

# **LREC 2008 Tutorial**

## **Non-verbal Information Resources for Constructive Dialogue Management**

### **Content**

Human speech differs from written communication in several ways, most important differences being the use of intonation, speaking rate, and phonation style to indicate speaker states, attitudes, and intentions, both towards the listener and with respect to the discourse. Current interaction technology is still largely based on more formal styles of speech that are closer to the written mode than to interactive conversational speech. However, we also need to take into account the multimodal aspect of communication, as domestic users of the technology expect it to be able to respond to their normal modes of everyday speech interaction.

In order to process this separate channel of speech information along with gestural expressions, the existing technologies will need to be adapted and new modules added. Specifically, new algorithms will be required for the detection and recognition of short back-channel utterances, frequent or idiomatic phrases, laughs and other affect bursts, and other non-verbal speech sounds. These developments will eventually enable a speaker-independent estimation of speaker states and discourse intentions by detection of the long-term variations in prosodic and voice-quality attributes carried by these short utterances. From the small changes across time in the prosodic characteristics of these frequent and unobtrusive nonverbal speech sounds, the various speaker states can be estimated in speaker-independent manner. Similarly, by mimicking these small dynamic changes, the equivalent affective states can be signaled in speech synthesis.

The resulting models of interaction will provide us with a better understanding of the feedback phenomena that are used in human natural conversations for constructing a shared context and to build social rapport. The experimental models can then be used for designing and developing dialogue systems that interact with human users using natural language in an intuitive manner. Intuitive communication strategies can improve the rigid and simple interactions that present-day systems exhibit, and thus the research also encourages interdisciplinary research where human and social sciences look into technological possibilities of application to the design and construction of interactive systems.

The tutorial will focus on methods, problems and challenges in the analysis of non-verbal and verbal communication in spoken interaction. It is grounded in research that combines speech signal analysis with interactional aspects of dialogue management. We will survey issues related to fluent communication starting from the processing of spoken language and going through various experiments on non-verbal communication to technological challenges in modelling and including these research results into interactive applications. The participants will also become familiar with the analysis and annotation work.

The tutorial assumes that the audience is familiar with basic research topics in Speech and Language Technology, but it does not require specific knowledge concerning non-verbal communication. The tutorial will give a self-sufficient overview of the concepts and techniques before continuing with the specific topics and problems of speech signals and their observation.

The tutorial will consist of two 1,5 hour lectures with a 30 minutes coffee break in between. The lectures will be structured as follows:

- 1 Part I: Introduction to non-verbal communication
  - State-of-the-art and motivation
  - Review of the different approaches to robust and efficient interaction management
  - Resources for monitoring non-verbal interactions in a discourse
- 2 Part II: Application of non-verbal aspects in constructive dialogue systems
  - Basic concepts and techniques for spoken interaction management
  - Fluency and robustness in conversational applications
  - Questions and discussion

## Tutorial speakers

Professor Nick Campbell

National Institute of Information and Communications Technology  
Keihanna Science City, Kyoto, Japan

Email: [nick@nict.go.jp](mailto:nick@nict.go.jp)

web: <http://feast.atr.jp/nick/>

Nick Campbell received his PhD in Experimental Psychology from the University of Sussex in the U.K. and is currently engaged as a Chief Researcher in the Department of Acoustics and Speech Research at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan, where he also serves as Research Director for the JST/CREST Expressive Speech Processing and the SCOPE ``Robot's Ears" projects. He was first invited as a Research Fellow at the IBM UK Scientific Centre, where he developed algorithms for speech synthesis, and later at the AT&T Bell Laboratories where he worked on the synthesis of Japanese. He served as Senior Linguist at the Edinburgh University Centre for Speech Technology Research before joining ATR in 1990. His research interests are based on large speech databases, and include non-verbal speech processing, concatenative speech synthesis, and prosodic information modelling. He is Visiting Professor at the Nara Institute of Science & Technology (NAIST) and at Kobe University in Japan.

Adjunct Professor Kristiina Jokinen

Department of Computer Science  
University of Tampere

Email: [kristiina.jokinen@helsinki.fi](mailto:kristiina.jokinen@helsinki.fi)

web: <http://www.ling.helsinki.fi/~kjokinen/>

Kristiina Jokinen is Adjunct Professor of Interaction Technology at the University of Tampere, and Visiting Professor of Intelligent User Interfaces at the University of Tartu, Estonia. Her research concerns human-computer interaction, spoken dialogue systems, and cooperative response planning, and her current interests include adaptation in speech interfaces, multimodal communication, and conversational systems. She has played a leading role in several academic and industrial research projects, and is currently writing her book "Constructive Dialogue Management - Speech Interaction and Rational Agents" (John Wiley & Sons, forthcoming). She is the secretary of SIGDial, the ACL/ISCA Special Interest Group for Discourse and Dialogue.