

A Multi-modal Perspective for the Artistic Evaluation of Robotic Dance Performances

Luca Giuliani, Allegra De Filippo, Andrea Borghesi, Paola Mello and Michela Milano

Department of Computer Science and Engineering, University of Bologna

Abstract

In recent years, the interplay between creativity and Artificial Intelligence (AI) has been gaining more and more attention from the research community. Within this vast area, humanoid robots have been successfully used in artistic research areas, and many works have studied and implemented systems for robotic dance. However, only few works take into account the human evaluation of these artistic outputs. For this aim, we start from a recent work focused on defining criteria for the evaluation of robotic dance performances, and we analyze a further crucial aspect such as the need for a *multi-modal perspective*: the musical element needs to blend seamlessly and organically with the choreography, in accordance to the judgments expressed by human evaluators. Based on the analysis of results, musical elements emerged as having a large impact on the artistic evaluation. For this reason, the final purpose of this work is twofold: we would like both to explore new creative paths where humans ideas can be broadened by the use of AI software, but also to help both human choreographers and AI algorithms to create dance performances and music with a major impact on the audience.

Keywords

Artistic Evaluation, Robotic Dance Choreography, Multi-Modal Creativity, Educational AI

1. Introduction

The adoption of humanoid robots in artistic research areas showed several successful applications, ranging from the generation of artificially generated music, images, and even tentative literature [1, 2]. Dance choreography is an area where the potential application of artificial intelligence has raised interest in recent years [3, 4], both choreographies for human performers and robotic ones, which have been showed to be well suited also to educational purposes [5]. The creation of *artificial art* is a new and expanding field where multiple challenges remain wide open, ranging from the actual acceptance of artificial creations as artistic products [6] to the evaluation of the quality of the generated works of art [7]. For instance, it is usually the case that humans are not kept in the creative loop, especially when it comes to the evaluation of these artistic performances.


We tackle this challenge and address the complex task of artistic evaluation by trying to extract some human-employed criteria for robotic dance performances evaluation. This work is the continuation of a previous research on the artistic evaluation of humanoid robotic dances [8]. Differently from the majority of studies in the field, which focuses on the automation of robotic dance performances [9, 10, 11], we want to involve humans in both the creation and

CREAI 2022, Workshop on Artificial Intelligence and Creativity, Nov.28– Dec.02, 2022, Udine, Italy

✉ luca.giuliani13@unibo.it (L. Giuliani)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

evaluation of the artistic composition (i.e., the robotic dance choreographies). In order to do that, in [8] we have organized a competition aimed at the development of robotic choreographies which has been held within the context of a Master Degree course on Fundamentals of Artificial Intelligence¹ at the University of Bologna (Italy). On a voluntary basis, students could decide to participate in the competition by developing their own algorithm for building the choreography; participants were also asked to fill a questionnaire to evaluate each of the proposed performances and a winner was decided based on the responses. Alongside, answers to the questionnaire are used to build a public dataset which maps specific features of each choreography to the judgement provided by the audience. We then trained different Machine Learning (ML) models to learn relationship between choreography features and evaluation score; we have also inspected these models to rank the importance of each feature, to help both human choreographers and AI algorithms to understand which aspects of a performance are more likely to increase its appreciation from the audience.

This analysis revealed that one of the aspects with greater impact on the human evaluation is the musical element incorporated in the dance. Hence, we decided to explore new directions involving *multi-modal* creativity. The idea is to propose that the creative process of developing a robotic choreography during the challenge should be effectively paired with music generation techniques, based on the feature importance analysis. In this way, both human artists and AI algorithms would be able to better integrate the music within the performance, thus achieving more impact on the audience.

2. Related Work

2.1. Humanoid Robotic Dance Automation

Humanoid robots have been employed in a variety of social tasks including, but not limited to, healthcare [12] and education [13, 14]. Likewise, dance is considered to be an important part of our social activities, therefore many research works on the development of robotic dance automation started to emerge aiming both at entertainment and human-machine interaction. Thanks to their ability to move in a physical environment and to mimic human-like actions, humanoid robots such as NAO² are the perfect candidate for these tasks. However, among the collection of works that have been proposed about robotic dances [10, 11, 15, 16, 17], the majority of them focused solely on the implementation of balanced and coordinated choreographies, without taking into account how a human audience would eventually evaluate it.

2.2. Music Generation Techniques

As revealed by [8], the music element has a great impact in choreography appreciation, suggesting that creating musical score in accordance to the robotic dance is crucial. However, to the best of our knowledge, no works has explicitly addressed this challenge. Moreover, automatic music generation is still in its infancy, as it is a very complex task which involves multiple experts with different roles, starting from an initial phase of composition up to the final mixing

¹<https://www.unibo.it/en/teaching/course-unit-catalogue/course-unit/2022/446566>

²<https://www.aldebaran.com/en/nao>

and mastering step. See for instance, Ji et al. [18] who propose a taxonomy based on three incremental levels: the score generation, the performance generation, and the audio generation.

A wide plethora of different techniques ranging from Genetic Algorithms [19, 20] to Swarm Intelligence [21], Markov Chains [22], and Artificial Neural Networks [23, 24] have been successfully applied in the past years both to the most common unconstrained generative task and to some of its variations – e.g., melody generation based on a given harmonic progression [25, 26] and vice versa [27], but also real-time improvisation along with a human performer [28]. However, mainly due to the higher complexity of audio signals with respect to scores, as well as a lack of proper audio generation datasets, most of the research have been focused on the top level only rather than developing a full, end-to-end approach.

2.3. Artistic Evaluation of Performances

Due to its subjective nature, the evaluation of art works and performances is known to be a very complex task. As pointed out in [29], being there no objective methodology to compare the results of models involved in the production of new art pieces, the vast majority of the approaches is based on surveying human responses with ad-hoc questions and scales.

As a matter of fact, various measurement tools have been proposed in the last decades in order to assess the quality of dance performances [30, 31]; still, most of them are specifically designed to evaluate human skills rather than mechanical subjects. An exception can be found in recent works by Gemeinboeck [32, 33, 34], which specifically focus on the perception of robotic body and movements; similarly, the Likert-like questionnaire proposed in [35] intentionally addresses the evaluation of robotic dances. This last questionnaire in particular, integrated with additional questions regarding the use of the surrounding space and the overall theatricality of the choreography, as well as its level of human reproducibility, was used in our previous work [8] in order to assess the importance of various aspects of a robotic dance performance in its success among social groups coming from different academic backgrounds.

Similarly, as regards evaluation techniques, the field of automatic music generation is dominated by listening tests followed by Likert-like questionnaires about the pleasantness and naturalness of the proposed audio track [36, 37, 38]. An interesting usage of this kind of evaluation can be found in [29], where the authors gathered data from human annotation of more than a thousand computer-generated records based on a 1-to-5 scale and eventually used this data to build a classifier aimed at helping the generative model to filter out unpleasant tracks before returning them to the users. For a broader perspective on machine-generated music evaluation, [18] provides an extensive review of the most important techniques.

3. Robotic Choreography Creation

As illustrated in [8], during the last years, we proposed the idea of a *challenge aimed at the creation of robotic choreographies*. This was motivated by the belief that a synergy between technical and creative areas, such as those of AI and dance performances respectively, would increase students' participation as well as providing them with new perspectives and transversal skills. The competition was based on voluntary participation, with students covering both the

roles of choreography creators and audience; the winning team was elected during a final voting day.

Participants of the challenge are asked to work in group on the development of a choreography within a simulated NAO robot environment. To encourage diversification among groups, no limitation about viable technique is given, although the larger part of performances were built upon those seen during the course, i.e., Planning, Search Strategy, and Constraints. The whole set of competition rules and guidelines has been described in [8]; in any case, what it is expected from students is to implement a system which, given a fixed initial state, a goal, and a set of requirements – e.g., duration of the choreography, mandatory positions, etc. –, is able to return a sequence of actions and transitions that respect the balancing and coordination constraints of the robot, thus avoiding any kind of possible inconsistency between subsequent positions.

3.1. Multi-modal Creation Perspective

So far, the challenge focused exclusively on its dance component. Indeed, no requirements on background music was given, leaving each group the freedom to select any piece of music suitable for the choreography – provided, of course, that it respected the total time limit. However, we plan to start employing a new modality, that is asking teams to generate their background music using state-of-the-art AI-based tools. This has the dual aim of exploring multi-modal art creation – thus increasing the levels of expressive freedom of students – while, simultaneously, allowing team members to freely synchronize their choreography with the underlying music beat, an aspect which has been found to be positively correlated with a higher aesthetic appeal of gymnastics and acrobatics performances [39, 40, 41].

Thanks to the emergence of recent neural architectures based on improved attention mechanisms, multi-modal inputs started to be employed in all kind of generative tasks [42], including robotic dance. [43] integrates visual and non-visual information to increase robots' self-awareness on aesthetics movements, similarly to what a human would do in front of a mirror; [44] exploits an integrated input source of past motion and music in order to better condition the generative process; [45] focuses on the integration of music context to help the model create a performance which correctly matches the musical elements. However, to the best of our knowledge, nobody within the field of robotic dance automation has ever made attempts to use this variety of input sources neither for an educational purpose nor to develop new forms of human-machine interactions where the AI software does not perform end-to-end generative tasks, but rather it serves as an enabling tool to improve human creative skills. Indeed, with this new challenge modalities, we expect students to form larger groups where members can focus on the artistic task they prefer, hopefully allowing them to increase both their creative and teamwork abilities.

4. Robotic Choreography Evaluation

As previously stated in 2.3, the evaluation of art works and performances is a complex task. This entails that dance researchers and choreographers lack of a formal knowledge that might help them to create more aesthetically pleasing performances. For this reason, two different audiences – one with a scientific (*S*) and one an artistic (*A*) background – were asked to evaluate

the challenge participants' works respectively to both the structure of the choreography and the behaviour of the robot performer [8]. We collected this data into two different datasets, one per type of audience, and eventually analyzed it with machine learning methods in order to extract the features that best predicted a success for the performance. The two datasets are publicly available³; both of them are composed by 403 records – 31 attendees to whom 13 choreographies were submitted –, each one encoding 20 different features concerning various aspects of a choreography such as, e.g., its duration, the number of movements it contains, the background music genre, the AI technique which was used, etc.

4.1. Feature Importance Analysis

The data analysis phase started with a straight-forward prediction task aimed at assessing whether or not machine learning models would have been able to extract informative patterns from the input features. Four different symbolic and explanatory models were taken into account in order to handle both the low amount of data and the open issue of interpretability in sub-symbolic AI. Among all, Gradient Boosting performed the best, thus it came as a natural choice for the subsequent task of feature importance analysis.

This analysis highlighted interesting trends. Some features such as time duration, music bpm and music genre showed a higher impact both in the *S* and in the *A* datasets (see 1), confirming our idea that background music strongly influences the pleasantness of a robotic dance performance. Similarly, other music-based features exhibited not only the significance of audio information in the overall evaluation, but also how this significance differs between the two social groups, with certain music genres influencing the scientific group more than the artistic one, and vice versa. Apart from those, even features directly assessing choreographic elements confirmed that members of the two audiences systematically leaned on certain aspects of the performance rather than others when judging its aesthetic quality. A more exhaustive analysis of the audience responses, along with a more detailed description of the evaluation questionnaire and dataset structure, can be found in [8].

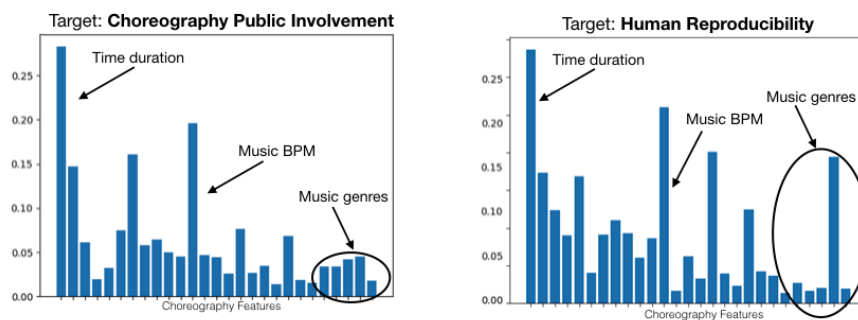


Figure 1: This figure illustrates the feature importance analysis for two target analyzed in [8], respectively the choreography public involvement and the human reproducibility as an average of both the analyzed audiences (i.e., *A* and *S*).

³<https://github.com/ProjectsAI/NAOPlanningChallenge/tree/main/datasets>

4.2. Multi-modal Evaluation Perspective

As already mentioned in 2.2, fully automated music generation tools seem to be far from being achieved in a short time span, especially if our aim is to keep the human in the loop. Nonetheless, not only we can try to experiment with some existing techniques for the sake of educational purposes and based on the previous feature importance analysis, but also we should design proper ways to assess the artistic quality of both the generated music and the overall multi-modal artistic product.

As in any other creative task, most of the evaluation methodologies for automated music generation are based on subjective judgements [18], thus they require human involvement. However, annotators are rarely asked to express an aesthetic opinion on the track, but rather they are simply required to discern machine-generated from human-generated pieces of music. In our case, the technical aspects of the song must be given for granted, with the audience being explicitly asked to assess its artistic traits. So far, works regarding the aesthetic evaluation of music mainly focused on the music performance [46], while in our case the performance is strictly choreographic and does not comprise any live musical act; quite the opposite, in our case we do not even care about the songs on their own, but rather on how well do they conform to the choreographic elements in terms of rhythm and mood. For this reason, we will need to extend the set of questionnaire topics taking into account this specific integration aspects between dance and music. Previous results demonstrate the strength of integrating complementary sources of information in different tasks such as artistic recognition [47] and creation [48], and also indicate the potential of applying multi-modal approaches within specific research areas like the artistic evaluation one.

5. Conclusions

The role of AI in creative activities is becoming more and more central. Within the field of robotic dance performance, this is manifested by the large number of studies on computational dance automation and humanoid robotic dance. We claim that this research area can particularly benefit from AI when humans are kept in the loop, both in the creative and in the evaluation process. Furthermore, starting from one of our previous works aimed at discovering correlations between certain features of a choreography and its success among a human audience, we address the issue of multi-modality. Indeed, since musical elements emerged as having a large impact from our feature importance analysis, we argue that the creative aspect must comprise music generation as well. Even though very few AI-based tools exist for assisting non-professionals to create musical tracks from scratch, we provide some future directions on how to build such integration and, eventually, how to evaluate it with appropriate questionnaires. The final purpose is twofold, since we would like both to explore new creative paths where humans ideas can be broadened by the use of AI software and to foster awareness about which aspects of a choreography correlate with a major impact on the audience.

Acknowledgements

This work has been partially supported by European ICT-48-2020 Project TAILOR (g.a. 952215). We thank the performing robots group⁴.

References

- [1] R. López de Mántaras, *Artificial intelligence and the arts: Toward computational creativity* (2016).
- [2] M. Mazzone, A. Elgammal, *Art, creativity, and the potential of artificial intelligence*, in: *Arts*, volume 8, MDPI, 2019, p. 26.
- [3] F. Sagasti, *Information technology and the arts: the evolution of computer choreography during the last half century*, *Dance Chronicle* 42 (2019) 1–52.
- [4] A. Plone, *The influence of artificial intelligence in dance choreography* (2019).
- [5] S. Anwar, N. A. Bascou, M. Menekse, A. Kardgar, *A systematic review of studies on educational robotics*, *Journal of Pre-College Engineering Education Research (J-PEER)* 9 (2019) 2.
- [6] J.-W. Hong, N. M. Curran, *Artificial intelligence, artists, and art: attitudes toward art-work produced by humans vs. artificial intelligence*, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 15 (2019) 1–16.
- [7] Y. Wang, H. Ma, *The value evaluation of artificial intelligence works of art*, in: *2019 International Joint Conference on Information, Media and Engineering (IJCIME)*, IEEE, 2019, pp. 445–449.
- [8] A. De Filippo, P. Mello, M. Milano, *Do you like dancing robots? AI can tell you why*, in: *PAIS 2022*, IOS Press, 2022. URL: <https://doi.org/10.3233/faia220064>. doi:10.3233/faia220064.
- [9] A. Manfrè, I. Infantino, F. Vella, S. Gaglio, *An automatic system for humanoid dance creation*, *Biologically Inspired Cognitive Architectures* 15 (2016) 1–9. URL: <https://doi.org/10.1016/j.bica.2015.09.009>. doi:10.1016/j.bica.2015.09.009.
- [10] O. E. Ramos, N. Mansard, O. Stasse, C. Benazeth, S. Hak, L. Saab, *Dancing humanoid robots: Systematic use of OSID to compute dynamically consistent movements following a motion capture pattern*, *IEEE Robotics & Automation Magazine* 22 (2015) 16–26. URL: <https://doi.org/10.1109/mra.2015.2415048>. doi:10.1109/mra.2015.2415048.
- [11] K. Shinozaki, A. Iwatani, R. Nakatsu, *Concept and construction of a dance robot system*, in: *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts - DIMEA '07*, ACM Press, 2007. URL: <https://doi.org/10.1145/1306813.1306848>. doi:10.1145/1306813.1306848.
- [12] A. Joseph, B. Christian, A. A. Abiodun, F. Oyawale, *A review on humanoid robotics in healthcare*, *MATEC Web of Conferences* 153 (2018) 02004. URL: <https://doi.org/10.1051/mateconf/201815302004>. doi:10.1051/mateconf/201815302004.
- [13] O. Mubin, C. J. Stevens, S. Shahid, A. A. Mahmud, J.-J. Dong, *A REVIEW OF THE APPLICABILITY OF ROBOTS IN EDUCATION*, *Technology for Education and Learning* 1 (2013).

⁴<https://site.unibo.it/performingrobots/en>

URL: <https://doi.org/10.2316%2Fjournal.209.2013.1.209-0015>. doi:10.2316/journal.209.2013.1.209-0015.

- [14] A. K. Pandey, R. Gelin, Humanoid robots in education: A short review, in: *Humanoid Robotics: A Reference*, Springer Netherlands, 2018, pp. 2617–2632. URL: https://doi.org/10.1007%2F978-94-007-6046-2_113. doi:10.1007/978-94-007-6046-2_113.
- [15] K. Shinozaki, A. Iwatani, R. Nakatsu, Construction and evaluation of a robot dance system, in: *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, 2008. URL: <https://doi.org/10.1109%2Froman.2008.4600693>. doi:10.1109/roman.2008.4600693.
- [16] D. Grunberg, R. Ellenberg, Y. Kim, P. Oh, Creating an autonomous dancing robot, in: *Proceedings of the 2009 International Conference on Hybrid Information Technology - ICHIT '09*, ACM Press, 2009. URL: <https://doi.org/10.1145%2F1644993.1645035>. doi:10.1145/1644993.1645035.
- [17] C. Angulo, J. Comas, D. Pardo, Aibo JukeBox – a robot dance interactive experience, in: *Advances in Computational Intelligence*, Springer Berlin Heidelberg, 2011, pp. 605–612. URL: https://doi.org/10.1007%2F978-3-642-21498-1_76. doi:10.1007/978-3-642-21498-1_76.
- [18] S. Ji, J. Luo, X. Yang, A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions, *CoRR abs/2011.06801* (2020). URL: <https://arxiv.org/abs/2011.06801>. arXiv:2011.06801.
- [19] A. Gartland-Jones, MusicBlox: A real-time algorithmic composition system incorporating a distributed interactive genetic algorithm, in: *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2003, pp. 490–501. URL: https://doi.org/10.1007%2F3-540-36605-9_45. doi:10.1007/3-540-36605-9_45.
- [20] C. Rizzuti, E. Bilotta, P. Pantano, A GA-based control strategy to create music with a chaotic system, in: *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2009, pp. 585–590. URL: https://doi.org/10.1007%2F978-3-642-01129-0_66. doi:10.1007/978-3-642-01129-0_66.
- [21] F. Mauceri, S. M. Majercik, A swarm environment for experimental performance and improvisation, in: *Computational Intelligence in Music, Sound, Art and Design*, Springer International Publishing, 2017, pp. 190–200. URL: https://doi.org/10.1007%2F978-3-319-55750-2_13. doi:10.1007/978-3-319-55750-2_13.
- [22] A. S. Ramanto, N. U. Maulidevi, Markov chain based procedural music generator with user chosen mood compatibility, *International Journal of Asia Digital Art and Design Association* 21 (2017) 19–24. URL: http://dx.doi.org/10.20668/adada.21.1_19. doi:10.20668/adada.21.1_19.
- [23] J. Wu, C. Hu, Y. Wang, X. Hu, J. Zhu, A hierarchical recurrent neural network for symbolic melody generation, *IEEE Transactions on Cybernetics* 50 (2020) 2749–2757. URL: <https://doi.org/10.1109%2Ftcyb.2019.2953194>. doi:10.1109/tcyb.2019.2953194.
- [24] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, D. Eck, Music transformer, 2018. URL: <https://arxiv.org/abs/1809.04281>. doi:10.48550/ARXIV.1809.04281.
- [25] J. A. Biles, GenJam: An interactive genetic algorithm jazz improviser, *The Journal of the Acoustical Society of America* 102 (1997) 3181–3181. URL: <https://doi.org/10.1121%2F1>.

420841. doi:10.1121/1.420841.

- [26] M. Kikuchi, Y. Osana, Automatic melody generation considering chord progression by genetic algorithm, in: 2014 Sixth World Congress on Nature and Biologically Inspired Computing (NaBIC 2014), IEEE, 2014. URL: <https://doi.org/10.1109/nabic.2014.6921876>. doi:10.1109/nabic.2014.6921876.
- [27] H. Lim, S. Rhyu, K. Lee, Chord generation from symbolic melody using BLSTM networks, in: S. J. Cunningham, Z. Duan, X. Hu, D. Turnbull (Eds.), Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017, Suzhou, China, October 23-27, 2017, 2017, pp. 621–627. URL: https://ismir2017.smcnus.org/wp-content/uploads/2017/10/134_Paper.pdf.
- [28] F. Pachet, P. Roy, J. Moreira, M. d’Inverno, Reflexive loopers for solo musical improvisation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2013. URL: <https://doi.org/10.1145/2470654.2481303>. doi:10.1145/2470654.2481303.
- [29] I. P. Yamshchikov, A. Tikhonov, Music generation with variational recurrent autoencoder supported by history, SN Applied Sciences 2 (2020). URL: <https://doi.org/10.1007/s42452-020-03715-w>. doi:10.1007/s42452-020-03715-w.
- [30] S. Chatfield, W. Byrnes, Correlational analysis of aesthetic competency, skill acquisition and physiologic capabilities of modern dancers, in: 5th Hong Kong International Dance Conference Papers, 1990, pp. 79–100.
- [31] D. Krasnow, S. J. Chatfield, Development of the “performance competence evaluation measure”: assessing qualitative aspects of dance performance, Journal of Dance Medicine & Science 13 (2009) 101–107.
- [32] P. Gemeinboeck, R. Saunders, Movement matters, in: Proceedings of the 4th International Conference on Movement Computing, ACM, 2017. URL: <https://doi.org/10.1145/3077981.3078035>. doi:10.1145/3077981.3078035.
- [33] P. Gemeinboeck, R. Saunders, Dancing with the nonhuman, Thinking in the world. London: Bloomsbury Academic (2019).
- [34] P. Gemeinboeck, The aesthetics of encounter: A relational-performative design approach to human-robot interaction, Frontiers in Robotics and AI 7 (2021). URL: <https://doi.org/10.3389/frobt.2020.577900>. doi:10.3389/frobt.2020.577900.
- [35] J. L. Oliveira, L. P. Reis, B. M. Faria, F. Gouyon, An empiric evaluation of a real-time robot dancing framework based on multi-modal events, TELKOMNIKA Indonesian Journal of Electrical Engineering 10 (2012). URL: <https://doi.org/10.11591/telkomnika.v10i8.1327>. doi:10.11591/telkomnika.v10i8.1327.
- [36] L. Yang, S. Chou, Y. Yang, Midinet: A convolutional generative adversarial network for symbolic-domain music generation, in: S. J. Cunningham, Z. Duan, X. Hu, D. Turnbull (Eds.), Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017, Suzhou, China, October 23-27, 2017, 2017, pp. 324–331. URL: https://ismir2017.smcnus.org/wp-content/uploads/2017/10/226_Paper.pdf.
- [37] H.-M. Liu, Y.-H. Yang, Lead sheet generation and arrangement by conditional generative adversarial network, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2018. URL: <https://doi.org/10.1109/icmla.2018.00114>. doi:10.1109/icmla.2018.00114.

- [38] Y. Yan, E. Lustig, J. VanderStel, Z. Duan, Part-invariant model for music generation and harmonization, in: E. Gómez, X. Hu, E. Humphrey, E. Benetos (Eds.), Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France, September 23-27, 2018, 2018, pp. 204–210. URL: http://ismir2018.ircam.fr/doc/pdfs/293_Paper.pdf.
- [39] Q. Cao, X. Chen, R. Song, H. Jiang, G. Yang, Z. Cao, Multi-modal experience inspired ai creation, arXiv preprint arXiv:2209.02427 (2022).
- [40] F. Veit, L. Riedel, D. Jeraj, Does jumping to the beat result in better ratings from gymnastics experts?, *Journal of Human Sport and Exercise* 17 (2021). URL: <https://doi.org/10.14198%2Fjhse.2022.174.17>. doi:10.14198/jhse.2022.174.17.
- [41] F. Veit, J. Veit, T. Heinen, The influence of music on judges' evaluation of complex skills in gymnastics, *European Journal of Sport Sciences* 1 (2022) 1–7. URL: <https://doi.org/10.24018%2Fejsport.2022.1.5.31>. doi:10.24018/ejsport.2022.1.5.31.
- [42] M. Suzuki, Y. Matsuo, A survey of multimodal deep generative models, *Advanced Robotics* 36 (2022) 261–278. URL: <https://doi.org/10.1080%2F01691864.2022.2035253>. doi:10.1080/01691864.2022.2035253.
- [43] J. Li, H. Peng, H. Hu, Z. Luo, C. Tang, Multimodal information fusion for automatic aesthetics evaluation of robotic dance poses, *International Journal of Social Robotics* 12 (2019) 5–20. URL: <https://doi.org/10.1007%2Fs12369-019-00535-w>. doi:10.1007/s12369-019-00535-w.
- [44] K. Kritsis, A. Gkiokas, A. Pikrakis, V. Katsouros, Attention-based multimodal feature fusion for dance motion generation, in: Proceedings of the 2021 International Conference on Multimodal Interaction, ACM, 2021. URL: <https://doi.org/10.1145%2F3462244.3479961>. doi:10.1145/3462244.3479961.
- [45] G. Valle-Pérez, G. E. Henter, J. Beskow, A. Holzapfel, P.-Y. Oudeyer, S. Alexanderson, Transflower, *ACM Transactions on Graphics* 40 (2021) 1–14. URL: <https://doi.org/10.1145%2F3478513.3480570>. doi:10.1145/3478513.3480570.
- [46] E. Gordon, Rating scales and their uses for measuring and evaluating achievement in music performance, GIA Publications, 2002.
- [47] M. Wysoczanska, T. Trzcinski, Multimodal dance recognition., in: VISIGRAPP (5: VISAPP), 2020, pp. 558–565.
- [48] F. Ofli, E. Erzin, Y. Yemez, A. M. Tekalp, Multi-modal analysis of dance performances for music-driven choreography synthesis, in: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2010, pp. 2466–2469.