Perceiving Artistic Expression: A Formal Exploration of Performance Art Salsa

KAYHAN ÖZCIMDER[1], (Member, IEEE), ZHAODAN KONG[2], (Member, IEEE), SHUAI WANG[3], (Student Member, IEEE), AND JOHN BAILLIEULL[4], (Fellow, IEEE)

[1]Department of Mechanical and Aerospace Engineering, Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
[2]Department of Mechanical and Aerospace Engineering, University of California at Davis, Davis, CA 95616, USA
[3]Division of Systems Engineering, Boston University, Boston, MA 02215, USA
[4]Division of Systems Engineering, Department of Mechanical Engineering, Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA

Corresponding author: John Baillieul (johnb@bu.edu)

ABSTRACT This paper studies artistic expression in human movement by exploring the performance art form salsa. The motions of a salsa performance are constructed as concatenations of motion primitives, each of which specifies the movement of the dance pair over the course of eight musical beats. To analyze the syntax of artistic expression, the choreography of dance performances is represented by a transition model that is based on humanoid robot representations of the dancers. In order to assess the quality of a performance, two distinct metrics are explored. By integrating the performance metrics into the proposed transition system, it is possible to create an algorithm that is capable of autonomously recognizing the dance moves and evaluating the quality of the performance with a score. To validate the model, a dance pair performed four distinct salsa dance sequences observed by an artificially intelligent (AI) judge. The video recordings of the performances are also shown to a dance audience for evaluation. By looking at the correlation between the dance audience and the AI judge's scores, we conclude that the proposed model performs well in evaluating the artistic merit of the dance.

INDEX TERMS Artificial intelligence, art, non-verbal communication, discrete-event systems.

I. INTRODUCTION

How can we measure 'success' in human or animal group behavior? In team sports, such as basketball, American football or soccer, 'success' can simply be defined as an execution of a team strategy that results in a score on offense or prevents the opponent team from scoring on defense. In animal group behavior, success can be defined as the protection of the group members from predators or finding a food source. However, if one considers performance art, such as dance, the definition of success might not be that clear, since the overall goal in art forms is not as explicit compared with those in sports or animal behavior. Nevertheless, we observe that, in dance competitions, such as the popular television show *Dancing with the Stars*, the judges' scores consistently agree with each other, which means that they may use similar performance metrics in judging the performances.[1] This brings up the question of whether there exists a formal way of evaluating a dance performance. In what follows, this question is examined by using a transition system model representation of salsa dance and by incorporating performance metrics to evaluate salsa performances.

A vast amount of work on understanding animal and human collective behavior has been published, e.g. [1]–[8]. One subject that the researchers are interested in is how each member within a group behaves individually to accomplish a shared group objective. In study [1], Young *et al.* show that a group of starling birds can maintain the cohesion of the group by each individual only interacting with its closest neighbors. The study [3] shows that a bee in a hive can use dance-like motions to communicate the distance and direction of a food source to other bees. Leonard *et al.* [2] discuss *flock logic* wherein a group of people is constrained to nearest neighbor interactions based on simple rules assigned to individuals. It is argued that constrained local interactions of the group drive the whole group to generate recognizable motion patterns.

---

[1]It could also mean that the judges influence each other. An AI judge would presumably be immune from such influence.

Lessons learned from animal and human collective motion have inspired applications to build artificial multi-agent systems [9]–[11]. The study [9] shows how to generate a trajectory that is synchronized with the musical beats for each individual quadcopter so that the whole group can perform a choreographed robotic dance. In [10] and [11] ballet is used as an art form to study how to construct ballet movements for humanoid robots. A state transition model is built with each state representing a transition between two ballet poses. The model is then used to generate ballet phrases. In order to define success, the term 'expressivity' is introduced as a figure of merit to convert the movement generation problem into an optimization problem.

However, one interesting challenge that remains to be addressed is understanding the interactions between humans and artificially intelligent systems. There exist studies such as [12]–[15] that investigate the social element of human robot interactions in various settings such as communication and artistic reflection. In line with these studies, here we study a setting in which an artificially intelligent (AI) observer is required to assess artistic merit in human dance performances based on criteria learned from human judges.

In conducting a formal study of this particular human-AI interaction, we investigate the dance form salsa. We explore two performance metrics in salsa dance via formal constructs. The first one is purely related to the skill level of the performers. Different performance levels can be captured by the number of moves a dancer can perform. In what follows, we propose two different levels. They are beginner level and intermediate level, each of which are represented by a transition model. Different from the ballet as studied in [10], a salsa performance involves a pair of dancers, a leader and a follower. The leader is generally a male dancer who is responsible for choosing the sequence to be performed and for signaling his decisions by gestures and motions. The follower is generally a female dancer who is responsible for executing the corresponding moves that are communicated by the leader. The second performance metric aims to quantify the artistic appeal of an execution that is perceived by the audiences. Here we introduce, two distinct components of measured artistic appeal. The first component measures the energy consumed by the dancers in an execution. It is shown that more energetic performances are more likely to be favored by audiences. The second component measures the diversity of the moves in a performance. Audiences generally prefer a performance that does not involve too much repetition. Finally, the proposed metrics will be used to construct an artificially intelligent (AI) judge that is capable of recognizing the dance moves and evaluating the quality of a dance performance.

Parts of the discussion that follow are based on our previously published work in the American Control Conference (ACC) [17], IEEE Conference on Decision and Control (CDC) [18] and [19]. The performance metrics and the transition system representation of salsa dancers have previously been used in the studies [17], [18] to investigate

a *forward problem*: generating choreographed automated dance sequences by humanoid robots. In this study, we investigate an *inverse problem* which involves the evaluation of a dance execution.

A. WHY SALSA?

Salsa is a Latin dance form which is popular around the world [20]. Different from other dances, which are generally the result of years of practice, two dancers without any prior practice can perform and enjoy salsa. This is achieved by a universal set of moves and communication signals that can be easily learned by both the leader and the follower dancers. Hence, equipped with the prior knowledge of the moves, the dancers can perform salsa as long as the leader as a decision maker executes the correct gestures to communicate and the follower correctly estimates the upcoming moves during the dance. Salsa can be seen as a particular type of collective motion in which the collective goal is to perform an artistically appealing dance while each individual has to fulfill his/her role as a leader or a follower.

In order to study salsa formally, in this paper we are going to use two mathematical models that use two key features of salsa. The first feature is that every distinct move in salsa has to be performed in eight musical beats by the dancers. This enables us to discretize a salsa performance into moves of eight beat intervals and to assign a letter to each move from a finite-sized alphabet $\mathcal{M} := \{A, B, \ldots\}$. By this method, a salsa performance can be represented as a concatenation of letters (one might think this as similar to a DNA sequence in biology). The second feature of salsa is the characteristics of leader-follower interaction. The leader (generally a male dancer) is responsible to communicate with the follower (generally a female dancer) by using gestures in order to signal his move decisions. Here, we use $\mathcal{S} := \{S_A, S_B, S_C, \ldots\}$ to represent the collection of the signals communicated by the leader to the follower to signal the corresponding moves from the set $\mathcal{M}$. For instance, the leader can push and pull the arms of the follower (S_A) to signal backward and forward steps (move A) [17].

II. SALSA WITH TOPOLOGICAL CONSTRAINTS

We define two different levels of salsa performances based on the size of the move set. *Beginner Level Salsa* (BLS) is defined as the performance with four fundamental moves which are assumed to be the foundation of advanced level salsa performances as well. In BLS every move starts and ends with the same pose, the leader chooses the move sequence from the set of four moves $\mathcal{M}_{BLS} := \{A, B, C, D\}$ without any constraints on the dance move transitions. *Intermediate Level Salsa* (ILS) is defined as the performance extending the alphabet of possible moves from four to eleven in order to capture advanced level salsa performance (Fig.1).

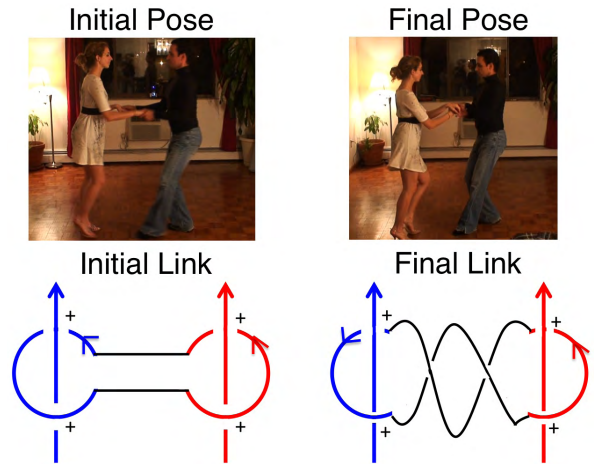$$\mathcal{M}_{ILS} := \{A, B, C, D, J, K, M, N, O, P, T\}. \quad (1)$$

Notice that $\mathcal{M}_{BLS} \subset \mathcal{M}_{ILS}$. The move set $\mathcal{M}_{ILS}$ is illustrated in Fig. 1. Each move involves an initial (p_i) and a final
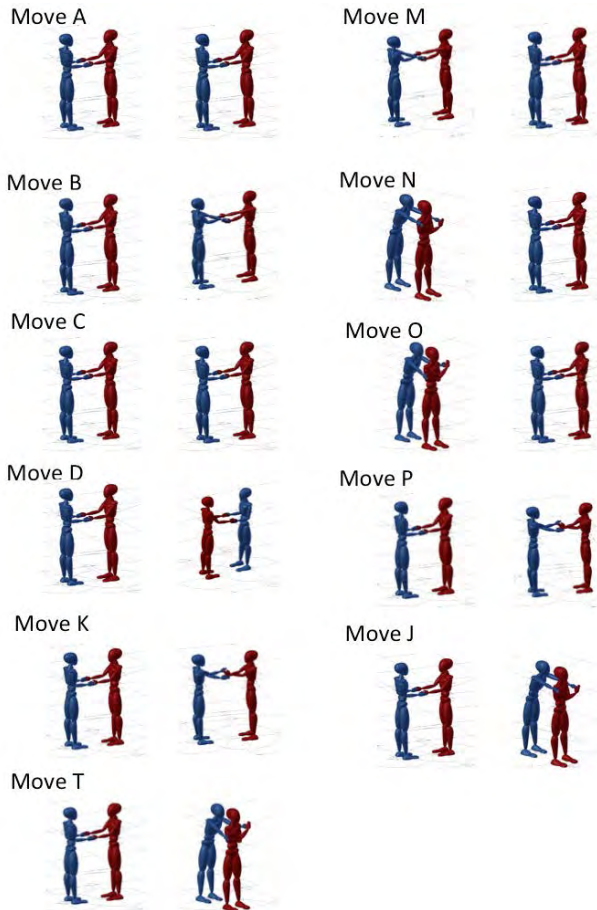
FIGURE 1. Initial (p_i) and final (p_f) pose illustrations for the eleven moves in Intermediate Level Salsa. The blue agent represents the leader and red agent represents the follower. For instance, move K starts with the initial pose shown on the figure and during the execution the follower rotates 2π in counter clock wise direction without breaking the hand contact with the leader which results as the corresponding final pose. More details about the moves and their physical descriptions can be found in [19].


FIGURE 2. An example illustration of the topological link representation of the initial and final pose of move B in ILS. Dancers and their arms are represented by a three component link with a fixed orientation. The link diagram representation is used to represent the dancers' arms by distinguishing the arm overpasses and underpasses. Link invariants, such as *linking number* and *Alexander Polynomials* are calculated by investigating each crossing and assigning positive or negative signs with respect to the overpasses or underpasses for each link diagram. More details about the topological features of the link diagram representations for the dance poses can be found in [19] and [23].

(p_f) pose that occur in ILS. One key difference between BLS and ILS is that in ILS the dancers are not allowed to break hand contact through the performance. Hand constraint plays a major role in investigating move transitions based on simple rules expressed in the language of topological knot theory [19], [21].

To study the leader-follower interactions in ILS, each pose can be described by a link in the topological knot theory, [22] (Fig. 2). The link diagram representation allows us to understand how move transitions are restricted by arm constraints in ILS. Since every move in $\mathcal{M}_{ILS}$ will end with a corresponding final link, the next admissible move has to have the same initial link for the transition to occur. This defines the admissible move transitions which are depicted in Fig. 3. Another problem is to investigate the transformation from an initial to final link diagram in a move. The interconnection between the physical dance moves and link transformations is studied by introducing a physical operator $*(.,.)$. In this operator, the first component represents the follower's

rotation (in radians) and the second component represents the direction of the rotation (CW for clockwise and CCW for counterclockwise).

For instance, in move A, both the leader and the follower dancer step forward and backward. Since move A does not involve any follower rotation, it is represented by $*(0, 0)$. However, in move B, the follower rotates 2π in clockwise direction which is represented by $*(2\pi, CW)$. The connection between the physical moves and the topological link transformations are discussed in more details in [19]. In salsa the leader is assumed to be the decision maker for the sequence construction, and the allowable dance move transitions decided by the leader are based on the syntactic requirement of matching the topology of the initial pose of a move with the topology of the final pose of the preceding move.

One may observe from the Fig. 3 that there exist both deterministic and nondeterministic transitions between the moves in ILS. For instance, the moves T, J, K, B are followed deterministically by the moves O, N, P, M, respectively. On the other hand, the moves in the set $\{A, C, D\}$ can be followed by any move from the set $\{A, C, D, T, J, K, B\}$. Differing from Beginner Level Salsa, arm constraints play a major role for the leader's decision making for move transitions. Fig. 3 gives an insight on the dance move transitions that appear in a real performance. If we assume that a dance pair starts a salsa performance with move A, which has a final pose with the dance pair holding hands without any arm crossings, then the leader has an admissible set $\{A, C, D, T, J, K, B\}$ of move transitions. Notice that all of the moves in the set have the same initial pose coinciding with move A's final pose.
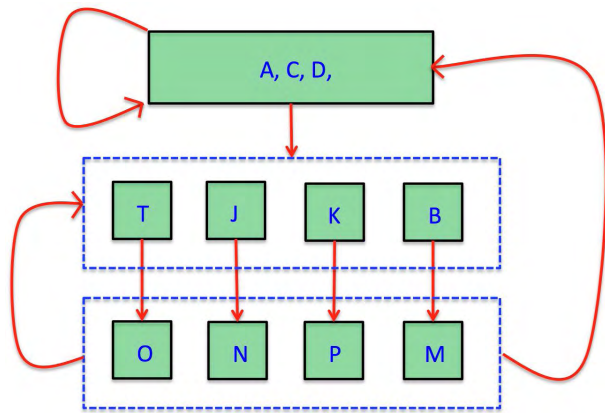
FIGURE 3. The allowable transitions for each move in ILS based on the physical/topological constraints. Admissible transitions are defined for the equivalence of the final link of the move executed and the initial link of the upcoming move. In the figure red arrows illustrate the admissible move transitions including deterministic and nondeterministic transitions. The admissible set is found by using the topological link diagram equivalences of the poses of the moves in ILS. For an admissible transition, the final link of a move and the initial link of the next must be equivalent. In the physical world, this corresponds to the moves that can be executed by the dance pair despite the arm constraints.
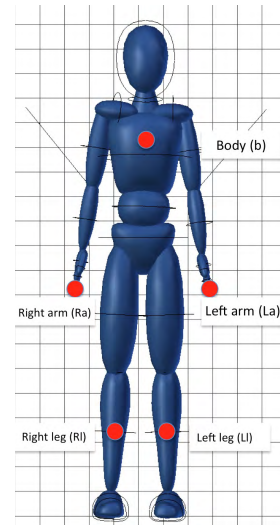

FIGURE 4. A humanoid robot that represents an actual dancer. The red circles represent five fiducial points that are named as the Body, Left/Right arm and Left/Right Leg. The number of fiducial points are chosen to distinguish the moves in Beginner Level Salsa and Intermediate Level Salsa.

On the other hand, move B involves the follower dancer's 2π rotation in the clockwise direction. Thus, it has to be followed deterministically by move M which is the follower's 2π rotation in counter clockwise direction. The deterministic transition is a result of the arm constraint since arm crossings in the final pose of move B make it physically unfeasible for the follower to rotate more in the CW direction. More detail about the topological constraints on dance move transitions can be found in [19]. The allowable transitions in BLS and ILS (Fig. 3) can be used to build a state transition model representation of the dancers.

III. SALSA AS A TRANSITION SYSTEM MODEL
In this section, we introduce a finite state machine representation of a pair of salsa dancers. We first define two separate transition system representations for the leader dancer, Bob, and the follower dancer, Alice. Finally, we combine them into a single model through interactions between the two, which is realized through the signal sent from Bob to Alice.

In order to describe the dancers' movement as well as track their motions, we use a humanoid robot to represent different parts of a dancer's body as illustrated in Fig. 4. It is assumed that there exist *fiducial points* (red circles in Fig. 4) that represent tracked points on the humanoid robot representation. The number and locations of the fiducial points are chosen in order to be able to distinguish the moves performed in BLS and ILS.

We define a state q to represent a pose in dance which is a discretization of a move, and a salsa performance becomes a concatenation of such poses. Moreover, q_{Al} and q_{Bo} represent the discrete states of the leader and the follower in a pose, respectively. This framework is similar to the work presented in [10] which shows the construction of a transition model for the poses that occur in a ballet warm up routine. Our study

differs from [10] in that we build a model of a pair of dancers. Such a model requires modeling the communication between the two dancers.

Definition 1: State transition model of a leader dancer and his communication strategy is given as

$$G_{Bo} = (Q_{Bo}, Act_{Bo}, \rightarrow_{Bo}, q_{Bo}^0, !). \qquad (2)$$

Q_{Bo} is the set of all possible states representing the initial p_i and final p_f dance poses (Fig.1). Act_{Bo} is the set of all possible actions (set of physical moves represented by the $*(.,.)$ operator). $q_{Bo}^0 \subseteq Q_{Bo}$ is the set of initial states. $!(Act_{Bo})$ is the formal representation of leader's decided action Act_{Bo} that is transmitted to the follower (using his gestures and motions). $\rightarrow_{Bo}: (q', a) \mapsto q$ denotes a transition relation (based on the topological constraints given in Sect.II. In this expression, $q' \in Q_{Bo}$ is the initial state, $a \in Act_{Bo}$ is the action that is executed by the leader and $q \in Q_{Bo}$ is the next state.

Similarly we can represent the follower dancer (Alice) by the transition system,

$$G_{Al} = (Q_{Al}, Act_{Al}, \rightarrow_{Al}, q_{Al}^0, ?), \qquad (3)$$

where definitions of the components Q_{Al}, Act_{Al}, q_{Al}^0 in Eqn. (3) are identical to those of the Eqn. (2) but the subscripts are replaced with Al to represent Alice. $?(Act_{Bo})$ is the representation of signal received by the follower such that she can make the proper move based on the signal she receives from the leader dancer. Further, her transition $\rightarrow_{Al}$ is defined as a mapping $(q', ?, a) \mapsto q''$. Based on $?(Act_{Bo})$, the signal she receives from the leader, she can execute an action a to move from state q' to q''.

Definition 2: The dance pair (Fig. 5) is represented by a single transition system by composing the two transition systems (2) and (3). The combined system is

$$G = (Q, Act, \rightarrow, q^0), \qquad (4)$$
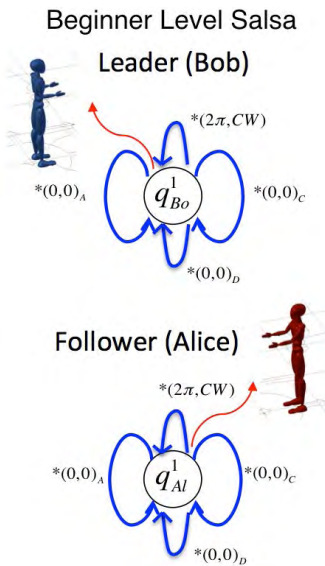
FIGURE 5. The transition system representation of BLS, [18]. All the moves in BLS have the same initial and final pose. There exists only one state for the state machine representation of the leader and follower. Blue arrows represent the transitions between the states which correspond to the physical dance moves that are defined by the $*(.,.)$ operator. Move A, C and D are represented by $*(0,0)$ since there is no rotation of the follower. However, move B is represented by $*(2\pi, CW)$ which corresponds to the follower's rotation of 2π in the clockwise direction.

where $Q = Q_{Bo} \times Q_{Al}, Act \subseteq Act_{Bo} \times Act_{Al}, q^0 \subseteq q^0_{Al} \times q^0_{Bo}$. The composition is achieved by synchronizing information sent by the leader and received by the follower. A detailed explanation of such composition can be found in textbooks on discrete systems such as [24] and [25]. Notice that a perfect synchronization is assumed here, i.e., the follower has no difficulty in interpreting the intended next move communicated by the leader.

It is assumed that there exists a synchronous message passing between these two transition systems such that follower can estimate the upcoming move perfectly without an error [18]. In this paper, we are going to use model (4) to recognize the sequences of movements that are being taken by a dancer pair (Fig. 6) from the recordings of the locations of the fiducial points.

The transition system model for a BLS based on (4) is shown in Fig. 5. In BLS all of the moves start and end with same pose ([17], [19]). Thus, in Fig. 5, q^1_{Bo} and q^1_{Al} represent the states which correspond to the poses of Alice and Bob, respectively. Blue arrows represent the state transitions based on the physical motions (described by the $*(.,.)$ operator) required to perform each move A, B, C and D. Since, the follower does not initiate a rotation in the moves A, C and D, the first and second components of the $*(.,.)$ operator are zeros even though they are distinct dance moves. Distinctions in these moves are represented by the subscripts $*(0, 0)_A$, $*(0, 0)_C, *(0, 0)_D$. Bob sends his decision transition to Alice through a signal $!(Act_{Bo})$ and Alice has the simultaneous transition based in the signal $?(Act_{Bo})$ she receives. For instance, if the agents perform move A, the leader has a transition $q^1_{Bo} \rightarrow q^1_{Bo}$ that is signaled to Alice. $!(*(0, 0)_A)$
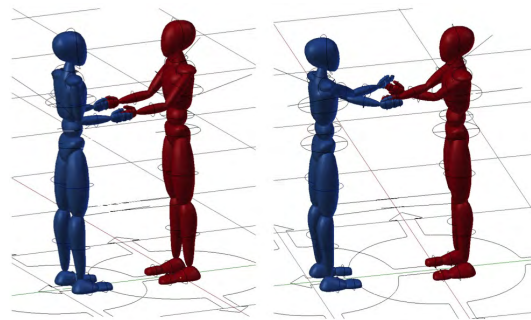
FIGURE 6. Robotic representations of the leader (blue) and follower dancer (red). The initial and final poses of move B in ILS are illustrated by the left and right sub-figure, respectively. Initial pose occurs in the first beat and final pose occurs in the last beat of a move performed in eight beats. The arm positions in the final pose corresponds to the followers rotation without breaking the hand contact.

is the formal representation of Bob pushing Alice's hand to make her move backward. Alice has the corresponding transition $q^1_{Al} \rightarrow q^1_{Al}$. Hence, the move A can be represented by a change of the state of the overall system (4) as $< q^1_{Bo}, q^1_{Al} > \rightarrow < q^1_{Bo}, q^1_{Al} >$.

The transition graph of the ILS is much more complex due to the constraints that force the dancers to keep hand contact through the dance. In Fig. 7, states of the leader and follower including the transitions for ILS (blue arrows) are depicted. The effect of the arm constraint can be observed from the final poses of move B, which are represented by q^1_{Bo} in BLS (in Fig. 5) and q^2_{Bo} (in Fig. 7) in ILS. In ILS, since the dancers are not allowed to break their hand contact, rotation with arm constraints will result in a different final pose in performing move B in ILS than move B in BLS although the dancers execute identical motions.

The associated dance poses in ILS shown in Fig. 1 are decomposed into the poses for the leader and the follower in Fig. 7. If one considers the initial and final poses illustrated in Fig. 1, same poses occur when the agents are in the states $< q^1_{Bo}, q^1_{Al} >$ and $< q^5_{Bo}, q^5_{Al} >$, respectively.

Using the transition models defined for BLS and ILS, a dance sequence can be observed by the following form,

$$[< q_{Bo}, q_{Al} >_{ji}, < q_{Bo}, q_{Al} >_{jf}],$$
$$[< q_{Bo}, q_{Al} >_{ji}, < q_{Bo}, q_{Al} >_{jf}], \dots \quad (5)$$

where i and f stand for the initial and final state (pose), respectively, and $j \in \mathcal{M}_{BLS}$ for BLS or $j \in \mathcal{M}_{ILS}$ for ILS and each bracket represents a move.

We use this bracket representation to distinguish the moves (letters) performed by the dancers. For such a purpose, the initial and final states of the leader and follower dancers need to be observed. Moreover, observation of the transition relations are needed in order to avoid ambiguities caused by the moves that start and end with the same pose. The recognized sequence of moves can be used to evaluate the 'perceived success' of the execution, which will be shown in the next section, as well as to eventually build an artificially
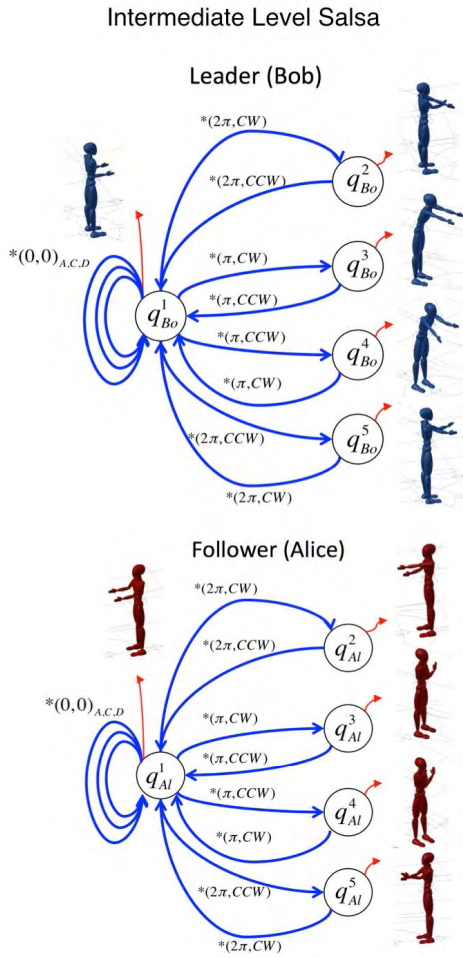
FIGURE 7. The transition system representation of ILS, [18]. Leader and follower have five distinct states to represent the distinct poses in ILS. Blue arrows represent the transitions between the states. Corresponding $*(.,.)$ operators for each transition are illustrated which involve the follower's 2π / π rotations in clockwise/counter clockwise directions. The notation $*(0,0)_{A,C,D}$ is used to represent the moves A, C and D that are already defined in BLS and that do not involve any follower rotation.

intelligent (AI) judge, which can recognize the moves and then give a score (Section V).

IV. METRICS TO EVALUATE SALSA

A formal model of a salsa dance performance enables us to solve two problems, a forward problem and an inverse problem. Based on the underlying structure of the leader-follower interactions during dance sequence generation, the forward problem aims at automatically generating choreographed dance sequences for the two robotic agents as well as a communication protocol to achieve a satisfactory performance [18]. In what follows we focus on solving the inverse problem, i.e., to understand the notion of 'perceived success' and what constitutes 'optimal dance'. We propose two metrics for BLS: *energy* and *phrase complexity* metrics. The energy metric is defined as the distance (in hectometers) covered by the dancers in the execution of the moves in $\mathcal{M}_{BLS}$. The total energy of a dance sequence is calculated by finding the frequencies of the letters and multiplying those

with their energy values. Phrase complexity is computed by finding the entropy of each 4-letter phrase in a sequence. Entropy of a phrase, [26], generated by the moves in BLS is

$$w = -\sum_{k=1}^{4} f_k \log_2 f_k, \quad (6)$$

where f_k is the frequency of each letter in a phrase. Then the *average phrase complexity* (W_{ave}) of a sequence is defined as the fraction of the sum of phrase entropy values to the number of phrases that appear in a sequence.

We further integrate these two metrics with the transition model (4) for the evaluation of a salsa performance. If we assume that a dance sequence is deconstructed as the bracket representation given in (5), then the deconstructed sequence can be evaluated based on the energy and phrase complexity metrics. To give an an example, assume that the agents Alice and Bob generate 20-letter long sequences. We partition a 20-letter sequence *Seq* into 4-letter phrases and assign a complexity value for each phrase as in the following.

$$Seq = [\underbrace{\overbrace{----}^{w_1}][\underbrace{----}_{4\text{ letters}}][----][----][----]}_{20\text{ letters}}$$

$$(7)$$

For this particular case, we define the entropy vector W such that $W = (w_1, \ldots, w_n)$ where $w_i \in \{0, 0.811, 1, 1.5, 2\}$, $i = 1, \ldots, n$ where n is the number of 4-letter phrases in a sequence ($n = 5$ in (7)). The possible values of w_i correspond to the entropy values calculated by the frequency of each letter in a 4-letter phrase in (6), e.g. phrase ($AAAA$) has phrase complexity $w = 0$ but ($BACD$) has phrase complexity $w = 2$. Below, we illustrate two possible sequences generated by dancers with the moves from BLS. The first sequence is generated randomly and the second sequence is generated such that the phrase complexity decreases through the sequence with $W_{Seq2} = (2, 1.5, 1, 0.811, 0)$.

*Seq*1: (ABCA)(BCAD)(BBBC)(BCDA)(BBDA)

*Seq*2: (ABDC)(BCAC)(CCBB)(DADD)(BBBB)

Similarly, an energy vector, $E = (e_1, \ldots, e_n)$ can be introduced for a sequence such that e_i, where $i = 1, 2, \ldots, n$, corresponds to the energy consumed by the agents in performing the ith, 4-letter phrase in a sequence. The *energy* metric for a sequence is defined as the total energy, $E_{total} = e_1 + \ldots + e_n$ consumed by the dancers to perform a dance sequence.

In order to compare the proposed metrics with humans' perceptions of artistic value, the video recordings of ten salsa sequences (performed by using the letters from the set $\mathcal{M}_{BSP}$) are shown to a dance audience who is asked to evaluate the performances. Strong correlations between these metrics with the audience's scores are reported and the details of the performance metrics are presented in Section V. High correlation between audience's scores and the consumed energy by the dancers (energy metric) implies that the dance audience
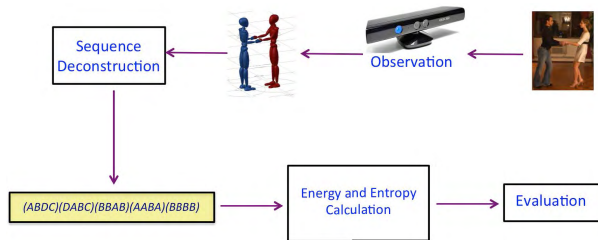
FIGURE 8. An artificially intelligent (AI) judge scheme to evaluate salsa. The AI judge involves two components: An *observation component* and an *evaluation component*. The *observation component* consists of a visual sensor to track the fiducial points on the humanoid representation of the human dancers. The tracked values are compared with a library of poses and move transitions. The sequence is deconstructed by recognizing the initial and final pose in a move as well as the transition between them. The deconstructed sequences are fed to a Score function in the *evaluation component* which assigns a score based on the Energy and Phrase Complexity metric (Eqn. (8)).

likes more energetic dance performances. This phenomenon is used by the choreographers by placing the most energetic dance sequence in the finale of a dance show in that it is believed that the last section of a show will be the most memorable by the audiences. However, if an energetic move is repeated many times, it may become boring for the audiences. Hence, a choreographer as well as dancers need to balance between the energy and diversity of a constructed dance sequence. Thus, the order of the moves in a sequence is also relevant, and this is captured in our model by the phrase complexity metric.

V. AN ARTIFICIALLY INTELLIGENT (AI) JUDGE

We are particularly interested in building an artificially intelligent (AI) judge that observes and evaluates the artistic success of a dance performance. The idea is similar to the judges that appear in the Olympic games or dance competitions. The judges in these contests have criteria that measure artistic reflection and also the complexity of the execution. It would be difficult for an AI judge to evaluate the warmth of a dancer's smile but instead it can evaluate artistic appeal of a performance by using the energy and complexity metrics. The overall scheme of an AI judge evaluating performance art is shown in Fig. 8.

Our AI judge has two components: an observation component and an evaluation component. We use the abstract model given in (4) to represent the leader and follower dancer as humanoid robots with tracked fiducial points. The goal of the observation component is then to estimate a sequence of states and the transition $(Q_o, \rightarrow_o)$ that best fit the observed sequence of tracked fiducial points and the model. Here, Q_o is the set of observed states including the initial and final states (poses) of the leader and the follower in a move and $\rightarrow_o$ is the observed transition between the initial and final state such that $\rightarrow_o: (q_i, a) \mapsto q_f$ where $q_i, q_f \in Q$ and $a \in Act_o$ represents the actions that are executed by the agents between the initial pose and final pose.

The purpose of the observation component is to deconstruct a salsa performance into a letter (move) sequence with the bracket representation proposed in (5). This is achieved by

detecting the x-y-z coordinates of the fiducial points defined in Section III. This is similar to the idea of template matching which is widely studied in computer vision [27]. Simply, the tracked points' coordinates are compared with the values that are contained in a library of poses with an allowed deviation δ. After detecting initial and final pose, the algorithm resets and starts to track the new move. In the previous sections, we have shown that there may be multiple moves with equivalent initial and final poses. To avoid ambiguities in dance move detection, we also include the observation of the transition $\rightarrow_o$. This transition is captured by tracking the velocities v_x, v_y and v_z for each fiducial point.

The evaluation component first decomposes the observed sequence into 4-letter phrases. It then calculates observed phrase complexity W_o and observed phrase energy E_o as described in Section IV. It finally computes the score of the observed sequence based on the following *Score* function, $Score = a.E_{total} + b.W_{ave} + c$. This function is a linear combination of E_{total} and W_{ave} where E_{total} is the sum of the energy values of the phrases and W_{ave} is the average phrase complexity that is calculated by dividing the total phrase complexity value by the total number of phrases that appear in a sequence.

The Score function is constructed as a linear function in that it fits the evaluations collected from a previous study which is reported in [17]. The study involved two dancers who performed ten distinct salsa sequences by using the moves in BLS. The dance sequences were recorded as video and shown to a dance audience who was asked to evaluate the videos with scores from 1 to 10. The scores of twenty judges were collected and the averages were calculated for each sequence. It is shown that the judges' scores are highly correlated to total energy E_{total} (with a correlation coefficient R=0.8) and average phrase complexity W_{ave} (R=0.75). Thus, in this study we use previous data as a training set to estimate the coefficients a, b and c in Score function (Fig. 9). The Score function learned from the data has the form

$$Score = -17.94 + 16.E_{total} + 0.833.W_{ave}. \qquad (8)$$

In order to validate our AI judge, we asked our experienced salsa dancers to perform four new dance sequences (each of them having 20 letters) by using the moves in BLS. All of the sequences are performed by the same two dancers in order to exclude the effect of artistic reflection of a dancer's personal demeanor. The sequences constructed by the dancers are given below.

$$V1: (BDCB)(DBCB)(DDBB)(CCDB)(DDBB)$$
$$V2: (BBBB)(BBBA)(ACAA)(AAAD)(AAAB)$$
$$V3: (AAAA)(ABAA)(BABA)(DACA)(DABC)$$
$$V4: (ABDC)(DABC)(BBAB)(AABA)(BBBB)$$

The video recordings of the dance sequences are shown to a dance audience with a random order. Average scores in 1-to-10 scale that are collected from 15 judges are, $Score_{V1} = 9.1$, $Score_{V2} = 3.09$, $Score_{V3} = 3.7$ and $Score_{V4} = 5.79$.
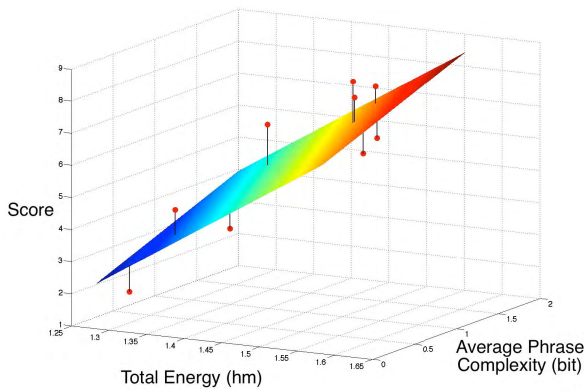
FIGURE 9. The least squares regression plane with coefficients estimated from the previous data [17]. The x-coordinate is the average phrase complexity of a dance sequence (bits), the y-coordinate is the total energy consumed by the dancers (hectometers) and the z-coordinate is the associated score assigned by the human judges.

FIGURE 10. A snapshot of the User Interface of the AI judge that uses Kinect C++ algorithm to track the fiducial points on the stick figure representations of the dancers. Recognized dance move is shown on the right bottom corner to the user. Video of the constructed AI judge detecting a dance sequence is available at https://www.youtube.com/watch?v=eHX26GGBB3A&edit=vd.

The same video recordings are fed into the AI judge. For the observation component, we use a Microsoft Kinect sensor in order to track the x-y-z coordinates of the fiducial points. Microsoft's open source C++ algorithm is modified for the purpose of our experiment such that two distinct libraries are contained in a movement library and transition library. The q vectors that represent the poses of the leader and the follower are integrated to the Kinect algorithm as a library so that the algorithm seeks to match the tracked coordinates of the fiducial points to one of the possible poses from the library with the maximum deviation δ. Moreover, a library of move transitions is incorporated into the C++ code which includes the deviations of the coordinates with respect to time for distinct physical moves described by the $*(.,.)$ operator. The initial pose shown in the upper-left corner of the Fig. 1 is incorporated into the algorithm as a trigger to start tracking fiducial points of the dancers (Fig. 4).

The timer starts and stops with the recognition of initial pose and final pose respectively. In Fig. 10, a snapshot of the algorithm is shown including the stick figure representation of a dancer and the detected letter which is illustrated in the right bottom corner. Finally, the algorithm computes the E_{total} and W_{ave} values for a recognized sequence which are then supplied to the Score function given in Eqn. (8). The dancers performed the sequences given as $V1$, $V2$, $V3$ and $V4$ in the view of the Microsoft Kinect sensor. Four sequences are deconstructed by the AI judge and average phrase complexity and energy values are computed for each sequence which are then fed to the Score function. The score values are calculated as 5.29, 2.88, 4.28, 3.69 for the sequences $V1$, $V2$, $V3$ and $V4$ respectively. The correlation coefficient between the audience's scores and the scores assigned by the AI judge is calculated as $R = 0.81$. Although the range of the scores given by the audience are much higher than the AI judge, the trend and the ranking of the scores are very similar. The difference between the scales may be a result of initial individual biases in the scale for the audience. Hence we conclude that the strong correlation implies that our judge
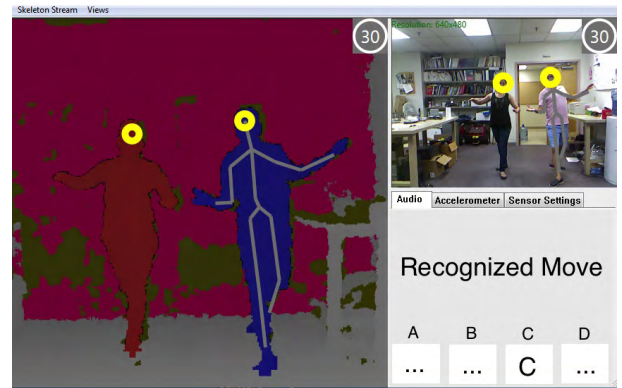
performs well enough in matching human's perceived relative artistic appeal for salsa performances.

VI. CONCLUSIONS

This study proposes a method to formally define artistic value in human collective motion. A popular form of performance art, salsa, is used as a prototype model for the analysis. The state transition model of salsa is constructed by using the states that correspond to the initial and final poses of the abstract representations of the salsa dancers. We introduce abstract representations of the dancers, and the transitions between states are introduced as the physical dance motions executed by the dancers. The model is then used as a base for the recognition of the dance moves and the evaluation of the detected sequences with respect to the metrics based on the energy and entropy (diversity) of the dance phrases. The general scheme of an artificially intelligent (AI) judge is introduced along with its components that we refer to as the observation and evaluation components. A score function is proposed that assigns a score based on the metrics calculated for the detected sequences. Finally, an implementation of the AI judge to evaluate BLS is shown by using the Microsoft Kinect sensor for fiducial point tracking and the C++ algorithm for the evaluation. Extending the applications to ILS is still on progress in that it requires a modified Energy metric since in ILS, arm and body movements have major influence on dancers' energy consumption and perceived artistic success.

The idea of an AI judge can be extended to other fields such as athletic competitions or other dance contests. One may think of a diving competition in the Olympics as an example. An AI judge can be constructed to capture the acrobatic motion sequences performed by the divers and the metrics can be modified to match the judging criteria for this particular contest. In all such applications an appropriate alphabet of motion primitives is needed to capture all of the possible moves. Ideally, an AI judge with fixed evaluation metrics can be potentially the most unbiased judge in any competition.

REFERENCES

[1] G. F. Young, L. Scardovi, A. Cavagna, I. Giardina, and N. E. Leonard, "Starling flock networks manage uncertainty in consensus at low cost," *PLoS Comput. Biol.*, vol. 9, no. 1, p. e1002894, 2013.

[2] N. E. Leonard *et al.*, "In the dance studio: Analysis of human flocking," in *Proc. Amer. Control Conf. (ACC)*, Chicago, IL, USA, Jun. 2012, pp. 4333–4338.

[3] K. von Frisch, T. D. Seeley, and L. E. Chadwick, *The Dance Language and Orientation of Bees*, 1st ed. Cambridge, MA, USA: Harvard Univ. Press, 1993.

[4] S. G. Reebs, "Can a minority of informed leaders determine the foraging movements of a fish shoal?" *Animal Behav.*, vol. 59, no. 2, pp. 403–409, 2000.

[5] T. D. Seeley and P. K. Visscher, "Group decision making in nest-site selection by honey bees," *Apidologie*, vol. 35, no. 2, pp. 101–116, 2004.

[6] M. Lindauer, "Communication in swarm-bees searching for a new home," *Nature*, vol. 179, pp. 63–66, Jan. 1957.

[7] J. M. Black, "Preflight signalling in swans: A mechanism for group cohesion and flock formation," *Ethology*, vol. 79, no. 2, pp. 143–157, 1988.

[8] R. W. Byrne, "How monkeys find their way: Leadership, coordination, and cognitive maps of African baboons," in *On the Move: How and Why Animals Travel in Groups*. Chicago, IL, USA: Univ. of Chicago Press, 2000, pp. 491–518.

[9] A. Schöllig, F. Augugliaro, and R. D'Andrea, "Improving tracking performance by learning from past data," Ph.D. dissertation, ETH Zurich, Zürich, Switzerland, 2012, p. 147.

[10] A. LaViers, Y. Chen, C. Belta, and M. Egerstedt, "Automatic sequencing of ballet poses," *IEEE Robot. Autom. Mag.*, vol. 18, no. 3, pp. 87–95, Sep. 2011.

[11] A. LaViers and M. Egerstedt, "Style based robotic motion," in *Proc. Amer. Control Conf. (ACC)*, Jun. 2012, pp. 4327–4332.

[12] E. Jochum, E. Vlachos, A. Christoffersen, S. G. Nielsen, I. A. Hameed, and Z.-H. Tan, "Using theatre to study interaction with care robots," *Int. J. Social Robot.*, vol. 8, no. 4, pp. 457–470, 2016.

[13] G. Ma, J. Gao, Z. Yu, X. Chen, Q. Huang, and Y. Liu, "Development of a socially interactive system with whole-body movements for BHR-4," *Int. J. Social Robot.*, vol. 8, no. 2, pp. 183–192, 2016.

[14] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, "To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability," *Int. J. Social Robot.*, vol. 5, no. 3, pp. 313–323, 2013.

[15] H. Knight, "Eight lessons learned about non-verbal interactions through robot theater," in *Proc. Int. Conf. Social Robot.* Berlin, Germany: Springer, Nov. 2011, pp. 42–51

[16] K. Özcimder, Z. Kong, S. Wang, and J. Baillieul. (Sep. 9, 2016). *Salsa Robotic Judge*. [Online]. Available: https://www.youtube.com/watch?v=eHX26GGBB3A&edit=vd

[17] J. Baillieul and K. Özcimder, "The control theory of motion-based communication: Problems in teaching robots to dance," in *Proc. IEEE Amer. Control Conf. (ACC)*, Jun. 2012, pp. 4319–4326.

[18] K. Özcimder, Z. Kong, and J. Baillieul, "Algorithmic approaches to artistic movement," in *Proc. 53rd IEEE Conf. Decision Control*, Dec. 2014, pp. 5373–5380.

[19] J. Baillieul and K. Özcimder, "Dancing robots: The control theory of communication through movement," in *Controls and Art*. Basel, Switzerland: Springer, 2014, pp. 51–72.

[20] J. Bosse, "Salsa dance and the transformation of style: An ethnographic study of movement and meaning in a cross-cultural context," *Dance Res. J.*, vol. 40, no. 1, pp. 45–64, 2008.

[21] J. W. Alexander, "Topological invariants of knots and links," *Trans. Amer. Math. Soc.*, vol. 30, no. 2, pp. 275–306, 1928.

[22] K. Murasugi, *Knot Theory and Its Applications*. Springer, 2007.

[23] K. Özcimder, "Communication through motion in dance with topological constraints," in *Proc. IEEE Amer. Control Conf.*, Jun. 2014, pp. 178–183.

[24] R. Alur, *Principles of Cyber-Physical Systems*. Cambridge, MA, USA: MIT Press, 2015.

[25] C. Baier, J. P. Katoen, and K. G. Larsen, *Principles of Model Checking*. Cambridge, MA, USA: MIT Press, 2008.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.

[27] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," in *Proc. IEEE Nonrigid Articulated Motion Workshop*, Jun. 1997, pp. 90–102.

KAYHAN ÖZCIMDER (M'15) received the B.Sc. degree from Gazi University, Ankara, Turkey, 2007, and the M.Sc. and Ph.D. degrees from Boston University in 2012 and 2015, respectively, all in mechanical engineering. He is currently an Associate Research Scholar with the Mechanical and Aerospace Engineering Department, Princeton Neuroscience Institute, Princeton University. He is also specialized in the applications of control theory. His research interests are discrete-event systems, bio-inspired control, and communication through motion in multi-agent robotic systems.

ZHAODAN KONG (M'11) received the bachelor's and master's degrees in astronautics and mechanics from the Harbin Institute of Technology, Harbin, China, in 2004 and 2006, respectively, and the Ph.D. degree in aerospace engineering with a minor in cognitive science from the University of Minnesota, Twin Cities, MN, USA, in 2011. Before joining the University of California (UC) at Davis in 2015, he was a Post-Doctoral Researcher with the Laboratory for Intelligent Mechatronic Systems and the Hybrid and Networked Systems Lab, Boston University. He is currently an Assistant Professor in mechanical and aerospace engineering with UC Davis. His research focuses on control theory, machine learning, formal methods, and their applications to human–machine systems, cyber-physical systems, and neural engineering.
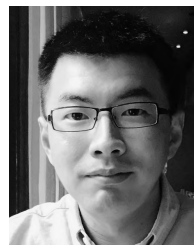
SHUAI WANG (S'15) received the B.S and M.S. degrees in electrical engineering from Xi'an Jiaotong University in 2010 and 2013, respectively, and the Ph.D. degree in systems engineering from Boston University, Boston, MA, USA, in 2018. His research interests include electricity market design, control systems, graph theory, and stochastic optimization.

JOHN BAILLIEULL (M'83–F'93) received the Ph.D. degree in applied mathematics from Harvard University, Cambridge, MA, USA, in 1975, under the supervision of R. W. Brockett. He has published research dealing with robotics, the control of mechanical systems, and mathematical system theory in all of its manifestations and applications. His early work dealt with connections between optimal control theory and what came to be called sub-Riemannian geometry. Other early work treated the control of nonlinear systems modeled by homogeneous polynomial differential equations. Such systems describe, for example, the controlled dynamics of a rigid body. His main controllability theorem applied the concept of finiteness embodied in the Hilbert basis theorem to develop a controllability condition that could be verified by checking the rank of an explicit finite dimensional operator. He has involved in the qualitative theory of electric energy systems, and over the past decade, he has turned his attention to the study of information-based control. Together with K. Li, he has explored source coding of feedback signals that are designed to provide optimally robust performance in the face of time-varying feedback channel capacity. He has also collaborated with W. S. Wong on research aimed at understanding information complexity in terms of the energy required for two (or more) agents to communicate through the dynamics of a control system. He is currently a Distinguished Professor of engineering with Boston University. He is a fellow of the International Federation of Automatic Control and the Society for Industrial and Applied Mathematics.

• • •