

Automated Synthesis of Expressive Speech for Military Training

W. Lewis Johnson & Shrikanth Narayanan

Interactive virtual humans have an essential role to play in the type of experiential learning environments that the Institute for Creative Technologies is creating. Ideally virtual humans have both a synthetic body, created using computer animation techniques, and a synthetic voice, that allows the virtual human to converse with the learner and with other virtual human characters.

Rapid progress is being made in methods for controlling motion of synthetic bodies for virtual humans, using keyframed animation, motion capture, inverse kinematics, biomimetic techniques, and combinations of the above. Progress with synthetic voices has not been nearly as impressive. Although a number of text to speech (TTS) products exist, they tend to sound monotonous and have problems with intelligibility. Successful applications of TTS products to virtual humans focus on applications such as news reading where a dry, artificial delivery is acceptable. But most tellingly, the entertainment industry has avoided use of TTS, and hires voice actors to record voices for animated characters. This approach may be acceptable for animated films, but it constrains the interactivity of computer-based entertainment and training applications, since the repertoire of utterances that the virtual human can say is limited to the set of recorded utterances. As virtual humans become more flexible and adaptable in their behavior, this limitation will become increasingly significant.

This project aims to make significant progress in effective voice synthesis for virtual humans, building on current TTS technology. Our goal is high-quality *expressive speech*, i.e., speech that is intended to convey or evoke a particular emotion and affect. Expressive speech is essential in military settings; for example, military commanders are trained to assume a “leadership voice” which both projects confidence and inspires action. Proper expression is particularly important for training applications, in order to create virtual human coaches that can respond to and critique trainee actions in a meaningful way. TTS research, which comes mainly out of the telephony industry, has made relatively little progress in the area of expressive speech, which is less important for ordinary telephone conversations.