Abstract

Speech is one of the most important means of communication for humans. With computer systems becoming more and more part of our daily lives, it therefore seems reasonable to develop computer programs that can communicate with human users via speech. These programs are known as spoken dialogue systems. A central part of human dialogue is the organisation of speech into alternating turns. Previous research has shown that humans are highly skilled in anticipating the end of the previous turn to prepare a timely response and keep the conversation fluid. Replicating this behavior is thus also important for enabling natural, fluent and responsive interaction with spoken dialogue systems.

This dissertation puts the focus on the analysis of nonverbal qualities in dialogue to help bridge this gap. In particular, the influence of speaker personality types on nonverbal qualities is examined. Results of this analysis could help to adapt turn-taking models to the personality type of their users to simulate more natural conversation. Furthermore, in order to examine the predictive power of nonverbal qualities on turn-taking, a logistic regression model is presented that makes continuous turn-taking decisions based on nonverbal qualities. To achieve this, a multimodal dataset of spoken English task-based dialogues is utilised. Next to annotations of the turns and other nonverbal qualities such as gaze and laughter, the dataset also contains personality scores for the participants based on the Big Five model. This dissertation finds evidence that the openness trait of the Big Five model, influences the total time spoken in a dialogue, the average time between turns and the amount of times a person gets interrupted during dialogue. There is also evidence that the extroversion factor influences the amount of gaps in speech a person leaves during dialogue. The presented continuous turn-taking model does not outperform a last-known value baseline.