

Interactive expressive virtual characters: challenges for conducting experimental studies about multimodal social interaction

M Courgeon¹, O Grynszpan², S Buisine³, J-C Martin¹

¹LIMSI-CNRS,
BP 133, Orsay, FRANCE

²CNRS USR 3246, Université Pierre et Marie Curie,
Paris, FRANCE

³Arts et Métiers Paris Tech, LCPI,
151 boulevard de l'Hôpital, 75013 Paris, FRANCE

courgeon@limsi.fr, ouriel.grynszpan@upmc.fr, stephanie.buisine@ensam.eu, martin@limsi.fr

¹www.limsi.fr, ²www.centre-emotion.upmc.fr, ³www.ensam.fr

ABSTRACT

Advanced studies about social interaction address several challenges of virtual character research. In this paper, we focus on the two following capacities of virtual characters that are the focus of research in human-computer interaction and affective computing research: 1) real-time social interaction, and 2) multimodal expression of social signals. We explain the current challenges with respect to these two capacities and survey how some of them are used in experimental studies with users having Autism Spectrum Disorders (ASD).

1. INTRODUCTION

Interactive virtual characters are expected to lead to an intuitive interaction via multiple communicative functions, such as the expression of social signals via their facial expressions, speech and postures. These characters are often used as tools for conducting experimental studies on the way users perceive facial expressions of emotions. Multiple studies make use of canned animations of non-interactive virtual characters and ask subjects to report how they perceive the virtual character.

Research related to the dynamic generation of virtual agents' nonverbal behaviors stresses the importance of defining their temporal coordination with speech-based communication. One challenge for virtual agents platforms is to control very precisely the synchronization of communication channels (Gratch et al, 2002). In terms of software architecture, this implies simultaneous generation of these various communication channels from a unique representation (e.g. facial expressions should not be derived from the speech content but must be generated simultaneously). BEAT (Behavior Expression Animation Toolkit (Cassell et al. 2001)) is an example of a framework allowing the automatic generation of animations synchronizing speech synthesis, voice intonation, eyebrow movements, gaze direction, and hand gestures. From a functional standpoint (Scherer, 1980), facial expressions can take on semantic (e.g. to emphasize or substitute for a word), syntactic (e.g. nodding, raising eyebrows to emphasize parts of the speech flow), dialogic (e.g. gazes to regulate speech turns) or pragmatic (e.g. expressing the speaker's personality, emotions or attitudes) functions in a conversation. Rules for coordination of facial expression with speech depend on these functions (Krahmer and Swerts, 2009).

Anthropomorphic virtual characters are also being used in conjunction with eye-tracking technology to study social gaze in humans (Wilms et al., 2010) and to train children with ASD (Lahiri et al, 2011). Eyetrackers enable to control on-line the gaze direction of users and to change correspondingly the displayed information.

Several studies in the field of affective computing use virtual characters and a categorical approach to emotion in which a single expression of an emotion category, such as anger, is displayed in a single modality (e.g. facial expression). Some cognitive theories of emotion suggest a more complex dynamics in the emotional process and the corresponding display of facial expressions. For example, the Component Process Model suggests that current events are appraised according to a sequential flow of criteria and that corresponding facial signs are displayed sequentially (Scherer 2010). The MARC platform features a model

of appraisal adapted from this Component Process Model (Courgeon et al. 2008). Facial signs of appraisal are displayed during a real-time interactive game with the users.

Furthermore, virtual characters need to be able to express emotions in several modalities in a coordinated fashion. For example, facial expressions need to be coordinated with speech, lip-sync and bodily expressions during congruent and incongruent combinations (de Gelder et al. in Press).

The goal of this future vision paper is to survey the current challenges for designing advanced interactive virtual characters endowed with social interaction capabilities and their current and future application with users with ASD for experimental studies about social interaction.

2. METHODS AND RESULTS

In this paper, we argue in favour of using an empirical methodology based on series of experimental designs implemented with groups of typical participants and participants having cognitive disabilities. This scaffolding process enables validating hypotheses regarding human-computer interaction that are instrumental in designing appropriate software for users with disabilities. Although it requires conducting several experiments prior to the development of the proper training application used for treatment and its underlying interactive model, it seems useful to avoid software design premises based on typical populations, but that would prove inappropriate for autism (Grynszpan et al., 2007).

We hereafter illustrate this approach using examples of such experimental series.

2.1 *Perception of the coordination between speech and facial expressions*

We conducted an empirical study seeking to explore the influence of the temporal coordination between speech and facial expressions of emotions on the perception of these multimodal expression by users (measuring their performance in this task, the perceived realism of behavior, and user preferences) (Buisine et al. 2010). We generated five different conditions of temporal coordination between facial expression of a virtual character and speech: facial expression displayed before a speech utterance, at the beginning of the utterance, throughout, at the end of, or following the utterance. Subjects recognized emotions most efficiently when facial expressions were displayed at the end of the spoken sentence. However, the combination users viewed as most realistic, preferred over others, was the display of the facial expression throughout speech utterance.

These results yielded graphic design guidelines for developing expressive virtual humans used for training social dialog understanding in ASD. Considering the outcomes for typical individuals, we reasoned that multimodal expression of emotion would be optimized when the facial expression was displayed throughout speech utterance and remained after it ended. Given difficulties in ASD regarding emotion recognition, this temporal pattern was thought to most likely enhance their ability to interact with the virtual character.

2.2 *Perception of the coordination between speech and facial expressions*

In a second study using a gaze-contingent graphic display, we developed a novel method for investigating social gaze during face-to-face encounters with a realistic virtual human that could both speak and produce facial expressions of emotions (Grynszpan et al., 2011). Experimentations carried out with 13 adults and adolescents having High Functioning Autism Spectrum Disorders (HFASD) provided evidence for alterations in the ability of individuals with HFASD to self-monitor their gaze in a social context (Grynszpan et al, 2012a,b). This empirical evaluation also suggested that, although the comprehension scores of participants with ASD were beneath those of matched typical participants, they were able to improve over the course of the experimental trials. Their performances appeared to be correlated with the time spent looking at the facial expressions when their visual field was restrained to a viewing window around their focal point using a gaze-contingent eye-tracking system. Those results support the use of human-computer interfaces that combined multimodal interactive characters with eye-tracking technology.

3. CONCLUSIONS

In order to set-up the experiments described above, our MARC platform was extended and applied to include advanced features about real-time social interaction and multimodal expression of emotions. We are also exploring the use of such technology with depressed patients and the MARC platform was already applied to social anxiety (Vanhala et al, 2012).

Our next step will be to use realistic expressive virtual characters to conduct additional experimental studies about other social interaction capacities related to joint attention, realism of appearance, and feeling of presence.

Acknowledgements: This work was supported by a grant provided by a partnership between the two following foundations: Fondation de France and Fondation Adrienne et Pierre Sommer (Project #2007 005874).

4. REFERENCES

- S Buisine, Y Wang and O Grynszpan (2010). Empirical investigation of the temporal relations between speech and facial expressions of emotion, *Journal on Multimodal User Interfaces*, **3**, pp. 263-270.
- J Cassell, H Vilhjálmsón and T Bickmore (2001) BEAT: the Behavior Expression Animation Toolkit, Proc. SIGGRAPH'01, New York, pp. 477–486.
- M Courgeon, C Clavel, N Tan and J-C Martin (2011) Front View vs. Side View of Facial and Postural Expressions of Emotions in a Virtual Character, *Transactions on Edutainment (TOE)*, **VI**, pp. 132-143.
- B de Gelder and J van den Stock (2011) Real faces, real emotions: perceiving facial expressions in naturalistic contexts of voices, bodies and scenes, In *The handbook of face perception* (A Calder, G Rhodes, M Johnson, J Haxby, Eds.), Oxford University Press, USA, pp. 535-550.
- J Gratch, J Rickel, E André, N Badler, J Cassell and E Petajan (2002) Creating interactive virtual humans: some assembly required, *IEEE Intell Syst*, **17**, pp. 54–63.
- O Grynszpan, J-C Martin and J Nadel (2007) Exploring the Influence of Task Assignment and Output Modalities on Computerized training for Autism, *Interaction Studies*, **8**, 2, pp. 241-266.
- O Grynszpan, J Nadel, J Constant, F Le Barillier, N Carbonell, J Simonin and J-C Martin (2011). A New Virtual Environment Paradigm for High-Functioning Autism Intended to Help Attentional Disengagement in a Social Context, *Journal of Physical Therapy Education*, **25**, 1, pp. 42–47.
- O Grynszpan, J Nadel, J-C Martin, J Simonin, P Bailleul, Y Wang, D Gepner and J Constant (2012a) Self-monitoring of gaze in high functioning autism, *Journal of Autism and Developmental Disorders*, pp. 1-9.
- O Grynszpan, J Simonin, J-C Martin and J Nadel (2012b) Investigating social gaze as an action-perception online performance, *Front. Hum. Neurosci.*, **6**, 94.
- E Krahmer and M Swerts (2009) Audiovisual prosody—introduction to the special issue, *Lang. Speech*, **52**, pp. 129–133
- U Lahiri, Z Warren and N Sarkar (2011). Design of a gaze-sensitive virtual social interactive system for children with autism, *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, **19**, 4, pp. 443–452.
- K R Scherer (1980) The functions of nonverbal signs in conversation, In: *The social and physiological contexts of language* (H Giles and R St Clair, Eds), LEA, New York, pp. 225–243.
- K R Scherer (2010) The Component Process Model: architecture for a comprehensive computational model of emergent emotion, In: *A Blueprint for Affective Computing* (K R Scherer, T Bänziger, E Roesch, Eds), Oxford University Press, pp. 47-70.
- T Vanhala, V Surakka, M Courgeon and J-C Martin (2012) Voluntary Facial Activations Regulate Physiological Arousal and Subjective Experiences During Virtual Social Stimulation, *ACM Journal Transactions on Applied Perception (TAP)*, **9**, 1.
- M Wilms, L Schilbach, U Pfeiffer, G Bente, G R Fink and K Vogeley (2010) It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience, *Social cognitive and affective neuroscience*, **5**, 1, pp. 98-107.