Walk the Talk: Gestures in Mobile Interaction

Zeynep Yücel[1(✉)], Francesco Zanlungo[2], and Masahiro Shiomi[2]

[1] Department of Computer Science, Okayama University, Okayama, Japan
zeynep@okayama-u.ac.jp
[2] Intelligent Robotics and Communication Laboratories, ATR, Kyoto, Japan
{zanlungo,m-shiomi}@atr.jp

Abstract. This study aims at describing navigation guidelines and concerning analytic motion models for a mobile interaction robot, which moves together with a human partner. We address particularly the impact of gestures on the coupled motion of this human-robot pair.

We pose that the robot needs to adjust its navigation in accordance to its gestures in a natural manner (mimicking human-human locomotion). In order to justify this suggestion, we first examine the motion patterns of real-world pedestrian dyads in accordance to 4 affective components of interaction (i.e. gestures). Three benchmark variables are derived from pedestrian trajectories and their behavior is investigated with respect to three conditions: (i) presence/absence of isolated gestures, (ii) varying number of simultaneously performed (i.e. concurring) gestures, (iii) varying size of the environment.

It is observed empirically and proven quantitatively that there is a significant difference in the benchmark variables between presence and absence of the gestures, whereas no prominent variation exists in regard to the type of gesture or the number of concurring gestures. Moreover, size of the environment is shown to be a crucial factor in sustainability of the group structure.

Subsequently, we propose analytic models to represent these behavioral variations and prove that our models attain significant accuracy in reflecting the distinctions. Finally, we propose an implementation scheme for integrating the analytic models to practical applications. Our results bear the potential of serving as navigation guidelines for the robot so as to provide a more natural interaction experience for the human counterpart of a robot-pedestrian group on-the-move.

Keywords: Affective communication · Gestures

1 Introduction

It is unanimously accepted that gestures constitute a fundamental component of non-verbal human-human interaction [1,2]. Drawing the listener's attention [3] and providing a natural and intuitive means for delivering various feelings, expectations, intentions and even communicating personality traits [4] can be listed among the prominent contributions of gestures to human-human communication [3].

A. Kheddar et al. (Eds.): ICSR 2017, LNAI 10652, pp. 220–230, 2017.
https://doi.org/10.1007/978-3-319-70022-9_22

Depending on the research domain, the term "gesture" may refer to different actions or behaviors. Taking a purely mechanical standpoint, gestures can be defined as the integration of several body parts (or their coupling) in the embodiment of interaction. According to this definition, hands are shown to play a bigger than other body parts, but overall they contribute to roughly half of the entire human gesturing, where the rest involves head, fingers, foot, or objects [5]. In contrast to this purely mechanical approach, in social signal processing, gestures are treated taking in consideration their implications. For instance, a hand waving gesture can refer to acceptance or rejection depending on the affective/attitudinal/cognitive state of the performer. In this study, we take a similar approach to the latter one by not containing ourselves to (mechanical) "movements", but considering gestures as a set of bodily "actions", that serve useful in communication.

The impact of gestures on human-robot interaction are subject to a detailed treatment as well [6]. Most studies take a robot stand-point and address face-to-face communication between a robot and a human in stationary settings such as around a table [7]. In such scenarios, they aim designing recognition methods for identifying and interpreting human gestures. Some studies also propose replication methods for the recognized gestures [8,9], which often target hand movements such as pointing, waving, covering etc.

In this respect, Salem et al. differ from the mainstream studies by considering a more dynamical interaction, where the human-robot pair moves around in a domestic environment or classroom [10,11]. They show that a robot, which employs gestures along with speech, is perceived by humans as more friendly, engaged, and competent.

Such effects of gestures are assumed to be pertinent in mobile human-robot interaction as well [12]. Here, we consider a mobility pattern as in service or assistive robotics applications, which may take place in dynamic public spaces, such as shopping mall, museum, or station. In this kind of continuously evolving environment, some aspects of communication such as staying in the field of view of the partner while avoiding obstacles or sustaining joint attention during motion, turn it into a challenge for a mobile robot to couple its gestures with its locomotion [13,14].

In that respect, this paper focuses particularly on pedestrian interaction (i.e. human-human interaction on-the-move) and studies the motion patterns of actual (i.e. uninstructed) dyads, which perform (or not) a number of gestures. Three variables are derived from trajectories and their distributions are investigated against three conditions: (i) the presence/absence of the isolated gestures and their types, (ii) number of concurring gestures and (iii) size of the environment. The empirical distributions are represented using analytic models, which are shown to reflect the distinctions between various gesturing patterns. Finally, we suggest guidelines for navigation of an intelligent platform (a social robot) based on the proposed models for reproducing human-like locomotion. Our guidelines and models have the potential of improving the navigation of a social robot to attain the capability of adjusting its locomotion in accordance to its gestures.

2 Data

We gather a dataset using a network of range scanners and video cameras in an underground pedestrian network in the Umeda area, Osaka, Japan (see Figs. 1 and 2(a)). The observation space (henceforth referred as the *environment*) is composed of a large main street of 7×60 [m] and an intersecting side-street of 3×20 [m].

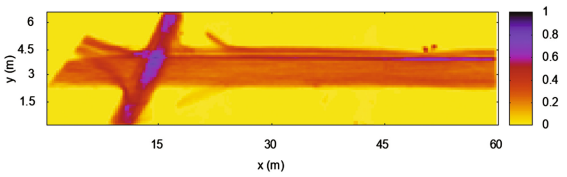


Fig. 1. Normalized density map of the environment.

In 6 h, we observe over 12.000 pedestrians and, among those, we choose to focus on the 634 dyads (1268 pedestrians). The reason for this choice is two fold: (i) there are more samples of 2-people (2p) groups than larger groups, so their motion can be characterized with better accuracy, (ii) they are shown to constitute the basic building block of all (larger) social groups [15].

2.1 Gestures of Interest

As explained in Sect. 1, we do not consider gestures to be mere mechanical movements but rather as actions of body parts, which play a role in communication. Therefore, taking a similar approach to Vinciarelli et al. [16], we focus on the following four gestures, which are listed among the behavioral cues associated to some of the most important social behaviors:

i. G1: Speech
ii. G2: Gaze (at partner)
iii. G3: Joint attention (mutual gaze at common target point)
iv. G4: Physical contact

In order to asses the influence of each individual gesture and several coupled gestures, we define the concept of *isolated* and *concurring* gestures. A gesture is said to be performed concurringly, if there is at least one other gesture along. Otherwise it is called an isolated gesture.

As for ground truth of gesture performance, we use the annotations provided by two human coders. They view the videos of the environment recorded at several locations, which are selected to get a good frontal view of the pedestrians moving at different directions (see Fig. 2(a) for a sample video frame). Based on these recordings, they label each dyad in the environment with respect to absence or presence

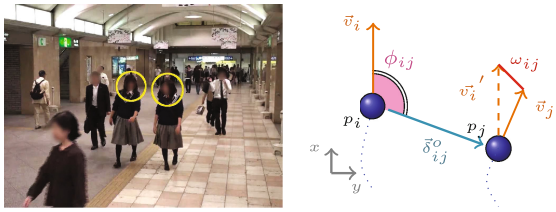
of the four gestures (G1 $\sim$ G4). In addition, the complementary case of no-gestures is automatically attributed to all remaining pairs and denoted by G0[1].

The inter-rater reliability of the coding process is assessed through Cohen's κ statistics (accounting for prevalence) [17,18]. Bearing in mind the considerable subjectivity of the task, the coders are found to have a satisfactory rate of agreement with $\kappa = 0.62$.

2.2 Definition of Benchmark Variables

The pedestrian trajectories are computed from the range data using the particle-filter based tracking method of [19]. Subsequently, three benchmark variables are derived from the trajectories. Consider that p_i and p_j constitute a dyad (see Fig. 2). On this sample dyad, the benchmark variables are described as follows:

i. Interpersonal distance, δ_{ij}, is the distance from p_i to p_j.
ii. Alignment, ϕ_{ij}, is the angle between the velocity of p_i (i.e. v_i), and δ_{ij}.
iii. Velocity difference, ω_{ij}, is the magnitude of vector $v_i - v_j$.



(a) (b)

Fig. 2. (a) A sample interacting dyad and (b) definition of the benchmark variables.

These variables are chosen as benchmark since they enable the representation of stationary (positional) and non-stationary (dynamic) relations of the dyad. Namely, δ_{ij} and ϕ_{ij} address the positional relation by describing the relative locations in terms of proximity and orientation, respectively, whereas ω_{ij} addresses the dynamic relation by describing the relative motion.

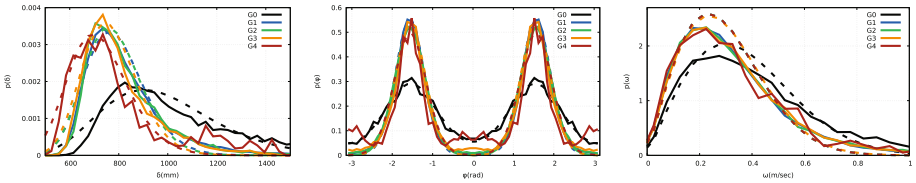
3 Dependence of Benchmark Variables on Gestures

This section considers several conditions on internal and external dynamics of interaction. The internal dynamics include the type of the gestures and the number of concurring gestures, whereas the external dynamics relate the environmental factors (i.e. the size of the space). In what follows, the distributions of the benchmark variables are illustrated for each case. Several inferences are drawn from empirical observations and then verified quantitatively.

[1] If a dyad performs more than one gesture (concurring gestures), it is assigned multiple labels.

3.1 Comparing No-Gesture Case to Isolated Gesture Case

We first consider the isolated performances of the four gestures and contrast them to G0 in order to asses the effect of each individual gesture on motion. Figure 3 illustrates the distributions of δ, ϕ and ω for G0 $\sim$ G4 in solid lines. It is clear that there is a distinction between G0 and G1 $\sim$ G4. Namely, if a gesture accompanies interaction, the peers move in closer proximity (Fig. 3(a)), with their velocities increasingly aligned (Fig. 3(b)) and the difference of velocities shrinking (Fig. 3(c)).



(a) (b) (c)

Fig. 3. Distributions of (a) δ, (b) ϕ and (c) ω for no-gesture case (G0) and isolated gestures (G1 $\sim$ G4) are given in solid lines. The dashed lines represent concerning models.

For confirming these empirical observations in a quantitative manner, we carry out an Anova regarding the five cases of G0 $\sim$ G4 (see first row of Table 1). The p-values turn out to be lower than 10^{-4}, which proves that the difference between the five distributions (with standard level of significance as 0.01)[2].

Next, we compare the curves of G1 $\sim$ G4 (omitting G0) so as to understand whether different gestures have different impact on motion patterns. By examining the curves in Fig. 3, we conclude that the effect of each individual gesture is to a similar degree.

An Anova involving G1 $\sim$ G4 supports this observation for δ and ω (see second row of Table 1). Namely, the p-values turn out to be 0.82 and 0.98 indicating that the difference is not statistically significant. On the other hand, the p-value regarding ϕ is determined to be 0.01, which is significant. This is due to the fact that in some cases for smooth avoidance, the dyad moves in a single file (one person following the other), which does not affect δ or ω but introduces a $\pi/4$ phase shift on ϕ.

3.2 Comparing Varying Number of Concurring Gestures

Next, we focus on concurring gestures and compute the distribution of the variables for dyads performing 1, 2 and 3 gestures concurring[3]. Both from Fig. 4 and

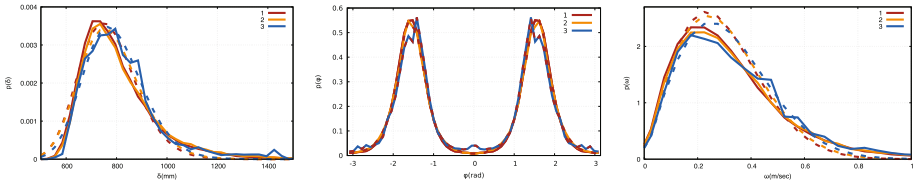
[2] For Anova relating ϕ, we use only the values $\phi \in [0, \pi/2]$ to be able to highlight differences in spread between distributions with the same average value $\approx \pi/2$.

[3] There was no dyad which performed all 4 gestures at once.

Table 1. Anova comparing various sets of curves from Figs. 3 and 4 (S_B and S_L stand for broad and limiting environment, respectively).

Effecting factor	Variables		
	δ	ϕ	ω
Presence of isolated gestures	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
Type of isolated gestures	0.82	0.01	0.98
Number of concurring gestures	0.10	0.08	0.37
Presence of gestures in S_B	0.004	$< 10^{-4}$	$< 10^{-4}$
Presence of gestures in S_L	0.005	0.01	0.17

Table 1 (see the third row), it is concluded that the difference between the distributions is not statistically significant. Combining this result with the previous one described in Sect. 3.1, we can state that the distributions present distinction with respect to the presence/absence of a gesture but not with the number of concurring gestures.



(a) (b) (c)

Fig. 4. Distributions of (a) δ, (b) ϕ and (c) ω with varying number of concurring gestures.

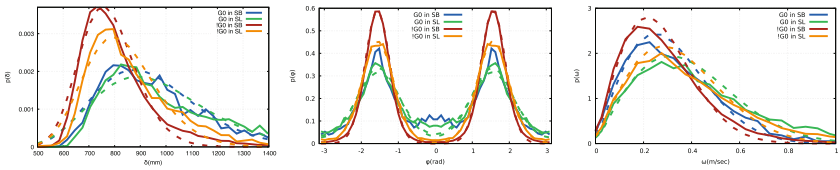
Having confirmed that the significance of presence of gestures irrespective of the number of overlapping performances, in the rest of the analysis, we consider not only isolated but also concurring gestures without paying regard to the number of gestures performed. Namely, we contrast the absence of gestures G0 to presence (of any number of) gestures denoted by !G0.

3.3 Impact of Environmental Factors

We suggest that a *broad space* (S_B) gives the possibility to the dyads to keep the group structure solid against mobile obstacles. Namely, they can avoid other pedestrian groups or individuals, as they keep their interaction continuous and reflecting the effect of gestures on their group structure. Thus, !G0 is expected to reflect the effect of gestures on group motion clearly and to be significantly different than G0 in S_B.

On the other hand, a *limiting space* (S_L) does not let avoidance of the obstacles without changing the group structure. Thus, it forces the dyad to first interrupt their interaction, then re-arrange itself for smooth avoidance and finally recover their former structure to continue interaction. Therefore, !G0 cannot reflect the effect of gestures clearly and the difference between G0 and !G0 is expected to be less prominent in S_L.

In our setting, we consider the large and narrow streets in Fig. 5 to act as S_B and S_L, respectively. Figure 5 illustrates the pattern of the benchmark variables for G0 and !G0 in S_B and S_L. It is clear from the empirical observations that G0 and !G0 present a larger difference in S_B (depicted by blue and red curves respectively), in comparison to the pattern for !G0 (depicted by green and orange curves).



(a) (b) (c)

Fig. 5. Distributions of (a) δ, (b) ϕ and (c) ω for G0 and !G0 in S_B and S_L. (Color figure online)

Figure 5 confirms that dyads which do not perform any gestures (G0), are not affected by the limiting property of the environment severely, since they can change the group structure for adapting to the avoidance conditions. On the other hand, the dyads, which perform gestures (i.e. !G0), can comfortably keep their structure intact in broad environments, whereas they are severely subject to ambient disturbances in limiting environments. We find this result particularly interesting. Finally, we verify these observations quantitatively. It is shown in the last two rows of Table 1 that the difference between G0 and !G0 is more prominent in S_B, and thus leads to smaller p-values as opposed to S_L.

4 Modeling of the Benchmark Variables

This section briefly introduces the three models employed in characterizing δ, ϕ and ω. Subsequently, we present the models distributions in comparison to empirical distributions and illustrate that they provide a very good approximation. In addition, we demonstrate the model parameters and interpret them in relation to our expectations.

For modeling δ, we adopt the approach of [15], which introduces a potential to describe the position dynamics of two pedestrians in an interacting group. According to their analysis, based on statistical physics methods, the distribution for δ is given by,

$$p(\delta|\beta, r_0) = \delta \exp\left(\frac{-\beta(\delta - r_0)^2}{\delta r_0}\right). \quad (1)$$

Here the parameter r_0 is the preferred distance for social interaction, while the larger β is, the more ordered the system is (the distribution is narrowly centered on r_0).

For ϕ, generalizing the results of [15], we assume a *von Mises* distribution, i.e. analogous of a Gaussian for a circular random variable,

$$\Delta(\phi|\nu, \kappa) = \frac{\exp(\kappa \cos(\phi - \nu))}{2\pi I_0(\kappa)}, \quad (2)$$

where ν denotes the mean value, κ is analogous of $1/\sigma^2$ in the normal distribution, and I_0 is the modified Bessel function of the first kind with order zero [20]. Since we do not pay regard to be positioned on the right or left side with respect to motion direction, both pedestrians in a pair are treated in the same manner. In addition, adopting an equally weighted linear combination of two von Mises distributions centered at $\pm\pi$ [21],

$$\Phi(\phi|\nu, \kappa) = 0.5p(\phi|\nu, \kappa) + 0.5p(\phi|\nu + \pi, \kappa). \quad (3)$$

For modeling ω, we use the fact that pedestrians in a social group move towards the same target, even though there may be slight deviations due to avoidance. Therefore, the expected value of ω is always around 0. Thus, we approximate the x and y components of ω with a Gaussian with 0 mean, which makes ω come from a Rayleigh distribution.

$$\Omega(\omega|\sigma) = \frac{\omega}{\sigma^2} \exp \frac{-\omega^2}{2\sigma^2}. \quad (4)$$

We use Δ, Φ and Ω described by Eqs. 1, 3, and 4 to model the distributions of the benchmark variables. For that purpose, we minimize the squared error between the empirical observations and corresponding models with a golden section search. By this way, the parameters which represent the empirical observations with highest accuracy are found as in Table 2, where the corresponding curves are illustrated in dashes on Figs. 3, 4 and 5. These curves prove that the effect of gestures on dyad motion can be captured by the proposed motion models with significant accuracy.

In Table 2, we present the parameters for two cases, where we detected a significant difference against an effecting factor (presence/absence of gestures and environment).

These values in Table 2 are in agreement with the inferences drawn from Table 1. Namely, the parameters governing the models of G0 are quite different than the ones of G1 $\sim$ G4, whereas the parameters concerning G1 $\sim$ G4 are similar within each other[4].

[4] Note that in modeling ϕ, the distinguishing effect is pertained by only κ, whereas μ is always around $\pi/2$, as expected.

Table 2. Parameters governing the models in Figs. 3 and 5.

		Model parameters				
		δ		ϕ		ω
		β	r_0	μ	κ	σ
Gestures	G0	8.3	850	1.54	2.4	0.30
	G1	20.8	743	1.56	8.0	0.23
	G2	22	740	1.56	7.0	0.23
	G3	22.7	728	1.55	7.4	0.24
	G4	16.7	674	1.56	6.5	0.23
Environment	G0 in S_B	10.3	834	1.54	3.1	0.25
	!G0 in S_B	25.1	740	1.56	9.0	0.21
	G0 in S_L	8.8	839	1.53	2.7	0.31
	!G0 in S_L	17.1	773	1.56	5.4	0.28

In addition, the remarks made in Sect. 3.3 are very easy to understand by looking at the model parameters. Namely, our observations from Fig. 5 imply that behavioral variation between G0 and !G0 in S_B is larger than their behavioral variation in S_L. This means that the model parameters need to have a larger difference in S_B than in S_L. This is indeed the case. For instance, while modeling δ, in case of the regularity term β we have $|10.3 - 25.1| > |8.8 - 17.1|$ and preferred distance term r_0 we have $|834 - 740| > |839 - 773|$. Similar conclusions are drawn by comparing the model parameters relating other benchmark variables. Therefore, the limiting effect of the environment on the behavior of gesture performing dyads is confirmed once more.

5 Integration of Motion Models into Social Robot Navigation

This study investigates the probability distribution of three benchmark variables (distance δ, alignment ϕ and velocity difference ω) in pedestrian dyads, and their dependence on the presence of gestures, and proposed mathematical models for such distributions. These models may be used to modify the navigation of social robots in order to provide a more natural interaction experience for their human partners. Although in this work we do not explicitly implement such a model in actual mobile robots, here we provide some guidelines concerning how such an implementation may be done.

One possible way of doing it is to use the models as terms in a cost function for the next velocity command of the robot [22]. Namely, the possible choices for the next (v, θ) (linear an angular velocity) command of the robot are discretized, and for each possible choice at the future position of the robot (and the corresponding estimated values of δ, ϕ and ω) are computed. The corresponding values of the model pdfs are then used as terms in the cost function of the

robot (higher pdf values correspond to a higher probability assigned to the corresponding (v, θ) choice. Of course, the cost function will include other terms (e.g. obstacle avoidance) and the inclusion of all relevant terms and computation of appropriate weights for each term is a non-trivial implementation problem, but from a conceptual viewpoint the introduction of the studied benchmark variables in such a cost function is straightforward.

The procedure above is standard in robotics. In the pedestrian simulation community, such models are often used to introduce a potential, and the acceleration of the agent is then derived as the gradient of the potential [15]. A detailed description of how to apply such a gradient based pedestrian model (Social Force Model) to robots moving in group is presented in another contribution to this conference.

6 Conclusion

This study fills the void of precise quantitative evaluation of effect of gestures on mobile interaction. We derive three benchmark variables from pedestrian trajectories and investigate their distributions with respect to presence/absence of gestures, number of concurring gestures and environment size. They are represented with three models, whose parameters are shown to reflect the distinguishing properties with significant accuracy.

- When humans perform gestures along with locomotion, they walk in closer proximity, with their velocities increasingly aligned and difference of velocities shrinking.
- The significant effect is pertained by the presence/absence of gesture(s) and not the type of gestures or the number of concurring gestures.
- The environment has an effect on the proposed variables proportional to the degree of its spatial limitation.
- The proposed models can be used to replicate similar behaviors in a robot by calibrating for the desired gesture patterns.

Together with the proposed implementation guidelines, our results bear the potential of enabling a social robot to provide a more natural interaction experience for the human counterpart of a robot-pedestrian group on-the-move.

Acknowledgments. This study was supported by JSPS KAKENHI Grant Numbers 15H05322 and 16K12505.

References

1. Knapp, M.L., Hall, J.A., Horgan, T.G.: Nonverbal Communication in Human Interaction. Cengage Learning, Boston (2013)
2. Streeck, J., Knapp, M.L.: The interaction of visual and verbal features in human communication. In: Poyatos, F. (ed.) Advances in Nonverbal Communication, vol. 10, pp. 3–23. Benjamins, Amsterdam (1992)

3. Hostetter, A.B.: When do gestures communicate? A meta-analysis. Psychol. Bull. **137**(2), 297 (2011)
4. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 222–235. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15892-6_24
5. Karam, M.: Ph.D. thesis: A framework for research and design of gesture-based human-computer interactions. Ph.D. dissertation, University of Southampton (2006)
6. Breazeal, C., Kidd, C.D., Thomaz, A.L., Hoffman, G., Berlin, M.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: IROS, pp. 708–713 (2005)
7. Rautaray, S.S., Agrawal, A.: Vision based hand gesture recognition for human computer interaction: a survey. Artif. Intell. Rev. **43**(1), 1–54 (2015)
8. Gleeson, B., MacLean, K., Haddadi, A., Croft, E., Alcazar, J.: Gestures for industry: intuitive human-robot communication from human observation. In: HRI, pp. 349–356 (2013)
9. Matuszek, C., Bo, L., Zettlemoyer, L., Fox, D.: Learning from unscripted deictic gesture and language for human-robot interactions. In: AAAI, pp. 2556–2563 (2014)
10. Salem, M., Rohlfing, K., Kopp, S., Joublin, F.: A friendly gesture: investigating the effect of multimodal robot behavior in human-robot interaction. In: RO-MAN, pp. 247–252 (2011)
11. Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., Joublin, F.: Generation and evaluation of communicative robot gesture. IJSR **4**(2), 201–217 (2012)
12. Haddington, P., Mondada, L., Nevile, M.: Interaction and Mobility: Language and the Body in Motion, vol. 20. Walter De Gruyter, Berlin (2013)
13. Mead, R., Matarić, M.: Autonomous human-robot proxemics: socially aware navigation based on interaction potential. Auton. Robots **41**, 1189–1201 (2016)
14. Gullberg, M., Holmqvist, K.: Keeping an eye on gestures: visual perception of gestures in face-to-face communication. Pragmat. Cogn. **7**(1), 35–63 (1999)
15. Zanlungo, F., Ikeda, T., Kanda, T.: Potential for the dynamics of pedestrians in a socially interacting group. Phys. Rev. E **89**(1), 012811 (2014)
16. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: survey of an emerging domain. Image Vis. Comput. **27**(12), 1743–1759 (2009)
17. Cohen, J.: Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bulle. **70**(4), 213 (1968)
18. Di Eugenio, B., Glass, M.: The kappa statistic: a second look. Comput. Linguist. **30**(1), 95–101 (2004)
19. Glas, D., Miyashita, T., Ishiguro, H., Hagita, N.: Laser-based tracking of human position and orientation using parametric shape modeling. Adv. Robot. **23**(4), 405–428 (2009)
20. Mardia, K.V., Jupp, P.E.: Directional Statistics. Wiley, Hoboken (2000)
21. Yücel, Z., Zanlungo, F., Ikeda, T., Miyashita, T., Hagita, N.: Deciphering the crowd: modeling and identification of pedestrian group motion. Sensors **13**(1), 875–897 (2013)
22. Murakami, R., Morales Saiki, L.Y., Satake, S., Kanda, T., Ishiguro, H.: Destination unknown: walking side-by-side without knowing the goal. In: HRI, pp. 471–478. ACM (2014)