A New Approach to Emotion Generation and Expression

Mei Yii Lim and Ruth Aylett

School of Mathematical and Computer Sciences,
Heriot Watt University,
Edinburgh, EH14 4AS, Scotland
{myl, ruth}@macs.hw.ac.uk

Abstract. This paper describes a new and flexible approach to emotion generation and facial mapping that forgoes emotions labelling and facial action units. A biologically inspired architecture of emotion that allows regulation of an agent emotions and behaviour is presented. The resulting emotions are expressed through a 2D animated character, representing a context-aware mobile tour guide, guiding visitors touring an outdoor attraction. We mapped three facial features: the eyes, the mouth and the eyebrows onto the arousal and valence dimensions where each dimension influences a facial feature more strongly than the others. The focus of this paper is on facial expression. We provide a review of related work, a description of our approach and finally concluding remarks.

1 Introduction

The Affective Guide is a context-aware mobile tour guide implemented on a PDA, taking advantage of wireless hotspots and bluetooth access points for communication with a server. It has a Global Positioning System for location detection and an integrated text-to-speech system. The user interacts using the graphical interface and receives output by means of text, audio and visual animation. The guide is capable of navigating the user to a destination, acknowledging their arrival and personalising stories based on the user's feedback. Its internal processing is controlled by an emergent emotional model.

2 Emergent Emotion Model

The emergent emotional model is based on the 'PSI' model [1]. It is unique because emotions are not explicitly defined but evolve from modulation of perception, motivation, action-selection, planning and memory access. This model bridges the gap between models that focus solely on physiological-level of emotions (eg. [2, 3]) and those that concentrate on higher-appraisal level (eg. [4, 5]). The emotions of the guide are triggered by conditions that are evaluated as being of significance to its 'well-being', establishing the desired relation between the guide and its interaction environment. Basically, the guide has two built-in motivators to maintain - *level of competence* and *level of certainty*. The level of competence refers to its ability to cope with the user's perspectives on an issue or event whereas the level of certainty is the degree of predictability of the user interests. Figure 1 illustrates the emergent emotion model.

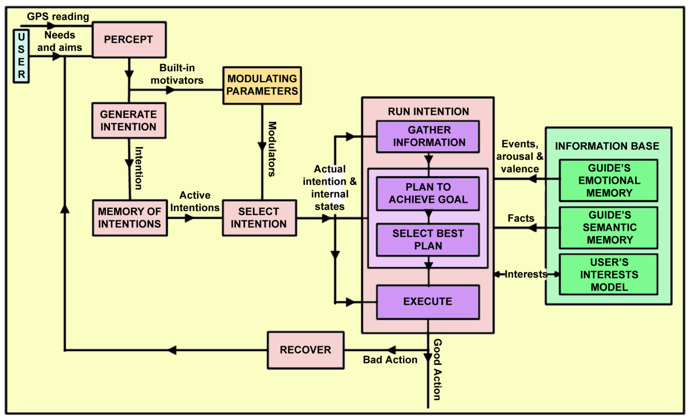

Fig. 1. The Emergent Emotion Model

The guide continuously reads user inputs (feedback on the degree of interest in the story and degree of agreement with the guide's argument), system feedback and the GPS information. Then, it generates an intention and stores it in a memory of intentions together with the built-in motivators. The guide has three possible intentions - update its beliefs about the user's interests; adjust the story presentation; or perform storytelling. More than one intention can be active at the same time and depending on the importance of the need, one of the active intentions is selected for execution. Based on its emotional state and the selected intention, the guide decides whether to explore for more information, to design a plan using the available information or to run an existing plan. The guide's emotional state is affected by modulators such as arousal level, resolution level and selection threshold; influenced by the built-in motivators. Arousal level refers to the agent's readiness to act while resolution level determines the carefulness of the guide's behaviour. The selection threshold on the other hand, is the limit competing motives have to cross in order to become active. Arousal is directly proportional to selection threshold while inversely related to the resolution level. For example, if the guide's prediction about the user's interests is right (high certainty) but the user's perspective is in conflict with the guide's viewpoint (low competence), then the arousal level of the guide will increase. The resolution level decreases while the selection threshold increases. In this case, the guide has some difficulty in coping with the differing perspective and it is motivated to concentrate on the specific goal, adjusting the presentation of the story appropriately by giving a more general view on the issues.

For story generation, the guide takes into consideration the user's interests, its own interests and its current memory activation. Its long-term memory is made up of both semantic and emotional memories. While the semantic memory contain facts, emotional memory is a memory for events that have emotional impact on the guide. The emotional memory reflects the guide's ideology and is tagged with 'arousal' and 'valence' [6] tags analogous to the *Emotional Tagging* concept [7], which recorded the guide's emotional states for an event. When interacting with the user, the guide will be engaged in meaningful reconstruction of its own past, at the same time presenting facts about the site of attraction. The recollective experience of the guide is related to the evocation of previously experienced emotions through the activation of the emotion tags, resulting in re-experiencing of emotions, though there might be a slight variation due to the input from the user. A more detailed description of the emotion model and the system can be found in [8].

3 Facial Expressions

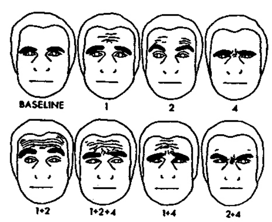

Fig. 2. The three FACS action units in the brow area and their combinations. AU 1 (action of inner frontalis) raises the inner corners of the eyebrows, AU 2 (action of the outer frontalis) raises the outer portion of the eyebrows and AU 4 (action of procerus, corrugators, and depressor supercilii) pulls the eyebrows down and together (from [9])

Having internal states, the guide needs a more obvious mechanism for expression in addition to its behaviours. The most common means is through facial expressions. Facial expressions can be studied either by direct measurement of facial activity or through an observer's judgments [10]. One of the most well-known measurement system is the Facial Action Coding System (FACS) [11]. FACS' chief use is for scoring facial actions seen on motion records and still photographs. Each *Action Unit* (AU) can involve more than one muscle change. Figure 2 shows an example of three AUs and the combination of these AUs to create different expressions. Although FACS can be used to discriminate positive from negative emotional expressions, and the intensity of these expressions to some extent, some questions remain unanswered on the accuracy of the resulting emotional inferences. Consequently, we have taken almost the opposite position.

Rather than requiring the user to learn the mechanics of facial movement, we emphasize the perception of a static expression. A FACS mapping would be too complex in terms of effort and resource requirements for the Affective Guide. We are not interested in modelling facial muscle movements or in providing a comprehensive set of facial changes. Besides, the AUs would be impossible to replicate using a 2-dimensional face.

Facial behaviour has been commonly categorised using two judgment procedures: *categorical*, involving basic emotion categories and *dimensional*, involving scales or dimensions that underlie the emotion categories. A particular emotion can usually be represented by more than one facial expression, depending on intensity, and the discrete approach suffers from the flaw of rigidity, with its one-to-one mapping. The dimensional approach eliminates this restriction, conveying a wider range of affective messages. Using the dimensional approach, the only consistent finding across experiments for classifiying facial behaviour is the pleasant-unpleasantness and activation or intensity dimensions [12]. The evidence for any dimension beyond that is weak. Consequently, the pleasure-arousal dimension can be seen in many attempts to map facial expressions to emotions (eg. [13, 14, 15]).

Breazeal [13] built a robot called *Kismet* that has the ability to express nine emotions through its facial expressions. *Kismet* has a set of hand-crafted releasers, the antecedent conditions to specific emotive responses with associated *[A, V, S]* (arousal, valence and stance) profiles. The six basic emotions used sit at the extremes of each dimension: surprise at high arousal, fatigue at low arousal, content at positive valence, unhappiness at negative valence, stern at closed stance and accepting at open stance. We see this approach of labelling each facial expression with an emotion term as inflexible and insufficient to accommodate human complex social interaction signals. Although facial expressions can provide accurate information about the occurrence of pleasant versus unpleasant emotional states, distinctions among particular positive or negative emotions is an issue yet to be addressed [10]. Hence, the emotion terms associated with the facial expressions may require further verification. This is not a problem in our case because we do not label emotions, thus, do not require accuracy of emotion judgment.

Russell [14] provides a bottom-up approach, avoiding the interpretation of facial expression on the level of emotion categories. Russell's mapping of Ekman's facial action units onto the pleasure-arousal space is illustrated in Figure 3. From the diagram, it can be seen that displeasure is signaled by downturned mouth and furrowed brows, while pleasure is conveyed by smooth brows and an upturned mouth. Lowered eyelids signal lower arousal while widened eyes, raised brows and increased muscle activity are signs of increasing arousal. Using Russell's approach, [15] modeled 25 AUs as a system of morph targets on 3-dimensional faces. In a study, observers gave pleasure and arousal rating to the single facial muscle components on the facial expressions displayed by a *Face Randomiser*. The regression factor scores for pleasure and arousal on the single muscles were calculated and an activation space of these muscles in Russell's

Circumplex Model was developed. Figure 4 illustrates the reconstruction of facial expression along the pleasure and arousal dimensions. The authors later found a multiform distribution of categorical emotions in the spaces. The resulting emotions can occur within contradictory contexts, for example feeling angry under happy circumstances. Even so, this exploratory work serves as an inspiration to our research because this approach can generate virtually endless combinations of facial expressions.

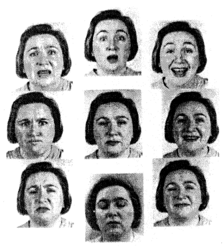

Fig. 3. Russell's pleasure-arousal space for facial expression (from [14])

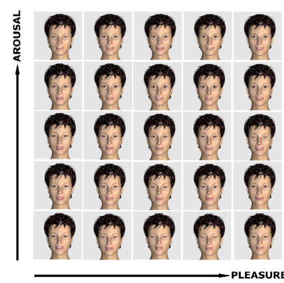

Fig. 4. The complete reconstructed pleasure and arousal space (from [15])

A much simpler implementation can be seen in [16]. Schubert developed *EmotionFace*, a software interface to display emotions expressed by music. A simple schematic face with eye opening representing arousal and mouth movement representing valence is used. Both eye and mouth changes are calculated based on parabolic functions. A high arousal will increase the eye opening while a low arousal will reduce it. A positive valence deepens the mouth up-concavity while negative valence flips the mouth in the opposite direction.

4 Our Approach to Emotion Expression

The approach taken here is to build the Affective Guide based on patterns that are familiar to the user but without explicit labelling of the expressed emotions. We argue that the interpretation of facial expressions depends on individuals and context of occurrence, hence put the task in the hand of the observer. The internal states of the Affective Guide are reflected through a simple 2-dimensional animated talking head. From the review in the previous section, it can be observed that arousal has the greatest impact on the eyes. On the other hand, valence affects the mouth curvature. As well as the eyes and mouth, we take into consideration the eyebrows, which have been found to be as influential in facial expression recognition [17]. The movement of the eyebrows alone may be enough to form some emotional expressions. For example, referring to Figure 2 again, in the expression of sadness, AU 1 or a combination of 1+4 occurs; in surprise, the combination 1+2 occurs; and in anger, AU 4 occurs. By combining the eyebrows with other facial movements, the entire range of human emotions can be produced.

Using the information gathered, we propose a simple novel approach of facial expression mapping onto the emotional space. This approach is flexible and able to produce an infinite range of expressions. We mapped meaningful facial features: eyes, mouth and eyebrows onto the two emotional dimensions, valence and arousal. Variation in these values will move the different facial features dynamically in real time producing different emotions. The valence value moving from negative to positive will move the lip curvature from a downturn U to an upturn U as depicted in Figure 5. A value from 0 to 0.5 denotes negative valence while a value greater than 0.5 represents positive valence. In the case of extreme pleasure, the cheek raiser is visible below the eyes. In contrast, for extreme displeasure, wrinkles are formed beside the wing of each nostril due to the action of the naso-labial fold. Along the arousal dimension, the size of eye opening increases with increasing arousal and reduces with decreasing arousal as shown in Figure 6.

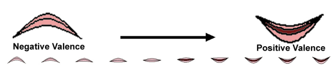

Fig. 5. The lip curvature change along the valence dimension

Fig. 6. The eyes opening along the arousal dimension

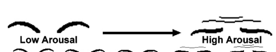

Fig. 7. Movement of the eyebrows along the arousal dimension when valence is positive

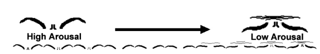

Fig. 8. Movement of the eyebrows along the arousal dimension when valence is negative

As for the eyebrows, they are influenced by both the arousal and valence values. Under positive valence, when arousal is low to medium (less than 0.5), the eyebrows will have a slight V curve. The eyebrows become more and more relaxed and straightened with increasing arousal. When arousal is very high (more than 0.8), the eyebrows will be raised slightly and the raised inner eyebrows cause delicate wrinkles to be formed across the forehead as presented in Figure 7. On the other hand, Figure 8 shows that the opposite takes place when valence is negative. If arousal is very high, the inner eyebrows curve downwards forming a V. The knitting of the inner eyebrows causes furrows to be observed especially when the arousal becomes too high. On the other hand, as arousal decreases, the curve will smoothen and become more and more relaxed. Nevertheless, when arousal falls below neutral (less than 0.5), the curve becomes an upturned V and in the extreme case, forms vertical furrows above the base of the nose, with transverse wrinkles across the forehead. The resulting facial expressions along the arousal and valence dimensions are shown in Figure 9. From the figure, it can be seen that a wide variation of facial expressions can be generated. Although only simple animation is applied, the nuances of the underlying emotional assessment (arousal and valence) can be reflected. This approach provides a robust and flexible architecture where emotion expressions are not limited to the basic discrete expressions but span a continuous space.

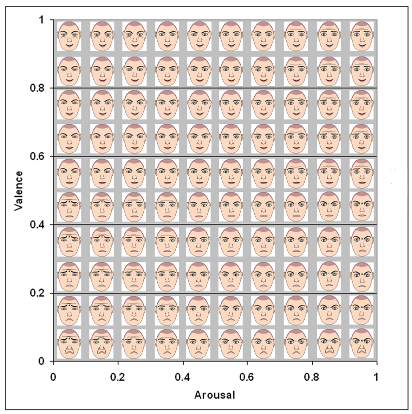

Fig. 9. Different facial expressions on the arousal-valence space

5 Conclusion

The Emergent Emotion Model is a biologically inspired model allowing creation of a believable guide. It regulates the guide's internal states and these subsequently affect the guide's behaviour and processing strategy. The emotions of the Affective Guide are not explicitly defined but emerge from modulation of information processing. In the current version of the Affective Guide, animation of the facial expressions is performed by swapping facial feature frames. Even with such a limited number of frames per feature, a vast combinations of facial expressions were generated, implying that an infinite variation could be achieved if morphing is possible.

Acknowledgements

Work supported by the European Union's Sixth Framework Programme, in the IST (Information Society Technologies) Thematic Priotity IST-2002-2.3.1.6 Multimodal Interfaces, HUMAINE (Human-Machine Interaction Network on Emotion, http://emotion-research.net) (Contract no. 507422). The authors are solely responsible for the content of this publication. It does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of data appearing therein.

References

[1] Dörner, D., Hille, K.: Articial souls: Motivated emotional robots. In: Proceedings of the International Conference on Systems, Man and Cybernetics. (1995) 3828–3832

[2] Cañamero, D.: A hormonal model of emotions for behavior control. In: VUB AI-Lab Memo 97-06, Vrije Universiteit Brussel, Belgium (1997)

[3] Velásquez, J.: A computational framework for emotion-based control. In: Proceeding of the Grounding Emotions in Adaptive Systems Workshop, SAB '98, Zurich, Switzerland (1998)

[4] Ortony, A., Clore, G., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge, UK (1988)

[5] Scherer, K.R.: Appraisal considered as a process of multilevel sequential checking. In Scherer, K.R., Schorr, A., Johnstone, T., eds.: Appraisal Processes in Emotion: Theory, Methods, Research. Oxford University Press, New York (2001) 92–120

[6] Kensinger, E.A., Corkin, S.: Two routes to emotional memory: Distinct neural processes for valence and arousal. PNAS **101**(9) (Mar, 2 2004) 3310–3315

[7] Richter-Levin, G., Akirav, I.: Emotional tagging of memory formation - in the search for neural mechanisms. Brain Research Reviews **43** (2003) 247–256

[8] Lim, M.Y., Aylett, R.: Intelligent mobile tour guide. In: Symposium on Narrative AI and Intelligent Serious Games for Education, AISB'07, Newcastle, UK (April 2-4 2007)

[9] Hager, J.C., Ekman, P.: The inner and outer meanings of facial expressions. In Cacioppo, J.T., Petty, R.E., eds.: Social Psychophysiology: A Sourcebook. The Guilford Press, New York (1983) Available: http://face-and-emotion.com/dataface/misctext/inner_outer.html.

[10] Ekman, P., Oster, H.: Facial expressions of emotion. Annual Review of Psychology **30** (1979) 527–554

[11] Ekman, P., Friesen, W.V.: Measuring facial movement with the facial action coding system. In: Emotion In The Human Face. 2nd edn. Cambridge University Press (1982) 178–211

[12] Ekman, P., Friesen, W.V., Ellsworth, P.: What emotion categories or dimensions can observers judge from facial behavior? In: Emotion In The Human Face. 2nd edn. Cambridge University Press (1982) 39–55

[13] Breazeal, C.: Emotion and sociable humanoid robots. International Journal of Human-Computer Studies **59** (2003) 119–155

[14] Russell, J.A.: Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective. In Russell, J.A., Fernaández-Dolz, J.M., eds.: The Psychology of Facial Expression. Cambridge University Press (1997) 295–320

[15] Grammer, K., Oberzaucher, E.: The reconstruction of facial expressions in embodied systems: New approaches to an old problem. ZiF: Mitteilungen (2006)

[16] Schubert, E.: Emotionface: Prototype facial expression display of emotion in music. In: Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia (Jul, 6-9 2004)

[17] Sadr, J., Jarudi, I., Sinha, P.: The role of eyebrows in face recognition. Perception **32** (2003) 285–293