Towards interactive speech synthesis; an example of robot-human dialogues in a spontaneous environment

Plenary speech by

Nick Campbell
Trinity College, Dublin

Speech synthesis is (by definition) *NOT* spontaneous. However, speech synthesis is increasingly being used in situations where spontaneous speech is common.

In the past, the main challenges for speech synthesis have been voice quality and prosody prediction, but we argue that nowadays the most important goal for synthesis technology research is for the system to know WHEN to speak, and to be able to parse the reaction of any listener(s) present. The second priority perhaps is to know WHAT to speak; in the sense that someimes it is necessary to repeat or paraphrase an utterance to facilitate smoother communication.

In order to explore the possibilities of Interactive Speech Synthesis, we are developoing a sentient dialogue system (Cara) which is able to monitor the cognitive states of its partner through sensing of vocal and physical dynamics throughout the conversation.

Work with the JST/ESP Expressive Speech Corpus has shown us that tone-of-voice is a key factor in displaying cognitive state changes and interpersonal dynamics. Multimodal signal processing as tested in the TableTalk and D64 data collections allows us to monitor similar reactive changes in body-posture and gestural dynamics.

Together with the HMI Research Group at UTwente, we have developed a sensitive Receptionist Robot that is able to manage short task-based conversations and to be aware of and cope with third parties that may intrude on the dialogue.

With DFKI and colleagues in the Metalogue Project we developed a computer dialogue system that sensed MetaCognitive processes in public speakers, and in the Joker Project with LIMSI and other European partners we are developing a joking conversational robot for elderly or socially deprived people.

Current work at the Speech Comunication Lab in Dublin includes extending the Herme robot for interactive short social dialogues and testing autonomous dialogue sensing mechanisms for timing and content control in potential entertainment or customer-care applications.

This invited talk will present the findings of our recent research into Interactive Speech Synthesis for Spontaneous Interactions and will describe our thinking behind future developments and uses of this exciting new technology.