DURMS
Durham Undergraduate
Research in Music & Science

Investigating Whether Music Composed by Artificial Intelligence can Elicit the Same Strength of Emotional Response as Music Composed by Humans

Romy Barratt
Durham University

ABSTRACT

This study explores whether music composed by Artificial Intelligence (AI) can elicit the same strength of emotional response as music composed by humans. The study provides a holistic overview of participants and evaluates the effect of musicianship on the perception of AI music. A survey gathering quantitative and qualitative data was distributed. The survey presented 3 extracts (2 AI, 1 human composed) followed by questions in response to the music heard. The participants attempted to determine the nature of the composer of the extracts and were given a chance to change their rankings of emotional response to the extracts once the nature of the composer was revealed. Significance testing using ANOVA was conducted and thematic analysis of qualitative data explained the results. Results found that the nature of the composer had no impact on the strength of emotional response of participants and that participants were unable to distinguish between AI and human composed extracts. Further research should gather data from a larger sample size and investigate the response of composers.

1. INTRODUCTION

Artificial Intelligence (AI) composers are achieving increasing recognition across the music industry. AI has the ability to create their own compositions and act a tool in aiding humans create their own works (Micchi et al., 2021), and creative activities (Hong et al. 2022). 'The precise boundaries of AI are elusive since they are related to what people feel is "intelligent behavior"' (Roads, 1985: p163). For centuries, listeners have attributed personalities and motivators to the creators of the music they listen to (Audry, 2019: p5). This study explores whether listeners are able to experience the same strength of emotional response from an AI composer as a human composer when they are not aware of the nature of the composer. Furthermore the study explores the perception of the AI composer as creative, able to create aesthetically pleasing music and the musical elements it uses. The significance of musicianship on the ability to discern AI and human composers will be tested. By carefully designing the survey to ensure that the term AI is only mentioned after the evaluation of the pieces, the study is able to evaluate whether a pre-conceived bias towards AI music changes participants perception and evaluation of the extracts.

Due to the rapid progression in the ability of AI music to learn from human resources, the first hypothesis of this study is, H_1: Participants will not be able to distinguish AI composed music from human composed music. This hypothesis is drawn from the findings of Tigre & Maw who determined that an awareness of the composer being AI had little impact on participants enjoyment of the music (2020). There has been much disagreement between academics as to whether AI can elicit the same strength of emotional response as human composers. Tubadji et al. argue that AI composers cannot fulfill the emotional needs of a listener, even when they are unaware of the nature of the composer (2021: 7) and lacks the human emotionality embedded in the music (2021: 1). Through Rogers et al.'s discovery that music is capable of shaping shopping habits and the possibility of creating an AI biometric to imitate this suggests that AI is capable of eliciting the same strength of emotion as human music (2021). Zulić corroborates this discovery, arguing that 'AI can create art that emulates human emotions' (2019: 112). This led to the second hypothesis, H_2: The nature of the composer will have no significance on the strength of emotional response to the extracts felt by the participants. Music professionals displayed a more negative response towards AI music than non-music professionals in the study conducted by Tigre & Maw (2021). Furthermore, music professionals exhibited a more intense emotional response to music when listening to music compared with amateurs (Mikutta et al., 2014: 102). This has led to the third hypothesis, H_3: Musicians will have a stronger emotional response to music composed by humans than non-musicians. One of the limitations of the study conducted by Tigre & Maw is that the limited range of their age demographic could have affected the results, introducing a potential age bias (2021). The current study has endeavoured to increase the age range to ensure the results are representative of a wide distribution of ages. Zulić discusses the capabilities of AI composition platforms such as AIVA as being able to write beautiful and emotional music (2019: 104). On the other hand, Hertzmann presents the view that art requires 'human intent, inspiration, a desire to express something' (2018: 1), which counter Zulić's views on AI being capable of reproducing human intent in art. The fourth hypothesis of

this study therefore determines that, H_4: There will be no significant relationship between participants' ranking of the strength of emotion the composers conveyed and the nature of the composer.

This study also investigates participants' perception of the different rankings of composer's creativity and ability to create aesthetically pleasing music without the knowledge of the nature of the composer. AI's ability to create aesthetically pleasing and creative music is widely disputed amongst academics. Hong et al. argue that it is the quality of the composition created that matters, not how it is made, showing that the creative process of the composition does not impact the quality of the resulting music (2022). A limitation of this study is the use of only quantitative data. Also gathering qualitative data would have allowed a more detailed understanding of participants' reaction to the music and a better evaluation of their thoughts. To allow for more extensive research on participants' opinions, the current study has gathered a combination of quantitative and qualitative data. This has led to the hypothesis, H_5: There will be no significant relationship between participants' ranking of music as aesthetically pleasing and the nature of the composer. In recent years, technology has been constantly evolving to allow computer programmes to generate new algorithms allowing AI to generate novel artworks (Chamberlain et al., 2018: 178). Therefore, H_6: There will be no significance in participants' ranking of perceived composer creativity in human and AI composed music.

Hong et al. tested participant's reaction to AI composed music in relation to the violation theory which found participants to possess a preconceived bias of AI music and enjoyed music less when they believed it had been composed by AI (2021: 1928). This counters the findings of Tigre & Maw who found that an awareness of AI authorship had little impact on participants' enjoyment of the music (2021: 144). However, Agudo et al. conclude similar results to Hong et al., that participants reported a stronger emotional response to music they believed to have been composed by a human, indicating a pre-conceived bias (2022: 4). A limitation of this study was the use of a multimedia video shown to participants. The effects of only AI composed music cannot be fully evaluated as multimedia was an external factor which influenced participants' perception of artwork. Tubadji et al. corroborates Agudo et al.'s results finding that participants preferred music composed by humans as opposed to AI when they knew the nature of the composer (2021: 1). The evidence for negative bias remains unclear (Ragot et al., 2020: 2). A further study by Hong et al. found that participants who were less willing to accept AI products were less likely to have a positive experience of them. (2021). One of the limitations of this study is that a genre appreciation test was not conducted before the study and therefore a genre bias may have influenced the results. Therefore, H_7: Pre-conceived bias will affect participants' perception of the music after the nature of the composer is revealed. AI can analyse over 300,000 musical scores and create an original work using a mathematical formula (Barreau, 2018). Due to AI's ability to learn from and recreate musical devices and techniques used by humans, H_8: Participants will identify the same musical devices that elicit an emotional response in human and AI composed music.

2. METHOD

Design. A survey gathering qualitative and quantitative data was used to investigate whether participants reacted with a stronger emotional response to human composed music compared to AI composed music. The survey collected demographic data and established whether there was a genre bias for classical piano music amongst participants by assessing how much classical piano music participants listened to on a weekly basis and asked participants to select the genres of music they listen to. This eliminated a genre bias and lowered the likelihood that participants would already be well acquainted with the extracts and know the nature of the composer (Hong, 2021). The term Artificial Intelligence was not mentioned or included in any documents, information sheets or consent forms until after participants had listened to all three extracts and responded to the initial questions. This eliminated bias before participants listened to the extracts and allowed analysis on the impact of bias on the perception of AI composed music once the composer was revealed (Agudo, 2022: 5). Furthermore, this allowed analysis on the thought process behind why participants perceived pieces of music to be composed by AI/humans. Quantitative data was analysed through ANOVA tests, graphs and qualitative data was coded and grouped to support and explain the findings.

Participants. The survey was taken by 37 participants recruited via a Durham University Facebook forum and shared with friends via WhatsApp. None of the participants showed signs of a genre bias for classical piano music so all 37 participants were eligible with all participants ranking their knowledge of classical piano music below 3 out of 10. Participants were required to be at least 18 years old to complete the survey. Participants ranged 18-67 with M=25. 51.4% of participants were female (N=19), and 48.6% were male (N=18). The even split of genders ensures that the data is not affected by a gender bias. 64.9% of participants self-identified as musicians (N=24) and 35.1% self-identified as non-musicians (N=13). Having a large number of participants that identified as musicians allowed the study to conduct thorough research with an important sample size. One of the limitations

of the survey is the smaller number of non-musicians who completed the survey. Ideally, the study would have 10 more non-musicians participate to have an equal representation of musicians and non-musicians in the data.

Materials and Stimuli. This study was conducted via an online Qualtrics survey collecting qualitative and quantitative data. The survey consisted of three musical extracts, two of which had been composed by AI and one composed by a human. These extracts were from YouTube and can be found in Appendix A. The nature of the composers of the stimuli were not revealed to participants until they had attempted to discern which pieces were created by AI and which were created by humans.

Procedure. The survey consisted of 6 sections. The first section gathered demographic data and established whether participants showed signs of a genre bias towards classical piano music. The second, third and fourth sections were identical to each other. The music was presented to all participants in the same order. Firstly, they each consisted of the participants listening to the musical extract with no name or composer details. Next, participants were asked questions about their reactions to the music giving ratings using a Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree). Qualitative questions followed to allow participants to elaborate on their answers. The fifth section asked participants to rank the order of the extracts they responded to with the most strength of emotion. They were then asked to answer questions on their degree of agreement to questions about AI composers and music using a Likert scale as before. Next, participants were asked to elaborate on their answers with a series of qualitative questions. The sixth section asked participants to discern which of the extracts they had listened to had been composed by AI and which had been composed by a human. After revealing the true nature of the composer, participants were asked their degree of agreement using a Likert scale about the process of listening to the music and their differing opinions after learning the nature of the composer. Participants were asked to re-rank their strength of emotional reaction to the extracts and were given a chance to elaborate on their opinions. The terms AI and human composer were not mentioned until the fifth section, after the participants had listened to the extracts and answered all questions on their emotional reaction to the stimuli. This avoided suspicion on the nature of the composer and eliminated introducing bias into the data collected. No names were attached to the extracts to ensure that all of the data collected was based purely on an auditive and not visual source. Extracts were simply referred to by their number. This was done to combat the limitations identified in the study of Agudo et al. of using multimedia as their stimuli (2022). A complete copy of the questions asked can be found in Appendix B.

3. RESULTS

Results were evaluated through a series of hypotheses testing. Significance was tested through ANOVA tests coded in R. Results must be $p<.05$ to be significant and less than $p<.01$ to be very significant, or show practically certain relationship.

H1: Participants will not be able to distinguish AI composed music from human composed music. 5 participants were able to identify the nature of all the composers correctly. 14 identified 2 composers correctly, 7 identified 1 composer correctly and 10 participants identified 0 composers correctly. There is no correlation between these numbers showing that participants were not able to distinguish AI composed music from human composed music. Figure 1 shows the breakdown of musicians and non-musicians identification of the composers. This also demonstrates no correlation between the number of composers correctly identified and the number of participants who identified them. H1 is accepted. Upon analysing the data in percentage form, it is not possible to determine that musicians were more able to discern the difference between AI and human composers. Although a higher percentage of musicians identified 3 composers correctly, a higher percentage of non-musicians were able to identify 2 composers correctly (see Figure 2). There is no correlation between musicianship and the ability to identify the nature of the composer.

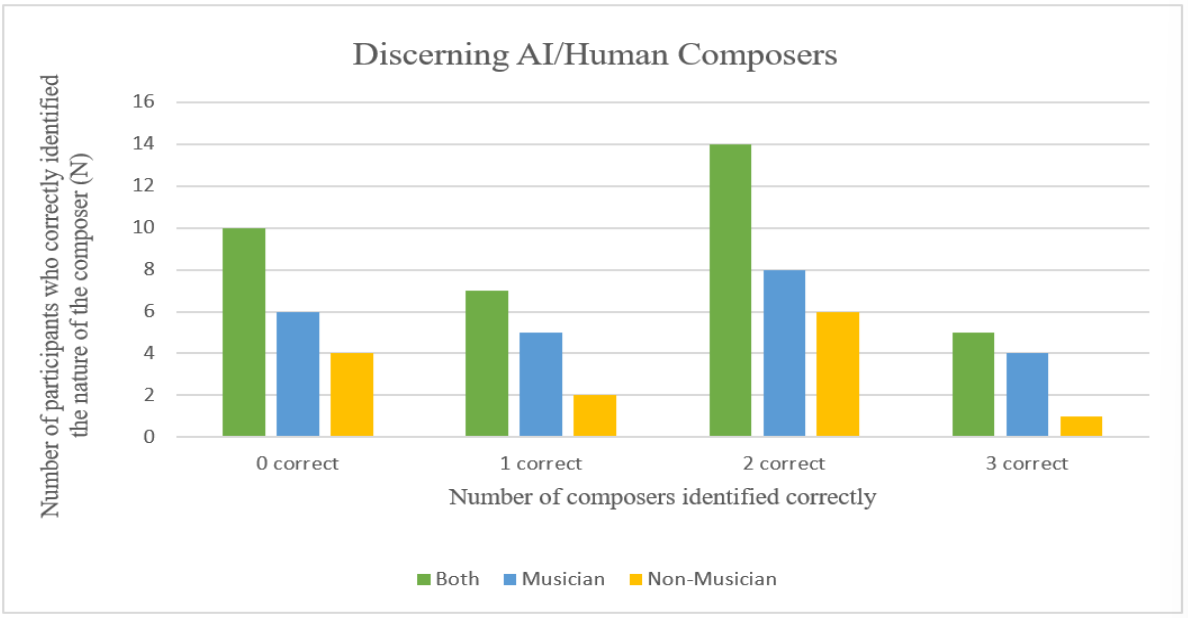


Figure 1. Bar chart showing the number of participants mapped against the number of composers correctly identified.

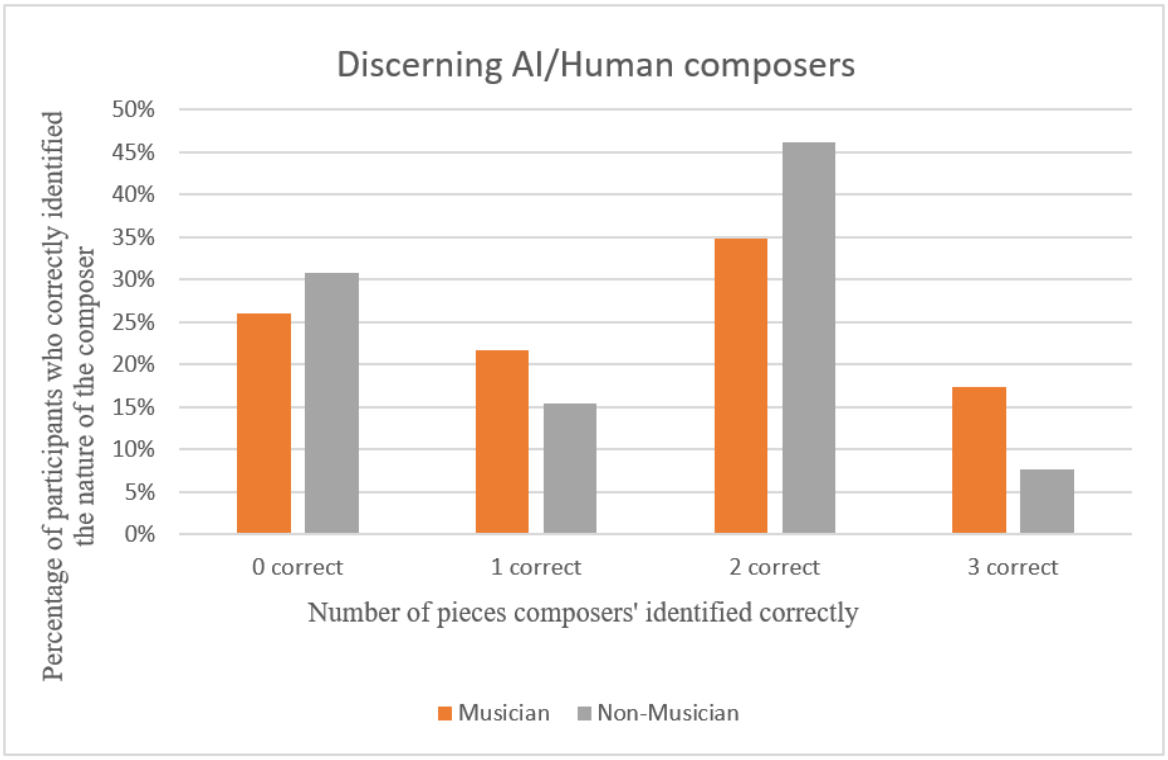


Figure 2. Bar chart showing the percentage of composers guessed correctly by musicians and non-musicians.

H2: The nature of the composer will have no significance on the strength of emotional response to the extracts felt by the participants. Through ANOVA testing of all participants, results determine the results were borderline significant ($F(1, 38)=3.06$, $p=.053$,). H2 is accepted as the results did not show statistical significance. The same significance test was conducted on only musicians, results also demonstrated that there was no significant relationship between the participants who identified as musicians' strength of emotional response to the music and the nature of the composer ($F(1, 25)=2.03$ and $p=.14$, $p>.05$). Through ANOVA significance testing of

participants who identified as non-musicians, results demonstrated the same findings ($F(1, 14)=1.95$ and $p=.14$, $p>.05$). The test found no significant relationship between the nature of the composer and the strength of emotion felt by participants who identified as non-musicians.

H3: Musicians will have a stronger emotional response to music composed by humans than non-musicians. Musicians' and non-musicians' emotional responses to human composed music, gathered through a Likert scale, were evaluated through analysis of the boxplot (see Figure 3). This boxplot was created using Python code (Appendix C). The spread of data is larger for musicians than non-musicans. The mean of the data for non-musicians is 3 (neither agree nor disagree) and the mean of data for musicians is 4 (agree). This shows that musicians reported a stronger emotional response to music composed by humans than non-musicians. H3 is accepted.

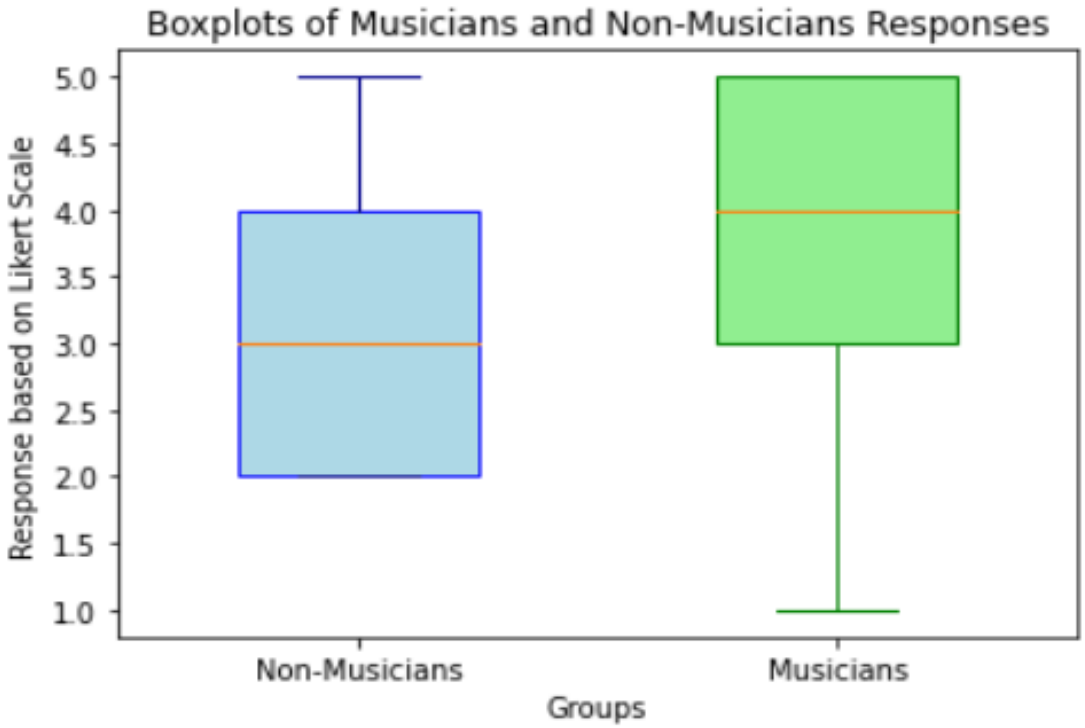


Figure 3. Boxplot showing the spread of emotional responses of musicians and non-musicians to human composed music.

H4: There will be no significant relationship between participants' ranking of the strength of emotion they felt that the composers conveyed and the nature of the composer. An ANOVA test found no significant interaction between the strength of emotion that participants felt the composers conveyed with the nature of the composer ($F(1, 38)=1$ and $p=.37$, $p>.05$). H4 is confirmed. Conducting an ANOVA test for musicians ($F(1, 25)=.78$ and $p=.46$, $p>.05$) and non-musicians ($F(1, 14)=1.14$ and $p=.34$, $p>.05$) showed that there was no significance between the strength of emotion that musicians and non-musicians felt the composer conveyed and the nature of the composer.

H5: There will be no significant relationship between participants' ranking of music as aesthetically pleasing and the nature of the composer. The results of the ANOVA test for comparing the participants' ranking of the extracts as aesthetically pleasing and the nature of the composer shows a highly significant interaction between the participants ranking of the extract as aesthetically pleasing and the nature of the composer ($F(1, 38)=7.61$ and $p=.001$, $p<.01$). Therefore, H5 is rejected. Creating a 2-way table reveals that the significance arose between Extract 2 and Extract 3 (see Table 1). This shows the nature of the composer to have an impact on participants reception of the music as aesthetically pleasing. Upon conducting an ANOVA test to test for the significance in musicians ranking of the extracts as aesthetically pleasing and the nature of the composer borderline significance can be found between musicians rankings and the nature of the composer ($F(1, 25)=3.17$ and $p=.051$, $p>.05$). The

ANOVA test conducted to test for significance in non-musicians rankings and the nature of the composer revealed an almost certain significance in the relationship ($F(1, 14)=6.24$ and $p=.007$, $p<.01$).

Table 1. 2-way ANOVA table testing the statistical significance between the relationship of the participants ranking of the music as aesthetically pleasing and the nature of the composer.

	Extract 1 (AI)	Extract 3 (Human)
Extract 3 (Human)	0.07	
Extract 2 (AI)	0.07	0.006

H6: There will be no significant difference in participants' ranking of perceived composer creativity in human and AI composed music. The results of an ANOVA test ($F(1, 38)=1.51$ and $p=.23$, $p>.05$) show that there is no significance between participants ranking of composer creativity and the nature of the composer as AI/human. H6 is accepted. A test conducted for musicians ($F(1, 25)=1.81$ and $p=.18$, $p>.05$) and non-musicians ($F(1, 14)=0.75$ and $p=.48$, $p>.05$) show no significance in the relationship between participants ranking of composer creativity and the nature of the composer for musicians or non-musicians.

H7: Pre-conceived bias will affect participants' perception of the music after the nature of the composer is revealed. After the nature of the composer was revealed to participants, 27% of participants (N=10) changed their ranking of which extract they reacted to with the strongest emotional intensity, 20% of musicians (N=5) and 38% of non-musicians (N=5). More participants would have to be gathered to test whether there is statistical significance between participants pre-conceived bias of a composer and their perception of the music after the nature of the composer is revealed. H7 is undetermined.

H8: Participants will identify the same devices that elicit an emotional response in human and AI composed music. By coding the qualitative data and organising the codes into themes, it is possible to compare the emerging musical devices across the three extracts. Almost identical themes emerged including dynamics, harmony, tonality, tempo, melody, rhythm and timbre (see Figure 4). Although tempo was most often mentioned in regards to extract 1, it was mentioned for all 3 extracts showing that the same musical devices across human and AI composers impact a pieces ability to convey emotion. A breakdown of the musical devices and their grouping into codes and themes can be found in Appendix D. H8 is accepted.

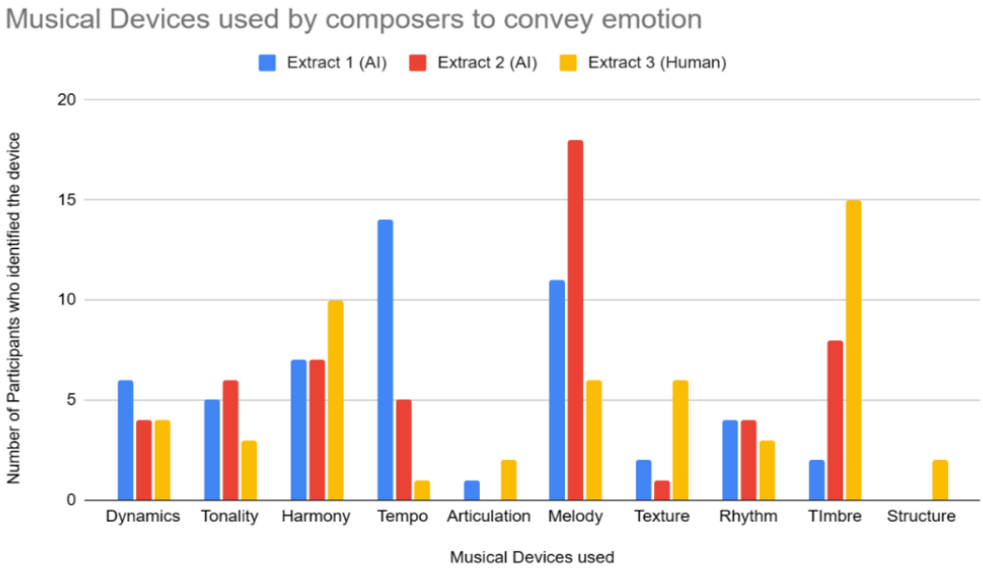


Figure 4. A bar graph showing the musical devices used by composers to convey emotion and the number of participants who identified the device in each extract.

Results suggest AI composers are capable of eliciting the same strength of emotional response as human composers. Although H2, The nature of the composer will have no significance on the strength of emotional response to the extracts felt by the participants, is accepted as no significance could be found between the strength of emotional response of the participants and the nature of the composer, qualitative data reveals that a greater range of emotions were evoked by the human composed extract (see Figure 7) compared to the AI extracts (see Figure 5 and Figure 6). A range of positive and negative emotions were elicited by the AI and human composed extracts. The human composed extract however demonstrated the greatest range of emotions with the qualitative data coded into 14 sections and the AI extracts could be coded into M=10 themes. The greater range of words used to describe the emotions experienced by participants demonstrates that human composed music is capable of eliciting a greater range of emotions than AI composed music.

Figure 5. Emotions elicited by Extract 1 (AI)

Figure 6. Emotions elicited by Extract 2 (AI)

Figure 7. Emotions elicited by Extract 3 (Human)

Upon evaluating the musical devices used by composers which were identified by participants as eliciting an emotional response, the coding and theming of the qualitative data reveals that AI relied more on tempo and melody to generate an emotional response whereas human composers relied more on timbre and texture. Participants also identified the use of structure by human composers whereas this was not identified in the AI compositions. Converting the codes into word clouds reveals that both AI and human composers relied on a sense of rising and falling to create an emotional response (see Figure 8, Figure 9 and Figure 10).

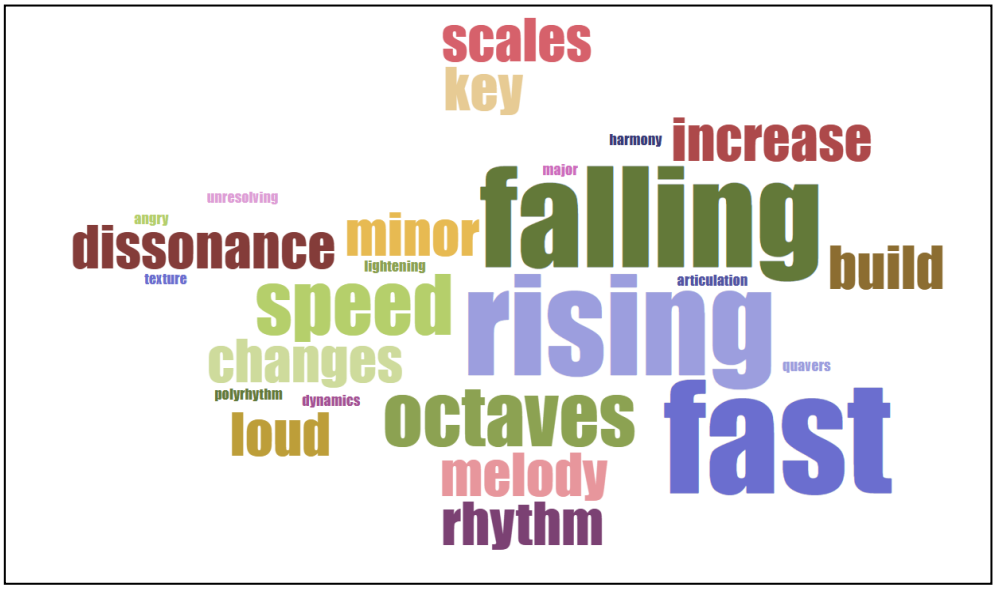


Figure 8. A word cloud showing the qualitative data codes of musical devices identified by participants in Extract 1 (AI)

Figure 9. A word cloud showing the qualitative data codes of musical devices identified by participants in Extract 2 (AI)

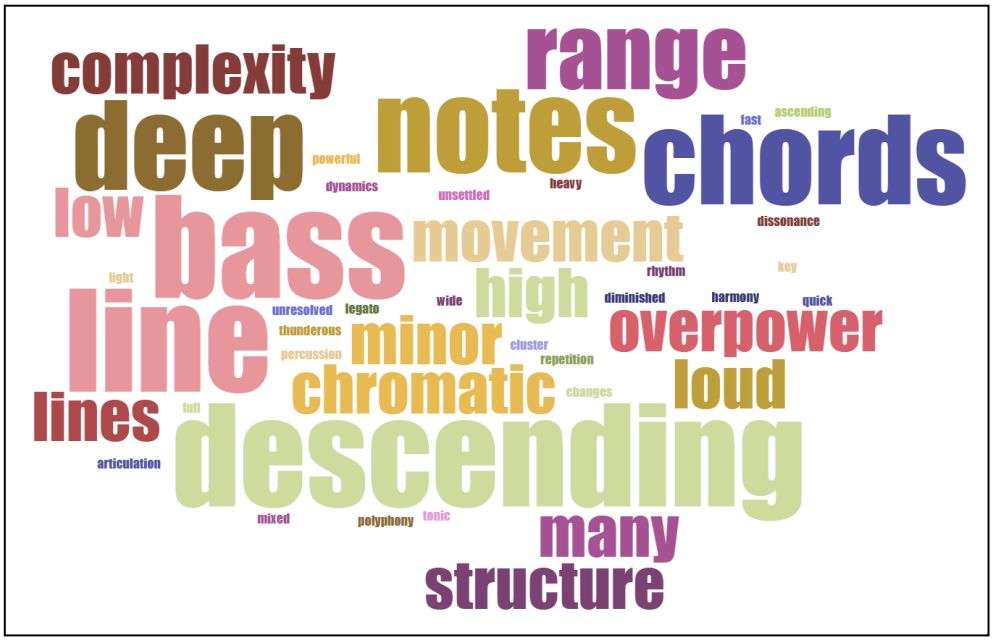


Figure 10. A word cloud showing the qualitative data codes of musical devices identified by participants in Extract 3 (Human)

4. DISCUSSION

Participants were pleasantly surprised by the creative capabilities of AI and its ability to elicit an emotional response across musicians and non-musicians. Participant 2 states that they like the first 2 pieces more as they are impressed with the capabilities of AI. Participant 4 was surprised by the emotional intensity AI was able to convey. This counters the findings of Tubadji et al. that AI music lacks an essence of human emotionality (2021). The study supports Zulić's findings that AI can create art that emulates human emotions (2019). The bias found upon participants reordering of their preference of the music aligns with the findings of Hong et al. (2021) and Tigre & Maw (2021) who found that awareness of a composition being created by AI impacted people's enjoyment of the piece. This is true even after they have conveyed their appreciation for the piece with reordering of preferences favouring human composed music with the number of participants favouring it increasing by 3.

One of the limitations of the study was that number of participants did not provide conclusive data to be able to accept or reject H7. 27% participants (N=10), 20% of musicians (N=5) and 38% of non-musicians (N=5) changed their rankings after finding out the nature of the composer. Although 38% seems high, this was only in fact 5 participants and a number of factors could have led them to re-rank the extracts that were not necessarily due to bias. Results for H7 were therefore inconclusive due to this limitation. Furthermore, both of the AI extracts were sourced from the same AI composer. Although, this allowed for effective comparisons across the AI works and the human composed work by averaging the results of both extract 1 and 2 for comparison with extract 3. Further research should compare the results found from multiple AI and human composers to evaluate whether the results are consistent. This survey provides a good start into the exploration between the strength of emotion that AI can convey in comparison with human composers in a field that has been minimally explored and allows for future researchers to build on by comparing other composers. Another limitiation was the studys use of only one human composed extract. Having the same number of human and AI composed extracts could have revealed varying results.

Participants were not able to discern the difference between AI and human composed music and identified the same musical features that provoked emotion across all 3 extracts, revealing the most pertinent musical features to making an emotional and creative piece of music. There was no difference between the statistical tests conducted for musicians and non-musicians. This suggests that music professionals were unable to discern the difference between AI and human composed music, vouching for AI's capabilities in producing the same level of emotional response as human music. The reason for the greater data spread in H3 could be due to the greater number of musicians who completed the survey. Although the mean of musicians responses to human composed music was higher than non-musicians responses, a greater number of participants would be required to test whether this is statistically significant

The 3 extracts chosen possessed similar musical qualities and were all performed by the same instrument, piano. This eliminated the chance for discrepancies in eliciting emotions being due to an extreme range of music. Furthermore, a genre bias test was conducted at the start of the survey to ensure that participants did not have a bias for classical piano music and therefore potentially be acquainted with these works. No participants demonstrated a genre bias and their listening to these extracts for the first time ensured that their first interpretation of their emotional response to the music was recorded. Despite the pieces possessing similar musical qualities, H5 was rejected as significance could be found between the nature of the composer and the music being described as aesthetically pleasing. This explains the reasoning for participants ranking of Extract 3, the human composed extract as their favourite with N=15 ranking Extract 3 as their favourite, N=14 for Extract 2 and N=8 for Extract 1. Participants were not aware of the nature of the study as gathering information about music composed by AI. After the nature of the composer was revealed to participants, the reordering of their preferences exhibited pre-conceived bias by a small number of participants. This number is not significant and many participants found a new appreciation for AI composed music who did not reorder their preference of the extracts. The study shows that there is no significance between the nature of the composer and their ability to create aesthetically pleasing, creative and emotional music. The same musical devices were used across all three extracts and they elicited the same strength of emotional response in participants. Reactions gathered by qualitative data show that participants responded to the human composer with a greater range of emotions, but there was no statistical significance of the strength of emotional response to each composer.

5. CONCLUSION

In conclusion, this study has found that music composed by AI is capable of eliciting the same strength of emotional response as music composed by humans. Furthermore, musicianship had no impact on the results as there was no difference between tests for statistical significance across the hypothesis.

REFERENCES

Agudo, Ujué., Arrese, Miren., Liberal, Karlos G., and Matute, Helena. 2022. 'Assessing Emotion and Sensitivity of AI Artwork.' *Frontiers in Psychology* 13(879088): 1-9. doi:10.3389/fpsyg.2022.879088.

Audry, Sofian., and Ippolito, Jon. 2019. 'Can Artificial Intelligence Make Art without Artists? Ask the Viewer.' Arts 8(1): 1-8. doi:10.3390/arts8010035.

Barreau, Pierre. 2018. How AI could compose a personalised soundtrack to your life. TED. Retrieved from https://www.ted.com/talks/pierre_barreau_how_ai_could_compose_a_personalized_soundtrack_to_your_life?language=en.

Chamberlain, Rebecca., Mullin, Caitlin., Scheerlinck, Bram., & Wagemans, Johan. 2018. 'Putting the Art in Artificial: Aesthetic Responses to Computer-Generated Art.' *Psychology of Aesthetics, creativity, and the Arts* 12(2): 177-192. doi:10.1037/aca0000136.

Hertzmann, Aaron. 2018.' Can Computers Create Art?' Arts 7(2): 1-25. doi:10.3390/arts7020018.

Hong, Joo Wha., Peng, Qiyoa., and Williams, Dmitri. 2021. 'Are you ready for artificial Mozart and Skrillex? An experiment testing expectancy violation theory and AI music.' *New Media & Society* 23(7): 1920-1935. doi:10.1177/1461444820925798.

Hong, Joo Wha., Fischer, Katrin., Ha, Yul., and Zeng, Yilei. 2022. 'Human, I wrote a song for you: An experiment testing the influence of machines' attributed on the AI-composed music evaluation.' *Computers in Human Behaviour* 131(107239): 1-12. doi:10.1016/j.chb.2022.107239.

Micchi, Gianluca., Bigo, Louis., Giraud, Mathieu., Groult, Richard., & Levé, Florence. 2021. 'I Keep Counting: An Experiment in Human/AI Co-Creative Songwriting.' *Transactions of the International Society for Music Information Retrieval* 4(1): 263-275. doi:10.5334/tismir.93.

Mikutta, C. A., Maissen, G., Altorfer, A., Strik, W., and Koenig, T. 2014. 'Professional Musicians Listen Differently to Music.' *Neuroscience* 268: 102-111. doi:10.1016/j.neuroscience.2014.03.007.

Ragot, Martin., Martin, Nicholas., and Cojean, Salomé. 2020. 'AI-generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?' Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 1–10.

Roads, Curtis. 1985. 'Research in Music and Artifiial Intelligence.' *Computing Surveys* 17(2): 163-190.

Rogers, Waymond., Yeung, Fannie., Odindo, Christopher., and Degbey, William Y. 2021. 'Artificial intelligence-driven music biometrics influencing customers' retail buying behaviour.' *Journal of Business Research* 126: 401-414. doi:10.1016/j.jburses.2020.12.039.

Tigre Moura, Francisco., and Maw, Charlotte. 2021. 'Artificial intelligence became Beethoven: how do listeners and music professionals perceive artificially composed music?' *Journal of Consumer Marketing* 38(2): 137-146. doi:https://doi-org.ezphost.dur.ac.uk/10.1108/JCM-02-2020-3671.

Tubadji, Annie., Huang, Haoran., and Webber, Don J. 2021. 'Cultural promximity bias in AI-acceptability: The importamce of being human.' *Technological Forecasting & Social Change*, 173(121100): 1-14. doi:10.1016/j.techfore.2021.121100

Zulić, Harun. 2019. 'How AI can Change/Improve/Influence Music Composition, Performance and Education: Three Case Studies.' *INSAM Journal of Contemporary Music, Art and Technology* I(2), 100-114.

APPENDICES

Appendix A: Musical Stimuli

Extract 1: AI composed piece: Apocalypse
https://www.youtube.com/watch?v=YA35O6_Nvbk

Extract 2: AI composed piece: Supernova
https://www.youtube.com/watch?v=nuSUTaN16aw

Extract 3: Human composed piece (Alexander Scriabin): Etude Op. 8, No. 2
https://www.youtube.com/watch?v=6Oh8NHYPI5I

Appendix B: Survey Questions

Section 1:

1. What is your age?
2. What is your gender?
3. What is your level of musicianship
4. How many hours do you listen to music per week?
5. How many hours do you listen to classical piano music per week?
6. What genres of music do you enjoy listening to?

Section 2/3/4:

1. Please indicate how much you agree with the following statements (5 point Likert scale: strongly disagree, disagree, neither agree no disagree, agree, strongly agree):
    a. I felt strong emotional intensity when listening to this extract
    b. The composer is conveying emotion
    c. I found this piece of music aesthetically pleasing
    d. The composer is creative
    e. This piece of music made me feel nostalgic
    f. This piece of music made me feel joyful
    g. This piece of music made me feel determined
    h. This piece of music made me feel anxious
    i. This piece of music made me feel happy
2. What emotions did you feel when listening to the extract? (open question)
3. What parts of the music bought out those emotions? Why? (open question)
4. What emotions do you think the composer was attempting to convey and how did they do this? (open question)

Section 5:

1. Which extract do you feel you reacted with the strongest emotional intensity to? (1 being the most, 3 being the least)
2. Indicate your level of agreement with the following statements (5 point Likert scale: strongly disagree, disagree, neither agree nor disagree, agree, strongly agree):
    a. A composer's creativity is important to the composition
    b. A composer's expression of emotion is important to a composition
    c. Music composed by Artificial Intelligence (AI) is able to convey emotion
    d. Music composed by Artificial Intelligence (AI) is capable of conveying the same level of emotion as music composed by a human
    e. I prefer music composed by humans as opposed to music created by Artificial Intelligence (AI)
    f. I can tell the difference between music created by humans as opposed to music created by AI
    g. Music composed by a human is able to convey emotion
3. What are your opinions on music composed by Artificial Intelligence (AI)? (Open question)
4. How many hours do you listen to the following per week?
    a. Classical piano music composed by humans
    b. Classical piano music composed by AI
    c. Music composed by AI
5. Do you think that music composed by Artificial Intelliegence (AI) is capable of conveying the same level of emotion as music composed by a human and why/why not? (open question)
6. Do you believe that classical piano music composed by AI is different to other music composed by AI (and if so, what)? (open question)
7. Do you think that the instruments AI chooses to compose with impact the emotional intensity of the music and why? (open question)

Section 6

1. Please identify which extracts you believe to have been composed by Artificial Intelligence (AI) and which by humans
2. Why have you made this selection?
3. Please answer the following questions (5 point Likert scale: Definitely not, probably not, might or might not, probably yes, definitely yes)
    a. Knowing the nature of the composers of the above extracts, do you still feel the same way about the music?
    b. Do you still feel the same emotional intensity towards the pieces created by AI as before you knew this information?
    c. DO you feel that you have been lied to now that you know 2 of the pieces were made by AI?
4. Which extract do you feel you reacted with the strongest emotional intensity to? (1 being the most, 3 being the least)
5. Has your opinion changed on the pieces now knowing that they have been composed by AI and why/why not? (Open question)

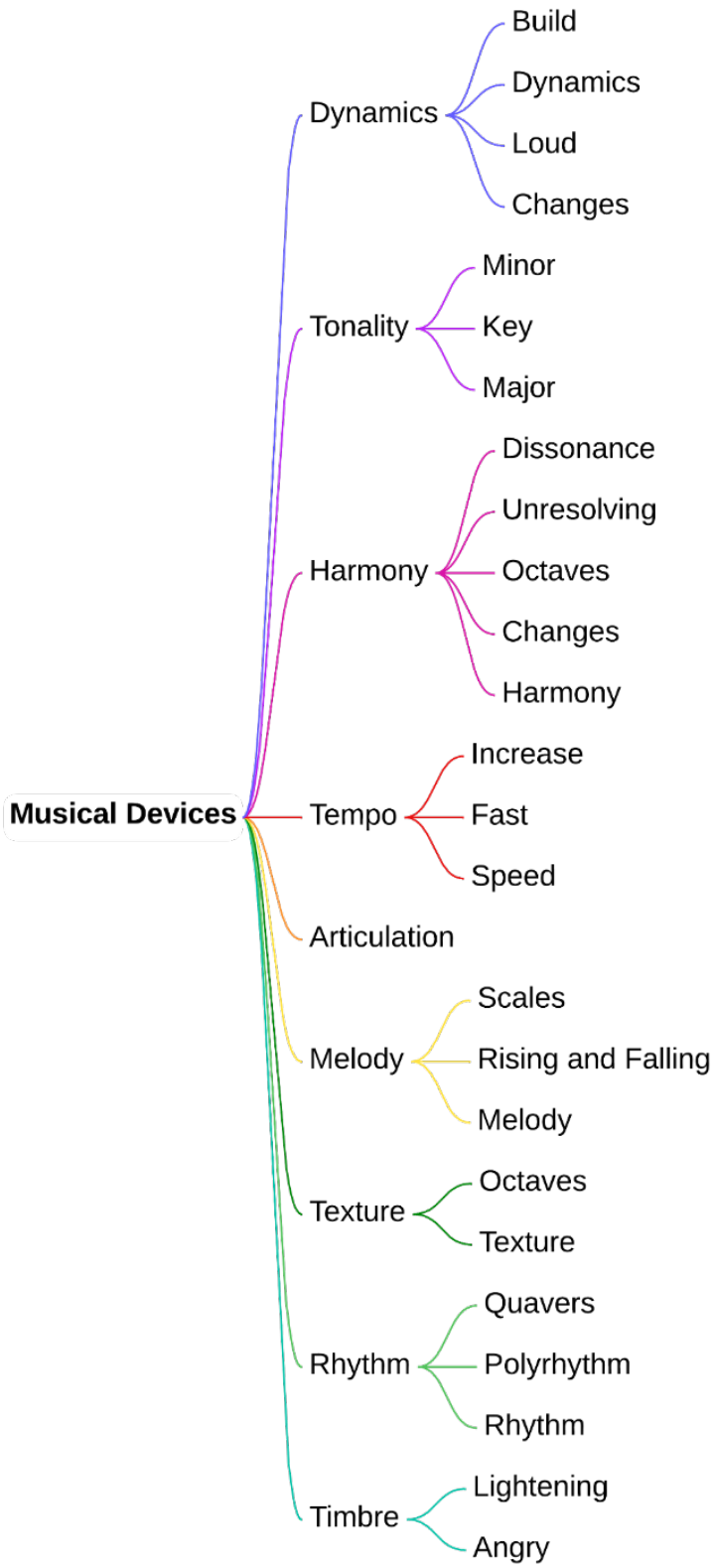
Appendix C: Python Code for the Box plots evaluating the strength of emotions felt by non-musicians and musicians when listening to music composed by a human.

box plot graph code and graph

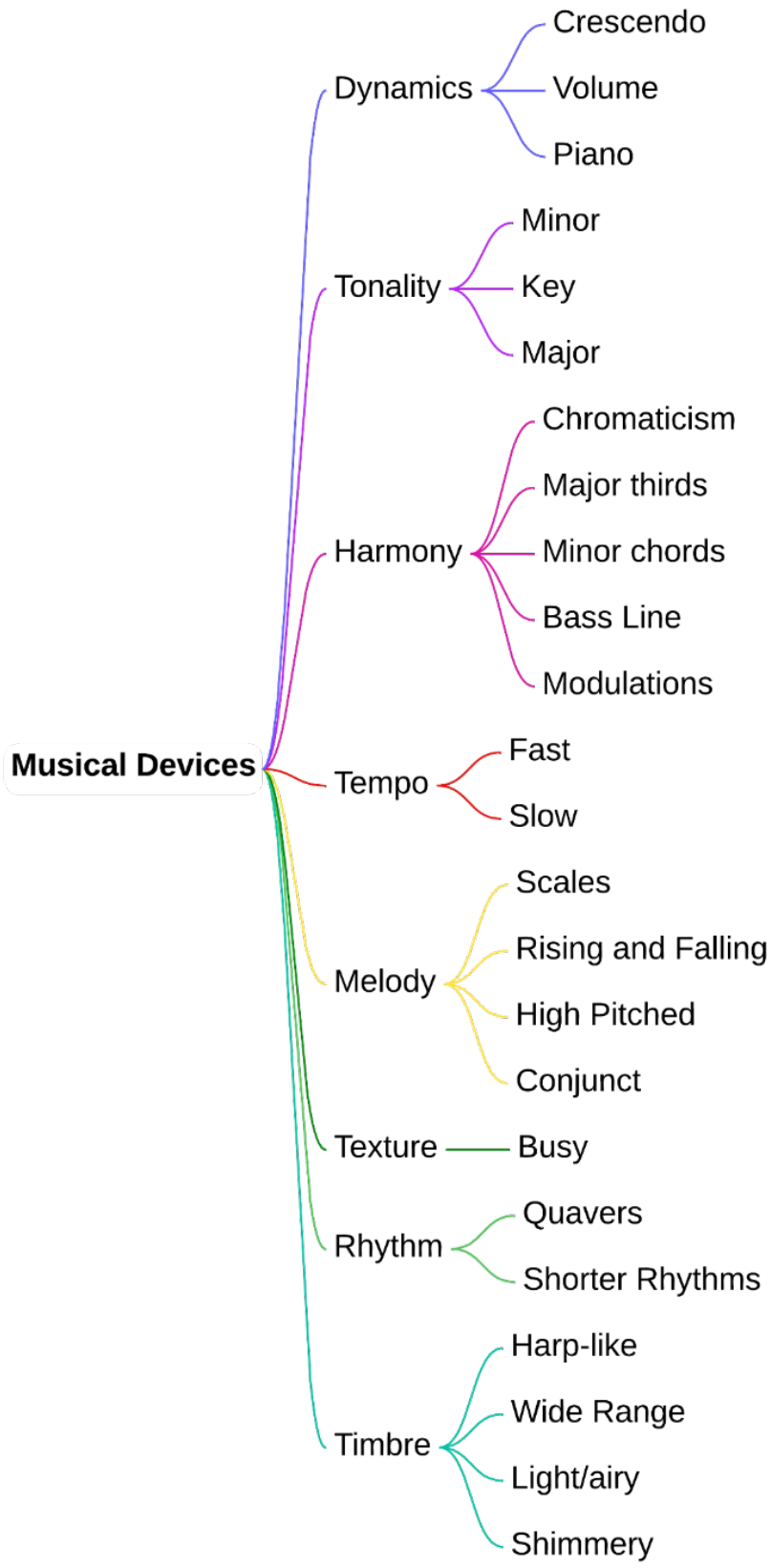
```
import matplotlib.pyplot as plt

import numpy as np


NM = np.array([3, 4, 2, 4, 3, 4, 4, 2, 5, 2, 4, 2, 3])

M = np.array([5, 3, 3, 4, 4, 5, 5, 2, 5, 5, 5, 3, 1, 3, 5, 4, 3, 4, 2, 3, 4, 4, 4, 4])


fig, ax = plt.subplots()


boxplot_NM = ax.boxplot(NM, positions=[1], patch_artist=True, boxprops=dict(facecolor='lightblue', color='blue'), widths=0.5)


plt.setp(boxplot_NM['whiskers'], color='darkblue')

plt.setp(boxplot_NM['caps'], color='darkblue')


boxplot_M = ax.boxplot(M, positions=[2], patch_artist=True, boxprops=dict(facecolor='lightgreen', color='green'),widths=0.5)


plt.setp(boxplot_M['whiskers'], color='green')

plt.setp(boxplot_M['caps'], color='green')


ax.set_xticks([1, 2])

ax.set_xticklabels(['Non-Musicians', 'Musicians'])


ax.set_xlabel('Groups')

ax.set_ylabel('Response based on Likert Scale')

ax.set_title('Boxplots of Musicians and Non-Musicians Responses')


plt.show()
```

Appendix D: Coding of the qualitative data collected on musical devices used by the composers identified by the participants.

Extract 1 (AI)



Extract 2 (AI)


Musical Devices
Dynamics
Crescendo
Volume
Piano
Tonality
Minor
Key
Major
Harmony
Chromaticism
Major thirds
Minor chords
Bass Line
Modulations
Tempo
Fast
Slow
Melody
Scales
Rising and Falling
High Pitched
Conjunct
Texture
Busy
Rhythm
Quavers
Shorter Rhythms
Timbre
Harp-like
Wide Range
Light/airy
Shimmery

Extract 3 (Human Composed)