CoDiS Lab

ISDS

Master Theses

Generating Non-Verbal Communication From Speech

Student: Manuel Rebol
Supervisor: Christian Gütl
Co-Supervisor: Krzysztof Pietroszek (American University, Washington DC)

The communication between human and virtual agents is becoming more important because of the increased usage of virtual environments in everyday life. Applications such as virtual learning platforms, online meetings and virtual assistance are gaining popularity. However, communication with virtual agents is often unintuitive and does not feel similar to talking with a person. Therefore, we focus on improving natural non-verbal human-agent communication in our work. We develop a speaker-specific model that predicts hand and arm gestures given input speech.

We model the non-deterministic relationship between speech and gestures with a Generative Adversarial Network (GAN). Inside the GAN, a motion discriminator forces the generator to predict only gestures that have human-like motion. To provide data for our model, we extract the gestures from in-the-wild videos using 3D human pose estimation algorithms. This allows us to automatically create a large speaker-specific dataset despite the lack of motion capture data.

We train our gesture model on speakers from show business and academia using publicly available video data. Once generated by our GAN, we animate the gestures on a virtual character. We evaluate the generated gestures by conducting a human user study. In the study, we compare our predictions with the original gestures and predictions from uncorrelated speech in two different tasks.

The results show that our generated gestures are indistinguishable from the original gestures when animated on a virtual character. In 53 % of the cases, participants found our generated gestures to be more natural compared to the original gestures. When compared with gestures from an uncorrelated speech, participants selected our gestures to be more correlated 65 % of times. Moreover, we show that our model performs well on speakers from both show business and academia.