Observing the Relationship Between Melodic Traits and their Effect on the Suitability of a Generated Melody

MG246981

Abstract—**Indie start-up studios often have very limited resources to produce games. Audio production is a field with few specialists compared to the other critical disciplines, and as such are an expensive investment for development. There is some research into automatic music generation, but it is limited to very specific use cases. Tools built specifically to assist small studios in music composition are limited and not feasibly available, requiring significant tailoring to make suitable for a game. Most deep-learning models for general music generation do not have an automatic evaluation method, as such a set of data to evaluate melodies is not readily available. The experiments for this research were conducted at Falmouth University to gather data to help provide a step towards an evaluation method. This study found that monophonic melodies with shorter steps between notes were generally more effective than those with larger pitch differences, appealing to a wider audience and being more viable as a musical line. Melodies with greater proportions of rests were found to have the opposite effect of reducing the appeal of a melody line. Such data is valuable when weighing what a melody requires from music generation AI, and could lead to an independent fitness function. This led to the conclusion that a gamified product could be distributed to many users and could lead producing a larger data set which could evaluate melodies with human character. Such a data set could be used to improve any AI model relying on a procedural set of rules.**

I. Introduction

MUSIC production for indie studios is often a difficulty, with little professional and academic development in the field[1]. A survey from 2014 found that per-project for indie studios, audio production cost $9,830 (£5,996.30 in 2014) [2]. Despite music production being a critical part of game production, it has remained inaccessible for developers without a background or developed skills in music or audio. This paper explores the use of an artefact which can assist in producing music for games and developing a data set for future evaluation methods. Machine generated art content has met a boom in recent years, and is proving to be an incredibly versatile tool [3]. Despite not receiving the same level of academic and industrial depth as AI image generation [4], AI generated music content is growing steadily [5], so it becomes much more possible to produce a reliable and viable music generation AI for new video game studios. The study takes place at Falmouth University, interacting with students on games courses to understand a link between musical intervals and the appeal of a monophonic melody through creative listening exercises with participants, tested by accumulating and analysing data from the use of a music generation algorithm. The results found that some melodic traits influence the appeal of it more than others. This study proposes a method to provide data towards an evaluation process that music generation models could benefit from when analysing their produced melodies and iterating upon them to make overall stronger results.

II. Literature Review

Algorithmic generation of music is a widely expanding field of study. Recent papers published explored the uses of such tools in improving quality of life for the elderly by providing an accessible way of producing one's own music[6] or supplementing the music composition process for users with a background or occupation in music [7]. Isolated music generation is also a topic of interest as it measures the similarity of purely computer driven music to human crafted music [8]. The findings in Liu et al [9] suggest that algorithmic music based on a human motif is met with positive responses from audiences interacting with these products. Such studies expand on the range of possibility in this area, and provide a tangible reference to the scope and progress into this subject. Music is an easily overlooked part of game production, and it is most likely linear and looping. The boundaries of graphics are constantly being pushed to achieve greater results. To further push away from traditional composition is to build a procedural tool which relies on live input from a player's actions and create novel tracks which accompany the game's progression [10]. The most appropriate models for this task are often debated and usually not discussed in depth together. In 2012, Mazurowski studied two different methodologies with different algorithms and explored them with accompanying artefacts. He concluded that reinforcement based learning like those of genetic algorithms was suitable as the path forward to investigating algorithmic composition [11].

A. Interactive Generation

The effectiveness of methods used for generation of musical phrases are a matter of debate. Genetic algorithms are a common approach to such a task, but are criticised due to a lack of reliability and being too resource intensive [12]. Genetic algorithms are also praised for their ability to produce appropriate and original works when used in conjunction with other techniques [13]. One particular study found favourable results in supplemental generation. The machine took in a melody in MIDI format as a base, and it took popular songs as an exemplar to produce a variety of accompanying chords. This is useful in composition procedures as it allows for fast experimentation and selection, and a small sample of amateur and professional audiences found that the generated results were enjoyable by both [14].

B. Isolated Generation

The usage of more automated methodology for generating pieces of music has been criticised by Masuda and Iba for producing pieces with very little control over them, suggesting that the research into this was less effective as a result [15]. A study supporting isolated generation was also produced, arguing that it would provide more inspiration for composers and that it would produce more pieces much faster [16]. Games relying on a procedural music system are also much more reliant on analysis and parameters rather than user input. A study in 2012 investigated music generation at run-time based off of the player's experiences and it concluded that: this process is rare and underdeveloped; that when developed further could produce a non-linear music model during play for a player [17]. Independent generation has been criticised, however, as machines currently not considered to be capable of thinking and therefore not capable of critical thought. Art is a subjective field, and requires a process AIs are not considered to possess [18]. Another study from 2018 has found as similar result, in that there is not yet an AI which can critically appraise music or an algorithm which can analyse a good piece [19] and must rely on a set of hard-coded rules set by a human. This is one of the gaps in data which this study aims to take a step towards filling.

C. State of Genetic Algorithms

Genetic algorithms are a type of evolutionary algorithm. Its character is attributed to the process of natural selection; an iterative method composed of selecting the best solutions and generating upon them [20]. These characteristics led to common use in optimisation problems, where resource use could be too intensive, such as The Travelling Salesman Problem. It reduces the need for many stages in brute-force testing [21]. Genetic algorithms, however, have some criticisms: the absolute best solution may not be found; the algorithm may also tend towards the wrong solution if it has marginally better performance than the other solutions in the right direction [22]. Genetic algorithms have a place in video game content generation. Such methods are useful in producing similar yet distinct level designs for games and often used at run-time to produce an evolving experience for the player [23].

D. Procedural Generation in Games

Procedural generation is an excellent example of AI generated content that is used in a games setting. It can be used as an offline or online tool for producing novel content for each player with minimal design input [24]. Such techniques are flexible, and can often be joined with other AI methods to produce new and interesting results; these tend to be more suitable for online generation due to their ability to adapt [25]. Each procedural generation function is usually bespoke - it must be tailored to each studio and often requires extensive research to achieve successful results [26]. Nevertheless, AI generated content is a powerful tool in game production with a solid place in the industry. Procedural music is much less common, but the clear pattern of increased use in procedural techniques suggests that an increase in interest is also emerging [27].

III. Research Questions

A. Do shorter intervals make up a better monophonic melody?

Procedural generation relies on a set of rules and processes to achieve a desired result. This set of rules may be set by machine learning or by a person. Machine learning generates its own understanding of how to achieve a desired result by analysing a large data set. Such a data set is difficult to acquire, and so the focus will be on finding a possible set of rules that could be used as a reasonable starting point for an independent evaluation function.

This study will investigate how a melody's intervals affect how appealing it is for listeners, and how that data can be applied. This will collect a viable criteria for generating melodies for algorithms by having participants rate each melody that is produced. Unlike machine learning from a series of examples, this would give the algorithm a degree of novel creation and immediate user input. By analysing the data from melodies created by users via the use of the artefact, it would provide an insight into what makes an effective melody in a vacuum for use in procedural generation.

B. Is there a particular ratio of notes to rests that is effective for melodies?

Rhythm is a critical part of music, and it is often created by varying the duration of notes in specific patterns. Rhythm is also supplemented by having moments of silence in the melody. An effective use of a rest is just as powerful as a strong swell in music [28]. This statement is only true if there is nuanced thought behind the decision to include a rest, but the music generation algorithm isn't capable of knowing the correct place of placing silence. This research question is posed to figure out if there is a specific proportion of rests which benefits the music, as it is an important part of evaluating a melody.

Using a similar process for the intervals question, this study will analyse the number of rests within the melody and plot it against the rating of the melody. The intervals are ignored for this test, as they do not have an effect on the rests.

C. Hypotheses

1) There is a positive relationship between short music intervals and the quality of a melody: The structure of a melody can be represented as a set of intervals - the distances between notes. The study focuses on these intervals as a way to measure the effectiveness of a melody. The data collected from the experiment provided a large

Hypothesis Number	H0	H1	Statistical Test
1	There is no relationship between the length of intervals and the quality of a melody	There is a positive relationship between the length of intervals and the quality of a melody	Spearman's R test
2	There is no relationship between the amount of rests in a melody and its quality	There is a positive relationship between the amount of rests in a melody and its quality	Spearman's R test

Fig. 1. Hypothesis table

quantity of melodies and ratings to correlate a relationship between intervals and fitness. It is expected that shorter intervals would make a better melody, as it would be easier to sing and have a more natural flow. Wang et al [29] suggest that shorter steps between notes leads to a more effective melody, as it leads to a more natural melody for singing.

The chosen statistical test for this was Spearman's R test, which would gauge the correlation between the number of given intervals and their corresponding rating. It was selected because it accepts the discrete ordinal predictor variable in this data set and can assess the relationship between two variables. For the purpose of simplifying the process of testing this hypothesis, the intervals have been grouped into groups of short (less than a tritone/half an octave) and long (tritone/half an octave or more) intervals. The desired p-value for accepting this hypothesis would be 0.05 or lower to indicate strong enough evidence.

2) There is a relationship between musical quality and the proportion of silence within it: A melody is composed of several beats, some of which are notes and some of which are rests - moments of silence. The number of rests within a melody can be plotted against its rating received to gain an insight into the relationship between silence and quality. Much like intervals, rests are a component of rhythm as they can change the way the beat is delivered in the melody. It can produce a syncopated feeling if it breaks up the notes in a poly-rhythmic way.

Spearman's R test has been chosen again for this, as we are comparing ordinal discrete data against more discrete data and want to learn of the relationship between them. A p-value of 0.05 is desired for the purpose of establishing evidence to support or refute the hypothesis.

IV. Methodology

The product is based on John Holland's genetic method [30], where chromosomes exchange genes to produce new items. Using the two-point genetic merge function, the melodies would be changed each generation, whilst maintaining some character from before. The experiments were all practical sessions involving participants to gather data. The genetic algorithm approach was chosen as it is useful as an evolutionary approach to music composition; allowing participants to develop melodies over a series of generations. This leads to a set of related melodies that users can enjoy the growth of in a manner that mimics melodic improvisation.

A. Agile Methodology

This study loosely worked as an Agile project; much of the work was divided into features and was added in specific phases. A time-bound plan was not solidly constructed, as there were gaps in the development from resource division. However, the primary features of this product were delivered in sprints. Despite not being part of a team project, the flexibility of Agile benefited the project when gaps in development were encountered as the incremental delivery and testing of features would prevent stagnation and the overall development cycle was managed more effectively [31]. The study was expected to change over time as new discoveries were made and work escaped scope, so this development lifecycle was also chosen as a measure to prevent the loss of efficiency as the study took form over time, with features gaining or losing priority.

B. Experimental Design

The purpose of this experiment was to collect data on the traits of a good melody. Letting participants have free use of the artefact will provide honest data because they are trying to achieve a result which is pleasing, rather than having to rate elements of the product. The process collected observational data in the form of integer arrays from continued generation of melodies without being intrusive for the participants, who would provide ratings for each melody in the form of integers from 0-100. The method proved to be engaging for participants, as they enjoyed trying to make better and better melodies. The experiment was designed to spur creative thought in participants, but there was not an expectation of any strong interaction between users and the artefact. Those participants who engaged the most provided more data than those who engaged less, and also enquired most into the research.

C. Artefact

The artefact[1] is an open-ended product designed to produce or assist in musical composition. The user can listen to melodies generated by the algorithm and assign to them a numbered rating. When the user regenerates the melodies, the assigned ratings are the genetic fitness for the purpose of breeding. Participants very quickly understood the product, and engaged well with it.

[1]Artefact repository available here: https://github.falmouth.ac.uk/MG246981/MG246981-Dissertation

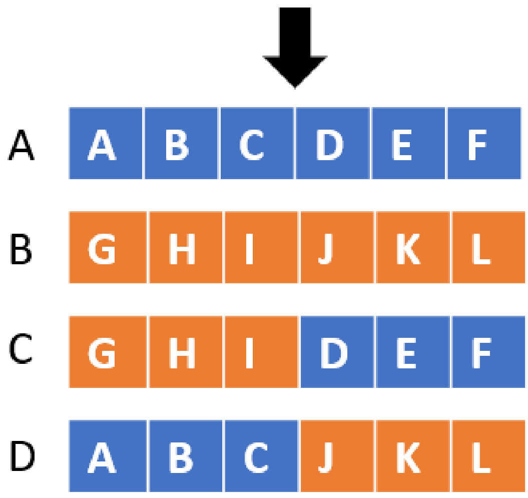

Fig. 2. Example of a one-point merge

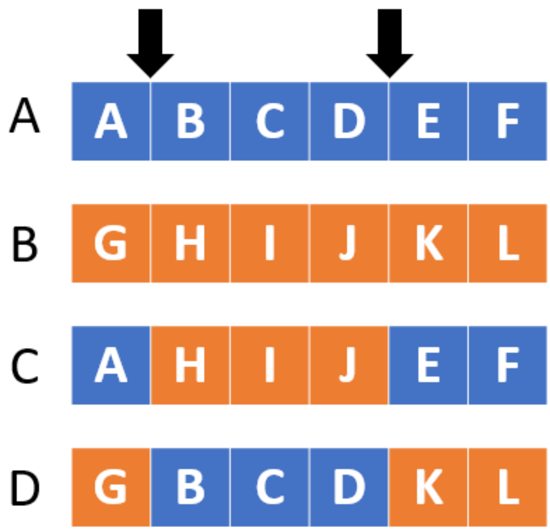

Fig. 3. Example of a two-point merge

The generation algorithm has an innate set of rules that it must use to produce a basic melody and avoid atonality. Its set of notes will only ever come from a configurable music key and its duration will never exceed two bars. This ensures that the first melody is, at a minimum, clean. If the notes were completely random pitches, it would lose sense and character. The tempo of the melodies can also be adjusted to change the character of the melody. For the duration of the experiments, the melodies were locked to the key of C major and 120 BPM (beats per minute), so that every test was controlled and does not discredit the experiment conditions.

Each melody begins as an integer array, stored in button object which can play back the melodies for users. Integer arrays were the choice to store the body of the melody as they would be much easier to manipulate with a genetic algorithm and simpler to store. Maestro - Midi Player Toolkit does not accept integer arrays for playback, so a function with an adapter design pattern [32] was produced to interpret it into a format which was usable by the playback tool.

In use, this artefact presented six melodies with corresponding sliders. These sliders would be used to rate the melody from 0 to 100 by the user to give it a fitness rating. Figure 3 shows an example of the user interface, with three melodies and corresponding sliders. At start, six random melodies are generated, and depending on the given rating, will be paired with another and bred. The breeding is done by John Holland's method, which involves one-point and two-point merges.

A slight bias may have been introduced as participants were expecting better results each generation, and such may falsely believe their later generations were better. Very few participants chose to create a new generation after listening to the initial population, and would often develop the melodies from low scores (0-20) to higher scores (50-80) when using the artefact. Participants would also report, however, that the quality of their melodies would drift after extended use of the product.

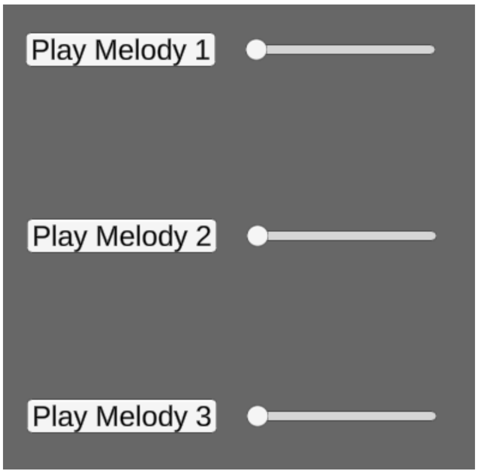

Fig. 4. Example of user interface

D. QA and Testing

This study operated with a V-shaped testing model [33]. This led to a sequential process of designing and implementing features alongside testing in an Agile manner. This model was selected as the thorough testing and clear definition of requirements was beneficial to the way the study was conducted. The design was abstracted first and checked against the requirements of the study. If a function failed during the testing phase, it would return to the design phase, as problems within the architecture are found and resolved, typically with a refactor. Each primary function was tested on a per-feature basis, so that development would not continue until the new code would be sufficiently tested and functioning. The unit tests as shown in the appendices were performed by manual calculation and comparison.

An informal pilot study was conducted with some colleagues to gain an informal appraisal of the suitability of the prototype artefact for testing without collecting any data. Multiple flaws were found with the initial design and this led to refactoring and re-evaluating the experiment

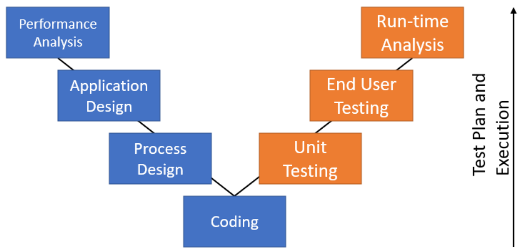

Fig. 5. The test plan model used for development

process. The non-deterministic nature of the algorithm led to a non-linear set of manual unit tests for the primary algorithm, and manual unit testing on the data collection and analysis functions. Once ethical approval was secured, some formal end-user testing was performed to ensure the artefact was deployable to a larger group of participants. Any input fields were either removed or changed for discrete inputs, so that erroneous values were impossible to input to improve the reliability of the artefact. By giving participants access to only that which is relevant in the artefact it removed any chance of accidental errors or unexpected problems arising.

E. Refactoring

The product has undergone multiple refactors, iterating upon the process to improve its capabilities and optimisation. The first attempt at the artefact was unsuccessful; the attempt was to produce the entire product and capabilities from scratch was too large a task in the space of time that was available. The second attempt was logged in the repository, and used Maestro - Midi Player Toolkit[2] to sequence and produce sound. This package was a challenge to implement with the algorithm, as the data structure used for the notes was not well suited for the genetic function of the product. This method of sequencing and playing relied on a hard-coded set of notes and beats, and was played in the Update() function in Unity. The first refactor had the set of notes be selected from an initial note, relying on the structure of a traditional natural major or minor key [34]. The first set of genetic methods was also implemented, and the product was usable outside of a testing environment. It still had several issues with playback and the architecture of the algorithm. The final refactor saw the playback be reworked into coroutines, which improved the potential of the artefact as it could more robustly handle and play back melodies. The capability of multiple notes became available. As part of the final refactor, the artefact environment limited user input to very discrete options relevant only to the experiment.

F. Data Collection

The data collected from the first experiment is a usage log of user melodies. It is headed by the user ID and the date, then followed by the saved melodies and their respective rating. The participants are aware that the process is there, but it is not intrusive and acts as a ”black box” for the user experience. This timeline of melodies was then read by a function which could arrange the data in a format which would make it possible to run statistical tests in R. These formats were the rests within the melodies and the intervals which made up the melodies. The bespoke saving function stores the data in a way that is easy to understand when glanced at even without additional data, and provides a quick expectation of preliminary results before more sophisticated data analysis.

The power analysis for this project was calculated through GPower 3.1, and was used to calculate sample size. Cohen’s effect size for Spearman’s r test was taken as 0.3 - a medium-size effect in GPower [35], [36]. With a value of $\alpha = 0.05$ and $\beta = 0.95$, this gave a result of 115 in total sample size. The total sample of participants was 24, and each participant gave a differing number of observations. The total number of observations recorded was 624. A larger group of participants was desired, but an online deployment of this artefact would have required a secure method for data collection which was out of scope for resources and time.

G. Research Philosophy

This study took a positivist approach to the task. The experiments were observational in nature; by allowing participants to produce the melodies that they enjoyed, it permitted this study to observe and analyse only what has happened in an objective way. By applying a mathematical approach to the music found in this study, it became possible to analyse patterns and possibly predict how people appraise the appeal of a melody line. The quantitative data allowed for a more grounded approach to the hypothesis testing for the purposes of this study.

V. Data Analysis

A. Hypothesis 1

The statistical process for the hypothesis *There is a positive relationship between short music intervals and the quality of a melody* involved a careful distribution of data. The collection of short intervals (less than a tritone) and long intervals (tritone or longer) are plotted against the rating they received. The grouping of short intervals and long intervals ensures greater reliability as it reduces the group sizes significantly. Major and minor variants of intervals exist [37], but it is important to mitigate the large number of possible groups in the data to reduce the spread of data. This measures the quality of melodies with groups of a particular proportion of intervals.

For hypothesis 1, the intervals must by collected from each melody. Each melody is stored as an array of integers with length 16. The difference between two notes is the distance in semitones and that can then be interpreted as an interval. Rests are ignored during this process. The interval occurrences are the independent variable, and the rating is the dependent variable.

B. Results

From the first experiment, all short intervals and long intervals were compiled into groups and plotted them against

[2]Maestro - Midi Player Toolkit available here: https://assetstore.unity.com/packages/tools/audio/maestro-midi-player-tool-kit-free-107994

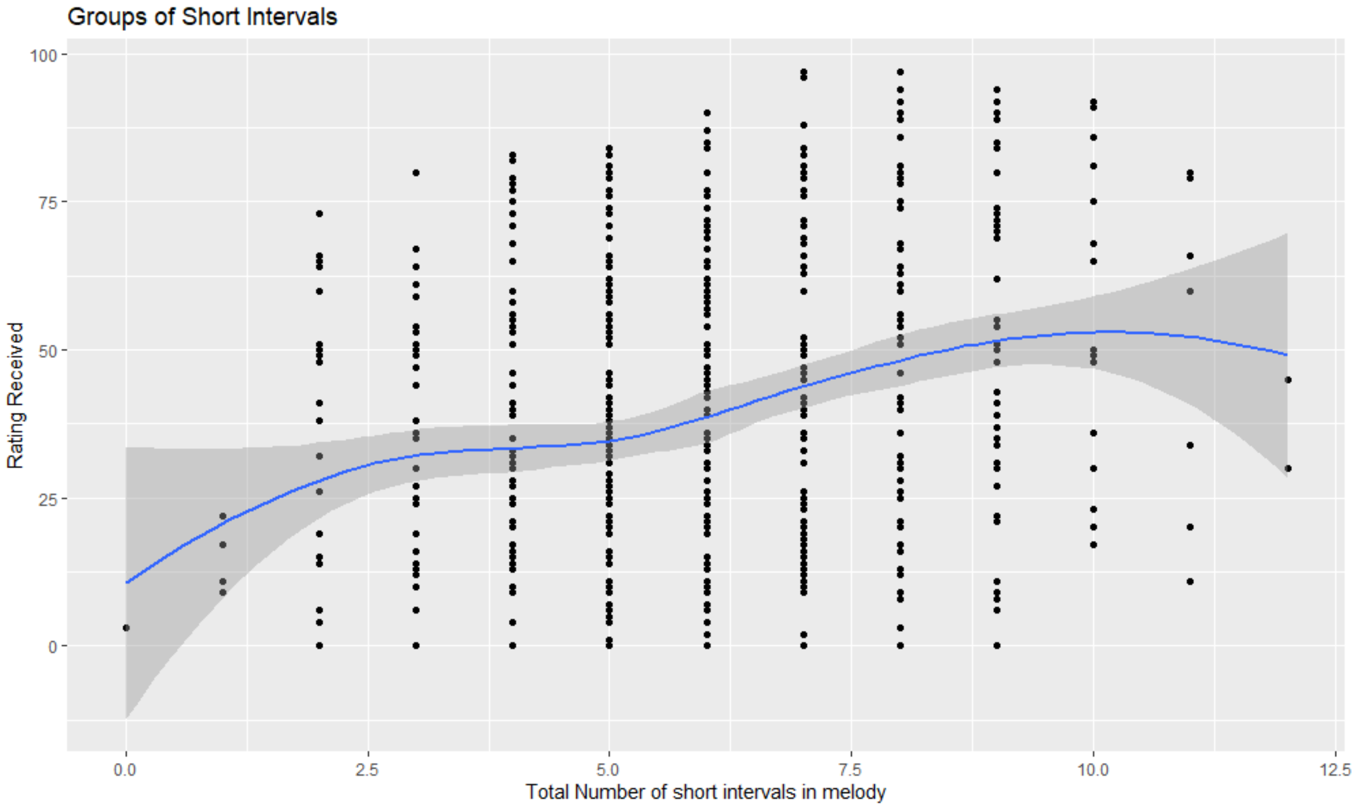

Fig. 6. The group of short intervals plotted against rating received

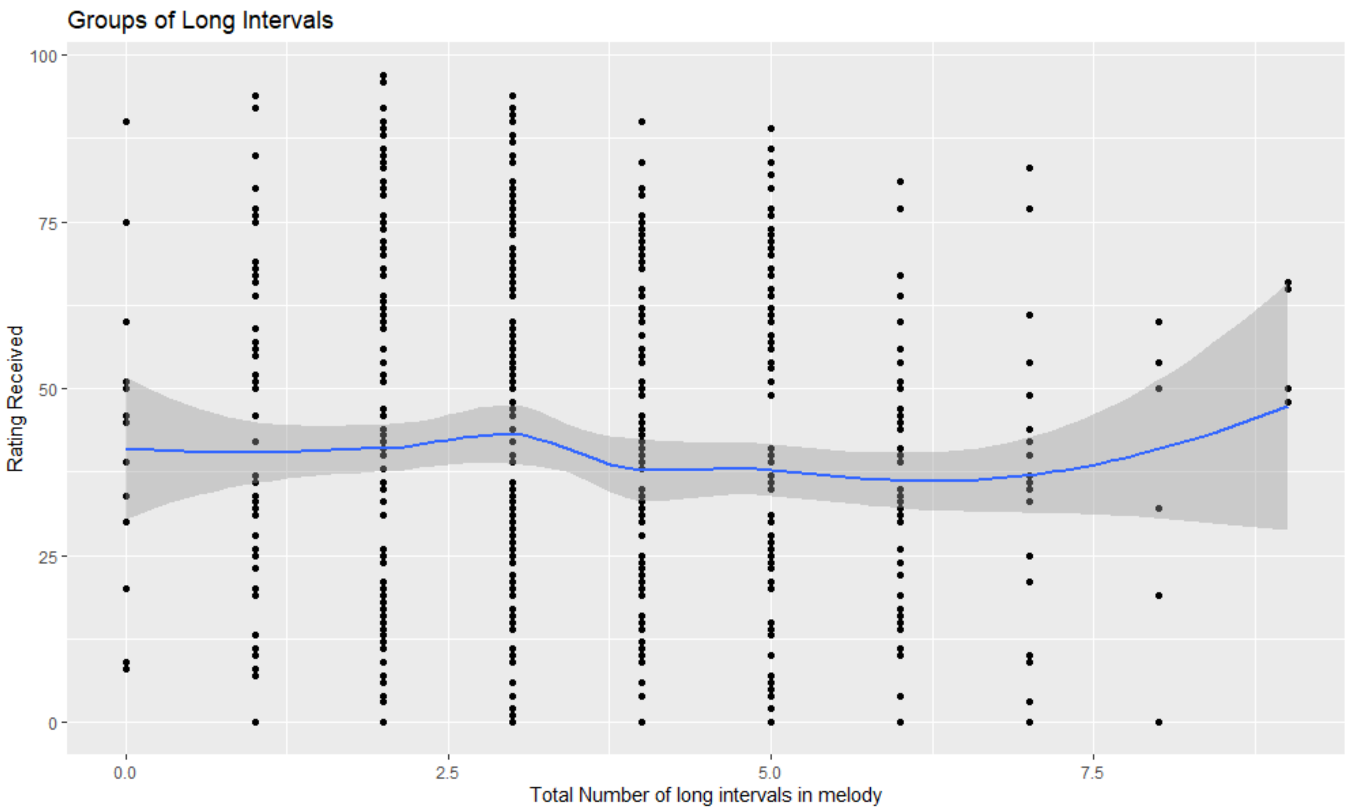

Fig. 7. The group of long intervals plotted against rating received

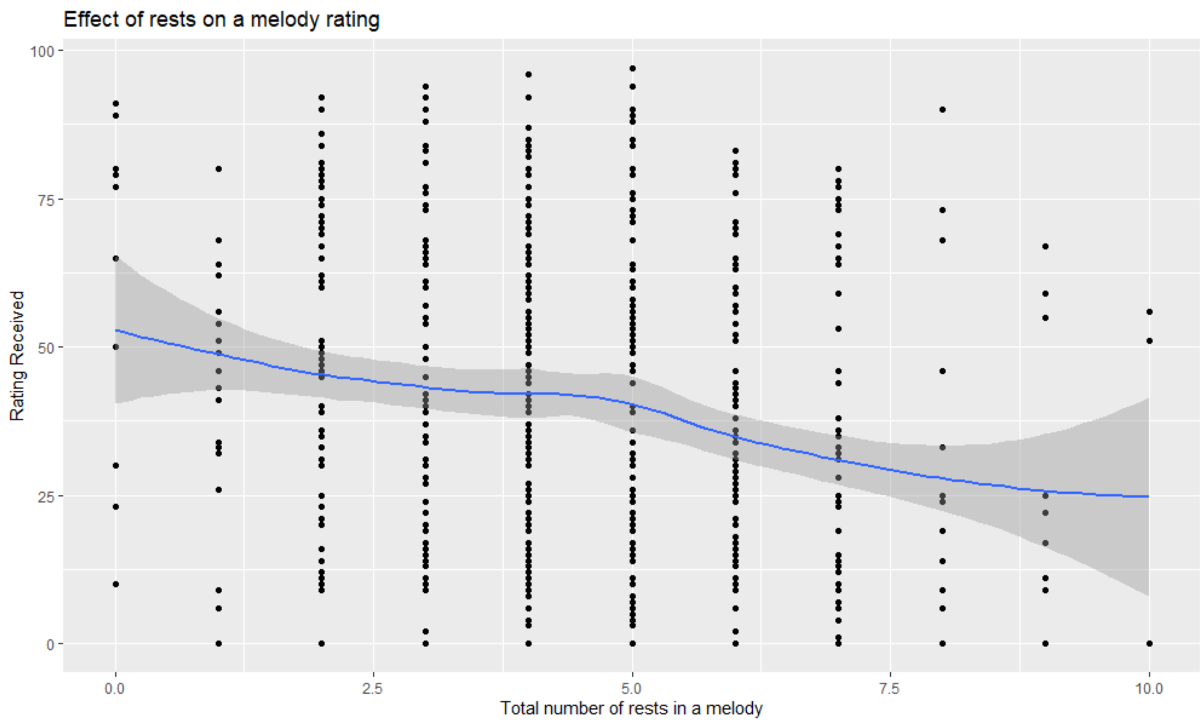

Fig. 8. The total number of rests per melody plotted against rating received

rating. With a LOESS line of best fit, a slight relationship is visible between the number of short intervals and the rating received in figure 6. Using Spearman's r test, the short intervals graph gave a p-value of 4.65e-10, and an r value of 0.246. The long intervals graph gave a p-value of 0.401, and an r value of -0.0337.

C. *Hypothesis 2*

The analysis process for the hypothesis *There is a positive relationship between the amount of rests in a melody and its quality* was similar to the one used in hypothesis 1. It was simpler as it only involved one group of data, and the relationship was surprisingly strong and in the opposite direction to what was expected as seen in figure 8. When tested for Spearman's R, the p-value was given as 2.28e-07, and the r-value was given as -0.205. The LOESS line of best fit shows a strong downward relationship and the low p-value suggests strong evidence to the fact that there is a negative relationship between the number of rests and the quality of a melody. With the evidence leaning far against the hypothesis, it must be refuted, as it is much more likely that there is a negative relationship.

VI. Discussion

A. *Findings of intervals*

The experiment for hypothesis 1 produced two scatter plots, see figure 6 and 7, which model a scatter plot of the sums of short and long intervals. The LOESS line of best fit shows that on figure 6, there is a slight positive relationship between the total occurrences of short intervals within a melody and its rating. The gradient coefficient is 0.246, which suggests a small to medium effect relationship. The p-value was incredibly small at 4.65e-10, which suggests very strong evidence that there is a non-zero relationship between short intervals and the quality of the melody. The strength of this relationship is about the level which was expected - the steps in a melody are only a single factor in the total success of a melody. The main components of a melody are: rhythm, pitch, timbre, texture and dynamics [38], and the findings from this experiment fall under pitch in that regard. This is not to say that a large number of short steps in the melody always tends towards the best fitness, as a melody would vary by the tone which it is trying to achieve, especially where top-quality music is required. The rating variable in this experiment only illustrates a participant's enjoyment of listening to the melody, but does not illustrate its suitability in other criteria, such as tone or setting.

Figure 7 finds that the number of long intervals has a very weak relationship. The high p-value of 0.401 indicates very weak evidence towards a relationship between the amount of long intervals and the quality of the melody. This was an unexpected result, as the number of long intervals is linked to the number of short intervals too. The graphs were expected to have opposite relationships, but that did not occur. The low average score throughout the graph, however, was expected in the long interval groups, as most effective melodies should be sung comfortably [29]. Melodies with higher jumps require greater technical skill to land effectively [39] so it would not fit one of the conditions for an appealing melody. There is, however, a slight increase in melodies where three occurrences of long intervals were found, suggesting that there may be an optimal amount of long intervals within a two-bar melody, but it is not possible to determine conclusively without a better data set.

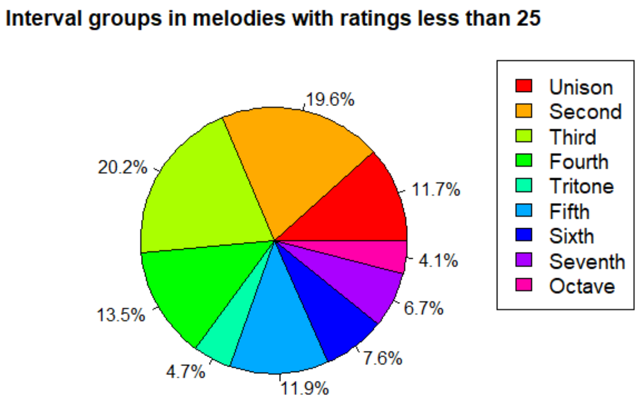

Fig. 9. Interval groups in the lower rating quartile

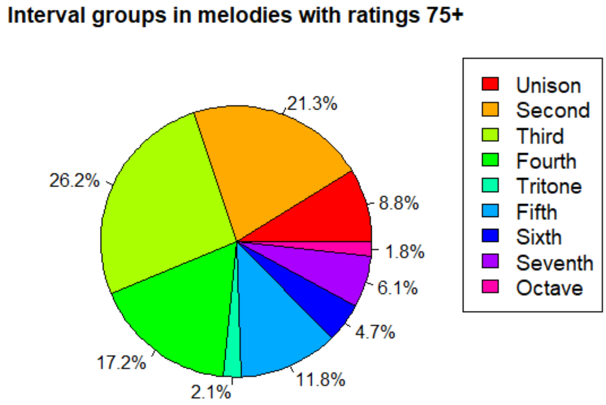

Fig. 10. Interval groups in the upper rating quartile

Spearman's R test is also not completely ideal, as it is non parametric and does not deal well with repeat values. This weakens the evidence it provides somewhat, but the supplementary graphs and pie charts should help establish the relationship between the intervals and the appeal of the melodies.

In figure 9 and figure 10, we can see the proportions of occurrences of interval groups in the top and bottom 25% of the melody groups. The bottom quartile has 226 melodies, and the upper quartile has 79 melodies. In the bottom quartile, long intervals make up 35% of the total intervals found across the melodies. In the upper quartile, long intervals make up only 26.5% of the total intervals found in the melodies. This suggests that long intervals have a slight negative relationship with the appeal of a melody, as there is a 8.5% difference in their presence. Short intervals make up 65% of the total intervals found in the lower quartile of melodies, and 73.5% of the upper quartile. This can be attributed to the fact that they are the most common occurrence due to the random generation process of the artefact, but the positive relationship is in line with the statistical test from earlier. Unfortunately the data sets have a large difference in samples, making this comparison weak.

With the evidence found towards the effect of shorter intervals on melodies, it becomes clear that hypothesis 1 can be accepted with a decent level of confidence. While the sample size is small and constrained by location and resources, the amount of observational data was large and the p-value was low enough to provide strong enough evidence. The visibility of the effect suggests that machines may be able to collect a data set with which to train themselves to appraise melodies much like people, supporting the future works of Lopez-Rincon et al [19] as the data found here could model how a machine understands a melody.

B. *Findings of rests*

In figure 8, we can see that there is a clear negative correlation between the number of rests within a melody. Whilst this led to the refutal of hypothesis 2, it shows that there is a general consensus when judging a monophonic melody and its proportion of rests. This further supports the idea of a tangible data set used by a machine to mimic human judgement of a melody and grant it an overall rating when paired with other data sets representing desirable and undesirable melodic traits. The negative relationship could also be attributed to the fact that the artefact places the rests randomly on the first generation, so the chance of a truly impactful rest is quite slim. It may not be entirely representative of the highest quality of melodies.

C. *Potential applications*

Using the data collected from this study to construct such a function would be flawed, as the data is incredibly bespoke and narrow for this study's purposes. However, the potential applications of a method similar to this one could provide a large enough data set when distributed amongst users and made into an gamified product. Such an artefact could be a game where users compose their own or generate random melodies, and these would be exchanged with other users who may rate them from all around the world. This would enable a large data set on the desirable traits of music to begin being collected. Once completed, an evaluation function using this data set could provide indie developers with the option of having appealing music be composed for their game without the large investment of an audio specialist. The music would not be especially powerful, but it would at least be pleasant to listen to by a larger audience and fill the gaps which would otherwise be replaced by less effective melodies.

VII. Ethical Considerations

A. *Informed Consent*

The safety and rights of participants was a priority when conducting this study. Participants must be fully aware of the research they are participating in and that their participation is completely voluntary, as set out in the Declaration of Helsinki and Nuremberg Code [40], [41]. The information sheet produced for participants had a clear description of the research, the purpose of the research and what their participation meant. It described what data was gathered and for what it would be used for. The form also detailed how participants would provide consent and how to withdraw from the study.

The consent form was a part of the artefact, which prevented any use unless the participant has understood and agreed to the information sheet and had the opportunity to ask questions, agreed that they understand they may withdraw from the study at any time without giving a reason, and that they agreed to take part in the study. The user must tick all the boxes before the use of the artefact is unlocked.

B. Data Handling and Withdrawal

The data collection process was designed to reduce any potential risk of confidentiality breaches. Users were never asked to share any contact information. Upon opening the artefact, the user is given a unique 8-character hash which they are advised to write down for their record. No identifying or personal data is stored, so this hash is required for anonymous differentiation of data. Should a participant want their data destroyed, they may contact the researcher or Falmouth university research team requesting their withdrawal and provide their hash. All data associated with that individual would then be destroyed immediately. All data which was collected was saved locally to the device with the artefact, but transferred personally by the researcher onto a secure Microsoft OneDrive under the University domain. No personal data was collected, and all other data was stored securely on GDPR-compliant cloud storage, so all GDPR regulations [42] were followed for the duration of this study.

C. Ethical Review Process

In order to secure ethics approval, this study underwent the medium risk ethics assessment by Falmouth University. The project was categorised and described in detail on its goals how it was going to be conducted. Risks to health, safety and reputation were assessed and listed explicitly with methods to mitigate any risks to participants, the research team and the university. There was a small chance of eye, ear and wrist strain, so the product was designed to be as safe as possible to use, with user-friendly design, dull colours and was prepared by the research team to have low volume. It was stated that participants could take a break at any time should any discomfort arise. The information and consent forms were developed in line with Falmouth University's code of conduct, then reviewed and approved.

D. Ethical Standard

In accordance with Falmouth University's Research and Knowledge Exchange integrity and ethics policy [43], this research has:

- Maintained transparency of data use with participants
- Underwent thorough ethical risk and ethics reviews
- Conformed to ethical, legal and professional obligations

The British Computer Society's code of conduct [44] was guide used for maintaining a high ethical and professional standard. To maintain a valuable contribution to computer science, this research has:

- Worked in the public interest
- Worked with professional competence and integrity
- Worked with duty in accordance with Falmouth University
- Upheld reputation and duty of computer science

E. Ethical Criticisms of AI Art

There have been debates on AI replacing human-sourced jobs and how it effectively removes the need for some professions [45]. Journalism has been influenced by the invention of AI news-writing algorithms. These algorithms are currently not much more than an assistance tool for journalists, but with constant developments on AI, their performance is improving rapidly. The efficiency and low cost of AI makes it a prime candidate for replacing much more expensive human workers in this field. AI is already superior to people in many tasks that involve logic or thinking with strict rules: such is the case especially with chess bots [46]. AI was put to the test by having it play against chess champion Garry Kasparov in 1997, with only a marginal success. It didn't take long for chess AI computing power to completely overtake human skill only some years later [47]. Despite AI being much better than people at playing chess, the community has remained very active and measures exist which prevent chess engines from being exploited for competitive success. Human ability has not been made obsolete despite initial objections from critics [48]. The objective of this research is not to replace any human jobs or outdo human ability but rather supplement and stimulate creative potential.

VIII. Limitations

This study had resource limitations when conducting experiments with participants. The population of this study is biased because the participants all came from the same location. Contacting external participants with unbiased sampling methods was too difficult to achieve, so the bias was not mitigated. The study also had a resource division limitation as it was conducted alongside other large projects. This constrained the complexity of the artefact and the experiment procedures. Much of the intended work that was cut is mentioned in the future work section, as it may be a worthwhile venture in future studies.

The music generation artefact was also limited in sound-playing potential as the free version of Maestro - Midi Player Tool Kit has limited features. The lower-quality MIDI sound banks which reduced the value of the second experiment, due to the mechanical sound of the notes. Ideally a more realistic MIDI sound bank would have been used to play back the melodies so that the playback wasn't dry and better reflected the intended conditions for these melodies.

IX. Future Work

A. Difference between human-composed and machine-composed music

The human character of music is unique and difficult to replicate in most AI generated mediums. Hong et al [49] suggests that presenting computer-generated music with humanlike qualities is a primary factor in the perception of its melodies, as that study found that those who accepted the AI as a musician enjoyed its music more than those who did not. This study originally wanted to also understand the difference in perception between human-composed and machine-composed music through a blind experiment where participants are asked to identify the melody as human or AI. This would help to understand what traits in melodies help make believable music, rather than just a pleasant tune which is enjoyable to listen to. It would also provide evidence towards the viability of an evaluation function as proposed earlier in this paper. The human character is also critical to the enjoyment of music, as the nature of people is imperfect and causes music to have its beauty, and replicating the mistakes or differences in timbre are just as important in producing a strong melody [50].

B. *Application of a prototype fitness function*

As proposed in the discussion and application section, a potential follow-up study could involve the creation of an evaluation algorithm relying on wider data set of melodies appraised by participants. This would help identify the true value of such a product for game developers if it could be tested within development environments or live game sessions. Pushing the boundaries of what is available as a tool for game development is a valuable and worthwhile contribution to the industry.

X. Conclusion

Despite the evidence being weak in some regards to the hypotheses presented in this study, there is a clear relationship between the appeal of a melody and some of the data produced in the experiment. The data set showing clear relationships from the experiment has provided an image of what could be expected if this method is applied when seeking a data set for an evaluation function. This study has found that monophonic melodies tend to be more effective when shorter intervals make up the majority of the steps in the melody, falling in line with other studies finding similar traits in effective melodies. However, the opposite was found for rests, and that the higher the proportion of silence within the melody, the less appealing it was to listeners. Whilst the data found in this study should not be used for such a function in its current state, it shows what such data may look like. This provides direction for a future iteration of such an artefact for more effective generation of melodies; one which excludes results which would be very unlikely to be successful, whilst encouraging candidates with a higher chance of success. Despite not being designed as an entertainment product, many participants found the experience engaging and involving, so a game with similar processes to this project could gather valuable data from a larger quantity of users.

XI. Acknowledgements

I would like to thank Matt Watkins, Staff Lecturer for Computing at Falmouth University for being my supervisor during this study. His guidance and support was incredibly valuable to this project and helped give direction when it was difficult to find.

I would also like to thank Dr Michael Scott, Head of Computing and Associate Professor at Falmouth University for his help and support in this research and brilliantly teaching this module.

I also present a great deal of gratitude to Liz Coulter-Smith, Staff Lecturer in Computing; Victoria Rees, Academic Liaison Librarian; Dr Joseph Walton-Rivers, Staff Lecturer in Computing and John Speakman, Staff Lecturer in Computer Graphics for offering up their time to provide me with invaluable feedback and advice at the time of writing this paper.

References

[1] J. P. Fernández-Cortés, "Ludomusicology," *Journal of Sound and Music in Games*, vol. 2, no. 4, pp. 13–35, 2021. DOI: 10.1525/jsmg.2021.2.4.13.

[2] B. Sinclair, *Game audio increasingly outsourced*, Sep. 2014. [Online]. Available: https://www.gamesindustry.biz/game-audio-increasingly-outsourced.

[3] *Ai can create!* Accessed December 2, 2022. [Online]. Available: https://reports.mintel.com/trends/#/observation/1119599?fromSearch=?freetext=AI%20AND%20Art&resultPosition=3.

[4] *Ai wrote this song*, Accessed December 2, 2022. [Online]. Available: https://reports.mintel.com/trends/#/observation/1034117?fromSearch=?freetext=AI%20AND%20Art&resultPosition=14.

[5] M. Civit, J. Civit-Masot, F. Cuadrado, and M. J. Escalona, "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends," *Expert Systems with Applications*, vol. 209, p. 118 190, Dec. 2022. DOI: 10.1016/j.eswa.2022.118190.

[6] P. Suthaphan, V. Boonrod, N. Kumyaito, and K. Tamee, "Music generator for elderly using deep learning," in *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, IEEE, Mar. 2021. DOI: 10.1109/ectidamtncon51128.2021.9425691.

[7] A. Y. H. Yap, H.-C. Soong, and S. S. H. Tse, "Real-time evolutionary music composition using JFUGUE and genetic algorithm," in *2021 IEEE 19th Student Conference on Research and Development (SCOReD)*, IEEE, Nov. 2021. DOI: 10.1109/scored53546.2021.9652786.

[8] L. R. de Azevedo Santos, C. N. S. Jr., and M. D. Costa-Abreu, "A methodology for procedural piano music composition with mood templates using genetic algorithms," in *11th International Conference of Pattern Recognition Systems (ICPRS 2021)*, Institution of Engineering and Technology, 2021. DOI: 10.1049/icp.2021.1435.

[9] A. Liu, J. Wang, J. Peng, *et al.*, "Composer4everyone: Automatic music generation with audio motif," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, Mar. 2019. DOI: 10.1109/mipr.2019.00101.

[10] J. Chan, J. J. Daza, W. Kwan, and A. Basu, "Facilitating player progression by implementing procedural music in videogames," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, Oct. 2017. DOI: 10.1109/smc.2017.8122969.

[11] Ł. Mazurowski, "Computer models for algorithmic music composition," in *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2012, pp. 733–737.

[12] P. Laine and M. Kuuskankare, "Genetic algorithms in musical style oriented generation," in *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*, IEEE. DOI: 10.1109/icec.1994.349942.

[13] P. Sheikholharam and M. Teshnehlab, "Music composition using combination of genetic algorithms and kohonen grammar," in *2008 International Symposium on Computational Intelligence and Design*, IEEE, Oct. 2008. DOI: 10.1109/iscid.2008.73.

[14] S. D. You and P.-S. Liu, "Automatic chord generation system using basic music theory and genetic algorithm," in *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, IEEE, May 2016. DOI: 10.1109/icce-tw.2016.7520919.

[15] N. Masuda and H. Iba, "Musical composition by interactive evolutionary computation and latent space modeling," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, Oct. 2018. DOI: 10.1109/smc.2018.00476.

[16] F. Mo, X. Wang, S. Li, and H. Qian, "A music generation model for robotic composers," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, Dec. 2018. DOI: 10.1109/robio.2018.8665078.

[17] D. Plans and D. Morelli, "Experience-driven procedural music generation for games," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 3, pp. 192–198, Sep. 2012. DOI: 10.1109/tciaig.2012.2212899.

[18] A. Hertzmann, "Can computers create art?" *Arts*, vol. 7, no. 2, p. 18, May 2018. DOI: 10.3390/arts7020018.

[19] O. Lopez-Rincon, O. Starostenko, and G. A.-S. Martin, "Algoritmic music composition based on artificial intelligence: A survey," in *2018 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, IEEE, Feb. 2018. DOI: 10.1109/conielecomp.2018.8327197.

[20] M. Mitchell, *An Introduction To Genetic Algorithms*. MIT Press, Mar. 1998, ISBN: 9780262631853.

[21] P. Guo, X. Wang, and Y. Han, "The enhanced genetic algorithms for the optimization design," in *2010 3rd International Conference on Biomedical Engineering and Informatics*, IEEE, Oct. 2010. DOI: 10.1109/bmei.2010.5639829.

[22] W. Banzhaf, L. Spector, and L. Sheneman, *Genetic Programming Theory and Practice XVI*. Springer, p. 234, ISBN: 9783030047344.

[23] A. S. Kholimi, A. Hamdani, and L. Husniah, "Automatic game world generation for platformer games using genetic algorithm," in *2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, Oct. 2018. DOI: 10.1109/eecsi.2018.8752741.

[24] Y. Zhang, G. Zhang, and X. Huang, "A survey of procedural content generation for games," in *2022 International Conference on Culture-Oriented Science and Technology (CoST)*, IEEE, Aug. 2022. DOI: 10.1109/cost57098.2022.00046.

[25] M. Kaidan, C. Y. Chu, T. Harada, and R. Thawonmas, "Procedural generation of angry birds levels that adapt to the player's skills using genetic algorithm," in *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*, IEEE, Oct. 2015. DOI: 10.1109/gcce.2015.7398674.

[26] B. M. F. Viana and S. R. dos Santos, "A survey of procedural dungeon generation," in *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, IEEE, Oct. 2019. DOI: 10.1109/sbgames.2019.00015.

[27] K. Collins, "An introduction to procedural music in video games," *Contemporary Music Review*, vol. 28, no. 1, pp. 5–15, Feb. 2009. DOI: 10.1080/07494460802663983.

[28] Z. Lissa, "Aesthetic functions of silence and rests in music," *The Journal of Aesthetics and Art Criticism*, vol. 22, no. 4, p. 443, 1964. DOI: 10.2307/427936.

[29] Z. Wang, S. Zhang, and X. Chen, "Exploring inherent properties of the monophonic melody of songs," Mar. 2020. arXiv: 2003.09287 [cs.SD].

[30] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–73, 1992, ISSN: 00368733, 19467087. [Online]. Available: http://www.jstor.org/stable/24939139 (visited on 03/28/2023).

[31] N. B. Moe, A. Aurum, and T. Dybå, "Challenges of shared decision-making: A multiple case study of agile software development," *Information and Software Technology*, vol. 54, no. 8, pp. 853–865, Aug. 2012. DOI: 10.1016/j.infsof.2011.11.006.

[32] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns, Elements of Reusable Object-Oriented Software (Addison-Wesley Professional Computing Series)*. Addison-Wesley Professional, 1995, p. 395, ISBN: 9780201633610.

[33] Smitha.vn, *An introduction to v model in testing: Zyxware*, Jul. 2016. [Online]. Available: https://www.zyxware.com/articles/5119/an-introduction-to-v-model-in-testing.

[34] S. M. Kostka, D. Payne, and B. Almén, *Tonal Harmony With an Introduction to Post-Tonal Music, With an Introduction to Post-Tonal Music*. McGraw-Hill Education, 2017, p. 704, ISBN: 9781259447099.

[35] G. M. Sullivan and R. Feinn, “Using effect size—or why the p value is not enough,” *Journal of Graduate Medical Education*, vol. 4, no. 3, pp. 279–282, Sep. 2012. DOI: 10.4300/jgme-d-12-00156.1.

[36] H. Kang, “Sample size determination and power analysis using the g*power software,” *Journal of Educational Evaluation for Health Professions*, vol. 18, p. 17, Jul. 2021. DOI: 10.3352/jeehp.2021.18.17.

[37] D. Farrant, *A guide to music intervals: The gaps between the notes*, Oct. 2022. [Online]. Available: https://hellomusictheory.com/learn/intervals/.

[38] M. H. Wennerstrom, V. L. Kliewer, and R. P. Delone, *Aspects of twentieth century music*. Prentice Hall, May 1975.

[39] P. Q. Pfordresher and S. Brown, “Poor-pitch singing in the absence of ”tone deafness”,” *Music Perception*, vol. 25, no. 2, pp. 95–115, Dec. 2007. DOI: 10.1525/mp.2007.25.2.95.

[40] W. M. Association., “World medical association declaration of helsinki. ethical principles for medical research involving human subjects.,” *Bulletin of the World Health Organization*, vol. 79, no. 4, pp. 373–374, 2001.

[41] “The nuremberg code (1947),” *BMJ*, vol. 313, no. 7070, Ed., p. 1448, 1996. DOI: 10.1136/bmj.313.7070.1448. eprint: https://www.bmj.com/content/313/7070/1448.1. [Online]. Available: https://www.bmj.com/content/313/7070/1448.1.

[42] *Guide to the uk general data protection regulation (uk gdpr)*. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/.

[43] K. S. David Prior, *Research and knowledge exchange integrity and ethics policy*, Accessed December 6, 2022, Sep. 2021. [Online]. Available: https://www.falmouth.ac.uk/sites/default/files/media/downloads/Research%20Ethics%20_%20Integrity%20Policy.pdf.

[44] T. C. I. f. I. BCS, *Bcs code of conduct*, last reviewed June 8, 2022, Jan. 1996. [Online]. Available: https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf.

[45] A. Miroshnichenko, “AI to bypass creativity. will robots replace journalists? (the answer is “yes”),” *Information*, vol. 9, no. 7, p. 183, Jul. 2018. DOI: 10.3390/info9070183.

[46] D. De Cremer and G. Kasparov, “Ai should augment human intelligence, not replace it,” *Harvard Business Review*, vol. 18, 2021.

[47] D. M. D. Iliescu, “The impact of artificial intelligence on the chess world,” *JMIR Serious Games*, vol. 8, no. 4, e24049, Dec. 2020. DOI: 10.2196/24049.

[48] D. Waldstein, *Chess thrives online despite pandemic*, May 2020. [Online]. Available: https://www.nytimes.com/2020/05/08/sports/coronavirus-chess-online-tournament.html.

[49] J.-W. Hong, K. Fischer, Y. Ha, and Y. Zeng, “Human, i wrote a song for you: An experiment testing the influence of machines’ attributes on the ai-composed music evaluation,” *Computers in Human Behavior*, vol. 131, p. 107 239, 2022, ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2022.107239. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563222000619.

[50] P. Townsend, “241Imperfections in music,” in *The Power of Imperfections: A Key to Technology, Love, Life and Survival*, Oxford University Press, Jan. 2022, ISBN: 9780192857477. DOI: 10.1093/oso/9780192857477.003.0015. eprint: https://academic.oup.com/book/0/chapter/337720666/chapter-pdf/42765996/oso-9780192857477-chapter-15.pdf. [Online]. Available: https://doi.org/10.1093/oso/9780192857477.003.0015.

Appendix A
Personal Reflection

This task was one of my largest academic undertakings. The process of having to think of research questions to answer and then design and deliver an experiment to collect data for analysis and write-up proved to be a stimulating challenge. The project was a fascinating task for me and I felt proud that I could delve into original research in an area which is less known. I found that there was integrity and value to the experiments conducted by me. I learned many useful lessons over the course of this project, and could see that there was always something to improve on at every stage of the study. The artefact was well received by my peers and it garnered some interest from those who were not musically or computer science minded. I did face several weaknesses during this project which became very clear: time management, experimental design and test planning.

A. Resource Division

One of the greatest challenges I had with this project was having to balance it alongside other projects. My group game project took up a great deal of my time as I took on many responsibilities throughout that project, applying myself to a great deal of parts within that project. The time I wished to spend on this project would often get funneled into one of the tasks required for the group game project. The work on that project was also much more inviting, as there was a greater variety of work which was more creatively stimulating; this would often draw me away from wanting to work on the study. After having to offset a lot of work over the weeks leading up to the deadline, I learned how important it was to properly keep time and divide it. After having timetabled my days, I found that my time was being spent more efficiently when working on this. I understand now how critical it is to plan a solid timeline of ”soft deadlines” from the beginning of a project to work more consistently over a longer period.

B. Experimental Design

This project would have benefited from more robust supplementary qualitative data. The subject of music is deeply subjective, and questionnaires to gain a worded insight into participants' opinions would have offered a valuable set of additional information which could have been used to supplement the data provided here. The data collected from the experiment was unfortunately limited to an extent as it was only a collection of two-bar melodies rated from 0-100, and was locked to the key of C major, 120BPM. The design could have been improved by having the algorithm favour specific numbers of intervals over a longer time to produce a better selection of data which would have be more representative of the intervals found in melodies.

C. Over-scoped experiment

I found myself wanting my experiment to work, without necessarily having it be fully supported by similar studies. The artefact was a culmination of my work for this specific subject on which I've worked on for months. The original plan I had for the project was out of scope, but I wanted to attempt to produce it, as it would have been a great personal step forward and contribution to this area of study. I was attached to the nature of the research and I was determined to make truly independent and original research; it was a personal ambition. This was an unrealistic goal for an undergraduate study and because it was my first attempt at a proper study. I realised over the course of my study that many experiments relied on other experiments that came before them, and that research is often a slow building process off of other studies. I could have improved my practice by conducting more thorough research on appropriate experimental designs and statistical plans with reference to other relevant studies in the field, and asking for peer or supervisor reviews of my design before progressing any further.

D. Disposition towards research

The project was incredibly demanding, even during the proposal stage. Having to familiarise myself with the state of the art in my chosen field and surrounding information was a tall task as there was so much information available but not all of it was relevant. Having to read through so many and accurately pick out relevant information proved to be a job which was difficult to focus on, as I was incredibly excited to produce my work and talk about it. I found my dedication to the project waxing and waning throughout its duration, especially as I took on more responsibilities in my other module. I realise now that my workload was badly balanced and coordinated for myself, which was detrimental to the quality of this project.

E. Test Plan

The project had challenged the way I approach software development as there was a very big emphasis on the reliability and validity of the artefact. I had to readjust the way I applied testing techniques because I had to provide proof. This was alien to me as I hadn't practiced such strict validation and verification methods before. Whilst I operated by good practice naturally and tested every time new code was added, I was still not used to having to keep a tight record of the tests I took. After having left some features later than I'd hoped, I realised the importance of incredibly thorough testing. It would have saved me some time as I wouldn't have had to fix bugs that I spotted right before deploying the artefact in my testing environment, then going home to fix the bugs. I was too used to my standard practice of light testing as I went along, rather than having a bulletproof test plan ready from the start. I could have improved my practice by researching multiple testing techniques more thoroughly and writing an exhaustive test plan before engaging with writing the code for the artefact.

F. Sophistication of the Algorithm

The algorithm itself was simpler than I'd liked, and would have benefited from implementing elitism. Melodies with low ratings would often persist through the generations, which participants would notice and mention and this principle in genetic algorithms would have helped with that. There was a great deal of new principles I had to learn and it was difficult to learn in a short span of time for the particular task I needed, and the process actually turned out a little more shallow than I had intended.

G. Concluding remarks

In the end, I am proud to have finished this study and underwent the academic process as it challenged the way I think and structure arguments, the way I approach software development and the way I manage my time. The experience was valuable in learning how to seek out new information and collate it. I have set out a collection of SMART objectives to help me prepare for my next study or project, which contain my greatest weaknesses during this project: time management, experimental design and test planning. I will do my best to improve my practices as I continue to work in computer science and keep learn about this field.

S	M	A	R	T
I will produce a plan in advance of the tasks which need to be completed	There will be a goal for each week of the project timeline, as evidenced in a planner or calendar app	I have some experience in producing plans for projects, but need to make them towards the beginning of projects	This will enable me to manage my time more effectively and produce more consistent work on the project	I anticipate completion in the first two weeks of the next project
S	M	A	R	T
I will produce a more sophisticated and robust experimental design before undertaking research	I will ask for colleague and supervisor reviews of my experimental design before producing any artefact	After having performed one study, I have greater experience in the requirements and design of experiments	This will ensure that my next study will be more reliable and robust	I expect that my next experimental design will be completed within one month of starting a project
S	M	A	R	T
I will produce a thorough and comprehensive test plan for the artefact before actually writing any part of it	I will have colleagues and my supervisor review my test plan before deploying it and beginning the artefact	After having produced one study, I understand the requirements and importance of test plans and have enough experience to write one next time	This will ensure that my next study is exhaustively validated and verified so that the study is more solid	The test plan should be completed in the first month, during the experimental design, of the project

APPENDIX B
STATISTICAL ADDENDUM

A. R code

```
library(ggplot2)
library(ggpubr)
library(reshape2)
library(matlib)
library(tidyverse)
library(dplyr)
#Get matrix of all intervals
allIntervals <- as.matrix(intervals[,1:9])
#Get matrix of short intervals (less than a tritone)
shortIntervals <- as.matrix(intervals[,1:4])
#Get matrix of long intervals (more than a tritone)
longIntervals <- as.matrix(intervals[,5:9])

#Get sum of short intervals in each melody line
shorts <- rowSums(shortIntervals)
#Get sum of long intervals in each melody line
longs <- rowSums(longIntervals)

#Get all ratings as a matrix
rating <- as.matrix(intervals)[,10]

#Tibble of short intervals columns and rating
intervalsTable <- tibble(shorts, rating)
#Tibble of long intervals columns and rating
longIntervalsTable <-tibble(longs,rating)
#Spearman's R Test on short interval data set
spearShort <- cor.test(shorts, rating, method=c("spearman"),
      exact=FALSE)
#Spearman's R Test on long intervals data set
spearLong <- cor.test(longs, rating, method=c("spearman"),
      exact=FALSE)
#exact was turned off for both spearman tests because of
      repeated data points found in the data set

#Plots short interval group against rating as points with a
      loess line of best fit
shortPlot <- ggplot(data = intervalsTable, aes(x = shorts, y =
      rating)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Groups of Short Intervals", x = "Total Number
      of short intervals in melody", y = "Rating Received"
      )

#Plots long interval group against rating as points with a
      loess line of best fit
longPlot <- ggplot(data = longIntervalsTable, aes(x = longs,
      y = rating)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Groups of Long Intervals", x = "Total Number
      of long intervals in melody", y = "Rating Received")

#print the plots
print(shortPlot)
print(longPlot)

#prepare all intervals groups and their rating for pie charts
pieIntervals <- tibble(cbind(allIntervals,rating))
#filters all columns fitting into upper and lower quartile
highRatings <- pieIntervals %>% filter(rating>=75)
lowRatings <- pieIntervals %>% filter(rating<=25)
view(lowRatings)
#Prepare labels in same order as data
labels <- c("Unison","Second","Third","Fourth","Tritone","
      Fifth","Sixth","Seventh","Octave")

#sum the total of each interval group in the two quartiles
highRatings <- colSums(highRatings)
lowRatings <- colSums(lowRatings)
#Calculate the percentage proportions of the upper quartile pie
piepercent <- paste0(round(100*highRatings[1:9]/sum(
      highRatings[1:9]),1),"%")
#pie chart for upper quartile ratings
pie(highRatings[1:9],labels = piepercent,
    main = "Interval groups in melodies with ratings 75+",
    col=rainbow(length(highRatings[1:9])))
#print legend for upper quartile pie
legend("topright",labels, cex=1.2, fill = rainbow(length(
      highRatings[1:9])))

#Calculate the percentage proportions of the lower quartile pie
piepercent <- paste0(round(100*lowRatings[1:9]/sum(
      lowRatings[1:9]),1),"%")
#Pie chart for lower quartile ratings
pie(lowRatings[1:9],labels = piepercent,
    main = "Interval groups in melodies with ratings less
        than 25",
    col=rainbow(length(lowRatings[1:9])))
#print legend for lower quartile pie
legend("topright",labels, cex=1.2, fill = rainbow(length(
      lowRatings[1:9])))

#plot number rests against rating
restsPlot <- ggplot(data=rests, aes(x = rests$Rests, y = rests$
      Rating)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(title = "Effect of rests on a melody rating", x = "Total
      number of rests in a melody", y = "Rating Received
      ")
#print plot
print(restsPlot)
#do spearman correlation of rests against rating
spearRest <- cor.test(rests$Rests, rests$Rating, method = "
      spearman")
```

B. Data stored by the artefact

```
Participant ID:b92d59ff Date:23/03/2023
    11:55:31
Date: 23/03/2023 11:58:02 | Generation
    1:
65 -1 -1 64 65 -1 -1 62 65 62 -1 -1 -1
    62 64 72 | Rating: 25
72 -1 64 -1 60 65 71 67 -1 67 60 72 62
    62 -1 60 | Rating: 40
62 67 71 -1 67 62 -1 69 62 71 62 72 60
    64 60 -1 | Rating: 45
72 -1 71 69 67 62 72 67 -1 60 69 -1 67
    65 67 71 | Rating: 35
64 67 62 60 60 65 72 62 67 72 67 69 72
    62 69 72 | Rating: 50
72 60 65 71 65 64 60 -1 65 60 72 62 64
    -1 71 65 | Rating: 40
Date: 23/03/2023 11:59:28 | Generation
    2:
62 67 71 -1 67 62 -1 69 62 71 62 72 72
    62 69 72 | Rating: 25
72 -1 64 -1 67 65 71 67 -1 67 60 72 62
    62 -1 60 | Rating: 40
72 60 65 71 65 64 60 -1 65 67 60 72 62
    62 -1 60 | Rating: 40
72 -1 71 69 67 62 72 69 -1 60 69 -1 67
    65 67 71 | Rating: 50
```

```
64 67 62 60 60 65 72 65 67 72 67 69 72
    62 69 72 | Rating: 30
72 -1 71 69 67 -1 -1 62 65 62 -1 -1 -1
    62 64 72 | Rating: 10
Date: 23/03/2023 12:00:24 | Generation
    3:
72 -1 64 -1 67 65 71 67 -1 60 69 -1 67
    65 67 71 | Rating: 50
-1 -1 64 -1 67 65 71 67 -1 67 60 72 62
    62 -1 60 | Rating: 10
64 67 62 60 60 64 60 -1 65 67 60 72 62
    62 -1 60 | Rating: 30
72 60 65 71 65 65 72 65 67 72 67 69 72
    62 69 72 | Rating: 10
72 -1 71 69 67 -1 -1 69 62 71 62 72 72
    62 69 72 | Rating: 10
62 67 71 -1 67 62 -1 62 65 62 -1 -1 -1
    62 64 72 | Rating: 10
Date: 23/03/2023 12:01:36 | Generation
    4:
72 -1 64 -1 67 65 71 67 -1 60 69 -1 67
    65 -1 71 | Rating: 10
72 -1 64 60 60 64 60 -1 65 67 60 72 62
    62 -1 60 | Rating: 50
72 60 65 71 65 65 71 67 -1 67 60 72 62
    62 -1 60 | Rating: 25
-1 -1 64 -1 67 65 72 65 67 72 67 69 72
    62 69 72 | Rating: 40
62 67 71 -1 67 62 -1 62 65 62 -1 -1 -1
    62 64 72 | Rating: 10
72 -1 71 69 67 -1 -1 69 62 71 62 72 72
    62 69 72 | Rating: 10
Date: 23/03/2023 12:03:21 | Generation
    5:
72 -1 64 -1 67 65 60 67 -1 60 69 -1 67
    65 -1 71 | Rating: 5
72 -1 64 60 60 64 60 -1 65 72 67 69 72
    62 69 72 | Rating: 75
72 -1 64 -1 67 65 71 67 -1 60 69 -1 67
    65 -1 60 | Rating: 25
72 60 65 71 65 65 71 67 -1 67 60 72 62
    62 -1 71 | Rating: 50
72 -1 71 69 67 -1 -1 69 65 62 -1 -1 -1
    62 64 72 | Rating: 25
62 67 71 -1 67 62 -1 62 62 71 62 72 72
    62 69 72 | Rating: 20
Date: 23/03/2023 12:04:00 | Finished
    melody: 72 -1 64 -1 67 65 60 67 -1
    60 69 -1 67 65 -1 72
```

	Rests	Rating
211	4	41
212	4	39
213	5	16
214	4	44
215	5	80
216	6	16
217	3	17
218	6	36
219	3	74
220	3	83
221	7	7
222	4	71
223	3	45
224	9	25
225	4	55
226	5	52
227	4	59
228	7	28
229	4	0
230	8	0
231	4	0

Fig. 11. Example of the rest data set

	Unison	Second	Third	Fourth	Tritone	Fifth	Sixth	Seventh	Octave	Rating
342	0	5	3	0	0	2	0	0	0	40
343	0	3	4	2	0	1	0	3	0	70
344	1	2	2	1	3	1	1	1	0	19
345	0	1	5	0	0	1	1	1	1	21
346	0	6	1	1	1	0	0	0	0	46
347	0	2	4	1	0	0	3	1	1	68
348	1	5	2	0	0	2	1	1	1	80
349	0	2	6	1	0	1	0	3	0	62
350	1	2	2	1	3	1	1	1	0	54
351	0	2	4	1	0	0	3	1	1	76
352	0	6	1	1	1	0	0	0	0	42
353	0	3	3	1	2	1	0	1	0	31
354	1	2	3	0	1	0	2	1	1	41
355	0	3	6	1	0	1	0	2	0	81
356	0	1	4	1	0	0	3	2	1	77
357	0	6	0	0	2	0	1	0	0	54
358	0	3	4	1	2	1	0	1	0	54
359	2	2	2	1	3	1	1	1	0	33
360	0	2	3	1	0	0	1	1	1	21
361	1	3	3	2	0	0	2	1	0	16
362	0	0	7	0	0	1	1	4	0	31
363	0	6	0	0	2	0	1	0	0	25

Fig. 12. Example of interval data set

Appendix C
Testing Addendum

Each unit test was performed manually before committing the feature. Some tests failed later during development and were noted in the commit log, but promptly fixed. Features not listed in the tables were still tested before deployment.

Test Case Type	Description	Test Step	Expected Result	Status
Functionality	Artefact should produce a random population of melodies which fit into the key	Compare produced melodies against possible melodies	All notes found within melodies should be	Pass
Usability	Artefact should be able to play back the melodies	Play all melodies from multiple generations	Melodies should play back correctly	Pass
Usability	Artefact should be able to rate the melodies	Observe the fitness melodies have when breeding	The melodies are appropriately ranked by fitness	Pass
Functionality	Artefact should merge melodies and produce correct child melodies	Compare child melodies against their respective parent melodies. Check for mutation.	All notes (except mutation cases) should contain notes from the parent melodies	Pass
Functionality	Artefact should correctly humanize melodies	Check that the MPTK events have varying durations and volumes	All MPTK events should have varying durations and volumes per playback	Pass
Functionality	Artefact should save the melodies in the correct format	Ensure that data saved is accurate and consistent	The data saved is the same as the data produced in use	Pass
Usability	Artefact should only unlock after consent given	Ensure artefact use is locked until consent is given	Artefact is inaccessible until consent given	Pass
Functionality	Artefact should correctly interpret interval data from file	Compare interval data against actual interval count of the melodies	The data analysed will be the same as the manual calculation	Pass
Functionality	Artefact should correctly interpret rest data from file	Compare rest data against actual rest data count of the melodies	The rest number will be the same as the manual calculation	Pass

Fig. 13. Table of unit tests performed

Melody data used for unit test	Do all notes belong to the key (Pass/Fail) [60,62,64,65,67,69,71,72,-1]
60 -1 67 69 72 62 69 64 -1 62 65 -1 -1 -1 72 -1	Pass
-1 64 -1 67 -1 -1 -1 69 -1 -1 64 -1 64 69 -1 64	Pass
69 60 67 60 -1 -1 60 72 -1 -1 -1 62 62 62 71 60	Pass
67 71 69 67 67 69 67 62 71 72 65 69 72 64 72 60	Pass
71 -1 65 69 71 71 -1 64 65 69 67 72 71 72 67 60	Pass
69 -1 64 69 -1 72 67 -1 -1 60 64 67 -1 -1 60 64	Pass
71 -1 65 69 71 71 -1 64 65 69 67 72 71 64 72 60	Pass

Fig. 14. Unit test of melody generation

Melody Used for Unit Test	Interval Test Data	Interval Expected Data	Pass/Fail
62 72 -1 62 69 -1 69 71 69 69 65 -1 64 -1 72 62	2,3,1,0,0,1,1,3,0	2,3,1,0,0,1,1,3,0	Pass
62 67 64 62 65 62 64 62 65 -1 -1 65 -1 -1 67 62	1,4,4,2,0,0,0,0,0	1,4,4,2,0,0,0,0,0	Pass
64 64 -1 72 67 71 71 65 65 -1 71 67 62 69 -1 62	3,0,2,2,2,2,1,0,0	3,0,2,2,2,2,1,0,0	Pass
60 62 -1 -1 69 -1 65 -1 -1 62 -1 71 71 64 69 72	1,1,3,1,0,2,1,0,0	1,1,3,1,0,2,1,0,0	Pass
62 71 60 71 64 -1 64 69 -1 -1 65 69 71 -1 67 71	1,1,4,1,0,1,1,2,0	1,1,4,1,0,1,1,2,0	Pass

Fig. 15. Unit test of interval analysis

Melody Used for Unit Test	Rest Test Data	Expected Rest Data	Pass/Fail
-1 69 72 65 72 60 62 71 -1 -1 72 72 67 65 60 -1	4	4	Pass
65 62 60 69 72 67 67 71 72 69 -1 65 67 -1 72 -1	3	3	Pass
72 64 60 71 -1 -1 62 71 67 -1 69 71 -1 62 -1 69	5	5	Pass
64 72 62 67 60 60 69 62 65 67 62 62 -1 71 62 -1	2	2	Pass
71 67 72 -1 -1 71 62 -1 64 67 -1 -1 62 -1 60 72	6	6	Pass
72 71 64 65 67 62 69 67 69 67 60 67 62 62 65 62	0	0	Pass
60 72 -1 -1 69 -1 64 -1 62 67 62 60 71 -1 60 -1	6	6	Pass
-1 62 60 67 67 64 60 65 60 69 64 62 -1 -1 60 62	3	3	Pass
69 69 64 -1 67 72 65 62 71 60 60 -1 -1 67 67 -1	4	4	Pass

Fig. 16. Unit test of rest analysis

APPENDIX D
ADDITIONAL MATERIAL

This section holds any additional miscellaneous appendices

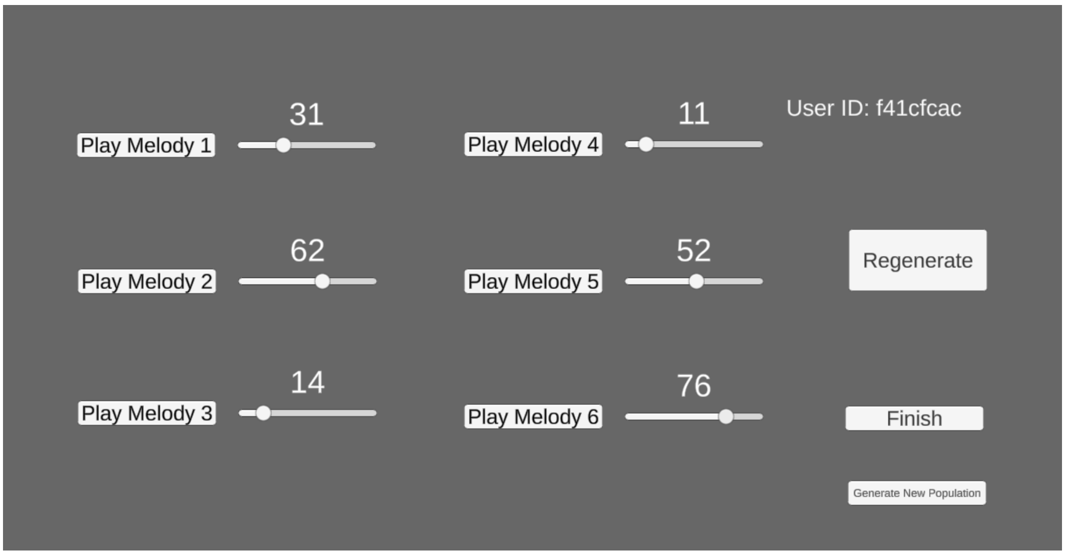

Fig. 17. The primary user interface as seen in the artefact

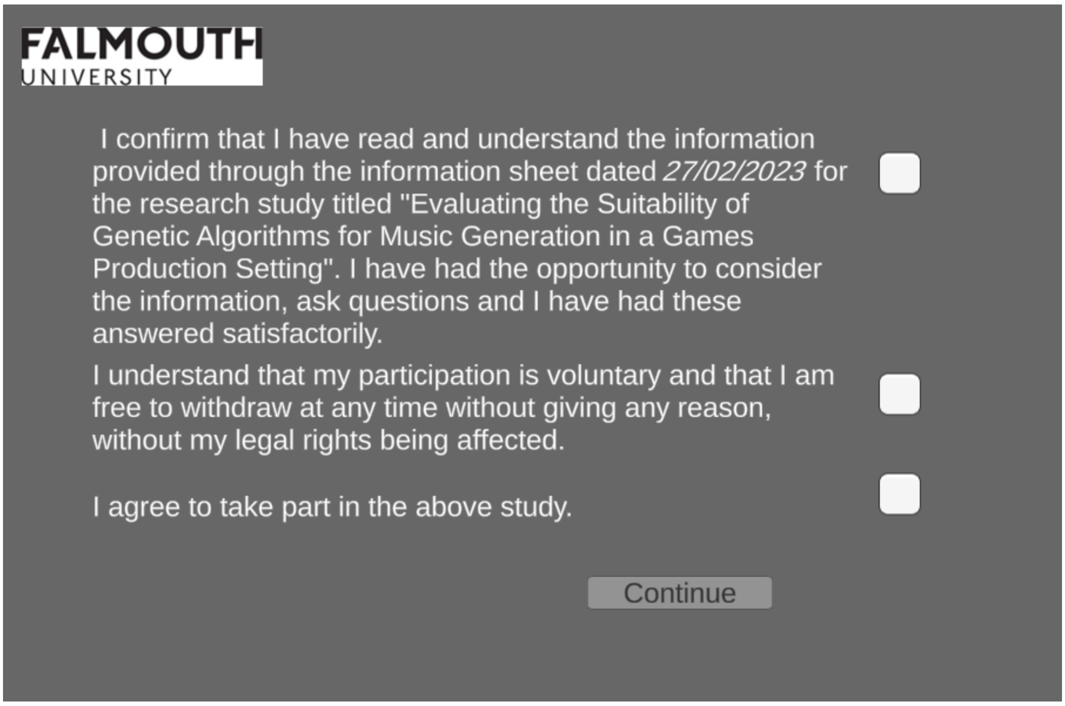

Fig. 18. Consent form as seen in the artefact