Towards a model of nonverbal leadership in unstructured joint physical activity

Radoslaw Niewiadomski
Casa Paganini - InfoMus
University of Genoa
Genoa, Italy
radoslaw.niewiadomski@dibris.unige.it

Maurizio Mancini
Casa Paganini - InfoMus
University of Genoa
Genoa, Italy
maurizio.mancini@unige.it

Lea Chauvigne
Department of Psychology, Neurosciences & Behaviour
McMaster University
Hamilton, Ontario
chauvila@mcmaster.ca

Antonio Camurri
Casa Paganini - InfoMus
University of Genoa
Genoa, Italy
antonio.camurri@unige.it

ABSTRACT

In this paper, we propose a set of algorithms to compute the cues of the nonverbal leadership in an unstructured joint physical activity, i.e., the joint activity of two or more interacting persons who perform some movements without a predefined sequence and without a predefined leader. An example of such activity can be a contact dance improvisation.

The paper is composed of three parts: cue set, dataset and algorithms. First, we propose a cue set of nonverbal leadership which is grounded on existing literature and studies. It is composed of eight cues that characterize the nonverbal behaviors of the leader in a joint physical activity.

In this paper we also introduce a new dataset. It consists of multimodal data (video, MoCap) of contact dance improvisations. Additionally, sensory deprivation conditions (vision and/or touch restraint) were introduced to collect the evidences of the various strategies used by leaders and followers during improvisation. The dataset was annotated by twenty-seven persons who carried out continuous annotation of leadership in the recorded material.

In the last part of the paper, we propose a set of algorithms that works on positional 3D data (i.e., joints' positions obtained from motion capture data of dancers). Each algorithm models one among the discussed cues of the nonverbal leadership.

CCS CONCEPTS

• **Human-centered computing** → **User model**; • **Applied computing** → *Performing art*;

KEYWORDS

Leadership, dance, nonverbal behavior, social signal processing

1 INTRODUCTION

The ability to determine who leads a group of interacting people is an important research topic both in Computer and Social Sciences. Several computational approaches were proposed to automatically determine the leadership [1, 8, 19, 22]. Most of them focus on human-human interactions in which the verbal communication is complemented by some nonverbal behaviors. A typical example is the meetings (e.g., [8, 19]). These papers exploit nonverbal cues of leadership that are related to the main activity (i.e., speaking and listening), such as the variations of prosody, the number of turn-taking, and so on. When investigating leadership in exclusively nonverbal interactions, researchers often focus on relatively simple physical activities, e.g., such as waving, moving a slider or bowing (e.g., [16]), or pre-defined sequences of movements (in other words a structured activities, e.g., tango improvisation [7]) .

Unlike the previous studies, we aim to examine nonverbal leadership in more complex and exclusively nonverbal interactions. The complete work plan is presented on Figure 1. Our long-term goal is to develop computational model of nonverbal leadership in unstructured joint full-body physical activity.

In particular, in this paper, we focus on parts which are highlighted in grey in Figure 1. We present: (1) a set of nonverbal leadership cues (Section 3), (2) the dataset that we created to test (in future works) our computational approach (Sections 4 and 5), and (3) the algorithms (Section 6) to compute the cues of nonverbal leadership.

Regarding the first item, the nonverbal behaviors in unstructured joint physical activity go beyond simple movement anticipation performed by the leader who is followed by other participants. In our work we aim to consider a large set of nonverbal cues
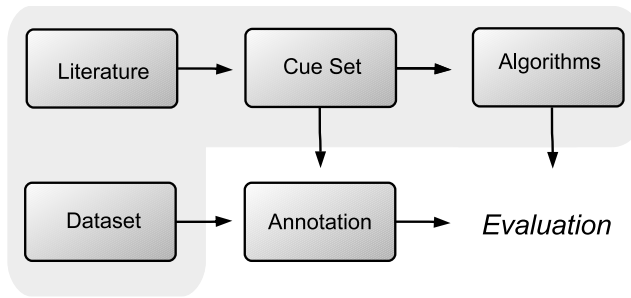
Figure 1: The overview of the work

that is used by human observers to determine who is a leader. An example of such a cue can be the movement stability. According to researchers [7, 16] leaders perform less hesitated and jittered motions than the followers do. We focus on cues that can be observed from full-body movements, therefore, we intentionally skip the cues that can be observed only in facial and audio modality. In Section 3 we describe eight such cues.

Regarding the second item, duos of contact dance improvisation are taken as a testbed to show the validity of our approach. Contact dance improvisation is a good example of joint unstructured physical activity without a predefined leader, where vision and touch can be used to communicate between the dancers. This is a perfect scenario to study nonverbal leadership. Consequently, a set of multimodal recordings (video, MoCap) was performed of the dance duos as well as continuous annotation of the leadership in recorded performances.

Finally, as for the third item, we propose a set of algorithms to compute the nonverbal leadership cues from motion capture data. The algorithms use 3D positions to compute the six out of eight cues. All algorithms work on very short time buffer allowing one to compute cues nearly in real-time. They were implemented in freely available EyesWeb XMI platform developed in InfoMus Lab at University of Genoa [1].

2 STATE OF THE ART

There is a growing interest assessing joint physical activity without designated leader. Most took advantage of the mirror game paradigm developed by Noy and colleagues [16], which consist in two individuals trying to move a slider or a ball along a string in synchrony. The authors showed that expert in joint improvisation are able to reach moment of togetherness where both players initiate the motion together in a predictive nature and with high synchrony. On the contrary, novices synchronize better when leader/follower roles are established, one initiating the motion and the other reacting to it [16]. However, when novices improvise without designated leader, both happen to lead at some point and to various degrees [25]. Furthermore, they coordinate better when their individual motion features are similar [21]. In these studies, the activities of the two players were limited to simple one dimension movement. In that case, the cue characterizing the leader is

the presence of a confident smooth motion while the follower oscillates around the leader's motion, being sometimes ahead and sometimes behind [5, 16].

To date, studies of nonverbal leadership with full-body motion are limited. In the research by Washburn and colleagues [23] a full body imitation task was considered. In that study [23], the leader was always ahead regardless of the expertise of the follower, even if experts displayed a better interpersonal coordination.

A number of studies was realized assessing the nonverbal cues of leadership in organizations. These studies showed that leaders were associated with enhanced facial expression, body openness, more prerogatives to touch, smaller interpersonal distance and increased gaze while speaking [4, 6, 20]. Eye gaze and proximity were specifically associated with emerging leader [3, 20]. Arm and shoulder movements particularly contribute to the perception of leadership [19]. Since it has been argued that the characteristics of leaders in organizations are quite similar to the ones of the leaders in dance [11, 24], the nonverbal leadership cues highlighted during social interactions can probably also be found in a full-body dance improvisation.

3 CUE SET

Having in mind our long-term aim of creating the computational model of nonverbal leadership, we consider the point of the view of the observer. That is, we are interested in the cues that are visible to the person who is not directly engaged in the interaction. We do not investigate the personal motivations (e.g., person willingness to be a leader in a given moment, her intentions, and so on). For the seek of simplicity in the rest of this paper we assume that the joint activity is performed by only two persons.

In Section 2 we briefly discussed different theoretical models and empirical studies on nonverbal leadership. These studies allowed us to define a set of cues that can be observed in full-body behaviors. Consequently, we propose the following cues as indicators of the nonverbal leadership: C1) Stability, C2) Precedence, C3) Suddenness, C4) Movement Richness, C5) Physical Guidance, C6) Gaze Direction, C7) Waiting, C8) Revoking. This cue set is the summary of cues that are found in the literature. Below we describe each of them in detail.

C1 - Stability. The leader's movements are more stable, confident and smooth, whereas the follower's movements are more hesitant and shaky. The leader, by deciding the next movement, has usually enough time to perform motor planning, and consequently she has more control of her movements, while the follower is more uncertain. The follower might have less balance and her movements can be more jerky.

In work dedicated to the analysis of the coordination patterns in collaborative dance improvisations based on tango, Kimmel and Preuschl mentioned stability as one of the cues of the leader [7]. When comparing the Center of Gravity velocities they noticed that followers' velocity curve over-oscillate around the leaders' velocity curve [7]. Noy and colleagues [16] reported similar findings, with leaders displaying confident motions and followers having more jittering motions. Sacheli and colleagues [18] found that leaders try to improve their signaling strategy by making more stable gestures.

C2 - Precedence. The leader's movements are ahead of the follower's movements. The leader starts moving, or changes direction or speed before the follower. Follower repeats the movement of the leader (or performs a similar one) with a certain delay.

This antecedent-subsequent relationship in joint physical activity is the most often considered indicator of leadership in the literature [11, 21, 22]. For instance, Varni and colleagues [22] computed the leadership index by analyzing chronemic aspect of the actions taken by the interaction partners, i.e., by measuring how often the leaders take precedence with respect to the others.

C3 - Suddenness. The leader's movements are more accentuated, and powerful. They are also more quick and sharp, while the follower's movements are more blurry (e.g., the start and end of the movement are less sharp and visible). Visible emphasis or additional acceleration can be observed in the movement of the leader. In the literature [15] the sudden movements are distinguished from the impulsive ones. The latter term means that the movement is realized without the preparation phase (e.g., avoidance movement), which is not true in the case of leader's movements.

Sacheli and colleagues [18] showed that leaders emphasized the kinematic features of their movements to improve her predictability.

C4 - Movement Richness. The leader moves more: she displays more rich, nuanced and varied movements. Her movements are more refined. Her different body parts may display more frequent and subtle direction changes, while the follower might not catch or reproduce every fine details.

This cue was considered e.g., by Sanchez-Cortes and colleagues [19] in their work dedicated to automatic detection of leadership in small groups. Two features considered in their model are: the amount of a person's body and head activity. Also according to the survey by Hall and colleagues [4] the leaders perform more gestures than the others.

C5 - Physical Guidance. The leader exert forces (pushes / pulls / presses / drags) onto the follower. During the physical contact with another, the leader can directly guide the follower's movements. Her limbs show more tension while the follower's limbs are more passive. When this cue is observed, the persons keep physical contact, and thus the trajectories of their movements often seem to be synchronized, but the shape of the limbs show that one person is using some force to influence the movement of the other person.

This cue is present in many couple dances where the leader physically guide the follower's movements by applying pressure on the follower's back or arms. In non-dance context, people with high-status have more prerogative to touch people with lower status than reciprocally [6]. This suggests that leaders are more likely to initiate touch and carry their intention toward the followers.

C6 - Gaze Direction. The persons' gaze direction is indicative of the leadership. The follower looks at, and pays attention to her partner but the leader does not. The follower needs to pay attention to know what movement to perform next. At the same time the leader movements are self-determined.

In his thesis, Colbert [3] identified eye gaze as the primary cue performed by followers toward an emerging leader. In small group meeting interactions: 1) time that a person is being watched by the other persons and 2) total time that a person is being watched non

reciprocally by any other persons, are the features used for automatic leader detection [2]. Sychns and Mohr [20] also suggested that the eye contact as an important marker of nonverbal leadership in many situations.

C7 - Waiting. The leader waits for the follower before continuing the movement. The leader stops or slows down her movement to allow the follower to catch up. Later the leader continues the movement. The leading impression is given since the leader takes into account what the other person does.

Matzdorf [10] postulates that "taking into account the Follower's feedback" is one of the responsibilities of the dancing leader. This cue was observed, for example, when analyzing the velocities of ankle movements in tango improvisations [7].

C8 - Revoking. When a leader ignores the follower's movement initiative, the follower may revoke or abruptly stop her (already started) action to catch up the leader. This means that the follower may significantly slow down, or even stop her movements, waiting for some signal from the leader about the movement to perform. The follower may notice that her movements are not followed by the other person, and, consequently, she may abort her "on-going" movement.

This cue was discussed by Kimmel and Preuschl [7] in their work on tango improvisations. According to them, the follower often cannot predict if the leader will continue with another step and so she prefers immobility in the case of "doubt" [7].

It is important to notice that some of the cues C1 - C8 may appear together but some of the combinations might be impossible. It means that a leader may display only one or more cues at the same time. For example F7 and F8 are mutually exclusive. Indeed, at a time the leader may slow down her movement to wait the follower, or ignore the follower's movements and impose her (=leader's) own.

4 DATA COLLECTION

To develop a model we organized a data collection, asking professional dancers to perform the contact dance, and more specifically, to mirror continuously each other's movements in duo. In this activity, both participants tend to start physical actions (or gestures) alternatively and imitate each other. The interaction between the dancers can be measured in the time (when) and spatial (how) domains. Spatially, body movement and position can be more or less similar (or symmetric). Temporally, the delay between movement initiation and movement response can vary. In addition, we manipulated the two sensory channels that allow the dancers to communicate with each other. It allows us to collect evidences of different strategies humans adopt to communicate their movements and future movement intentions to one another.

4.1 Task

Dancer pairs were asked to perform an improvised mirroring task in silence (no music, no verbal communication). The dancers might be able to hear each other (e.g., breath). They were instructed to mirror each without pre-designated leader or follower (they were not told who was the "real body" and who was the "reflected image"), so leadership may switch spontaneously. The dancers were asked to try to enhance as much as possible the communication

A   B

C   D
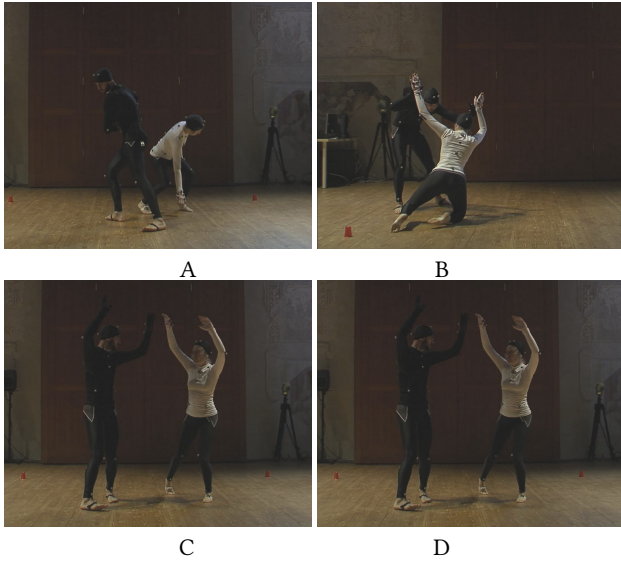
Figure 2: One couple of dancers in four different experimental conditions: A) no visual and no touch contact (NVNT), B) only touch contact and no visual contact (NVT), C) only visual contact, and no touch contact (VNT), D) both visual and touch contact (VT).

of the intention and expressivity. We put constraints on the way dancers could communicate to each other during their performance by manipulating their shared sensory channels in a 2-by-2 design: Visual contact (present/absence) x Touch contact (present/absent). This led to four different conditions (see Figure 2, pictures A, B, C and D, respectively). In condition:

- NVNT) No visual contact, and no touch contact is allowed,
- NVT) Only touch contact but not visual contact is allowed,
- VNT) Only visual contact but no touch contact is allowed,
- VT) Visual contact, and touch contact are allowed.

4.2 Data Recordings

Five professional dancers (2 males, 3 females) with common background in contemporary dance were recorded. Dancers performed the task in duo once with each one of the other dancers, leading to 10 pairs. The pairs performed each of the four conditions once, leading to 40 trials. The trials were performed in a random order. Each recording is between 60 and 90 seconds.

As can be seen on Figure 2 we recorded full-body motion capture data of the dancers to be used for automatic recognition of leadership. The data were captured with the Qualisys motion capture system at 100 Hz and synchronized with a video recording (1280x720, 50fps). The motion capture data contains 3D positions of twenty-four markers for each dancer (foot x2, ankle x2, knee x2, hip x2, hand x2, wrist x2, elbow x2, shoulder x2, sacrum, L1, C7, sternum (xiphoid), sternum (manubrium) and 3 on the head). The motion capture data were manually cleaned and polynomial and linear interpolations were used to fill the missing data.

5 ANNOTATION STUDY

Human annotation was performed on the data described in Section 4. The aims of this study were to check if 1) the majority of the annotators is able to perceive a single leader (i.e., whether they agree on whose is leading); 2) there are differences in the leadership annotation between the four conditions. For instance, in the case NVNT (no visual, no touch contact) the annotators should not be able to observe the leadership as frequently as in the other conditions, because the dancers probably move independent of each other.

Regarding the first item, we want ensure us that it is possible to perceive the leader only from the nonverbal behaviors before developing a computational model. Regarding the second item, we want to see if introducing the sensory deprivation has influenced the perception of the nonverbal leadership.

5.1 Participants

27 undergraduate students (7 males, mean age 18.96 years old) with varying dance experience rated the videos after giving their informed consent (McMaster University Research Ethics Board). They were recruited through an online participant registration system. One course credit or $5 was provided as the compensation.

5.2 Stimuli and Task

Three videos from each of the four conditions (i.e., 12 videos) were selected, all displaying male-female dance pairs. One additional video (vision/no-touch) was used for training. We selected a subset of 12 videos only so that the entire subset could be rated in a reasonable amount of time. Duration of the chosen videos is between 60 and 90 seconds. Participants watched the videos on a computer screen, sitting at 1m of the screen in an isolated room. The twelve videos were segmented into shorter clips so that the participants could remember the entire clips between the watching phase and the rating phase. Fifteen participants watched segmentations of five 18-second clips while the remaining 12 participants watched segmentations of six 15-second clips. All clips were presented twice in a row: at the first presentation (watching phase), participants had to watch and figure out who was leading at any given time, and at the second presentation (rating phase) they performed continuous rating. For a single video, all clips were presented in chronological order. Each participant watched all 12 videos in a randomized order. Specifically, participants were asked to perform a continuous rating of the leadership by holding down arrow keys on a keyboard. The annotation consists of the following four options:

- holding down the right arrow key when the female was the leader (FL),
- holding down the left arrow key when the male was the leader (ML),
- holding down the up arrow key when there was no leader (i.e., the dancers were moving independently) (NO),
- holding down the down arrow key when both were leaders (e.g., the movements by some parts of the body would suggest that one dancer is the leader, when the other parts would indicate the opposite) (BOTH).
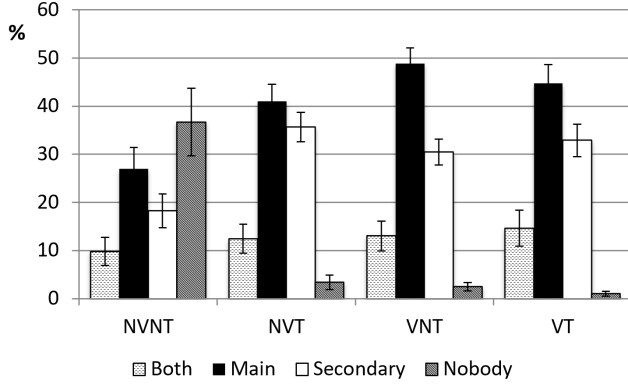
Figure 3: Proportions of answers in the annotation study. "Both" represents the case when both dancers are leading at the same time. "Main" and "Secondary" represent either the answers ML or FL, where Main represents the one that leads the most amount of time in a video, and secondary represents the other one. "Nobody" is when none of the dancers is leading. Error bars represent standard error of the mean.

We intentionally had a separate key to annotate BOTH case because pressing the left and right arrows together can occur by error without actually indicating that both dancers were leading. Before the experiment, participants completed a set of training trials. After completing the rating task, participants filled out a brief questionnaire containing demographics information, dance experience, and perceived difficulty of the rating task.

5.3 Results

For each video and each participant, we computed the percentage of frames in which each one of the four arrow keys was pressed. Figure 3 shows the mean across participants and videos for each condition. In each video, Main and Secondary represent either the answers ML or FL, where Main represents the one that leads the most amount of time in a video, and Secondary represents the other one (obviously the Main will always have higher score than the Secondary).

We performed statistical analyses on the continuous rating data to assess if the leadership rating differed depending on the presence/absence of the two sensory channels (touch and vision) shared between the dancers displayed in the videos. We combined the rating in three measures that were tested in three 2-ways ANOVAs: i) the proportion of "no leader" over all other ratings, ii) the proportion of "one leader" versus "both leading equally at the same time", and iii) the relative amount of time each dancer is leading. The 2 Touch (present/absent) x 2 Vision (present/ absent) repeated measures ANOVA on the proportion of "no leader" ratings showed a significant main effect of vision, $F(1, 26) = 26.4$, $p < 0.001$, a significant main effect of touch, $F(1, 26) = 36.4$, $p < 0.001$, and a significant interaction between vision and touch $F(1, 26) = 29.5$, $p < 0.001$. Indeed, the answer "no leader" was selected significantly more during the no-vision / no-touch condition (NVNT) than any other condition ($p < 0.001$ for all). Likewise, the proportion of "one leader" versus "both leading equally at the same time" shows
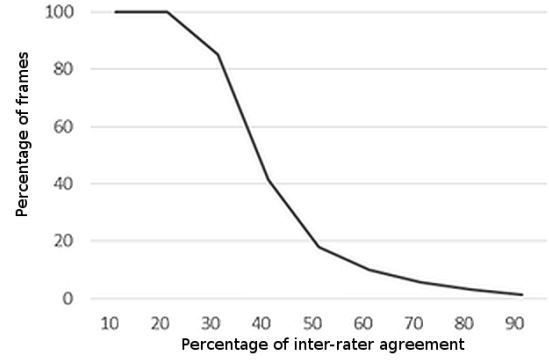
Figure 4: Amount of frames that reaches at least x% of inter-rater agreement

a significant interaction between vision and touch: $F(1, 26) = 4.6$, $p < 0.05$. The answer "Both" tended to be selected more than the answer "Male" or "Female" in the no-vision/touch condition (NVT) than other conditions but post-hoc failed significance. Finally, we tested if the two dancers leads the same amount of time versus if one leads more. We found a significant main effect of touch, $F(1, 26) = 5.1$, $p < 0.05$, and a significant interaction between vision and touch: $F(1, 26) = 9.4$, $p < 0.01$. The two dancers leaded more equally (taking turn) in the no-vision/touch condition (NVT) than the no-vision/no-touch condition (NVNT) ($p < 0.005$).

We also computed for each frame, the extent to which the raters agree on the selected answers. P_i is defined as:

$$P_i = \frac{1}{n(n-1)} \sum_{j=1}^{k} n_{ij}(n_{ij} - 1) \tag{1}$$

where n is the number of raters (here $n = 27$), the possible answers are indexed by j, $1 \leq j \leq k$, (here $k = 4$), and n_{ij} is the number of raters who assigned the j-th answer to the i-th frame. Figure 4 shows the proportion of frames in the corpus that reach a given degree of agreement between annotators.

Inter-rater agreement between the 27 annotators was measured by computing the number of participants that selected one of the four answers. The agreement was measured with the Fleiss κ for 27 annotations for each data frame. The average values for each condition are given in Table 1.

Table 1: Average Fleiss Kappa for each condition

Condition	NVNT	NVT	VNT	VT
Fleiss κ	0.11	0.15	0.26	0.15

5.4 Discussion

The continuous annotation of the corpus allowed us to fill the aims mentioned at the beginning of Section 5. According to the first part of the analyses, we found that for the majority of the data there is one leader (i.e., the labels NO and BOTH were chosen less often). However, when tactile and visual communication between

dancers is prevented (condition NVNT) no leader was observed for the most of the time.

In the second part of the analyses, we measured the inter-raters agreement. The results were not optimal. It should be noted, however, that we computed the agreement taking in account the 4 separate categories, but actually, the options overlay partially. Indeed, one rater could perceive one dancer as leader at a given time, whereas another rater would perceive that both dancers lead equally. Even if they chose different options, their answers overlap partially.

Consequently, in future we plan to conduct additional annotation study of the most significant segments (see Section 7 for details).

6 ALGORITHMS

In this section, we present the algorithms to compute 6 features F1 - F6, corresponding to the first 6 cues in Section 3. The last two cues (C7 and C8) where not taken into account for the moment, as they require the computation of the movement intention of both the leader and the follower on longer time buffers, which is a non-trivial task.

6.1 Absolute speed and curvature

Given a single joint's 3D trajectory, defined in a parametric way:

$$j(t) = (x(t), y(t), z(t)) \tag{2}$$

we compute the absolute tangential speed s:

$$s(t) = \sqrt{(x'^2 + y'^2 + z'^2)} \tag{3}$$

Similarly, 3D curvature k is computed as follows:

$$k = \frac{\sqrt{(z''y' - y''z')^2 + (x''z' - z''x')^2 + (y''x' - x''y')^2}}{(x'^2 + y'^2 + z'^2)^{(3/2)}} \tag{4}$$

Absolute tangential speed and curvature of a group of joints J are defined as the average absolute tangential speed and curvature of the single joints included in J.

6.2 Operators G and H

We now define two auxiliary operators $G_k(m, n)$ and $H(p, q)$. The first one corresponds to the amount of frames for a time interval in which the discrete variable m is greater than the discrete variable n, increased by a constant k, i.e.:

$$G_k(m, n) = \text{number of frames} \mid (m - n) > k \tag{5}$$

In other words, the above operator quantifies how long the value of one variable (m) is "substantially" higher than the value of another variable (n). The second operator $H(p, q)$ returns a ordered pair of values R_1 and R_2 such as:

$$\text{If} \quad ((p - q) > 0) \quad \text{Then} \quad R_1 = 1, R_2 = 0 \tag{6}$$

$$\text{Else} \quad R_2 = 1, R_1 = 0 \tag{7}$$

6.3 F1: Stability

Stability is extracted by computing the following three low-level features on the persons' group of joint (e.g., wrists, ankles or whole body): Variance (Var), Smoothness (Sm), and Variability (V).

The first one is the well-known statistical function. In our case, we apply it on the absolute tangential speed of a group of joints J.

To compute the second feature, i.e., Smoothness (Sm), we use the algorithm described in [12]. It is a simplified version of the Pearson correlation of trajectory curvature and absolute tangential speed:

$$Sm(s) = \frac{1}{ks} \tag{8}$$

where s and k are the absolute tangential speed and curvature. Variability (V) is computed as the difference between the actual joint velocity and the low-pass filtered velocity of the same joint (see [14] for more details).

Given two persons A and B and a group of joints J, we compute the above mentioned 3 low-level features, which results in 6 values. Next, we apply the G_k operator to the resulting values, that is:

$$GVar_J^A = G_{0.5}(Var(s_J^A), Var(s_J^B)) \tag{9}$$

$$GVar_J^B = G_{0.5}(Var(s_J^B), Var(s_J^A)) \tag{10}$$

The same computation is applied to Sm and V. Next we apply H operator:

$$\{A_J^{Var}, B_J^{Var}\} = H(GVar_J^B, GVar_J^A) \tag{11}$$

$$\{A_J^{Sm}, B_J^{Sm}\} = H(GSm_J^B, GSm_J^A) \tag{12}$$

$$\{A_J^V, B_J^V\} = H(GV_J^B, GV_J^A) \tag{13}$$

Finally, the *Stability* $F1_J^Z$ of the group of joints J of the person $Z, Z = \{A, B\}$ is:

$$F1_J^Z = max(Z_J^{Var}, Z_J^{Sm}, Z_J^V) \tag{14}$$

6.4 F2: Precedence

We base the computation of Precedence (F2) on the Event Synchronization (ES) algorithm [17]. Given two simultaneously measured discrete univariate time series x_n and y_n, $n = 1, ..., N$, the ES algorithm computes two quantities: Q, synchronization degree and q, being a measure of "delay asymmetry". The latter is equal to 1 if all events in x_n precede those in y_n, it is equal to -1 if all events in y_n precede those in x_n, it is equal to zero if all events in x_n occur at the same time of those in y_n.

Given two persons A and B and a group of joint J, we apply ES on the time series containing positions of local maxima of the speed s_J of A and B, obtaining the value of $q_J^{A,B}$.

Finally, the value of *Precedence* (F2) on a group of joints J is:

$$\text{If} \quad (q_J^{A,B} > 0) \quad \text{Then} \quad F2_J^A = 1, F2_J^B = 0 \tag{15}$$

$$\text{Else} \quad F2_J^A = 0, F2_J^B = 1 \tag{16}$$

6.5 F3: Suddenness

To compute the Suddenness we follow the approach proposed by Niewiadomski and colleagues in [15] which is based on alpha-stable distributions [9]. An alpha-stable distribution can be modeled by a probability density function characterized by four parameters $(\alpha, \beta, \gamma, \delta)$.

Given a discrete time series containing the tangential velocity of a group of joints J, we fit an alpha-stable distribution to the probability density function of the speed s_J of group J and we obtain the values: $\alpha_J, \beta_J, \gamma_J, \delta_J$. Next, for each J we scale s_J and we multiply it by δ_J:

$$Sud_J = \delta_J \frac{1 - \alpha_J}{2} \tag{17}$$

For two persons A and B and a group of joints J, we compute:

$$GSu_J^A = G_{0.5}(Su_J^A, Su_J^B) \tag{18}$$

$$GSu_J^B = G_{0.5}(Su_J^B, Su_J^A) \tag{19}$$

In the final step, *Suddenness* (F3) is computed as follows:

$$\{F3_J^A, F3_J^B\} = H(GSu_J^A, GSu_J^B) \tag{20}$$

6.6 F4: Movement Richness

We model this cue by computing the kinetic energy of the movement. Given a discrete time series containing the tangential absolute speed s of a group of joints J, we define KE:

$$KE_J = \frac{ms_J^2}{2} \tag{21}$$

where m is an approximation of the mass of J. Given persons A and B and a group of joints J, we compute:

$$GKE_J^A = G_{0.5}(KE_J^A, KE_J^B) \tag{22}$$

$$GKE_J^B = G_{0.5}(KE_J^B, KE_J^A) \tag{23}$$

Finally, *Movement Richness* (F4) is computed as follows:

$$\{F4_J^A, F4_J^B\} = H(GKE_J^A, GKE_J^B) \tag{24}$$

6.7 F5: Physical Guidance

This cue models the fact that a person is pushing the other one with her hands, or pulling the hands of the other person toward her. Given persons A and B, we compute the speed s_{xW}^Z, $x = \{L, R\}$, $W = \{A, B\}$, $Z = \{A, B\}$, $W \neq Z$ of Z's x (left/right) hand with respect to the W's body, e.g.: s_{RA}^B is a speed of A's right hand toward B.

To identify if persons' hands are in physical contact, we compute four distances d_{xW}^{xZ}, $x = \{L, R\}$, $W = \{A, B\}$, $Z = \{A, B\}$, $W \neq Z$ between all possible pairs of hands of A and B, e.g., d_{RB}^{RA} is distance between A's right and B's right hand. We find the speeds of pairs of hands which are closer each other than 15 cm:

If $(d_{RB}^{RA} < 15cm)$ $PC_1 = s_{RA}^B + s_{RB}^A$ Else $PC_1 = 0$ (25)

If $(d_{LB}^{LA} < 15cm)$ $PC_2 = s_{LA}^B + s_{LB}^A$ Else $PC_2 = 0$ (26)

If $(d_{RB}^{LA} < 15cm)$ $PC_3 = s_{LA}^B + s_{RB}^A$ Else $PC_3 = 0$ (27)

If $(d_{LB}^{RA} < 15cm)$ $PC_4 = s_{RA}^B + s_{LB}^A$ Else $PC_4 = 0$ (28)

and then we compute:

$$PC = \sum_{i=1}^{4} PC_i \tag{29}$$

$$PC_{rev} = -1PC \tag{30}$$

$$GPC = G_{0.5}(PC, PC_{rev}) \tag{31}$$

$$GPC_{rev} = G_{0.5}(PC_{rev}, PC) \tag{32}$$

Finally, *Physical Guidance* (F5) is obtained, as:

$$\{F5^A, F5^B\} = H(GPC, GPC_{rev}) \tag{33}$$

6.8 F6: Gaze Direction

We approximate the person's gaze with the direction of her head. Thus, we compute the line passing from the back and the front head middle point: line $gaze_A$ is representing A's gaze direction, and $gaze_B$ is representing B's gaze direction. We also compute the line dir_{AB} between A's and B's sternums.

We define the $gaze_{AtoB}$ (*gaze of A toward B*) as an angle between the lines $gaze_A$ and dir_{AB}, and we define $gaze_{BtoA}$ (*gaze of B toward A*) as the angle the lines $gaze_B$ and dir_{AB}. Next, we compute the number of frames fr_{AB}, such that A is gazing B, but B is not gazing A, using the following formula:

$$fr_{AB} = \text{ number of frames} \mid (gaze_{AtoB} < 45) \tag{34}$$

$$\text{and} \quad gaze_{BtoA} > gaze_{AtoB} + 10 \tag{35}$$

Analogical computation of this feature is done to compute the number of frames B is gazing A, i.e., fr_{BA}. Finally, *Gaze Direction* (F6) is computed:

$$\{F6^A, F6^B\} = H(fr_{AB}, fr_{BA}) \tag{36}$$

7 CONCLUSION AND FUTURE WORKS

In this paper, we proposed a set of cues of nonverbal leadership in unstructured joint physical activity. It consists of eight cues that can be observed in the nonverbal behaviors of leaders and followers. We also presented a dataset of dance improvisations which contains video and full-body motion capture data of five dancers. The data was rated by 27 persons with continuous annotation of leadership. The results of the annotation and motion capture data will be made publicly available in the near future for research purposes.

The most significant contributions of this paper are:

- a cue set of nonverbal leadership in unstructured joint physical activity,
- a set of algorithms to compute these cues;

- a dataset of dance improvisations including sensory deprivation conditions;
- statistical analysis showing that sensory deprivation conditions influence the nonverbal communication in the couple, and consequently perceived nonverbal leadership.

Even if this work was realized in the context of contemporary dance, the proposed algorithms can be applied in various different contexts such as: i) the analysis of group sport activities, performing musicians, and other social activities, or ii) building models of the joint activities for interactive robots and virtual agents, e.g., in learning scenarios.

As explained in Section 1 and in Figure 1, the final aim is the creation of a computational model of nonverbal leadership. The next step is to evaluate the algorithms proposed in Section 6. The validation will consist of: i) additional annotation of the segments; ii) computation of the cues on motion capture data corresponding to the same segments. We will conduct the second annotation study in which some segments of recordings described in Section 5 will be shown to raters, who will be asked to list, which cues, between C1-C8, they would have observed in the videos. The same segments will be processed using the algorithms proposed in Section 6. In the final step, the results of the human annotation and the automated computation will be compared to measure the agreement.

Future work will also include extension of the cue set. As mention in the paper by Moreland [13], which showed that dyadic interactions proceeds different than group interactions, the proposed algorithms will be extended also to deal with the data of more than two persons at a time.

ACKNOWLEDGMENTS

REFERENCES

[1] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2017. Multi-task Learning of Social Psychology Assessments and Nonverbal Features for Automatic Leadership Identification. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*. ACM, New York, NY, USA, 451–455. https://doi.org/10.1145/3136755.3136812

[2] Cigdem Beyan, Nicolò Carissimi, Francesca Capozzi, Sebastiano Vascon, Matteo Bustreo, Antonio Pierro, Cristina Becchio, and Vittorio Murino. 2016. Detecting Emergent Leader in a Meeting Environment Using Nonverbal Visual Features Only. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*. ACM, New York, NY, USA, 317–324. https://doi.org/10.1145/2993148.2993175

[3] David Colbert. 2007. Nonverbal Cues of the Leadership Selection Process: Leadership Selection in a Small Group. (2007).

[4] Judith A. Hall, Erik J. Coats, and Lavonia Smith LeBeau. 2005. A Nonverbal Behavior and the Vertical Dimension of Social Relations: A Meta-Analysis. *Psychological Bulletin* 131, 6 (2005), 898–924.

[5] Yuval Hart, Lior Noy, Rinat Feniger-Schaal, Avraham E. Mayo, and Uri Alon. 2014. Individuality and Togetherness in Joint Improvised Motion. *PLOS ONE* 9, 2 (02 2014), 1–8. https://doi.org/10.1371/journal.pone.0087213

[6] Nancy M. Henley. 1973. Status and sex: Some touching observations. *Bulletin of the Psychonomic Society* 2 (1973), 91–93.

[7] Michael Kimmel and Emanuel Preuschl. 2016. *Dynamic Coordination Patterns in Tango Argentino: A Cross-Fertilization of Subjective Explication Methods and Motion Capture*. Springer International Publishing, Cham, 209–235. https://doi.org/10.1007/978-3-319-25739-6_10

[8] Ahmet Alp Kindiroglu, Lale Akarun, and Oya Aran. 2017. Multi-domain and multi-task prediction of extraversion and leadership from meeting videos. *EURASIP Journal on Image and Video Processing* 2017, 1 (21 Nov 2017), 77. https://doi.org/10.1186/s13640-017-0224-z

[9] Paul Lévy. 1925. *Calcul des probabilités*. Vol. 9. Gauthier-Villars Paris.

[10] Fides Matzdorf. 2005. You can tell your follower where to go, but you canŠt put them there: Leadership as partnership. In *1st Annual Conference on Leadership Research Re-thinking Leadership: New Directions in the Learning and Skills Sector?*

[11] Fides Matzdorf and Ramen Sen. 2016. Demanding Followers, Empowered Leaders: Dance As An "Embodied Metaphor" For Leader-Follower-Ship. *Organizational Aesthetics* 5, 1 (2016), 114–130.

[12] Barbara Mazzarino and Maurizio Mancini. 2009. The Need for Impulsivity & Smoothness-Improving HCI by Qualitatively Measuring New High-Level Human Motion Features.. In *SIGMAP*. 62–67.

[13] Richard L. Moreland. 2010. Are Dyads Really Groups? *Small Group Research* 41, 2 (2010), 251–267. https://doi.org/10.1177/1046496409358618

[14] Radoslaw Niewiadomski, Ksenia Kolykhalova, Stefano Piana, Paolo Alborno, Gualtiero Volpe, and Antonio Camurri. 2018. Analysis of Movement Quality in Full-Body Physical Activities. *accepted to: ACM Transaction On Interactive Intelligent Systems* (2018). https://doi.org/10.1145/3132369

[15] Radoslaw Niewiadomski, Maurizio Mancini, Gualtiero Volpe, and Antonio Camurri. 2015. Automated Detection of Impulsive Movements in HCI. In *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter (CHItaly 2015)*. ACM, New York, NY, USA, 166–169. https://doi.org/10.1145/2808435.2808466

[16] Lior Noy, Erez Dekel, and Uri Alon. 2011. The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences* 108, 52 (2011), 20947–20952. https://doi.org/10.1073/pnas.1108155108

[17] Rodrigo Quian Quiroga, Thomas Kreuz, and Peter Grassberger. 2002. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* 66 (Oct 2002), 041904. Issue 4. https://doi.org/10.1103/PhysRevE.66.041904

[18] Lucia Maria Sacheli, Emmanuele Tidoni, Enea Francesco Pavone, Salvatore Maria Aglioti, and Matteo Candidi. 2013. Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Experimental Brain Research* 226, 4 (01 May 2013), 473–486. https://doi.org/10.1007/s00221-013-3459-7

[19] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2012. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia* 14, 3 (June 2012), 816–832. https://doi.org/10.1109/TMM.2011.2181941

[20] Birgit Schyns and Gisela Mohr. 2004. Nonverbal Elements of Leadership Behaviour. *German Journal of Human Resource Research* 18, 3 (2004).

[21] Piotr Słowiński, Chao Zhai, Francesco Alderisio, Robin Salesse, Mathieu Gueugnon, Ludovic Marin, Benoit G. Bardy, Mario di Bernardo, and Krasimira Tsaneva-Atanasova. 2016. Dynamic similarity promotes interpersonal coordination in joint action. *Journal of The Royal Society Interface* 13, 116 (2016). https://doi.org/10.1098/rsif.2015.1093

[22] Giovanna Varni, Gualtiero Volpe, and Antonio Camurri. 2010. A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media. *IEEE Transactions on Multimedia* 12, 6 (Oct 2010), 576–590. https://doi.org/10.1109/TMM.2010.2052592

[23] Auriel Washburn, Mariana DeMarco, Simon de Vries, Kris Ariyabuddhiphongs, R. C. Schmidt, Michael J. Richardson, and Michael A. Riley. 2014. Dancers entrain more effectively than non-dancers to another actorŠs movements. *Frontiers in Human Neuroscience* 8 (2014), 800. https://doi.org/10.3389/fnhum.2014.00800

[24] David Zeitner, Nicolas Rowe, and Brad Jackson. 2016. Embodied and Embodiary Leadership: Experiential Learning in Dance and Leadership Education. *Organizational Aesthetics* 5, 1 (2016), 167–187.

[25] Chao Zhai, Francesco Alderisio, Piotr Słowiński, Krasimira Tsaneva-Atanasova, and Mario di Bernardo. 2016. Design of a Virtual Player for Joint Improvisation with Humans in the Mirror Game. *PLOS ONE* 11, 4 (04 2016), 1–17. https://doi.org/10.1371/journal.pone.0154361