

Facial animation for real-time conversing groups

Rachel McDonnell*

Graphics, Vision and Visualisation Group, Trinity College Dublin.



Figure 1: Our real-time crowd system with pedestrians and conversational groups.

1 Introduction

In recent years, many developments have been made in real-time crowds, with popular video games such as Ubisoft's *Assassins Creed* investing significant resources into their crowds. Research has focused heavily on agent locomotion, but rarely have there been efforts to integrate realistic groups of stationary humans. However, groups of idle people conversing are important for the realistic depiction of a crowded scene. Using motion captured data of real conversations in order to create these groups produces realistic results. However, if motion data or storage is limited, this results in many duplicated conversations, which might appear unrealistic. In [Ennis et al. 2010], we examined the circumstances under which combining and reusing segments of recorded conversations would appear realistic to the observer. The results of our experiments allowed us to integrate varied groups of conversing characters into our crowd. Until now our characters were animated with conversational body-motion alone (Figure 1). The lack of facial animation and expressions can be disturbing when the conversing characters are focussed on by the viewer. In this ongoing work, we aim to increase the plausibility of our conversing groups by adding Level Of Detail (LOD) facial animation, while maintaining interactive frame-rates.

2 System Implementation

Firstly, we integrated a set of next-gen characters into our system with high quality opacity, bump and diffuse texture maps, and a full set of facial bones (Figure 2). We then captured a corpus of conversational body motions and sounds from three male and three female actors. All conversations were natural and unscripted. Body motion data was captured using a 13 camera Vicon optical motion capture system, with 52 markers per actor. Speech was recorded using condenser microphones placed in front of each actor to record their part of the conversation. Facial and eye-movements were not captured during this session.

We used a commercial lip-sync tool to generate the mouth movements from the audio recorded during our capture session. These were then exported into our system and played back in real-time. Finite state machines were used to blend between conversations

at run-time. Periodic eye blinks occur on average every 4.8 seconds [Pelachaud et al. 1996], so we triggered an eye-blink event at run-time with a slight random offset to avoid a robotic result. In order to create natural motion, we separately motion-captured the eye-blinks of an actor during conversation, using small markers placed on their upper and lower eyelids. At run-time, we randomly select from this set of captured blinks.



Figure 2: Next-gen character with face & eye-blink motion applied.

Adding facial animation is important for realism but is computationally expensive when the number of characters increases, due to the additional bone transformation updates. We use motion LOD in order that only characters that are in close proximity to the camera have facial animation, and blend to a neutral pose for the characters that are not seen.

Acknowledgements

This work was sponsored by Science Foundation Ireland as part of the Metropolis project, led by Professor Carol O'Sullivan.

References

- ENNIS, C., McDONNELL, R., AND O'SULLIVAN, C. 2010. Seeing is believing: Body motion dominates in multisensory conversations. *ACM Transactions on Graphics* 29, 4, 91:1–91:9.
- PELACHAUD, C., BADLER, N. I., AND STEEDMAN, M. 1996. Generating facial expressions for speech. *Cognitive Science* 20, 1, 1–46.

*Rachel.McDonnell@cs.tcd.ie