Designing Emotionally Expressive Robots: A Comparative Study on the Perception of Communication Modalities

Christiana Tsiourti
University of Geneva
Geneva, Switzerland[1]

Astrid Weiss
TU Wien
Vienna, Austria[2]

Katarzyna Wac
University of Geneva
Geneva, Switzerland[1]

Markus Vincze
TU Wien
Vienna, Austria[2]

ABSTRACT
Socially assistive agents, be it virtual avatars or robots, need to engage in social interactions with humans and express their internal emotional states, goals, and desires. In this work, we conducted a comparative study to investigate how humans perceive emotional cues expressed by humanoid robots through five communication modalities (face, head, body, voice, locomotion) and examined whether the degree of a robot's human-like embodiment affects this perception. In an online survey, we asked people to identify emotions communicated by Pepper - a highly human-like robot and Hobbit – a robot with abstract humanlike features. A qualitative and quantitative data analysis confirmed the expressive power of the face, but also demonstrated that body expressions or even simple head and locomotion movements could convey emotional information. These findings suggest that emotion recognition accuracy varies as a function of the modality, and a higher degree of anthropomorphism does not necessarily lead to a higher level of recognition accuracy. Our results further the understanding of how people respond to single communication modalities and have implications for designing recognizable multimodal expressions for robots.

Author Keywords
HAI; HRI; Emotional Expression; Multi-Modal Interaction; Facial Expression; Body Motion; Social Robots.

ACM Classification Keywords
I.2.9 Robotics: Commercial robots and applications; H.5 Information Interfaces and Presentation (e.g., HCI); H.5.2 User Interfaces: Evaluation/methodology.

INTRODUCTION
In the last few years, a growing interest has been seen in the development of socially assistive agents, that interact in a natural and fluid way with humans in their everyday life. Applications are plentiful: household assistants, companions for children and elderly, partners in industries, guides in public spaces, educational tutors at school and so on [21]. In these complex interaction scenarios, the recognizability of an agent's emotional expressions strongly impacts the resulting social interaction [25]. An agent with a detailed internal emotional model but with poorly designed emotional expressions may be limited in its ability to interact effectively with a human user. For example, imagine a robot nurse intended to empathize with hospital patients. Although this robot may have the capacity to recognize the affective state of a user correctly, if it is not able to adequately react, the patient might misinterpret the robot's behavior and feel uncomfortable.

Human emotional communication involves the complex interaction of multiple sensory channels, or "modalities" (i.e., visual, auditory) in complementary and supplementary combinations [35]. In social Human-Agent Interaction (HAI), several modalities have been identified to support the emotional expression of agents including, among others, facial expressions [6], emotional speech [7], body posture and motion [34], head posture [4], and locomotion [39]. There is substantial work suggesting that people can recognize and distinguish between unimodal and multimodal emotional expressions of avatars and robots, in some cases even with similar accuracy to that of human emotional expressions [30]. However, the vast majority of the studies on human emotional perception of agents focus exclusively on isolated unimodal expressions, or multimodal expressions as a whole, failing to gather information on responses to each single component of the multimodal combination. As a result, it is still not well understood how people perceive robot emotional expressions conveyed through one modality versus another. For example, do people base their perceptions of specific emotions on one modality more than another? Does the face, body or voice dominate in perceptions of certain emotional expressions, or are they all equally effective? Without a direct comparison, we still have a limited understanding of how certain modalities contribute to the overall recognizability of synthetic emotional expressions and whether there is a close association between certain emotional expressions and modalities. If there are indeed associations between emotions and modalities, then these

HAI '17, October 17–20, 2017, Bielefeld, Germany

ACM ISBN 978-1-4503-5113-3/17/10…$15.00
https://doi.org/10.1145/3125739.3125744

[1] {Christiana.tsiourti,Katarzyna.wac}@unige.ch

[2] {Astrid.weiss,Vincze}@acin.tuwien.ac.at

Figure 1: The robots Pepper (left) and Hobbit (right).

may provide the foundation for a coherent set of design guidelines for more reliable and believable robot emotional expressions that enhance the quality of social HAI in daily-life applications.

In this paper, we build upon previous work suggesting that the face, head, body, voice, and locomotion are appropriate emotional expression modalities for robots. We validate and compare the communication efficacy of these five modalities according to human perception and recognition accuracy of robot emotional expressions. Our study is focused on three central research questions:

(1) Can people accurately recognize the emotional states of a robot, based on restricted unimodal expressions, and without any contextual information?

(2) Does emotion recognition accuracy vary as a function of communication modality, robot embodiment, and emotion being communicated?

(3) Does dispositional empathy modulate people's ability to decode and accurately recognize robot emotional expressions?

We developed a database of 22 videos with the robots Pepper and Hobbit [20] (see Figure 1), expressing three basic emotions (happiness, sadness, surprise) through five restricted communication modalities (face, head, body, voice, and locomotion). Next, we set up an online survey and used our database to evaluate peoples' ability to recognize the emotional expressions of the robots. We analyzed how effectively each modality can communicate emotional information, by comparing people´s qualitative (perception of the expression) and quantitative (recognition accuracy) responses to the expressions. We also analyzed a dispositional measure, people's empathy - defined as an affective response stemming from the understanding of another's emotional state or what the other person is feeling or would be expected to feel in a given situation [16]. Evidence suggests that individuals with a low level of dispositional empathy achieve lower accuracy in decoding facial expressions of humans [15] as well as robots [32].

RELATED WORK

There is a range of modalities for machine emotional expression including facial expressions [6, 32], body movement [28, 31], head movement [4], voice prosody [8, 14] and locomotion [39]. The face has been the most commonly used modality, and a number of studies show that recognition rates for basic emotions of both avatars and robots are substantially high [6, 32]. More recently, there is an increasing interest in understanding the role played by whole-body expressions. A handful of researchers focused on the use of robotic body language to display emotions, with a significant emphasis on the display of emotions through dance (i.e., Laban movement [28]). McColl et al. [31] evaluated the use of emotional body language for a human-like social robot, utilizing body movement and posture descriptors identified in human emotion research. Humans were able to recognize the robot's emotional body language with high recognition rates. A study by Beck et al. [4] highlights the fact that head position is an important variable that can increase the correct identification of some emotions.

The modality of speech also conveys emotional information, through explicit linguistic messages (e.g., utterances), and implicit paralinguistic cues that reflect the way the words are spoken (e.g., pitch, intensity, rate, timing) [26]. A number of HAI studies ([1, 8, 14]) propose designs for vocal patterns and specific utterances and evaluate how well human subjects perceive the intended emotions. Their results highlight the ability and importance of using vocal prosody to convey robot emotions through speech. Finally, studies also suggest a relation between locomotion parameters (e.g., acceleration, curvature) and the attribution of emotions, even in robots with no human degrees of freedom [39]. Few researchers also investigated the perception of bimodal [12, 40] or multimodal robot emotional expressions [1] (facial expressions, head-arm gestures, and speech), suggesting that ultimately the combined perception may be the most effective way to convey a robot's affective states.

In summary, most research on the development of robot emotional expression has focused either on human recognition of single modalities in isolation or the recognition of multimodal expressions at large. Very few studies have been carried out to understand the importance of one modality with respect to another, and to the best of our knowledge, there is no comprehensive comparison that investigates how the face, head, body, voice, and locomotion, contribute to the perception of particular robot emotional states.

DATABASE OF UNIMODAL EMOTIONAL EXPRESSIONS

In order to answer our three research questions, we created a database with 22 short videos of two robots conveying restricted unimodal emotional expressions (happiness, sadness, surprise) through five modalities: face, head, body, voice, and locomotion.

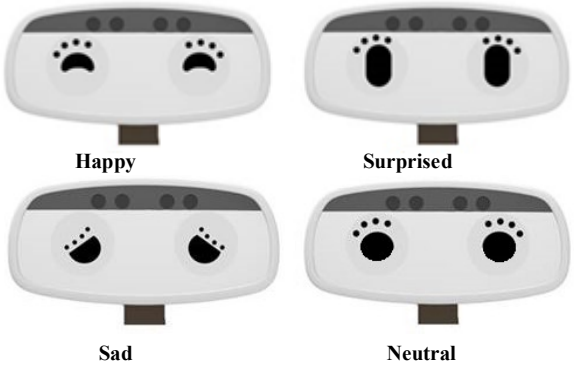

Figure 2: Facial expressions of the Hobbit robot.

Apparatus

We created our database using two robots (see Figure 1) with different embodiments and different levels of human-likeness, as detailed below. With this comparative approach, we were able to investigate how much human-likeness is necessary for people to ascribe social and emotional intelligence to a robot.

Pepper: A highly human-like robot with 20 degrees of freedom (DoF): 6 motors in each arm, 2 in the neck, 2 in the hips, 1 for the knee, and 3 for the wheels in an omnidirectional drive configuration. Pepper can talk and play sounds through its loud speakers and has several LEDs to convey information. Pepper has a static face.

Hobbit: A humanoid robot with 4 DoF: 2 motors for the wheels in a differential drive configuration, and two motors in the head for pan and tilt. Hobbit can talk and play sounds through its loudspeakers and has a display on its head to present facial expressions.

Expressed Emotions

We modeled three of the basic emotions [17]: happiness, sadness, and surprise. These three emotions were selected due to their previously demonstrated recognition through one or more of the modalities of interest in this study: face [32], head [4], body [31], voice [6] and locomotion [39]. Choosing emotions that are widely investigated allowed us to have reliable sources for movement and posture descriptors to design our expressions and permitted us to compare our results with previous studies on emotion perception and recognition. We also chose emotions which are mapped onto different quadrants of the valence-arousal space [38]; happiness has a positive valence, sadness has negative valence and surprise as a special case can have any valence from positive to negative. Finally, we chose "social emotions," namely emotions that serve a social and interpersonal function in interactions [23], since these are especially useful for socially assistive robots.

Design of unimodal emotional expressions

We designed analogous expressions for the two robotic platforms, despite the different number of DoF and the dynamics of movement of the two embodiments. An

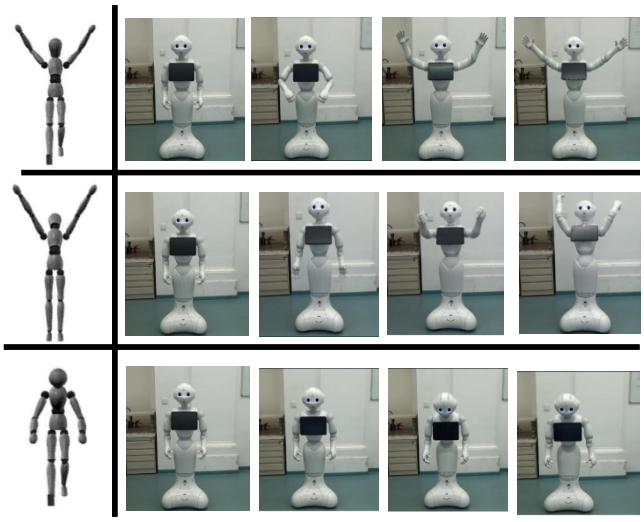

Figure 3: Body postures for happiness, surprise , and sadness, (top to bottom) [13] and implementation on Pepper.

expression started with the robot in a neutral position, next a transition to the actual emotion was shown, and then the robot returned to the neutral position. All the videos were recorded such that the viewer could see the profile of the robot and were about 10 seconds long. Even though we are aware that emotion perception is context dependent [10], in this study we did not want knowledge of the eliciting context to have an effect on the perception of the expressions. Therefore in the videos, the robot was acting in an emotionally neutral context. A number of other studies investigating the recognition of robot emotional expressions do not involve any context (e.g., [32]).

Facial expressions: Facial expressions were designed only for Hobbit since Pepper has a static face. The animations (happiness, sadness, surprise and neutral) were designed and validated in an expert evaluation study, within the EU project which developed the robot (see Figure 2).

Head and body expressions: Head expressions were designed for both robots. Body expressions were designed only for Pepper since Hobbit's DOF do not support body motion. All expressions were modeled after the way humans use their head, torso, and arms to express emotions. Sources for our design were Coulson [13], who has grounded basic emotions into specific features that describe the configuration of human body posture and Kleinsmith et al. [27], who describe motion dynamics (velocity and amplitude) linked to particular emotions. To create convincing and believable expressions, we started from the creation of expressive key poses [42] for each emotion (happiness, sadness, surprise). Each joint of the robots was carefully positioned to match the original pose as described by [13]. Once a key pose was implemented, the velocity and amplitude of body parts were modeled according to [27], creating the animated expressions (see Table 1). Happy movements were large, somewhat fast and relatively

fluid, while the sad movements were smaller, less energetic, and slower.

Vocal expressions: For the vocal expressions, we used brief non-linguistic vocalizations which lend themselves to be expressed with different emotional meanings and are language and culture independent [36]. The expressions were identical for both robots and were synthesized using a commercial Text-to-Speech (TTS) engine [11]: laughter vocalization (happiness), negative "oh" (sadness), and a sudden, short intake of breath (surprise).

Locomotion expressions: For the locomotion expressions, we identified features that could be modulated equally in both robots and composed them into expressions. Previous studies have shown that changes in speed and direction have an effect on the perception of robot affect [39]. Combining these two locomotion features, resulted in three expressions, which were designed for both robots: forward slow motion, forward fast motion, backward fast motion. We did not define a direct mapping between these expressions and emotions. Instead, we wanted to analyze the relationship between the features speed and direction and perceived emotion.

EVALUATION OF EMOTIONAL EXPRESSIONS

Study Design and Procedure

The study was set up as an online survey, which was distributed via various academic mailing lists. As an incentive for participation, three randomly selected participants won a 50€ Amazon gift card. Interested participants were directed to a website that gave instructions about the study. After giving consent, participants started the survey. The total time for completion was less than twenty minutes, and participants were able to save their results in between and take a break.

We used a within-subjects repeated measures design; all the participants watched all the face, head and locomotion expressions of Hobbit; and all the body, head, and locomotion expressions for Pepper. In the interest of keeping the survey short, and since the vocal expressions were identical for both robots, each participant was randomly assigned to watch either Hobbit's and Pepper's vocal expressions. For each participant, the sequence of the robots and the sequence of the emotional expressions were randomized to prevent carry-over effects from one robot or expression to the other. The same set of questions had to be completed for every video:

- **Question 1:** Participants were instructed to imagine that the robot is reacting to something that it is observing, and describe the robot's reaction, using their own words, in an open-ended question.
- **Question 2:** Participants were asked to select the most appropriate emotional label from a set of seven possible responses (Sadness, Happiness, Anger, Surprise, Neutral, I don't Know and Other). The labels Anger and Other were included in the list to uncover patterns of confusability and to reduce inflated accuracy rates due to the forced choice [22].

Emotion	Motion dynamics	Head movement	Body movement	
			Upper body	*Arms*
Sadness	Small, slow movements	Forward head bend	Forward chest bent	Arms at side of trunk
Happiness	Large, fast movements	Straight head	Straight trunk	Vertical and lateral extension
Surprise	Large, fast movements	Backward head bent	Backward chest bent	Vertical extension
Neutral	No movement	Straight head	Straight trunk	Arms at side of trunk

Table 1: Head and body expression descriptors for emotions.

After going through all the videos, participants were asked to provide basic socio-demographic information (age, gender, profession and previous experience with the robots). Next, participants were asked to fill in the Toronto Empathy Questionnaire [41], a 16-item self-assessment questionnaire assessing dispositional empathy.

DATA ANALYSIS AND RESULTS

Participants

A total of 170 people, from 30 countries, completed the online survey. All the participants watched the face, head, body and locomotion expression videos for both robots. 84 participants watched Pepper's vocal expression videos and 86 participants watched Hobbit's vocal expression videos. The age of the participants ranged between 18 and 69. 75 participants (44%) were female, 95 participants (56%) male. 68 participants (40%) reported having seen the Pepper robot in action before, and 21(12%) participants reported having seen the Hobbit robot.

Qualitative Results

To analyze participants' responses to the open-ended question (Question 1) we used thematic coding [5]. As the data was homogenous and no comparative quantification was intended, the coding was performed by one coder who categorized the answers. This categorization led to a thematic map of four main themes: **1)** internal emotional state of the robot, **2)** emotional behavior of the robot, **3)** non-emotional behavior of the robot, **4)** interactional experiences of the robot. In line with Heider and Simmel's well-known study on the attributional processes in perception [24], participants interpreted the robots as animated beings and often attributed the origin of their expressions to internal motives or needs that emerged as a consequence of interactional experiences, with human (e.g., the participant) or non-human actors (e.g., other robots) or objects of the environment. The interactions were also described in terms of valence (i.e., something positive/negative happened to the robot) or arousal (i.e., something sudden occurred in the environment).

Quantitative Results

We analyzed whether the emotions expressed by the robots were accurately recognized based on a closed question (Question 2), in which participants had to select the robot's emotion from a list of suggested terms. We worked with nominal data on independent groups: the modalities, emotions and robot platforms. To calculate the recognition accuracy of each emotional expression, the chosen answers for each video were compared against the ideal distribution by means of Chi-square tests. In other words, if 170 people had to identify the right emotion out of a list of seven different terms (Happiness, Sadness, Surprise, Anger, Neutral, I don't know, Other) and the robot was expressing "sadness", the ideal distribution would be 0, 170, 0, 0,0,0,0. We performed this analysis separately for each robot. Besides comparing each robot's expressions with the ideal distribution, the expressions of the two robots were also compared against each other to analyze the effects of the different robot embodiments. The detailed Chi-square results are presented in Table 2. Figures in bold show conditions with nonsignificant results ($p>0.01$), meaning that the assessment of the emotions did not differ significantly from the expected correct choice. We also analyzed the dominant emotion per expression, defined as the answer that was most frequently selected for a given video. Overall, in 69% of the videos, the dominant emotion was the correct one (87% for Hobbit and 50% for Pepper). The detailed results are presented in Table 3.

Modality	Group comparison	Happiness Video	Sadness Video	Surprise Video
Face	Hobbit vs. Ideal	**$\chi^2(6, N=170) = 15.30, p = .02$**	**$\chi^2(6, N=170) = 5.29, p = .51$**	**$\chi^2(6, N=170) = 1.86, p = .93$**
Head	Hobbit vs. Ideal	*N/A*	**$\chi^2(6, N=170) = 14.12, p = .03$**	$\chi^2(6, N=170) = 104.05, p = .00$
	Pepper vs. Ideal	*N/A*	$\chi^2(6, N=170) = 86.12, p = .00$	$\chi^2(6, N=170) = 62.41, p = .00$
	Hobbit vs. Pepper	*N/A*	**$\chi^2(6, N=170) = 0.00, p = 154.80$**	**$\chi^2(6, N=170) = 0.00, p = 112.35$**
Voice	Hobbit vs. Ideal	**$\chi^2(6, N=86) = 12.66, p = .05$**	$\chi^2(6, N=86) = 39.12, p = .00$	**$\chi^2(6, N=86) = 11.91, p = .06$**
	Pepper vs. Ideal	**$\chi^2(6, N=84) = 10.71, p = .10$**	$\chi^2(6, N=84) = 48.76, p = .00$	$\chi^2(6, N=84) = 20.01, p = .00$
	Hobbit vs. Pepper	**$\chi^2(6, N=84) = 0.14, p = 9.55$**	**$\chi^2(6, N=84/86) = 0.01, p = 18.34$**	**$\chi^2(6, N=84/86) = 0.03, p = 13.60$**
Body	Pepper vs. Ideal	$\chi^2(6, N=170) = 48.71, p = .00$	$\chi^2(6, N=170) = 144.99, p = .00$	$\chi^2(6, N=170) = 139.51, p = .00$

Table 2: Cumulated findings from the closed question on the emotional expression of the two robots (N=170). Figures in bold show conditions with non-significant results.

Comparison of modalities and emotions

Face modality emotion recognition: Participants were asked to recognize facial expressions of happiness, sadness, and surprise for Hobbit. The Chi-Square test for all the videos was not significant upon comparing the ideal distribution against the chosen emotions. In other words, there was no significant difference ($p>0.01$) in the assessment of the robot´s facial expression compared to the ideal distribution (see Table 2). These results show that our participants could accurately identify the three emotions conveyed by Hobbit through the face modality.

Head modality emotion recognition: Participants were asked to recognize head motion expressions of sadness and surprise, for both Hobbit and Pepper. For Hobbit's sadness expression, no significant difference could be found for the participant's answers and the ideal distribution, meaning that the participants accurately recognized the robot's emotion. However, this was not the case for Pepper's sadness head motion. As observed from the absolute numbers of emotion selection (see Table 3), only 49 of the 170 participants (29%) assigned the correct emotion to the sadness video and 51 participants (30%) selected the term "other." Among those, 28 (16%) stated that the robot's expression conveyed "embarrassment." Also, 39 (23%) participants selected the term "neutral." According to the qualitative results, those participants thought that the robot was merely turning its head down to look at or track an object, without expressing any emotional reaction. For the surprise head motion, the Chi-square test for both robots was significant, meaning that, the participants' assessments differed significantly from the expected correct choice (see Table 2). Nevertheless, in absolute numbers, the dominant emotion selected for Pepper was the correct one.

Voice modality emotion recognition: Participants were asked to recognize brief non-linguistic vocalizations of happiness, sadness, and surprise for either Hobbit (N=86) or Pepper (N=84). For both robots, the Chi-square test for happiness and surprise was not significant. In other words, the participants could accurately identify the happy and surprised vocal expressions. Participants could not identify the sadness vocal expression in either of the two robots. Still, as observed from the descriptive data, for Hobbit the dominant emotion was the correct one.

Body modality emotion recognition: Participants were asked to recognize body motion expressions of happiness, sadness, and surprise for Pepper. The Chi-Square test for the three emotions was significant upon comparing the ideal distribution against the chosen emotions, meaning that the participants' assessments differed significantly from the expected correct choices. Upon taking a detailed look at the descriptive data, it becomes evident that for the case of happiness, even though the emotion assessment was biased by mentions other than the correct emotion, the dominant emotion was the correct one. In the case of sadness, 91 out of the 170 participants (54%) associated the robot's expression with a "neutral" emotional state. For surprise,

only 16 participants (9%) assigned the correct emotion to the Pepper's expression and 45 participants (26%) said that the expression conveyed "anger."

Perception of locomotion modality: Participants were asked to assess three expressions for both robots (forward slow, forward fast and backward fast motion). The results suggest a relation between the locomotion parameter of direction and the attribution of emotions, while the type of embodiment had no effect. According to the absolute numbers of emotion selection, the dominant emotion for the "fast forward" and "slow forward" videos was "neutral" for both robots. In other words, motion in a forward direction was associated with a "neutral" emotional state, independently of speed and embodiment. The dominant emotion for the "fast backward" motion was "fear" for both robots.

Comparison of robot embodiments

Besides comparing the recognition of the robots' expressions against the ideal distribution, we also compared the recognition scores of the two robots against each other for the head and voice modalities (i.e., recognition of Pepper's sadness head motion vs. Hobbit's sadness head motion). Although there is some variation in the absolute numbers of emotion selection (see Table 3), the Chi-square tests were not significant neither for the head nor the voice modality (see Table 2). In other words, we found no significant difference in the correct recognition scores between the two platforms.

Toronto Empathy Questionnaire

The Toronto Empathy Questionnaire was calculated by reversing the inverted items and computing the summative score over all 16 items. The total score of the questionnaire can range from 0 (no empathy at all) to 64 (total empathy) [41]. Our analysis resulted in a mean value of 46.55 (SD = 6.76). There were no significant differences between female (mean= 47.97, SD=7.35) and male (mean = 45.44, SD = 6.08) participants. Our results for female participants were within the range given in the source of the questionnaire [41] (between 44 and 49 points); our results for male participants are slightly higher than the source (between 43 and 45 points), indicating that our male participants have a slightly higher dispositional empathy than average. We wanted to test if the participant's dispositional empathy modulated the assignment of the correct emotion in the recognition task. Namely, we were interested to see if the results achieved with the Chi-square tests would be improved if we excluded all ratings from participants with an empathy disposition below the Toronto Empathy Questionnaire threshold (43). As expected, removing the assessments of the 37 participants with below-threshold empathy scores improved all our Chi-square test results (as compared to the Chi-square values of Table 2). The Chi-square values for the remaining sample of 78% (133 participants) became slightly lower, and some significances went up, both indicating the similarity between the ideal

Assessed emotion	Facial expression N=170		Head motion N=170		Body motion N=170		Vocal expression N=86 /N = 84	
	Hobbit	Pepper	Hobbit	Pepper	Hobbit	Pepper	Hobbit	Pepper
Happiness Videos								
Happiness	**119**	*N/A*	*N/A*	*N/A*	*N/A*	**79**	**53**	**54**
Sadness	22	*N/A*	*N/A*	*N/A*	*N/A*	0	0	0
Surprise	2	*N/A*	*N/A*	*N/A*	*N/A*	19	1	2
Anger	0	*N/A*	*N/A*	*N/A*	*N/A*	2	1	0
Neutral	8	*N/A*	*N/A*	*N/A*	*N/A*	40	11	10
I don't know	6	*N/A*	*N/A*	*N/A*	*N/A*	13	8	3
Other	13	*N/A*	*N/A*	*N/A*	*N/A*	17	12	15
Sadness Videos								
Happiness	4	*N/A*	1	8	*N/A*	4	3	0
Sadness	**140**	*N/A*	**121**	49	*N/A*	13	**28**	20
Surprise	1	*N/A*	0	2	*N/A*	2	4	1
Anger	0	*N/A*	0	0	*N/A*	0	3	4
Neutral	4	*N/A*	13	39	*N/A*	**91**	18	15
I don't know	12	*N/A*	2	21	*N/A*	7	8	16
Other	9	*N/A*	32	**51**	*N/A*	53	22	**28**
Surprise Videos								
Happiness	2	*N/A*	8	25	*N/A*	19	0	1
Sadness	0	*N/A*	0	0	*N/A*	1	3	0
Surprise	**156**	*N/A*	37	**67**	*N/A*	16	**54**	**43**
Anger	0	*N/A*	2	1	*N/A*	**45**	0	0
Neutral	2	*N/A*	**78**	29	*N/A*	35	14	10
I don't know	2	*N/A*	19	27	*N/A*	29	9	22
Other	8	*N/A*	26	21	*N/A*	25	6	8

Table 3: Assessment of closed-question emotional terms. Absolute numbers of emotion selection for each modality.

distribution and the actual emotion assessment. In one case, for the voice modality, a statistically significant result could be achieved: Pepper´s surprise vocal expression was rated not significantly different from the ideal distribution (χ^2 (6, N = 133) =13.93, p = .03). In other words, participants with an average or high empathy disposition accurately identified this vocal expression.

DISCUSSION

For one of the first times in literature, we systematically compared peoples' perception of robotic emotional expressions conveyed through five different modalities

(face, head, body, voice, and locomotion) across a range of social emotions (happiness, sadness, surprise). Our approach aimed at assessing robot emotional expressions in terms of "*Can people accurately recognize a robot's emotions based on observed restricted unimodal expressions?*" Participants viewed context-free stimuli and were not provided with any information whatsoever about the interactional context. For each emotional expression, we identified the most frequently selected emotion (dominant emotion) and computed a recognition accuracy score, corresponding to the proportion of correct choices (i.e., the emotion intended by the robot). The accuracy scores were measured against the ideal distribution - a very strict criterion since it suggests "complete" recognition accuracy (impossible to obtain, given the individual variation in the perception and interpretation of emotional expressions).

Our findings provide several contributions towards the design of more reliable and believable robot emotional expressions. Although participants faced a highly challenging task, the quantitative analysis showed that the majority of participants interpret robot expressions in an accurate way. Moreover, the qualitative thematic analysis revealed that in addition to assigning an emotional interpretation to the robot's expressions, people tend to relate the robot's emotional behavior with interactive experiences. This finding underlines that situational information is especially influential for the attribution of emotion [10] and suggests that when people attribute emotional states to a robot, they create an "imaginary" interaction-context in their mind, to make sense of the robotic expressions they observe [24].

Variations in Recognition Accuracy as a Function of Communication Modality and Robot Embodiment

The recognition accuracy rates reported in this study exceeded the chance level, showing that both robots were capable of conveying the intended emotions using the modalities of the face, head, and voice. In addition, locomotion appears to be a suitable "add-on" modality to enhance the expressiveness of emotional behaviors, especially in terms of valence (positive/negative).

Anthropomorphism, reflected in a robot's appearance, behavior, and communication (e.g. modality), plays an important role in the design of socially intelligent robots [19]. In this study, we compared the same emotional behaviors and modalities across two platforms with very different levels of human like appearance. We found that a higher degree of human-likeness does not necessarily lead to a higher level of emotion recognition accuracy. Simple behaviors of the 4 DoF Hobbit robot led to similar recognition rates as the highly human-like 20 DoF Pepper robot. The results of our comparison suggest that emotional cues are transferable between embodiments with different expressive features and it is possible to convey emotional information without a fully expressive human-like body, with movable torso, legs or arms.

The recognition accuracy rates for the facial expression modality were the highest, compared to the other modalities. In addition to the high accuracy rates, the results were statistically significant for all three emotions. Remarkably, the Hobbit's surprise facial expression achieved the highest recognition accuracy rate (92%), despite the fact that robotic surprise facial expressions are typically less accurately recognized, often as a result of being confused with other emotions (i.e., fear) [32]. These results underscore the importance of facial communication of emotion and are consistent with previous findings regarding the utility of the face for conveying the internal affective states of robots (e.g.,[9, 32]). On the other hand, our findings add to recent work showing that the face is not uniformly the sole modality of emotional communication for robots (e.g., [4],[31]). Sadness and surprise were recognized with sufficient accuracy through restricted head motion expressions. These findings are in line with previous research by Beck et al. [4].

We also investigated the perception of brief non-linguistic vocalizations, and we have found that they seem to be good examples of accurately recognized expressions for discrete robot emotions. Many studies have examined the vocal characteristics of speech in hopes of defining a vocal signature for each basic emotion [37]. However, the fact that existing TTS engines are not yet prepared for efficiently synthesizing natural vocal acoustics makes it challenging for robots to convey the true meaning of a target emotion to human users. The results of our simple evaluation provide insight into the design of new experiments embedding non-linguistic vocalizations in a conversational setting to evaluate whether they can enhance the overall emotional expression.

Variations in Recognition Accuracy as a Function of Emotion

As expected from earlier reports in the literature, our findings show that certain emotions are better recognized than others, partly depending on the communication modality considered [2]. We found that the recognition scores for happiness were higher than those of other emotions, across all modalities. The most apparent discrepancies between the dominant answers and the expected correct choice were observed for Pepper's sadness expressions, across the head, body and voice modalities. A possible explanation might be related to the fact that Pepper's face configuration (i.e. the arrangement of the eyes and mouth) reflects happy facial expression. In the body and voice videos, it is possible that the participants experienced emotionally incongruent face-body and face-voice multimodal expressions, and the "happy" face was more salient than the other modalities. This is particularly striking in the voice condition, where the non-linguistic vocalizations of the robots were exactly the same, and yet the recognition rates for Pepper are significantly lower than those of Hobbit. As regards the less well-recognized sadness body expression for Pepper, the dominant emotion

was "neutral" (i.e., absence of emotion). The participants' qualitative responses to this expression suggest that the emotionality of the expression was not clearly perceived. Specifically, the robot's expression was interpreted as a sign of respect or submission without a clear emotional connotation (e.g., "*It seems that the robot is bowing to someone with respect*," "*The robot is expressing a submissive agreement*").

The qualitative and quantitative data indicate that the low recognition rate of Pepper's surprise body expression may be due to the ambivalent nature of this emotion (one can be positively and negatively surprised) coupled with the absence of an explicit emotion eliciting-event. The robot's reaction received both positive and negative connotations with the dominant emotion being "anger" (26%) followed by "neutral" (21%) – chosen by participants who had trouble interpreting the robot's behavior (e.g., "*the robot moved its arms, but I can't see any cause for that reaction*"). A smaller number of participants chose happiness (11%). A possible explanation for this concerns the fact that Pepper's surprise and happiness body expressions were designed based on two perceptually similar postures, as defined by Coulson [13] (see Figure 3). In a context-free scenario, participants may have assumed happiness as a default emotion for expressions with the arms lifted and extended to the sides. In fact, an analogous confusion pattern between joy and surprise was also observed in the original study by Coulson [13]. In general, confusion patterns were not seen in the face and voice modalities, most likely due to the significant differences between happy and surprised facial and vocal expressions.

Personality, Culture and Emotion Recognition

As expected from earlier reports in the literature [32], we found a correlation between the dispositional empathy score and the ability to perceive and accurately recognize emotional expressions of robots. Our findings further underline the fact that individual variations in emotion recognition skills can lead to a possible impairment in the interpretation of robot emotional expressions and consequently bias the quality of the social interaction. Further analysis of the results for participants scoring low in low dispositional empathy is likely to reveal interesting trends about the social perception of robots by individuals with empathy deficits disorders (e.g., Autism, Asperger).

There are numerous theories about the role of culture in human-human [18] and human-robot [29] affective interaction. Despite the large- cultural variability of our participants, it was not possible to extract any trends suggesting whether people' from different cultural backgrounds differ in their perception of robot expressions.

CONCLUSION AND FUTURE WORK

The acceptance of socially assistive agents, be it virtually-embodied avatars or physically-embodied robots, depends on their ability to convey information on their emotional state and intentions to their users. Past work has shown that the face, head, body, voice, and locomotion, are appropriate modalities to convey the internal emotional states, goals, and desires of agents. However, few studies have directly compared how people perceive emotional cues communicated through one modality versus another. Without a direct comparison, it remains unclear how each modality contributes to the overall recognizability of synthetic emotional expressions. In this paper, we presented a comparative study that investigates how people perceive five communication modalities (face, head, body, voice, locomotion) and whether a robot's anthropomorphic embodiment affects this perception. We developed a database of context-free unimodal emotional expressions for two robots with varying degrees of human-likeness (Pepper and Hobbit) and asked people to identify the communicated emotions in an online survey. A mixed-methods approach, combining statistical analysis of recognition scores and thematic analysis of qualitative data showed that emotion recognition accuracy varies as a function of the modality, the emotion being communicated, and the embodiment. Recognition accuracy scores were highest when participants viewed emotions conveyed via the face; however significant recognition accuracy scores were also achieved for the head, voice and body modalities, suggesting that the face is not uniformly the sole communication modality able to convey robot emotional information. We also found that a higher degree of human-likeness does not necessarily lead to a higher level of recognition accuracy, suggesting that unimodal emotional cues are transferable between embodiments with different expressive features.

This study is a first step towards the design of more reliable and believable multimodal expressions for socially assistive robots. There is evidence that when it comes to assessing emotional expressions regarding readability, videos can be used to provide fast and comparable assessments [4, 32]. Nevertheless, videos eliminate factors such as general properties, size, and operating noise of robots, which might influence people's perception [33]. In addition, recent research [3] indicates that a robot's physical presence affects human judgments of robots as social partners. Our future work will include a more ecologically valid paradigm where participants engage in face-to-face interaction with the robots. This setup will likely enhance the impact of the emotional expressions, potentially highlighting differences in the efficacy of the compared modalities. We also intend to study how different communicative modalities operate in synchronized combinations and assess whether and how additional modalities and contextual information influence people's perception of robots as emotionally intelligent social partners.

ACKNOWLEDGMENT

This work is supported by the Swiss National Science Foundation under grant P1GEP2_168609, and the EU projects GrowMeUp (H2020-643647), CoME (AAL-2014-7-127) and ANIMATE (AAL-2013-6-071).

REFERENCES

1. Aly, A. and Tapus, A. 2015. Multimodal Adapted Robot Behavior Synthesis within a Narrative Human-Robot Interaction. Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on (2015), 2986–2993.
2. App, B. et al. 2011. Nonverbal channel use in communication of emotion: How may depend on why. Emotion. 11, 3 (Jun. 2011), 603–617.
3. Bainbridge, W.A. et al. 2011. The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. International Journal of Social Robotics. 3, 1 (Jan. 2011), 41–52.
4. Beck, A. et al. 2010. Towards an Affect Space for robots to display emotional body language. 19th IEEE International Symposium on Robot and Human Interactive Communication Principe. (Sep. 2010), 464–469.
5. Boyatzis, R.E. 1998. Transforming qualitative information: Thematic analysis and code development. sage.
6. Breazeal, C. 2003. Emotion and sociable humanoid robots. International Journal of Human-Computer Studies. 59, 1-2 (Jul. 2003), 119–155.
7. Breazeal, C. 2001. Emotive qualities in robot speech. Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. (2001), 1388–1394.
8. Breazeal, C. and Aryananda, L. 2000. Recognition of Affective Communicative Intent in Robot-Directed Speech. Autonomous Robots. 12, 1 (2000), 83–104.
9. Breazeal, C.L. 2002. Designing sociable robots. MIT Press.
10. Carroll, J.M. and Russell, J.A. 1996. Do facial expressions signal specific emotions? Judging emotion from the face in context. Journal of Personality and Social Psychology. (1996).
11. CereProc Text-to-Speech Engine: https://www.cereproc.com/.
12. Costa, S. et al. 2013. Facial Expressions and Gestures to Convey Emotions with a Humanoid Robot. Social Robotics. Springer International Publishing. 542–551.
13. Coulson, M. 2004. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. Journal of Nonverbal Behavior. 28, 2 (2004), 117–139.
14. Crumpton, J. and Bethel, C.L. 2016. A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech. International Journal of Social Robotics. 8, 2 (Apr. 2016), 271–285.
15. Dapretto, M. et al. 2006. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. Nature Neuroscience. 9, 1 (Jan. 2006), 28–30.
16. Eisenberg, N. et al. 2014. Empathy-related responding and cognition: A “chicken and the egg” dilemma. Handbook of Moral Behavior and Development. (2014).
17. Ekman, P. 1970. Universal facial expressions of emotion. California Mental Health Research Digest. 8, 4 (1970), 151–158.
18. Elfenbein, H.A. and Ambady, N. 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychological bulletin. 128, 2 (Mar. 2002), 203–35.
19. Fink, J. 2012. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. Springer, Berlin, Heidelberg. 199–208.
20. Fischinger, D. et al. 2013. Hobbit-the mutual care robot. Workshop on Assistance and Service Robotics in a Human Environment Workshop in conjunction with IEEE/RSJ International Conference on Intelligent Robots and Systems (2013).
21. Fong, T. et al. 2003. A survey of socially interactive robots. Robotics and Autonomous Systems. 42, 3-4 (Mar. 2003), 143–166.
22. Frank, M.G. and Stennett, J. 2001. The forced-choice paradigm and the perception of facial expressions of emotion. Journal of personality and social psychology. 80, 1 (Jan. 2001), 75–85.
23. Hareli, S. and Parkinson, B. 2008. What’s Social About Social Emotions? Journal for the Theory of Social Behaviour. 38, 2 (Jun. 2008), 131–156.
24. Heider, F. and Simmel, M. 1944. An Experimental Study of Apparent Behavior. The American Journal of Psychology. 57, 2 (Apr. 1944), 243.
25. Hudlicka, E. 2003. To feel or not to feel: The role of affect in human–computer interaction. International Journal of Human-Computer Studies. 59, 1-2 (Jul. 2003), 1–32.
26. Juslin, P.N. and Scherer, K.R. 2005. Vocal Expression of Affect. The new Handbook of Methods in Nonverbal Behavior Research. Oxford University Press. 65–135.
27. Kleinsmith, A. and Bianchi-Berthouze, N. 2013. Affective Body Expression Perception and Recognition: A Survey. IEEE Transactions on Affective Computing. 4, 1 (Jan. 2013), 15–33.
28. Knight, H. and Simmons, R. 2016. Laban head-motions convey robot state: A call for robot body language. 2016 IEEE International Conference on Robotics and Automation (ICRA) (May 2016), 2881–2888.

29. Li, D. et al. 2010. A Cross-cultural Study: Effect of Robot Appearance and Task. International Journal of Social Robotics. 2, 2 (Jun. 2010), 175–186.
30. Mavridis, N. 2015. A review of verbal and non-verbal human–robot interactive communication. Robotics and Autonomous Systems. 63, (Jan. 2015), 22–35.
31. McColl, D. and Nejat, G. 2014. Recognizing Emotional Body Language Displayed by a Human-like Social Robot. International Journal of Social Robotics. 6, 2 (Apr. 2014), 261–280.
32. Mirnig, N. et al. 2014. Can You Read My Face? International Journal of Social Robotics. 7, 1 (Nov. 2014), 63–76.
33. Moore, D. et al. 2017. Making Noise Intentional. Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17 (New York, New York, USA, 2017), 12–21.
34. Mumm, J. and Mutlu, B. 2011. Human-robot proxemics. Proceedings of the 6th international conference on Human-robot interaction - HRI '11 (New York, New York, USA, Mar. 2011), 331.
35. Partan, S. and Marler, P. 1999. Communication goes multimodal. Science (New York, N.Y.). 283, 5406 (Feb. 1999), 1272–3.
36. Read, R. and Belpaeme, T. 2012. How to use non-linguistic utterances to convey emotion in child-robot interaction. Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12 (New York, New York, USA, 2012), 219.
37. Russell, J.A. et al. 2003. Facial and vocal expressions of emotion. Annual review of psychology. 54, (Jan. 2003), 329–49.
38. Russell, J.A. and A., J. 1980. A circumplex model of affect. Journal of Personality and Social Psychology. 39, 6 (1980), 1161–1178.
39. Saerbeck, M. and Bartneck, C. 2010. Perception of affect elicited by robot motion. Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction - HRI '10. (2010), 53.
40. Salem, M. et al. 2011. A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. 2011 RO-MAN (Jul. 2011), 247–252.
41. Spreng, R.N. et al. 2009. The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. Journal of personality assessment. 91, 1 (Jan. 2009), 62–71.
42. Thomas, F. et al. 1995. The illusion of life : Disney animation. Hyperion.