

Attitude towards and evaluation of computer-generated music in music listeners and musicians.

Anna Aljanaki
University of Tartu
Institute of Computer Science
Tartu, Estonia
anna.aljanaki@ut.ee

Abstract

Despite accepting the involvement of computers in such important areas as finance, self-driving and surveillance, society still struggles with the concept of computer as an originator of art, and has difficulties with defining value and place of computer-generated art. In this paper we will explore the questions of value of computer music from the perspectives of music listeners and professional musicians with different backgrounds in a series of open-ended interviews. Several solutions for the problem of lack of intentional agency are proposed.

1 Introduction

The emergence of computer creativity has challenged our understanding of the value and goals of creativity in general, including human creativity. Within just a few decades, the importance of information technologies and artificial intelligence in our society has drastically increased, affecting the way we take important decisions, communicate, and stay informed. The society seemed to accept that change without major resistance, accepting the role of a computer as a tool and an assistant. This acceptance happens even when the stakes are high (e.g., in autonomous driving) Wang et al. [2020].

However, there is a general reluctance to accept the competence of a computer in artistic domains: music, literature and visual art, and there is bias against computer-produced art (Moffat and Kelly [2006]). In Ragot et al. [2020], a sample of 565 participants evaluated paintings created by humans and AI on four criteria: liking, perceived beauty, novelty, and meaning. Paintings perceived as being drawn by human were evaluated significantly more highly than those perceived as being made by AI, irregardless of whether they were actually produced by human or AI. In Knotts and Collins [2020], surveying musicians revealed that there is general scepticism about the current state of music AI and that its improvement in order to replicate human creativity is an undesirable goal. It is of course very understandable that musicians seek a creative partnership with technology, and do not wish to be replaced by it.

Nonetheless, a lot of effort in computer music creativity is currently directed precisely towards the goal of reaching human level in music composition (see Fernández and Vico [2013], Ji et al. [2020]). In order for the society in general, and musical community in particular, to benefit from these efforts in the same way that we have already benefited from the information technologies in other areas, we need to understand what is computer generated music value proposition, whether it is different from that of music composed by human composers. Can the same value system that humans use to experience music be applied to music made by a computer? Can the same evaluation criteria be applied to both human composed and computer generated music? Which factors are behind the bias against computer generated art? These questions are under explored, and we think that open-ended approach can provide the answers in the best way. In this paper we will explore these topics through interviews and open-ended questions.

2 Related work

In this part we will very succinctly cover existing research on evaluation of computer generated music and acceptance of such technology among the musicians and general public.

2.1 Computer music evaluation

Evaluation of computer creativity usually involves it being compared to human creativity. Algorithmic composition systems, whether based on machine learning or on rule-based systems, are usually trained by learning from examples of music produced by human composers (for review, see Fernández and Vico [2013], Ji et al. [2020]), and, in case of SOTA systems, artificial neural networks, their loss metric is related to the ability to predict the next symbol as similarly as possible to an existing musical piece written by a human composer (e.g., Huang et al. [2018], Dhariwal et al. [2020]), with the goal of composing novel music that is as close as possible to the golden standard, e.g., human composed music. Although this trivial and well-established design of music generation systems and their evaluation may seem to be the only sensible option, this is only so if we accept for a fact that computer generated music needs be similar to human composed music. However, the goal might as well be to generate music as dissimilar as possible from currently existing music, but still having aesthetic or utilitarian value. In other domains, it has been shown that letting computer algorithms solve the problem can lead to counter-intuitive but efficient solutions that are difficult for humans to find (Keane and Brown [1996]).

One of the most straightforward evaluation methods is a test, where the subject needs to guess whether a piece was written by a human or by a computer. More sophisticated and rigorous systems might compare the statistical properties of computer generated music with the properties found in a corpus of human written music (Ens and Pasquier [2018], Yin et al. [2021]). E.g., in Sturm and Ben-Tal [2017], Bob Sturm and Oded Ben-Tal compare the statistics (distributions of pitches, metres, modes) of generated transcriptions with the statistics displayed by the training set.

Evaluation using human listeners is of paramount importance for creative output. It might involve listening to music and providing subjective rating, or measuring the galvanic skin response in order to determine whether algorithmic music is achieving its goal of inducing a desired emotional response in the listener (Daly et al. [2015]). For more detail on generative music evaluation methods, we refer the reader to a comprehensive survey Ji et al. [2020].

2.2 Reception of computer music

Attitude towards computer generated music is an understudied topic. In Simoni [2018], reception theory is applied to the reaction of three subjects to algorithmic music using a cognitive-affective model with a goal of finding whether more exposure to computer-generated music increases its acceptance. In (Chamberlain et al. [2018], Ragot et al. [2020]) the evidence of bias against computer-generated art is shown. In Tigre Moura and Maw [2021], several hypothesis regarding perception of computer generated music were tested. It was found that music listeners believe it is less acceptable for AI-composed music to be used in high involvement contexts when compared to music professionals. Both in this study and in Pasquier et al. [2016] and Eigenfeldt et al. [2012], the bias against computer generated music was investigated, but was not confirmed experimentally.

3 Methods

We collect data from two samples of participants. In the first part, we analyze the reactions of non-musicians to computer generated and human composed music. In the second section, four musicians with different backgrounds are interviewed on their perception of computer generated music.

3.1 Musical stimuli

We picked two musical pieces for the purposes of fueling discussion about the value of computer generated music. Both are solo piano pieces. The first piece is named "Piano Distance" and was composed by a Japanese composer Toru Takemitsu in 1961. It lacks stable metre and conventional

Table 1: Results of the evaluation by non-musicians

Musical piece	Perceived as human composed	Perceived as computer music
"Piano Distance"	5.18 ± 2.65	4.50 ± 2.39
PerformanceRNN	6.50 ± 2.00	6.20 ± 1.31

harmonic hierarchy, and sounds rather dissonant and unconventional (akin to serialist music). The second piece was generated by a PerformanceRNN (Simon and Oore [2017]). It is much more conventionally sounding, reminiscent of romantic period in classical music the network was trained on, though it displays (not in a very noticeable way) the problems characteristic of simple music generation methods with a small receptive field: some pitches are out of harmonic context, the rhythm is sometimes broken.

3.2 Questioning the non-musicians

Non-musicians are randomly split into two groups. First group listens to the two piano pieces in sequence and is correctly told which one is composed by a human and which is produced by an algorithm. Second group is told that a piece by Takemitsu is actually composed by a computer and a piece by the performanceRNN is composed by a human. Both groups are asked to evaluate both pieces on a scale from 1 to 10 and describe their reaction to them and are asked whether they think they should apply the same criteria when evaluating human-composed and computer-generated music.

3.3 Interviews with musicians

During the interview, musicians are asked four questions: 1. Imagine that you are in a jury of a music composition contest. Which criteria would you apply when evaluating human composed and algorithm composed music? The same two pieces above are used in order to facilitate the discussion. 2. Do you think we should apply the same criteria during the contest to computers and humans? 3. Is music composition a goal oriented process and should we keep a goal in mind when evaluating?

The discussion is not limited to these questions and any tangential topics are discussed.

4 Results

In this section we will summarize the findings.

4.1 Questioning the non-musicians

In each group, there were 10 participants (a total sample of 20), recruited based on their interest in music. In Table 1, we see the ratings both groups gave to each piece on a scale from 1 to 10. Overall, when pieces were perceived as composed by a human, the ratings were slightly higher, though the difference was not statistically significant on a t-test in either case. Almost unanimously (18/20) the participants expressed an opinion that the same evaluation criteria should be applied to both computer generated and human composed music.

However, the participants' actual reaction contained some reservations as to the possibility of using the same criteria. Several participants noted, that music written by a human has a creative intent and expresses emotions, and computer music does not (i.e., lack of intentional agency Stein [1974]): *"For me, it is of paramount value to which extent the composer is capable of expressing their deeply personal emotional experience in their composition."*, *"The music affects me most when I sense a similarity in what the composer has been experiencing and what I have experienced in my life."*, *"My criteria are intentionality and story-telling in music"*, *"When the music is written by a human, I am willing to search deeper for the meaning in it."*, *"This piece sounds shallow and lacks ideas, but I might be influenced by the knowledge it is computer generated, because I am not expecting to find deep ideas in computer generated music."*

Secondly, half of the participants (10/20) have named subjective enjoyment, pleasure as one of the main criteria in music evaluation, and said they did not care whether music was computer-generated or not provided it was enjoyable.

Third most important criteria brought forth by 12 out of 20 participants was that music needs to conform with the conventional (apparently, for Western music) standards in terms of its melodic, harmonic and rhythmic properties.

5 Interviews with musicians

Similarly to Knotts and Collins [2020], all the musicians in our study thought that computer creativity should augment human one, not replace it. For the sake of brevity, we will compress the dialogue format of an interview into monologues, keeping only the most interesting parts.

Participant 1, electronic music composer: "Music composition is a process driven by an internal impetus to express an emotion or to tell a story. The computer lacks this impetus, but this is a possibility for human and computer creativity to augment each other. Human should act as an external stimulus to the computer to start creating, by providing this necessary initial impetus."

Participant 2, pianist and violinist with extensive orchestral experience: "There are two sides to music - composition and performance, and they need to tell the same story in a cohesive way. In computer music this is very often not the case. In Takemitsu's piece I could immediately see how dynamics, harmony, even room acoustics - everything is interrelated and serves to convey the same idea of restlessness and anguish. This interconnectedness of all of the expressive vocabulary - choreography, dynamics, dialogue - is what we need to strive for in computer generated music. In my opinion, we do not need a computer to generate the 42nd symphony of Mozart. Our goal should be to enrich the experience for humans composing and performing music, because the ultimate pleasure is in interacting with music actively and we shouldn't be giving all the active roles away to the computer. As for the lack of intentional agency in computers, music based solely on statistical principles instead of emotional motivation is nothing new. Think Xenakis. Computer music does not need to imitate human music in every way, we might find new meaning in music based completely on statistical principles, which might well be alien to human nature, but still have value."

Participant 3, piano teacher: "When evaluating computer-generated music, we can take the musical goal into account. Where does the goal come from? It is defined by a human operating a computer. Human defines what a composition needs to express, and then they can subjectively evaluate how well the algorithm was able to express the intention it was given."

Participant 4, semi-professional guitar player: "Computer music sounds like a draft, a collection of disjointed ideas. In my opinion, it does not need to be assembled by a computer, human can take inspiration from computer generated ideas and put them together in a meaningful way. Only human touch will help these abandoned ideas obtain meaning. In general, I do not mind mistakes, pauses, little details such as scratching of the finger on the string, it brings music to life for me. There is something artificial and unpleasant about perfectly sounding music."

6 Conclusion

Professional musicians (PM) and music listeners (ML) see the role of computers differently. ML would like to see computers passing the musical Turing test as soon as possible, whereas PM would like computers to become their creative partners. Both PM and ML find that computers need to be given external motivation to compose emotionally compelling music. Three possible solutions are proposed: using expressive intent fulfillment as an evaluation metric, using a human composer to combine raw material coming from the computer with creative intent, or abandoning creative intent altogether to explore the possibility of computer music that is unlike human music.

Acknowledgments and Disclosure of Funding

This research has been financed by European Social Fund via "ICT programme" measure

References

- R. Chamberlain, C. Mullin, B. Scheerlinck, and J. Wagemans. Putting the art in artificial: Aesthetic responses to computer-generated art. 12(2):177–192, 2018.
- I. Daly, A. Malik, J. Weaver, F. Hwang, S. J. Nasuto, D. Williams, A. Kirke, and E. Miranda. Towards human-computer music interaction: Evaluation of an affectively-driven music generator via galvanic skin response measures. In *2015 7th Computer Science and Electronic Engineering Conference (CEECE)*, pages 87–92, 2015. doi: 10.1109/CEECE.2015.7332705.
- P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever. Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*, 2020.
- A. Eigenfeldt, A. Burnett, and P. Pasquier. Evaluating musical metacreation in a live performance context. pages 140–144, 01 2012.
- J. Ens and P. Pasquier. A cross-domain analytic evaluation methodology for style imitation. In *Proc. Int. Conf. Computational Creativity*, Salamanca, Spain, June 2018.
- J. D. Fernández and F. Vico. Ai methods in algorithmic composition: A comprehensive survey. *J. Artif. Int. Res.*, 48(1):513–582, oct 2013. ISSN 1076-9757.
- C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. M. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck. Music transformer. *arXiv: Learning*, 2018.
- S. Ji, J. Luo, and X. Yang. A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions. *arXiv preprint arXiv:2011.06801*, 2020.
- A. J. Keane and S. Brown. The design of a satellite boom with enhanced vibration performance using genetic algorithm techniques. In *Conference on adaptive computing in engineering design and control*, page 107–113, 1996.
- S. Knotts and N. Collins. A survey on the uptake of music ai software. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 499–504, 2020.
- D. Moffat and M. Kelly. An investigation into people’s bias against computational creativity in music composition. In *Proc. Workshop on Computational Creativity*, 2006.
- P. Pasquier, A. Burnett, N. G. Thomas, J. B. Maxwell, A. Eigenfeldt, and T. Loughin. Investigating listener bias against musical metacreativity. In *Proc. ICCCI*, 2016.
- M. Ragot, N. Martin, and S. Cojean. Ai-generated vs. human artworks. a perception bias towards artificial intelligence? In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA ’20, page 1–10. Association for Computing Machinery, 2020.
- I. Simon and S. Oore. Performance rnn: Generating music with expressive timing and dynamics. <https://magenta.tensorflow.org/performance-rnn>, 2017.
- M. Simoni. The audience reception of algorithmic music. In *The Oxford Handbook of Algorithmic Music*. 2018.
- M. Stein. Stimulating creativity. 1974.
- B. L. Sturm and O. Ben-Tal. Taking the models back to music practice: Evaluating generative transcription models built using deep learning. *Journal of Creative Music Systems*, 2(1), 2017.
- F. Tigre Moura and C. Maw. Artificial intelligence became beethoven: how do listeners and music professionals perceive artificially composed music? *Journal of Consumer Marketing*, 38(2):137–146, 2021.
- S. Wang, Z. Jiang, R. B. Noland, and A. S. Mondschein. Attitudes towards privately-owned and shared autonomous vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 72:297–306, 2020.
- Z. Yin, F. Reuben, S. Stepney, and T. Collins. “a good algorithm does not steal – it imitates”: The originality report as a means of measuring when a music generation algorithm copies too much. In *Proc. EvoMUSART*, 2021.