



HAL
open science

Exploring Engagement in Human-Robot Interaction through the Quantification of Human Spontaneous Movement

Isabel Casso, Hendry Ferreira Chame, Patrick Henaff, Yvonne N.
Delevoye-Turrell

► **To cite this version:**

Isabel Casso, Hendry Ferreira Chame, Patrick Henaff, Yvonne N. Delevoye-Turrell. Exploring Engagement in Human-Robot Interaction through the Quantification of Human Spontaneous Movement. The 33rd IEEE International Conference on Robot and Human Interactive Communication - IEEE RO-MAN 2024, Aug 2024, Pasadena (CA), United States. 10.1109/RO-MAN60168.2024.10731439 . hal-04606893

HAL Id: hal-04606893

<https://hal.science/hal-04606893v1>

Submitted on 8 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring Engagement in Human-Robot Interaction through the Quantification of Human Spontaneous Movement

Isabel Casso¹, Hendry Ferreira Chame², Patrick Hénaff^{2,3}, Yvonne Delevoye-Turell^{1,4}

Abstract—Developments in the field of social robotics open interesting opportunities for applications in healthcare, education, and services. For this, studying *engagement* in human-robot interaction (HRI) is crucial for improving the quality of interactive experiences. Questionnaires are powerful in describing voluntary behavior; however, engagement is often an implicit non-voluntary behavior that reaches awareness only once initiated. Inspired by research in cognitive psychology, we propose a behavioral feature to quantify engagement in HRI through the measurement of spontaneous movement and spectral wavelet analysis. For this, we conducted an experiment during which participants listened to sad stories narrated by a moving social robot. Throughout the experiment, we tracked the participants’ spontaneous and non-voluntary sway movements with a motion capture system. The experiments were conducted with three robotic platforms (Buddy, Pepper, and Nao). Results showed that spontaneous body sway can be modulated by social robots within no-goal-oriented interaction. This opens up the application possibility of our methodology to various modes of HRI such as collaboration.

I. INTRODUCTION

Advances in social robotics constitute interesting opportunities for applications in healthcare, education, entertainment, and services. These robots are expected to be useful to humans, hence studying the design properties that offer optimal *engagement* in human-robot interaction (HRI) is crucial. Engagement has been defined as a process by which the perceived connection between participants in interaction is initiated, maintained, and ended [1]. However, the definition of engagement differs in the literature across modes of interaction (e.g., Human-Human (HHI), Human-Computer (HCI), Human-Robot). In their survey, Salam and colleagues determined engagement to be dependent on emotional, cognitive, and behavioral factors [2]. Common estimation features across the various interaction modes may exist, specifically, the quantification of spontaneous movement was considered as a non-verbal behavioral feature that would offer the means to infer human engagement. Studies in HHI have reported that spontaneous rhythmic movements like head nodding and body sway emerge continuously during interaction and are triggered without individuals’ awareness [3], [4]. Human spontaneous movements have also been quantified along with posture to assess engagement in HRI [5], [6]. These

HHI and HRI studies resorted to temporal motion analysis from motion capture or video recordings, suggesting that the quantification of spontaneous movement can serve as an additional indicator of engagement to other non-verbal behavioral indicators such as gaze, or face expression processing [7], [8].

The main contribution of our work is to propose a methodology based on wavelet spectral analysis to quantify spontaneous movement as a behavioral metric that can characterize engagement in HRI. To test our methodology, we defined a social (i.e., no goal-oriented) face-to-face interaction experiment where participants were invited to listen to sad stories told by a robot that exhibited emotional body language. Concurrently, their movements were tracked with a motion capture system. We hypothesized that the presence of a robot would affect spontaneous human sway and that spontaneous sway during interaction would be related to the social traits attributed to the robot. We also examined how different robot platforms might impact participants’ spontaneous swaying based on robots’ morphological differences and emotional body postures. Additionally, we compared these results with a control condition for which no robot was present.

This document is organized as follows. In Section II related works based on spectral analysis of non-verbal behavior are discussed for contextualizing the proposal of our methodology. Section III provides an algorithmic description of our method. In Sections IV and V the experiment is presented and the results are discussed, respectively. Finally, in Section VII the conclusions and research perspectives of our study are presented.

II. PREVIOUS WORKS

Within HRI, studies have considered the quantity of movement as a behavioral feature to determine the level of human engagement. In [5], [6], authors quantified spontaneous movement by computing the average motion of silhouette pixels on video frames. A disadvantage of this method might be the analysis of static frames, which could overlook the dynamic changes in human motion. As studies show, humans in interaction often exhibit sporadic and irregular movements (e.g. ballistic motions for pointing, changes in posture), and rhythmic movements (e.g. head nodding, body sway) emerging at different frequencies. HHI studies have proposed to quantify movement with spectral analysis of time series in the frequency domain using the Fourier transform [9], [10]. A limitation of this technique is that it assumes the time series to present a constant regular pattern, which does not

¹ Univ. Lille, CNRS, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France yvonne.delevoye@univ-lille.fr

² Université de Lorraine, CNRS, LORIA, F-54000 Nancy, France. hendry.ferreira-chame@loria.fr

³ ENIB - École Nationale d’Ingénieurs de Brest, Lab-STICC UMR CNRS 6285, Brest, France patrick.henaff@enib.fr

⁴ Senior member 2023 of the Institut Universitaire de France

correspond to situations of irregular motion which is non-stationary in nature. Alternatively, spectral analysis through the wavelet transform does not require time series to be stationary. Thus, it provides a time-frequency representation allowing the identification of frequency components and power spectrum magnitude at different time intervals [11]. For example, in [12] cross wavelet spectral analysis was employed for movement coordination. In [11] interpersonal synchrony patterns were evaluated. These studies suggest that spectral analysis through the wavelet transform is suitable for the quantification of movement characteristics such as temporal variability and intensity. To summarize, inspired by research in HHI, we propose to employ continuous-wavelet transform analysis to quantify spontaneous movement in HRI. We propose that the magnitude of the power spectrum can serve as an indicator of spontaneous movement emergence and thus, be taken as an objective feature for engagement inference. In the section below, we provide an algorithmic description of our methodology.

III. MATHEMATICAL MODEL

Wavelet analysis has been applied in several fields [13]. The wavelet transform is used to analyze time series, which contains non-stationary power at different frequencies. From the notation proposed in [14], let the discrete signal X_n with $n \in [0, N)$ represent a sequence of measurements of a participant's motion (e.g. a point at the torso) with respect to a reference frame fixed on the environment, obtained at an equal time spacing δt . The continuous wavelet transform of X_n is defined such that

$$W_n(s) = \sum_{i=0}^{N-1} X_i \psi * \left[\frac{(i-n)\delta t}{s} \right] \quad (1)$$

with η a non-dimensional parameter, s a scale factor, $*$ representing the complex conjugate, and ψ a wavelet function, which is taken as the Morlet wavelet, defined so

$$\psi(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2} \quad (2)$$

with ω_0 a non-dimensional frequency that must be taken (usually $\omega_0 = 6$) to satisfy the admissibility condition for a wavelet, which has zero mean and be localized in both frequency and time-space.

The wavelet transform $W_n(s)$ is a complex signal that can be separated into a real part (amplitude) $|W_n(s)|$ and an imaginary (phase) $\tan^{-1}[\Im\{W_n(s)\}/\Re\{W_n(s)\}]$. Then, the power spectrum can be calculated such as

$$P(s) = |W_n(s)|^2 \quad (3)$$

Sometimes in wavelet analysis it is required rectification of bias in the wavelet power spectrum [15]. The presentation of this problem deserves a detailed mathematical treatment which for simplicity is not given here.

We propose the algorithm *spontaneous movement engagement* (SME) as a behavioral metric for engagement inference in HRI through the processing pipeline shown in Algorithm

1. The procedures referred to are presented in pseudo-formal language in Algorithm 2 and described in detail in Section IV.

Algorithm 1

```

1: procedure SME
2:    $A \leftarrow initialize()$ 
3:   for  $r \in \{\text{Buddy, Nao, Pepper}\}$  do
4:     for  $f \in \{\text{Low, Medium, High}\}$  do  $\triangleright$  robot sway condition
5:        $P_{r,f} \leftarrow initialize()$ 
6:       for  $j \in G_r$  do  $\triangleright$  participants exp. group  $G_r$ 
7:          $X_{r,f,j} \leftarrow BuildSequence()$ 
8:          $P_{r,f,j} \leftarrow WaveletAnalysis(X_{r,f,j})$ 
9:        $A_{r,f} \leftarrow StatisticalAnalysis(P_{r,f})$ 
10:  return  $A$ 

```

Algorithm 2

```

1: procedure BUILDSEQUENCE ()
2:    $X_j^{3D} \leftarrow acquireData()$   $\triangleright$  for participant  $j$ 
3:    $\hat{X}_j^{3D} \leftarrow computeKinematics(X_j^{3D})$ 
4:    $\tilde{X}_j^{3D} \leftarrow filterOutliers(\hat{X}_j^{3D})$ 
5:    $X_j^{1D} \leftarrow reduceDimension(\tilde{X}_j^{3D})$ 
6:    $\hat{X}_j^{1D} \leftarrow downSample(X_j^{1D})$ 
7:    $X_j \leftarrow filterData(\hat{X}_j^{1D})$ 
8:   return  $X_j$ 

```

```

1: procedure WAVELETANALYSIS ( $X_j$ )
2:    $\gamma \leftarrow selectParameters()$ 
3:    $W_{n,j}(s) \leftarrow computeCWT(X_j, \gamma)$   $\triangleright$  Eqs. (1), (2)
4:    $P_j(s) \leftarrow computePS(W_{n,j}(s))$   $\triangleright$  Eq. (3)
5:    $\tilde{P}_j(s) \leftarrow scalePS(P_j(s))$   $\triangleright$  see [15]
6:    $\tilde{P}_j(s) \leftarrow T^{-1} \sum_{t=1}^T \tilde{P}_{j,t}(s)$   $\triangleright$  with experiment time  $T$ 
7:   if  $\tilde{P}_j(s) < \gamma$  then  $\triangleright$  test for aberrant motion
8:     return  $\tilde{P}_j(s)$ 
9:   else
10:    return  $\emptyset$ 

```

```

1: procedure STATISTICALANALYSIS ( $P_{r,f}$ )
2:    $A_{r,f} \leftarrow N^{-1} \sum_j^{j \in G_r} P_{r,f,j}$   $\triangleright$  participants' mean
3:    $\hat{A}_{r,f} \leftarrow filterBandPass(A_{r,f}, K)$   $\triangleright$  with bandwidth  $K$ 
4:    $\tilde{A}_{r,f} \leftarrow Y_N^{-1} \sum_y \hat{A}_{r,f,y}$   $\triangleright$  mean in freq. range  $Y$ 
5:   return  $\tilde{A}_{r,f}$ 

```

IV. EXPERIMENTS

A. Materials

As detailed in Table I and shown in Fig. 1, three robot platforms were used in the experiments, having distinct morphology, actuation modalities, and size. The software components included: Qualysis Track Manager to obtain motion capture data of human movement, the PyCWT module [16] was used for wavelet analysis, the library *naoqi* version 2.1 was used to program the robots Nao and Pepper. All programs were developed in Python language both version 2.7 and 3.

Hardware	Description
Robot Buddy	4 degrees of freedom (DOF), 0.56 m height, a cone-like base connected to an oval head, it has an animated face in a touchpad on the head.
Robot Nao	25 DOF, 0.57 m in height, presents a humanoid morphology showing bipedal walk and two arms, it also has motionless eyes and a mouth and its eyes can blink.
Robot Pepper	20 DOF, 1.2 m tall, featuring arms and a base with no legs. The head has eyes that emulate blinking, but the mouth remains fixed.
Computer	x64-based PC, i5-1235U processor, RAM 32 GB.
Qualisys	Six infrared cameras recorded 3D motion at 200 Hz of four reflective markers placed on the shoulders, base of the neck, and forehead of participants (see Fig. 1A)

TABLE I: Hardware components

B. Emotional body cues of robots

Considering our task was to listen to sad stories told by the robots, they were programmed to show behavioral cues corresponding with a negative affective valence. As such, the specificity of their actuation systems was carefully considered (see Fig. 1).

Buddy was programmed to exhibit sway motion in its sagittal plane (i.e. the plane XZ facing the human) while showing sad facial expressions on its touchpad head. The body morphology above the hip is relatively similar for Nao and Pepper, hence they were programmed to sway in a combined motion between the head and torso tilt, which intended to imitate human cowering movements generally associated with sadness [17]. However, there were some differences in their behavior. Nao held its arms against the torso, but this posture was not replicated for Pepper (see Fig. 1-B) since placing Pepper’s arms against the torso resulted in increased mass around its center of gravity, causing the hip’s actuators to overheat triggering warnings alarms, which could distract participants.

Sway motion was controlled in position, from a sinusoidal signal with a constant frequency and amplitude depending on the robot, sampled at 20 Hz. Three periodicity levels were selected to display different energy levels on a negative valence [18]: Low = 8 s (0.13 Hz), Medium = 4 s (0.25 Hz), and High = 2 s (0.5 Hz).

Three sad stories (mean length of 3 min) told in French from a first-person perspective were written and recorded with the *text-to-speech* method of Buddy supported by ReadSpeaker. The generated voice pitch had a fundamental frequency of 220 Hz. To control the impact of differences in robot voice synthesis, Buddy’s generated stories were stored and played back in Nao and Pepper.

C. Participants

Four independent groups of participants $N=76$ were recruited ($n_{Buddy}=20$, $n_{Nao}=20$, $n_{Pepper}=21$, $n_{No-Robot}=15$). All participants were fluent in French language, undergrad and post-grad psychology students, aged 18 to 35 years.

Participants were informed about the experiment’s general procedure at least 24 hours before their inclusion and gave their signed consent. All participants were naive to the objective of the study.

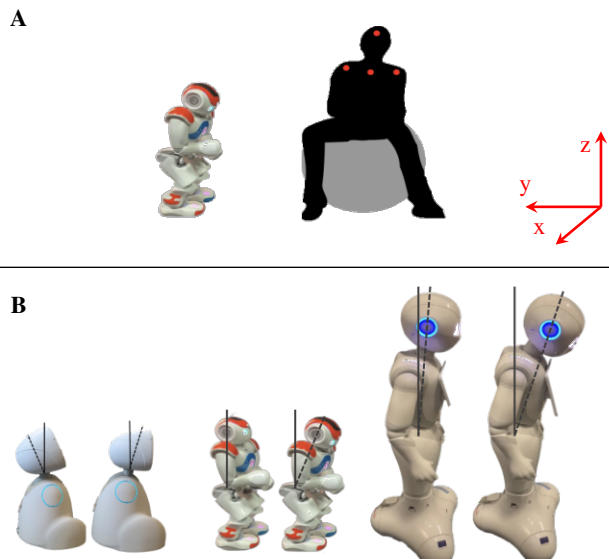


Fig. 1: (A) Illustration of the posture adopted by participants during the experiment. The red dots on the head, shoulders, and base of the neck, represent the tracked markers in relation to a fixed reference frame. (B) Emotional body language programmed on robots Buddy (left), Nao (center), and Pepper (right).

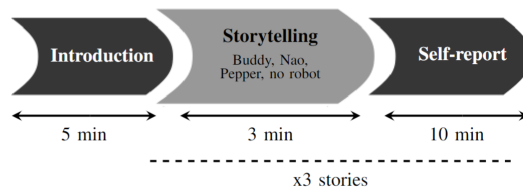


Fig. 2: Experimental protocol. The experiment videos are available at: <https://tinyurl.com/4nffm2dk>

The experimental procedure and inclusion of participants followed the strict code of ethics of the WMA Declaration of Helsinki for research involving human participants.

D. Experimental protocol

We conducted experiments in which participants were grouped for interaction with one of the three robots. As a control condition, a group of participants listened to three sad stories from a speaker (No-Robot condition).

The experiment structure is presented in Fig. 2. During the introduction phase, participants were instructed to sit on an ergonomic ball. We opted for this type of seating to allow for a wider range of motion and enable the non-voluntary spontaneous back-and-forth oscillation (i.e., sway) of their body, which is often restricted by conventional chairs. Then, participants watched a short video of 1 min duration to contextualize the experiment.

Panel A of Fig.1 shows the posture adopted by participants when facing the corresponding robot. During the storytelling phase, participants were instructed to listen to the stories told by the robot. No specific instructions were given to

participants other than to listen to the stories. Participants were not informed either about the valence of the stories or that the narration would be from the first-person perspective. During the storytelling phase, the robot played the three stories always in the same order, while displaying emotional body cues which were counterbalanced for every participant.

The environmental conditions of the experimental room were identical for all robots. Since Pepper is almost twice the size of the other robots, a platform was used to elevate the participants' position to have an equivalent view direction slope to the robot.

After every story narrated by the robot (Self-report phase), participants completed the Godspeed questionnaire [19] for the categories: *anthropomorphism*, *likeability*, *animacy*, and *perceived intelligence*. Semantic differential scales were used (e.g., *Fake-Natural*, *Apathetic-Responsive*, *Unfriendly-Friendly*). From a pilot evaluation, we decided to exclude the category *Perceived Safety*, since this category was insufficiently discriminating for the small-sized social robots considered.

E. Spontaneous Movement Engagement (SME)

This section explains in detail the procedures involved in the proposed method SME by referring to a specific line shown in Algorithm 2. For the *BuildSequence* procedure, data was acquired at 200 Hz (line 2), and markers velocity and acceleration were estimated (line 3). A filtering procedure was applied to extract all absolute acceleration peaks above 1000 mm/s^2 so filtered points were substituted by interpolation (line 4). Data in 3D was reduced to 1D by tacking the Euclidean distance between consecutive acquisitions of the torso's estimated location (line 5). Data was then down-sampled from 200 Hz to 25 Hz (line 6). A third-order Butterworth low-pass filter with a cutoff frequency of 11 Hz to the 3D was applied, followed by a highpass third-order Butterworth filtering with a threshold of 0.03 Hz to remove the low-frequency trend (line 7). The resulting data from previous steps correspond to the motion sequence for participant j . The Morlet mother wavelet was chosen with $\omega_0 = 6$ (see Eq. (2)). The procedure *WaveletAnalysis* involved selecting the wavelet transform parameters (line 2). To calculate the transform in Eq. (1) (line 3), we used the method *wavelet.cwt* of the module PyCWT, with parameters: time step $dt = 0.05$, starting scale $s_0 = 0.25$, time increment between scales $dj = 0.08$ sec, and the number of powers of two with dj sub-octave intervals $J = 100$. The wavelet coefficients were then squared to obtain the wavelet power spectrum (PS) conforming to (3) (line 4). To correct for the effects of the wavelet transform on the PS, the latter was multiplied by the inverse of the scale at each frequency (see [15]) and by a scaling factor (line 6). Only sway movements were quantified (i.e., rocking back and forth), while posture corrections such as stretching were not taken into account in this analysis. Participants who exhibited excessive movements during the adaptation period of the *introduction* phase (see Fig. 2) were excluded from the experiment (with threshold $\gamma = 0.6$, see line 7).

For the procedure *StatisticalAnalysis*, the power spectra for participants were averaged according to the robot condition (Buddy, Nao, Pepper, No robot) and robot sway periodicity condition (Low, Medium, High). We calculated the local minima of the PS to identify the significant frequency windows for the bandwidth filter [20]. The parameter K (see line 4) was selected so 90% of the PS was contained in the passing bandwidth.

V. RESULTS

Results are presented for the spontaneous movement quantification according to Algorithm 1, followed by the participants' responses to the Godspeed questionnaire.

A. Spontaneous human movement

Fig. 3 presents an example of the procedure *WaveletAnalysis* (see Algorithm 2) applied to participants who displayed spontaneous sway motion. In Fig. 4 we present an example of the operation *filterBandPass* (line 3), to determine the bandwidth for statistical analysis. Our analysis determined that SME tended to be higher when all robots moved at an 8s periodicity (0.13 Hz) in the sagittal plane.

A one-way ANOVA analysis indicated that the robot's sway motion globally exerted an influence on human spontaneous motion ($F(2,31) = 6.06$, $p = .006$, $\eta_p^2 = 0.28$). A Tukey posthoc analysis confirmed that the spontaneous movement of participants who interacted with Buddy had higher SME than those who interacted with Pepper ($t = 1.91$, $p = .035$) and Nao ($t = 3.43$, $p = .005$).

Concurrently, those who interacted with Pepper had higher SME than those who interacted with Nao ($t = 2.74$, $p = .006$). The post-hoc analysis did not reveal a statistical difference between the group that did not interact with a robot and the robot groups. Specifically, the no-robot group demonstrated higher SME variability among the three groups ($M = 0.10$, $SD = 0.14$), and tended to be lower than the Buddy group ($M = 0.15$, $SD = 0.10$) ($t = -0.95$, $p = 0.77$). The no-robot group tended to have higher SME than the Nao group ($M = 0.06$, $SD = 0.03$) ($t = 1.59$, $p = 0.39$), and than the Pepper group ($M = 0.09$, $SD = 0.03$) ($t = 0.89$, $p = 0.81$).

B. Robot social traits

A one-way ANOVA determined an interaction of Godspeed categories and Robots ($F(5,33) = 5.9$, $p < .001$, $\eta_p^2 = 0.16$). A Tukey posthoc test revealed statistical differences in all categories except for Anthropomorphism. We found Buddy ($t = 7.42$, $p < .001$) and Pepper ($t = 6.12$, $p < .001$) were perceived to be more animated than Nao, as shown in Fig. 5. As for the Likeability category, Buddy ($t = 7.13$, $p < .001$) and Pepper ($t = 7.85$, $p < .001$) were perceived to be more likable than Nao. Finally, in the Perceived Intelligence category, Buddy ($t = 8.07$, $p < .001$) and Pepper ($t = 8.52$, $p < .001$) were perceived to be more intelligent than Nao. Table II presents mean results for each category and robot.

In Fig. 6, we present the correlation matrix showing the relationship between the Godspeed categories, robot size, and Spontaneous Movement engagement (SME).

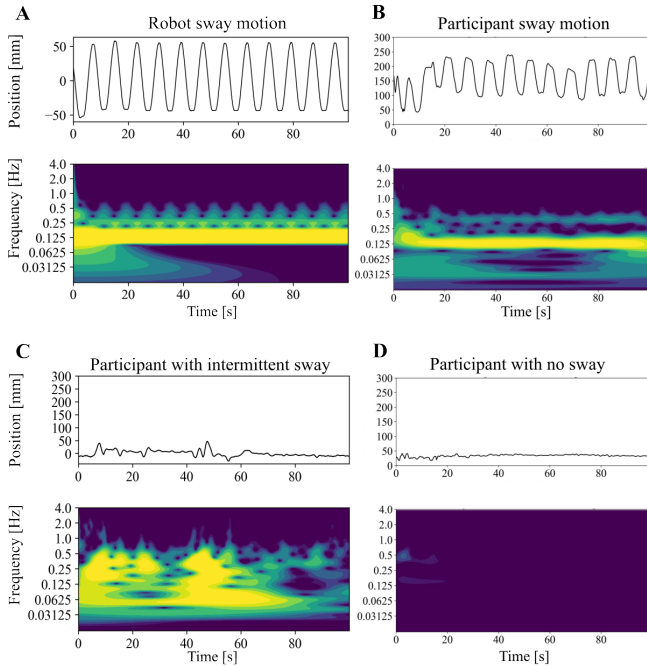


Fig. 3: Movement analysis examples. The upper panel of each figure presents the relative 3D position of the reflective motion marker in mm. The lower panel shows the power spectra heat map, where high and low power spectra are highlighted by bright and dark colors, respectively. High-power spectra indicate more and larger movements, while low-power spectra indicate an inhibition or absence of movement.

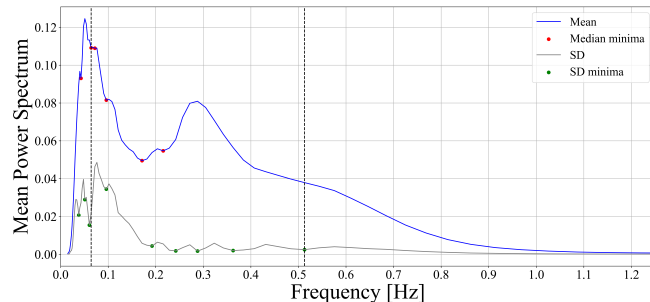


Fig. 4: Local minima identification example. Red and green dots correspond to the local minima in the mean and standard deviation curves of the power spectra, respectively. The dotted lines indicate the limits of the selected bandwidth for statistical analysis [20].

Category	Buddy M (SD)	Nao M (SD)	Pepper M (SD)
<i>Animacy</i>	3.3 (0.48)	2.26 (0.44)	3.29 (0.5)
<i>Likeability</i>	3.14 (0.73)	2.1 (0.46)	2.89 (0.63)
<i>Perceived intelligence</i>	3.41 (0.78)	2.24 (0.43)	3.36 (0.75)

TABLE II: Mean scores for Godspeed social trait categories.

VI. DISCUSSION

The social traits attributed to Buddy and Pepper were higher than Nao’s, which could suggest a preference for either a small non-humanoid robot with animated facial expressions or a preference for a bigger humanoid robot,

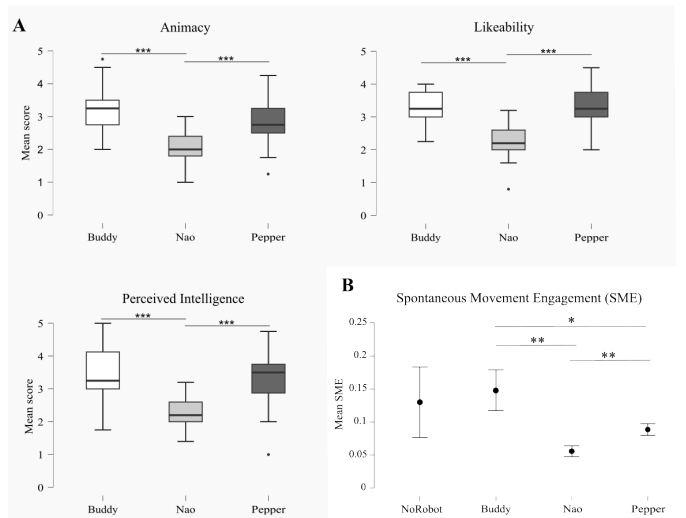


Fig. 5: (A) Results on the Animacy, Likeability, and Perceived Intelligence of the Godspeed questionnaire. Boxplots represent each robot. (B) SME from left to right: No-robot, Buddy, Nao, and Pepper. * $p < .05$, ** $p < .01$, *** $p < .001$



Fig. 6: Correlation matrix of the four Godspeed categories, robot size, and the SME of humans.

within social interaction with a negative (i.e., sad) context. Our results on human movement show that an 8s periodicity (Low condition) of robot movement has the potential to trigger spontaneous sway in participants without them being aware, while the other two periodicities (2s and 4s, for Medium and High conditions, respectively) seemed to have an inhibitory effect on SME. Significant differences in SME (see Fig. 5) indicate that Buddy elicited more spontaneous movement than the other robots. Comparing the figures of robot social traits and spontaneous movement (Part A and B of Fig. 5, respectively) we observed both Buddy and Pepper groups exhibited higher Godspeed scores and more spontaneous movement compared to the Nao group. Even if no statistical differences were found between the SME of the robot groups and the no-robot group, the latter tended to present a higher variability which could be explained by the lack of a visual anchor during the storytelling phase. Also, a possible explanation for Nao’s low SME and Godspeed scores could be attributable to factors such as the negative interaction context and its small humanoid morphology, as some participants tended to characterize it as a distressed

child. We observed a strong positive correlation between human movement and Animacy, Likeability, and Perceived Intelligence in Fig. 6. We found a strong positive correlation between SME and the Anthropomorphism category, suggesting that more anthropomorphized robots lead to increased spontaneous movement in participants. Furthermore, we noticed a moderate negative correlation between Anthropomorphism and robot size, indicating that smaller robots tended to be more anthropomorphized, which may have influenced participants' behavior. This can be explained by the fact that Nao has arms, legs, and a torso while Buddy has animated eyes and mouth.

VII. CONCLUSION

Our study has provided evidence that the quantification of human spontaneous movement during HRI is a non-invasive technique that can be used as a non-verbal behavioral feature for engagement inference. This method is complementary to other features such as gaze and facial expression analysis. Our results on human sway complemented the perceived social traits of the robots and revealed behavioral differences between groups. Results also showed that spontaneous movement can be modulated by the presence of social robots, in a face-to-face no-goal-oriented interaction. We reported that some categories in the Goodspeed questionnaire were perhaps redundant for estimating engagement, thus as a next step we plan to employ a more comprehensive tool (e.g., RoSAS [21]). Also, a larger experimental sample with non-university members should be studied to strengthen our results. We also acknowledge a potential confounding factor, such as the presence of an animated face on Buddy, which suggests the need to include another robot with animated facial expressions (e.g. Furhat). Finally, more research is required to explain the variability of sway in the no-robot group. Considering that context has been reported to influence engagement, we do not disregard the potential impact of the negative context on participants. Additional data shows that participants experienced lower valence and arousal after the experiment, and all groups perceived the stories as sad. Therefore, we aim to investigate how a positive interaction can affect SME. Although our methodology was applied in a listening task with no goal, we consider that our methodology has the potential to be applied to other modes of HRI (e.g., collaboration, guide-and-follow). In future studies, we plan to test the proposed metric in engagement inference where the robotic system adapts its movement to the SME of humans in real-time, instead of having established body cues. We hypothesize engagement will be stronger as the robotic system will simulate behavioral dynamics observed in HHI.

ACKNOWLEDGMENT

This work was funded by the 80-Prime 2021 program financed by CNRS granted to YDT. We thank the CNRS Federation (FR CNRS 2052 SCV), Equipex+ Continuum as well as the École des Mines de Nancy (FR) for access to equipment and the use of Nao. Thank you to Laurent Ott (SCALAB, UMR9193) and the engineers from CreativLab

Robotex 2.0 (U. of Lorraine, FR) for technical support. Finally, we thank the students who helped IC during the data collection and recruitment.

REFERENCES

- [1] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1-2, pp. 140–164, 2005.
- [2] H. Salam, O. Celiktutan, H. Gunes, and M. Chetouani, "Automatic context-aware inference of engagement in hmi: A survey," *IEEE Transactions on Affective Computing*, vol. 2, pp. 445–464, 2023.
- [3] W. Tschacher, G. M. Rees, and F. Ramseyer, "Nonverbal synchrony and affect in dyadic interactions," *Frontiers in psychology*, vol. 5, p. 1323, 2014.
- [4] A. Chang, H. E. Kragness, W. Tsou, D. J. Bosnyak, A. Thiede, and L. J. Trainor, "Body sway predicts romantic interest in speed dating," *Social Cognitive and Affective Neuroscience*, vol. 16, no. 1-2, pp. 185–192, 2021.
- [5] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani, "Fully automatic analysis of engagement and its relationship to personality in human-robot interactions," *IEEE Access*, vol. 5, pp. 705–721, 2016.
- [6] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *6th international conference on HRI*, 2011.
- [7] C. Oertel, C. De Looze, S. Scherer, A. Windmann, P. Wagner, and N. Campbell, "Towards the automatic detection of involvement in conversation," in *Analysis of Verbal and Nonverbal Communication and Enactment.COST International Conference, 2010*, 2011.
- [8] C. Oertel, G. Castellano, M. Chetouani, J. Nasir, M. Obaid, C. Pelachaud, and C. Peters, "Engagement in human-agent interaction: An overview," *Frontiers in Robotics and AI*, vol. 7, p. 92, 2020.
- [9] R. C. Schmidt, S. Morr, P. Fitzpatrick, and M. J. Richardson, "Measuring the dynamics of interactional synchrony," *Journal of Nonverbal Behavior*, vol. 36, pp. 263–279, 2012.
- [10] J. Issartel, T. Bardainne, P. Gailliot, and L. Marin, "The relevance of the cross-wavelet transform in the analysis of human interaction—a tutorial," *Frontiers in psychology*, vol. 5, p. 111259, 2015.
- [11] K. Fujiwara and I. Daibo, "Evaluating interpersonal synchrony: Wavelet transform toward an unstructured conversation," *Frontiers in psychology*, vol. 7, p. 516, 2016.
- [12] A. E. Walton, M. J. Richardson, P. Langland-Hassan, and A. Chemero, "Improvisation and the self-organization of multiple musical bodies," *Frontiers in psychology*, vol. 6, p. 130211, 2015.
- [13] M. Farge, "Wavelet transforms and their applications to turbulence," *Annual review of fluid mechanics*, vol. 24, no. 1, pp. 395–458, 1992.
- [14] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.
- [15] Y. Liu, X. San Liang, and R. H. Weisberg, "Rectification of the bias in the wavelet power spectrum," *Journal of Atmospheric and Oceanic Technology*, vol. 24, no. 12, pp. 2093–2102, 2007.
- [16] Krieger, Freij, Brazhe, Torrence, Compo, and contributors., "Python module for continuous wavelet spectral analysis," 2017.
- [17] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception*, vol. 33, no. 6, pp. 717–746, 2004.
- [18] M. Bretan, G. Hoffman, and G. Weinberg, "Emotionally expressive dynamic physical behaviors in robots," *International Journal of Human-Computer Studies*, vol. 78, pp. 1–16, 2015.
- [19] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.
- [20] S. Sozzi, A. Nardone, and M. Schieppati, "Specific posture-stabilising effects of vision and touch are revealed by distinct changes of body oscillation frequencies," *Front. in neurology*, vol. 12, p. 756984, 2021.
- [21] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas) development and validation," in *2017 ACM/IEEE Conference on human-robot interaction*, pp. 254–262, 2017.