Towards Symbiotic Creativity: A Methodological Approach to Compare Human and AI Robotic Dance Creations

Allegra De Filippo*, **Luca Giuliani**, **Eleonora Mancini**, **Andrea Borghesi**, **Paola Mello** and **Michela Milano**
DISI, University of Bologna

Abstract

Artificial Intelligence (AI) has gradually attracted attention in the field of artistic creation, resulting in a debate on the evaluation of AI artistic outputs. However, there is a lack of common criteria for objective artistic evaluation both of human and AI creations. This is a frequent issue in the field of dance, where different performance metrics focus either on evaluating human or computational skills separately. This work proposes a methodological approach for the artistic evaluation of both AI and human artistic creations in the field of robotic dance. First, we define a series of common initial constraints to create robotic dance choreographies in a balanced initial setting, in collaboration with a group of human dancers and choreographer. Then, we compare both creation processes through a human audience evaluation. Finally, we investigate which choreography aspects (e.g., the music genre) have the largest impact on the evaluation, and we provide useful guidelines and future research directions for the analysis of interconnections between AI and human dance creation.

1 Introduction

Dance is an area where the potential application of Artificial Intelligence (AI) has raised interest, and humanoid robots have been successfully used thanks to their human-like aspect. Several works studied and implemented systems for robotic dance creation [Joshi and Chakrabarty, 2021], such as fine-balanced robotic dance movements with a human performer, dance motion imitation techniques, or generative techniques through visual observation. Most of these works focused on the automation of various aspects of dance creation, while only few studies have proposed evaluation metrics for robotic dance. Recent research works also aim at understanding the key differences between human and machine creation processes for different artistic disciplines, including dance [Peng *et al.*, 2021]. Overall, there is an emerging consensus about the need of exploring the relation between robotic dance creation and its aesthetic evaluation.

In this perspective, our research targets the differences between human and AI artistic creation, in particular focusing on the aesthetic evaluation of robotic dance performances. We propose a methodological approach by defining a series of steps in order to carry out the analysis in a balanced initial setting that allows to fairly compare both creation processes. Our methodological approach relies on three macro-phases: 1) a common setting definition, 2) choreography creation (either by AI or human), 3) a common evaluation phase.

We tested our methodological approach on a case study based on AI- and human-created choreographies for a NAO robot[1], in collaboration with a group dancers from the International Higher Education Academy C&C[2] led by the choreographer Carlo Massari. Firstly, we defined a common setting of constraints and positions to be satisfied by both the AI algorithm and the human choreographer. Then we performed all the choreographies with a NAO robot. The evaluation of the performances was then conducted with the help of a human audience unaware of the choreography creation processes, and using a shared evaluation scheme.

Using these results, we then investigated the general preference of the audience (human rather than AI choreographies), the importance of the music genre on the evaluation, and, more broadly, which aspects of the choreography have the largest impact on the evaluation. We also provide guidelines to replicate our analysis in the same artistic domain with different settings, and we propose future directions to further analyze the interplay between AI and human dance creation.

From our perspective, this work allows to better identify strengths and weaknesses of both human and AI creation processes. This can trigger several research directions such as the exploration of new art forms for human dance creation and symbiotic creation processes where human and AI can collaborate to design new dance performances.

2 Background

Human Dance Creation Generally, when choreographers create a dance, they start from a particular idea or stimulus from which they generate the sequence of movements [Schiphorst *et al.*, 1990]. Thus, each path is unique and based on 1) the choreographer's experience, 2) the initial idea, and

*allegra.defilippo@unibo.it

[1] https://www.aldebaran.com/en/nao
[2] https://www.ceccompany.org/?lang=it

3) the interaction with dancers and other experts, who inspire modifications to the initial idea, up to the creation of the final choreography. Indeed, the creative process is both interactive and iterative [Schiphorst *et al.*, 1990]. [Ciolfi Felice *et al.*, 2016; Singh *et al.*, 2011] show that, despite the creative process of each choreographer being unique and changing over time, common choreographic aspects can be defined, representing ideas at different levels of abstraction, and combining them with a set of base operations, such as rearranging elements and establishing transitions.

Formally, systems such as the Laban [Von Laban, 1975] or Benesh [Benesh and Benesh, 1977] notation have been defined. These methods are used mostly by large dance companies who can afford a full-time annotator. Contemporary choreographers rarely use these systems, that are designed to document finished work and not for annotating the first stages of exploration. More often, the initial stage of the creative process is done in the choreographers' mind, where they can explore ideas, which are then refined after a first embodiment through the dancers. According to [Heyward, 2015], choreographers often capture intermediate phases of their work with intermediate artefacts, such as inspiring images, annotated sketches, and video clips of dance fragments.

Codified forms of dances composed by precise patterns of movement exist. This is the case of Noh dance [Wolz, 1975], where each dance is composed of highly detailed "kata" (patterns of movement) joined together to form a flowing sequence. The creative process, in these cases, is aimed at creating a choreography as a sequence of kata that is capable of giving artistic expressiveness and musical coherence to the entire performance. In this work, we mainly refer to this codified form of dance.

Computational Dance Automation In recent years, many researchers proposed methods to automate partial aspects of dance, from dance notation to choreography, and from dance capture to dance generation [Sagasti, 2019; Joshi and Chakrabarty, 2021].

Especially in dance, where physical movement is a key factor, the use of robots is continually expanding thanks to their humanoid shape. Many works have studied and implemented systems for robotic dances, ranging from 1) humanoid robots performing fine-balanced dance movements with a human performer [Ramos *et al.*, 2015; Shinozaki *et al.*, 2007; Shinozaki *et al.*, 2008]; 2) experiments with robot motions that automatically coordinate to the music beat using a real-time music signal over which the humanoid robot has to dance autonomously [Grunberg *et al.*, 2010]; 3) dance motion imitation techniques for humanoid robots aimed at generating dance movements adequate to the music rhythm through visual observation [Angulo *et al.*, 2011]. Recent works [Liu *et al.*, 2020; Wang, 2022] are focused on the representation of choreographies as a sequence of basic and simple positions for robotic dances, according to Noh dance structure.

All these works focused on the creation of robotic dances through human-robot interaction, without taking into account a human evaluation of these artistic creations.

Dance Evaluation Approaches Defining a criterion for objective evaluation of dance performances is a complex task. In modern dance, well-known metrics for the evaluation of qualitative aspects of dance performances such as Aesthetic Competence Evaluation (ACE) [Chatfield and Byrnes, 1990] or Performance Competence Evaluation Measure (PCEM) [Krasnow and Chatfield, 2009] have been defined as a standard in this context. In particular, they consider aspects such as technique, space, phrasing and presence, time and energy, and a focus on the physical qualities of the performer, through three different levels of judgment for each area of evaluation.

Dancing robots are getting abler and abler at performing many kinds of robotic dances [Aucouturier *et al.*, 2008] thanks to improvements in mechanics and control. In this perspective, based on the existing measures for evaluating human dance skills, many works have proposed evaluation metrics for robotic dance. [Oliveira *et al.*, 2012] introduced a framework to evaluate robotic dance performances based on a Likert [Likert, 1932] questionnaire. With a major focus on robotic poses, [Manfrè *et al.*, 2017; Peng *et al.*, 2019] discuss aesthetic evaluation processes of robotic dance poses, for improving choreography creation. [Saffiotti *et al.*, 2020] focuses on the combination of human and robotic performers, by proposing a collaboration model evaluated using a Likert questionnaire to assess the harmony with the music, the harmony between the two performers, and the overall judgment. A more horizontal approach for merging the two evaluation perspectives (human and robotic) can be found in [De Filippo *et al.*, 2022b], where the evaluation questionnaire is expanded with aspects such as the overall theatricality of the robotic dance choreography, the degree of human reproducibility, and the use of the surrounding space in the dance performance.

It must be noted that all these works focus on the aesthetic evaluation of dance performances but disregard the actual creation process which, on the contrary, is an important feature to be taken into account [Hong and Curran, 2019]. Conversely, in our work we analyze how a human or a computational creation process can differently affect the aesthetic evaluation of a robotic dance performance.

3 Methodological Approach

In this work, we propose a methodological approach to analyze the differences between human- and AI-based artistic creation, in particular focusing on the aesthetic evaluation of robotic dance performances. We define and formalize the steps needed to carry out the analysis in a thorough and technically sound manner, based on a balanced initial setting to equally compare both the creation processes. Then, we experiment our approach on a case study based on AI- and human-created choreographies for a NAO robot. Finally, we compare the obtained choreographies based on a common evaluation phase to provide useful guidelines in the analysis of the interconnections between AI and human dance creation.

As illustrated in Figure 1, our methodological approach is based on three macro-phases: 1) common setting definition, 2) choreography creation based either on AI or human creation process, 3) common evaluation phase.

In details, we proceed with the following steps for experimenting our approach on a case study based on robotic dance choreographies:

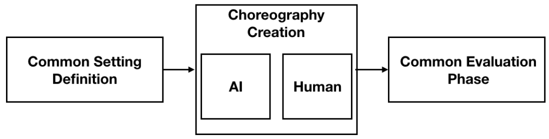

Figure 1: Schema of our methodological approach

1. Common setting definition:
   1 a) Definition of a common set of basic positions inspired by Noh dance [Wolz, 1975].
   1 b) Definition of a common set of constraints to be satisfied in terms of (e.g.,) initial and final position and number of mandatory positions to be executed during the dance performance; a choreography is represented as sequence of positions [Liu *et al.*, 2020; Singh *et al.*, 2011; Wolz, 1975].
   1 c) Definition of a common set of music tracks with fixed time duration; tracks belong to a subset of different music genres, picked for being highly recognizable based on their specific characteristics [Sturm, 2012; Scaringella *et al.*, 2006] (i.e., folk, electronic, classical, rock and latin).
2. Choreography creation:
   2 a) Creation of 5 robotic choreographies with a basic AI technique taking into account music and position constraints as inputs.
   2 b) Creation of 5 robotic choreographies by a human choreographer taking into account *the same* constraints as the AI algorithm.
3. Common evaluation phase:
   3 a) Creation of an ad-hoc questionnaire based on the state of the art [De Filippo *et al.*, 2022b; Oliveira *et al.*, 2012];
   3 b) Execution of a between-subjects experiment where the 10 choreographies (grouped as C_AI and C_H, all performed by a NAO robot) are evaluated by two different groups of people with both artistic and scientific knowledge.
   3 c) Statistical analysis of the questionnaire results in order to answer our research questions.

3.1 Common Setting Definition

The problem description, given to both our AI algorithm and the human choreographer, states an initial and a final position. We also provide a repository of robot positions[3], with all the mandatory and intermediate positions (e.g., sit position, stand position, etc.) already implemented for the NAO robot. In particular, we provide both the code and a detailed description and video demo for each position, in order to clarify them to the human choreographer. A focus on the choreography creation phase is depicted in Figure 2 by highlighting the common initial setting.

The shared constraints are the following:

[3] https://github.com/ProjectsAI/ComparativeArtisticEvaluation/tree/main/codePositions

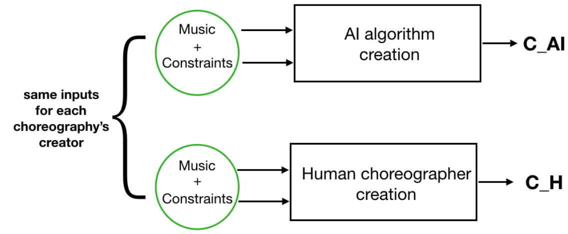

Figure 2: Choreography creation

- each choreography must start with a specific initial position and must end in a specific final position;
- the total duration must be of exactly 2 minutes;
- each choreography must contain at least each mandatory position;
- intermediate positions can be used to move between mandatory positions;
- each choreography must contain at least 5 different intermediate positions;
- the sequence of positions must avoid possible incompatibilities between two consecutive positions;
- each choreography must be associated with one of the 5 different genre music tracks provided.

For each creation process (AI or human), the final goal is to generate 5 different choreographies (one per genre) satisfying the given constraints. Repetitions of positions is allowed if the minimum number of intermediate and mandatory positions is satisfied. All the choreographies are available in a public repository[4]. More details can be found in Section 3.2. In Figure 3 we show an example of two choreographies (represented as sequences of positions for the same music track), one created by our AI algorithm and one by a human choreographer, respecting the same constraints.

3.2 Choreography Creation

Human Creation A group of 30 dancers from the International Higher Education C&C Academy led by the choreographer Carlo Massari[5] worked on a 3-days creative workshop to prepare the robotic choreographies. The teams of choreographer and dancers were asked to create choreographies starting from the same input of our AI algorithm. The choreographer starts from the music track to plan a sequence of positions that satisfies the required constraints. This is a well-established approach in the artistic and theatrical dance scenario, as for the case of Noh dance. In more details, we organized the creation process in 3 macro-phases: 1) definition phase: understand the problem definition, 2) preparation phase: strategize the design choices, 3) execution phase: implement the solution. In phase 1, the choreographer and the dancers were provided with a pool of positions (video and textual description), the duration (in seconds) for each position on NAO robot, the constraints, and audio tracks. Then,

[4] https://github.com/ProjectsAI/ComparativeArtisticEvaluation/tree/main

[5] https://www.ceccompany.org/?lang=it

AI ALGORITHM OUTPUT	HUMAN CHOREOGRAPHER OUTPUT
MUSIC TRACK 4	MUSIC TRACK 4
'OB_STAND INIT',	'OB_STAND INIT',
'DIAGONAL LEFT',	'UNION ARMS',
'ROTATION FOOT LEFT LEG',	'OB_SIT',
'OB_STAND',	'ROTATION HANDGUN',
'RIGHT ARM',	'OB_SIT RELAX',
'OB_HELLO',	'ROTATION HANDGUN',
'SING WITH ME',	'OB_SIT',
'SUPERMAN',	'ROTATION HANDGUN',
'OB_STAND ZERO',	'ROTATION FOOT LEFT LEG',
'ROTATION FOOT LEFT LEG',	'OB_HELLO',
'SING WITH ME',	'SPRINKLER',
'OB_SIT',	'OB_STAND',
'SUPERMAN',	'OB_WIPE FOREHEAD',
'OB_SIT RELAX',	'OB_STAND ZERO',
'BIRTHDAY DANCE',	'WORKOUT',
'ROTATION HANDGUN',	'OB_WIPE FOREHEAD',
'ROTATION FOOT RIGHT LEG',	'OB_WIPE FOREHEAD',
'OB_WIPE FOREHEAD',	'OB_WIPE FOREHEAD',
'OB_CROUCH'	'OB_WIPE FOREHEAD',
	'OB_CROUCH'

Figure 3: Sequence of positions generated by the AI algorithm and the human choreographer. In particular, we have the initial and final positions (blue), the mandatory positions (green and at least 6), and the intermediate positions (black and at least 5).

in phase 2, the choreographer experimented with the entire sequence of positions in collaboration with the dancers, by analyzing different position patterns with the different music tracks and by classifying them based on different stylistic choices. Finally, for phase 3, they provided us with a written sequence of positions (per music track) that satisfies all the required constraints and to be codified and then performed by a NAO robot.

A significant difference with the typical creative process lies in the fact that we are bounding the dance performance through arbitrary and somewhat unnatural constraints; this contrasts with the usual approach of human choreographers, but it allows to define a common initial setting for a more balanced evaluation of both the (AI and human-created) performances.

AI Creation The creation process is structured into three macro-phases that mirror the human creation process. The algorithm begins with defining the problem, then considers factors like dance-music synchronization and position patterns (preparation), finally creates a position sequence for a NAO robot that satisfies the required constraints (execution).

Specifically, the algorithm is based on simulated annealing [Kirkpatrick *et al.*, 1983]. For each move in the sequence, a neighborhood is defined and used to retrieve the next move according to a value function. The value function takes into account the aesthetic constraints defined in the preparation phase. During the preparation phase, design choices are carried out in order to account for constraint satisfaction, similarly to how a choreographer would make choices to respect constraints in human choreography. These choices involve parameters used in the value function, such as:

- α: the song's amplitude, which is used to establish different intervals associated with various song-related moments (e.g., an α value above a predetermined threshold denotes an intense or explosive chorus-related moment).
- cm: the class of a move, which is defined according to factors including the robot's ability to move fluidly and the execution time of the movement.
- $move\_match$: a fitness function, which expresses how well does a given move $m \in cm$ fits in a time frame characterized by a certain amplitude α. The higher the value, the more accurately the move is associated with the amplitude, with ideal matches represented by certain pairings (α, cm).
- $time\_match$: a second fitness function, which quantifies the effectiveness of a particular move based on the transition time from the previous move. Smaller transition times are associated with higher values.

Algorithm 1 Execution

```
procedure RUN SIMULATED ANNEALING
    while optimal_solution ≠ FOUND do
        Expand neighborhood
        Choose a move m from the neighbourhood
        Evaluate the chosen move through v(cm, α)
    end while
    return optimal_solution
end procedure
```

In the execution phase (Algorithm 1), simulated annealing is repeatedly run until all the constraints of the problem described in Section 3.1 and the aesthetic constraints mentioned earlier are satisfied. On the basis of the aforementioned parameters, the algorithm specifically seeks to optimize the value function $v(cm, \alpha)$, which is compute by averaging the values of the two fitness functions $move\_match$ and $time\_match$. Simulated annealing, by definition, makes sure that the likelihood of accepting moves that are worse for the optimization process (i.e., moves with a low score) reduces with each iteration. The optimal solution is the sequence of moves for which all the constraints of the problem described in Section 3.1 and the aesthetic constraints mentioned earlier are satisfied. Even though we opted for such a solution, *the execution phase can be thought of as a black-box procedure which could be replaced by any other algorithm, provided that the same constraints are respected.*

3.3 Common Evaluation

Evaluation Questionnaire The evaluation phase is conducted to investigate the reactions and perceptions of the audience of the robotic performances created. The methodological tools used for data collection are participant observation and questionnaires. Based on [De Filippo *et al.*, 2022b] we define a survey[6] to evaluate the robotic choreographies generated by the algorithm and the choreographer. We propose two different questionnaires, i.e., one per creation process (AI or Human). Each questionnaire is composed by 5 pages (one per different choreography music), and each page is composed by a video demo of the robotic choreography and a list of questions, one for each evaluation target. Each participant anonymously vote the proposed choreographies, providing a score for all targets on a Likert scale (from 1 to 5). The targets are: (1) Storytelling; (2) Rhythm; (3) Movement Technique; (4)

[6] https://forms.gle/BkPF2mpjX3QGyjfL8

Public Involvement; (5) Space Use; (6) Human Characterization; (7) Human Reproducibility. For each target, we propose a specific question to the user.

Results Collection and Audience We conduct a between-subjects experiment for the evaluation phase; in this kind of design each user is exposed to a single subset of choreographies (AI generation vs human generation). This allows to evaluate the differences between both subsets of choreographies without any conditioning in the evaluation phase: the audience is not affected by any notion about the experimental setting [Charness *et al.*, 2012]. We administered a questionnaire pertaining to the choreographies to different groups of subjects. To avoid decision fatigue [Pignatiello *et al.*, 2020], questions order is randomized in each questionnaire. The audience has been selected to be equally distributed among users with scientific, artistic or both backgrounds. We collected the questionnaire responses and then analyzed them, starting with the identification of the emerging choreography features in relation with the evaluation targets.

Dataset Construction For each choreography in input, we extract 15 features. We collect them according to the evaluation targets, relying on state-of-the-art analysis [De Filippo *et al.*, 2022b]. For each choreography, the following information are stored[7]: 1) *nDifferentMovements*: the number of different movements of the choreography; 2) *nTotalMovements*: the number of total movements (with optional repetitions) of the choreography; 3) *movementDifficulty*: the degree of difficulty of the moves; the allowed levels (i.e., low, medium, high) are mapped to the interval $md \in [1, 3]$; 4) *acrobaticMovements*: the level of acrobatic movements, $am \in [1, 3]$; 5) *movementsRepetition* (mr): the level of movement repetitions, $mr \in [1, 3]$; 6) *humanMovements* (h): the level of human movement presence, $h \in [1, 3]$; 7) *balance* (b): the level of balance movements, $b \in [1, 3]$; 8) *speed* (s): the degree of movement speed, $s \in [1, 3]$; 9) *bodyPartsCombination* (bc): the level of combinations involving different body parts, $bc \in [1, 3]$; 10) *musicBPM* (bpm): the number of Beats Per Minute; 11) *headMovement* (hm): the level of combinations involving the head, $hm \in [1, 3]$; 12) *armsMovement* (arm): the level of combinations involving the arms, $arm \in [1, 3]$; 13) *handsMovement* (hdm): the level of hands movement presence, $hdm \in [1, 3]$; 14) *legsMovement* (lm): the level of legs movement presence, $lm \in [1, 3]$; 15) *feetMovement* (fm): the level of feet movement presence, $fm \in [1, 3]$.

The 7 evaluation targets follow the survey questions (see Section 3.3).

4 Experimental Analysis

For our experimental part, we examined user scores through three different analyses, to answer to our research questions defined in Section 3. First, we investigated score preferences for different artistic creation strategies, focusing on specific evaluation targets. Second, we examined score preferences for different music genres, again focusing on specific targets. Finally, we delved into the analysis of the choreography features that mostly influenced the audience's evaluation.

[7] https://github.com/ProjectsAI/ComparativeArtisticEvaluation/tree/main/datasets

Our research questions are the following:

- **RQ1:** Does the audience prefer AI-created choreographies or Human-created ones?
- **RQ2:** Does the music genre influence the evaluation of the audience?
- **RQ3:** Which choreography features have the largest impact on the evaluation targets?

4.1 Experimental Setting

The audience is composed by 68 participants with a scientific and/or artistic background. Each user successfully completed only one questionnaire (AI or Human). Participants interacted with an anonymous questionnaire, as explained in Section 3.3. We conduct a between-subjects experiment with two groups composed by 34 participants. Each group voted on 5 choreographies for a total of 340 observations (considering both AI- and human-generated choreographies). In this setting, each user is exposed either to only AI-created choreographies or to only human-generated ones.

For our experimental phase, we define independent variables in order to observe and measure their effects on our dependent variables. In details, our independent variables are 1) the creation process of the choreography and 2) the music genre; the dependent variables are the evaluation targets that we want to measure, as reported in Section 3.3. To address RQ1, we compared ,for each evaluation target, the average ratings received for AI-created choreographies to the average ratings received for Human-created choreographies. To address RQ2, we further split the data based on the music genre accompanying the choreography. Finally, as regards RQ3, we used linear regression to identify the features (see Section 3.3) with the strongest influence on the evaluation target.

4.2 AI and Human Creation Process Comparison

First, we compared the average ratings received for AI-created choreographies to the average ratings received for human-created choreographies.

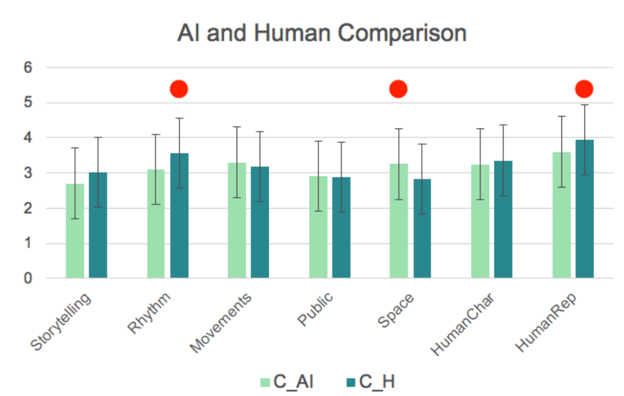

Figure 4: Average scores and deviation standard per evaluation target, both for C_AI (light green) and C_H (dark green). The statistically significant gaps between C_AI and C_H are highlighted with a red circle.

We used a two-sided Mann-Whitney U rank test [Rosner and Grove, 1999] on the two independent samples of collected scores, since the requirement of normal distribution for the t-test is not met for these samples. The test results showed that the evaluation targets for which we have a statistically significant preference in term of scores are Rhythm ($p = 0.008$), Space ($p = 0.002$), and Human Reproducibility ($p = 0.004$). As for the Storytelling, we can observe a gap that is not statistically significant (with $p = 0.184$), but that deserves to be further analyzed with (e.g.,) a larger sample of observations.

This suggested that: (1) participants significantly preferred Human-created choreography in terms of rhythmic coherence with the music; (2) participants significantly preferred AI-created choreography based on the use of space – the space is perceived to be better used in AI choreographies. This might be explained by the fact that human choreographers are used to work with real environments (the actual stage with its confined spaces, the position of the dancers and audience, etc) and thus might somewhat project these usual, implicit constraints into their choreographies, underutilizing the "artificial" space where the robot is dancing. (3) As for the Human Reproducibility target, we asked if the choreography can be easily reproduced by a human performer, and the results showed that evaluators ranked Human-created choreographies as more reproducible.

4.3 Music Genre Comparison

To address RQ2, we further split the data based on the music genre. Previous test revealed that there is statistically significant correlation between rhythmic coherence and music. We used again a two-sided Mann-Whitney U rank test on the two independent samples of collected score. All statistically significant values are highlighted with (*).

The test results (see Table 1) showed that folk is the music genre with a statistically significant preference in term of scores for almost all the targets, with significant values ($p < 0.05$) for Rhythm, Space, and Human Reproducibility. For this music genre, the results are in line with the trends emerged during the first experiment (Section 4.2), with a statistically significant gap for C_AI related to the use of Space, and for C_H related to Rhythm and Human Reproducibility, while the inversion of the results trend can be seen only for the Storytelling target that shows a significant preference for C_AI.

As for the electronic genre, a significant preference (for C_H) only emerged for the Human Reproducibility target, again similarly to the results of the first experiment.

As for the Latin and Rock genres, the Storytelling target shows again a significant preference for C_H in line with the first experiment. This holds for the targets describing the coherence with the rhythm and the use of space as well.

Finally, for the classical genre, the significant preference (for C_AI) only emerged for the use of Space target, exhibiting again the trend of the first experiment.

Moreover, this second set of experiments confirms that, in general, similar score values can be observed for two targets: Movements (related to the technique and fluidity of movements) and Public (related to the public involvement). Average preferences connected to these targets are similar for C_AI and C_H regardless of music genre. This can be explained by the common initial setting that limits both the movements choice and the degree of creativity allowed for public involvement.

Music Genre	Evaluation Target	C_AI avg	C_AI std	C_H avg	C_H std
Folk	Storytelling	**3.117** (*)	1.174	2.705	1.404
	Rhythm	3.264	0.994	**3.941** (*)	0.776
	Movements	3.411	0.988	3.265	0.931
	Public	2.764	1.102	2.763	1.371
	Space	**3.294** (*)	1.243	2.529	1.236
	Human Char	3.088	1.025	**3.471** (*)	1.079
	Human Rep	3.676	1.224	**4.471** (*)	0.928
Electronic	Storytelling	2.911	1.190	2.794	1.343
	Rhythm	3.588	1.076	3.911	0.865
	Movements	3.323	1.036	3.176	0.903
	Public	3.088	0.900	2.941	1.099
	Space	3.029	0.869	2.911	1.055
	Human Char	3.352	1.011	3.382	1.101
	Human Rep	3.617	1.015	**4.117** (*)	0.913
Rock	Storytelling	2.647	1.097	**3.089** (*)	1.164
	Rhythm	3.323	1.006	3.058	1.253
	Movements	3.441	0.823	3.205	1.174
	Public	3.029	1.086	2.823	1.381
	Space	**3.147** (*)	0.857	2.558	1.210
	Human Char	3.235	1.046	3.265	0.931
	Human Rep	3.617	1.044	3.500	1.308
Latin	Storytelling	2.470	1.022	**3.265** (*)	1.238
	Rhythm	3.000	1.231	**3.735** (*)	0.931
	Movements	3.205	1.008	3.088	1.311
	Public	2.970	1.193	3.029	1.313
	Space	3.088	1.055	3.147	1.209
	Human Char	3.176	1.058	3.323	1.272
	Human Rep	3.441	0.859	**4.088** (*)	0.865
Classical	Storytelling	2.852	1.282	3.205	1.409
	Rhythm	3.205	0.977	3.177	1.192
	Movements	3.118	0.913	3.205	1.409
	Public	2.676	1.065	2.852	1.258
	Space	**3.529** (*)	0.928	3.000	1.044
	Human Char	3.382	0.953	3.323	1.272
	Human Rep	4.029	0.834	3.558	1.330

Table 1: Average scores and standard deviation per evaluation target based on music genre, both for C_AI and C_H. We highlight statistically significant values with (*).

4.4 Feature Choreography Analysis

Finally, we tackle RQ3 and we employ a linear regression model to identify the choreography features with the largest influence on the evaluation targets. We consider the set of features listed in Section 3.3 which were extracted for each choreography. We split this information based on C_AI and C_H in the following tables and we show only statistically significance values ($p < 0.05$) for at least one target.

In Table 2 and Table 3 we highlight statistically significant correlations (both positive and negative) between the features

Choreography features **C_AI**	Storytelling (coef)	Rhythm (coef)	Movements (coef)	Public (coef)	Space (coef)	HumanChar (coef)	HumanRep (coef)
nTotalMov	**.156** (*)	**.205** (*)	**.121** (*)	**.120** (*)	**.216** (*)	**.227** (*)	**.280** (*)
movDifficulty	**-.096** (*)	**-.114** (*)	**-.075** (*)	**-.060** (*)	**-.063** (*)	**-.077** (*)	**-.092** (*)
acrobaticMov	**-.056** (*)	**-.075** (*)	**-.044** (*)	**-.041** (*)	**-.049** (*)	**-.061** (*)	**-.071** (*)
movementsRep	.075	**.190** (*)	**.066** (*)	.109	**.157** (*)	**.218** (*)	**.241** (*)
humanMov	.105	.054	.082	.041	-.089	-.074	-.109
balance	**-.086** (*)	**-.160** (*)	**-.072** (*)	**-.084** (*)	**-.101** (*)	**-.143** (*)	**-.163** (*)
speed	-.019	**-.052** (*)	-.019	**-.041** (*)	-.010	**-.044** (*)	**-.029** (*)
bodyPartsComb	-.003	-.049	.005	**.065** (*)	.024	**.079** (*)	.045
musicBPM	-.005	.003	**.007** (*)	**.009** (*)	**-.008** (*)	-.002	**-.009** (*)
armsMov	-.048	**-.072** (*)	-.035	.000	-.064	-.048	**-.097** (*)
handsMov	-.007	.174	.006	.132	**.266** (*)	**.349** (*)	**.374** (*)
feetMov	.013	-.032	.001	**-.062** (*)	.036	-.034	.026
r2	.037	.032	.017	.022	.032	.012	.037

Table 2: Results obtained with the linear regression models. Each model predicts evaluation targets based on AI-created choreography (C_AI) features. We selected those features presenting at least one statistically significant (*) gap for the evaluation targets.

Choreography features **C_H**	Storytelling (coef)	Rhythm (coef)	Movements (coef)	Public (coef)	Space (coef)	HumanChar (coef)	HumanRep (coef)
nTotalMov	**.060** (*)	**.064** (*)	.038	**.064** (*)	**.103** (*)	**.045** (*)	.034
movDifficulty	-.037	**-.076** (*)	**.-069** (*)	.004	-.030	.043	**-.046** (*)
acrobaticMov	-.057	-.003	.023	-.004	-.050	.009	-.028
movementsRep	-.016	-.015	.004	.020	.020	-.006	-.068
humanMov	-.041	**.077** (*)	-.024	-.013	-.019	-.007	**.085** (*)
balance	**.127** (*)	-.060	.038	.013	.052	.035	.009
speed	**.068** (*)	.015	-.009	.026	**.080** (*)	.004	.019
bodyPartsComb	**-.105** (*)	**-.076** (*)	-.006	-.018	**-.072** (*)	-.001	**.057** (*)
musicBPM	**.014** (*)	**.008** (*)	**.011** (*)	**.009** (*)	.006	**.011** (*)	**.014** (*)
armsMov	**-.082** (*)	**.082** (*)	.005	.005	-.032	-.006	.050
handsMov	**.150** (*)	-.054	-.014	-.004	.092	.030	.001
legsMov	-.001	.013	**.023** (*)	-.001	-.019	**.021** (*)	**.032** (*)
feetMov	.053	**-.068** (*)	**.056** (*)	.016	.007	.035	-.053
r2	.029	.121	.003	.005	.044	.004	.105

Table 3: Results obtained with the linear regression models. Each model predicts evaluation targets based on Human-created choreography (C_H) features. We selected those features presenting at least one statistically significant (*) gap for the evaluation targets.

and the different evaluation targets. Many interesting correlations can be found: (1) we observe that the level of movements difficulty is negatively correlated to the evaluation targets both for C_AI and C_H; (2) accordingly, the total number of movements in the choreography is positively correlated with the evaluation targets; (3) music BPM is positively correlated with the evaluation targets (only) for C_H.

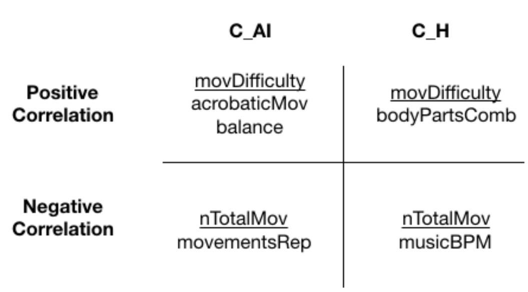

Figure 5: Features with a major impact on all the evaluation targets

Figure 5 shows the features with a major impact on all the evaluation targets based on the creation process (i.e., AI or human), by helping to define some common guidelines.

5 Discussion and Conclusions

In this work, we propose a methodological approach to analyze the differences between human- and AI-based artistic creation, in particular focusing on the aesthetic evaluation of robotic dance performances. We define a series of steps to carry out the analysis in a balanced initial setting that allows to fairly compare both creation processes. Our approach is based on three macro-phases: (1) definition of a common setting, (2) choreography creation based either on AI or human creation process, (3) common choreography evaluation. Related to the emerging trends of the final evaluation phase, we suggest future interdisciplinary research directions.

Firstly, related to RQ1 (Does the audience prefer AI-created choreographies or Human-created ones?), we plan to analyze a further step in the experimental part, by providing the parallel execution of both the AI- and Human-made choreography to the audience. The idea is to propose a within-subject analysis to a different group of participants, in order to understand if these results are confirmed also by users that are exposed to both the choreographies at the same time. We plan to make two versions of this experiment by proposing either only human or robot performers.

Related to RQ2 (Does the music genre influence the evaluation of the audience?), we have observed that the target associated to the rhythm showed a statistically significant gap in most of the experiments, so we plan to implement new choreographies with different music genres and also with different audio tracks of the same music genre, to further validate our results. Moreover, w.r.t. the feature associated to the number of different movements, we observed its positive correlation with most of the evaluation targets (RQ3). This result suggests that choreographies with a greater variety of different positions are preferred in general, but also that a longer duration of the choreography could be preferred as well ([De Filippo *et al.*, 2022b]). In this direction, we plan to provide different common settings (e.g., time duration, rhythm, movements) for further experiments with human choreographers.

Finally, the idea is to provide more useful indications to both human choreographers and AI algorithms to create robotic choreographies based on positive/negative correlations between choreography features and evaluation targets. This can trigger the exploration of new art forms for human dance creation and symbiotic creation processes where human and AI can collaborate to design new dance performances. Moreover, this work can also provide a useful starting benchmark [De Filippo *et al.*, 2022a] for training Machine Learning models to predict the relationship between the input features and the predicted target for new choreographies.

Acknowledgements

We thank Matteo Casari, Carlo Massari C&C Company, and the performing robots group. This work is partially supported by PNRR - M4C2 - (PE00000013) - "FAIR - Future Artificial Intelligence Research" Spoke 8 "Pervasive AI", funded by the European Commission under the NextGeneration EU programme.

References

[Angulo *et al.*, 2011] Cecilio Angulo, Joan Comas, and Diego Pardo. Aibo jukebox–a robot dance interactive experience. In *International Work-Conference on Artificial Neural Networks*, pages 605–612. Springer, 2011.

[Aucouturier *et al.*, 2008] Jean-Julien Aucouturier, Katsushi Ikeuchi, Hirohisa Hirukawa, Shin'ichiro Nakaoka, Takaaki Shiratori, Shunsuke Kudoh, Fumio Kanehiro, Tetsuya Ogata, Hideki Kozima, Hiroshi G Okuno, et al. Cheek to chip: Dancing robots and ai's future. *IEEE Intelligent Systems*, 23(2):74–84, 2008.

[Benesh and Benesh, 1977] Rudolf Benesh and Joan Benesh. *Reading dance: The birth of choreology*. London: Souvenir Press, 1977.

[Charness *et al.*, 2012] Gary Charness, Uri Gneezy, and Michael A Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, 81(1):1–8, 2012.

[Chatfield and Byrnes, 1990] SJ Chatfield and WC Byrnes. Correlational analysis of aesthetic competency, skill acquisition and physiologic capabilities of modern dancers. In *5th Hong Kong International Dance Conference Papers*, pages 79–100, 1990.

[Ciolfi Felice *et al.*, 2016] Marianela Ciolfi Felice, Sarah Fdili Alaoui, and Wendy E. Mackay. How do choreographers craft dance? designing for a choreographer-technology partnership. In *Proceedings of the 3rd International Symposium on Movement and Computing*, MOCO '16, New York, NY, USA, 2016. Association for Computing Machinery.

[De Filippo *et al.*, 2022a] Allegra De Filippo, Andrea Borghesi, Andrea Boscarino, and Michela Milano. Hada: An automated tool for hardware dimensioning of ai applications. *Knowledge-Based Systems*, 251:109199, 2022.

[De Filippo *et al.*, 2022b] Allegra De Filippo, Paola Mello, and Michela Milano. Do you like dancing robots? ai can tell you why. In *PAIS 2022*, pages 45–58. IOS Press, 2022.

[Grunberg *et al.*, 2010] David Grunberg, Robert Ellenberg, In Hyuek Kim, Jun Ho Oh, Paul Y Oh, and Youngmoo E Kim. Development of an autonomous dancing robot. *International Journal of Hybrid Information Technology*, 3(2):33–43, 2010.

[Heyward, 2015] Anna Heyward. How to write a dance. *The Paris Review: The Daily, February*, 4, 2015.

[Hong and Curran, 2019] Joo-Wha Hong and Nathaniel Ming Curran. Artificial intelligence, artists, and art: attitudes toward artwork produced by humans vs. artificial intelligence. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 15(2s):1–16, 2019.

[Joshi and Chakrabarty, 2021] Manish Joshi and Sangeeta Chakrabarty. An extensive review of computational dance automation techniques and applications. *Proceedings of the Royal Society A*, 477(2251):20210071, 2021.

[Kirkpatrick *et al.*, 1983] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[Krasnow and Chatfield, 2009] Donna Krasnow and Steven J Chatfield. Development of the "performance competence evaluation measure": assessing qualitative aspects of dance performance. *Journal of Dance Medicine & Science*, 13(4):101–107, 2009.

[Likert, 1932] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[Liu *et al.*, 2020] Yuechang Liu, Dongbo Xie, Hankz Hankui Zhuo, and Liqian Lai. Plan2dance: Planning based choreographing from music. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13624–13625, 2020.

[Manfrè *et al.*, 2017] Adriano Manfrè, Ignazio Infantino, Agnese Augello, Giovanni Pilato, and Filippo Vella. Learning by demonstration for a dancing robot within a computational creativity framework. In *2017 First IEEE International Conference on Robotic Computing (IRC)*, pages 434–439. IEEE, 2017.

[Oliveira *et al.*, 2012] João Lobato Oliveira, Luis Paulo Reis, and Brígida Mónica Faria. Empiric evaluation of robot dancing framework based. *TELKOMNIKA*, 10(8):1701–8, 2012.

[Peng *et al.*, 2019] Hua Peng, Jing Li, Huosheng Hu, Liping Zhao, Sheng Feng, and Keli Hu. Feature fusion based automatic aesthetics evaluation of robotic dance poses. *Robotics and Autonomous Systems*, 111:99–109, 2019.

[Peng *et al.*, 2021] Hua Peng, Jinghao Hu, Haitao Wang, Hui Ren, Cong Sun, Huosheng Hu, and Jing Li. Multiple visual feature integration based automatic aesthetics evaluation of robotic dance motions. *Information*, 12(3):95, 2021.

[Pignatiello *et al.*, 2020] Grant A Pignatiello, Richard J Martin, and Ronald L Hickman Jr. Decision fatigue: A conceptual analysis. *Journal of health psychology*, 25(1):123–135, 2020.

[Ramos *et al.*, 2015] Oscar E Ramos, Nicolas Mansard, Olivier Stasse, Christophe Benazeth, Sovannara Hak, and Layale Saab. Dancing humanoid robots: Systematic use of osid to compute dynamically consistent movements following a motion capture pattern. *IEEE Robotics & Automation Magazine*, 22(4):16–26, 2015.

[Rosner and Grove, 1999] B Rosner and D Grove. Use of the mann–whitney u-test for clustered data. *Statistics in medicine*, 18(11):1387–1400, 1999.

[Saffiotti *et al.*, 2020] Alessandro Saffiotti, Peter Fogel, Peter Knudsen, Luis de Miranda, and Oscar Thörn. On human-ai collaboration in artistic performance. In *First International Workshop on New Foundations for Human-Centered AI (NeHuAI) co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), Santiago de Compostella, Spain, September 4, 2020*, pages 38–43. CEUR-WS, 2020.

[Sagasti, 2019] Francisco Sagasti. Information technology and the arts: the evolution of computer choreography during the last half century. *Dance Chronicle*, 42(1):1–52, 2019.

[Scaringella *et al.*, 2006] Nicolas Scaringella, Giorgio Zoia, and Daniel Mlynek. Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23(2):133–141, 2006.

[Schiphorst *et al.*, 1990] T. Schiphorst, T. Calvert, C. Lee, C. Welman, and S. Gaudet. Tools for interaction with the creative process of composition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, page 167–174, New York, NY, USA, 1990. Association for Computing Machinery.

[Shinozaki *et al.*, 2007] Kuniya Shinozaki, Akitsugu Iwatani, and Ryohei Nakatsu. Concept and construction of a dance robot system. In *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, pages 161–164, 2007.

[Shinozaki *et al.*, 2008] Kuniya Shinozaki, Akitsugu Iwatani, and Ryohei Nakatsu. Construction and evaluation of a robot dance system. In *Entertainment Computing Symposium*, pages 83–94. Springer, 2008.

[Singh *et al.*, 2011] Vikash Singh, Celine Latulipe, Erin Carroll, and Danielle Lottridge. The choreographer's notebook: a video annotation system for dancers and choreographers. In *Proceedings of the 8th ACM Conference on Creativity and Cognition*, pages 197–206, 2011.

[Sturm, 2012] Bob L Sturm. Two systems for automatic music genre recognition: What are they really recognizing? In *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pages 69–74, 2012.

[Von Laban, 1975] Rudolf Von Laban. *Laban's principles of dance and movement notation*. Princeton Book Company Pub, 1975.

[Wang, 2022] Zhigang Wang. Music choreography algorithm based on feature matching and fragment segmentation. *Computational Intelligence and Neuroscience*, 2022, 2022.

[Wolz, 1975] CARL Wolz. Dance in the noh theatre. *The World of Music*, 17(3):26–32, 1975.