Development of animated facial expressions to express emotions in a robot: RobotIcon

Lukáš Danev*, Marten Hamann†, Nicolas Fricke‡, Tobias Hollarek§ and Dennys Paillacho¶‖
Facultat d'Informàtica de Barcelona, Universitat Politècnica de Catalunya (UPC)
Jordi Girona 1-3. Barcelona, Spain
*lukas.danev@est.fib.upc.edu, † marten.hamann@googlemail.com,
‡ nicolas.fricke@student.hpi.uni-potsdam.de, § t.hollarek@mytum.de
‖Department of Automatic Control, Universitat Politècnica de Catalunya
Jordi Girona 1-3. Barcelona, Spain
dennys.paillacho@upc.edu
¶Electrical and Computer Engineering Department, CIDIS, Escuela Superior Politécnica del Litoral (ESPOL)
Km 30.5 via Perimetral. Guayaquil, Ecuador
dpailla@fiec.espol.edu.ec

***Abstract*—Robotics is a field which does not only become more and more important in industry and production but is becoming part of many peoples every day life and as such, the interaction between robots and humans is an essential field of research. In this work, we present a face – being an important part of a natural communication – and show that animated facial expressions can be used to express the robot's emotions. To do so, we developed a set of animated facial expressions to be shown on the screen of the MASHI robot, a telepresence robot with a wide-screen display mounted on top. A survey was conducted to evaluate the recognition value of the facial expressions as well as to evaluate whether displaying the full face or rather just the eye-region affects this recognition rate. The evaluation of this survey shows that most of our emotions are recognized better if a mouth is provided. Yet, the recognition rate is still acceptable without displaying the mouth. In summary, all emotions were well recognized and therefore, displaying these on robots like MASHI will contribute to a more natural human-robot interaction.**

Keywords*—human-robot interaction, social robotics, robot face, robot animated facial expressions, robot gestures

I. Introduction

Robots have been playing an important role mainly in industrial settings for quite some time now. These robots, however, have been set in a cage following a preimplemented execution plan. With the availability of high computing power even in cheap machines, it is becoming feasible to design intelligent agents that are able to adapt to their environment. In this new generation of robots the focus is on the ability of interacting with humans in a socially natural way [1, 2].To provide these robots with such a capability to express emotions in a simplified and easier readable manner, there can be represented facial expressions as emoticons or simplified avatars.

By now, it has been shown that our brain developed in a way in which we now react to emoticons in the same way as we would react to emotion expressing real faces [3]. This means that even very abstract ways of expressing emotions can be understood and are processed in a natural way by humans.

In Mirning et al. [4] the authors were able to show the effectiveness of facial expressions and using emotions in human robot interaction to improve the engagement of humans in the communication with the robot. In a game scenario they tested users playing against a robot that did not express emotions against the same robot with a mounted tablet screen on which a face was displayed that expressed emotions according to the game situations. They were able to show that the robot with a face was perceived as more intelligent and that the given task was perceived as more attractive to the participants if facial feedback was provided.

In [5], Russell *et al.* have developed a smart phone app that in its current status detects keywords and expresses emotions. The emotions are displayed at the smart phone screen using a single pair of animated eyes which can express six basic emotions. The app is specifically developed to give an expression mechanism to the MU-L8 robot where the phone is mounted to the head of the robot. However, the researchers have not evaluated empirically if the emotions they express are actually recognized by users as intended.

Another project uses a hardware head to express emotions where a thorough evaluation of the expressions has been conducted by evaluating a test scenario with elementary school children as participants [6]. Their results show that basic emotions can usually be recognized by a very high percentage of people whereas more subtle or complicated emotions like disgust or fear are harder to recognize out of context. One issue of this particular robot head is that the mouth itself cannot change. This might influence the poor recognition rates for some of the emotions and indicate that a mouth does enhance the expressiveness of a facial expression.

The aim of this work is to analyze the degree to which people can recognize comic-like facial expressions of a robot, especially when these facial expressions include or not the

This work has been partially supported by the Ecuadorian Ministry of Higher Education, Science, Technology and Innovation - SENESCYT through the Grant "Convocatoria Abierta 2012".

978-1-5386-3894-1/17/$31.00

mouth.

A description of the robotic platform where detailed in the next section. In section III, details about the software implementation of facial robot expressions where described. The evaluation of the facial expressions used to express robot emotions where developed in section IV. Finally, in section V conclusions and future works where provided.

II. Robotic Platform

Our solution will run on the MASHI platform. MASHI is a telepresence robot developed for the study of human-robot interaction [7].

A. Technical Description

MASHI currently is equipped with, among other devices, a screen which can display a website with WebRTC and has a camera as well as a microphone and speakers attached. It is an approximately human scale robot. Its base is a plattform of approximately half a meter diameter which has two motorized wheels to the right and the left of the plattform. In addition one non motorized wheel is attached in the front. On this plattform the controlling electronics for the wheels as well as a battery and a laptop are placed. Also on this plattform is a pipe that goes about 1.5m up. On its top the display, speaker and microphone are mounted. The display also has motors to more it right and left and tilt it around a small angle. So MASHI's screen is in the position where a human being would have his eyes. It is approximately 7" in size and is a wide angle screen.

B. Robot's Social Features

Our project focuses on MASHI's facial expressions. We are aiming at giving MASHI an own face, with which the robot can express emotions. These emotions can either be controlled by a human operator through a browser interface or they can be produced by MASHI autonomously.

In addition MASHIS most important social feature is his ability to move. Proxemics are a very powerful tool in influencing how a conversation is experienced on an emotional level for humans. Therefore it is very important for social robots to apply to these rules.

C. Appearance

The appearance of MASHI's face is important on different dimensions. On the one hand, the appearance of the face could be used to resemble someone the conversation partner knows or relates to. This way, emotions might be easier to read for the conversation partner, since an emotional bond exists. On the other hand, the complexity and the closeness to a real face conveys the impression that MASHI has the same capabilities as a real human being would have. Such impressions will be misleading, and therefore diminish the experience users will have when interacting with MASHI. Therefore, a face should be a stylized version of a real face. Since MASHI's screen is wide angled and in the position where a human would have the eyes, it needs to be evaluated whether it is more convenient for a conversation partner to show the entire avatar on the screen or rather only show the eyes with eyebrows and maybe the corners of the mouth. If the latter solution should prove more suitable, it needs to be evaluated whether these features are sufficient to display recognizable emotions.

D. Communicative Skills and Interactive Behaviour

MASHI is equipped with a speaker and a microphone, that allows the verbal communication between the teleoperator and the user. Furthermore, MASHI has nonverbal communication abilities. The teleoperator can send two types of visual information: A video stream showing his face or an artificial face, that is controlled by the teleoperator. With the video stream, it should be possible to transfer every facial expression or gestures like nodding or head-shaking. The artificial face can express several basic emotions mainly by using the mouth, the eyes and the eyebrows.

The emotions evaluated here are: happy, sad, angry, neutral, and sleepy. While the first three are arguably emotions which may effectively be used within a conversation and neutral is the facial expression made when no emotion is being displayed, the sleepy state was introduced to be able to display that MASHI's operator may be absent at the moment. This way, users may more easily understand why the robot does not respond immediately.

III. Implementation

The implementation has been built on top of an already existing architecture. The already existing implementation is built with NodeJS[1], which is a cross-platform runtime environment allowing for easy creation of server-side and network applications. NodeJS builts on the Google Chrome JavaScript API[2], for which the code is composed of JavaScript, HTML and CSS.

A. System Architecture

The system architecture is subdivided into an *operator part*, which runs on the operator's computer and provides the control interface for MASHI, and a *robot part* which runs on MASHI, receives the operator's commands and executes them. The implementation of the RobotIcon affects both of these components, since the operator needs to control MASHI's emotions, which then in turn need to be displayed on the robot.

Since the existing code base knew only one mode of operation – the telepresence mode, where the video stream of the operator is transmitted to and displayed on MASHI – both sides needed to be adapted to now support a second, new mode of operation: the RobotIcon mode, where an artificial face is being displayed on MASHI and it's appearance, in terms of the emotions being displayed, can be controlled from the operator side.

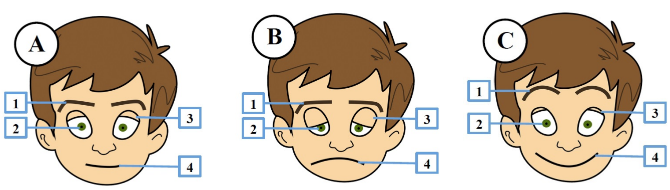
Two main tasks are done at the operator side - controlling what is displayed on the MASHI's face and comunicating it the robot side. Already at the current stage of the project there are a lot of choices for the user and it was a challenging task to design a graphical user interface which would not be overwhelming for the user. The GUI is built using HTML, JavaScript and CSS.

[1]NodeJS: http://nodejs.org/

[2]Google Chrome JavaScript: https://developer.chrome.com/extensions/api_other

The existing code was extended to support the control of the emotions as well as toggling between the two modes of operation. To implement the RobotIcon mode, and allowing for switching between the two modes without a user intervention required on MASHI, this architecture needed to be changed. Therefore, a toggling mechanism was implemented within the MASHI view. This mechanism is encapsulated in the $fullscreen.js$-file which is being loaded by the HTML page and is available through a global variable $fullscreenControl$ on the robot side. The initial view remains to be the known menu view, but by calling $fullscreenControl.setVideoFullscreen()$, the menu view is hidden and the operator's video stream fills the whole page. And by calling $fullscreenControl.setRoboticonFullscreen()$, the new RobotIcon view can then be activated, hiding the video completely and displaying only the artificial face on the full available screen. Therefore, if MASHI now is started, the operator only has to set the web browser on the robot's screen to fullscreen (with MASHI's current setup e.g. by pressing the $F11$-key). After this, all toggling of the operation modes can now be done programmatically from the operator's side.

B. Implementation of the RobotIcon



(a) (b) (c)

Fig. 1: Different expressions of the artificial face for MASHI with indication of the animatable regions. (A) shows a neutral expression, (B) a sad emotion, and (C) shows a happy emotion. Within the face, eleven different attributes can be altered for changing the emotion: (1) the shape and angle of each eyebrow, (2) the height of each eyelid, (3) the direction and intensity of the gaze of each eye, and (4) the shape of the mouth.

As explained in section II, the artificial face on MASHI should help users to relate to the robot, to feel empathy and strengthen social bonding. Therefore, it is crucial that the face looks and behaves human alike. Yet, we deliberately decided to use a face which is *alike* a human face, but in fact is not a real human face. We decided against actual pictures of faces and for simplified, comic-styled face. This not only allows for easier recognition of the expressed emotions, but also avoids conveying a wrong message – that the robot would have abilities only human beings have. Figure 1 shows the face we developed for MASHI. It has eleven attributes which can be altered to change the displayed emotion:

- the shape and angle of each eyebrow,
- the height of each eyelid,
- the direction and intensity of the gaze of each eye, and
- the shape of the mouth.

For being able to display such graphics within the web browser the MASHI interface runs on, different technologies can be used. The main three technologies we took into consideration were 1) interchangeable static images, 2) the HTML5 canvas[3] element, or 3) using a scalable vector graphic (SVG)[4]. Since we early on decided that for a more realistic emotional representation, the transitions between the emotions are an important aspect, the option of using static images was soon rejected. While allowing manipulation of the image via JavaScript, the HTML5 canvas element does not provide a scene graph. This has the effect, that the canvas does not know about any elements being represented – everything is just a pixel – which makes the manipulation of certain elements extremly complicated. Therefore, we decided to use SVG as the underlying technology.

SVG images are being defined by XML[5] files. Within an SVG file, paths, basic shapes – such as circles, ellipses, or rectangles –, or text can be defined. Attributes, like fill color or stroke width, to these elements can be defined for each single element and each element within the SVG can be directly addressed and manipulated.

For easier manipulation of the SVG DOM[6] tree, we employed the open source JavaScript library Snap.svg[7]. With this library, the properties of the elements within the graphic can easily be manipulated and animations can automatically be generated between two transition points.

Choosing SVG as the underlying technology also allows us to easily change other attributes within the image – such as eye- or hair color. Also adding a completely different face, for example a female character, later in the development process is possible without having to dramatically change the existing logic. Such a new face would simply need to be a seperate SVG file with the same structure as the existing file.

The $roboticon.js$ library was implemented to encapsulate the logic for the face manipulation. It has to be included on the HTML page displaying the RobotIcon. The RobotIcon SVG graphic is loaded into the HTML page using the HTML5 $object$ element. When loaded, the $roboticon.js$ library searches for the $object$ element with the $roboticon$ class set. This element then is being manipulated by the library. To now interact with the RobotIcon, the library exposes the globally accessible variable $RobotIcon$, which in turn exposes methods to animate the RobotIcon.

We have decided to use JSON[8] as the description language for the facial expressions. The format of the JSON file is explained in section III-C. The JSON can then be applied to the artificial face by calling $RobotIcon.parseAndApplyJson(jsonString)$, which then applies the new facial expression values to the displayed RobotIcon and generates a smooth animation between the previous and the new emotion.

To make the RobotIcon seem more human alike, we have implemented a time interval based eye blinking. We have

[3]Canvas is an HTML element, which represents a bitmap surface and can be altered using JavaScript. See http://www.w3.org/TR/html-markup/canvas.html for more information.

[4]Find more on SVG here: http://www.w3.org/Graphics/SVG/

[5]Extensible Markup Language

[6]Document Object Model

[7]Snap.svg homepage: http://snapsvg.io/

[8]JavaScript Object Notation, find more under: http://www.json.org/

found, that communication partners engage more and that the emotions expressed benefit from such details. This interval can later also be used to even better express emotions, like blinking more often when a nervous emotion is shown.

As described in section II-C, it needs to be evaluated whether the full face or rather only the eye-region should be displayed on MASHI's display. This has been analyzed in a user study, which is being presented in section IV. In the implementation we therefore have added a toggling mechanism, which easy switching between the two modes. By calling $RobotIcon.changeDisplayMode(mode)$ and passing either string $"whole\_face"$ or $"eyes\_only"$ as mode parameter, the view mode can be switched. The RobotIcon remains it's facial expression even if the view mode is switched within the operation mode.

Section III-D will give a description of the user interface the operator can use to control all the parameters changing the face' appearance.

C. Communication

Based on the in section III-A described architecture, we had to design a way to communicate the new commands from the operator side to the robot. Building on the already existing code, this is being done with the NodeJS socket connection. To now transmit emotions, we have, as already mentioned in the previous section III-B, decided to use JSON as a way of describing the parameters of the facial expression. Since JSON is represented as a string it can be transmitted using the already existing architecture.

D. Graphical User Interface on Operator Side

In the graphical user interface (GUI) on the operator side, a small window on the top left corner shows what actually is being displayed on MASHI. This way, the operator can assess what users actually see when interacting with the robot. Beneath that, the operator can choose between different operating modes. In the *Menu* mode, MASHI shows a menu, where the user can select different settings or send and receive text messages to or from the operator. In the *Operator* mode, MASHI displays the video, that is recorded of the operator in fullscreen.

The third mode displays an artificial face on MASHI's screen. In this mode, the operator can choose, if the whole face or only the region around the eyes should be displayed. Showing only the eye region of the face makes the appearance more consistent with the hardware of MASHI, because of MASHI's widescreen display and the speaker being located under the display. A discussion of which mode should rather be used will be presented in section IV. Despite the display mode, the user can choose from six different emotions (happy, sad, angry, uncertain, neutral, sleepy) to be displayed and adjust the intensity of the current emotion. The neutral emotion can be used as the basic emotion during communication. The happy, sad, angry and uncertain emotions are very expressive and useful to show reactions to specific inputs. Sleepiness can be expressed, when the robot is currently not controlled by the operator, because he is busy or not at the operating computer. The intensity slider gives the operator an additional degree of freedom.

The operator can also control the robot's movements with either four different buttons or alternatively four keys on the keyboard. The keyboard should be preferred, when the operator is using a computer. The buttons can be used, when a smartphone or a tablet is used to control the robot. The robot can only be controlled, when the *Control robot* mode is turned on. If this mode is turned of, it is possible to send text messages to MASHI's display. The largest window of the GUI shows the video record from MASHI's webcam. It has to be large, because it makes navigation easier.

IV. Evaluation

In this evaluation we focus on to which extend the emotions we created are recognizable. In order to reach wide public and scale the study, we decided to use an online questionnaire. In our study we also try to answer the question what difference the displaying of the full face opposed to the eye-only region makes in terms of recognizability. This is an important information which could contribute to whether to show the full face with the mouth or whether to show only area around the eyes and therefore whether to suggest an alteration of the hardware setup of MASHI.

A. Methodology

The study was conducted online through questionnaire on Google Forms[9]. It was convenient for the participants and it gave us more information than we would have been able to gather in a manually conducted study. Apart from the basic information about participants, the questionnaire had four major parts. Each participant was asked for gender, nationality, age group and whether the participant have prior knowledge about Human Robot Interaction or have worked in the field of Human Robot Interaction. In each of the four major parts were participants presented with pictures showing five basic emotions and were asked to recognize them. The basic emotion shown on the pictures were namely happy, sad, sleepy, neutral and angry.

The main study then had four parts. In the first part, participants were shown pictures displaying only the area around the eyes while the mouth was not visible (see the figure 2a). Participants were asked to recognize emotion shown on the picture and write it into a free-text field. In the second part, participants were shown exactly the same pictures as in the first part but this time the participants were presented also with a list of our five basic emotions. Participants were asked to select how strongly the picture expresses each emotion on scale from 1 to 7. The full scale from 1 to 7 was translated into "not at all", "very little", "a bit", "some", "visibly", "strongly" and "very strongly" respectively (see the figure 3).

The third and fourth part were analogous to the first and second, with the difference that now the full face including the mouth was shown (see the figure 2b). This way, we can see how important the mouth is for recognizing the correct emotion. The drawback of this ordering is that participants were primed in the free-text part three to write the emotions which they saw as the choices in part two. This is a trade-off we chose deliberately, because if we would have chosen

[9]Google Forms: http://www.google.com/forms/about/

an ordering in which first the two parts are free text followed by two parts with choices, we would prime participants by showing them mouth before letting them choose how strongly are emotions expressed only by eyes.

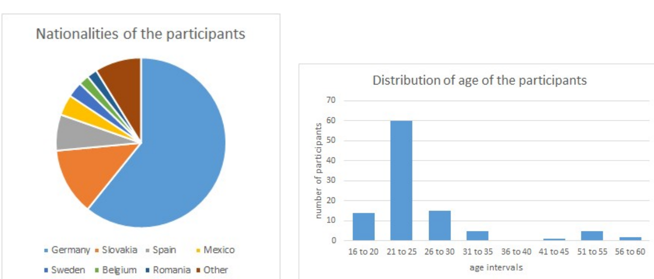
(a) Eyes only (b) Full face

Fig. 2: Face area shown in the questionnaire.

Fig. 3: Choices of emotions and intensities in the questionnaire for part two and four

B. Data analysis

Over one hundred participants[10] took part in our study. One third of our participants were female and two thirds were male. About one fifth of our participants had a prior knowledge in the field of human robot interaction. Our study attracted people from 16 countries but the most participants come from Germany, Slovakia and Spain which together counts for over 80% of our participants. The distribution of nationalities is shown as pie chart in figure 4a. The most numerous age group of our participants was from 21 years to 25 years. The detailed distribution of age groups of our participants is shown in figure 4b.



(a) (b)

Fig. 4: Group distribution: (a) by nationality and (b) by age

[10]In total, 102 participants took part in the study.

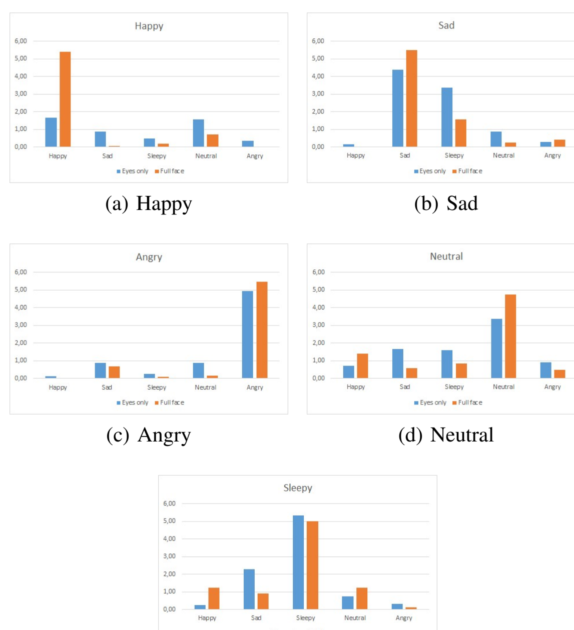
From a first analysis of the free-text answers a three to four times higher diversity was given when participants were presented with the eyes only part in comparison to the full face part. Table I shows the percentages of people who entered a word in the free-text field recognizing the displayed emotion correctly, with eyes only part in the first line and the full face in the second line. At a first glance, very high values for the full face, and slightly smaller values for the eyes only part can be observed. An outlier can be found when looking at the happy emotion with the eyes only being displayed.

TABLE I: The percentages of people who correctly recognized each emotion in free text parts

	Happy	Sad	Sleepy	Neutral	Angry
Eyes only	1%	51%	63%	28%	65%
Full face	92%	97%	86%	86%	97%

The figures 5a and 5b, as well as 5c, 5d, and 5e show the average values of how strongly each emotion was said to be present in each picture. The higher the bar is, the more the emotion is being recognized within the picture. The blue columns show the results from the eyes only part while the orange columns correspond to the full face expression.

For the happy emotion there was the biggest difference between showing the mouth or nor. Actually over 80% of participants wrote surprise instead of happy in free text answers for eyes only part. By showing the smiling face the performance improved rapidly from 1 percent to 92 percent and only 3 percent of participant wrote surprise when there was a mouth. One can see that it is difficult recognize happy emotion also by looking at the big difference in size of the bars in figure 5a.



(a) Happy (b) Sad (c) Angry (d) Neutral (e) Sleepy

Fig. 5: How strong is each emotion expressed.

The sad emotion turn out to be easily confusable with sleepy emotion. Although the sadness was the most popular answer in free text part, it was closely followed by the tired and bored. Again the addition of the mouth to the sad eyes removed the doubts of which emotion is expressed. We can see it both from the improvement by 46% in free text part and also by much smaller orange bar compare to blue bar for sleepy emotion in figure 5b. Nonetheless we can see that the sad emotion is strongly present in both cases.

For the sleepy emotion there was the smallest difference between the eyes only and the full face. It is the only one emotion which had stronger presence of sleepiness without the mouth than with mouth as can be seen on figure 5e. The main reason for that is that our mouth for sleepy emotion was not expressing the right emotion. Some kind of open or yawning mouth would be much better choice than just a single line which is currently used as mouth.

The neutral emotion earned the position of the worst recognizable emotion. In the free text part for eyes only there was the highest variety of answers and all together over 60 different ones. When the participants were presented with the choices (see the figure 5d) the performance was much better as others emotion were not present there either.

The angry emotion was on the other hand the best recognizable emotion. Almost in all metrics it has the highest performance. Only from eyes region 65% of participants correctly recognized the emotion and with the addition of the mouth the result was close to hundred. Also from figure 5c can be seen that the presence of other emotion is negligible compare to the angry emotion.

To sum up we are pleased with the results of the study. It confirms that the emotion can be recognized only from the region around the eyes and with the addition of the mouth we got very accurate results. The most difficultly recognizable emotion is the happy emotion when the mouth is not show and in both case the neutral emotion. The angry emotion was the most recognizable.

V. Conclusion

In the context of this work, we have developed an artificial comic face to be displayed on the teleoperated robot MASHI, which is supposed to act as a guide in a community center. It is possible to display six different emotions and adjust the corresponding intensity using the face. We compared to different viewing modes, where we either displayed the whole face or only the eye region of the face. Our evaluations has shown, that it is highly beneficial to display the whole face in order to transmit all emotions properly, especially for showing happiness. However with very careful design and further evaluation with users it is possible to express at least some emotions using only the eyes.

The design of the system over an open source architecture allows future improvements. One possible improvement would be to add more emotions to allow a more distinct facial communication, which could be realized easily by using our interface for generating emotions. Another improvement could be made by adding the function of individualizing the appearance of the face from the operator side, because in the current version there is only one certain male face available which might not correspond to the operator's appearance.

Despite that, alternative user interfaces should be evaluated. One possibility would be to generate emotions without buttons, but by representing the operator's emotions autonomously.

The social presence of the robot could be even further enhanced by adding a motion component to the face like nodding or tilting the face. Using the eyes to focus the user or other objects could also make the robot more socially present. Also, gestures could be employed to emphasize emotions.

References

[1] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 143 – 166, 2003, socially Interactive Robots. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092188900200372X

[2] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: A survey," *Foundations and Trends in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2008. [Online]. Available: http://dx.doi.org/10.1561/1100000005

[3] O. Churches, M. Nicholls, M. Thiessen, M. Kohler, and H. Keage, "Emoticons in mind: An event-related potential study," *Social Neuroscience*, vol. 9, no. 2, pp. 196–202, 2014, pMID: 24387045. [Online]. Available: http://dx.doi.org/10.1080/17470919.2013.873737

[4] N. Mirnig, Y. K. Tan, T. W. Chang, Y. W. Chua, T. A. Dung, H. Li, and M. Tscheligi, "Screen feedback in human-robot interaction: How to enhance robot expressiveness," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE, 2014, pp. 224–230.

[5] E. Russell, A. Stroud, J. Christian, D. Ramgoolam, and A. B. Williams, "Smile: A portable humanoid robot emotion interface," in *9th ACM/IEEE International Conference on Human-Robot Interaction, Workshop on Applications for Emotional Robots, HRI14, Bielefeld University, Germany*, 2014.

[6] J. Kedzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz, "Emys emotive head of a social robot," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 237–249, 2013.

[7] D. Paillacho, C. Angulo, and M. Díaz-Boladeras, "An exploratory study of group-robot social interactions in a cultural center," in *IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2015 Workshop on Designing and Evaluating Social Robots for Public Settings*, 2015, pp. 44–48.