Author's version. The final publication is available at link.springer.com

Bodily Mood Expression: Recognize Moods from Functional Behaviors of Humanoid Robots

Junchao Xu[1], Joost Broekens[1], Koen Hindriks[1], Mark A. Neerincx[1,2]

[1]Intelligent Systems, EEMCS, Delft University of Technology, 2628 CD, Delft, the Netherlands
{junchaoxu86, k.v.hindriks, joost.broekens}@gmail.com
[2]Human Factors, TNO, 3769 DE, Soesterberg, the Netherlands
mark.neerincx@tno.nl

Abstract. Our goal is to develop bodily mood expression that can be used during the execution of functional behaviors for humanoid social robots. Our model generates such expression by stylizing behaviors through modulating behavior parameters within functional bounds. We have applied this approach to two behaviors, waving and pointing, and obtained parameter settings corresponding to different moods and interrelations between parameters from a design experiment. This paper reports an evaluation of the parameter settings in a recognition experiment under three conditions: modulating all parameters, only important parameters, and only unimportant parameters. The results show that valence and arousal can be well recognized when the important parameters were modulated. Modulating only the unimportant parameters is promising to express weak moods. Speed parameters, repetition, and head-up-down were found to correlate with arousal, while speed parameters may correlate more with valence than arousal when they are slow.

Keywords: mood expression, nonverbal behavioral cues, body language, social robots, human robot interaction (HRI).

1 Introduction

Nonverbal expression of affect, as a key ability of social robots, helps humans to understand robots' internal states (e.g., emotions, moods, beliefs, and intentions) and improves the life-like quality of robots [1]. Besides facial expression, bodily expression is a major communication channel of affect. Experimental studies showed that people can recognize these expressions (e.g., [2], [3], [4], and [5]). Furthermore, bodily expression improved humans' recognition of robots' emotion ([2], [3]). In addition, bodily expression is important for robots that lack facial features (e.g., NAO and ASIMO). One way of constructing bodily expression is to build from scratch by "mimicking" humans' behaviors (static postures and dynamic movements). These bodily expressions are typically designed as explicit behaviors. They usually consist of body actions that express emotions deliberately. For example, raising both hands shows happiness [2]; arms akimbo shows anger [3]; covering eyes by hands shows

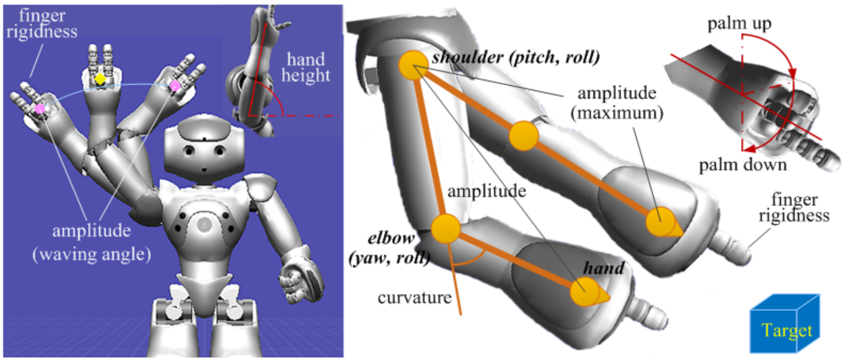

Fig. 1. The parameterized waving (left) and pointing (right) behaviors: our model contains three pose parameters of the arm shown in the figure, two pose parameters of head (head-vertical and head-horizontal), and four motion parameters containing motion-speed, decay-speed, repetition, and hold-time. More details can be found in [6].

fear [4]. However, these body actions rise and dissipate quickly and do not extend over time. Thus, we believe that this type of expression is suitable for expressing emotions, but not moods. Moreover, these body actions may interrupt functional behaviors. For example, a robot cannot express excitement while it is pointing to the object or person that makes it excited by raising both hands. Our work aims at integrating bodily expression of mood with functional behaviors, e.g., task execution, communicative gestures, walking, etc. To this end, we parameterized functional behaviors so that modulating parameters can generate affective cues. Hence, moods can be reflected from the same behavior executed in different "styles", rather than the behavior "contents" per se. As a result, mood can be expressed continuously over time, even when robots are executing tasks. Therefore, we believe that this method is suitable for mood expression. Moreover, bodily mood expression may enhance the affective interaction by prolonging it and providing more modalities.

We investigated our behavior model with a humanoid robot NAO, with interests in whether parameter modulation can be effectively applied to a robotic platform for showing mood. In particular, mood is expressed less explicitly through our approach. In addition, we studied high-DOF functional behaviors, allowing us to define more parameters that may enrich the mood expression. We are also interested how behavior parameters can be combined to show different moods. In previous work [6], our model has been applied to two functional behaviors, waving and pointing (Fig. 1), and we obtained general design principles about the relations between mood variables and behavior parameter modulation from a design experiment, in which participants were asked to design mood expression according to five levels of valence labeled by very-unhappy, unhappy, neutral, happy, and very-happy. In addition, the relative importance and the interrelations between parameters were investigated [7]. Table 1 summarizes the main findings. It is not clear whether people can recognize moods in the presence of behavior functions, since people may devote their attention to behavior functions. This paper reports the findings of a study on people's recognition of the mood expressions resulted from the design experiment, and whether the conclusions of the design experiment correspond to people's perceptions.

Table 1. The design principles and parameter importance.

Waving	HandHeight	Finger	Amp	Rep	HoldTime	DecaySpd	MotSpd	HeadVer.	HeadHor.
Relation[1]	+	+	+	+		+	+	+	
Import.[2]	2		5	4			3	1	
Pointing	**PalmDir.**	**Finger**	**Amp**	**Rep**	**HoldTime**	**DecaySpd**	**MotSpd**	**HeadVer.**	**HeadHor.**
Relation[1]	+		+	*	*	+	+	+	
Import.[2]			2				3	1	

*[1] */+ denotes significant correlations with valence; + denotes increase with valence;*
[2] The number denotes the importance: small - important; unnumbered - unimportant.

Several studies addressed behavior parameter modulation. Wallbott [8] studied the emotional bodily movements and postures of actors/actresses. His study indicated that the body movement “qualities” can reflect emotions. Laban movement analysis [9] models body movements from different aspects, e.g., effort and shape. Chi et al. [11] developed EMOTE framework for synthesizing expressive gestures of virtual agents. An evaluation of effort elements showed that trained observers can recognize the displayed effort at a moderate rate, whereas this study also indicated that prominent effort elements may mask other elements when they are showed in combination. In contrast to EMOTE, which performs as a post process of pre-generated behaviors, Pelachaud et al. [9] modifies gestures before the computing the animation. They characterized behavior expressivity using six parameters: spatial, temporal, fluidity, power, overall activation, and repetition. Their model was applied to an embodied conversational agent for communicating cognitive and affective states through modulated gestures. Evaluation showed that spatial and temporal extents received high recognition rate, but power and fluidity quite low; abrupt and vigorous received high recognition rate but not for sluggish. To achieve a better concord between mood expression and behavior functions, our approach defines behavior parameters while defining the behavior functional profile, so behaviors are also modified first and then the robot joints are computed.

2 Experiment Design and Hypotheses

The recognition experiment first evaluated whether participants can differentiate the five valence levels from modulated behaviors of the design experiment [6]. Second, we tested whether people’s recognition is different when modulating different parameter (sub)sets according to the relative importance [7]: 1) all parameters (APS); 2) only important parameters (IPS), which are numbered in Table 1; and 3) only unimportant parameters (UPS), which are unnumbered. We expect that modulating only the IPS parameters can still express moods without reducing the recognition rate considerably. Although statistical results and participants’ ranks showed that the importance of the UPS parameters was low, but participants did modify them during the design. Thus, we suspect that the UPS parameters can express “weak” moods, which are more implicit and less intense, so we tested whether modulating only the UPS parameters can still express moods. Moreover, the behavior naturalness is one of participants' design criteria in the design experiment. Thus, we suspect that modulating all the parameters may result in more natural behaviors than modulating only IPS parameters. Hence, the behavior naturalness was assessed in the recognition experiment. Therefore, our hypotheses are formulated as follows:

H1. People can distinguish different valence levels from modulated behaviors when all behavior parameters (APS) are modulated. The relationship between parameter settings and perceived valence levels is consistent with the relationship found in the design experiment;

H2. People can perceive different levels of valence when only important parameters (IPS) are modulated; People can still recognize the valence when only modulating unimportant parameters (UPS), but the recognition rate is worse than the APS and IPS conditions;

H3. The behaviors generated by modulating all the parameters (APS) are perceived more natural than the ones generated by modulating only the important parameters (IPS).

The test settings (videos can be found at our website[1]) for the recognition experiment is based on the average setting obtained from the design experiment [6]. An average setting may be not the best design due to the inconsistency and unnaturalness caused by mixture of different individual designs [15]. In our case, the diversity on the arousal dimension is averaged out: for the negative valence, most participants designed sadness (low arousal), but a part of participants designed anger (high arousal). Therefore, we corrected the weighted settings in terms of consistency and naturalness within the boundary of the design principles found in the design experiment, and added anger to recover the diversity on the arousal dimension.

Besides, we tested whether people can perceive arousal from the test settings, since the participants of the design experiment did consider the arousal dimension as just mentioned. We also studied whether parameter sets influence the recognition of arousal. Note that the important parameters (IPS) were obtained from the task where participants were asked to design mood expression only according to the valence. The importance may be only or more in regard with the valence. Thus, whether the parameter sets influence the perception of arousal was unclear.

Paired comparison was used to test how well people perceived valence and arousal from behaviors under APS, IPS and UPS conditions (H1, H2). This method provides more precise results in interval scales than a direct scaling, because it transforms the scaling task, which is difficult for humans, to a comparison task [12, 13]. Participants were asked to compare (not paired comparison) the naturalness of generated behaviors corresponding to each mood under the IPS and APS conditions respectively (H3). The notions of valence and arousal were explained to participants before the experiment using categorical emotion labels and SAM manikins. Naturalness was explained mainly in terms of natural interaction. Participants were provided a user interface for inputting answers and proceeding with the experiment. Two grey NAO robots were used to perform behaviors modified by two moods simultaneously to reduce participants' cognitive workload. Waving and pointing were arranged in a counter-balanced order. For each behavior, the six moods were presented in pairs in a random order, and they were presented under APS, IPS, and UPS conditions

---

[1] http://ii.tudelft.nl/~junchao/mood_expr_recog.html

Table 2. The correlation (Pearson *r*) between parameters and valence or arousal.

Waving	HandHeight	Finger	Amp	Rep	HoldTime	Decay	MotSpd	HeadVer.	HeadHor.
Valence	0.889	0.936	0.966	0.858		0.848	0.848	0.950	
Arousal				0.653	-0.976	0.977	0.977	0.797	
Pointing	**PalmDir.**	**Finger**	**Amp**	**Rep**	**HoldTime**	**Decay**	**MotSpd**	**HeadVer.**	**HeadHor.**
Valence	0.507		0.914	0.810	0.315	0.927	0.927	0.984	
Arousal				0.923		0.978	0.978	0.924	

successively. 26 participants (13 females and 13 males) were recruited from Delft University of Technology. The participants' ages ranged from 21 to 35 years (M = 28.6, SD = 3.3). 13 participants are Chinese, and the other 13 are not. All the participants signed the informed consent form. A pre-experiment questionnaire confirmed that the participants had little experience of designing robots or animated characters. Each participant received a gift as compensation for their time.

3 Analysis and Result

The method based on Thurstone model from [13] was used to analyze the paired comparison data. To see how well participants recognized the moods under the APS, IPS, and UPS conditions, only the mood factor was input into the analysis. For convenience, all results are combined and illustrated in Fig. 2. Assuming that valence and arousal are orthogonal [14], the tested moods are denoted in the valence-arousal space (Fig. 2). First, we interpret the recognition of valence from the five settings derived from the design experiment; second, we interpret the recognition of arousal; finally, we interpret the additional mood anger.

To analyze the recognition of valence (H1), we first looked at the results under APS condition (Fig. 2a, b). Regardless of anger (interpreted later), for both behaviors the valence of each pair of moods was significantly differentiated by participants except for unhappy and neutral pointing. This result shows that people can recognize the valence from the behavioral cues in general (H1). Pearson correlations between parameter values and the perceived valence scales (Fig. 2a, b) were computed. The results (Table 2) show that the relationship between parameters and perceived valence is generally consistent (H1) with the findings of the design experiment (Table 1).

Secondly, we interpret how participants' recognition under the IPS and UPS conditions differs from the APS condition (H2). To this end, we added the parameter set condition as a factor [13] to the paired comparison analysis. The overall result (Table 3) affirms that the parameter set condition influenced participants' perception significantly for both behaviors with regard to valence. In addition, we compared the parameter set conditions in pairs using the same method above. For both behaviors, there are no significant differences between the parameter set APS and IPS (Table 3), which suggests that modulating only the important parameters is capable of expressing valence almost equally well as modulating all the parameters. The generated scale of valence under IPS condition is similar with the APS condition (Fig. 2c, d). The only difference is that the happy and very-happy pointing were

Table 3. Significant differences of recognition between parameter set conditions.

		Overall	APS vs. IPS	APS vs. UPS	IPS vs. UPS
Waving	Valence	p < 0.001***	p = 0.205	p < 0.001***	p < 0.001***
	Arousal	p < 0.010*	p = 0.931	p = 0.026*	p = 0.001**
Pointing	Valence	p < 0.001***	p = 0.671	p < 0.001***	p < 0.001***
	Arousal	p = 0.006**	p = 0.879	p = 0.011*	p = 0.001**

**p<0.05, **p<0.01, ***p<0.001*

differentiated under the APS condition but not for the IPS condition. Possible reason is that repetition increased for very-happy under the APS condition, but not for the IPS condition, since repetition was rated unimportant in previous study. Further study is needed to address whether repetition is important to valence in different situations.

The recognition of valence under APS and IPS conditions is significantly better than UPS condition (Table 3). The high-arousal moods (anger, happy, and very-happy) and neutral were less successfully differentiated by participants for waving (Fig. 2e). Similar results were obtained for pointing (Fig. 2f). Besides, the unhappy and neutral pointing were not significantly differentiated. This suggests that none of the UPS parameters is sufficient to present the valence of high-arousal moods. However, as we hypothesized, some moods can still be recognized even without modulating important parameters. The valence of unhappy and very-unhappy waving was significantly differentiated from waving of neutral and high-arousal moods (Fig. 2e). The long hold-time, slow decay-speed, head turning away from both hand and the front distinguished the unhappy and very-unhappy. We exclude finger since few participants mentioned it in the post-questionnaire. For pointing, the valence of very-unhappy was significantly differentiated from other moods except unhappy. Thus, we conclude that the UPS parameters are promising for "weak" mood expressions for at least two valence levels: positive and negative.

Results also show that participants recognized arousal levels from behaviors. Under APS (Fig. 2a, b) and IPS conditions (Fig. 2c, d), the arousal of high-arousal moods and neutral was significantly differentiated for both behaviors, regardless of anger (integrated later). The arousal of low-arousal moods (unhappy and very-unhappy) was significantly differentiated from high-arousal moods and neutral for waving, whereas only very-unhappy was significantly differentiated form high-arousal moods and neutral for pointing. Statistically, there are no significant differences of perceived arousal between the APS and IPS conditions for both behaviors (Table 3), which suggests that the IPS parameters are capable to express arousal equally well as the APS parameters. However, the perceived arousal under UPS condition differs significantly from either the APS or IPS condition (Table 3). For waving (Fig. 2e), the arousal of very-happy and anger significantly differentiated from neutral, whereas other high-arousal moods were not. Possible reasons are that the zero hold-time and fast decay-speed of angry and very-happy waving made the overall movement fast and fluent, resulting in the perception of a high arousal. The arousal of high-arousal moods was better recognized for the pointing behavior than waving behavior. For pointing (Fig. 2f), the arousal of all high-arousal moods was significantly differentiated from neutral, and very-happy was differentiated from happy. Fast decay-speed and high repetition may account for this. This suggests that the decay-speed and repetition correlate more with arousal than valence. They were actually considered unimportant to valence.

The arousal of unhappy and very-unhappy waving was significantly differentiated from other moods (Fig. 2e), but unhappy and very-unhappy were not differentiated from each other. For pointing, the arousal of neither unhappy nor very-unhappy was significantly differentiated from neutral (Fig. 2f). In fact, the arousal between unhappy and very-unhappy was not significantly differentiated for both behaviors under all conditions, but their valence was significantly differentiated under APS and IPS conditions. The arousal-correlated parameters (e.g., speed, repetition) seem not able to render arousal for low-arousal moods. Back to the UPS condition (Fig. 2e, f), we found that the very slow decay-speed distinguished the valence of very-unhappy from neutral. It seems that the speed parameter like decay-speed may correlate more with valence when it is slow, whereas correlates more with arousal when it is fast.

The recognition of angry waving showed the promise of expressing anger through parameter modulation. The valence of anger was perceived as negative for all conditions (Fig. 2a, c, e), although it was not significantly differentiated from neutral under APS and UPS conditions. Surprisingly, the valence was better differentiated from neutral under the IPS condition (Fig. 2c). We considered that the longer hold-time under IPS condition caused the movement jerkier resulting in a more negative perception, whereas the zero hold-time and the faster decay-speed under APS condition made the movement smoother resulting in a relative more positive perception. Furthermore, the head turned away from the moving hand in the APS condition, which made the robot seem to avoid the eye-contact resulting in a feeling of fear, while fear has a more positive valence than anger [14]. The valence of anger was recognized better for waving than pointing, since it was recognized as positive for pointing under all conditions. Perhaps, the presence of arousal (by large amplitude, repeated movements, and fast speed) in angry pointing was dominant and masked the expression of negative valence, which led people to consider the mood as excitement.

As discussed before, the arousal of anger was recognized significantly higher than neutral and low-arousal moods for both behaviors under all conditions. Interestingly, the perceived arousal of angry pointing and waving under UPS condition was as high as very-happy (Fig. 2e, f), whereas in other conditions it is significantly lower than very-happy. Possible reason is that most parameters were set to the same value between these two moods under the UPS condition. However, the only element that made the arousal of the very-happy pointing under APS and IPS conditions higher than angry pointing is the high-raised head. This suggests that head-up-down correlates with arousal. According to the above discussion, we summarize the parameters that correlate with arousal in Table 2, where Pearson correlation was computed between parameter values and the perceived arousal scale.

Binomial tests were used to analyze whether behaviors under the APS condition was perceived more natural than the IPS condition (H3). Participants’ choices between the APS and IPS conditions are not significantly above chance level for each mood and behavior. Thus, our study did not show that modulating UPS parameters improves the behavior naturalness. We also tested the effect of gender and culture (Chinese and Non-Chinese) by adding them as a factor into paired comparison analysis separately. The results do not show any significant differences between gender and culture conditions.

4 Discussion

The modulation of the important parameters expresses moods better than unimportant parameters. Most important parameters like hand-height, amplitude, motion-speed, and repetition are "global" parameters, which influence the overall movement. Changing these parameters will alter the movement appearance noticeably. Head position also has strong effect on affect expression [5], probably because the head is a special body part that people usually pay attention to during interaction. The unimportant parameters are "local" parameters that influence only a small region of the body parts (e.g., finger-rigidness, palm-up-down) or a short period (e.g., hold-time and decay-speed are temporally local) of the whole movement. Thus, they may not produce sufficient affective cues or people may not even notice them. Hence, behaviors with more "global" parameters may be more affectively versatile, For example, waving has higher expressivity than pointing. In fact, moods were recognized better through waving than pointing in general.

Interactions may exist between valence and arousal. According to Table 1 and Table 2, parameters like motion-speed, head-up-down and repetition of waving were found to correlate with both valence and arousal. In addition, a 5-point Likert scale (from 1: "extremely disagree" to 5: "extremely agree") post-experiment questionnaire suggests that the participants generally agreed on that valence and arousal are related. The mean rating is 3.85 (SD=0.88). Several studies also reported that valence and arousal are not orthogonal [16]. The interaction between valence and arousal should be taken into account when we design mood expressions.

Our model is possible to be generalized to other behaviors in terms of the relations between behavior parameters and mood variables. As in our model parameters are defined at the stage of constructing behavior functional profiles, parameters are dependent on behavior functions. Thus, the same parameters may have different meanings for different behaviors. Despite the differences, design principles may still hold. For example, although the amplitude is the swing angle for waving but the arm extension for pointing, larger amplitude corresponds with a positive mood for both behaviors. However, design principles may also be different for the same parameters. For example, the hold-time means smoothness for waving but persistence for pointing. Hence, shorter hold-time (smoother movement) corresponds with a positive mood for waving, whereas longer hold-time (more persistent) of pointing generally expresses a positive mood. We suggest designers pay attention to the meaning of a parameter for specific behaviors when modulating the parameter to express mood.

5 Conclusion and Future Work

This paper presents a study on people's recognition of humanoid robots' bodily mood expression through behavior parameter modulation. The results indicated that five valence levels can be expressed through parameter modulation for the two behaviors studied. Arousal can also be expressed with at least four levels. The important parameters that influence the behavior overall have a major effect on both valence and arousal. The unimportant parameters can express "weak" moods for at least two levels of valence and three levels of arousal for both behaviors, but no effect on naturalness of these parameters was observed. The speed parameters, repetition, and head-up-down were found to correlate with arousal. Speed parameters are capable to render arousal when they are fast, but not when they are slow. In the future, we will improve the angry pointing and study the relation between the pointing direction and mood expression. While mood expressions via parameter modulation can be recognized in an experimental setting, whether people can recognize them correctly, even notice, in real HRI scenarios still remains a question. We will apply the design principles into more behaviors used in HRI and address the question in the future.

References

1. Breazeal, C., Takanishi, A., Kobayashi, T.: Social robots that interact with people. Springer Handbook of Robotics, pp. 1349–1369. Springer, Berlin, (2008)
2. Zecca, M., Mizoguchi, Y., Endo, K., Iida, F., Kawabata, Y., Endo, N., Itoh, K., Takanishi, A.: Whole body emotion expressions for kobian humanoid robot - preliminary experiments with different emotional patterns -. in RO-MAN, pp. 381–386. IEEE, (2009)
3. Hirth J., Schmitz N., Berns K.: Towards social robots: Designing an emotion-based architecture. Int. J. Social Robotics, vol. 3, pp. 273–290, (2011)
4. Häring, M., Bee, N., Andre, E.: Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. in RO-MAN, pp. 204–209. IEEE, Aug. (2011)
5. Beck, A., Stevens, B., Bard, K.A., Cañamero L.: Emotional body language displayed by artificial agents. ACM Trans. Interact. Intell. Syst. vol. 2, no. 1, pp. 1–29. (2012).
6. Xu, J., Broekens, J., Hindriks, K., Neerincx, M.A.: Mood Expression through Parameterized Functional Behavior of Robots. in press, ROMAN, Gyeongju, Republic of Korea, (2013)
7. Xu, J., Broekens, J., Hindriks, K., Neerincx, M.A.: The Relative Importance and Interrelations between Behavior Parameters for Robots' Mood Expression. in press, ACII, Geneva, Switzerland, (2013)
8. Wallbott, H.: Bodily expression of emotion. European J. Social Psychology, vol. 28, no. 6, pp. 879–896. (1998)
9. Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Communication, vol. 51, no. 7, pp. 630–639, (2009)
10. von Laban, R., Ullmann, L.: The Mastery of Movement. 4th ed. Dance Books, Limited, (2011)
11. Chi, D., Costa, M., Zhao, L., Badler, N.: The emote model for effort and shape. in SIGGRAPH. pp. 173–182. ACM, (2000)
12. Engeldrum, P.G.: Psychometric scaling: a toolkit for imaging systems development. Imcotek Press (2000).
13. Rajae-Joordens, R., Engel, J.: Paired comparisons in visual perception studies using small sample sizes. Displays (2005).
14. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology. 39 pp. 1161-1178 (1980).
15. Bergmann, K., Kopp, S., Eyssel, F.: Individualized gesturing outperforms average gesturing - evaluating gesture production in virtual humans. In: Intelligent Virtual Agents, pp. 104-117. Springer (2010)
16. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International Affective Picture System (IAPS): Technical Manual and Affective Ratings. (1997)

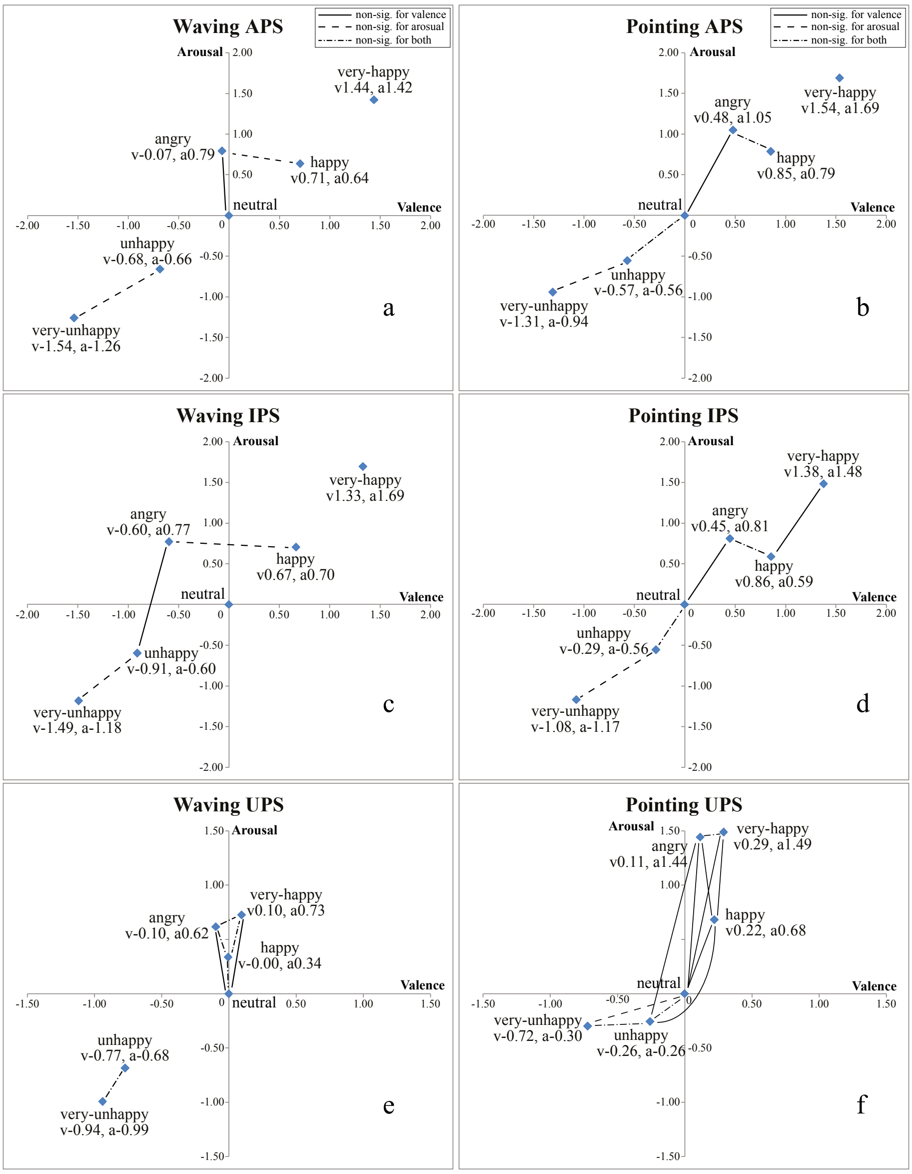

Fig. 2. The figure shows the position of each perceived mood in the valence-arousal space under the APS, IPS and UPS conditions for the waving and pointing behaviors. The valence or arousal of unconnected moods was significantly differentiated, while for the connected ones either valence or arousal or both was not significantly differentiated.