Mimebot - Investigating the Expressibility of Non-Verbal Communication Across Agent Embodiments

SIMON ALEXANDERSON, KTH - Speech, Music and Hearing
CAROL O'SULLIVAN, Trinity College Dublin
MICHAEL NEFF, University of California, Davis
JONAS BESKOW, KTH - Speech, Music and Hearing

Unlike their human counterparts, artificial agents such as robots and game characters may be deployed with a large variety of face and body configurations. Some have articulated bodies but lack facial features, and others may be talking heads ending at the neck. Generally, they have many fewer degrees of freedom than humans through which they must express themselves, and there will inevitably be a filtering effect when mapping human motion onto the agent. In this paper, we investigate filtering effects on three types of embodiments, a) an agent with a body but no facial features, b) an agent with a head only and c) an agent with a body and a face. We performed a full performance capture of a mime actor enacting short interactions varying the non-verbal expression along five dimensions (e.g. level of frustration and level of certainty) for each of the three embodiments. We performed a crowd sourced evaluation experiment comparing the video of the actor to the video of an animated robot for the different embodiments and dimensions. Our findings suggest that the face is especially important to pinpoint emotional reactions, but is also most volatile to filtering effects. The body motion on the other hand had more diverse interpretations, but tended to preserve the interpretation after mapping, and thus proved to be more resilient to filtering.

CCS Concepts: •**Computing methodologies** → **Motion capture;** *Motion processing;*

Additional Key Words and Phrases: Motion Capture, Perception, Cross-mapping

ACM Reference format:
Simon Alexanderson, Carol O'Sullivan, Michael Neff, and Jonas Beskow. 2017. Mimebot - Investigating the Expressibility of Non-Verbal Communication Across Agent Embodiments. *ACM Transactions on Applied Perception* 1, 1, Article 1 (January 2017), 13 pages.
DOI: 10.1145/3127590

1 INTRODUCTION

In animation, characters often convey non-verbal expressions that can be clearly interpreted by a human even when the character embodiment is different from that of a human. In robotics, there has been an increasing interest in robots that attempt to interact with humans on 'human-like' terms, e.g. using language, facial expression and gesture. Similar to animated characters, robots may be deployed with a large variety of face and body configurations. Typically, they have significantly fewer degrees of freedom than a human body. Common robot platforms either have articulated bodies and limited facial features, or have heads ending at the neck.

The general question addressed in this work is how to *quantify* and *improve* the non-verbal expressive capability of diverse embodiments (robots and other artificial agents), for the purposes of more engaging and efficient interaction with humans. The ability to quantify the non-verbal expressive range of a given robot configuration

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. XXXX-XX/2017/1-ART1 $15.00
DOI: 10.1145/3127590

is crucial, not only in order to be able to maximize the expressibility of an existing robot, but also to investigate new robot embodiments, e.g. what degrees of freedom are most important to implement in a physical system.

Our underlying assumption is that in order to compensate for the lack of degrees of freedom, we need to emphasize clarity in the motion. Our approach is to map motion captured from an actor to different (virtual) robot embodiments. By comparing the original and mapped motion, we are able to measure, through perceptual experiments, how much information is retained in the different embodiments.

Our dataset contains non-verbal communication reflecting different internal states of the robot, and was recorded using a professional mime artist. We primarily target states important for human/robot communication such as displaying levels of certainty [27] and attentiveness [22] but also affective states important for social bonding and empathy [6]. We collected three data sets where the actor specifically was asked to use a) only head and face (Face-only), b) only body (Body-only) and c) face and body (Face-and-body) to express a variety of inner states.

In this paper, we present our framework for data collection and a first study on the emotion subset of the data. The study investigates how emotions are perceived across three different embodiments of a humanoid robot-like character, using the customized data-set for each embodiment. The evaluation experiments are performed using virtual embodiments rather than existing robot platforms.

The main contributions of this paper regard the way expressive motion is studied across embodiments and the framework in which the motion is applied to the embodiments. Instead of recording natural human motion, we record motion specifically tailored to fit the different embodiments. In this sense, we favor believability and appropriateness over realism. Another contribution of this paper is the application of mime acting techniques to the field of robots and agents. While there has been a large interest in stylized motion in both the robotics and computer graphics communities, the predominance of work has been carried out on manually crafted animation, and the relevance of mime in the field of artificial agents has not been covered. We believe that mime actors could serve as a valuable source to generate high quality, stylized motion suitable for data-driven methods requiring large amounts of data.

2 RELATED WORK

It is well known that people can identify human motion and intention from extremely simplified representations such as point-light displays [14]. Atkinson et al. [2] showed that also basic emotion can be recognized through such displays. The perception of emotional body language has also been studied on virtual characters. McDonnell et al. [19], [20] recorded motion capture data of the six basic emotions of Ekman [9] performed by an actor, and mapped it to virtual characters in different categories: realistic, cartoon, zombie and a wooden mannequin. Their results suggested that the perception of emotions are not affected by the appearance of the character. In a similar setup, Beck et al. [5] found that emotions were perceived more strongly on a real actor than on a virtual agent, and that stylized emotions were perceived more strongly than natural.

Hyde et al. [13] studied the effects of damped and exaggerated facial motion in virtual characters rendered realistically or in a cartoon style and found that realistically rendered characters benefited from exaggerated motion and cartoon rendered characters benefited from damped. In a study by [21], images of illustrated characters showing emotions of fear and anger were manipulated to congruent and incongruent stimuli by exchanging the faces in the images. The results indicated that the judgments of facial expression were biased towards the expression of the body in the incongruent cases.

In the field of robotics, perception of emotion has been studied on a variety of existing robot platforms. Emotional body laguage for the Nao robot was studied by [5] and [12]. For a non-humanoid sphere-shaped robot (Maru), sound, colors, and vibrations were used to express emotional behaviors [26].

Other researchers have adopted principles of animation [30] to create expressive motion for robots. Gielniak et al. [11] propose an algorithm to exaggerate gestural movements and found positive effects on the memory

and level of engagement of an interaction with the Simon robot acting as storyteller. They also found that the robot was perceived more cartoon like and entertaining to interact with. [28] used the animation principles of anticipation and reaction for the PR2 robot which was shown to give impact on readability and perceived intension. The work by Ribeiro and Paiva [25] takes inspiration from animation and puppetry in the creation of emotional expressions for the EMYS robot. While our approach also encompasses the use of stylization to increase the comprehensibility of non-verbal communication, we use the principles and techniques of mime acting as a data source. This enables us to create larger data sets and faster turn-around times in the design process.

Some studies have evaluated the effects of emotional expressiveness with task based experiments, such as negotiation [7] or persuasion tasks [31], rather than questionnaires and self-ratings. Such metrics can be used to get an indirect measurement of specific inner states (e.g. how negotiation performance correlates with expression of happiness or anger). In our study we needed a more direct metric applicable to every inner state and embodiment in the database, and rate the emotions on the same scales as when produced by the mime actor.

2.1 Mime acting

A key issue in developing appropriate movement stimuli from motion capture is to decide on the performers from whom to source the data. It is tempting to use laypeople as subjects since their movements are likely natural and all people make consistent use of nonverbal communication. However, this is often a poor choice as laypeople may not have the clarity of movement required, especially when aspects of their performance are lost in the capture process, and they may send mixed signals.

As explored in this work, mime actors appear to offer a particularly good source for movement data. Mime is generally performed in silence, so mime actors have extensive training in using their bodies, and their bodies alone, to communicate [16]. Their work is based on the same twin principles of simplification and exaggeration that underlie effective animation [30]. Mimes are trained to find the most extreme form of expression, from which they can pull back as needed, to go beyond natural gestures and to never be vague in order to generate clarity in expression [16]. They may slow timing [15] and exaggerate balance [3] to make communication more clear and generate interest. Importantly, the movements in mime must always be justified and react to the context and performer's surrounding space [16]. Robot actions must also be based on this sensitivity to context and it is why our stimuli are defined as reactions to statements that establish a context. Finally, mime actors will often use movement to create a physical environment that that is unseen, for example, showing the caring of a suitcase through movement, when there is no physical suitcase on stage. They use displays of counterbalance to generate the illusion of physical objects [8]. Such adjustments may be particularly important in the context of robotics, where the physical abilities of the robot are likely unknown, but human-like compensations will read clearly to subjects.

3 METHOD AND MATERIAL

3.1 Script preparation

Together with the mime actor (lecturer in mime acting at the Stockholm School of Dramatic Arts), and a director of mime and puppet shows, short scripted interactions were prepared for two hypothetical scenarios of human-robot interaction. In the first scenario the robot is used as a waiter/bartender in a restaurant, and in the second as a personal assistant for household use. We considered the following inner states important to be able to display non-verbally: certainty, attentiveness, engagement, and positive and negative affect. The first three states were chosen as they are important for conversational dynamics and have strong audiovisual features [27], [22], and the latter two for increasing the social bonding [6]. The two affective states were divided into scales according to the cross-diagonal of the circumplex model of emotion [24], one ranging from sad to joyful and the other ranging

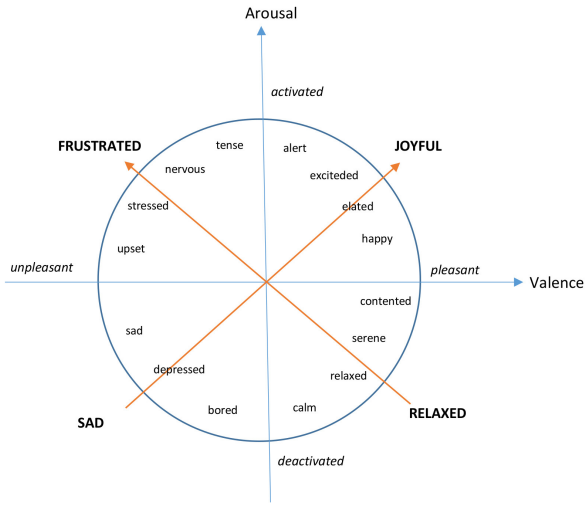

Fig. 1. Circumplex model of emotion.

Table 1. Example interactions for the five inner states. The robot lines are marked in bold

Inner State	Example script
Uncertain-certain	How many pills did I take today |**One in the morning two in the afternoon**
Unattentive-attentive	Wake me up at 9 am |**mmm-hm** |And remind me to go shopping |**uh-huh** | We need something for Kate |**ok**
Not engaged-engaged	**Take a right after the church** |mmm-hm |**keep going until the roundabout** |mmm-hm ...
Relaxed-frustrated	Can we have another round |**Sorry the bar closed at one**
Sad-joyful	Do you remember the meeting with Jill last week? |**Oh yes**

from relaxed to frustrated, see Figure 1. The same categories have previously been used by robot researchers in [26]. In total, our scripts contain five to seven interactions for each of the inner states. Table 1 shows some example interactions in each series.

We designed each interaction explicitly to be context neutral and expressible with different variations of the inner state. As an example, the answer to the question "Who was the third president of the United States?" may be expressed both with high and low levels of certainty, and the interaction "Interlocutor: -Do you remember the meeting with Jill last week? Robot: -Yes" may be expressed with both sadness and joy. All affective states contain an emotional reaction and start from neutral expression. In the previous example, the emotional reaction follows the question about the memory of the meeting. We preferred this to constant states of emotion as it gives context to the emotion under the principle of "reaction creates action" [16]. As the mime artist expressed it "the emotion spreads through the body like a drop of color hitting the surface of a glass of water".

Fig. 2. Marker sets used in the recordings. a) Full set used in the range-of-motion recordings; b)-d) marker sets used the Body-only, Face-and-body and Face-only data sets.

3.2 Data recording

In order to obtain highly detailed data of the mime actor's performance, we performed a full performance capture (including fingers, face and eye-movements) in our motion capture lab. The lab is equipped with a NaturalPoint Optitrack motion capture system (Motive 1.9.0) with 16 Prime 41 cameras. The cameras have a resolution of 4 mega-pixels and were operated with a frame rate of 120 fps. The number and placement of retro-reflective markers varied between the data sets for the three embodiments as shown in Figure 2. For the Face-only and the Face-and-body data sets we captured the actor's eye-movements with Tobii Glasses 2 eye-tracking glasses at a sampling rate of 100 Hz. We placed four markers on the glasses, two symmetrically placed on the front and two along the sidepieces to calculate the world space coordinates of the gaze target. This allowed the gaze tracking to be robust to slipping of the glasses during recordings. In addition to the motion capture and gaze data, we recorded full HD video using a JVC Everigo camera (full HD at 25 fps), stereoscopic video with a ZED camera (2 x full HD at 30 fps) and audio with a Studio Projects C1 large diaphragm condenser microphone and an RME FireFace 800 external sound card. To provide a sync-signal for all data-streams, we used a clapboard equipped with reflective markers at the start and end of each recording.

The recordings were carried out in separate stretches for each inner state and embodiment, and contained all interactions and intensity levels for that state. The intensity levels were carried out in levels 3,4,5, then 3,2,1 (for example neutral, frustrated, very frustrated, then neutral, relaxed, very relaxed). This allowed the actor to use the neutral expression as starting point for each scale. Between interactions, the actor could see the script for the next interaction on a prompt, and after the interaction was performed he was asked to repeated it until there was a common consensus among the actor and two supervising researchers that the expressed non-verbal communication reflected the intended level and state.

For the Face-only data set, the actor was sitting down and asked to use the head and face to express the inner states. For the Body-only data set, the actor wears a white neutral mask. The use of a neutral mask is a traditional mime technique employed to induce an acting style where everything is conveyed with body motion. It also serves a purpose in our evaluation experiments by blocking off impressions from the face for comparisons with the Body-only condition. For the Face-and-body data set, the actor was asked to express himself with the complete body.

3.3 Data synchronization, cleanup and completion

After the recordings were completed the marker data was labeled and gaps were filled. We used the software provided by the motion capture system, Optitrack Motive 1.9.0, for labeling and gap filling of the markers on the

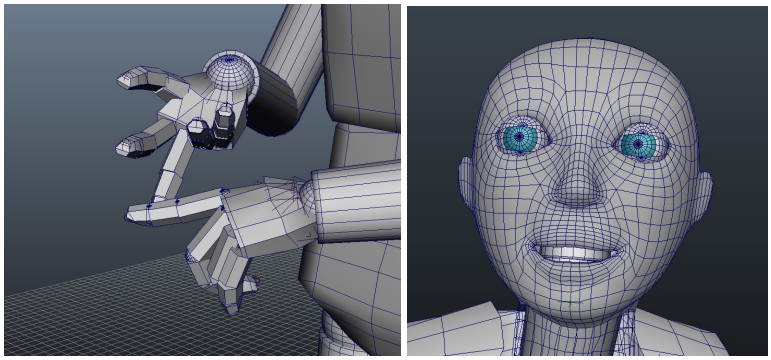

Fig. 3. Examples of retargeted fingers and face.

body. For the markers attached to the hand and the face, which are notoriously time consuming to clean, we used the methods of Alexanderson et al. [1] for automatic labeling, and of Baumann et al. [4] for automatic gap-filling.

For the marker sets with reduced numbers of markers on the fingers and/or face, we used the data from the range-of-motion (RoM) to reconstruct the remaining markers using Kernel Canonical Correlation Analysis (kCCA). We used a similar implementation as in [10], where kCCA was used for data-driven polygon mesh deformation by estimating bone transformations from a small set of control points. For the finger data in the Face-and-body data set, we reconstructed each of the markers on the proximal phalanges individually using the fingertip marker on the same finger as input data. For the Face-and-body and the Face-only data sets we reconstructed all the extra markers in the RoM set except the markers on the lower eye-lid, which proved to have too much uncorrelated motion to yield a satisfying result. In the reconstruction of face markers, we used all markers as input data to the kCCA algorithm.

3.4 Character creation, solving and retargeting

We created and rigged a 3D model of a humanoid robot-like character using Autodesk Maya. Two different versions of the head were created, one with facial features for use on the embodiments with face and one blank mask without any facial features for use on the one without (Figure 4). The proportions of the character's skeleton were directly taken from the motion capture system. To avoid uncanny valley effects, we designed the faces to have a stylized look inspired by robots with a medium mechano-humanness score [18], and used simple non-photo realistic textures [17].

The solving and retargeting from the marker and gaze data to the character rig was performed using three different methods: The skeleton motion for the body (excluding fingers) was solved with the standard Autodesk Motion Builder procedure, the finger motion was solved with inverse kinematics applied to a hand skeleton using the IKinema Action plugin to Autodesk Maya, and the face was solved using Softimage Face Robot. We used stabilized marker data expressed in a local coordinate system following the hands and head for the face and finger solving. As the Face Robot system uses slightly different marker placements than our setup, we estimated the locations of missing markers in the following way. Center lip markers (upper and lower) were approximated as the center of spline curve arc, using the corner and the upper/lower lip markers as control points. Nostril markers were approximated as the center of the line between the sneer markers and the nose tip. Ear, fore-head and nose bridge markers where approximated to be static. As we wanted the face to have a stylized look, we did not use the wrinkle paint feature in Face Robot. Eye movements were added as aim constraints to follow the gaze target data from the eye-trackers, and we inserted blinks at the places where there were gaps in the eye-tracking data. As a final step, we let the upper eye-lid follow the pitch rotation of the eyeballs according to

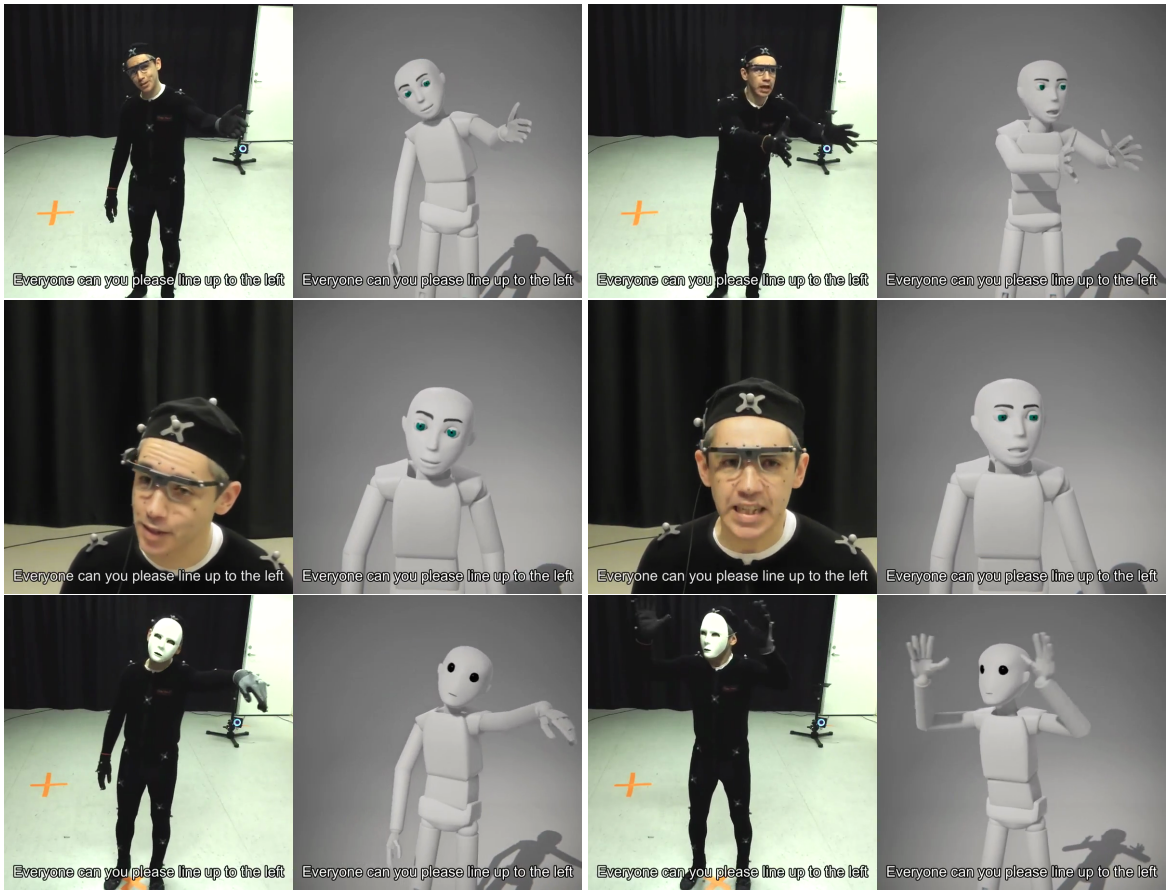

Fig. 4. Corresponding images of the recorded mime actor and the retargeted virtual character for the three data-sets. Upper: Face-and-body, Mid: Face-only and Lower: Body-only. The images show the extreme levels for the Relaxed-frustrated series.

common animation practice [23]. The final videos were rendered in the same frame rate as the video recordings, 25 fps. Examples of the retargeted fingers and face are shown in Figure 3. Still shots from the recording and corresponding rendered animations are shown in Figure 4.

4 CROWDSOURCED PERCEPTUAL STUDY

We executed two sets of experiments on the emotional sub-set of the recorded data. The first experiment was a prestudy performed on the videos of the mime actor, where we assessed how well the intended emotions are recognized and selected the most salient interactions. In the second experiment, we evaluated the filtering effects of the cross-mappings to the virtual embodiments and made a comparison between embodiments. In all experiments, we removed the sound and replaced it with subtitles. This was done to remove the influence of prosodic features in the evaluations. The experiments were conducted on a crowdsourcing platform built for academic studies, Prolific Academics.

Table 2. The three most frequently used words to describe the emotions shown in the selected interactions. Number of occurrences are given in parenthesis.

Emotion	Face-and-body	Face-only	Body-only
Sad	sad (8), disappointed (4), regret (4)	sad (6), disappointed (4), annoyed (4)	bored (6), disappointed (4), frustrated (3)
Joyful	happy (21), joy (9), pleased (3)	happy (25), joy (6), understanding (4)	surprised (12), excited (6), happy (5)
Relaxed	bemused (4), relaxed (3), confused (3)	bored (5), relaxed (4), confused (3)	bored (7), confused (7), annoyed (5)
Frustrated	frustrated (6), annoyed (6), impatient (4)	angry (14), frustrated (9), annoyed (9))	angry (14), frustrated (11), annoyed (8)

Table 3. Number and percentage of unique words used to describe the emotions shown in the selected interactions. Lower percentage values means more unanimous identification.

Emotion	Face-and-body	Face-only	Body-only
Sad	39 (66%)	56 (71%)	51 (75%)
Joyful	18 (36%)	38 (48%)	46 (65%)
Frustrated	27 (51%)	40 (47%)	31 (46%)
Relaxed	38 (72%)	62 (76%)	50 (67%)

4.1 Prestudy

To assess how well the acted emotions are identified and to select the best interactions, we conducted a prestudy on the video clips from the recordings.

We performed three different experiments in the prestudy, one for each embodiment (Face-and-body, Face-only and Body-only). In each experiment, the participants were shown videos of the real actor performing the extreme levels 1 and 5 (i.e. *very* relaxed, frustrated, sad and joyful) of all interactions from the two series of inner states Sad-joyful and Relaxed-frustrated. The participants were asked to write down the emotion underlying the interaction shown in the video in a free text form. Multiple answers were allowed if they thought several emotions fit. We recruited 26 native English speaking participants between ages 18-65 years for each experiment, out of which 19 (10 male, 9 female) completed the test for the Face-and-body, 25 (12 male, 13 female) for the Face-only, and 23 (10 male, 13 female) for the Body-only experiment. The participants were screened to not have taken part in any of the other experiments in the prestudy.

After the experiment was completed, we performed a sentiment analysis of the answers using the Linguistic Inquiry and Word Count (LIWC) [29], which is a text analysis software that categorizes words into linguistic, topical and psychological categories. LIWC gives affective ratings for *posemo*, *negemo*, *anxiety*, *anger* and *sadness*. To select the most salient interactions across embodiments, we concatenated the free-text answers for each interaction from the three experiments and calculated the LIWC scores. The two interactions with largest difference in *posemo* score were selected from the Sad-joyful series and the two interactions with largest difference in *anger* score were selected from the Relaxed-frustrated series. Table 2 shows the most frequently used words to describe the emotions shown in the selected interactions, and Table 3 shows the amount and percentage of unique words used in the descriptions. As can be seen in Table 2, all the descriptions were assigned to the intended quadrants in the circumplex model (Figure 1) except the *relaxed* Body-only video, which was

Table 4. Correlation between intensity levels and ratings (Real/Anim). Asterisk denotes significant correlation to the p<0.01 level.

Emotion	Face-and-body	Face-only	Body-only
Sad	-0.56* /-0.56*	-0.55* / -0.50*	-0.52* / -0.48*
Joyful	0.74* / 0.54*	0.81* / 0.67*	0.54* / 0.51*
Frustrated	0.56* / 0.24*	0.46* / 0.24*	0.32* / 0.34*
Relaxed	-0.44* / -0.33*	-0.55* / -0.12	-0.25* / -0.28*

generally interpreted to have more negative valence. The relatively higher percentage of unique text words in the answers suggest that there was more diversity in the reading of the sad and relaxed videos compared to the joyful and frustrated ones.

4.2 Evaluation

To evaluate how well the internal states and intensity levels are perceived for the three embodiments, we conducted a second set of experiments using the interactions selected in the prestudy. In three separate experiments, one for each embodiment, we performed evaluations of the videos of the real actor (condition Real) and the animated embodiment (condition Anim) along the two series of inner states Sad-joyful and Relaxed-frustrated. Each experiment contained 2 series of inner states, 2 selected interactions and 5 intensity levels, and thus 40 videos (20 in condition Real and 20 in condition Anim). We randomly mixed the Real and Anim videos into two sets of stimuli A and B, each containing 10 Real and 10 Anim videos. We then exposed half of the participants to the set A and the other half to set B, and asked them to rate the emotional reaction in the interactions along the four dimensions sadness, joy, relaxation and frustration on a five point scale ranging from 0-"not at all" to 4-"very much". Although the two video sets were different, all participants saw both animated and real examples of every embodiment/inner-state combination and the presentation order was randomized for each participant.

By adding the additional ratings along the non-intended dimensions we were able to investigate if the filtering effects not only change the intensity of the emotion being displayed along the intended dimension, but also if they change the perceived emotion towards the complementary axes. Forty native English speaking participants between ages 18-65 years were recruited for each experiment, out of which 35 (17 male, 18 female) completed the test for the Face-and-body, 30 (13 male, 17 female) for the Face-only, and 30 (16 male, 14 female) for the body only experiment. We ensured that no participants took part in any of the other experiments or in the prestudy.

4.3 Results

The average ratings along all emotional dimensions are presented in Figure 5. We performed pair-wise one-way Analysis of Variance (ANOVA) tests on the ratings of the videos in conditions Real and Anim. The ratings marked with an asterisk show significant changes (p<0.01) between conditions. As can be seen in the figure, the Face-and-body and Face-only embodiments show the most change between the real and animated conditions with significant drops in perceived joy, frustration and relaxation. The Body-only data set show no significant effects of condition along the intended emotions.

To quantify the degree to which the acted dimensions elicited corresponding gradual emotion judgments, we calculated correlations between acted intensity level and perceptual rating of corresponding emotions for each embodiment and visual condition. For the Sad-joyful series, correlation between the level (1-5) and the rating for *joy* and *sadness* were computed, and for the Relaxed-frustrated series, correlations between level and ratings for *relaxation* and *frustration* were computed. Correlations are displayed in Table 4. As expected, the emotions that go from high to low as the level increases (*sad* and *relaxed* respectively) yield negative correlations. The highest

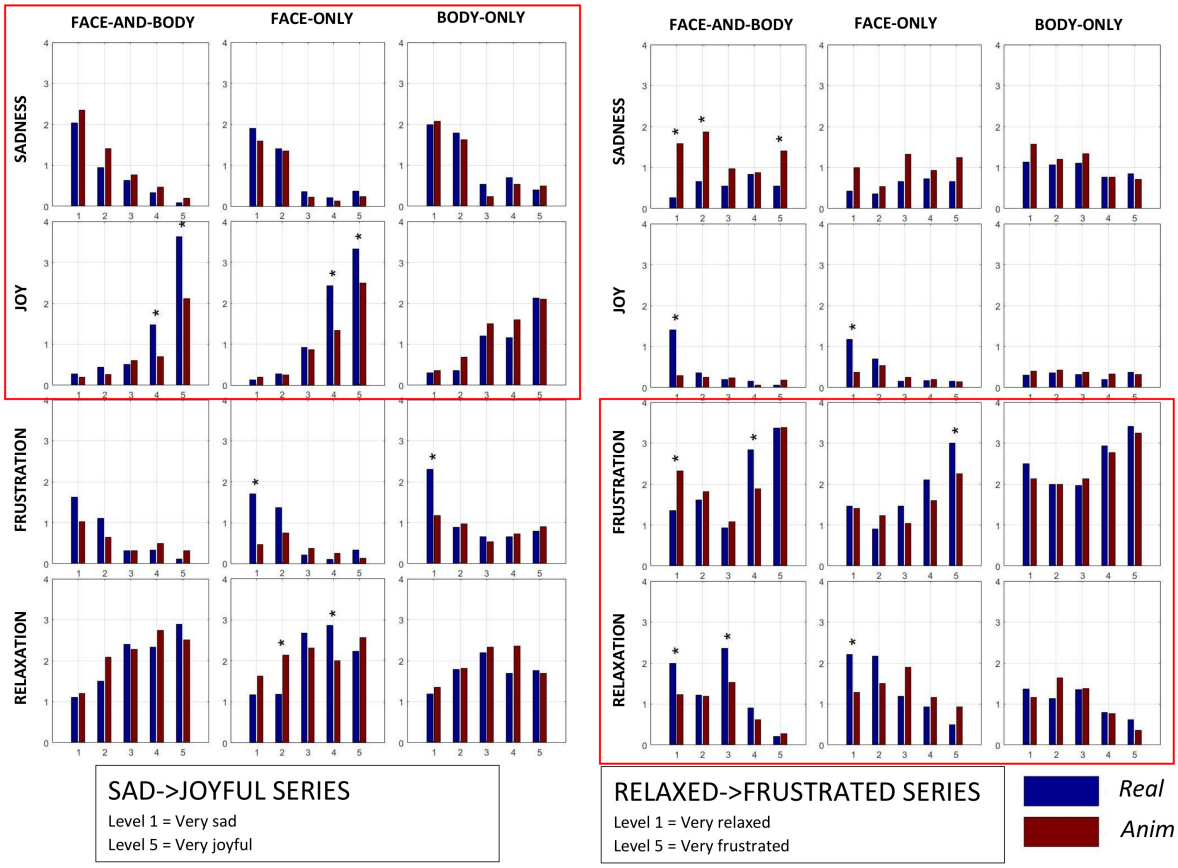

Fig. 5. Average ratings of of the real actor (blue) and and animated character (red) along the evaluated dimensions *Sadness*, *Joy*, *Frustration* and *Relaxation*. The x-axis shows the recorded intensity levels ranging from level 1 (very sad/relaxed) to level 5 (very joyful/frustrated). Left: Sad-joyful series, Right: Relaxed-frustrated. The red frames mark the ratings along the intended emotions. Asterisk denotes significance to the $p<0.01$ level.

correlation is 0.81 for *joy* in the Face-only (Real) condition. Over all, the Real condition yields higher or equal absolute correlations than Anim in most cases.

In order to further explore how the factors affected the participants' ability to perceive the internal states depicted, we performed a mixed design ANOVA with categorical predictor *Embodiment* (Body-only, Face-only, Face-and-body) and dependent variable *Condition* (Animated and Real versions of Sad-joyful and Relaxed-frustrated). Post-hoc analysis was conducted using Newman-Keuls tests to compare means and only values with $p < 0.05$ are reported as significant. In this analysis, we only consider the ratings of sadness for the two lowest intensities of the Sad-joyful stimuli; of joy for the two highest intensities, and similarly only considered relaxed and frustration ratings for the bottom and top Relaxed-frustrated intensity pairs respectively. We then averaged the resulting sad/joy and relaxed/frustration rating pairs to give one number for the intensity perceived for each depiction of internal state. The results are shown in Figure 6.

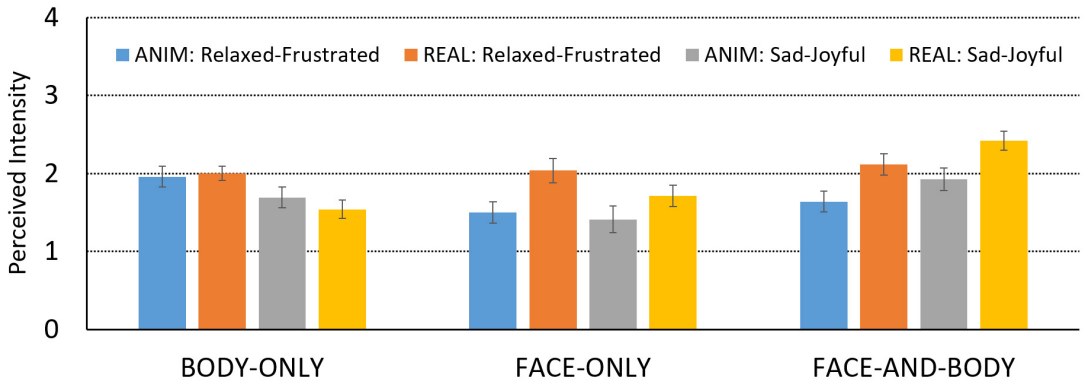

Fig. 6. Perceived intensity of real and animated inner states for the three embodiments.

We found a main effect of Embodiment ($F(2, 92) = 4.8160, p < .05$), where Face-and-body was rated as having a higher intensity overall than both Body-only and Face-only, which were rated equally. A main effect of Condition ($F(3, 276) = 6.1666, p < .0005$) showed that the perceived intensities of the real Relaxed-frustrated and Sad-joyful stimuli were higher than for their animated counterparts. Finally, a significant Embodiment*Condition interaction effect ($F(6, 276) = 4.6587, p < .0005$), shown in Figure 6, provides further insights into the differences between the factors. The real Face-and-body Sad-joyful condition was rated higher than all the others, and the animated Face-only Sad-joyful was rated the lowest.

5 DISCUSSION

In general, the evaluations show that both the mime and the agent animations were able to capture the intended emotions, and that the task of generating different levels of emotions was successful to a high degree. As indicated by the correlations in Table 4, both the mime and agent animations captured the Sad-joyful continuum better than the Relaxed-frustrated continuum. An explanation for this may come from the greater difficulties in generating the relaxed emotion. As seen in Table 2 from the prestudy and in Figure 5 from the evaluation, this emotion was especially affected when facial features were absent or affected by filtering.

We found a general loss of signal in the filtering from the mime motions to the agent animations as reflected in greater noise in the response curves (and lower correlations). The most dominant effects are found on the Face-and-body and Face-only embodiments, which show significant decreases in the perception of joy and frustration. We believe that a an important factor explaining this is the lack of detail in the capture of the eye-lids and the areas around the eyes. Especially in states of high activation, the widening of the eyes is important characteristic not captured in our setup. Another notable effect of filtering is found in the interactions showing the relaxed emotion for the Face-and-body data set; while the Real videos were rated as showing both joy and relaxation when the face was present, the Anim videos where rated as showing frustration and sadness for Face-and-body, with similar tendencies in the Face-only data set. An explanation of this may be that the facial movements are not captured and transfered as effectively to the robot character.

The ratings for the interactions intended to display sadness were not affected by filtering, showing no significant variation in ratings and similar correlations. Interestingly, the Body-only condition showed very similar results for the Real and Anim conditions. The only significant difference was an unintended, higher level of frustration in the “very sad” Body-only video that was not replicated in the Anim condition. The correlations are quite similar for the two conditions for Body-only. These results suggest that the impact of filtering on Body-only data

is much lower. While the inclusion of face data led to the highest overall correlations, it appears to either be harder to adequately capture or more impacted by filtering than the body data.

6 CONCLUSION AND FUTURE WORK

In this paper we have presented a dataset and a methodology for quantifying and improving the non-verbal expressive capability of different robot/agent embodiments. By recording highly expressive enactments of the same interaction at several positions along a continuous dimension corresponding to an internal state (e.g. a pair of contrastive emotions), mapping the recordings to a specific embodiment, and perceptually rating the results along the same dimension, we are able to get quantitative metrics on how well the given embodiment can be used to encode the dimension in question and how much the mapping from the real data to the embodiment impacts the effect – what we refer to as the filtering effect of the embodiment. A novel aspect of our recorded data set is the use of a professional mime actor in order to generate highly expressive motion. The perceptual rating study reported in this paper was carried out on a subset of the corpus containing affective states. Interesting results include that joy is more impacted by filtering than sadness when the embodiment includes a facial representation. Overall, the Body-only representation showed the least impact of filtering.

The dataset also contains non-affective communicative states (certainty, attentiveness and engagement). The next step will be to perform perceptual rating of these data in a similar manner. In future work we aim to investigate filtering effects in more detail by studying different embodiments and reducing or manipulating the degrees of freedom of the virtual robot. Questions such as *what is the effect of movable eye balls, two vs. three degrees of freedom in the neck, movable shoulders, etc.* may be addressed using the same experimental paradigm. Further on, we aim to carry out similar experiments with physical robot platforms, again using the same dataset and experiment design.

ACKNOWLEDGMENTS

The authors would like to thank the mime artist Alejandro Bonnet and Henrik Bäckbro for help in the preparation of the scripts and for providing practical and theoretical knowledge on mime and dramatic acting.

This work was funded by KTH/SRA ICT The Next Generation and Science foundation Ireland PI grant S.F.10/IN.1/13003.

REFERENCES

[1] Simon Alexanderson, Carol O'Sullivan, and Jonas Beskow. 2016. Robust online motion capture labeling of finger markers. In *Proceedings of the 9th International Conference on Motion in Games*. ACM, 7–13.

[2] Anthony P Atkinson, Winand H Dittrich, Andrew J Gemmell, and Andrew W Young. 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33, 6 (2004), 717–746.

[3] Eugenio Barba. 1991. Theatre Anthropolgoy. In *A Dictionary of Theatre Anthropology: The Secret Art of The Performer*, Eugenio Barba and Nicola Savarese (Eds.). Routledge, London.

[4] Jan Baumann, Björn Krüger, Arno Zinke, and Andreas Weber. 2011. Data-Driven Completion of Motion Capture Data.. In *VRIPHYS*. 111–118.

[5] Aryel Beck, Brett Stevens, Kim A Bard, and Lola Cañamero. 2012. Emotional body language displayed by artificial agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 1 (2012), 2.

[6] Timothy W Bickmore and Rosalind W Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 2 (2005), 293–327.

[7] Celso M de Melo, Peter Carnevale, and Jonathan Gratch. 2011. The effect of expression of anger and happiness in computer agents on negotiations with humans. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, 937–944.

[8] Jean Dorcy. 1961. *The Mime*. Robert Speller and Sons, Publishers, Inc. Translated by Robert Speeler, Jr. and Pierre de Fontnouvelle.

[9] Paul Ekman. 1992. Are there basic emotions? (1992).

[10] Wei-Wen Feng, Byung-Uck Kim, and Yizhou Yu. 2008. Real-time data driven deformation using kernel canonical correlation analysis. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 91.

[11] Michael J Gielniak and Andrea L Thomaz. 2012. Enhancing interaction through exaggerated motion synthesis. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 375–382.
[12] Markus Häring, Nikolaus Bee, and Elisabeth André. 2011. Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *Ro-Man, 2011 Ieee*. IEEE, 204–209.
[13] Jennifer Hyde, Elizabeth J Carter, Sara Kiesler, and Jessica K Hodgins. 2013. Perceptual effects of damped and exaggerated facial motion in animated characters. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 1–6.
[14] Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics* 14, 2 (1973), 201–211.
[15] Joan Lawson. 1957. *Mime: The Theory and Practice of Expressive Gesture With a Description of its Historical Development*. Sir Isaac Pitma and Sons Ltd., London. Drawings by Peter Revitt.
[16] Jacques Lecoq. 2009. *The moving body (Le corps poetique): Teaching creative theatre*. A&C Black.
[17] Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior* 25, 3 (2009), 695–710.
[18] Maya B Mathur and David B Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition* 146 (2016), 22–32.
[19] Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona Newell, and Carol O'Sullivan. 2008. Evaluating the emotional content of human motions on real and virtual characters. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 67–74.
[20] Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona N Newell, and Carol O'Sullivan. 2009. Investigating the role of body shape on the perception of emotion. *ACM Transactions on Applied Perception (TAP)* 6, 3 (2009), 14.
[21] Hanneke KM Meeren, Corné CRJ van Heijnsbergen, and Beatrice de Gelder. 2005. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America* 102, 45 (2005), 16518–16523.
[22] Catharine Oertel, José Lopes, Yu Yu, Kenneth A Funes Mora, Joakim Gustafson, Alan W Black, and Jean-Marc Odobez. 2016. Towards building an attentive artificial listener: on the perception of attentiveness in audio-visual feedback tokens. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 21–28.
[23] Jason Osipa. 2010. *Stop staring: facial modeling and animation done right*. John Wiley & Sons.
[24] Jonathan Posner, James A Russell, and Bradley S Peterson. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology* 17, 03 (2005), 715–734.
[25] Tiago Ribeiro and Ana Paiva. 2012. The illusion of robotic life: principles and practices of animation for robots. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*. IEEE, 383–390.
[26] Sichao Song and Seiji Yamada. 2017. Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2–11.
[27] Marc Swerts and Emiel Krahmer. 2005. Audiovisual prosody and feeling of knowing. *Journal of Memory and Language* 53, 1 (2005), 81–94.
[28] Leila Takayama, Doug Dooley, and Wendy Ju. 2011. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 69–76.
[29] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
[30] Frank Thomas, Ollie Johnston, and Frank. Thomas. 1995. *The illusion of life: Disney animation*. Hyperion New York.
[31] Yuqiong Wang, Gale Lucas, Peter Khooshabeh, Celso de Melo, and Jonathan Gratch. 2015. Effects of emotional expressions on persuasion. *Social Influence* 10, 4 (2015), 236–249.

Received July 2017