What a Feeling: Learning Facial Expressions and Emotions

What a feeling: Aprender Expressões Faciais e Emoções

Verónica Orvalho

Faculdade de Ciências da Universidade do Porto

IT - Instituto de Telecomunicaçoes

verónica.orvalho@dcc.fc.up.pt

José Miranda

Faculdade de Engenharia da Universidade do Porto

UDI/IPG - Unidade de Investigação para o Desenvolvimento do Interior/IP Guarda

Augusto Sousa

Faculdade de Engenharia da Universidade do Porto

INESC- Porto - Instituto de Engenharia de Sistemas e Computadores do Porto

Resumo

As pessoas que sofrem de autismo apresentam uma enorme dificuldade em reconhecer as expressões faciais. A nossa abordagem está focada num dos maiores problemas identificados neste tipo de doença: a capacidade de reconhecer os sentimentos dos outros. *What a feeling* é um videojogo cujo objectivo é melhorar a capacidade de individuos, social e emocionalmente diminuídos, no reconhecimento de emoções através da expressão facial. O jogo desenvolvido permite, através de um conjunto de exercícios, que qualquer pessoa de qualquer idade possa interagir com modelos 3D e aprender sobre as expressões da face.O jogo é baseado em síntese facial em tempo real. Este artigo descreve a mecânica da nossa metodología de aprendizagem e discute algumas linhas de orientação para trabalho futuro.

Palavras-chave: *animação facial, expressões faciais, autismo, videojogos*

Abstract

People with Autism Spectrum Disorders (ASD) find it difficult to understand facial expressions. We present a new approach that targets one of the core symptomatic deficits in ASD: the ability to recognize the feeling states of others. What a Feeling is a videogame that aims to improve the ability of socially and emotionally impaired individuals to recognize and respond to emotions conveyed by the face in a playful way. It enables people from all ages to interact with 3D avatars and learn facial expressions through a set of exercises. The game engine is based on real-time facial synthesis. This paper describes the core mechanics of our learning methodology and discusses future evaluation directions.

Keywords: *facial animation, facial expressions, autism spectrum disorders, serious games*

Figure 1: Skeleton facial rig and two facial poses (left); 3D characters from the game (middle); screen shot of the game (right).

1. Introduction

The face is the key element to convey emotion and plays an important role in verbal and non-verbal communication. Many efforts have been done to teach people with Autism Spectrum Disorders (ASD) to recognize facial expressions with varying results [Golan et al., 2008], but none focused on using real time facial synthesis. Most methodologies use Paul Ekman's approach based on photographs of facial expressions and FACS (Facial Action Coding System) [Ekman and Friesen, 1975]. Besides having severely limited interactivity, they fail to reproduce the dynamics of a facial expression: far from being a still image, it is the voluntary and involuntary contraction of muscles that produces different facial movements. These movements convey emotions from one individual to another, enabling non-verbal communication. Thus, we need to come up with an additional teaching method that includes facial motion and uses realtime facial synthesis.

We argue that current technological advances in character animation can substantially improve the way we teach people with ASD to recognize facial expressions and emotions.

We present a videogame designed to assist people with ASD to recognize facial expressions in a playful way, without induction of stress. The core technology builds on a facial synthesis method [Orvalho, 2007] we have developed, that eases the real time animation process. The main research challenges arise from the synchronization and realism problems, the support for the reusability of facial components, and the need for an avatar-user interaction model with real time response. The videogame contains a set of exercises embedded in the gameplay that reinforce the learning process and generate a real time avatar response based on direct user input or on a set of predefined rules. We also include a facial expression editor capable of displaying 3D characters in real-time. This allows the therapist to adjust or create new exercises on the fly, without the need of artistic or technical skills. The main outcome is a comprehensive system that analyzes the users input and enables avatars to respond to it in real-time. The avatars are 3D models with underlying anatomical behavior, and have a wide range of visual styles: human, cartoon or fantastic creature (see Figure 1).

2. Background

In previous work where people interact with a virtual character, the virtual character representations have been quite simple, and little or no attention was paid to facial expression. Here we intend to incorporate the full range of facial expressions based on a geometric, skeletal and muscle model, allowing for a much greater repertoire of possible expressions.

2.1. Clinical

According to [American-Psychiatric-Association, 2000], ASD are characterized by the presence of a markedly abnormal or impaired development in social interaction and communication and a marked restricted repertoire of activities and interests. The manifestations of ASD vary greatly, and the term ASD is used to reflect the heterogeneity of impairments, ranging from mild to severe. The loss in reciprocal social interaction is broad and persistent. Impairments in face processing are well documented in autism [Hughes, 2008]. Individuals with ASD experience a variety of difficulties and use atypical strategies in both face recognition and the identification of emotional facial expressions that persist through adulthood [Grossman and Tager-Flusberg, 2008]. ASD are chronic conditions, for which no complete cure has yet been found.

The general agreement is that early diagnosis followed by appropriate treatment can improve outcomes in later years for most individuals. Thus, the question of how various interventions may help to increase the individuals ability to function is highly relevant to families, health professionals, and policy makers. Computers house a large potential for helping affected individuals to acquire underdeveloped capacities [Bolte,]. A especially promising technique is video self modeling, where the motivation to watch oneself on the video may be enhanced by the portrayal of predominantly positive and successful behaviours, which may also increase attention and enhance self-efficacy [Bellini and Akullian, 2007].

2.2. Technological

Animating virtual characters is challenging, as the face is capable of producing about 5000 different expressions. Much effort has gone into retargeting or automatically creating the body rig of characters [Baran and Popovic, 2007; Hecker et al., 2008] and some good results have achieved. But these efforts focused mainly on mesh retargeting, and none on full facial rig retargeting. Mesh retargeting does not solve the problem of mapping an animation from a human to a non-realistic character, because of the diverse face proportions and geometry [Orvalho, 2007]. The core technology builds on our previous research, where we developed a facial synthesis method capable of transferring facial rig and animations between dissimilar characters [Orvalho, 2007].

2.3. Multimedia Solutions

There are several solutions currently available aimed at helping people with ASD improve recognition of facial expressions.

Mind Reading helps children and adults learn facial emotions and expressions. It provides an interactive guide and a video library of people showing different emotions. It also includes a set of quizzes and games to check the learning progress [Baron-Cohen et al., 2004].

The ***Transporters*** is an animation series designed to help children with autism to discover the world of emotions. The series involves 3D objects, which are mechanical vehicles with human faces showing emotional expressions in social context. The goal is to catch the children's attention with mechanical motion to encourage incidental social learning and increase awareness of the face. This project shows great progress in emotion recognition skills with autistic children [Baron-Cohen et al., 2007].

Children with ASD usually avoid eye contact with others. FaceSay focus in teaching where to look for facial cues, such as an eye gaze or a facial expression. The study [Biasini and Hopkins, 2007] found that children using FaceSay made significant improvement in reading facial expressions and emotions.

We believe that multimedia solutions that use lifelike stimuli can improve facial expression recognition with only relatively brief training period.

3. Repercussion

The What a Feeling videogame has important repercussions in at least three different areas:

Technological Our technology can speed up the character animation process as demonstrated in [Orvalho, 2007]. The videogame prototype allows real time preview and editing of facial expressions. It enables creating animations on the fly by scripting facial behaviors, using our facial expression analyzer and classifier system. Even non-experienced artists will be able to create high quality animations.

Clinical The videogame prototype will be used for validating the new approach: helping autistic people in identifying and expressing emotions. The novelty is the interactivity of the process, as opposed to previous methods that used still images.

Societal Integration of socially impaired people, which will help them become a contributing member of an advanced society.

4. What a Feeling: the videogame

The videogame prototype has three ways to control avatar behavior: avatars controlled by the system, avatars controlled by the patient and avatars controlled by the therapist. The patient and the therapist can directly manipulate the character by assigning it an expression in real time or by creating a new expression. The therapist can also create different exercises for facial expression recognition based on the core mechanics embedded in the game. This paper describes the core mechanics of the videogame and the facial synthesis game module, as it is the most relevant component of the game engine.

4.1. Facial Synthesis Game Engine

The core technology behind the prototype is based on the facial synthesis of 3D characters. It strives to solve the synchronization and realism problems, support reusability of facial components, and have an avatar-user interaction model with real time response. The game engine supports skeletal and blend shape animation.

Reproducing the subtleties of a face through animation requires developing a sophisticated facial rig. A rig is a set of controls that allows animating a character. This job becomes more challenging if we aim to animate the face in real time and have it respond to specific user input. For this project, we created a sophisticated facial rig that animates in real time. This is not a trivial task, as we wanted to reuse the rig in characters with different appearance, geometry and proportions. Thus, this rig becomes the foundation of the system pipeline. It will ensure that the characters animations follow a consistent artistic style, to ease the process of recognizing facial expressions and emotions. Using our rig retargeting technique [Orvalho, 2007], we automatically transfer the rig from the generic character to each game character.

The generic rig is built in three layers: bones, blend shapes and pre-set basic expressions. The **bone structure** has 34 main bones that control the geometry of the face. Figure 2 shows the bone distribution on the generic character and on one character from the videogame. The **blend shapes** serve as constraints when the user manipulates the characters, and guarantee the correct deformation of the character. Each bone has associated a set of blend shapes: when a bone moves in any of its axis the correspondent blend shapes are triggered. Finally, we have defined six **basic expressions** that can be easily combined to generate new expressions. The six basic expressions are: anger, disgust, fear, joy, sadness and surprise and can be combined to generate the full range of human expressions [Faigin, 1990; McLoud, 1994].

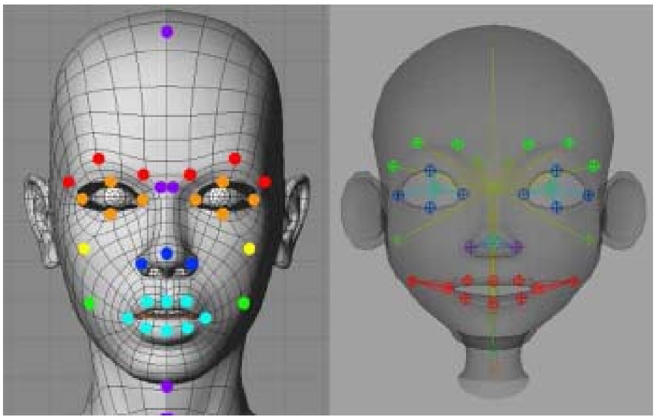

Figure 2: Generic rig bone distribution (left) - it includes 34 main bones: 5 violet bones to control the neck, head and upper nose, 6 red bones to control the eyebrows, 8 orange bones to control the eyes, 2 yellow bones to control the cheeks, 2 green bones to control the lower part of the cheeks, 8 light blue bones to control the mouth and 3 dark blue bones to control the nose. Bone distribution of a character from the videogame after transferring the rig from the generic character (rigth).

4.2. Core Game Mechanics

The driving force of this project is to teach how to visually recognize the emotion of others. It tackles one of the core symptomatic deficits in ASD: how to recognize facial emotions and expressions on different people. To accomplish this we included in the game design a playful interaction scheme to engage the patients while teaching them difficult concepts. The videogame stirs the cooperation between the therapist and the patient. It also stimulates the competition in the patient, which tries to beat previous scores. The cooperation and competition game techniques allow the patient to become more immersed in the game, in contrast to traditional facial recognition applications.

The game interaction model uses an avatar, represented by a face model, and the game perspective is third person.

The core mechanics are based on identifying “micro” and “subtle” expressions and analyzing the three emotion concepts: intensity, length and valence.

Intensity: subtle expression vs. strong expression Emotions usually vary in intensity. Strong expressions involve the whole face in large facial movements that can be obvious. Subtle expressions involve small facial movements that often appear only in one region of the face, such as just the brows, eyelids, cheeks, nose or lips. These small movements occur gradually when an emotion begins, when an emotion is repressed, or when an emotion is deliberately masked.

Length: micro expression vs. macro expression Emotions vary in how long they last. A micro expression can be as short as 1/25 to 1/5 of a second. On the other hand, [Ekman and Friesen, 1975] gives the average time for an emotion to unfold to be 2-4 seconds (surprise is the shortest universal emotion). Micro expressions are signs of emotions just emerging, emotions expressed before the person displaying them knows what she is feeling, or emotions the person is trying to conceal.

Valence: positive expressions vs. negative expression Some expressions are termed positive, such as joy. Other expressions are termed negative, such as anger, fear, sad, and disgust. Some expressions are neutral in this spectrum, such as surprise.

There are specific rules for each game scenario that aim to enhance the experience of the learning process, without interfering the main goal of the videogame (see section 5 for a detailed description).

There is a time based system where the player has to complete the mini tasks and lessons to score points. The style of interaction is similar to *Guitar Hero* [http://www.guitarhero.com] that has successfully showed an enjoyable gameplay.

5. Results

What a Feeling is implemented in C++, using OGRE SDK [http://www.ogre3d.org/] for the graphic rendering engine and wxWidget [http://www.wxwidgets.org/]for the GUI interface. All 3D models were created using Maya [http://www.autodesk.com]. The game runs in real time and was tested on a PC with an Intel Core 2 Duo, with 2GB RAM and an ATI Radeon X1600 graphics card.

The videogame prototype [http://www.dcc.fc.up.pt/~veronica.orvalho/tlife/demo.rar] has two modes: one for the patient and one for the therapist. Figure 3 shows the patient mode and a close up of the therapist mode.

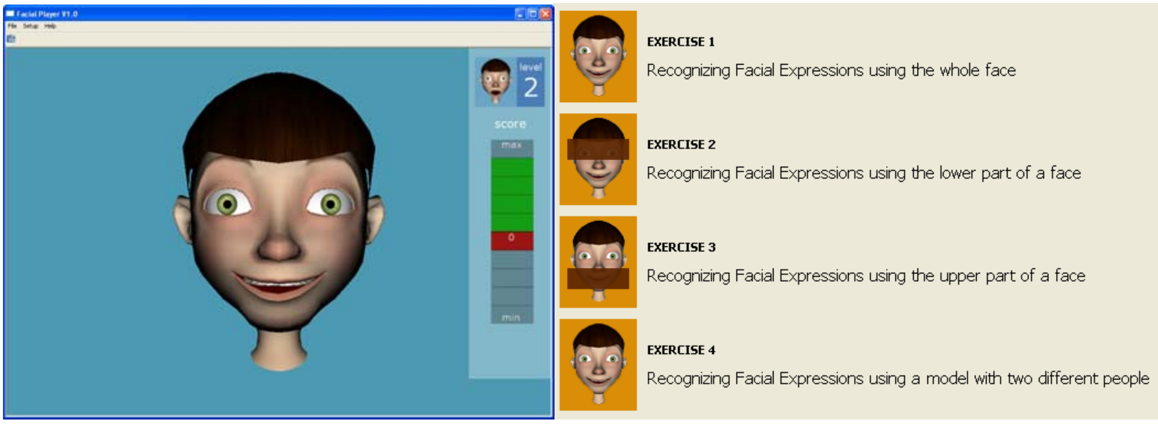

Figure 3: Patient Mode - main window with the current expression and left panel with auxiliary information (left); Therapist Mode - close up of different configurable scenarios available (right).

Patient mode *the game itself*

The main goal is to recognize the facial expression that the therapist has chosen in each exercise.

Therapist mode *the game setup and exercise definition*

The setup menu allows defining which expressions will show in the game and how fast they will appear (see section 4.2).

There are two types of exercises for the patient to perform:

Facial Expression Recognition the patient has to recognize a specific facial expression from a list of expressions that are displayed consecutively on the screen.

Facial Expression Creation the patient can create a specific expression following the therapist's instructions.

Each type of exercise can have different configurable scenarios where the avatar appears with a full face, half face or a mix face:

1. Full face: uses all the face to display the expressions. This scenario allows teaching expressions associated to a specific type of avatar.
2. Half face: either the upper part or the lower part of the face is hidden. This scenario allows teaching a specific section of an expression.
3. Mix face: uses half of the face from one character and the other half of the face from another character. This scenario allows teaching expressions by dissociating them from the type of avatar.

The main challenge defined in the gameplay is directly related with the number of expressions and the speed at which they are displayed. The patient user interface consists of a main window that displays the avatar, and a left panel that shows the score status, the expression to identify and the difficulty level. It is critical to keep the user interface simple so the patient only needs to concentrate on recognizing the facial expressions. Figure 4 shows a screen shot of the patient user interface. Finally, to evaluate the performance of the patient at the end of the exercise, a report with the score appears on the screen. Figure 7 shows a screen shot of the game report.

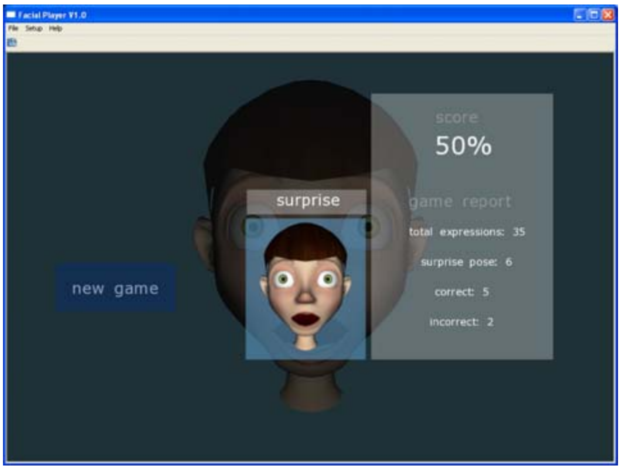

Figure 4: Patient Mode - end of exercise report.

6. Conclusion

What a Feeling is a videogame prototype that teaches how to recognize facial expressions. It allows patients to become very immersed in the experience, while also enabling therapists and relatives to easily create new exercises. Now, recognizing facial expressions is something to be enjoyed rather than a difficult chore.

Our next step is to test this approach in a real environment. So far, we have assembled an interdisciplinary team of experts with vast experience in psychological mental disorders.

Acknowledgements

Many thanks to Xenxo Alvarez and Face In Motion for providing the 3Dmodels. We also thank Pamela Davis for sharing with us her knowledge about the representation of facial emotions. This work is partially supported by Instituto de Telecomunicações (IT) [http://www.it.pt] and FCT [http://www.fct.mctes.pt].

References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision. Washington, DC: APA.

Baran, I., and Popovic, J.(2007). *Automatic rigging and animation of 3d characters*. ACM SIGGRAPH 2007.

Baron-Cohen, S., Golan, O., Chapman, E. and Granader, Y. (2007). *Transported into a world of emotion*. The Psychologist 20(2):76-77

Baron-Cohen, S., Golan, O., Wheelwright, S., and Hill, J. J.(2004). *Mind reading: The interactive guide to emotions*. London: Jessica Kingsley Limited (www.jkp.com).

Bellini, S., and Akullian, J. (2007). *A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders*. Exceptional Children, 73, 261-284.

Biasini, F and Hopkins, M. (2007). *Demonstration and Evaluation of Avatar assistant: Encouraging Social Development in children with Autism Spectrum Disorders*. University of Alabama at Birmingham.

Ekman, P., and Friesen, W. (1975). *Unmasking the face. a guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Faigin, G. (1990). *The artist's complete guide to facial expression*. Wattson-Guptill.

Golan, O., Baron-Cohen, S., Ashwin, E., Day, K., McLintock, S., Leggett, V. (2008). *Facilitating emotional understanding and face-processing in young children with autism spectrum conditions using animations of vehicles with faces*. IMFAR.

Bolte, S. (date?). *Computer-based intervention in autism spectrum disorders*. O.T.Ryaskin (Ed.).

Grossman, R., and Tager-Flusberg, H. (2008). *Reading faces for information about words and emotions in adolescents with autism*. Research in Autism Spectrum Disorders 2.

Hecker, C., Raabe, B., Maynard, J., and Prooijen, K. (2008). *Real-time motion retargeting to highly varied user created morphologies*. ACM SIGGRAPH 2008.

Hughes, J. (2008.) *A review of recent reports on autism: 1000 studies*. Epilepsy and Behavior 13.

Mcloud, S. (1994). *Understanding comics*. Harper Ed.

Orvalho, V. (2007). *Reusable facial rigging and animation: Create once, use many*. PhD thesis, Universitat Politecnica de Catalunya.