case of the current project, there was a separation of the methodology into two phases. The first phase consisted of the formulation of the perceptual studies. In the second phase, the insights from these studies were used along with past literature and the use of music production workflow concepts, in order to develop the MMM Generator.

Perceptual Studies

With regards to the formulation phase, there were two main concepts that were followed and that allowed the creation of the studies, those of conceptualising the research and of experimenting at the initial creation stage. About the first, the primary objective was to create studies that allowed for a quantification of aesthetic properties, as an attempt to control and improve these properties for the MIDI creation. This objective called for a conceptualisation of what is being investigated, in order to obtain coherency in the project and maintain a logical structure in the controls of the MMM Generator during the second phase. The way to maintain coherency between the studies was to examine related concepts, and use the second study as a follow up to the first. Particularly, in Study 1 we examined the concept of pitch distributions in liking, and in Study 2 we explored pitch intervals in relation to cognitive load, which served as a follow up to the findings of Study 1 where an association of complexity with interestingness and liking was found.

The second concept in forming the studies has to do with the role of experimentation in creating initial ideas for exploration. In the broad literature there exist many choices if one wishes to identify gaps and work on them, but the choice of topics was dictated by the purpose of creating a music generator that used quantified and coded properties that are related to aesthetic appreciation. The topics of research were therefore determined by the kind of contribution that we wanted to make in the industry, as opposed to pursuing the examination of gaps as a first step and looking for ways to implement the outcomes into a software after the fact. In the case of the current thesis, ideas about how the statistical properties of pitch distributions could affect liking responses (and aid in the quality of the generated output) were initially produced through intuition, and preceded the creation of the related study. Then, the experimentation phase included the exploration of various ways to create a pitch generator that operated on the basis of generating through the use of statistical distributions and their parameters. The creation of such a pitch generating algorithm led to the possibility of systematically creating stimuli with statistical properties of controlled values, and consequently, the development of a hypothesis about their effects on liking ratings. At that point, we had substantial context that could allow us to connect these ideas to the literature and design the study. This process highlighted the benefit of connecting ideas from different disciplines. Namely, the connection of examining statistical distribution properties that was borrowed from descriptive statistics, with the relationship between the sense of balance and aesthetic appreciation from the field of music cognition.

Similarly to Study 1, an observation about how the continuation of a melody by adding an exact inversion of itself leads to the continuation sound a bit “off key” for some listeners, resulted in the investigation of how relationships between melodic parts can influence aesthetic perception. The next step was to experiment with the creation of algorithms that took a melody as input and generated altered (i.e. inverted) versions of it, which gave us the opportunity and the design idea to create controlled deviations from initial melodies, and systematically study the cognitive performance of listeners in understanding the relationship between the original and the progressively more complex altered melodies. These algorithms were then used in order to generate the stimuli of Study 2, which was developed in the context of literature that explores the relationship between cognitive load and liking in melodies.

Implementation of studies in the MMM Generator

The algorithms of the two studies, in terms of ways to control the melodic output and of the settings that worked the best in terms of liking responses as revealed in the studies, were fundamental for the development of the MMM Generator. Their consequent use in the product, while utilising the new information gained from the studies in the process, highlights the benefit of planning the research in accordance with the aim of a future implementation in the development of a product. Of course, this did not rule out the employment of pre-existing literature in music cognition that could be used in the creation of the plugin. As applied in the generation of stimuli for the perceptual studies, our interpretation in computer code of the archetypal melodic contours of Meyer (1973) was also used in the MMM Generator, in the section related to shaping the melodies. Further, we were inspired by the concept of the inverted U relationship (Berlyne, 1971) between likings responses and the complexity of a stimulus when choosing and limiting the lowest and highest values of the “tighten” slider, which controlled the complexity of the melodies by affecting pitch distances.

So far we have presented the link between research and the generator software, emphasising the significance of designing research that is made to fit the purpose of software development of this kind, along with using past research by interpreting it in code. However, there is not one particular way to implement the algorithms and findings of the perceptual studies into the MMM Generator. It is crucial that the design is focused towards the needs of the user base, which consists of both hobbyist and professional music producers. In the plugin market it is typical to develop products that appeal to both types of users, since hobbyists need simplicity due to a lack of expert understanding, and professionals have the same need in order to achieve time optimisation in their workflow. For this reason, the amount of success of the product in terms of output quality and usefulness ultimately depends not only on the quality of the algorithms and research, but also on the developer’s understanding and skills when it comes to contemporary music production techniques and workflow. Further, an understanding of the existing solutions in the market and the userbase’s perception of them is also critical, in order to identify gaps and good practices of the competitive software offerings, as it was shown in Study 3 of the current thesis.

Finally, an important aspect of the methodology in creating the plugin was to incorporate feedback from the users who downloaded and used the MMM Generator in their music making environment. In Study 3, we provided a design that includes a combination of scales and open ended questions that were then recoded into main themes and keywords, in order to analyse the users’ sentiment about the quality and usefulness of the software. This process was also especially useful for examining the possibility of further improvements of the plugin towards output quality and the users’ needs, through the use of text-based open structure interviews, and the questionnaire’s open ended questions.

5.2.3 Limitations and future steps

On the subject of this project's limitations, the main aspect is that the participants of the studies worked in an unsupervised setting, at their own place. The completion of the online questionnaires, interviews and tryout of the MMM Generator all happened remotely. It was a design decision led by the fact that online remote distribution of the questionnaires allows for a bigger dataset, as it enables more participants to join the study. As the big majority of the data collection happened during the covid lockdown periods, it was also the only choice that could be made so that the project could be completed.

The effect of a remote design is a possible constraint on achieving significant results, because of the possible existence of factors influencing the accuracy of responses, under the conditions in which the participation is taking place. In particular, an interesting outcome is that we could not confirm a strong association between complexity and liking in the perceptual studies. This could be caused by the design of the stimuli which did not reach extended complexity values, in order for an inverted U relationship to appear, and also by the remote and unsupervised setting in which the participants rated the stimuli and that could hinder the emergence of associations under unfocused or less than ideal listening settings. Further, as pattern recognition in Study 2 was not found to be affected by the changes in complexity levels, the idea of how the fulfilment of expectations could affect liking was not fully fleshed out, and it could be investigated further.

Regarding Study 3, there were cases of participants who reported that the plugin was not working. Some of the reports could have been raised by actual incompatibility between the plugin and the participant's computer system, but a part of them could have been caused by inability of the participant to properly install or operate the plugin. This was an issue causing us to lose some users that wanted to participate in the study and that is related to the remote nature of the design, because we could have provided better help in the testing process if the study had been done in a supervised and local setting. Further, through this study we did not extensively address differences between the degree to which the academic literature domain and the music production and songwriting domain contributed to the ratings and impressions of the listeners towards the MMM Generator. Also, a direct comparison of its quality and usefulness when compared to other generators in the market was not attempted. Finally, it is worth noting that there are factors inseparable to the nature of the study that could affect ratings. For example, when provided with a new technology, participants may view it under a favourable light due to the element of novelty. On the other hand, AI based music generation technology is thought to be perceived with a strong negative bias (Shank et. al, 2022).

To sum up, the span of the thesis points to further exploration of complexity, as an interesting direction for future research in the context of quantifying aesthetic aspects in music, and in its potential importance in the link between melodic expectations and liking. This direction could help to either find more results regarding optimal values for music generation processes relating to complexity, by using different study designs, or it could further challenge the relationship between complexity and liking that is supported in music cognition literature. More broadly, research in the framework we presented in the project, but beyond the aspect of pitch, could provide new ways to generate musical output and to evaluate the ways in which listeners experience aesthetic responses to music. Such routes could be the quantification of variables on music performance, and on rhythm complexity. Regarding future steps in software creation, an extra direction to implementing the ones described about the perceptual studies, would be the creation of a hybrid developmental approach that combines the neural network techniques presented in Chapter 4 with the approach we used for the MMM Generator in this project. This approach would be an augmentation of the software code that would enable novel and hybrid features, along with a broader variety of musical output in comparison to generation algorithms that are primarily based on one of the two original approaches.

5.3 Concluding remarks

We conclude that the thesis was successful in its aims of making a contribution to music cognition literature, to the music software industry, and to the methodology of creating a link between them. It showed an approach of using perceptual concepts that are related to aesthetic responses in music, in order to draw quantifiable elements that were examined in their effect on liking ratings of melodic excerpts. Such concepts were the processing of pitch distributions and perceived melodic complexity, which resulted in studies that offered insights about optimally distributed pitches (termed as the Uniformity Principle) and the association of perceived complexity with pitch intervals and tempo.

The use of these insights, along concepts found in past literature, constituted a basis for the development of a software that generated music parts in the context of the DAW environment. This was achieved with a further analysis concerning music generation plugins that are already released in the market. This analysis gave us an understanding of good practices, and identified gaps that could be addressed by using a software development approach presented in this thesis. A study that was conducted with the aim of understanding the usefulness and quality of the software we developed showed highly positive results in both of those domains. The software is currently under further development in order to be released commercially with [Redacted text], a leading DAW plugin company. Overall, this project provides new insights in quantitative music cognition, and a framework for research-focused plugin development that indicates a strongly positive outlook on the potential of using music cognition for the creation of innovative and high quality plugins.

6. References

Adams, C. R. (1976). Melodic contour typology. *Ethnomusicology*, 179-215.

Aitken, P. P. (1974). Judgments of pleasingness and interestingness as functions of visual complexity. *Journal of Experimental Psychology*, *103*(2), 240.

APA Dictionary of Psychology. (n.d.). Veridical. In *APA Dictionary of Psychology*. Retrieved September 15, 2022, from https://dictionary.apa.org/veridical

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829-839.

Baltar, F., & Brunet, I. (2012). Social research 2.0: virtual snowball sampling method using Facebook. *Internet research*. 22(1), 57-74.

Barnum, C. M., & Dragga, S. (2001). *Usability testing and research*. Allyn & Bacon.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83-100.

Barrouillet, P., Lépine, R., & Camos, V. (2008). Is the influence of working memory capacity on high-level cognition mediated by complexity or resource-dependent elementary processes?. *Psychonomic Bulletin & Review*, *15*(3), 528-534.

Beauvois, M. W. (2007). Quantifying aesthetic preference and perceived complexity for fractal melodies. *Music Perception*, *24*(3), 247-264.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception*, *12*(3), 353-364.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press. Pp. 455-528

Bretan, M., Weinberg, G., & Heck, L. (2016). A unit selection methodology for music generation using deep neural networks. *arXiv preprint arXiv:1612.03789*.

Briot, J. P., Hadjeres, G., & Pachet, F. D. (2017). Deep learning techniques for music generation--a survey. *arXiv preprint arXiv:1709.01620*.

Brown, S., & Jordania, J. (2013). Universals in the world's musics. Psychology of Music, 41(2), 229-248.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & psychophysics*, *8*(5), 279-286.

Berlyne, D. E. (1971). Aesthetics and psychobiology (pp. 86–89). New York, NY: Appleton-Century-Crofts.

Berlyne, D. E. (Ed.). (1974). *Studies in the new experimental aesthetics.* Washington, DC: Hemisphere.

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48.

Cambouropoulos, E. (1997). The role of similarity in categorisation: Music as a case study. In *Proceedings of the Third Triennial ESCOM conference* (pp. 533-538).

Captain Melody 5 - Mixed In Key. (2020). Mixed In Key. https://mixedinkey.com/captain-plugins/captain-melody/

Chew, E. (2002, September). The spiral array: An algorithm for determining key boundaries. In *International Conference on Music and Artificial Intelligence* (pp. 18-31). Springer, Berlin, Heidelberg.

Christenson, P. G., & Peterson, J. B. (1988). Genre and gender in the structure of music preferences. *Communication research*, *15*(3), 282-301.

Chu, H., Urtasun, R., & Fidler, S. (2016). Song from PI: A musically plausible network for pop music generation. *arXiv preprint arXiv:1611.03477*.

Colley, A. (2008). Young people's musical taste: Relationship with gender and gender‐related traits 1. *Journal of applied social psychology*, *38*(8), 2039-2055.

Collins, T., Laney, R., Willis, A., & Garthwaite, P. H. (2016). Developing and evaluating computational models of musical style. *AI EDAM*, *30*(1), 16-43.

Coleman, J. (1958). Relational analysis: The study of social organizations with survey methods. *Human organization*, *17*(4), 28-36.

Conklin, D. (2016). Chord sequence generation with semiotic patterns. *Journal of Mathematics and Music*, *10*(2), 92-106.

Cowby. (2018, June, 7). *Evabeat MELODY SAUCE alternative* [Online forum post]. KVR. https://www.kvraudio.com/forum/viewtopic.php?t=506062

[Redacted text]

Creswell, J.W., & Plano Clark, V.L. (2011). *Designing and Conducting Mixed Methods Research, 2nd Edition.* Sage Publications: Los Angeles

Delplanque, J., De Loof, E., Janssens, C., & Verguts, T. (2019). The sound of beauty: How complexity determines aesthetic preference. *Acta Psychologica*, *192*, 146-152.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*(3939), 1604-1605.

Dewitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception*, *3*(3), 259-274.

Doornbusch, P. (2004). Computer sound synthesis in 1951: The music of CSIRAC. *Computer Music Journal*, *28*(1), 10-25.

Dowling, W. J. (1970). Two Processes in Melody Recognition. *The Journal of the Acoustical Society of America*, *48*(1A), 89-89.

Dowling, W. J. (1971). Recognition of inversions of melodies and melodie contours. *Perception & Psychophysics*, *9*(3), 348-349.

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, *12*(5), 417-421.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology: A Journal of Research in Music Cognition*, *1*(1), 30.

Dubreuil, A. (2020). *Hands-on music generation with magenta: Explore the role of deep learning in music generation and assisted music composition*. Packt Publishing Ltd.

Drewes, F., & Högberg, J. (2007, May). An algebra for tree-based music generation. In *International Conference on Algebraic Informatics* (pp. 172-188). Springer, Berlin, Heidelberg.

Eck, D., & Schmidhuber, J. (2002, September). Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. In *Proceedings of the 12th IEEE workshop on neural networks for signal processing* (pp. 747-756). IEEE.

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music*, *34*(3), 337-371.

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review*, *11*(1).

Eichert, R., Schmidt, L., & Seifert, U. (1996, September). Logic, gestalt theory, and neural computation in research on auditory perceptual organization. In *Joint International Conference on Cognitive and Systematic Musicology* (pp. 70-88). Springer, Berlin, Heidelberg.

Fabius, O., & van Amersfoort, J. R. (2015). Variational recurrent auto-encoders, June 2015. *arXiv preprint arXiv:1412.6581*.

Farley, F. H., & Weinstock, C. A. (1980). Experimental aesthetics: Children's complexity preference in original art and photo reproductions. *Bulletin of the Psychonomic Society*, *15*(3), 194-196.

Fischer, H. (2011). *A history of the central limit theorem: from classical to modern probability theory*. New York: Springer.

Friberg, A., Colombo, V., Frydén, L., & Sundberg, J. (2000). Generating musical performances with Director Musices. *Computer Music Journal*, *24*(3), 23-29.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in cognitive psychology*, *2*(2), 145.

Fritz, T. H., Halfpaap, J., Grahl, S., Kirkland, A., & Villringer, A. (2013). Musical feedback during exercise machine workout enhances mood. *Frontiers in psychology*, *4*, 921.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* (pp. 326-366). MIT press.

Gordon, J., & Gridley, M. C. (2013). Musical preferences as a function of stimulus complexity of piano jazz. *Creativity Research Journal*, *25*(1), 143-146.

Gordon, K. (2020, March, 1). *5 Principles of Visual Design in UX*. Nngroup. https://www.nngroup.com/articles/principles-visual-design/

Hadjeres, G., Nielsen, F., & Pachet, F. (2017, November). GLSR-VAE: Geodesic latent space regularization for variational autoencoder architectures. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-7). IEEE.

Harrison, P., & Pearce, M. T. (2020). Simultaneous consonance in music perception and composition. *Psychological Review*, *127*(2), 216.

Herremans, D., Martens, D., & Sörensen, K. (2014). Dance hit song prediction. *Journal of New Music Research*, *43*(3), 291-302.

Hood, M. (1982). *The ethnomusicologist* (Vol. 140). Kent, Ohio: Kent State University Press.

Huang, C. Z. A., Hawthorne, C., Roberts, A., Dinculescu, M., Wexler, J., Hong, L., & Howcroft, J. (2019). The Bach doodle: Approachable music composition with machine learning at scale. *arXiv preprint arXiv:1907.06637*.

[Redacted text]

Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., ... & Eck, D. (2018). Music transformer. *arXiv preprint arXiv:1809.04281*.

Hughes, M., Riegel, S., Beutler, T., DeSantis, D., Gallagher, I., Haywood, K., … , & Slama, T., Ableton. (2022). *Ableton Reference Manual Version 11 for Windows and Mac.*

Hunt, S., Nash, C., & Mitchell, T. (2017, September). Thoughts on interactive generative music composition. In *2nd Conference on Computer Simulation of Musical Creativity, Milton Keynes, UK*.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* The MIT Press.

Imamoglu, Ç. (2000). Complexity, liking and familiarity: Architecture and Non-architecture Turkish students' assessments of traditional and modern house facades. *Journal of Environmental Psychology*, *20*(1), 5-16.

Jaques, N., Gu, S., Bahdanau, D., Hernández-Lobato, J. M., Turner, R. E., & Eck, D. (2017, July). Sequence tutor: Conservative fine-tuning of sequence generation models with kl-control. In *International Conference on Machine Learning* (pp. 1645-1654).

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 369.

Keller, P. E., & Schubert, E. (2011). Cognitive and affective judgements of syncopated musical themes. *Advances in Cognitive Psychology*, *7*, 142-156.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians?. *Journal of Basic and Clinical Physiology and Pharmacology*, *12*(2), 125-144.

Koelsch, S. (2006). Significance of Broca's area and ventral premotor cortex for music-syntactic processing. *Cortex*, *42*(4), 518-520.

Kozmik Kandi. (2021, July, 8). *Making Music with Magenta Studio - An interview with a musician* [Video]. YouTube. URL https://www.youtube.com/watch?v=Sd1g7ZuX06Y

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *51*(4), 336-353.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review*, *89*(4), 334-368.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological science*, *1*(1), 70-73.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, *17*(2), 151-195.

Kubovy, M. & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80, 97-126

Kulkarni, R., Gaikwad, R., Sugandhi, R., Kulkarni, P., & Kone, S. (2019). Survey on deep learning in music using GAN. *International Journal of Engineering Research & Technology*, *8*(9), 646-648.

Lattner, S., Grachten, M., & Widmer, G. (2018). Imposing higher-level structure in polyphonic music generation using convolutional restricted boltzmann machines and constraints. *Journal of Creative Music Systems*, *2*, 1-31.

Lerdahl, F., & Jackendoff, R. S. (1996). *A Generative Theory of Tonal Music, reissue, with a new preface*. MIT press.

Levitin, D. J. (2002). *Foundations of cognitive psychology: core readings*. MIT press. pp. 481-498

Levitin, D. J., Chordia, P., & Menon, V. (2012). Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proceedings of the National Academy of Sciences*, *109*(10), 3716-3720.

Lewis, J. R. (1992, October). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the human factors society annual meeting* (Vol. 36, No. 16, pp. 1259-1260). Sage CA: Los Angeles, CA: Sage Publications.

Liu, Y., Liu, G., Wei, D., Li, Q., Yuan, G., Wu, S., ... & Zhao, X. (2018). Effects of musical tempo on musicians' and non-musicians' emotional experience when listening to music. *Frontiers in Psychology*, *9*, 2118.

Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, *34*(1), 41-64.

Lyu, Q., Wu, Z., Zhu, J., & Meng, H. (2015, June). Modelling high-dimensional sequences with lstm-rtrbm: Application to polyphonic music generation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence* (pp. 4138-4139).

Manaris, B., Vaughan, D., Wagner, C., Romero, J., & Davis, R. B. (2003, April). Evolutionary music and the Zipf-Mandelbrot law: Developing fitness functions for pleasant music. In *Workshops on Applications of Evolutionary Computation* (pp. 522-534). Springer, Berlin, Heidelberg.

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception*, *22*(4), 663-714.

Margulis, E. H., & Simchy-Gross, R. (2016). Repetition enhances the musicality of randomly generated tone sequences. *Music Perception: An Interdisciplinary Journal*, *33*(4), 509-514.

Marin, M. M., & Leder, H. (2013). Examining complexity across domains: relating subjective and objective measures of affective environmental scenes, paintings and music. *PloS one*, *8*(8), e72412.

Martens, G., De Meyer, H., De Baets, B., Leman, M., Lesaffre, M., & Martens, J. P. (2005). Tree-based versus distance-based key recognition in musical audio. *Soft Computing*, *9*(8), 565-574.

Medina, D., & Barraza, P. (2019). Efficiency of attentional networks in musicians and non-musicians. *Heliyon*, *5*(3), e01315.

Melody Sauce | Evabeat MIDI VST Plugin | DAW Plugin. (2020). Evabeat. https://www.evabeat.com/

Meyer, L.B. (1956). *Emotion and meaning in music.* University of Chicago Press: Chicago.

Meyer, L. B. (1973). *Explaining Music: Essays and Explorations.* University of California Press: California

Micallef Grimaud, A., & Eerola, T. (2021). EmoteControl: an interactive system for real-time control of emotional expression in music. *Personal and Ubiquitous Computing*, *25*(4), 677-689.

Narmour, E. (1989). The “genetic code” of melody: Cognitive structures generated by the implication-realization model. *Contemporary Music Review*, *4*(1), 45-63.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press.

Ni, Y., Santos-Rodriguez, R., Mcvicar, M., & De Bie, T. (2011, December). Hit song science once again a science. In *4th International Workshop on Machine Learning and Music*.

Nielsen, J. (1994, April). Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 152-158).

Nierhaus, G. (2009). *Algorithmic composition: paradigms of automated music generation*. Springer Science & Business Media.

Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Patel, A. D. (2010). Music, biological evolution, and the brain. In C. Levander & C. Henry (Eds.). *Emerging disciplines* (pp. 91–144). Houston, TX: Rice University Press.

Payne, C. (2019). MuseNet. *OpenAI Blog*, https://openai. com/blog/musenet.

Pearce, M. T., & G. A. Wiggins. (2004, August). Rethinking gestalt influences on melodic expectancy. In *Proceedings of the 8th International Conference on Music Perception and Cognition*, (pp. 367-371).

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378-395.

Pechmann, T. & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition. 20* (3)., 314-320.

Pembrook, R. G. (1987). The effect of vocalization on melodic memory conservation. *Journal of Research in Music Education*, *35*(3), 155-169.

Qpbeats. (2019, November, 4). *Just Bought Captain Plugins, Anyone have experience with Captain Melody?* [Online forum post]. Reddit. https://www.reddit.com/r/edmproduction/comments/drhwv2/just_bought_captain_plugins_anyone_have/

Raybaut, P. (2009). Spyder-documentation. *Available online at: pythonhosted.org*.

Riff Generation - In Session Audio. (2020). In Session Audio. https://insessionaudio.com/products/riff-generation/

Ritossa, D. A., & Rickard, N. S. (2004). The relative utility of ‘pleasantness’ and ‘liking’ dimensions in predicting the emotions expressed by music. *Psychology of Music*, *32*(1), 5-22.

Roberts, A., Engel, J., & Eck, D. (2017, December). Hierarchical variational autoencoders for music. In *NIPS Workshop on Machine Learning for Creativity and Design* (Vol. 3).

Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., ... & Eck, D. (2019). Magenta studio: Augmenting creativity with deep learning in ableton live. In *Proceedings of the 7th International Workshop on Musical Metacreation*.

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018, July). A hierarchical latent vector model for learning long-term structure in music. In *International conference on machine learning* (pp. 4364-4373). PMLR.

[Redacted text]

Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, *83*(2), 164-175.

Russell, P. A. (1982). Relationships between judgements of the complexity, pleasingness and interestingness of music. *Current Psychology*, *2*(1), 195-201.

Russell, P. A. (1994). Preferability, pleasingness, and interestingness: Relationships between evaluative judgements in empirical aesthetics. *Empirical Studies of the Arts*, *12*(2), 141-157.

SadowickProduction. (2019, February, 15). *Ableton Live 10 & Google AI - Magenta Studio* [Video]. YouTube. URL https://www.youtube.com/watch?v=MmL6RAZN_yg

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception*, *14*(3), 295-318.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: tests of children and adults. *Journal of Experimental Psychology: General*, *131*(4), 511.

Seeger, C. (1960). On the moods of a music-logic. *Journal of the American Musicological Society*, *13*(1/3), 224-261.

Shah, P., & Miyake, A. (1999). Models of Working Memory--An Introduction.

Shank, D. B., Stefanik, C., Stuhlsatz, C., Kacirek, K., & Belfi, A. M. (2022). AI composer bias: Listeners like music less when they think it was composed by an AI. Journal of Experimental Psychology: Applied.

Sherman, A., Grabowecky, M., & Suzuki, S. (2015). In the working memory of the beholder: Art appreciation is enhanced when visual complexity is compatible with working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 898-903.

Silverman, M. J. (2012). Effects of melodic complexity and rhythm on working memory as measured by digit recall performance. *Music and Medicine, 4*(1), 22–27.

Simchy-Gross, R., & Margulis, E. H. (2018). The sound-to-music illusion: Repetition can musicalize nonspeech sounds. *Music & Science, 1*. 2059204317731992.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129-138.

Smith, D., & Wood, C. (1981, October). The 'USI', or Universal Synthesizer Interface. In *Audio Engineering Society Convention 70*. Audio Engineering Society.

Sorensen, A., & Gardner, H. (2010, October). Programming with time: cyber-physical programming with impromptu. In *Proceedings of the ACM international conference on Object oriented programming systems languages and applications* (pp. 822-834).

Stathakopoulos, V. (2005). *Market research methods* (pp. 133-146). Stamoulis Publications: Athens.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Taruskin, R., (2009). *Oxford History of Western Music: Volume 5*. New York: Oxford University Press (p. 56)

Tatanka. (2018, October, 20). *Simmo: I think these are completely different tools. Riff Generator is more a very powerful arp/sequencer kind of thing* [Comment on the online forum post *In Session Audio Fluid Series and Riff Generation*]. KVR. https://www.kvraudio.com/forum/viewtopic.php?t=491729&start=30

Temperley, D. (2007). *Music and probability*. Mit Press. pp 79-98

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, *32*(2), 418-444.

Temperley, D., & Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception*, *25*(3), 193-212.

Thompson, L.M., & Yankeelov, M.J. (2012). Music and the Phonological Loop. *Proceedings of the 12th International Conference on Music Perception and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, 980-986.

Tidwell, J., Brewer, C., & Valencia-Brooks, A. (2020). *Designing Interfaces: Patterns for Effective Interaction Design.* O'Reilly.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological review*, *107*(4), 885.

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140096.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child development*, 821-830.

Van Der Weij, B., Pearce, M. T., & Honing, H. (2017). A probabilistic model of meter perception: Simulating enculturation. *Frontiers in psychology*, *8*, 824.

Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral science*, *11*(2), 105-114.

Von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, *18*(1), 59-85.

Vos, P. G., & Van Geenen, E. W. (1996). A parallel-processing key-finding model. *Music Perception*, *14*(2), 185-223.

Voss, R. F., & Clarke, J. (1978). ”1/f noise”in music: Music from 1/f noise. *The Journal of the Acoustical Society of America*, *63*(1), 258-263.

Weaver Beats. (2019, March, 16). *Melody Sauce VST Review* [Video]. YouTube. URL https://www.youtube.com/watch?v=CTiU6b_UEWk

Werts, D. (1997). Chord Progression Data Obtained in an Experiment by J. Bharucha and C. Krumhansl. *Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology*, *1317*, 200.

Whorley, R. P., & Conklin, D. (2016). Music generation from statistical models of harmony. *Journal of New Music Research*, *45*(2), 160-183.

Williams, D., Kirke, A., Miranda, E., Daly, I., Hwang, F., Weaver, J., & Nasuto, S. (2017). Affective calibration of musical feature sets in an emotionally intelligent music composition system. *ACM Transactions on Applied Perception (TAP)*, *14*(3), 1-13.

Williamson, V. J., Mitchell, T., Hitch, G. J., & Baddeley, A. D. (2010). Musicians’ memory for verbal and tonal materials under conditions of irrelevant sound. *Psychology of Music*, *38*(3), 331-350.

Wilson, S., Cottle, D., & Collins, N. (2011). *The SuperCollider Book*. The MIT Press.

Wu, D., Kendrick, K. M., Levitin, D. J., Li, C., & Yao, D. (2015). Bach is the father of harmony: revealed by a 1/f fluctuation analysis across musical genres. *PLoS One*, *10*(11), e0142431.

Yang, L. C., Chou, S. Y., & Yang, Y. H. (2017). MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. *arXiv preprint arXiv:1703.10847*.

Yoshino, I., & Abe, J. I. (2004). Cognitive modeling of key interpretation in melody perception 1. *Japanese Psychological Research*, *46*(4), 283-297.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2, Pt.2), 1–27.

Ziv, N., & Keydar, E. (2009). The relationship between creative potential, aesthetic response to music, and musical preferences. *Creativity Research Journal*, *21*(1), 125-133.

7. Appendix

Questionnaire of Study 1.

Melody Rating Questionnaire

*Required

Information Sheet

Date: 27/04/2019
Project title: Statistical properties of melodies and their effect on melody perception.

Introduction: You are being invited to take part in a research project being carried out by a researcher at the University of Sheffield. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

1. What is the project's purpose?
The study aims to develop a better understanding of how quantitative properties of melodies can influence people's perception of them. The result of this study can help in the formation of theory about statistical regularities and their effect on the appreciation of melodies.

2. What will happen to me if I take part?
The experiment is carried out via a Google Forms questionnaire at your location. The experiment will involve listening to and answering questions about melodies, as well as some demographic questions. The experiment should take about 15 minutes to complete.

3. What are the possible disadvantages and risks of taking part?
Taking part in this research is entirely voluntary. If you decide to take part, you will be asked to agree to a consent form. You can still withdraw at any time without giving a reason. There are no possible disadvantages or risks of taking part other than problems related to the volume of sounds. It will be beneficial to adjust the volume of your listening device to a comfortable level in order to avoid any possible irritation by the stimuli.

4. What if something goes wrong and I wish to complain about the research?
If something goes wrong you can raise a complaint with the research project lead or the University. For this research project, PhD student Alex Stamatiadis is the overall lead for the whole project (astamatiades1@sheffield.ac.uk). If you feel your complaint was not handled satisfactorily, you can also contact the University's Registrar and Secretary at registrar@sheffield.ac.uk.

5. Will my taking part in this project be kept confidential?
All the information that we collect during the course of the research will be kept strictly confidential and will only be accessible to members of the research team. The data you will provide will not contain identifiable information, and you will not be able to be identified in any reports or publications.

6. What will happen to the data collected, and the results of the research project?
We will aim to publish the results in academic journals. They will also be presented at academic conferences and will be used in future quantitative studies in relation to people's responses to melodies.

7. Who has ethically reviewed the project?
The project has been approved by the Department of Music's Ethics Review Committee at the University of Sheffield.

8. Who is organising and funding the research?
The project is being undertaken and self-funded by Alex Stamatiadis, a PhD student in the Music Department, and supervised by Dr Renee Timmers.

9. Contact for further information
Please contact Alex Stamatiades, astamatiades1@sheffield.ac.uk, or Dr Renee Timmers, project supervisor, at r.timmers@sheffield.ac.uk

Thank you for reading this. We hope you would like to take part in this research project.

Consent Form: Statistical properties of melodies and their effect on melody perception.

Please tick the appropriate boxes.

Consent Form

By selecting the box below, you agree to the following:

• I have read the Information Sheet and understand its contents.
• I understand that I am free to withdraw from the research at any time, and I do not have to provide a reason.
• I confirm that I am willing to be a participant in the above research study.
• I note that my data will be anonymous.

9. *

Tick all that apply.

☐ I agree.

Type Arp

Melody 1

 http://youtube.com/watch?v=ujW6cSti4jc

10. How much do you like this melody? *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

11. How complex does this melody sound?

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

12. How interesting does this melody sound?

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

Melody 2


http://youtube.com/watch?v=-mzpsMQHV2w

13. How much do you like this melody? *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

14. How complex does this melody sound?

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

15. How interesting does this melody sound?

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

130. What is your music education level? *

Mark only one oval.

- No music education.
- Self-taught musician.
- Music high school.
- Higher education in music.
- Postgraduate education in music.

131. What music genres do you like?

Tick all that apply.

- [ ] Pop / Top 40
- [ ] R&B / Hip-Hop
- [ ] Rock
- [ ] Dance / Electronic
- [ ] International / Ethnic
- [ ] Other

132. What is your age group?

Mark only one oval.

- 18-24
- 25-44
- 45-64
- 65+

133. What is your gender?

Mark only one oval.

- Female
- Male

Questionnaire of Study 2.

Information Sheet

Date: 02/02/2020

Project title: Statistical properties of melodies and their effect on melody perception.

Introduction: You are being invited to take part in a research project being carried out by a researcher at the University of Sheffield. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

1. What is the project's purpose?
The study aims to develop a better understanding of how mathematical and psychological properties of melodies can influence people's perception of them. The result of this study can help us to understand better how the relations between notes affect the appreciation of melodies.

2. What will happen to me if I take part?
The experiment is carried out via a Google Forms questionnaire at your location. The experiment will involve listening to and answering questions about melodies, as well as some demographic questions. The experiment should take about 15 minutes to complete.

3. What are the possible disadvantages and risks of taking part?
Taking part in this research is entirely voluntary. If you decide to take part, you will be asked to agree to a consent form. You can still withdraw at any time without giving a reason. There are no possible disadvantages or risks of taking part, you only need to make sure that you feel comfortable with the volume level.

4. What if something goes wrong and I wish to complain about the research?
If something goes wrong you can raise a complaint with the research project lead or the University. For this research project, PhD student Alex Stamatiadis is the overall lead for the whole project (astamatiades1@sheffield.ac.uk). If you feel your complaint was not handled satisfactorily, you can also contact the University's Registrar and Secretary at registrar@sheffield.ac.uk.

5. Will my taking part in this project be kept confidential?
All the information that we collect during the course of the research will be kept strictly confidential and will only be accessible to members of the research team. The data you will provide will not contain identifiable information, and you will not be able to be identified in any reports or publications.

6. What will happen to the data collected, and the results of the research project?
We will aim to publish the results in academic journals. They will also be presented at academic conferences and will be used in future quantitative studies in relation to people's responses to melodies.

7. Who has ethically reviewed the project?
The project has been approved by the Department of Music's Ethics Review Committee at the University of Sheffield.

8. Who is organising and funding the research?
The project is being undertaken and self-funded by Alex Stamatiadis, a PhD student in the Music Department, and supervised by Dr Renee Timmers.

9. Contact for further information
Please contact Alex Stamatiades, astamatiades1@sheffield.ac.uk, or Dr Renee Timmers, project supervisor, at r.timmers@sheffield.ac.uk. Questions can also be directed to Dr Simon Keegan-Phipps, Head of Department of Music at s.keegan-phipps@sheffield.ac.uk.

Information related to ethics and data management:
According to data protection legislation, we are required to inform you that the legal basis we are applying in order to process your personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)). Further information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly. This study has received ethical approval from the Department of Music, and is conducted in accordance with the research ethics guidelines of The University of Sheffield.

4. I understand that my taking part is voluntary and that I can withdraw from the study at any time by closing the questionnaire and not submitting the data. *

Mark only one oval.

- Yes
- No

How my information will be used during and after the project

5. I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form. *

Mark only one oval.

- Yes
- No

6. I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form. *

Mark only one oval.

- Yes
- No

7. I give permission for the data that I provide to be deposited in Google Drive database so it can be used for future research and learning. *

Mark only one oval.

- Yes
- No

So that the information you provide can be used legally by the researchers

8. I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield. *

Mark only one oval.

- Yes
- No

Consent Form

By selecting the box below, you agree to the following:

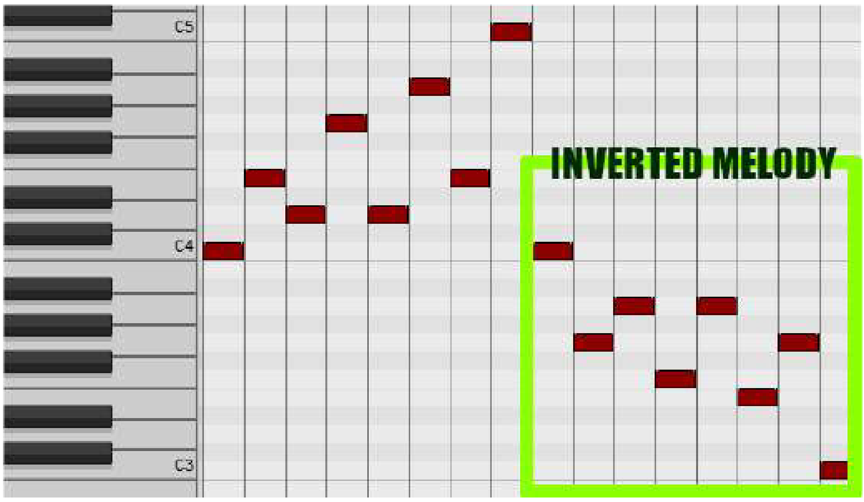
- I have read the Information Sheet and understand its contents.
- I confirm that I am willing to be a participant in the above research study.
- I note that my data will be anonymous.

9. *

Tick all that apply.

- [ ] I agree.

In the questionnaire you will listen to different kinds of melodies and rate them according to how much you like each of them. These melodies contain two sections: The first half contains a musical phrase, and the second half is the same phrase but inverted (some melodies will have more exact inversions than others). You will also be asked to rate the level of preciseness of inversion of that second part of the melody. To familiarise yourself, an example of what is the second, inverted part of a melody is given below. The notes inside the green box are the second half of the whole melody, i.e. the inverted part.



Melodic Type 1

Melody 1

http://youtube.com/watch?v=qrykL2zS3jA

10. How much do you like this melody? *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

11. How much of an exact inversion of the first part is the second part of the melody? *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

12. How complex does this melody sound? *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

Melody 2


http://youtube.com/watch?v=Z4jxQ4SFOso

13. How much do you like this melody? *

Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

64. What is your age group?

Mark only one oval.

- 18-24
- 25-44
- 45-64
- 65+

65. What is your gender?

Mark only one oval.

- Female
- Male
- Other

Questionnaire of Study 3.

Evaluation of the Music Generator

*Required

Information Sheet
Date: 21/01/2021
Project title: Evaluation of the Music Generator.

Introduction: You are being invited to take part in a research project being carried out by a researcher at the University of Sheffield. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

1. What is the project's purpose?
The study aims to develop a better understanding of how music producers evaluate the Generator, and how they would work with it. You will be asked questions related to the usefulness of the Generator, and your preferences when it comes to using it. The results of this study can help us understand better to what degree can the utilisation of academic research on music cognition benefit the development of music creation software.

2. What will happen to me if I take part?
The experiment is carried out via a Google Forms questionnaire at your location. The experiment will involve working with the generator for at least 30 minutes, and answering questions about your evaluation of it (10 to 15 minutes).

3. What are the possible disadvantages and risks of taking part?
Taking part in this research is entirely voluntary. If you decide to take part, you will be asked to agree to a consent form. You can still withdraw at any time without giving a reason. There are no possible disadvantages or risks of taking part.

4. What if something goes wrong and I wish to complain about the research?
If something goes wrong you can raise a complaint with the research project lead or the University. For this research project, PhD student Alex Stamatiadis is the overall lead for the whole project (astamatiades1@sheffield.ac.uk).

5. Will my taking part in this project be kept confidential?
All the information that we collect during the course of the research will be kept strictly confidential and will only be accessible to members of the research team. The data you will provide will not contain identifiable information, and you will not be able to be identified in any reports or publications.

6. What will happen to the data collected, and the results of the research project?
We will aim to publish the results in academic journals. They will also be presented at academic conferences and will be used in future studies in relation to the use of academic literature for the creation of music software.

7. Who has ethically reviewed the project?
The project has been approved by the Department of Music's Ethics Review Committee at the University of Sheffield.

8. Who is organising and funding the research?
The project is being undertaken and self-funded by Alex Stamatiadis, a PhD student in the Music Department, and supervised by Prof Renee Timmers.

9. Contact for further information
Please contact Alex Stamatiades, astamatiades1@sheffield.ac.uk, or Prof Renee Timmers, project supervisor, at r.timmers@sheffield.ac.uk. Questions can also be directed to Dr Simon Keegan-Phipps, Head of Department of Music at s.keegan-phipps@sheffield.ac.uk.

Information related to ethics and data management:
According to data protection legislation, we are required to inform you that the legal basis we are applying in order to process your personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)). Further information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly. This study has received ethical approval from the Department of Music, and is conducted in accordance with the research ethics guidelines of The University of Sheffield.

Thank you for reading this. We hope you would like to take part in this research project.

Consent Form: Evaluation of the Music Generator.

Please tick the appropriate boxes.

Taking Part in the Project

1. I have read and understood the project information sheet dated 21/01/2021. (If you will answer No to this question please do not proceed with this consent form until you are fully aware of what your participation in the project will mean.) *

Mark only one oval.

- Yes
- No

2. I have been given the opportunity to ask questions about the project (I have been given the e-mail address of the researcher). *

Mark only one oval.

- Yes
- No

3. I agree to take part in the project. I understand that taking part in the project will include answering questions about my evaluation of the Generator. *

Mark only one oval.

- Yes
- No

4. I understand that my taking part is voluntary and that I can withdraw from the study at any time by closing the questionnaire and not submitting the data. *

Mark only one oval.

- Yes
- No

How my information will be used during and after the project

5. I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form. *

Mark only one oval.

- Yes
- No

6. I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form. *

Mark only one oval.

- Yes
- No

7. I give permission for the anonymous data that I provide to be deposited in Google Drive database so it can be used for future research and learning. *

Mark only one oval.

- Yes
- No

So that the information you provide can be used legally by the researchers

8. I agree to assign the copyright I hold in any materials generated (= my written feedback) as part of this project to The University of Sheffield. *

Mark only one oval.

- Yes
- No

Consent Form

By selecting the box below, you agree to the following:

• I have read the Information Sheet and understand its contents.
• I confirm that I am willing to be a participant in the above research study.
• I note that my data will be anonymous.

9. *

Tick all that apply.

- [ ] I agree.

10. Having used the Generator for at least 30 minutes, what are your general first impressions? *

Usefulness	From 0 to 10 (less to more), how much do these words / phrases describe your view on the Generator?

11. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unuseful	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Useful

12. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Uninteresting	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Interesting

13. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Difficult to use	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Easy to use

14. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Ordinary	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Unique

15. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unoriginal	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Creative

16. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Uninspiring	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Inspiring

17. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Cluttered	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Clean

18. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Doesn't provide usable results quickly	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Provides usable results quicl

19. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Doesn't provide the features I need	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Provides the features I need

20. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Disrupts workflow	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Fits workflow

21. *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Doesn't do what I need for my purpose	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Does what I need for my pu

22. What aspect of the Generator did you find most useful? *

23. What aspect of the Generator did you find least useful? *

24. If you had the opportunity to improve something about the Generator to make it more useful? What would you improve? *

Use cases

25. Can the use of a plugin like this fit into your workflow? If yes, how? If no, why? *

26. How would you mostly use a generator like this? *

Mark only one oval.

- ◯ Mostly for starting new tracks.
- ◯ Mostly for filling parts in tracks that I have already started.
- ◯ An equal mixture of both.

27. What is your opinion about features such as "Balance" / "Tighten", that are vague in what they do on the technical side, and instead focus on what is the perceived effect of them? Would you have a preference between technical and perception based feature labelling? *

Table 21: Exact frequencies of respondents' ratings in Study 3 questionnaire.

Variable	Frequencies of Rating Values										Median Rating
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
Useful	0	0	1	0	0	1	5	3	0	6	8
Interesting	0	0	0	0	2	2	1	3	1	7	9
Easy To Use	0	1	0	0	3	1	1	4	3	3	8
Unique	0	1	0	1	1	2	2	2	2	5	8
Creative	0	0	1	1	2	1	1	4	1	5	8
Inspiring	0	0	0	1	1	0	3	2	4	5	9
Clean	0	1	0	1	1	1	1	5	3	3	8
Quick	1	0	0	1	2	0	1	4	2	5	8
Features I Need	1	0	2	4	1	1	0	0	4	3	6
Fits Workflow	0	1	0	1	3	2	1	1	2	5	8
Fits My Purposes	0	1	3	0	2	3	0	1	1	5	6