Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound

Diana Löffler
University of Würzburg, Chair for Psychological Ergonomics
Würzburg, Germany
diana.loeffler@uni-wuerzburg.de

Nina Schmidt
University of Würzburg, Chair for Psychological Ergonomics
Würzburg, Germany
nina.schmidt@stud-mail.uni-wuerzburg.de

Robert Tscharn
University of Würzburg, Chair for Psychological Ergonomics
Würzburg, Germany
robert.tscharn@uni-wuerzburg.de

ABSTRACT

Artificial emotion display is a key feature of social robots to communicate internal states and behaviors in familiar human terms. While humanoid robots can draw on signals such as facial expressions or voice, emotions in appearance-constrained robots can only be conveyed through less-anthropomorphic output channels. While previous work focused on identifying specific expressional designs to convey a particular emotion, little work has been done to quantify the information content of different modalities and how they become effective in combination. Based on emotion metaphors that capture mental models of emotions, we systematically designed and validated a set of 28 different uni- and multimodal expressions for the basic emotions joy, sadness, fear and anger using the most common output modalities color, motion and sound. Classification accuracy and users' confidence of emotion assignment were evaluated in an empirical study with 33 participants and a robot probe. The findings are distilled into a set of recommendations about which modalities are most effective in communicating basic artificial emotion. Combining color with planar motion offered the overall best cost/benefit ratio by making use of redundant multimodal coding. Furthermore, modalities differed in their degree of effectiveness to communicate single emotions. Joy was best conveyed via color and motion, sadness via sound, fear via motion and anger via color.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**;

KEYWORDS

Robotics, social robots, emotion, color, sound, motion, human-robot interaction, non-humanoids, human-agent interaction, multimodal interaction, affective computing

ACM Reference Format:
Diana Löffler, Nina Schmidt, and Robert Tscharn. 2018. Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. In *Proceedings of HRI'18.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3171221.3171261

HRI'18, March 5th–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-4953-6/18/03…$15.00
https://doi.org/10.1145/3171221.3171261

1 INTRODUCTION

People tend to interact with computers and other media in a fundamentally social way [43]. Robotics leverages this phenomenon by designing physically embodied agents that engage with humans on a social level, promoting their anthropomorphization during interaction [10]. A crucial part of this social interaction is the mutual multimodal perception and interpretation of emotion [36]. Benefits of expressing artificial emotions by a robot are to establish social bonds, facilitate empathic inferences about the current internal state in familiar human terms and guiding subsequent behavior [11,17,18]. While humanoid robots can express artificial emotion by borrowing from immediately recognizable anthropomorphic features like facial expression, gestures or speech [8,39], many yet commercially available robots do not possess these output modalities. To communicate affect, such appearance-constrained robots have to draw on modalities like color, sound or vibration instead [5,6,49]. While previous work on affective communication using less-anthropomorphic channels focused on evaluating emotion recognition of specific uni- and multimodal expressions, no attempt has been made on *quantifying* the effectiveness of single and combined modalities on conveying affect. However, an explicit multimodal analysis would allow deriving design recommendations that can be adapted to a variety of appearance-constrained social robots to create tailored expressions that maximize emotion identification and users' confidence in emotion recognition.

In this paper, we aim to provide roboticists with guidance on the effective expression of artificial emotion using the most common output modalities of appearance-constrained social robots: *color, motion* and *sound.* To this aim, first, we review related design efforts in the multimodal expression of emotion with a focus on appearance-constrained social robots and discuss their shortcomings. Next, we create a set of 57 unimodal affective expressions based on literature search to express the four basic emotions joy, sadness, fear and anger utilizing the three modalities on a robot probe. The readability of this initial set is validated in a video-based manipulation check with 22 participants. The most effective expressions for each modality are then selected, systematically combined and evaluated in a lab-based study with 33 participants. As a result, the contribution of each modality and modality-interaction effects on recognizing single and combined emotions can be quantified. We conclude with a set of design recommendations about which modalities are most effective for specific kinds of emotion communication.

Our findings help roboticists to effectively convey affect through readily available output channels in appearance-constrained social robots facilitating intuitive Human-Robot Interaction (HRI).

2 RELATED WORK

2.1 Design Focusing on Output Modalities

Motivated by empirical findings that 80% of human communication is encoded in facial expressions and body movements [45], many studies on the role of nonverbal cues for affective communication in social robots emerged [5]. As a large number of robots does not possess anthropomorphic modalities such as gaze or gesture [5,6,49], HRI researchers increasingly seek to explore the more artificial output channels of appearance-constrained robots like motion [8,44,46] and body language [1,3,13,19,35,37,52,54], orientation [8,9], color and light [48-50], sound [25,26,49,53], or vibration [49]. An advantage of using less-anthropomorphic modalities is also to set more appropriate expectations about the robot's capabilities, as for example a speaking robot could be assumed to be able to maintain a conversation [2], resulting in frustration if this expectation is not met [34].

As mappings between these modalities and emotional expressions are less straightforward compared to mimicking anthropomorphic signals like facial expressions, a common strategy is to borrow from related fields such as animal behavior, as animals apparently communicate their emotions on a less detailed level. However, humans are still able to read the expressed emotions, e.g. from different dog barks, even if they do not own a dog [42]. In an HRI-related example, Singh and Young mapped a robot's emotional state to its doglike tail movements [47]. Such expression designs are, on the one hand, highly dependent on a robots' specific form factor and, on the other hand, should be thoroughly tested for readability, user acceptance and transferability, otherwise low identification rates and user confusion can result.

Other works that do not directly borrow from human or animal models often provide mixed design recommendations for which modalities are most suitable to express certain kinds of emotions. For example, while Song and Yamada conclude that emotions were best identified with their designed color expressions for a simple-shaped robot [49], Häring et al. report on color as an unreliable component for affective expression in humanoid robot NAO [19]. These inconsistencies show that transferring specific expressional designs for certain output modalities between different robots and/or contexts is only possible to a limited extent. As there is a general lack of a methodology that supports the effective translation of emotions to less-anthropomorphic output modalities, arbitrary mappings and reliance on lay theories are favored.

Another example is the common misconception that Plutchik's *wheel of emotions* [41] is a theoretically sound model that maps emotions to colors, while instead, it is only an infographic that used the color wheel to better illustrate the variations and relationships among different emotions. System designs based on such arbitrary assignments of colors to emotions inevitably result in ambiguous cues that hamper intuitive HRI. Thus, expression designs that focus on output modalities rather than understanding the particulars of each emotion need to be rigorously validated to increase transferability.

Moreover, since the contribution of each modality and possible interaction effects between modalities have not been quantified yet, roboticists have little guidance on which modality or modality combination is most effective in communicating emotion, and whether if it is worthwhile to use more than one modality for the expression of affect.

2.2 Design Focusing on Understanding Emotion

An alternative approach to express artificial emotion is to first understand the users' mental models of emotions and then translate them into effective designs. Previous work pursuing this approach follows the notion of embodied cognition, which has become a major paradigm for theorizing and doing research in HRI in recent years [16,52].

A prominent approach of embodied cognition is the cognitive-linguistics theory of conceptual metaphor [28,29], postulating that users' mental models of abstract ideas like emotions can be captured through analysis of metaphors in discourse. For example, when talking about emotions, we use expressions such as *warm joy*, suggesting that joy is conceptualized in terms of physical temperature. These mental models expressed in conceptual metaphors also have behavioral consequences. For example, the perception of temperature differences impacts judgements of joyfulness, even if the temperature is only indicated by color [32]. Thus, when talking about abstract concepts (e.g., joy), we borrow the experiential entailments from concrete physical target domains (i.e., temperature), which allows us to describe and reason about abstract concepts as if they had concrete properties [7].

Conceptual metaphors such as JOY IS WARM temperature can now be used to translate emotions into physical parameters that can be expressed with a robots' output modalities and are intuitively understood by the user as they incorporate the users' mental model [21,33]. Although this cognitive-linguistics approach extracts the mental models through metaphors in discourse and therefore seems to be very culture-specific, basic conceptual metaphors, especially those of emotions, are very little dependent on culture [31] and similar expressions can be found in many languages [27]. In his book Metaphor and Emotion: Language, Culture, and Body in Human Feeling, Kövecses summarizes a wealth of cross-cultural metaphorical speech about different emotions [27] on which roboticists can draw on.

- Joy is metaphorically conceptualized as UP (cheering him up, six feet off the ground, heaven on earth), LIGHT (bright smile), ACTIVE (alive with joy) and WARM (that warmed my spirits).
- Sadness is understood in terms of DOWN (he brought me down with his remarks), DARKNESS (dark mood, lacking brightness in the mood), COLD (losing his father put his fire out), PASSIVENESS (disheartening news), BLUE (feeling blue) and BURDEN (he staggered under the pain).
- Fear metaphors include HIDDEN ENEMY (fear slowly crept up on him), BURDEN (fear weighted heavily on them), NATURAL FORCE (she was engulfed by panic) and DARKNESS (being overshadowed by fear, eyes being dark with fear).

- Anger is metaphorically conceptualized as HOT FLUID IN A CONTAINER (boiling with anger), OPPONENT IN A STRUGGLE (overcome by anger), AGGRESSIVE ANIMAL BEHAVIOR (don't snap at me) and RED (red with rage, seeing red).

Following conceptual metaphors in the design process has been proven flexible enough to facilitate intuitive interaction with various user interfaces, such as graphical, tangible and gesture [22,32,33]. As this method of deriving design features of emotion expressions from language is quick and low-cost, e.g., compared to co-designing with actors, we aim to extend it to HRI. As affective expressions inspired by emotion metaphors have a high chance to be intuitively understood by the user, they represent suitable candidates to quantitatively compare the information content of different emotion expression modalities and how they might become effective in combination.

3 EXPRESSION DESIGN

Following the emotion metaphors compiled by Kövecses [27], we created multiple unimodal expressions for each of selected output modalities and target emotions. We then validated these variants in a video-based manipulation check to select the most effective expressions for the subsequent user evaluation. This iterative procedure was required since the emotion metaphors do indicate modality characteristics that are crucial to understanding an emotion, but not precise parameters. The selection of output modalities, target emotions, as well as the design and validation of the candidate expressions with our robot probe are described in the following.

3.1 Selection of Output Modalities

To determine which output modalities are most frequently used in current commercially available social robots and related research prototypes, we conducted an informal online search on Google and Google Scholar using keywords such as "robot" or "robots today" in English and German. In addition, technology news websites (e.g., golem.de) and online shops (e.g., robotshop.com) were scanned. The search brought up 59 different robots (see Figure 1) that were examined for their output modalities.

The output modalities were determined from product descriptions, instruction manuals or inferred from accompanying videos showing the robot in action. If there was no information available whether a specific robot possesses a certain output modality, we coded it as if the robot did not have this output modality. Although our survey included humanoid as well as non-humanoid social robots, the top three output modalities were less-anthropomorphic and comprised colored light (82%), motion (80%) and sound (56%). Anthropomorphic output modalities like voice (35%), facial expression (25%) or gaze (13%) were less frequently used. Out of the 59 robots, 32% possessed a display. Based on these results, we selected color, motion and sound as the most frequently available output modalities for our emotion expression designs.

3.2 Target Emotions

The literature on the classification of different types of emotions is comprehensive and manifold. A common distinction divides emotions into primary and secondary. Primary emotions are innate and experienced for short time periods, oftentimes as a reaction

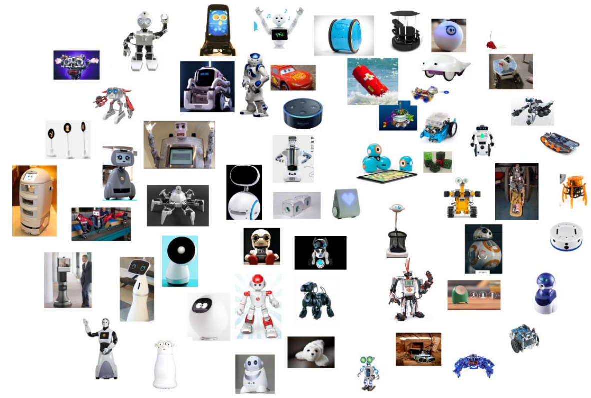

Figure 1: The 59 robots examined for output modalities.

to external stimuli, and are shared across cultures. These emotions offer the greatest potential for intuitive HRI and are thus studied in most of the theoretical and empirical work. Primary emotions are joy, sadness, anger, fear, surprise, and disgust [15]. Secondary emotions, on the contrary, are acquired and do not have a universal corresponding facial expression. Guilt, regret, pride and jealousy are examples of secondary emotions.As suggested by more recent research on facial expressions, the primary emotions probably originated from a basic set of four: joy, sadness, and two mixed categories of anger/disgust and fear/surprise [24]. These four emotions have a very distinct facial expression that is recognized by all humans independent of their cultural background. However, the mixed categories of anger/disgust and fear/surprise are only later discriminated into separate emotions. According to findings from developmental psychology, young children have difficulties in ascribing the emotions disgust and surprise, but not anger and fear [20]. Based on these insights, we focus on the target emotions joy, sadness, fear and anger in the following.

3.3 Candidate Expressions

3.3.1 Joy. To select appropriate colors for joy, we refer to the emotion metaphors JOY IS LIGHT and JOY IS WARM, indicating that joy is cross-culturally associated with light and warm hues, intense saturation and high brightness [31]. Thus, we selected different pink and yellow hues with varying degrees of intense saturation and high brightness (HSB values: 45/100/100, 45/60/100, 50/100/100, 50/60/100, 315/100/100, 315/60/100). To express joy through movement, the emotion metaphors JOY IS UP and JOY IS ACTIVE provide hints to what is important when conveying this emotion. According to our survey, most robots can only perform horizontal movements, so we focused on the metaphor joy is active and translated it into planar motions. We created variants of dance-like and "purposeless" movements that should be perceived as alive and active. The emotion metaphors joy is up inspired the expression design for sound. As voice output is much less common than sound in currently available social robots, we decided to only use tone patterns to convey the target emotions. Joy was operationalized as high pitched sounds (300-900 Hz, 200-480ms beeps) and longer inter-beep-intervals (250-440ms) [42]. All sound patterns were created with Audacity. All tones consisted of a sine waveform generated by the built-in generator *chirp* or *tone*. Adjustments were made using the effects tool fade in/out and change tempo. All sounds had an amplitude change from 0.1-1 and lasted for 3s.

3.3.2 Sadness. To choose appropriate colors to represent sadness, the emotion metaphors SADNESS IS DARKNESS and SADNESS IS BLUE were used, which indicate a link between short wavelength hues, low saturation and reduced brightness [31]. Based on this, we created the following shades of dark blue to represent sadness (HSB values): 240/100/50, 240/100/40, 240/50/50, 240/40/40. Regarding movement, sadness is indicated by a lack of vitality, passiveness and reticence as expressed in the emotion metaphors SADNESS IS DOWN, SADNESS IS PASSIVENESS and SADNESS IS A BURDEN. Thus, the designed movements variants contained slow, barely noticeable motions with rotations away from the user. The sound for sadness was inspired by the emotion metaphors SADNESS IS DOWN and SADNESS IS PASSIVENESS, resulting in "slow" sounds similar to a slow speaking rate [39] and a falling tone [39,49] (100-400 Hz with 0.1-1 amplitude).

3.3.3 Fear. Colors for fear are taken from the emotion metaphor FEAR IS DARKNESS. This led us to the operationalization as black and gray colors (HSB values 0/0/20, 0/0/40). The emotion metaphor FEAR IS A HIDDEN ENEMY was translated into "hiding" and "escaping" movements, as expressed in *being hounded by fear*. To choose appropriate sounds for the emotion of fear, we created several sound files with high pitched tones [42] and a fast rate [39] (800-850 Hz with short 100ms beeps and 50ms inter-beep-intervals) according to the emotion metaphors FEAR IS A HIDDEN ENEMY and FEAR IS A NATURAL FORCE.

3.3.4 Anger. Suitable colors for anger are derived from the emotion metaphor ANGER IS SEEING RED and operationalized as different shades of red (HSB values: 360/100/100,360/100/80, 360/100/70). Inspired by the emotion metaphor ANGER IS HOT FLUID IN A CONTAINER, ANGER IS AN OPPONENT IN A STRUGGLE and ANGER IS AGGRESSIVE ANIMAL BEHAVIOR, we designed several shaking and forward movements that mimic a bursting container, an "inner fight" as well as fast motion towards the user. The sound files for the anger emotion were inspired by the emotion metaphor ANGER IS AGGRESSIVE ANIMAL BEHAVIOR, resulting in low pitched sounds and short inter-beep-intervals [42] (80-150 Hz beeps of 100-200ms and 50ms inter-beep-intervals).

3.4 Social Robot Probe

To be able to validate the candidate expressions with users, we designed a social robot probe capable of performing planar movements and displaying colored light and sound. We took great care of a simplistic design without anthropomorphic, zoomorphic or too mechanical elements to minimize potential biases in affective attribution. Another design requirement was that the robot should be able to navigate as silently as possible, since the noise of servomotors has been recently shown to shape aural impressions of robotic motion [49], which would be a confounding variable. The final form was mostly determined by the robots' mechanical interior and desired functionality. To enable the robot to perform planar movements on a table in front of the participants in the user study, a small turning cycle and high agility were achieved through

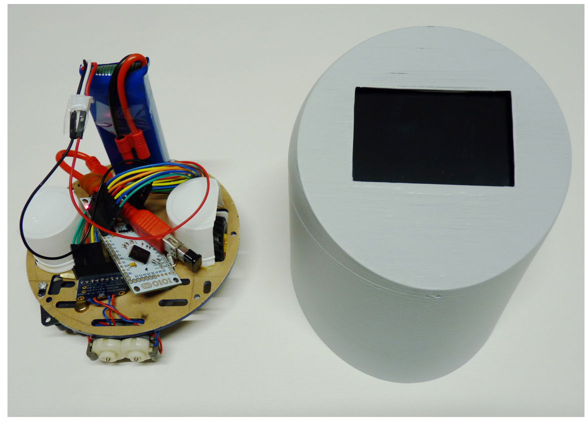

Figure 2: Robot interior (left) and case with display (right).

two wheels and a ball caster from the P*ololu Round Robot Chassis Kit*. An Android phone (LG L40) was used as the main processor connected via Bluetooth to an IOIO OTG Board with the Adafruit TB6612 Motor Driver Breakout Board to control the engines. The Android phone was also used to display colors and play sounds and was attached to an inclined plane on the robot's 3d printed gray plastic case (see Figure 2). An Android app was used to control the robot during the user study and a 7.4V 2000mAh 25C Lipo Battery served as energy supply.

3.5 Manipulation Check

To gather initial user feedback on the designed set of 57 unimodal affective expressions (15 colors, 20 sounds, 22 motions), an online video-based survey was created with Unipark. A convenient sample of 22 undergraduates from a German university was tasked to rate each of the presented unimodal stimuli on a 7-point Likert scale from "appropriate" to "not appropriate" regarding how well the expression conveyed the four emotions joy, sadness, fear and anger. The reason why only unimodal expressions were selected for this initial validation is that we wanted to quickly narrow down the stimulus material from 57 to 12 expressions for the final evaluation, as testing 57 expressions will all possible combinations would have exceeded the length of one study by a multiple. Moreover, as unimodal emotion expression is more challenging that multimodal, expressions that perform worse in this validation are also less likely to perform well in combination.

The stimuli were presented to participants in random order. Subjects were requested to turn on sound and increase volume on their computers, maximize brightness on their screen, disable color changing programs like night shift, and enable video plugins. Their settings were tested as they had to enter a word that was spoken in a sound file and a number that was shown in a video.

The expression that scored best for each emotion and modality was then selected for the final user study (see Table 1). Through the manipulation check, valuable insights regarding the study procedure and instructions of the participants could be gained. For example, sounds designed to express fear were often confounded with anger and vice versa. This might have been due to a confusion of the participants whether they should indicate how they felt (e.g.,

Table 1: Final expression designs tested in the user evaluation.

Emotion	Color (HSB)	Motion	Sound
Joy	45/100/100	fast rotations, circular movements	4 tones with 300-900Hz, amplitude 0.1-1 for 200-480ms, 300ms pause
Sadness	230/40/40	slowly rotating away from the user	falling beep sound 300-100Hz, amplitude 1-0.5 over 3000ms
Fear	0/0/20	jumpy movements away from the user	alternating tone (800Hz, amplitude 0.1-1) for 100ms, 50ms pause
Anger	0/100/100	shaking movements towards the user	3 tones with 80-150Hz, amplitude 0.1-1, for 100-200ms, 200ms inter-beep-intervals

frightened by the robot's angry beeps) or which emotion the robot tried to convey. We transferred these insights to the study design of the final evaluation.

4 USER EVALUATION

We set up a lab-based user study to quantify the effectiveness of each modality (color, motion and sound) in communicating the four basic emotions as well as the interplay when multiple modalities are combined. The overall goal is to identify a) the significance of each modality in communicating emotion and b) possible interaction effects between modalities for single emotions as well as for the whole set combined. These results will be used to derive design recommendations about which modalities are most effective in communicating specific basic emotions and which are best to express all four. "Effectiveness" was operationalized as objective classification accuracy (whether a presentation modality was classified as the intended emotion or not) as well as subjective confidence in the classification. According to literature, the perceived confidence should result in higher ratings for multimodal expressions, e.g., color and sound, compared to unimodal expressions, e.g., color [4]. The participants, study procedure and results of the user evaluation are described in the following.

4.1 Participants

33 German students (17 female) were recruited and participated for course credit or 5 Euro. The participants' ages ranged from 18 to 28 years (M = 22.51 years, SD = 2.44). All participants had no reported defective color vision and passed all of 15 selected plates of the Ishihara Color Test [23] that tested for the most common red-green color impairment. Collected control variables include mood, interest in technology, interest in robots and emotional intelligence. The participants' self-reported mood was rated as M = 2.38, SD = 0.58 (5-point Likert scale from 1="positive mood" to 5="negative mood"). On 7-point Likert scales participants rated their interest in technology as rather high (M = 5.84, SD = 1.42), and their interest in robots as medium high (M = 4.59, SD = 1.64). None of the participants owned a robot. The participants' emotional intelligence was rated on thirty 7-point Likert scale items as M = 3.78 (SD = 0.26), with high values indicating high emotional intelligence.

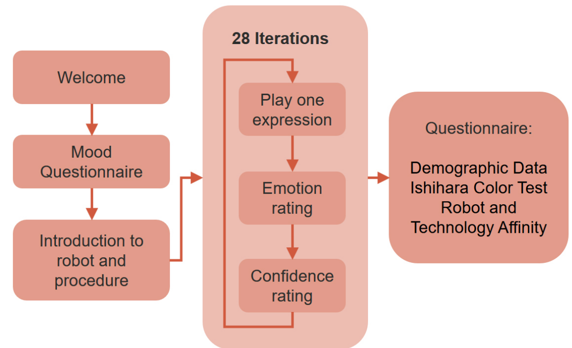

Figure 3: Procedure of the user evaluation.

4.2 Procedure

The study took place in an artificially lit room (220 Lux) with covered windows, white walls and a wooden table covered in a white tablecloth. After being welcomed and signing the consent form, participants were seated at the table and asked to complete a mood questionnaire [40], see Figure 3. After that, they were introduced to the robot and received oral instruction about the study. Subjects were told that in each experimental trial they will see the robot expressing a certain emotion using color, motion and/or sound. The final set of parameters for the three modalities (color, motion, sound) and selected emotions (joy, sadness, fear, anger) based on the results of the single-channel video-based validation study are depicted in Table 1. The participants were tasked to select the most appropriate emotion out of the four given categories joy, sadness, fear and anger, and rate their confidence on a 7-point Likert scale from 1="not confident at all" to 7="very confident". The target emotion expression could be replayed as often as the participants liked. This procedure was repeated 28 times. To avoid sequence effects, the stimuli were presented in random order. After finishing all trials, the participants were asked to respond to 15 selected plates of the Ishihara Color Test [23] to test for color vision deficiency and fill out a final questionnaire asking for demographic data, technology and robot affinity, as well as emotional intelligence. The whole experiment lasted about 15 minutes.

4.3 Results

4.3.1 Data analysis. Figure 4 visualizes how each of the 28 expressions (differing in presentation modality) was categorized into the four target emotions. Table 2, right columns, show the mean classification accuracy and the participants' confidence ratings aggregated for each presentation modality. The interplay between presentation modalities, emotion classification accuracy and confidence in the assignments will be analyzed and presented in two steps. First, the overall significance of each modality and possible modality-interaction effects are determined for communicating the combined set of emotions. Second, the significance of each modality and interaction effects when combining modalities are analyzed for single emotions.

To this aim, binary logistic Generalized Estimating Equations (GEE) models were utilized. This approach was chosen as the model can handle correlated data that arises from repeated measures of the same individuals over time as well as binary outcome data types (classification accuracy), and is more flexible for missing data compared to other approaches [30]. The results of the GEE model are presented as Odds Ratios. Ratios greater than 1 are interpreted as increasing the likelihood of the to-be-predicted event, whereas ratios less than 1 are interpreted as decreasing the likelihood of the to-be-predicted event (in this case classification accuracy). Ratios equal to 1 are not significantly associated with increased or decreased likelihood of classification accuracy.

4.3.2 Modality Significance for Combined Emotions. The GEE model for combined emotions contained the seven presentation modalities (color, motion, sound, color-motion, color-sound, motion-sound, color-motion-sound) as dummy-coded predictors for classification accuracy for each emotion and the control variables age, sex, mood, technology interest, robot interest and emotional intelligence as covariates. The theoretical assumption that each of the four emotion categories would have been chosen at chance level (25%) if no presentation modality would have been shown to the participants was chosen as the reference category.

Table 2, left columns, summarize the computations for parameter estimates for the GEE model. GEE results show that all presentation modalities were strong predictors of classification accuracy. Color alone was the weakest predictor for emotion classification, whereas motion and sound performed equally well. Combining two modalities improved classification accuracy, with color-motion being the strongest predictor, increasing the likelihood that an emotion was correctly classified by 1:31.4 times. Combining all three emotions did not markedly improve classification accuracy compared to the bimodal color-motion presentation style. Sex, age, mood, emotional intelligence and interest in technology were not significantly associated with classification accuracy. In contrast, when subjects described themselves as being more interested in robots, the emotion classification accuracy was slightly increased by 1:1.21 times.

To assess the influence of the number of modalities on confidence ratings, a repeated measures ANOVA was performed and revealed a significant influence of the number of modalities on confidence ratings, $F(2,64) = 88.52$, $p<.001$, $\eta_p^2 = .73$. To analyze differences between one, two, and three modalities, planned Helmert-contrasts were calculated. Confidence ratings for unimodally conveyed emotions were significantly lower than for bi- or trimodal emotions, $F(1,32) = 125.6$, $p<.001$, $\eta_p^2 = .80$, and confidence for bimodal emotions was significantly lower than for trimodal emotions, $F(1,32) = 41.7$, $p<.001$, $\eta_p^2 = .57$. This means that combining different modalities increases the participants perceived confidence in emotion categorization.

4.3.3 Modality Significance for Single Emotions. To determine the significance of each modality and possible interaction effects between modalities in communicating a specific emotion, a GEE

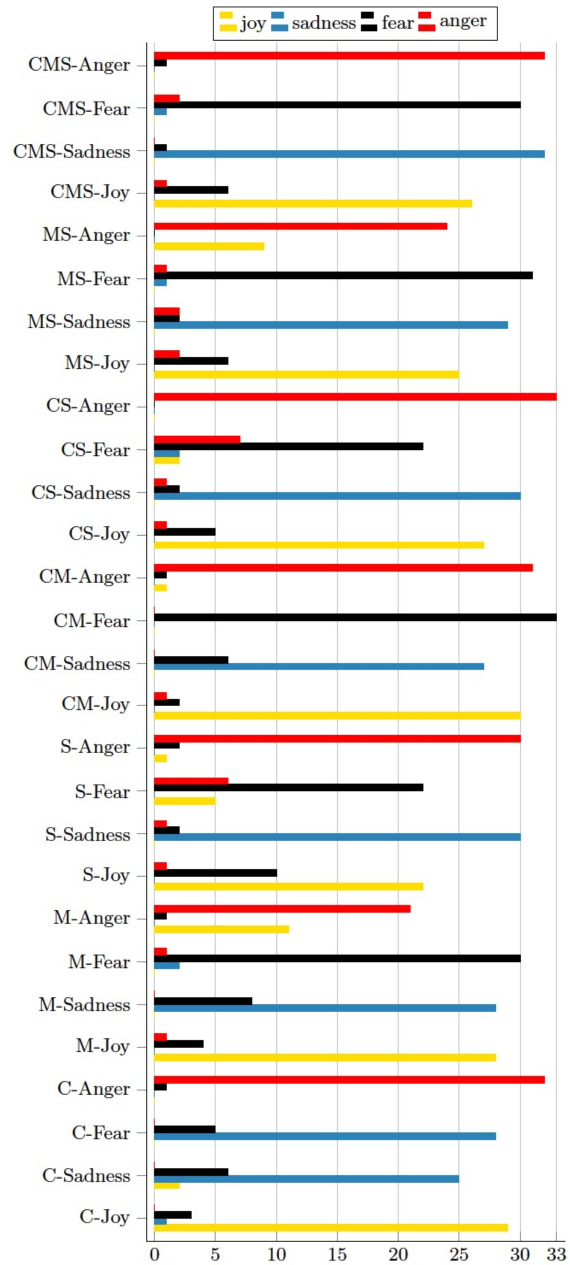

Figure 4: Emotion categorization of 28 different presentation styles across three modalities for the 33 participants.

model is calculated for each emotion containing the seven presentation modalities (color, motion, sound, color-motion, color-sound, motion-sound, color-motion-sound) as dummy-coded predictors for classification accuracy. Because of the overall non-significant effects of the control variables and the weak predictive power of "interest for robots", the control variables were dispensed in the emotion-specific models in favor of clarity. The presentation modality with the highest classification accuracy was chosen as reference category for each emotion-specific GEE model. To estimate the impact of the seven presentation modalities on the metric variable "perceived confidence in emotion classification", an analogue GLM was computed for each emotion. The results of the analyses of modality significance for single emotions are described in the following.

Joy. The emotion "joy" received the highest mean classification accuracy in the color-motion presentation modality, so this was chosen as the reference category of the emotion-specific GEE and GLM models. Compared to this reference, the channels sound and motion-sound significantly reduced the likelihood of an accurate classification by 1:5.0 and 1:3.2, respectively. Regarding perceived confidence of the classification, the unimodal channels color and sound received significantly lower ratings than the bimodal channel color-motion.

From these results, it can be concluded that sound does not communicate the emotion of joy well. Color is the strongest predictor, followed by motion. As color and motion alone are quite prone to classification errors, both channels should be combined to leverage the multimodal advantage of higher classification accuracy (91%) and higher perceived confidence in the classification. However, adding sound to create a trimodal expression does not add to classification accuracy nor perceived confidence.

Sadness. The emotion "sadness" received the highest mean classification accuracy in the trimodal color-motion-sound presentation modality, so this was chosen as the reference category of the emotion-specific GEE and GLM models. Compared to this reference, the channels color, motion and color-motion significantly reduced the likelihood of an accurate classification by 1:10.2, 1: 10.2, and 1:7.1, respectively. The perceived confidence of the classification was lower for the unimodal presentations color, motion, sound, and for the bimodal presentation color-motion.

From this, it can be concluded that the main driver for communicating the emotion sadness is sound. Whenever sound is involved, the classification accuracy reaches a maximum of 93% on average. Adding either color or motion raises the classification confidence compared to sound only, but a trimodal presentation does not add a significant increase in perceived confidence.

Fear. The emotion "fear" received the highest mean classification accuracy in the bimodal presentation modality color-motion, so this was chosen as the reference category of the emotion-specific GEE and GLM models. Compared to this reference, the presentation styles color, sound and color-sound significantly reduced the likelihood of classification accuracy by 1:166.7, 1:15.9, and 1:15.9, respectively. The confidence ratings decreased in presentation modalities color, sound and color-sound. Combining all three modalities did not result in an additional confidence benefit.

From these results, it can be concluded that the emotion fear is best communicated by motion. Although color as a single channel should be avoided as participants misinterpreted the black color as "sadness", the combination together with motion disambiguates the expression and leads to a perfect classification accuracy of 100%. The confidence ratings are highest in all presentation modalities involving motion, and adding further modalities is not beneficial.

Anger. The emotion "anger" received the highest mean classification accuracy in the color-sound presentation modality, so this was chosen as the reference category of the emotion-specific GEE and GLM models. Compared to the reference, the presentation modalities motion and motion-sound significantly reduced the likelihood of classification accuracy by 1:18.2 and 1:12.0, respectively. Regarding perceived confidence of the classification, the unimodal channels color, motion, sound and the bimodal presentation modality motion-sound decreased perceived classification confidence.

These results show that the emotion anger is best communicated by color, followed by sound. Motion is less suited to communicate this emotion. Whenever color is involved as presentation modality, classification accuracy reaches 97% on average. Combining color and sound leads to a significant increase in confidence ratings compared to color alone, but combining all three modalities does not further benefit perceived classification confidence.

Table 2: Beta-Coefficients, Odds Ratios and [95% CIs] of the GEE model predicting whether the emotions were correctly classified together with mean classification accuracy in %, confidence ratings and (SDs) aggregated for each modality.

	b	Odds	acc.	conf.
Constant	-2.12	0.12		
	[-5.38, 1.14]	[0.01, 3.14]		
Color	1.90*	6.69	69 (47)	4.39 (1.94)
	[1.65, 2.15]	[5.20, 8.60]		
Motion	2.43*	11,4	79 (41)	4.69 (1.50)
	[2.05, 2.81]	[7.80, 16.66]		
Sound	2.48*	11,95	79 (41)	4.47 (1.64)
	[2.00, 2.96]	[7.41, 19.27]		
Color-Motion	3.45*	31,37	92 (28)	5.37 (1.42)
	[2.81, 4.09]	[16.53, 59.53]		
Color-Sound	2.86*	17,37	85 (36)	5.16 (1.48)
	[2.41, 3.30]	[11.12, 27.54]		
Motion-Sound	2.74*	15,47	83 (38)	5.09 (1.54)
	[2.32, 3.16]	[10.17, 23.54]		
Color-Motion-Sound	3.54*	34,55	91 (29)	5.79 (1.37)
	[2.93, 4.15]	[18.78, 63.58]		
Sex (Female)	-0.21	0.82		
	[-0.65, 0.24]	[0.52, 1.27]		
Age	-0.54	0.95		
	[-0.13, 0.02]	[0.88, 1.02]		
Mood	-0.23	0.80		
	[-0.56, 0.10]	[0.57, 1.11]		
Emotional Intelligence	0.62	1.85		
	[-0.11, 1.34]	[0.90, 3.82]		
Interest in Technology	-0.06	0.94		
	[-0.29, 0.17]	[0.75, 1.18]		
Interest in Robots	0.19*	1,21		
	[0.04, 0.34]	[1.04, 1.41]		

5 DISCUSSION

Based on the results from our user evaluation, we propose a set of design recommendations that can guide roboticists on effectively expressing emotions with the most common output modalities of appearance-constrained social robots.

(1) **Output modalities differ in their degree of effectiveness to communicate basic emotions.** While the modalities color, motion, sound and combinations thereof are all

suitable for expressing the basic emotions joy, sadness, fear and anger (82% in classification accuracy on average), they differ substantially in effectiveness, ranging from 15-100%.

(2) **Multimodal communication of basic emotions improves classification accuracy and confidence.** Unimodal expressions are correctly identified in 76% of cases. Combining two modalities improves classification accuracy to 87% and largely increases people's confidence that they have determined the robot's correct emotion. Combining more than two modalities does not significantly improve classification accuracy, but further increases confidence in the assignment.

(3) **Combining color and motion provides the best cost/benefit ratio in communicating basic emotions.** If robot designers have to make a choice between output channels, the motion channel is recommended if only one modality is used for emotion communication (79% classification accuracy). If two modalities can be chosen, a combination of the most widespread output channels color and motion is recommended (92% classification accuracy).

(4) **Output channels should be chosen based on which emotions are going to be conveyed.** As no single modality covers all emotions equally well, robot designers should prioritize modalities based on the emotions they would like to communicate. Joy is best expressed through color (88%), followed by motion (85%). As both unimodal expressions are quite prone to errors, they should be combined (91%). Sadness is best communicated by sound (91%), fear by motion (91%) and anger by color (97%).

(5) **Following emotion metaphors is a systematic, quick and low-cost way to design effective multimodal expressions of artificial emotion.** Joy is understood in terms of WARM, LIGHT and UP and can be expressed through yellow color and high pitched tones, or a combination thereof. Sadness is conceptualized as down and can be conveyed by a slowly falling sound. Fear is understood as the flight from a HIDDEN ENEMY, which can be designed as jumpy movements away from the user. Anger is conceptualized as SEEING RED, which can be conveyed through red color.

Our quantitative comparison of the information content of different output modalities for the communication of affect could show that different emotions use different distributions of expression channels, showing parallels non-verbal human and animal communication. For example, the universal emotion fear evolved to prevent the organism from exposure to harmful stimuli. This could explain why a stylized flight motion pattern is the main predictor to identify this emotion correctly. On the other hand, the dominant impact of color on the classification of the anger emotion probably originates from the physiological correspondence between anger and facial flushing [12], that is further strengthened by metaphorical speech (*seeing red*).

Another finding worth discussing is that participants were less confident in their attribution of affect and categorized expressions less accurately when emotions were presented unimodally. Adding more modalities increased confidence ratings and decreased error rates, a finding in line with the literature on multimodal interaction, as the information provided by one source is used to resolve ambiguities in another, thereby reducing errors and uncertainties [4].

This leads us to the discussions of limitations of this study. Like other empirical work on the readability of affective expressions of social robots, the emotion categories were prescribed. If an expression did not clearly match one of the four target emotions, participants could use the exclusion principle, as they were required to provide an answer in each trial. Because of this forced choice, the results of this study overestimate the true performance. Nevertheless, if the expressed emotions would have been perceived as poor operationalizations, low confidence ratings and random emotion category assignments would have resulted, which was not the case. Moreover, the designs were based on emotion metaphors to increase their perceived meaningfulness and subsequently validated in a manipulation check. Another limitation is the small and not representative sample size of 33 German undergraduates. Future studies are needed that verify the designed emotion expressions in other cultures and user populations with other sociocultural norms that may influence emotional expression [17].

6 CONCLUSION

In this paper, we quantified the significance of the three most common output modalities color, motion and sound for expressing basic artificial emotion in an appearance-constrained social robot. By following emotion metaphors that capture mental models of emotions accessed by analysis of metaphorical language, we created 57 unimodal expressions for the emotions joy, sadness, fear and anger. This initial set was validated in a video-based study with 22 participants and a social robot probe. The best expressions for each modality and emotion were selected and systematically combined with the other modalities to create a final set of 28 unimodal and multimodal affective expressions. In a controlled lab experiment involving 33 participants, classification accuracy and perceived confidence in emotion assignment were measured. The findings are distilled into five design recommendations for effective multimodal expression of artificial emotion via three low-cost output channels that are already present in most robots.

This work significantly extends previous research on artificial emotion expression by quantifying multimodal communication, making the findings adaptable to a variety of social robots. An open question for further research is whether our design recommendations also support the effective expression of emotion in humanoid robots. Also, future studies are needed that verify other emotions like surprise or anticipation, quantify the information content of other promising modalities like temperature or vertical movements, and explore the impact of sociocultural factors, also in more applied contexts than this study could provide.

ACKNOWLEDGMENTS

We thank our anonymous reviewers of HRI 2018 and Program Committee member Heather Knight for their invaluable comments on restructuring a prior draft of this work.

REFERENCES

[1] 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: 717-746.
[2] Bartneck, Christoph, and Jodi Forlizzi. A design-centred framework for social human-robot interaction. In RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), 591-594.
[3] Aryel Beck, Lola Canamero, and Kim A. Bard. 2010. Towards an Affect Space for robots to display emotional body language. In 19th International Symposium in Robot and Human Interactive Communication, 464-469.
[4] Paul Bertelson and De Gelder. 2004. The psychology of multimodal perception. In Crossmodal space and crossmodal attention. 141-177.
[5] Cindy L. Bethel and Robin R. Murphy. 2008. Survey of Non-facial/Non-verbal Affective Expressions for Appearance- Constrained Robots. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38, 1: 83-92.
[6] Cindy L. Bethel and Robin R. Murphy. 2006. Affective expression in appearance constrained robots. In Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI'06, 327.
[7] Anna M. Borghi and Ferdinand Binkofski. 2014. Language, Languages, and Abstract Concepts. Wiley Interdisciplinary Reviews: Cognitive Science 5, 2: 111-124.
[8] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. International Journal of Human-Computer Studies 59, 1-2: 119-155.
[9] Cynthia Breazeal. 2004. Social interactions in HRI: the robot view. In IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 181-186.
[10] Cynthia Breazeal. 2003. Toward sociable robots. Robotics and Autonomous Systems 42, 3-4: 167-175.
[11] Cynthia Breazeal and Rodney Brooks. 2005. Robot Emotion: A Functional Perspective. In Who Needs Emotions? Oxford University Press, 271-310.
[12] Mark a Changizi, Qiong Zhang, and Shinsuke Shimojo. 2006. Bare skin, blood and the evolution of primate colour vision. Biology letters 2, 2: 217-221.
[13] Mark Coulson. 2004. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. Journal of Nonverbal Behavior 28, 2: 117-139.
[14] Kerstin Dautenhahn. 2007. Socially intelligent robots: dimensions of human-robot interaction. Philosophical Transactions of the Royal Society of London B: Biological Sciences 362, 1480.
[15] Paul Ekman. 1992. An argument for basic emotions. Cognition & Emotion 6, 3: 169-200.
[16] Fiske, Susan T., Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: warmth and competence. Trends in Cognitive Sciences 11, 2: 77-83.
[17] Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. Robotics and Autonomous Systems 42, 3-4: 143-166.
[18] Gross, James. 2013. Handbook of Emotion Regulation. Guilford publications.
[19] Häring, Markus, Nikolaus Bee, and Elisabeth Andre. 2011. Creation and Evaluation of Emotion Expression with body movement, sound and eye color for humanoid robots. In Ro-Man, 204-209.
[20] Harris, Paul L. 1989. Children and emotion: The development of psychological understanding. Cambridge, MA: Basil Blackwell.
[21] Jörn Hurtienne and Lucienne Blessing. 2007. Design for Intuitive Use-Testing image schema theory for user interface design. In 16th International Conference on Engineering Design, 1-12.
[22] Jörn Hurtienne, and Lucienne Blessing. 2007. Design for Intuitive Use - Testing image schema theory for user interface design. The International Conference on Engineering Design.
[23] Shinobu Ishihara. 1917. Tests for Colour Blindness. Handaya Hongo Harukich, Tokyo.
[24] Jack, Rachael E., Oliver GB Garrod, and Philippe G. Schyns. 2014. Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. Current Biology 24, 2: 187-192.
[25] Takanori Komatsu. 2005. Toward Making Humans Empathize with Artificial Agents by Means of Subtle Expressions. In ACII, 458-465.
[26] Takanori Komatsu and Seiji Yamada. 2007. How do robotic agents' appearances affect people's interpretations of the agents' attitudes? In CHI'07 extended abstracts on Human factors in computing systems - CHI'07, 2519.
[27] Zoltan. Kövecses. 2003. Metaphor and emotionâĂŕ: language, culture, and the body in human feeling. Cambridge University Press.
[28] George Lakoff and Mark Johnson. 1997. Metaphors We Live By. The production of reality: essays and reading on social interaction: 124-134.
[29] Mark J Landau, Lucas A Keefer, and Brian P Meier. 2011. Wringing the Perceptual Rags: Reply to IJzerman and Koole (2011). Psychological Bulletin 137, 2: 362-365.
[30] Kung-Yee Liang and Scott L. Zeger. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 1: 13-22.
[31] Diana Löffler. 2017. Color, Metaphor and Culture: Empirical Foundations for User Interface Design. PhD thesis. University of Würzburg.
[32] Diana Löffler, L. Arlt, T. Toriizuka, Robert Tscharn, and J Hurtienne. 2016. Substituting Color for Haptic Attributes in Conceptual Metaphors for Tangible Interaction Design. In TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, 118-125.
[33] Diana Löffler, Anne Hess, Andreas Maier, Jörn Hurtienne, and Hartmut Schmitt. 2013. Developing intuitive user interfaces by integrating users' mental models into requirements engineering. In Proceedings of the 27th International BCS Human Computer Interaction Conference, 15.
[34] Manja Lohse. 2011. Bridging the gap between users' expectations and system evaluations. In 2011 RO-MAN, 485-490.
[35] Marco de Meijer. 1989. The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behavior 13, 4: 247-268.

[36] Francois Michaud, Paolo Pirjanian, Jonathan Audet, and Dominic LÃĪtourneau. 2000. Artificial Emotion and Social Robotics. In Distributed Autonomous Robotic Systems 4. Springer Japan, Tokyo, 121-130.

[37] Jerome Monceaux, Joffrey Becker, Celine Boudier, and Alexandre Mazel. 2009. Demonstration: first steps in emotional expression of the humanoid robot nao. Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI.

[38] Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. 2017. Making Noise Intentional: A Study of Servo Sound Perception. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI'17, 12-21.

[39] Murray, Iain R., and John L. Arnott. 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. The Journal of the Acoustical Society of America 93, 2: 1097-108.

[40] singer, Eva. 2003. IN THE MOOD. Eine Untersuchung zum Mood Management in der Fernsehnutzung. Ludwig- Maximilians-Universität München.

[41] Plutchik, Robert. 2003. Emotions and Life: Perspectives from Psychology, Biology, and Evolution. American Psychological Association.

[42] Peter Pongracz, Csaba Molnar, Adam Miklosi, V. Csanyi, T. Leitner, and D.M. Rumbaugh. 2006. Acoustic parameters of dog barks carry emotional information for humans. Applied Animal Behaviour Science 100, 3-4: 228-240.

[43] Reeves, Byron, and Clifford Ivar Nass. 1996. How people treat computers, television, and new media like real people and places.

[44] Mark Scheeff, John Pinto, Kris Rahardja, Scott Snibbe, and Robert Tow. 2002. Experiences with Sparky, a Social Robot. In Socially Intelligent Agents. Kluwer Academic Publishers, Boston, 173-180.

[45] Jocelyn Scheirer, Jocelyn Scheirer, and Rosalind W. Picard. 1999. Affective objects. MIT Media Laboratory Perceptual Computing Section: 2000.

[46] Shimokawa, Takuji, and Tetsuo Sawaragi. 2011. Acquiring communicative motor acts of social robot using interactive evolutionary computation. In IEEE International Conference on Systems, Man, and Cybernetics, 1396-1401.

[47] Ashish Singh and James E. Young. 2012. Animal-inspired human-robot interaction. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI'12, 237.

[48] M. V. Sokolova and A. Fernandez-Caballero. 2015. A review on the role of color and light in affective computing. Applied Sciences 5, 3: 275-293.

[49] Song, Sichao, and Seiji Yamada. 2017. Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. ACM, 2017.

[50] Terada, Kazunori, Atsushi Yamauchi, and Akira Ito. 2012. Artificial emotion expression for a robot by dynamic color change. In RO-MAN, 314-321.

[51] Joshua Wainer, David Feil-seifer, Dylan Shell, and Maja Mataric. 2006. The role of physical embodiment in human-robot interaction. In ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, 117-122.

[52] Harald G. Wallbott and Harald G Wallbott. 1998. Bodily expression of emotion. European Journal of Social Psychology 28, 6: 879-896.

[53] Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. 2016. Review of Semantic-Free Utterances in Social Human-Robot Interaction. International Journal of Human-Computer Interaction 32, 1: 63-85.

[54] Massimiliano Zecca, Yu Mizoguchi, K. Endo, F. Iida, Y. Kawabata, Nobutsuna Endo, Kazuko Itoh, and Atsuo Takanishi. 2009. Whole body emotion expressions for KOBIAN humanoid robot - preliminary experiments with different Emotional patterns. In RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication, 381-386.