The Mimic Game:
Real-time Recognition and Imitation of Emotional Facial Expressions

Nicolas Stoiber* Orange Labs/Supelec
Olivier Aubault Orange Labs
Renaud Seguier Supelec SCEE
Gaspard Breton Orange Labs

We present an interactive animated agent system that mimics the facial expressions of a person facing a webcam. Related studies have pursued the same objective from a *retargeting* perspective, by capturing the facial deformations of a source face to apply them on a virtual character one frame at a time. Our system differs from these approaches in that it works at the semantic level, and not the morphological one. The system first captures the current facial configuration of the user and recognizes his expressive state. The agent then autonomously produces a facial expression based on the identified state, using its own deformation and dynamic characteristics. This behavior is therefore referred to as mimicking instead of retargeting or cloning. Rigid head movements are also captured and mimicked by the agent.

1 Facial Motion Capture

The markerless capture of the user's facial configuration is based on an Active Appearance Model (AAM) [Sattar et al. 2007]. AAM are commonly used to describe the appearance changes of deformable objects using geometric shape and texture information. They are particularly well-adapted to facial deformations which cause both shape changes (movement of the facial elements) and pixel intensity changes (wrinkles). Our particular AAM model is trained on a person-specific facial expression, and accurately tracks both rigid head movements and non-rigid expressive deformations.

2 Facial Expression Recognition

The tracker's output (a high-dimensional vector describing the current facial configuration) does not have any semantic meaning. Interpreting the facial deformations implies converting this configuration vector to a more meaningful representation, related to higher-level concepts. Previous emotion recognition methods have traditionally relied on theoretical representations of human emotion [Ekman 1982]. These representations were associated to facial deformation data through supervised learning. Instead, we propose a more objective representation of emotional facial expressions, extracted from the data itself. We form an optimized 2D approximation of the nonlinear space of facial expressions, using a manifold fitting technique on a database of expressive images [Stoiber et al. 2009]. The obtained 2D manifold can be unfolded to form a simple yet consistent visual representation of the space of facial expressions. Moreover, this representation is convenient for semantic interpretations.

Our real-time imitation system eventually projects the high-dimensional features detected by the AAM on the 2D manifold. The result is displayed as a point on its disk-shaped unfolded representation (see figure 1).

3 Facial Expression Synthesis

The disk-shaped intuitive representation of figure 1 can also be used as a command interface for the virtual character [Stoiber et al. 2009]. The system can thus easily select target facial expressions for the agent to display, according to the identified state of the user. Ultimately, a dynamics-aware animation system computes the facial movements necessary to transition from the current expression to the target one.

Our animation system relies on a collection of input-output nonlinear motion models inspired by automatic control science. Any target expression can be given as input to the motion models, and the corresponding facial movements are outputted in real time; inputs can be changed on-the-fly, providing a good reactivity to evolving targets. The models' parameters are learned from real facial motion data, ensuring the naturalness of the produced animations. Our motion models also integrate a stochastic component that accounts for the inherent variability of human motion.

Figure 1: *Snapshot of the real-time imitation system. Facial deformations and head movements are captured on the human face (left), and facial deformation are interpreted as a point in the simplified facial expressions space (top right). The virtual character then reproduces the position and the expressive state of the user (right) using its own motion models.*

4 Conclusion

Our embodied agent successfully mimics the expressions of the user (see the accompanying video). Thanks to the recognition step, the system does not just capture the facial deformations, but is able to interpret them. Therefore, instead of simply imitating the user, the agent could trigger more constructive reactions, such as expressing surprise when the user looks sad or angry. In real-world applications, other information channels such as speech or gaze direction need to be accounted for. Our system would however help define more realistic agent behaviors by efficiently handling the recognition and the synthesis of facial cues.

This work was funded by the Companions project sponsored by the European Commission as part of the Information Society Technologies program (IST-FP6-034434).

References

EKMAN, P. 1982. *Emotion in the Human Face*. Cambridge University Press, New York.

SATTAR, A., AIDAROUS, Y., LE GALLOU, S., AND SEGUIER, R. 2007. Face alignment by 2.5d active appearance model optimized by simplex. *ICVS*.

STOIBER, N., SEGUIER, R., AND BRETON, G. 2009. Automatic design of a control interface for a synthetic face. *Proceedings of the conference on Intelligent User Interfaces*, 207–216.

*e-mail: nicolas.stoiber@orange-ftgroup.com

SIGGRAPH 2010, Los Angeles, California, July 25 – 29, 2010.
ISBN 978-1-4503-0210-4/10/0007