



# Affect Display Recognition Through Tactile and Visual Stimuli in a Social Robot

Sara Marques-Villarroya<sup>(✉)</sup>, Juan Jose Gamboa-Montero,  
Cristina Jumela-Yedra, Jose Carlos Castillo, and Miguel Angel Salichs

University Carlos III, Madrid, Spain  
{smarques, jgamboa, jocastil, salichs}@ing.uc3m.es

**Abstract.** New technologies are nowadays an important part of human communication and interaction. While text, facial, and voice recognition have become increasingly fluid in recent years, thanks to the development of machine learning algorithms, recognising and expressing sensations or moods via multimodal recognition is a field that the literature could further explore. This situation introduces a new challenge to social robots. In this work, the authors study how a combination of visual and tactile stimuli influences people's perceptions of affect display and seeks to apply these findings to a social robot. In the experiments, the subjects had to determine the perceived valence and arousal of simultaneously being exposed to the two stimuli mentioned above. The analysis revealed that the combination of touch and facial expression significantly influences the valence and arousal perceived by users. Based on these findings, this work includes an application for the robot to determine the user's *affect display* in real-time.

**Keywords:** Social robotics · Multimodal interaction · Affect display recognition · Affective touch · Facial expression

## 1 Introducción

The way humans communicate has changed in recent years with the emergence of new technologies. As a result, there is a need to enhance the interactions between humans and these technologies. Recognising and expressing emotions by analysing the perceived stimuli constitutes another step toward achieving a natural interaction. As Beale and Peter studied, emotions are produced in interpersonal relationships after the first few interactions, implying that it is a gradual process that takes time [3]. As a result, when discussing devices with which the user will interact, the ability to perceive emotions is an added value because it may generate a sense of trust. This feature becomes essential when discussing personal assistance or education applications. In this sense, social robots stand out among those devices with educational or assistive care functions.

According to Henschel et al. [10], “a social robot must be able to interact bidirectionally, display thoughts and feelings, be socially aware of its surroundings, provide

*social support and demonstrate autonomy*". With these considerations in mind, to make a robot *socially aware of its surroundings* and thus *interact bidirectionally*, it appears necessary to equip such devices with the ability to recognise the user's *affect display*: the expression of the user's internal emotional estate<sup>1</sup>. Based on this drive to improve social robots, the main goal of this work is to study how a combination of visual and tactile stimuli influences people's perceptions of affect display and how to apply these findings to a social robot. Specifically, we propose an application that recognises the perceived user's affect display.

With respect to the works in the literature focused on recognising human reactions to stimuli, Diekhoff et al. [6] examined how certain images with fearful facial expressions created a bias in participants that altered their perception of emotion recognition in neutral faces. Vasconcelos et al. [19] investigated the accuracy with which experimentees recognised vocal emotions from nonverbal human vocalisations. Regarding tactile stimuli, it is worth mentioning the study by Tsalamal et al. [18] in which the authors evaluated the influence of a haptic stimulus on visual stimuli. To do so, participants indicated the valence level suggested by various facial expressions. At the same time, a stream of air was applied with varying degrees of intensity to their left arm. The authors concluded that the tactile stimuli significantly influenced the experimentees' valence perception.

When considering how to capture the user's affect display during human-robot interaction, we discover that much of the literature focuses on visual and auditory stimuli. Huang et al. [11], for example, attempted to recognise emotions during human-computer interaction by combining facial detection with an analysis of the user's electroencephalographs. Similarly, Breazeal et al. [4] investigated the recognition of a user's affective communicative intent without focusing on the prosodic patterns of the speech. Finally, despite being scarce, research such as that of Yohanan [20], Altun [1], Andreasson [2] or Teysser [17] validate the relevance of tactile stimuli analysis when analysing the user's affect display using a social robot.

The remainder of the paper is structured as follows: The methodology used to obtain the data used in this study is shown in Sect. 2, and the results are presented and discussed in Sect. 3. Section 4 describes the integration of an affect display recognition application in a robotic platform using the data gathered in the previous sections. Finally, Sect. 5 highlights and discusses the main findings of this work.

## 2 Experimental Study

To endow a social robot with the ability to respond to the user's affect display, we must first understand how people perceive those same stimuli. In a typical interaction environment, stimuli tend to appear grouped rather than individually. As a result, evaluating just a stimulus alone could lead to inaccurate results. Based

---

<sup>1</sup> In this work, we will use the definition of affect display introduced by Yohanan et al. [20]. We must clarify that the authors acknowledge that this expression could be faked, but these nuances are out of the scope of the paper.

on this premise, a study was planned to collect and analyse the valence and arousal perceived by users when exposed to the target stimuli simultaneously. The visual ones were presented through the appearance of different images on a screen, while the experimenter provided tactile stimuli to make it appear as natural as possible. The users then input their perception of the valence and arousal level produced by these two stimuli into a graphical user interface specifically designed to automate the data gathering and ease the subsequent analysis.

## 2.1 Conditions and Stimuli Studied

We define seven kinds of touch stimuli in this study based on their duration, intensity, and form. We chose them following the ideas from the article by Silvera et al. [16], which condenses Yohanan’s [20] gestures into the six most essential touches during HRI. To adapt this list to the social robot (see Sect. 4.1), we removed the ‘push’ gesture because it is irrelevant when interacting with our desktop robot, and ‘pat’ for being almost imperceptible using the touch gesture detector introduced in the same section. We also added three more types of contacts considered interesting in HRI: ‘tickle’ and ‘rub’ frequently appear in everyday interactions, such as those with a pet. We also added ‘hit’ despite its negative connotation since we expected it to have more extreme valence and arousal values, which could help to have a more diverse set of gestures. Table 1 summarises the set of touch gestures used in the experiment along with comprehensive definitions.

**Table 1.** Definitions of the touch gestures used for this experiment.

Gesture	Definition
Stroke	Move your hand with gentle pressure
Rub	Move the hand repeatedly with firm pressure
Tickle	Touch with light finger movements
Scratch	Rub with the fingernails
Tap	Strike the with a quick light blow or blows using one or more fingers
Slap	Quickly and sharply strike with an open hand
Hit	Deliver a blow with either a closed fist or the side or back of your hand

Regarding facial expressions, we will use Paul Ekman’s six basic emotions [7] for the simple expressions, also adding a ‘neutral’ one. The following expressions with their abbreviations were used in this study: angry (AN), afraid (AF), disgusted (DI), sad (SAD), neutral (NE), surprised (SU), and happy (HAP). In this experiment, we used images from the Karolinska Directed Emotional Faces (KDEF) database [5] Combining the sets of touch and facial stimuli, we obtained a total of 49 unique combinations. To eliminate bias, we created five cases, each made up of 20 randomly chosen touch and face combinations. Each users was presented one of these cases, trying to ensure balance among cases instances for our dataset.

## 2.2 Experimental Setup

The study on affect display included 50 subjects, 29 of them were male and 34 were under 30 years old. None of the participants had any prior knowledge of the experimental procedure, user interface, or any of the images shown during the study. Participants were exposed to the two types of stimuli at the same time: A picture of a person’s face with a specific facial expression appeared on the application screen (see Fig. 1), and, simultaneously, the experimenter performed a touch gesture on the user’s left arm. The experimenter was behind an opaque screen, and their arm was covered with a surgical glove and a long sleeve to prevent the subject from guessing their age or gender.

As Fig. 1 shows, the results of valence and arousal levels are plotted on the X and Y axes, representing Russell’s circumplex [13]. Both levels have a range of  $-100$  to  $100$ . The  $-100$  scale represents the most unpleasant in terms of valence and the most relaxing in terms of arousal, whereas  $100$  represents a very pleasant and high arousal level. To modify the values of valence and arousal, the interface included two sliders attach to each axis, which the user could move freely. After that, the user pressed the “OK” button to continue to the next pair of stimuli. The experiment lasted five to seven minutes on average, with 20 image and touch combinations performed in each case.



Fig. 1. Graphic interface designed for the experiment.

## 3 Analysis of Results

The goal of the general analysis of the results from the tests performed on the 50 users is to find a relationship between tactile and visual stimuli and the levels of valence and arousal. First, we ensured that all data had a normal distribution using the Shapiro-Wilk method [15]. Then, we performed an ANOVA analysis, which allowed to compare the differences between the means of the different groups. In our case, by performing an ANOVA on the influence of the combination of touch and expression on the value of valence and arousal, we discovered that the combination of the two stimuli had a significant impact ( $p < 0.05$ ) on the affect display perceived by the user (both in valence and arousal). Similarly, we investigated whether the interaction of the stimuli’s combination with the

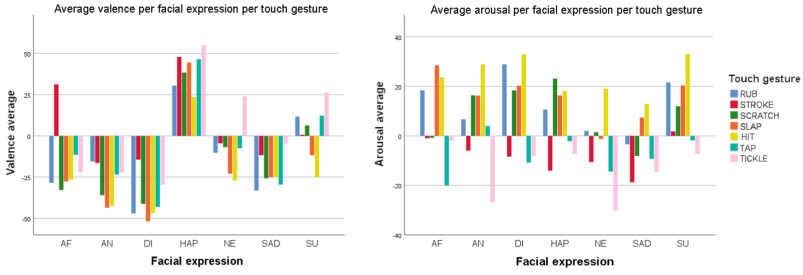
**Table 2.** Results obtained with the multivariate ANOVA study.

Interaction group	Variable	p-value
Stimuli combination	Valence	<b>1,8896E-84</b>
	Arousal	<b>6,6621E-19</b>
Stimuli * age	Valence	0,442862
	Arousal	0,399626
Stimuli * gender	Valence	0,496526
	Arousal	0,942287
Stimuli * age * gender	Valence	0,539040
	Arousal	0,489812

participants' age (under/over 30 years old) and/or gender influenced their perception of the stimuli. The ANOVA on this interaction produced non-significant results ( $p > 0.05$ ). Table 2 shows the outcomes of the ANOVA study.

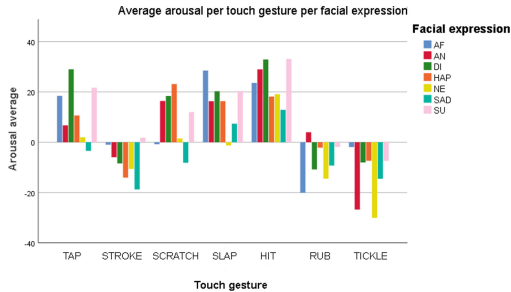
With these findings, we obtained the means for each combination of stimuli, yielding the results depicted in Fig. 2. These graphs show the mean valence (left graph) and arousal (right graph) obtained for each gesture and facial expression combination. The ANOVA analysis shows that the combination of stimuli significantly influences valence and arousal; however, there are no significant differences when the users' age and/or gender are taken into account. Looking at the results on the left side of Fig. 2, which shows the average valence obtained in each combination, we can see that the face emotions 'afraid', 'angry', 'disgusted', 'neutral' and 'sad' have primarily negative values, outweighing the tactile information. These results are consistent with the fact that these facial expressions are commonly associated with negative emotions. However, in the case of the 'afraid' face, the valence obtained from the 'stroke' gesture is positive. Therefore, while facial expressions are relevant in the perception of affect display, they can be affected by the contact performed at that moment, turning an unpleasant feeling into a pleasant one. The same effect can be seen with the 'happy' expression, which aids in perceiving all gestures as pleasant. We can see, however, that the more abrupt gestures, such as 'hits', achieve a lower level of valence than the rest of the touches studied. In the case of the 'surprised' facial expression, we can see diverse outcomes. Because the level of valence of 'surprised' emotion in Russell's circumflex is low, it can be considered as pleasant or unpleasant expression depending on the user. In this case, where the facial expression is unimportant, we can see how the touch gestures significantly modulate the valence, ranging between 26 and  $-25$ .

Complementarily, on the right side of Fig. 2, we can see how the arousal results in more uneven values for each facial expression. For this reason, we decided to group the results by the kind of touch gesture instead to try to find some patterns, which resulted in Fig. 3. The figure shows that looking at the touch gestures, the results are more aligned, implying that for the arousal variable, the type of gesture is more significant than the facial expression, in contrast to the data obtained with the valence. In this case, we can see that the 'tap', 'scratch', 'slap', and 'hit'



**Fig. 2.** Average values of valence (left) and arousal (right) gathered in the experiment. The horizontal axis shows the facial expressions afraid (AF), angry (AN), disgusted (DI), happy (HAP), neutral (NE), sad (SAD) and surprised (SU).

gestures are primarily positive, whereas the ‘stroke’, ‘rub’, and ‘tickle’ gestures are mainly negative. These outcomes are linked to the definitions of each of the gestures. While ‘tap’, ‘scratch’, ‘slap’, and ‘hit’ are gestures that involve applying pressure to the user’s arm, where the intensity is brief but intense, ‘stroke’, ‘rub’, and ‘tickle’ imply a soft gesture on the user with less pressure, resulting in a negative arousal value. In this analysis, we also noticed that, as with valence, the visual stimuli have some influences on the user’s perception. In the case of ‘tap’, for example, we see that arousal drops to negative values in the presence of ‘sad’ facial expressions, just as it does with ‘scratch’. Finally, we created the *affect\_display* database with all the valence and arousal results, which the robot will use to estimate the user’s affect display.



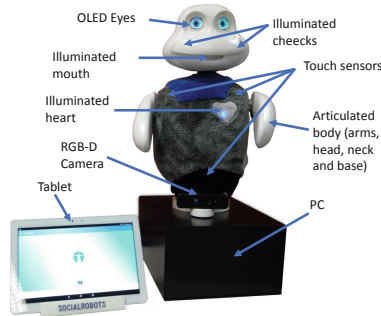
**Fig. 3.** Average arousal values (y-axis) as a function of touch gesture (x-axis) and facial expression (color).

## 4 Integration in a Social Robot

This section describes an application that allows the robot to recognise and respond to various communicative intentions expressed by the user. This application was created using the results presented in Sect. 3. The current section contains a brief description of the robotic platform, the designed application, and some preliminary results.

## 4.1 The Robotic Platform

The Mini robot, developed by the UC3M RoboticsLab [14], was originally conceived to perform cognitive stimulation and companionship tasks with elderly people. The robot integrates a series of social skills, such as playing different games, storytelling, and making jokes. It can interact with the user by proactively proposing activities based on user preferences, learning from their tastes, and adapting to them.



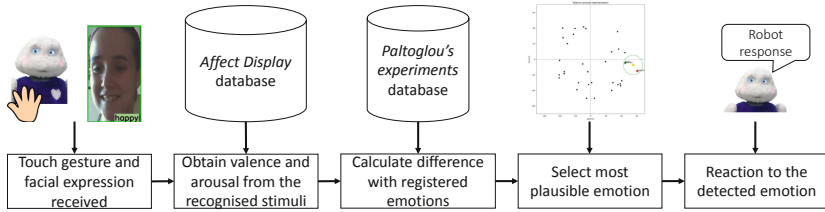
**Fig. 4.** The social robot mini.

The Mini robot has OLED screens in its eyes that allow it to look in different directions and express emotions. It also has LED lighting on the cheeks, mouth, and heart to make it more expressive. Mini has five motors that allow it to move its arms, head, neck, and base (see Fig. 4). It has piezoelectric microphones and capacitive sensors on its arms and belly to detect tactile stimuli. As for perceiving visual stimuli, it has an RGB-D camera on its base.

## 4.2 Design of an Application for Affect Display Recognition and Reaction

Figure 5 shows the application flowchart developed to recognise the users' affect display and react accordingly. For stimuli detection, the robot uses, on the one hand, the detector developed by Gamboa et al. [8] for touch gesture detection and, on the other hand, for facial expression recognition, the *emotions-recognition-retail-0003*<sup>2</sup> detector, based on the neural network developed by Intel. When the robot detects both stimuli, it attempts to recognise the user's affect display by loading the data from the *affect\_display* database.

<sup>2</sup> Emotion recognition network: [https://docs.openvinotoolkit.org/latest/\\_models\\_intel\\_emotions\\_recognition\\_retail\\_0003\\_description\\_emotions\\_recognition\\_retail\\_0003.html](https://docs.openvinotoolkit.org/latest/_models_intel_emotions_recognition_retail_0003_description_emotions_recognition_retail_0003.html).



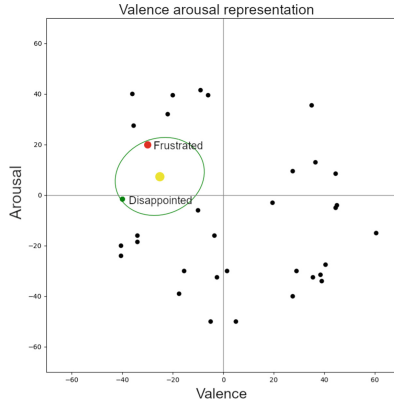
**Fig. 5.** Flow diagram representing the affect display recognition skill we propose in this work.

We decided to derive the 2-dimensional coordinates (valence and arousal) of the 35 emotions described in Russell's circumplex [13] from the works of Gobron et al. [9] and Paltoglou et al. [12]. Then, we calculate the Euclidean distance between the current valence and arousal values and those obtained in Paltoglou's experiments. Furthermore, we broaden the search area by leveraging detector uncertainty. Based on the results, we adjust the valence search range based on the confidence of the facial expression detector. The confidence of the touch detector, on the other hand, is used to rescale the arousal axis. Figure 6 depicts an example of the detector output when attempting to recognise the user's affect display with a tactile gesture 'slap' and a facial expression 'sad' with 75% and 90% confidence, respectively. In black, we see the 35 possible emotions from Paltoglou's experiment, and in yellow, the point obtained from our experiments with the perceived stimulus combination. The red dot represents the closest affect display, and therefore the one selected by the robot. The green dot represents the user's potential affect displays. Finally, the green ellipse represents the robot's search area. We use the distance between the yellow point and the closest emotion as the initial radius, and the ellipse's angle corresponds to the angle between the yellow and green dots. Then, we added the detectors' uncertainty, with a weighted Y-axis from the touch detector confidence and an X-axis from the vision detector confidence. Because the touch detector's confidence is lower in the example, the Y-axis is longer than the X-axis.

Finally, the robot will select the perceived emotion and react to it verbally. To filter possible errors of the detector, the robot notifies the user if there are more than five possible emotions within the search ellipse, which is more than 15% of options from which it can select. In this case, the robot informs the user that it does not know the emotion the user is conveying. We recorded a video<sup>3</sup> to demonstrate the social robot recognising the affect display of the user.

<sup>3</sup> Working example video: <https://youtu.be/jrv8bY0ssUI>.





**Fig. 6.** Outcome of one of the searches conducted during the robot tests as a result of a combination of a ‘slap’ and a ‘sad’ face (yellow dot). The selected emotion is “frustrated” (red point) is ‘dissappointed’ (green dot). (Color figure online)

## 5 Conclusions

This paper studies how a combination of visual and tactile stimuli influences people’s perceptions of affect display and seeks to apply these findings to a social robot. In this case, we experimented with 50 users to determine the perceived valence and arousal when simultaneously exposed to a combination of seven touch gestures and seven facial expressions. The data analysis revealed that the combination of touch and facial expression significantly affects the valence and arousal perceived by users ( $p < 0.05$ ). Specifically, the analysis showed that facial expression had more influence over the perceived valence, while the touch gesture had more impact on the arousal. Based on these results, we developed an application for the robot to determine the user’s affect display at any given time.

In future research, the number of users will be increased to conduct a more generalised study, emphasising the cultural differences between subjects. In addition, we plan to incorporate a machine learning system based on a regressor to predict the affect display more robustly, thus avoiding to rely on the mean values to make the estimation.

**Acknowledgements.** The research leading to these results has received funding from the projects: Robots Sociales para Estimulación Física, Cognitiva y Afectiva de Mayores (ROSES), RTI2018-096338-B-I00, funded by the Ministerio de Ciencia, Innovación y Universidades; Robots sociales para mitigar la soledad y el aislamiento en mayores (SOROLI), PID2021-123941OA-I00, funded by Agencia Estatal de Investigación (AEI), Spanish Ministerio de Ciencia e Innovación; the project PLEC2021-007819, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, and RoboCity2030-DIH-CM, Madrid Robotics Digital Innovation

Hub, S2018/NMT-4331, funded by “Programas de Actividades I+D en la Comunidad de Madrid” and cofunded by the European Social Funds (FSE) of the EU.

## References

1. Altun, K., MacLean, K.E.: Recognizing affect in human touch of a robot. *Pattern Recogn. Lett.* **66**, 31–40 (2015)
2. Andreasson, R., Alenljung, B., Billing, E., Lowe, R.: Affective touch in human-robot interaction: conveying emotion to the nao robot. *Int. J. Soc. Robot.* **10**(4), 473–491 (2018)
3. Beale, R., Peter, C.: The role of affect and emotion in HCI. In: Peter, C., Beale, R. (eds.) *Affect and Emotion in Human-Computer Interaction*. LNCS, vol. 4868, pp. 1–11. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-85099-1\\_1](https://doi.org/10.1007/978-3-540-85099-1_1)
4. Breazeal, C., Aryananda, L.: Recognition of affective communicative intent in robot-directed speech. *Auton. Robot.* **12**(1), 83–104 (2002)
5. Calvo, M.G., Lundqvist, D.: Facial expressions of emotion (kdef): Identification under different display-duration conditions. *Behav. Res. Methods* **40**(1), 109–115 (2008)
6. Diekhof, E.K., Kipshagen, H.E., Falkai, P., Dechent, P., Baudewig, J., Gruber, O.: The power of imagination-how anticipatory mental imagery alters perceptual processing of fearful facial expressions. *Neuroimage* **54**(2), 1703–1714 (2011)
7. Ekman, P.: Basic emotions. *Handbook Cogn. Emotion* **98**(45–60), 16 (1999)
8. Gamboa-Montero, J.J., Alonso-Martin, F., Castillo, J.C., Malfaz, M., Salichs, M.A.: Detecting, locating and recognising human touches in social robots with contact microphones. *Eng. Appl. Artif. Intell.* **92**, 103670 (2020)
9. Gobron, S., Ahn, J., Paltoglou, G., Thelwall, M., Thalmann, D.: From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text. *Vis. Comput.* **26**(6), 505–519 (2010)
10. Henschel, A., Laban, G., Cross, E.S.: What makes a robot social? a review of social robots from science fiction to a home or hospital near you. *Current Robot. Reports* **2**(1), 9–19 (2021)
11. Huang, Y., Yang, J., Liu, S., Pan, J.: Combining facial expressions and electroencephalography to enhance emotion recognition. *Future Internet* **11**(5), 105 (2019)
12. Paltoglou, G., Thelwall, M.: Seeing stars of valence and arousal in blog posts. *IEEE Trans. Affect. Comput.* **4**(1), 116–123 (2012)
13. Russell, J.A.: A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**(6), 1161 (1980)
14. Salichs, M.A., et al.: Mini: a new social robot for the elderly. *Int. J. Soc. Robot.* **12**(6), 1231–1249 (2020)
15. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). *Biometrika* **52**(3–4), 591–611 (1965)
16. Silvera-Tawil, D., Rye, D., Velonaki, M.: Interpretation of social touch on an artificial arm covered with an eit-based sensitive skin. *Int. J. Soc. Robot.* **6**(4), 489–505 (2014)
17. Teyssier, M., Bailly, G., Pelachaud, C., Lecolinet, E.: Conveying emotions through device-initiated touch. *IEEE Trans. Affect. Comput.* **13**, 1477–1488 (2020)
18. Tsalamlal, M.Y., Amorim, M.A., Martin, J.C., Ammi, M.: Combining facial expression and touch for perceiving emotional valence. *IEEE Trans. Affect. Comput.* **9**(4), 437–449 (2016)

19. Vasconcelos, M., Dias, M., Soares, A.P., Pinheiro, A.P.: What is the melody of that voice? probing unbiased recognition accuracy with the montreal affective voices. *J. Nonverbal Behav.* **41**(3), 239–267 (2017)
20. Yohanan, S., MacLean, K.E.: The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature. *Int. J. Soc. Robot.* **4**(2), 163–180 (2012)