# Investigating the Influence of Embodiment on Facial Mimicry in HRI Using Computer Vision-Based Measures

Maike Paetzel[1], Giovanna Varni[2], Isabelle Hupont[2],
Mohamed Chetouani[2], Christopher Peters[3] and Ginevra Castellano[1]

***Abstract*— Mimicry plays an important role in social interaction. In human communication, it is used to establish rapport and bonding both with other humans, as well as robots and virtual characters. However, little is known about the underlying factors that elicit mimicry in humans when interacting with a robot. In this work, we study the influence of embodiment on participants' ability to mimic a social character. Participants were asked to intentionally mimic the laughing behavior of the Furhat mixed embodied robotic head and a 2D virtual version of the same character. To explore the effect of embodiment, we present two novel approaches to automatically assess people's ability to mimic based solely on videos of their facial expressions. In contrast to participants' self-assessment, the analysis of video recordings suggests a better ability to mimic when people interact with the 2D embodiment.**

## I. Introduction

Mimicry – "the tendency to imitate facially, vocally or posturally people with whom we are interacting" [1] – serves important functions in social interactions, like establishing rapport, understanding people's emotional state and supporting mutual behavioral coordination. Several kinds of mimicry have been identified and examined in human-human social interactions (e.g. [2], [3]), revealing, for example, how mimicking behaviors are more frequently adopted in groups with close interpersonal relationships. Studies on human-human interactions even suggest that mimicry has a major influence on building rapport with others, both for the person imitating and the one being imitated [4].

Recently, researchers started to investigate how these findings relate to artificial agents and social robots. They found, for instance, that being mimicked by a social agent during an interaction can increase the likability and the rapport in the human partner [5][6]. However, even though the most basic form of mimicry in human-human interaction is the mirroring of facial expressions [7], this modality is still under-explored in human-robot interaction (HRI).

There are several factors that may influence human interaction behaviors in HRI and, in particular, the ability to mimic an agent [8]. Most related work suggests that a physical embodiment increases task success in HRI (e.g. [9], [10]). On the contrary, Bennett and Šabanović [11] found that the detection accuracy of facial expressions is higher in a digital avatar compared to a physical embodiment, which is supported by recent work about pain perception in social characters [12]. However, these studies are in contexts that are not related to mimicry. One of the few works looking at the effect of the embodiment on mimicry was performed by Hofree et al. [13]. They found the spontaneous mimicry of participants imitating an android robot to be stronger compared to a video recording of the robot, but a virtual replica of the robot was not included in the experiment.

The diversity of findings in related work shows that many open questions still remain regarding the influence of embodiment in HRI. In this work, we therefore explore embodiment as a potential influence on people's ability to mimic an agent by using a live interaction with the mixed embodied robot head *Furhat* [14] and a fully virtual 2D embodiment of the same character. The virtual agent is expected to have a higher social presence and might therefore elicit mimicry better compared to a video recording of the character as used by Hofree et al. [13], among others.

In order to address our objective, we designed an experiment in which participants were asked to intentionally mimic laughing artificial characters. We use laughter as a multimodal social signal that, as mimicry, displays a shared affiliative state and facilitates sociability and mutual cooperation [15]. Furthermore, the multimodal nature of laughter makes for an interesting and challenging stimulus both in the synthesis and the analysis of the expression.

In their work on mimicking an android robot, Hofree et al. [13] make use of facial electromyography to assess people's mimicry. This kind of technology is able to accurately capture complex facial muscle activation associated with specific facial expressions and their fast dynamics, but it is invasive.

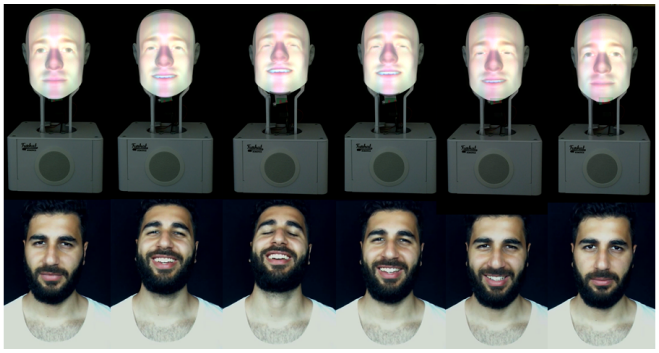

Fig. 1. Laughter synthesis on Furhat and one exemplar mimicry episode.

[1]Uppsala University, Department of Information Technology, Sweden {maike.paetzel, ginevra.castellano}@it.uu.se
[2]Université Pierre et Marie Curie, Institut des Systèmes Intelligents et de Robotique, Paris, France {hupont, varni, chetouani}@isir.upmc.fr
[3]KTH Royal Institute of Technology, Department of Computational Science Technology, Sweden chpeters@kth.se

Thus, in order to guarantee a natural interaction, new solutions using on-board sensors, for example video-cameras, are still missing. In this paper, we explore and evaluate two novel computer vision-based measures to automatically assess the quality of the user's facial mimicry of an artificial agent during interactions. In addition to an expert rating of the video recordings and participants' self-assessment, the two automated approaches are used to investigate the influence of embodiment when mimicking a laughing agent.

## II. Methodology

In this paper, we are empirically addressing the following research question: What influence does the mixed embodied robot Furhat have on the ability to mimic a social agent compared to a 2D control condition of the same character?

In order to investigate the influence of the embodiment on the ability to mimic, we propose two novel measures to automatically judge the ability of the user's mimicry. This is early exploratory work and the literature does not give a clear indication whether people will be better capable of mimicking the 2D or the 3D version of the character.

### A. Experimental design

We designed a within-subject experiment with the independent variable *type of embodiment* in which participants were asked to intentionally mimic an artificial character. Even though most mimicry in humans is performed unconsciously [16], due to the goal of this paper to investigate the influence of embodiment on the ability to mimic, we explicitly asked participants to imitate the social agent.

The character was presented as a *3D embodiment projected on a Furhat robot* and a *2D embodiment on a screen* as comparison. Joyful laughter was chosen since it is understandable regardless of the social context and is widely validated in terms of morphology (i.e. face, voice) [17].

The stimulus included vocal features, facial features and a head movement, which makes it an interesting subject of study in the automatic assessment of mimicry. As demonstrated in the case of human facial expressions [18], it is expected that participants are able to mimic this purely positive expression very well regardless of their relationship towards the origin of emotion.

### B. Participants

From the 24 students recruited to participate in the experiment, 3 had to be excluded either because the data analysis suggested a misunderstanding of the task or due to technical issues during the experiment. The 21 participants included in this paper (age: M = 26.38, SD = 4.79, min = 22, max = 37) were all, with the exception of one, enrolled in Computer Science or related subjects at Uppsala University. 28.6% of the participants identified with a female and 66.7% with a male gender (participants were allowed to withhold this information). All participants had at least a high-school degree, advanced English language ability and, according to self disclosure, most had little or no experience in acting.

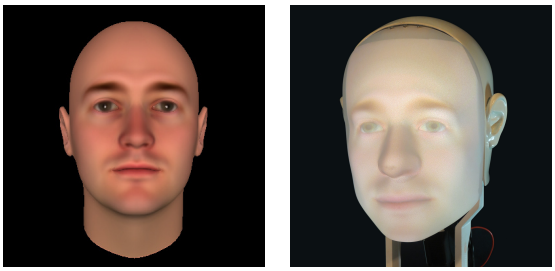

Fig. 2. The 2D (left) and 3D (right) version of the character.

### C. Apparatus and stimuli

A male virtual face created in FaceGen Modeller [19] is used in this study. It is presented to the participants in two different embodiments with approximately the same size: A 2D representation on a screen and a 3D back-projected version on a Furhat robot [14] (Figure 2). Furhat is equipped with a rigid mask of a male face on which the texture is projected from within. The virtual face was displayed on a screen with a resolution of 1600 x 1200 in portrait orientation. The character presents the main stimulus of joyful laughter as well as three alternative behaviors which were added in order to make it more difficult for participants to discern the focus of the study. In both the 2D and the 3D version, facial expressions were realized through the animation of the virtual face mesh. However, while head movements in the 2D face are achieved through virtual rotations, the Furhat robot has two motors with which it conducted real yaw and pitch head movements. Both embodiments have audio speakers attached beneath the face.

**Synthesis of the stimulus**

The laughter stimulus is composed of audio and the related facial expression. The CereProc synthesizer voice *William* [20], the standard voice of the Furhat robot, has six different laughter samples available. Since these laughter types have neither been classified nor previously been evaluated according to what type of laughter they represent, all six stimuli were presented to 30 participants (36.67% female, age M= 25.53, SD=5.66) in an online pre-study. One audio sample which the majority of participants (63.3%) defined as *joyful laughter* (length 0.5 sec) was selected for the main study.

In contrast to the audio stimulus, virtual facial expression of laughter is grounded in studies from psychology [21]. Our synthesis of the joyful laughter is based on the work by Ruch et al. [17], who describe joyful laughter according to the Facial Action Coding System (FACS) [22]. FACS decomposes facial expressions into small anatomically-based components called Action Units (AUs), and each AU intensity can be encoded in a 5-level scale ranging from "A" (trace) to "E" (maximum). Laughter mainly involves AU6 "*cheek raiser*", AU12 "*lip corner puller*", AU25 "*lips part*" and AU26 "*jaw drop*". Particularly, joyful laughter implies the activation of (Action Unit - intensity) AU6-C, AU12-C, AU25-C and AU26-C, in addition to a head back movement.

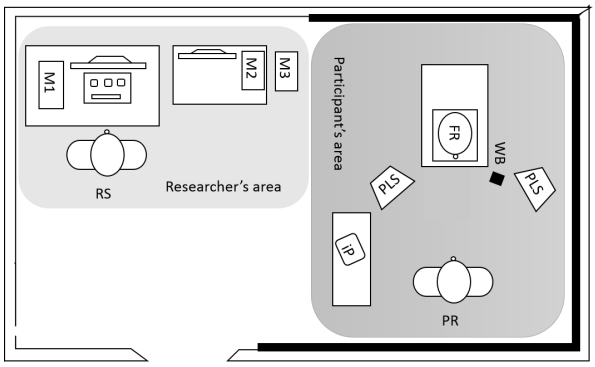


Fig. 3. A sketch of the experimental setup.

The descriptions of laughter in Ruch et al. [17] are however static. To synthesize the stimulus used in the study, we modulated this static description with the dynamics imposed by the audio stimulus. The resulting expression was validated by two experts on virtual characters and facial expressions.

### D. Experimental setup

The experimental sessions took place in a laboratory room at Uppsala University. Figure 3 illustrates the setup. The walls of the room were covered by black curtains to provide a homogeneous background and guarantee good visibility of the faces generated by the Furhat robot. The participant (PR) was standing in the participants' area at a distance of about 100 cm from the Furhat robot (FR) or the screen with the virtual face (VF). This value falls in the *personal space* of the participant according to the Hall's theory [23]. FR/VF was placed on a table at approximately 170 cm from the ground. A 800 x 600 resolution webcam (WB) operating at 30 fps was set atop a tripod to frontally frame the faces of participants during the sessions. Good lighting was guaranteed through a professional lighting system (PLS).

The Furhat robot and the sensors run on 3 different machines (M1 - M3). The applications for recording data streams in a synchronized way was developed using the EyesWeb XMI software platform [24]. Finally, an iPad (iP) was available on an additional table on the left side of the participant for obtaining information for the questionnaires. A researcher (RS) was present in the room to supervise the experiment and manage safety.

### E. Procedure

Prior to the on-site experiment, participants filled out an online questionnaire containing demographic and personality questions, which included an assessment of their level of *gelotophobia* (*"the fear of being laughed at"*) [25]. One participant with a mean gelotophobia rating beyond the standard cut-off value of 2.5 was excluded from participation.

After arriving at the session, participants were informed about the experimental procedure and signed a consent form. They then entered the participant's area where they were placed facing the character (FR/VF) and the webcam which would record their facial expressions. Participants were instructed to mimic the character's behavior in terms of facial expressions, head movements and voice. Once the participant was ready to start, the 2D or 3D embodiment of the character displayed the first trial behavior. A beep tone indicated the start and end of the displayed behavior. Participants were given 8 seconds to mimic the stimulus. A third beep tone indicated the end of that phase, after which they answered a questionnaire (Q1) about their self-assessed mimicry performance (cf. Section III).

Once finished with (Q1), the same behavior was displayed for a second and third time, each followed by a mimicry and questionnaire response phase. After finishing (Q1) for the third trial, the iPad guided participants to two additional questionnaires inquiring in-depth about the participant's perception of the stimulus. After responding to the last of those questionnaires, the same sequence of tasks was started for the next behavior displayed by the character. For each embodiment, four different behaviors were displayed: three alternative behaviors and the joyful laughter behavior, which was shown in either the second or the fourth position of the behavior sequence. The initial embodiment and the position of the laughter behavior in the sequence (second or fourth) was determined using Latin square to minimize ordering influences on the results. Participants were then given a five minute break, while the experimenter set up the other embodiment. After the break, the second mimicry session started. It included the same four behaviors, mimicry sessions and questionnaires as the previous embodiment in the same order. In the end of the experiment, a final questionnaire was presented to participants in which they were asked to elaborate on their experience in two free-text open questions.

## III. Human assessment of facial mimicry

### A. Participants' self-assessment

In questionnaire (Q1), participants were asked to rate *how well* they mimicked the character, *how much effort* the mimicry took them and *how comfortable* they felt mimicking the character. Each measure was formulated as a statement (e.g., "I mimicked the character very well.") and participants rated their agreement with this on a 5-point Likert scale.

Participant's self-assessment data was normalized using Min-Max scaling prior to the analysis in order to abstract from individual bias in people's ratings. The scaling was performed on all 6 trials per question and participant.

### B. Experts' assessment

Two researchers, trained in the generation and analysis of facial expressions, were asked to independently rate the video recordings of the mimicry episodes. The rating was based on the statement: "How well did the participant mimic the facial expression of laughter displayed by the character?" and performed on a 5-point Likert scale. The inter-rater reliability between the ratings provided by the two researchers was assessed through a two-way mixed, consistency average-measure ICC (Intraclass Correlation Coefficient) [26]. The ICC value of 0.77 was in the excellent range, with with a 95% confidence interval $0.684 < ICC < 0.832$ (F(121,121)

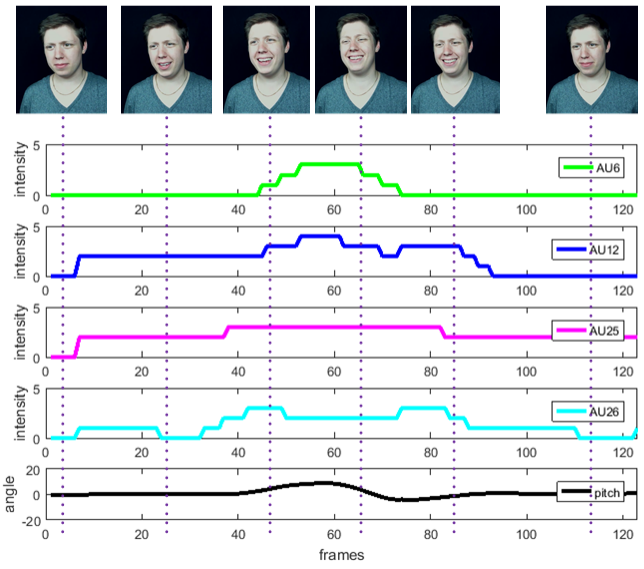


Fig. 4. Time series of AU6, AU12, AU25, AU26 and head pose pitch angle. AU intensities were assigned a discrete numerical value from 1 to 5 (*no activation* was coded as 0).

= 7.62, $p < .001$. This value shows that the two independent coders introduced only a small amount of measurement error. For further analysis, the average between the two experts' ratings is used.

## IV. Automated approaches to facial mimicry

In this section, two approaches to automatically assess people's ability to mimic are proposed. The novelty of both of these approaches is that they use only the video streams of a frontal video-camera framing the face of a human partner. The first approach (*intrapersonal*) is based only on the analysis of head pose and facial expression features the human mimicker performed, while the second one (*interpersonal*) exploits features from the interpersonal dynamics of mimicry, that is it takes into account the head pose and the facial expressions performed by the human mimicker and the agent mimicked.

Participants were given 8 seconds to mimic the character. However, mimicry episodes were generally shorter so the video recordings were manually cut according to the following criteria: (1) initial and final mimicry frames were detected; (2) a second before and after those frames was used as the start and end of the cut video, respectively.

Computer vision and machine learning techniques were used to automatically extract information about participants' head poses and facial expressions from the videos. A frame per frame detection of the head pitch angle and the intensity of AU6, AU12, AU25 and AU26 was performed for all recorded mimicry episodes. These AUs and pitch have been chosen as they are the most characteristic of the facial expression of laughter. AU intensities were detected through a system previously developed by the authors in [27], while head pitch was computed by using the IntraFace software package [28]. The final time series obtained were then smoothed to remove noisy samples. For AUs, a centered moving average smoothing technique was employed (window size: 10 frames), and pitch was low-pass filtered by using a $3^{rd}$ order Savitzky-Golay filter (window size: 19 frames). Extra post-processing was performed for pitch: the average of the pitch values for the first 15 frames was considered as an initial bias level and this constant value was subtracted from all the samples. Figure 4 shows the 5 time series extracted from a mimicry episode.

TABLE I

Intrapersonal features extracted per mimicry episode.

| Type | Feature name |
|---|---|
| **AU features** | mean activation<br>standard deviation of activation<br>maximum activation value<br>median of activation |
| **Head pose features** | mean pitch value<br>standard deviation of pitch<br>maximum pitch value<br>minimum pitch value<br>median value of pitch |

### A. Intrapersonal approach

To obtain a behavioral representation of joyful laughter, 21 features were computed from the time series. Table I displays the list of features, which are the following:

- **AU features** (computed for AU6, AU12, AU25 and AU26): *Mean*, *standard deviation*, *maximum value* and *median of activation*, computed over the whole AU time series representing the mimicry episode, were re-scaled in the range [0,1].
- **Head pose features:** *Mean*, *standard deviation*, *maximum*, *minimum* and *median* values were computed over the whole pitch time series. Prior to computation, the pitch time series was re-scaled to [-1,1] by dividing its values by 30, as the IntraFace head pose tracker has a [-30°,30°] pitch detection range [28].

By using the aforementioned features, a total of 126 feature vectors of 21 dimensions were created from our experimental data, one per mimicry episode. We included 4 trials collected from a discarded participant in this processing step, as he correctly understood the task but due to technical issues two of his trials with the 2D embodiment could not be recorded. Due to the imbalance between 2D and 3D trials, we decided to remove the user from the analysis of the embodiment effect. However, the user's recordings can still improve the pre-processing for the automated analysis.

The inspection of such a large number of high-dimensional information is noisy and its interpretation difficult. To overcome this problem, we applied manifold learning techniques [29] to reduce the dimensionality of our feature vectors from 21 to 2. In that way, each mimicry episode $m$ is represented as an XY point $P_m$ in a projected space.

The Multi-Dimensional Scaling (MDS) algorithm provided by the *Scikit-Learn* toolkit [30] was used. MDS

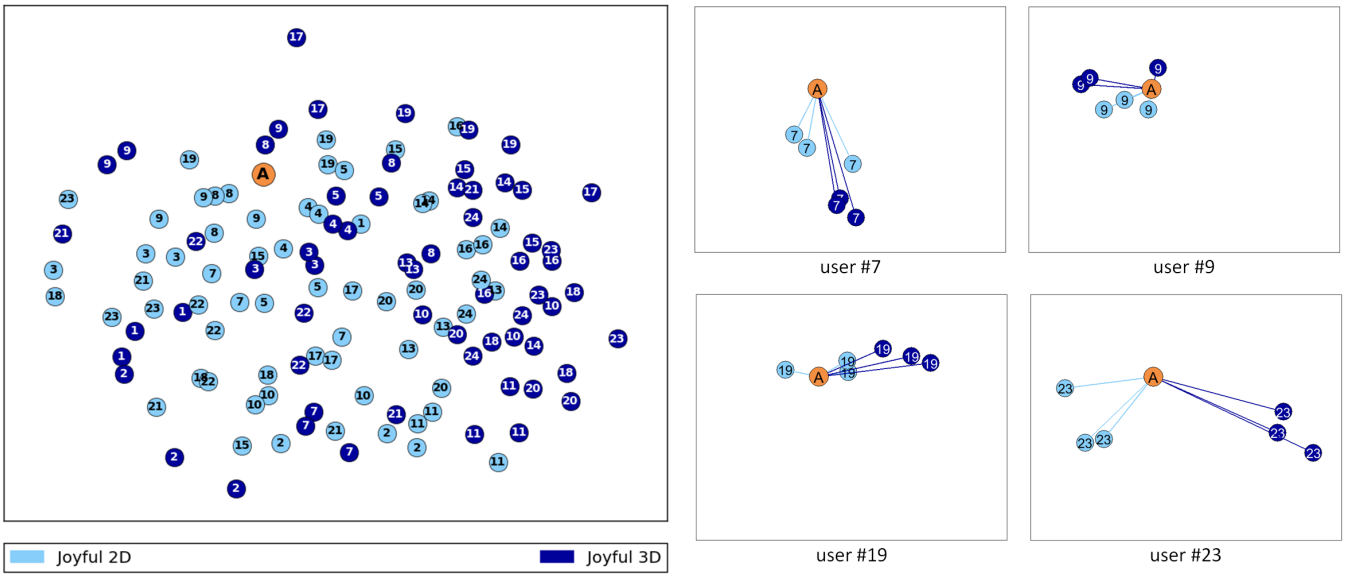


Fig. 5. MDS projection of intrapersonal mimicry vectors. Left: Each point in the embedded space represents a laughter mimicry episode *m*. User ID is included inside the dots. The projected point $P_A$ corresponding to agent's behavior is colored in orange and marked with letter "A". Right: Specific users' mimicry episodes, where blue lines represent distances $d_m$.

is especially appropriate for this purpose as it seeks a low-dimensional representation of the data in which the distances correspond well to the distances in the original high-dimensional space. Consequently, points representing mimicry episodes with similar behavior patterns will tend to be closer in the projected space, while trials with different behaviors should be much more distant. The agent's original behavior was also projected into this space, resulting in a point $P_A$. Finally, Euclidean distances $d_m$ between each mimicry point $P_m$ and $P_A$ were computed as intrapersonal measures of mimicry. Figure 5 depicts the projection of the data into the two-dimensional space.

### B. Interpersonal approach

Investigating mimicry also implies the study of the shared dynamics of the mimicker and of the mimicked partner. Due to the nonlinearity of the human behavior, traditional correlation-based analyses are not suitable to capture all the relevant aspects of the interpersonal dynamics. For this reason, we adopted a nonlinear method, the Cross-Recurrence Quantification Analysis (CRQA) [31]. It quantifies dependencies between two time series describing two dynamical systems in a generic feature space. CRQA is based on the concept of cross-recurrence introduced through the Cross-Recurrence Plot (CRP), a square/rectangular black and white area spanned by the two time series. Black points correspond to the times the two systems co-visit the same area in the feature space, whereas white points correspond to the times at which each system runs in a different area. The mathematical expression of a CRP is the cross-recurrence matrix (CR):

$$CR_{i,j}^{\vec{f1},\vec{f2}}(\epsilon) = \Theta(\epsilon - \|\vec{f1}_i - \vec{f2}_j\|),\ i = 1...N, j = 1...M \quad (1)$$

where $\vec{f1}$ and $\vec{f2} \in \mathbb{R}^d$ are the d-dimensional time series of the two systems having N and M samples, respectively; $\epsilon$ is the threshold to claim closeness between two points, $\Theta(.)$ is the Heaviside function and $\|.\|$ is a norm. CRQA holds on also by using categorical data. A CRP can be qualitatively analyzed at graphical level looking at the black points patterns in the plot. These patterns are hints of the joint dynamics of the two systems. CRQA enables a quantitative analysis of these patterns.

Participants (the mimickers) and the Furhat robot/virtual face (the mimicked) were described through a multivariate time series of the AUs and head pitch. Two CRPs and the corresponding cross-recurrence matrices were built for each trial both for the AUs ($CRP_{AUs}$, $CR_{AUs}$) and for the head pitch ($CRP_{hp}$, $CR_{hp}$). To build these plots, the time series of the agents were rescaled in amplitude to be comparable with those extracted from the participants. Then, a zero-padding of the shortest time series between that of the participants and that of the Furhat robot/virtual face was done to guarantee the time series had the same length. A city block distance was used for both the plots, $\epsilon_{AUs}$=6 was adopted for the AUs (this results in a maximum error tolerance of 30% on the whole set of AUs), $\epsilon_{hp}$=0.2 was adopted for the pitch (this results in a maximum error tolerance for the pitch of 6 degrees). In order to create a single plot, $CRP_{AUs\_hp}$, $CRP_{AUs}$ and $CRP_{hp}$ were merged together by computing the Hadamards product $CR_{AUs\_hp}$=$CR_{AUs} \circ CR_{hp}$. According to the cut of video recordings, $CRP_{AUs\_hp}$ was cropped by removing the first second and the last second + the zero-padded segment. Figure 6 shows an exemplary $CRP_{AUs\_hp}$.

The following CRQA measures (see [31, 32] for more details) were computed on this plot to explore the extent to which a participant was able to mimic the agent:

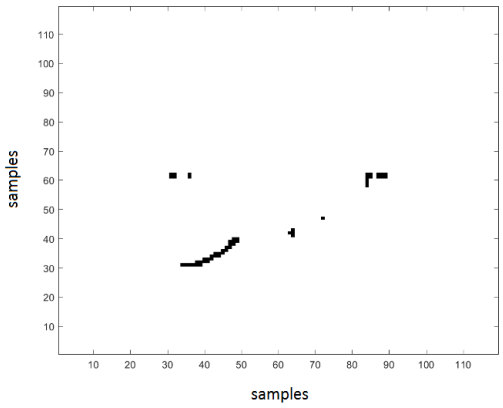


Fig. 6. An exemplary $CRP_{AUs\_hp}$ obtained during a trial with Furhat.

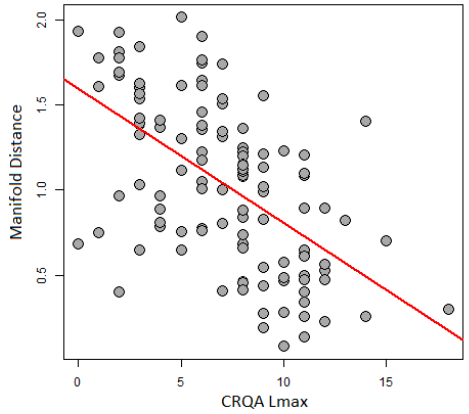


Fig. 7. Scatterplot of the correlation between the manifold distance and CRQA $L_{max}$ per trial. The linear regression is indicated by a red line.

**cross-Recurrence Rate** ($cRR$)
$cRR$ is defined as:

$$cRR(\epsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} CR_{i,j}(\epsilon) \tag{2}$$

and measures the density of recurrence points in CRP. It corresponds to the ratio between the number of the matrix elements "shared" by the mimicker and the mimicked and the number of available elements (i.e. all the elements of the matrix). It represents the overall extent to which the mimicker and the mimicked shared the same values of AUs and pitch. To claim mimicry, it is necessary but not sufficient to have a certain cRR: for example, the single isolated black points that can appear in the plot are taken into account in the cRR's computation but they are hints of randomness. Therefore, information about how the recurrence point are structured in the plot are also needed.

**Average diagonal line length** ($L$) **and length of the longest line** ($L_{max}$)
$L_\tau$ is the average length of the diagonal lines parallel to the main diagonal line in CRP. It measures how stable the mimicry behavior was: a high value of L indicates that the mimicker was able to repeat long sequences of AUs and pitch, whereas a low value of L indicates that the mimicker was able to repeat only short sequences of AUs and pitch. However, L-based measures do not take account of dynamical deviations between the behavior of the mimicker and the mimicked. A mimicker could have activated the AUs and reproduced the pitch exactly, but he or she could have performed the behavior with a different speed with respect to the mimicked. These deviations in the dynamics result in the curving of CRP diagonal lines. To solve this issue, we also computed the following measure.

**Length of the longest curved line** ($S_{max}$)
It computed the longest curved lines complying with these particular path connections: $(CR_{i-1,j-1}, CR_{i-2,j-1}, CR_{i-1,j-2})$. This implies also accounting for behavior performed with a 2x and a 0.5x speed deviation.

CRQA analysis was mainly carried out using the Python SyncPy library for interpersonal synchrony analysis [33].

## V. Results

From the total 126 data points (21 participants x 3 trials x 2 embodiments), four samples were removed due to technical failures. Corresponding trials using the other embodiment were manually removed from the data set in order to enable paired significance testing. This resulted in 118 data points which are considered in the analyses.

A Shapiro-Wilk normality test shows that the self-assessment data, expert rating data and all measures from the automated analysis are not normally distributed, except for $L_{max}$ ($W = .979$, $p = .061$), which only shows a light trend with respect to being normally distributed. In the following, we will use the non-parametric Wilcoxon rank sum test and the Spearman's rank correlation, both of which are robust to data that is not normally distributed.

A Wilcoxon rank sum test between the first and the third trial on all measures from the automated analysis, the self-assessment and the expert rating showed no influence of the trial on participant's ability to mimic the character ($p > .25$ for all measures), suggesting that no learning effect exists.

### A. Suitability of automated approaches to facial mimicry

The results from the *intrapersonal approach* and the *interpersonal approach* are highly negatively correlated with each other in all dimensions: manifold distance and $cRR$ ($r_S = -.628$, $p < .001$), manifold distance and $L_m$ ($r_S = -.496$, $p < .001$), manifold distance and $L_{max}$ ($r_S = -.585$, $p < .001$) and manifold distance and $S_{max}$ ($r_S = -.515$, $p < .001$). This suggests that a higher distance in the manifold data, which can be interpreted as a weaker ability to mimic the character, is also leading to a lower rating in the CRQA measures, which again relates to a weaker mimicry performance. Figure 7 shows an exemplary correlation.

All four CRQA measures are positively correlated with the average expert rating ($cRR$: $r_S = .117$, $p = .209$,

$L_m$: $r_S = .055$, $p = .563$, $L_{max}$: $r_S = .209$, $p = .023$, $S_{max}$: $r_S = .152$, $p = .101$). However, only the correlation between the expert rating and $L_{max}$ is significant. Similarly, the expert rating and the manifold distance are negatively correlated, $r_S = -.177$, $p = .055$, with a strong trend towards significance.

The participants' self-assessed ability to mimic and most CRQA measures are significantly positively correlated as well ($cRR$: $r_S = .202$, $p = .029$, $L_m$: $r_S = .154$, $p = .105$, $L_{max}$: $r_S = .198$, $p = .032$, $S_{max}$: $r_S = .274$, $p = .003$). The negative correlation between the manifold distance and participants' self assessed ability to mimic is only small and not significant ($r_S = -.066$, $p = .478$). However, we also only see a very small correlation between the participants' self-assessment and expert ratings ($r_S = .056$, $p = .546$), suggesting that the participants' self-assessment might not be very accurate and therefore is not a reliable assessment of the automated measures.

### B. Influence of the embodiment

Participants assessed their own ability to mimic to be slightly better in 3D compared to 2D, but the effect of embodiment is not significant using a paired Wilcoxon test ($Z = 230.5$, $p = .25$). Participants also reported they used slightly more effort when mimicking the 3D character compared to the 2D character ($Z = 133$, $p = .111$), but again the influence is not significant. Similarly, we do not see any influence of the embodiment on the self-assessed comfort during mimicry ($Z = 244.5$, $p = .812$). In addition, the embodiment shows no significant influence on the experts' assessment of the mimicry ($Z = 561.5$, $p = .459$).

In comparison to the manual rating of the mimicry both by the participants and the experts, which shows a slightly better rating for the 3D embodiment compared to 2D, the automated approaches to assess people's ability to mimic show exactly the opposite influence. The distance $d_m$ of the participants' trial in the intrapersonal approach is higher in 3D ($M = 1.08$, $SD = 0.06$) than in 2D ($M = 0.99$, $SD = 0.06$) and a paired Wilcoxon signed rank test shows that the effect of embodiment is significant, $Z = 589$, $p = .026$. The CRQA metrics show the same trend in the average ratings, for example, the average $cRR$ is slightly lower in 3D ($M = 5.6, SD = 0.7$) compared to 2D ($M = 6.4, SD = 0.7$), but the influence of embodiment is not significant (e.g. $cRR$: $Z = 1050$, $p = .214$).

## VI. Discussion

Based on the high correlation consistency between the intrapersonal and the interpersonal approach as well as the significant correlation with the expert rating in the important measure of CRQA, $L_{max}$, and the strong trend with the manifold distance, we consider *both suggested approaches to be valid automated metrics to assess people's ability to mimic a character's facial expressions.*

The influence of the mixed embodied robot platform Furhat on participants' ability to mimic the character is arguable. The participants' self-assessment suggest a slightly better ability to mimic the 3D Furhat platform compared to the 2D virtual counterpart. Even though this trend is not significant, the importance of the embodiment becomes evident in the written qualitative assessment. Here, *participants subjectively clearly favor the 3D over the 2D platform*, because "every movement of the eyes and small micro-expression were much clearer and noticeable in 3D", "due to the tangible face in 3D" or because it "was easy to follow". One participant even noted that the "2D character was not as pleasant to mimic as the 3D character". From all the participants who provided comments on the embodiment, only one subjectively preferred the 2D version.

Interestingly, our *automated measures suggest that participants objectively mimicked the 2D character better than the 3D version*. In the manifold projection, the distance of the 3D trials to Furhat is significantly higher than for the 2D trials. Our CRQA analysis points in the same direction, even though the influence of embodiment is not significant. We believe that with a larger number of participants the evidence in the CRQA assessment would become even clearer.

The automatically computed higher task success in 2D is interesting because it does not confirm related work in which participants had a higher task success when interacting with a 3D embodiment [9][10]. Hofree et al. [13] even found spontaneous mimicry to be stronger in the 3D embodiment compared to a 2D video recording of the same character. Our findings are rather in line with recent work suggesting that facial expressions are more easily detectable in a virtual version of a character [11][12]. Bennett and Šabanović [11] suggested it might be easier to maintain FACS fidelity in the digital version of the character which could explain their findings. Our work, however, shows that the FACS fidelity alone cannot fully explain a preference for a 2D embodiment, since the accuracy of the FACS codes was as high in our 3D embodiment as in the 2D version. Since we use a mixed embodied robot platform, an alternate potential explanation for our finding could be that small details are difficult to detect in 3D due to the slightly blurry projection onto the physical mask. In addition, the movement of the robot head might be more difficult to follow and and the noise potentially distracting due to the use of servos, which is not a challenge for the virtual character animation as it has the potential to be smoother and more quiet and life-like. Mixed embodied platforms are a comparably new technology and our findings highlight the importance of further exploring such platforms to better understand their dynamics.

The deviation from Hofree et al. [13] might also be related to the social presence of the virtual character, which could be higher in a live interaction with a character compared to a video recording. Future work is necessary to further investigate the influence of the embodiment, especially by including a real 3D robot platform and a video recording of the robot. We then aim to explicitly assess the character's social presence in the future to explore the link between the ability to mimic and the social presence of the embodiment.

Our findings highlight the importance of including both subjective as well as objective measures when assessing

task performance in social interactions with robots. Even if the task performance might be objectively higher in one embodiment, the perception of participants' own task success is relevant because it has potential influence on their future interactions. *Our study shows that an objectively higher task success is not necessarily related to the perceived task success and should therefore be assessed separately*.

Apart from the choice of embodiment, the study presented in this paper could be extended by using a more diverse set of stimuli. In our paper, we focused on laughter, which presents an important social signal especially in mimicry situations and it combines facial expressions, vocal features and head movements. For future work we will include a broader set of stimuli to ensure our findings are valid for the mimicry of an agent in general. In addition, we want to explore further influence factors on the ability to mimic a social agent and investigate in what way unintentional mimicry is related to the success in an intentional mimicry task.

## VII. Conclusions

In this paper, we present an experiment where people were asked to mimic a joyful laughter when interacting with two different types of embodiment: A Furhat mixed embodied robot platform and a 2D version of the same character. We introduce two novel approaches to automatically assess people's ability to mimic the character solely based on frontal video recordings which have a high correlation with each other and with two experts manually assessing the videos.

The two automated approaches suggest that people are better able to mimic the 2D representation of the character, while they subjectively prefer the 3D over the 2D embodiment. This finding is relevant, because it reveals that objective and subjective task-success are not necessarily related to each other. However both are important for designing successful long-term interactions with social agents.

## ACKNOWLEDGMENT

Thanks to L. Oestreicher for Fig. 2. C. Peters' work is partly supported by the European Commission (EC) Horizon 2020 ICT 644204 project ProsocialLearn. The work received fundings from ROMEO2 and Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements dAvenir programme under reference ANR-11-IDEX-0004-02. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

## References

[1] P. Bourgeois and U. Hess, "The impact of social context on mimicry," *Biological Psychology*, vol. 77, no. 3, pp. 343–352, 2008.

[2] U. Hess and A. Fischer, "Emotional mimicry as social regulation," *Personality and Social Psychology Review*, vol. 17, no. 2, pp. 142–157, 2013.

[3] U.Hess and A. Fischer, "Emotional mimicry: Why and when we mimic emotions," *Social and Personality Psychology Compass*, vol. 8, no. 2, pp. 45–57, 2014.

[4] M. Stel and R. Vonk, "Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction," *British Journal of Psychology*, vol. 101, no. 2, pp. 311–323, 2010.

[5] J. Gratch *et al.*, "Virtual rapport," in *International Workshop on Intelligent Virtual Agents*, 2006, pp. 14–27.

[6] L. D. Riek, P. C. Paul, and P. Robinson, "When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry," *Journal on Multimodal User Interfaces*, vol. 3, pp. 99–108, 2010.

[7] L. Brothers, "The neural basis of primate social communication," *Motivation and emotion*, vol. 14, no. 2, pp. 81–91, 1990.

[8] S. Boucenna *et al.*, "Robots learn to recognize individuals from imitative encounters with people and avatars," *Scientific reports*, vol. 6, 2016.

[9] D. Leyzberg *et al.*, "The physical presence of a robot tutor increases cognitive learning gains," in *CogSci*, 2012.

[10] J. Fasola and M. Mataric, "A socially assistive robot exercise coach for the elderly," *Journal of Human-Robot Interaction*, vol. 2, no. 2, pp. 3–32, 2013.

[11] C. C. Bennett and S. Šabanović, "Deriving minimal features for human-like facial expressions in robotic faces," *International Journal of Social Robotics*, vol. 6, no. 3, pp. 367–381, 2014.

[12] M. Moosaei, S. K. Das, D. O. Popa, and L. D. Riek, "Using facially expressive robots to calibrate clinical pain perception," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 32–41.

[13] G. Hofree, P. Ruvolo, M. S. Bartlett, and P. Winkielman, "Bridging the mechanical and the human mind: spontaneous mimicry of a physically present android," *PloS one*, vol. 9, no. 7, p. e99934, 2014.

[14] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems*, 2012, pp. 114–130.

[15] R. Dunbar, "Mind the gap: or why humans aren't just great apes," 2008.

[16] G. Dumas *et al.*, "Inter-brain synchronization during social interaction," *PloS One*, vol. 5, no. 8, 2010.

[17] W. F. Ruch, J. Hofmann, and T. Platt, "Investigating facial features of four types of laughter in historic illustrations," *The European Journal of Humour Research*, vol. 1, no. 1, pp. 99–118, 2013.

[18] P. Bourgeois and U. Hess, "The impact of social context on mimicry," *Biological Psychology*, vol. 77, no. 3, pp. 343–352, 2008.

[19] "FaceGen Modeller," http://facegen.com/, accessed: 2016-05-08.

[20] "CereProc," https://www.cereproc.com/, accessed: 2016-05-08.

[21] W. Ruch and P. Ekman, "The expressive patterns of laughter," in *Emotion, Qualia and Consciousness*, A. Kaszniak, Ed. Ed. Tokyo, Japan: World Scientific, 2001, pp. 426–443.

[22] J. C. Hager, P. Ekman, and W. V. Friesen, "Facial Action Coding System," *Salt Lake City, UT: A Human Face*, 2002.

[23] E. Hall, *The Hidden Dimension*. Anchor Books New York, 1969.

[24] S. Piana *et al.*, "Automated analysis of non-verbal expressive gesture," in *Human Aspects in Ambient Intelligence*. Atlantis Press, 2013, pp. 41–54.

[25] W. Ruch and R. T. Proyer, "Who is gelotophobic? assessment criteria for the fear of being laughed at," *Swiss Journal of Psychology*, vol. 67, no. 1, pp. 19–27, 2008.

[26] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology." *Psychological assessment*, vol. 6, no. 4, p. 284, 1994.

[27] I. Hupont and M. Chetouani, "Region-based facial representation for real-time action units intensity detection across datasets," *Pattern Analysis and Applications*, in press.

[28] X. Xiong and F. Torre, "Supervised descent method and its applications to face alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 532–539.

[29] M. Vlachos *et al.*, "Non-linear dimensionality reduction techniques for classification and visualization," in *8th International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 645–651.

[30] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 56, pp. 237 – 329, 2007.

[32] J. Serra, X. Serra, and R. G. Andrzejak, "Cross recurrence quantification for cover song identification," *New Journal of Physics*, vol. 11, no. 9, p. 093017, 2009.

[33] G. Varni, M. Avril, A. Usta, and M. Chetouani, "Syncpy: a unified open-source analytic library for synchrony," in *Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, 2015, pp. 41–47.