Perceptually Valid Facial Expressions for Character-based Applications

Ali Arya
School of Information Technology, Carleton University, Canada
Steve DiPaola
School of Interactive Arts and Technology, Simon Fraser University, Canada
Avi Parush
Department of Psychology, Carleton University, Canada

Abstract

This paper addresses the problem of creating facial expression of mixed emotions in a perceptually valid way. The research has been done in the context of a "game-like" health and education applications aimed at studying social competency and facial expression awareness in autistic children as well as native language learning, but the results can be applied to many other applications such as games with need for dynamic facial expressions or tools for automating the creation of facial animations. Most existing methods for creating facial expressions of mixed emotions use operations like averaging to create the combined effect of two universal emotions. Such methods may be mathematically justifiable but are not necessarily valid from a perceptual point of view. The research reported here starts by user experiments aiming at understanding how people combine facial actions to express mixed emotions, and how the viewers perceive a set of facial actions in terms of underlying emotions. Using the results of these experiments and a three-dimensional emotion model, we associate facial actions to dimensions and regions in the emotion space, and create a facial expression based on the location of the mixed emotion in the three-dimensional space. We call these regionalized facial actions "facial expression units".

Keywords: face, expression, perception, animation, game

1 Introduction

The human face is a rich source of information regarding underlying emotional states. Facial expressions are crucial in showing the emotions as well as increasing the quality of communication and speech comprehension. Perceiving these expressions is a social skill that people develop from early ages. The lack of this skill has been observed in people suffering from autism and other brain disorders, and can cause serious issues in social communication. Our research is linked to a project that aims at studying social competency and emotion awareness in autistic children. The project, as briefly introduced later, is developing a game-like tool that engages children in a variety of activities in order to assess their social skills. It uses computer generated scenarios involving characters showing different emotions and personality traits which children need to interact with. Proper generation of facial expressions in a perceptually valid way is essential to this and many other computer games that require non-verbal communication particularly if these expressions are the result of non-scripted dynamic events. Even in non-real time applications a tool that allows automated creation of facial expressions can help animators. We briefly discuss a learning game project for languages of the aboriginal people of Canada as one example in the area (The Autism and Aboriginal Art projects are out of the scope of this paper and will only be reviewed briefly to illustrate possible applications of our facial expression model). The detailed study of facial actions involved in the expression of the six universal emotions (joy, sadness, anger, fear, surprise, and disgust) [12] has helped the computer graphics community develop realistic facial animations. Yet the visual mechanisms by which these facial expressions are altered or combined to convey more complicated emotional states remain less well understood by behavioural psychologists and animators. Examples of such emotional states are two emotions being felt at the same time (the primary focus of this paper), transition from one emotional state to another, and one emotion being masked by another one.

Table 1. Examples of Action Units for Universal Expression of Joy and Sadness [13]

Emotion	Action Units
Joy	eye-lid tightened cheeks raised brows slightly lowered wrinkles around eyes lip-corners raised and stretched
Sadness	upper eye-lid lowered lower eye-lid raised inner brows raised lip corners lowered lower lip raised slightly cheeks raised

The lack of strong theoretical basis for combining facial actions to present complex emotional states has resulted in the use of ad-hoc methods for blending facial expressions (i.e. creating a facial expression that represents an emotional state in which more than one emotion is being felt), as discussed in Section 2. They mainly consider the facial movements for transient or combined expressions a simple mathematical function of the main expressions involved. The methods that have emerged are therefore computationally tractable, but the question of their "perceptual" and "psychological" validity has not yet been answered properly. An example of such methods is calculating a weighted average of facial actions in the blended expressions [11] where a weight equal to the strength of each expression in the blend is applied to the value of facial actions in the individual expression to find the "blended" values. Figure 1 illustrates the problem with non-perception-based methods of creating and blending facial expressions. Figures 1(a) and 1(b) show the expression of universal emotions "surprise" and "anger", using the Ekman's suggested facial action units for these emotions [13]. Figure 1(c) shows a weighted average of action units to create a blend of these two emotions, to be used for transition between them or simply expressing a mixed emotion. It can be noticed that the process results in facial actions which may be acceptable as mathematical average of the two expressions but are not perceptually valid, i.e. they are not what viewer perceive the proper expression of a mixed emotion (particularly lower lip movement).

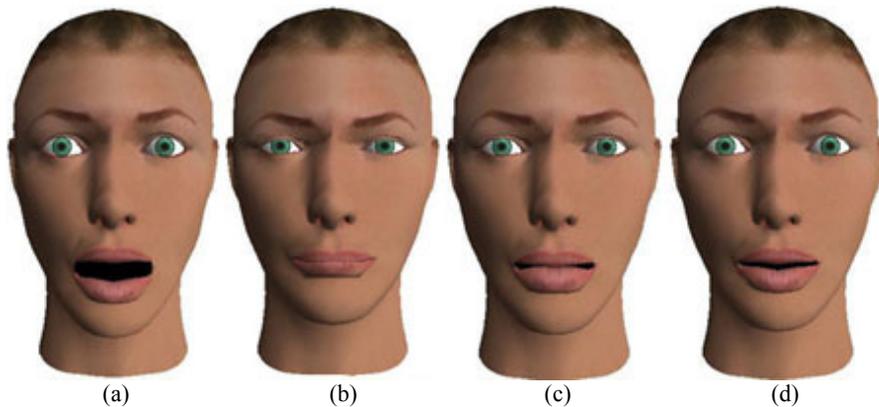| (a) | (b) | (c) | (d) |

Figure 1. (a) Surprise, (b) Anger, (c) Transition between Surprise and Anger using simple averaging (interpolation), and (d) Blending based on emotion parameters and their associated facial actions which is more "perceptually valid".

On the other hand, dimensional models have been suggested for emotions, and the locations of emotions in these multi-dimensional or parameter spaces have been studied by behavioural psychologists [26,29]. Arousal (level of excitement) and Valence (level of pleasantness) are common dimensions suggested in this regard [26] and will be discussed in Sections 2 and 3. Unfortunately, the emotional values of facial actions (i.e. how they relate to emotion dimensions) have not been studied enough. For example, lowering the eyebrows has been observed in facial expression of more than one emotion (See Table 1). But it is not clear if the lower-eyebrow facial action has any specific emotional value of its own, such as being related to high Valence or low Arousal.

Our research aims at understanding how people perceive expressions, especially when mixed, and how they combine facial actions to express mixed emotions. Mixed emotions can mean superimposition of two emotions felt at the same time, one emotion being masked by another one, and a transition from one emotion to another (which can have a different visual result). This paper is focused mainly on the first case but can also be used, with possible extensions, in other cases. The proposed approach is based on two principles which form the basis for our main contributions:
1- Selecting facial actions needed for expression of mixed emotions based on perception experiments where users show how they use and perceive facial actions with regards to mixed emotions
2- Associating facial action to emotion parameters rather than specific emotional states, so the facial actions for each expression (single or mixed emotion) are selected based on parameter values (i.e. the location of intended expression in the parameter space) rather than pure mathematical operations on two independent expressions

We consider facial expressions as points in a three-dimensional emotion space, with Valence, Arousal, and Agency (being self-oriented) as the dimensions or parameters (See Section 3). As discussed later, multi-dimensional emotion spaces have been studied in behavioural psychology, for example by Scherer [29]. Based on Ekman's decomposition of universal expressions to smaller facial action units, we define "expression unit" as a facial action, an emotion dimension, and an association between them in form of an activation rule (e.g. if dimension value is high, then the action will be negative low). We create the visual appearance of a combined expression by mapping it into the 3D space and activating the expression units based on the value of emotion parameters at that point. The user experiments are done in two phases:
1- A group of participants generate facial expressions for mixed emotions using a 3D facial animation software.
2- A second group of participants verify the facial expressions of the first group based on their own perception of the emotions in generated images.

Our approach is unique in that it uses perceptual validity rather than pure mathematical operations to suggest the way facial actions are combined. Also associating facial actions to emotion parameters instead of emotional states, especially when based on the user experiments, is one of our contributions. This is particularly significant when we notice that facial actions for mixed emotions and the emotional value of individual facial actions have not been studied in much depth by behavioural psychologists. Most of the existing approaches "assume" that what happens in a mixed emotion is a combination of what happens in either one of those emotions. This combination can be a simple linear averaging or a more complicated one, but still a combination of elements coming from those individual emotions. Our primary research objective was to question such assumptions by running user experiences for mixed emotions and see how facial actions change from individual emotions to a mixed state. Naturally, if we want to study and consider facial actions in mixed emotions not as a function of individual emotions, then we need to find other variables that control the facial actions, and that is where the need for associating facial actions to emotion parameters instead of individual emotions becomes clear. By associating facial actions to dimensions of emotion space, the actions in a mixed emotion can simply be a function of dimensional values and not individual emotions. Finally, we propose a method of using these research findings in a facial animation system, although we believe that the user study by itself can be used by animators. In Section 2 some existing works related to our research will be reviewed. Sections 3 and 4 will discuss the major concepts in our approach and our experimental methodology. Some experimental results, sample applications, and concluding remarks will be presented in Sections 5, 6 and 7.

2 Related Work

Facial Action Coding System (FACS) by Ekman and Friesen [12] was one of the first systematic efforts to determine the smallest possible facial actions, or Action Units (AUs), and associate them to facial expressions. MPEG-4 standard [3] uses a similar approach to define Facial Animation Parameters (FAPs). Ekman [13] showed that joy, sadness, anger, fear, surprise and disgust have universal facial expressions that can be described in terms of a series of action units (Table 1). Although behavioural psychologists have done further studies on emotions, and computer graphics researchers have used this idea extensively [23], mixed emotions (and transition between universal emotions) and particularly their facial expression have not been studied enough to provide a theoretically sound basis for animation purposes.

On the other hand, the study of emotions and establishing parameters that all emotional states can be defined in terms of their values have been other areas of active research in behavioural psychology. Thayer [33] suggested two parameters of Energy and Stress that form a two-dimensional mood space shown in Figure 2(a), while Russell [26,27] defined the location of common emotional states in a circumplex in 2D space with Arousal and Valence as dimensions (or parameters).
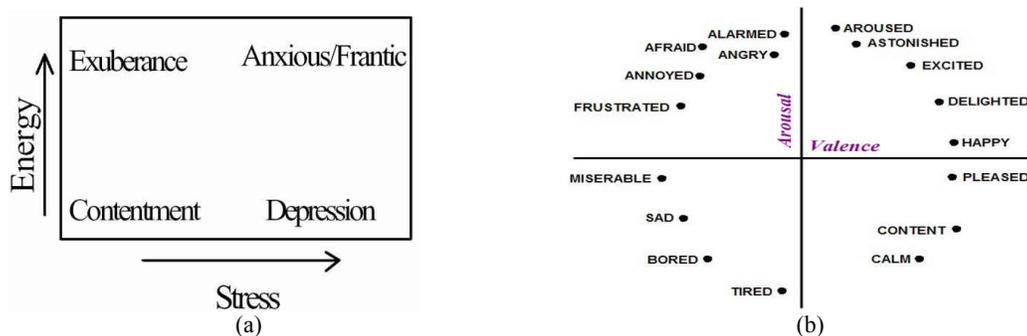

Figure 2. Emotion Spaces: (a) Thayer, (b) Russell

Some researchers have investigated the need for more than two parameters to define an effective emotional space. This is based on the fact that even with equal values for Arousal and Valence we can have emotions that are completely different, as shown by Scherer [29]. The best examples among common emotions are fear and anger which are not easily distinguishable in a 2D space [16]. Mehrabian [20] suggested Control as the third parameter, while Uncertainty and Agency have been suggested by Tiedens and Linton [34] and Ruth et al. [28]. Stance is another suggested dimension used by Breazeal et al. [6] to create an emotional robot. Control, Agency and Stance are conceptually close as they somehow represent the level of inwardness vs. outwardness of the emotion. Figure 3 illustrates the 3D emotion space with Agency/Stance used as the 3rd dimension. Alberch et al [1] suggested Agency as the third dimension and gave quantitative values for projecting standard emotions to these three dimensions. Becker and Wachsmuth [4] suggested Dominance/Control as the third dimension which is similar to Agency but less appropriate as an emotional parameter, as it is more of a personality parameter.

On the other hand, few researchers have studied the emotional effect of individual facial actions. One of these few examples is the study done by Smith and Scott [32] that investigates such emotional effects but as shown in Table 2, it is not exactly with respect to emotion parameters but more of a general emotional effect. Kaiser and Wehrle [15] used the results by Scherer [29] on Appraisal theory (analysis of emotions in terms of dimensional space rather than discrete "basic" emotions) and develop a system that generates facial expressions using five Stimulus Evaluation Checks (SECs) such as novelty and pleasantness of the external event. Their perceptual study is not directly linking facial actions to dimensions of emotion space and the choice of SECs instead

of a 2D or 3D emotion space can be hard to visualize and correlate to facial actions (examples being individual's ability to influence the event or his/her needs). Grammer and Oberzaucher [14] reported a perceptual study involving users rating individual facial action with respect to Arousal and Valence, and then using the results for perceptually valid facial animation. Although close to what we are suggesting, their research is limited to a 2D emotion space and does not consider the effect of facial actions working together to create an expression. Individual facial action can cause unreliable perceptions during the rating, as shown later in our results. In our study, the objective is to understand the emotional effect of each individual, but participants rate a series of expressions that represent different combination of facial actions. Analysis of these ratings can give more reliable data on the affective meaning of each facial action while considering their inter-relation.
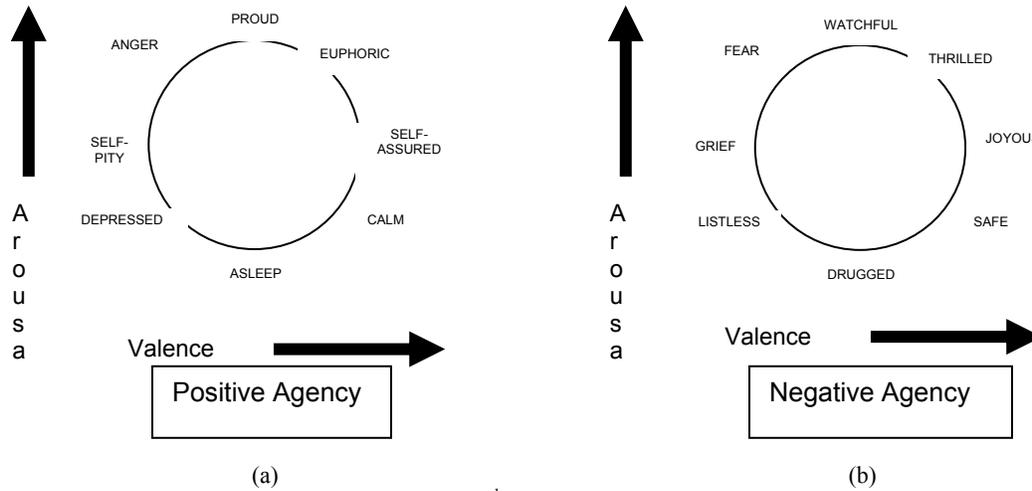

(a)   (b)

Figure 3. 2D Circumplex Mapped Along the 3rd Dimension (Agency or Stance). Positive Agency (a) creates self-oriented emotions like anger, but negative values for Agency (b) will result in other-oriented emotions like fear with similar Arousal and Valence.

Morphing [5,31] as a popular method for creating animation between key images, have long been used for facial animation involving the expression of universal emotions [7,17,24]. This involves having the images of the individual facial expressions, and creating transitions and/or blend by morphing between them. In cases where only a neutral image (or base 3D state) is available, 2D or 3D warping [38] has been used to create new expressions using the neutral face and facial actions that are learned for expressing different emotions [17,37]. The blend of two expressions, or transition between them, is most commonly done by a weighted average of the expressions, regardless of the mathematical model s used for each of them [11,22,25,41]. Yang and Jia [39] have suggested the use of MAX operator when two expressions use the same action, while bilinear approaches have been proposed by other researchers [9,10,36] that divide the face space into face geometry (or content) and emotion (or style) sub-spaces but still within the emotion sub-space use linear combination of individual expressions to create a blend. Byun and Badler [8] used a series of motion analysis parameters (space, time, flow, and weight) to modify an input stream of MPEG-4 FAPs, but do not directly address the issue of blending universal emotions.

Table 2. Emotional Effect of Facial Action [32]. Up arrows indicate that activation of the facial action increases the corresponding emotional effect while down arrows mean the opposite. Darker arrows show stronger relation.

	Facial Action							
Meaning	Eyebrow Frown	Raise Eyebrows	Raise upper Eyelid	Raise Lower Eyelid	Up Turn Lip Corners	Open Mouth	Tighten Mouth	Raise Chin
Pleasantness	↓				↑	↑	↓	↓
Goal Obstacle/Discrepancy	↑							
Anticipated Effort	↑							
Attentional Activity		↑	↑					
Certainty		↓		↑		↑		
Novelty		↑	↑					
Personal Agency/Control		↓	↓			↓		

Martin et al [19] considered two types of blends (superposition and masking). This as mentioned before is a valuable point. They proposed 6 dimensions or parameters which are overlapping and hard-to-visualize. Their work also considered seven facial areas each corresponding to one of the blended expressions. This assumption as discussed before is a significant weakness in most of existing approaches. Tsapatsoulis et al [35] and Albrecht et al [1] suggested interpolation based on 3 emotion parameters but AUs are still associated to basic emotions not parameters and their approaches are not based on user experiments for perception of mixed emotions. Some researchers such as Mufti and Khanam [21] have suggested the use fuzzy rule-based systems [40] for recognition and synthesis of facial expressions, in order to deal with the uncertainty involved. No direct link has been established between these rules and a perceptual study of facial actions.

The common drawbacks of these methods are that (1) the blending and transitions although mathematically correct do not have a perceptual basis, and (2) the facial actions are associated with individual expressions and then mixed based on how strong that expression is. The facial expression of an emotional state that is a mix of anger and surprise, for example, is not necessarily a direct combination of expressions for anger and surprise. A perceptual analysis of the effect of facial actions and the location of the emotional state within the emotion space is needed to decide which facial actions need to be present at the facial expression.

3 Facial Expression Units

In a previous study by the authors [2] it has been demonstrated that facial actions can be associated to personality parameters, Dominance and Affiliation. This can help create perceptually valid facial animations in which the designer defines the personality type by setting the personality parameters and the system automatically selects facial actions that correspond to that setting. Continuing with that idea, one may question what the emotional values of individual facial actions are. In other words, the question is "Can facial actions, in addition to their combined role in expressing emotions, have individual associations to emotion parameters?" If such associations exist, we can define any emotional state (e.g. universal emotions or any combination of them) as a point in multi-dimensional emotion space and then activate proper facial actions to express that state without the need to mathematically mix facial actions not knowing what the perceptual effect of the result will be.

A facial Expression Unit (EU) consists of a facial action associated with the dimensions of a 3D emotional space based on perception-based experiments, and used as building block for facial expression of emotions. Underlying EU-based facial expression of emotions is a dimensional emotion space. The 3D space shown in Figure 3 is an effective choice as it separates universal emotions with the least number of parameters that can be visualized, i.e. related to visual actions as opposed to parameters like novelty used by Smith and Scott [32]. Three dimensions of the emotion space are arousal (the level of excitement), valence (the level of pleasantness), and agency (the level of self or other-directedness). Expression Units are FACS Action Units which are associated to emotion dimensions. In other terms, an EU is a pair of AU-dimension with proper association that in our suggested method is through a fuzzy rule set. By this definition, EU is not exactly a "new" concept as it is basically a FACS AU, but we believe it is "functionally new" as it is now linked to emotion dimensions through rules of activation.

Table 3. Dimensional Values for Standard Emotions

Emotion	Arousal	Valence	Agency
Joy	17.3	42.2	12.5
Sadness	-17.2	-40.1	-52.4
Fear	14.8	-44.4	-79.4
Anger	34.0	-35.6	20.0
Disgust	60.0	-60.0	30.0
Surprise	60	0	0

Using EUs allows creating mixed emotions by mapping the emotional state to the 3D emotion space and then activating the facial actions that are more likely to cause the perception of desired emotions. To locate the emotional state in the 3D space, we can either set the arousal-valence-agency parameters directly, or choose the values for universal emotions which can be translated to dimension values. We use the same dimensional mapping that Albercht et al [1] have used for standard emotions in order to place these emotions in 3D space (Table 3). The dimension values will define a 3D vector for each standard emotion corresponding to its full activation. The user can specify a percentage value for each emotion that simply gives a length to that vector, and the final location of a mixed emotion can be found by adding the vectors for the standard emotions and creating a final vector. If e_i is the percent value for the ith standard emotion and v_{ij} is the strength of that emotion on j^{th} dimension as shown in Table 3 ($1<j<3$), the final 3D vector representing the intended emotional state in 3D space can be found using the following equation where d_j is the component on j^{th} dimension:

$$d_j = \sum_{j=1}^{6} e_i \times v_{i,j}$$

(1)

Once the dimension values have been set, the association between dimensions and the facial actions will be used to activate proper facial action on the face. The process of associating facial actions to emotion dimensions is discussed in Section 4. As we

will see, such association is mainly based on the correlation of facial actions generated and the perceived emotions in a series of user experiments. Due to uncertain, imprecise, and non-linear nature of these associations, we propose a fuzzy system [40] for this part. Fuzzy systems are rule-based decision-making systems that define rule variables in terms of fuzzy (qualitative or linguistic) values such as Low and High. Membership functions determine the degree to which the non-fuzzy value of a variable belongs to each fuzzy label. For each emotion dimension, a series of fuzzy labels (Low, Medium, High and Very High, for Negative and Positive sides, plus Zero) are defined as illustrated in Figure 4. The non-fuzzy (numeric) values of emotion dimension (d_j in Equation.1) will act as input to fuzzy membership functions that determine how much each fuzzy label will be activated. For example a value of 20 in Figure 4 will result in a membership degree of 100% in Positive-Low and zero in other fuzzy labels, while 30 will result in about 80% in Positive-Low, 20% in Positive-Medium, and zero in others.
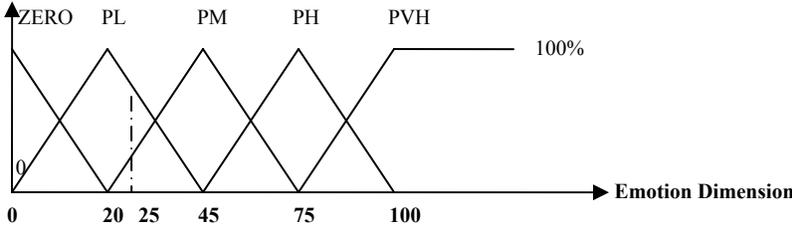

Figure 4. Fuzzy Membership Functions. L, M, and H stand for Low, Medium, and High. P and N mean Positive and Negative. The negative and positive sides are defined symmetrically.

Based on perceptual user experiments discussed in Section 4, each fuzzy label is associated with a series of facial actions with nominal values. For example, Positive-High Arousal can cause the Open_Jaw action by 40%. The dimensional values of a mixed emotion will be used to calculate the membership values in each fuzzy label which multiplied by its nominal output will activate a facial action to a certain level. Depending on associations observed in experiments, the fuzzy rule can have one or two fuzzy label as input, for example (Output activation levels are examples based on perceptual analysis data):

IF Arousal is PositiveHigh THEN OpenJaw is 40% active
IF Arousal is PositiveHigh AND Agency is PositiveMedium THEN SqueezeEyebrow is 25% active

For each rule the overall membership will be the minimum of all input membership values. This overall value, multiplied by the nominal output value, will determine the activation level of the output facial action.

$$a = a_{nom} \times MIN(m_i)$$ (2)

4 Perception-based EU Associations

The key concept in EU-based approach is the use of perceptually valid combination of facial actions to create expression of mixed emotions. To achieve this, the foundation of our approach is the study of how people combine facial actions when expressing mixed emotions as opposed to previous studies which were based on individual emotions. In order to do this we asked users to use an MPEG4-compatible 3D facial animation system to create facial images that best represent short emotional stories. For example, the following story was intended to show a mix of surprise and joy:

Story 1: *It was already getting dark. Walking back home from shopping, he was thinking "This sucks! For the first time, I got money to buy my dad the Christmas present I know he'd love but he can't be here." He knew dinner would be ready soon and his mother had promised to make all his favourites, but Christmas dinner couldn't taste good without "dad." He turned into their street, looked at their house, and suddenly stopped. The sight of his old man's car, parked in front of the house, was the most amazing gift he'd ever received!*

Our facial animation system allowed users to read the story and use the MPAG-4 FAPs to create a facial expression. The users were then asked to rate the universal emotions for the same story, to be converted to three emotional dimensions using the Equation 1. Using these ratings, we are able to not rely on any assumptions regarding the affective content of the stories, although our results showed that the intended emotions were very close to the perceived emotions from the stories. By having a series of users, and a variety of stories for each of them, we generated a database of mixed emotions and their respective facial actions. Statistical analysis of this database, i.e. the correlation of actions and dimensions, provided us with EU associations needed for fuzzy rule set.

To make the results more reliable, we ran a second series of user experiments where the images created by the first group together with the stories were shown to another group of users. They were also asked to rate the universal emotions for the stories, and also rate how successfully the images represented the emotional state of the stories. The images which were not perceived as the intended emotional state (i.e. the images for which two top perceived emotions were not the same as those

intended by the group-1 participants) were removed from the group-1 database, so they would not affect our perception-based EU associations.

5 Experimental Results

The experiments were done with volunteers from university students, 30 participants for group 1 and 20 for group 2. Each participant was shown three stories that intended to have mixed emotions of joy-surprise (story 1 shown in Section 4), anger-sadness (story 2), and fear-disgust (story 3).

Story 2: *Two things you should know about him are that he doesn't have a pet and he doesn't like chicken. Only recently, he realized that these two are in fact related. When he was 5 years old, he had two chickens. He had a few but couldn't keep most of them from crows and neighbourhood cats until these last two. He fed them almost everything, made them a house, and played with them all the time. They grew big and strong so much that his parents started to complain about how hard it was to keep them in their small yard, but he wouldn't let them go. One day, when he came back from school, he couldn't find the chickens. He went to the kitchen to ask his mother where they were but didn't have to. They had chicken for dinner.*

Story 3: *He was starting to feel lost. He had been walking on the trail, or what he thought was the trail, for an hour and had not seen a sign of his friends. It was dark and cold, and hard to find the way among the trees. Suddenly he heard a sound: someone or something was moving. The sound was followed by a wild and painful scream. He ran as fast as he could, hitting branches and hardly keeping himself from falling. When he stopped he found the surrounding familiar. He was there before. There was a low moaning sound. He moved ahead and that was when he saw it. The little deer was attacked and his body torn and mutilated, covered with blood and the remains of some internal organs he did not want to recognize, but alive. Whatever did this was still close by.*

Group 1 of participants created 90 images for three stories, as illustrated in Figure 5. The emotional ratings for stories are converted to dimensional values using the method explained in Section 3. Figure 6 shows the distribution of facial actions along the dimensions for sample images. After running the experiments with group 2, and assessing the success of these images in causing the desired perception, 25 of these images were removed from the database due to difference between emotional rating. Tables 3 and 4 summarize the resulting data in terms of average values and correlation between facial actions and emotion dimensions in successful images. An image is considered successful if (1) the participants have in both groups have selected the same emotions for corresponding story, and (2) group 2 participant has rated the image more than 50% for the story.

Figure 5. Group-1 Images. From left to right: Stories 1, 2, and 3

Table 5 shows that there is a high level of correlation between facial actions and emotion dimensions in cases like lower-raise-mouth-corner and valence, while the same action does not have significant correlation with agency. A low correlation means the facial action can happen anywhere along that dimension and does not have a clear dependence on the dimension. A high correlation, on the other hand, means if we move along that dimension there is an order in occurrence of the action. In other terms, a high correlation shows that the facial action has a clear emotional value on that dimension. After detecting the high-correlation pairs of AU-dimension, we calculated the typical fuzzy values of facial actions in those pairs along the related dimensions and in each region specified by the fuzzy labels, as shown in Table 6. For example the first row contains these rules:

IF Arousal=PM THEN OpenCloseJaw=-30%

These are in some cases combined with a second fuzzy label to form double-input rules. This happens when activation of a facial action is not found to be dependent on one dimension only but a combination of dimensional values, for example:

IF Arousal=PM AND Agency<PL THEN OpenCloseJaw=-30%

This will help in negative agency cases such as fear that have high arousal.

Table 4. Average and Standard Deviation of Participant Data (left-side facial action values similar to the right-side)

Story	1 (AVG)	1 (STDEV)	2 (AVG)	2 (STDEV)	3 (AVG)	3 (STDEV)
Joy	92.66	10.46	1.74	4.82	0	0
Sadness	10.83	16.21	75.44	19.47	38.05	29.29
Anger	1.83	4.11	63.22	23.88	13.66	18.59
Fear	0	0	5.57	13.1	85.09	16.69
Disgust	1.09	5.1	58.4	31.01	44.96	30.14
Surprise	79.74	21.28	36.4	29.5	28.57	27.34
Arousal	54.99	10.95	57.34	25.93	72.24	26.03
Valence	33.3	7.89	-89.06	27.5	-84.41	31.18
Agency	6.18	8.46	-13.26	12.37	-70.79	20.34
OpenCloseJaw	-29.59	15.39	-15.14	12.89	-40.51	22.1
LowerRaiseTopLipMiddle	-0.36	0.81	-11.32	9.68	6.17	12.64
LowerRaiseBottomLipMiddle	0.71	1.55	17.35	13.06	5.7	10.66
StretchRightMouthCorner	62.95	27.46	2.38	12.87	-6.72	23.08
LowerRaiseRightMouthCorner	35.31	16.34	-62.09	35.64	-30.97	34.28
CloseRightEyelidTop	-4.51	6.72	-16.63	11.05	-15.5	7.52
CloseRightEyelidBottom	0.65	2.93	18.53	13.51	22	16.08
LowerRaiseRightEyebrowInner	14.22	8.05	-61.08	53.35	8.76	35.11
LowerRaiseRightEyebrowMiddle	21.98	12.72	-82.59	52.54	-23.32	41.68
LowerRaiseRightEyebrowOuter	33.09	16.21	-73.84	49.11	-48.79	47.36
SqueezeRightEyebrow	0.08	2.59	0.5	10.89	-30.97	19.41

Table 5. Correlation between Emotion Parameters and Facial Actions. Values between -1 and +1. Zero means no correlation.

AU	Correlation with Arousal	Correlation with Valence	Correlation with Agency
OpenCloseJaw	-0.53433	0.107024	0.433232
LowerRaiseTopLip	0.07523	0.123375	-0.31481
LowerRaiseBottomLip	0.363023	-0.59434	0.042889
StretchMouthCorner	-0.16874	0.708319	0.521789
LowerRaiseMouthCorner	-0.27075	0.817912	0.335051
CloseEyelidTop	-0.33643	0.670673	0.336042
CloseEyelidBottom	0.519491	-0.7365	-0.42089
LowerRaiseEyebrow	-0.33987	0.730664	0.081778
SqueezeEyebrow	-0.35346	0.390539	0.695211

Table 6. Fuzzy Rules for Activating Facial Actions

| AU	Value for Fuzzy Labels NVH(1) to PVH(9)								
Arousal	NVH	NH	NM	NL	Zero	PL	PM	PH	PVH
OpenCloseJaw	0	0	0	0	0	0	-30	-50	-60
CloseEyelidTop	0	0	0	0	0	0	0	25	30
CloseEyelidBottom	0	0	0	0	0	0	0	-20	-25
Valence	NVH	NH	NM	NL	Zero	PL	PM	PH	PVH
LowerRaiseBottomLip	50	40	30	15	0	0	0	0	0
StretchMouthCorner	-15	-10	-5	-5	0	30	40	50	60
LowerRaiseMouthCorner	-50	-50	-40	-20	0	20	40	50	50
CloseEyelidTop	-30	-25	-20	-5	0	-5	-20	-25	-30
CloseEyelidBottom	26	17	8	4	0	2	15	20	25
LowerRaiseEyebrow	-80	-70	-60	-30	0	15	25	35	45
Agency	NVH	NH	NM	NL	Zero	PL	PM	PH	PVH
OpenCloseJaw	-31	-20	0	0	0	0	0	0	0
StretchMouthCorner	-10	-10	-5	0	0	-5	-10	-15	-15
SqueezeEyebrow	0	0	0	0	0	35	50	90	90
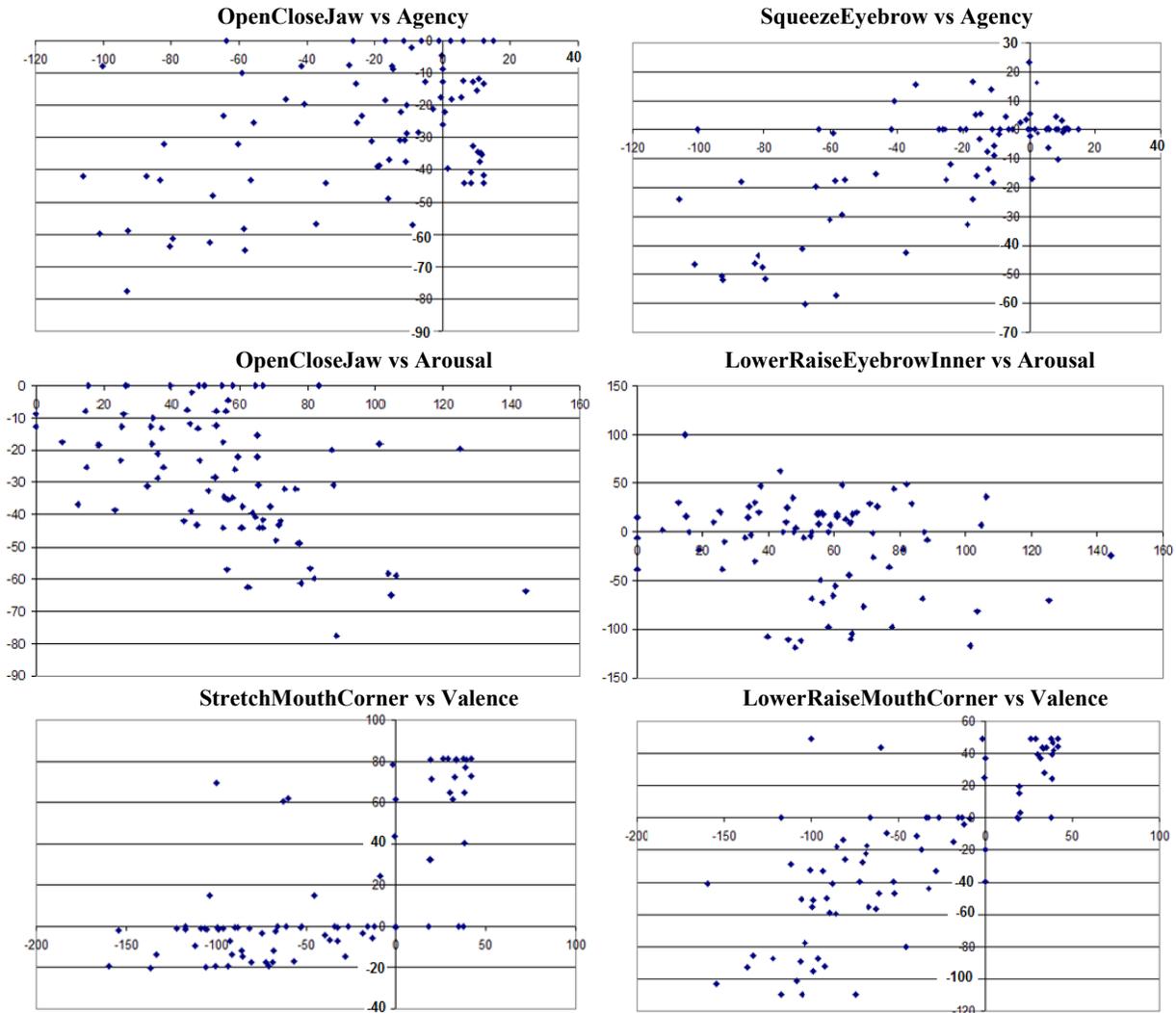
Figure 6. Example of distribution of facial action (Y-axis) along dimensions (X-axis). The data ranges are different as they are set based on available data. For example in the Open-Jaw case (top left), positive values correspond to closing of the jaw. Since the neutral face already has the mouth closed, no expression has used a positive value for this action.

To illustrate the effectiveness of EU-based method, we compared the participant images with expressions created by EU method and weighted averaging. Figure 7 illustrates the result of applying EUs to create facial expression of mixed emotions. The left image is created using the average facial actions used by participants as listed in Table 4. In other term, the image represents the average visual perception of participants for that mixed emotional state. The middle image is created based on the average values of Arousal, Valence, and Agency (Table 4) and activating facial actions as suggested by EU approach for those emotional values. Finally the right image is generated by using the average level of standard emotions specified by participants (the first 6 rows in Table 4) as weight and activating facial actions in the blend using the weighted average (i.e. activation of each action is calculated by multiplying its nominal value in the standard emotion and the weight for that emotion). Looking at details of movements in eyebrow, eyelid and mouth regions, we can see that EU's result in a much more perceptually valid result, i.e. closer to participants' suggestions. Further tests of the EU-based method are underway through sample applications.

Figure 7. Example of EU-based facial expressions. Rows 1 to 3 correspond to stories 1 to 3, and from left to right images are based on average participant input, EU method, and weighted average of activated AUs.

6 Sample Applications

Additive Uses:
Most traditional computer game systems use morph target-based facial models which can provide the advantage of a small data set (i.e. typically expression and phoneme target models) combined with open ended surface details since the morph targets can be individually sculpted at model creation time. Morph target-based facial systems, however do not provide the parametric control and perceptual validity needed for communication, health and research-based serious game development. For instance, one serious game application that we are involved in requires the system to allow expression animation, such as time based multiple expression layers, to be decoupled from voice animation, as well as to be combined to any voice, emotion or expression layers in perceptually valid ways. We refer to this as an *additive* use of expression units. The lip-sync movements are calculated using a different system that determines the position of mouth area vertices. The facial actions for expressions are calculated using EUs and then "modulated" over the lip-sync actions.

In conjunction with the Canadian Parks department, we have prototyped and are in design discussions for the creation of a language learning game that exposes youth on native reservations to characters who speak in their specific native language. The goal of this effort is to foster learning and exposure to the local native language by creating customized video games where characters in the game speak their local native language via perceptually valid facial avatars. The hope is that the game project, with other efforts, sparks interest in the local native language in upcoming generations and will help to stop the disappearance of hundreds of native languages. The issue for the parks department is that there are thousands of native languages in Canada, making specifically authoring a game for each language cost prohibitive. By using our parameterized facial system, with the EU-based approach outlined in this paper, it is possible to make one template game, with full emotive and expressive 3D characters where different voice and lip-sync parameters can be layered in, either as a simple pre-process or in real-time. It is quite feasible then, given a detailed dialog script and character back-story description, to employ a local native speaker from each target native community for a few days of voice recording to voice the script which then can be automatically layered into the facial expression to create the finished lip-sync characters. This allows for one general e-learning production effort to create a base video game (complete with emotional character interactively reacting to programmed game scenarios) that can spawn thousands of language variants at reasonable cost. Figure 8 depicts example screen shots from a prototype animation coordinated with the Parks Department where a man speaking in an arbitrary local native language can be layered with perceptually valid expressions under real-time program control. In this prototype, the character changes into a native mask while he is speaking.
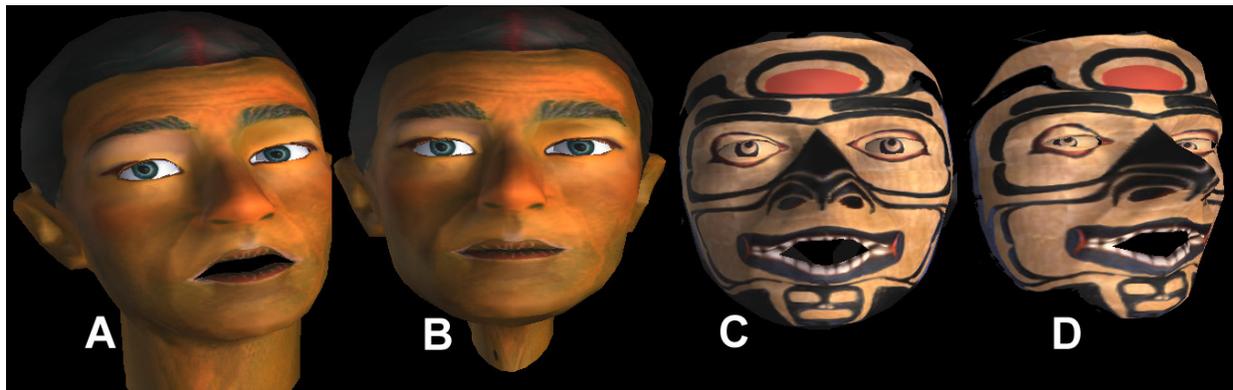

Figure 8. Screen shots of real-time interactive animation layering additive voice, expression-unit and facial unit layers.

Deconstructive Uses:
Autism is one of the most commonly diagnosed developmental disorders in children [30]. It is characterized by severe impairment in the domains of social, communicative, cognitive, and behavioral functioning. Despite the inconsistent profiles across individuals with autism, recent research reflects a developing consensus that the most central behavioral symptom of autism is the impairment of social behavior, such as impairment of the social use of gaze, failure of joint attention interaction, lack of interest in peers, difficulty initiating communication, preference for solitary play. Research has shown that individuals with autism fail to understand the emotional state expressed by another person. Whether individuals with autism rely on isolated parts of the face, such as eyes or bushy eyebrows, to recognize faces rather than whole faces is not clear. That is because static images do not provide realistic assessment of face processing and emotion recognition. Our research helps study social competency and facial expression awareness in autistic children by combining our configurable 3D facial communication input system with a real-time eye gaze measurement system. This study involves varying the visual and social complexity of a dynamic 3D face in discrete steps to measure facial awareness and social competency (emotional recognition). The goal of the work is to both: 1) better understand this issue from a scientific and clinical point of view and 2) to build a computer based communication and serious game systems that take advantage of the clinical and technical results of the research. Our study investigates whether different image rendering styles as well as varying levels of expressions scenarios, such as percent of direct eye gaze and normal to dampened mouth movement, can affect autistic children's visual perception and their social competence. We are using 'The

Awareness of Social Inference Test' (TASIT) [18] to measure results. The TASIT test comprises videoed sequences in three parts assessing (1) emotion recognition, (2) the ability to understand when a conversational inference such as sarcasm is being made, and (3) the ability to differentiate between different kinds of counterfactual comments (lies and sarcasm). This work will be in conjunction with ongoing work using TASIT video sequences.

By presenting each social sequence under controlled facial animations with differing rendering styles and levels of expression situations on high developed autistic teenagers, we can 1) monitor eye-gaze patterns and retention time along with 2) how well they correctly identify the emotion associated with TASIT animated sequence. Unlike the TASIT video of real-time actors, our system now allows us to understand each sequence deconstructed into their building block expressions. This allows detailed output measurements for every expression unit and the ability to identify gross expressions into their component parts. We then can control and modify any level of expression unit to isolate and validate problem areas. Currently most autism individuals miss important eye/eyebrow expression cues by overly gazing at the quick movement in mouth area. Our first set of experiments will systematically dampen mouth movements in perceptually valid ways, so the lip sync will look natural, but will less and less overwhelm the subject. We then can track at what level of mouth movement dampening, balancing eye gaze between eye, mouth and other head areas returns. Other experiments involve varying the percentage of direct eye gaze as well as changing the level of realism of rendering from very realistic to non photorealistic drawn like rendering (see Figure 9). The goal is to understand how the complicated factors of social complexity versus visual complexity effect face to face communication in autism subjects. Because autism is a spectrum, individuals vary widely in how they perceive faces. Being able to deconstruct, output and modify expression units at any level of the animated 3D face animations allows us to conduct very detailed and modifiable experiments as well as easily transfer knowledge from the experiments over to consumer learning serious games toolkits for autism.

Figure 9. TASIT social sequences using expression units 3D rendered realistically (left) and in a drawing style (right).

7 Conclusion

We proposed Expression Units as a method to create perceptually valid facial expressions for blending and transition of universal emotions. We have aimed at studying the individual emotional effect of facial actions. Our approach is based on (1) the analysis of viewers' perception of facial actions, and (2) associating facial actions to emotion dimensions rather than individual emotions. We use a three-dimensional emotion space consisting of arousal, valence, and agency.

We showed that many facial actions have clear correlation with these dimensions and used these correlations to define expression units as facial actions that are controlled by emotion dimensions. The resulting "expression units" are in general compatible with previous findings (Table-1 for example). There are some exceptions such as lowering eyebrows which should happen at a low level with high valence, but in our finding they are slightly raised instead. This can be explained by common misconception in our participants and the low impact of this action compared to other actions such as those in mouth area, for perception. Also, our data does not include regions in emotional space such as low arousal (not used by six universal emotions).

The proposed fuzzy rule-based system is what can be implemented in any programmable facial animation system where animators can set the dimension values and the system uses rules shown in Table 6 to automatically activate the facial actions to create the expression. Our results can be used in two ways: (1) by animators as guidelines for mixing facial expressions, and (2) by developers of facial animation systems to implement algorithms for automating process in a perceptually valid form. More studies will allow us to understand the mechanism of mixing facial expressions better and generate more efficient EU associations. We particularly need to use more stories that cover full 3D emotional space and also more participants to create a stronger statistical sample set.

References

[1] Albrecht, I., M. Schröder, J. Haber, H. Seidel, "Mixed feelings: expression of non-basic emotions in a muscle-based talking head," *Virtual Reality* 8(4): 201-212, 2005

[2] Arya, A., S. DiPaola, L. Jefferies, and J. T. Enns, "Socially communicative characters for interactive applications," *14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG-2006)*, University of West Bohemia, Plzen, Czech Republic, January 30 - February 3, 2006

[3] Battista, S., et al. MPEG-4: "A Multimedia Standard for the Third Millennium," *IEEE Multimedia*, vol. 6, no. 4, IEEE Press, 1999

[4] Becker, C., and I. Wachsmuth, "Modeling Primary and Secondary Emotions for a Believable Communication Agent." *Proceedings of International Workshop on Emotion and Computing.* 2006

[5] Beier T. and S. Neely, "Feature-based image metamorphosis," *Computer Graphics*, ACM SIGGRAPH, July 1992

[6] Breazeal, C., "Affective interaction between humans and robots," *Proceedings of the 2001 European Conference on Artificial Life.* 2001

[7] Bregler, C., et al., "Video rewrite: Driving visual speech with audio," *Computer Graphics*, ACM SIGGRAPH. 1997

[8] Byun M, and N. I. Badler, "FacEMOTE: Qualitative Parametric Modifiers for Facial Animations," *Proceedings of ACM Computer Animation Symposium.* 2002

[9] Chang, J. et al. "Facial Expression Analysis and Synthesis : a Bilinear Approach," *Proceedings of International Conference on Information Acquisition.* 2007

[10] Chuang, E. S. et al. "Facial expression space learning," *Proceedings of 10th Pacific Conference on Computer Graphics and Applications.* 2002

[11] Deng, Z. et al., "Expressive Facial Animation Synthesis by Learning Speech Co-Articulation and Expression Spaces," *IEEE Transactions on Visualization and Computer Graphics*, Volume 12, Issue 6, Nov.-Dec. 2006 Page(s):1523 – 1534. 2006

[12] Ekman, P. , and W. V. Friesen, *Facial Action Coding System,* Consulting Psychologists Press Inc. 1978

[13] Ekman, P. *Emotions Revealed,* Times Books. 2003

[14] Grammer, K., and E. Oberzaucher, "The Reconstruction of Facial Expressions in Embodied Systems New Approaches to an Old Problem,"

[15] Kaiser, S., and T. Wehrle, "The Role of Facial Expression in Intra-individual and Inter-individual Emotion Regulation,"

[16] Lerner, J.S., and D. Keltner, "Beyond valence: Toward a model of emotion-specific influences on judgment and choice," *Cognition and Emotion*, 14, 473–493. 2000

[17] Lin, D.T., and H. Huang, "Facial Expression Morphing and Animation with Local Warping Methods," *Proceedings of 10th International Conference on Image Analysis and Processing.* 1999

[18] McDonald, S. et al, "The Ecological Validity of TASIT: A Test of Social Perception," *Neuropsychological Rehabilitation*, Volume 14, Number 3, July 2004

[19] Martin, J.C., R. Niewiadomski, L. Devillers, S. Buisine, C. Pelachaud, "Multimodal complex emotions: Gesture expressivity and blended facial expressions," *International Journal of Humanoid Robotics*, Special Edition "Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids", 2006

[20] Mehrabian, A. "Framework for a comprehensive description and measurement of emotional states," *Genetic, Social, and General Psychology Monographs*, 121:339–361. 1995

[21] Mufti, M. and A. Khanam, "Fuzzy Rule-Based Facial Expression Recognition," *Proceedings of International Conference on Intelligent Agents, Web Technologies and Internet Commerce.* 2006

[22] Naiwala P. et al., "Interactive Analysis and Synthesis of Facial Expressions based on Personal Facial Expression Space," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition.* 2004.

[23] Parke, F. I., and K. Waters, *Computer Facial Animation*. A. K. Peters. 2000

[24] Pighin, F., et al., "Synthesizing realistic facial expressions from photographs," *Proceedings of Computer Graphics, Annual Conference Series.* 1998

[25] Pyun, H., et al., "An Example-Based Approach for Facial Expression Cloning," *Proceedings of Eurographics/SIGGRAPH Symposium on Computer Animation.* 2003

[26] Russell, J. A., "A circumplex model of affect," *Journal of Personality and Social Psychology*, 39, 1161–1178. 1980

[27] Russell, J. A., "Core Affect and the Construction of Emotions," *Psychological Review*, 110: 145-72. 2003

[28] Ruth, J. A. et al., "Linking Thoughts to Feelings: Investigating Cognitive Appraisals and Consumption Emotions in a Mixed Emotions Context," *Journal of the Academy of Marketing Science*, 30 (1), 44-58. 2002

[29] Scherer, K. R. et al., "What determines a feeling's position in affective space," *Cognition and Emotion*, 20 (1), 92-113. 2006

[30] Schreibman, L., The Science and Fiction of Autism, Harvard University Press, 2007

[31] Seize, S. M., and C. R. Dyer, "View morphing," *Computer Graphics*, pages 21–30. ACM SIGGRAPH, August 1996

[32] Smith, C. A. and H. S. Scott, "A componential approach to the meaning of facial expressions," *J. A. Russell & J. M. Fernandez-Dols (Eds.), The psychology of facial expression,* pp. 229-254. New York: Cambridge University Press. 1997

[33] Thayer, R.E., *The Biopsychology of Mood and Arousal*, Oxford University Press, New York. 1989

[34] Tiedens, L. Z. and S. Linton, "Judgment under emotional certainty and uncertainty: The effects of specific emotions and their associated certainty appraisals on cognitive processing," *Journal of Personality and Social Psychology*, 81, 973-988. 2001

[35] Tsapatsoulis, N., A. Raouzaiou, S. Kollias, R. Cowie and E. Douglas-Cowie, "Emotion Recognition and Synthesis based on MPEG-4 FAPs," *Igor Pandzic, R. Forchheimer (eds), MPEG-4 Facial Animation*, John Wiley & Sons, UK, 2002

[36] Wang, H and N. Ahuja, "Facial expression decomposition," *Proceedings of 9th IEEE International Conference on Computer Vision.* 2003

[37] Williams, L. "Performace-driven facial animation," *Computer Graphics*, pages 235–242. ACM SIGGRAPH, August 1990

[38] Wolberg, G., *Digital Image Warping*. IEEE Computer Society Press, 1990

[39] Yang, B. and P. Jia P., "Synthesis of Combined Facial Expressions Using Anatomy-based Model," *Proceedings of IMACS Multiconference on Computational Engineering in Systems Applications.* 2006

[40] Zadeh, L. A., "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 1. 1973

[41] Zalewski L. and S. Gong, S., "2D Statistical Models of Facial Expressions for Realistic 3D Avatar Animation," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 2005