



**KTH Computer Science
and Communication**

**KTH Computer Science
and Communication**

Investigating Communicative Feedback Phenomena across Languages and Modalities

LOREDANA CERRATO

Doctoral Thesis
Stockholm, Sweden 2007

KTH Computer Science and Communication
Department of Speech, Music and Hearing
10044 Stockholm, Sweden

GSLT Graduate School of Language technology
Faculty of Arts, Göteborg University
40530 Göteborg, Sweden

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan framlägges till offentlig granskning för avläggande av filosofi doktorexamen i tal och musikkommunikation tisdagen den 15 maj 2007 klockan 15.00 i sal F3, Sing-Sing, Kungliga Tekniska Högskolan, Lindstedtsvägen 26, Stockholm

Abstract

This thesis deals with human communicative behaviour related to feedback, analysed across languages (Italian and Swedish), modalities (auditory versus visual) and different communicative situations (human-human versus human-machine dialogues).

The aim of this study is to give more insight into how humans use communicative behaviour related to feedback and at the same time to suggest a method to collect valuable data that can be useful to control facial and head movements related to visual feedback in synthetic conversational agents.

The study of human communicative behaviour necessitates the good quality of the materials under analysis, the support of reliable software packages for the audio-visual analysis and a specific coding scheme for the annotation of the phenomena under observation.

The materials used for the investigations presented in this thesis span from spontaneous conversations video recorded in real communicative situations, and semi-spontaneous dialogues obtained with different eliciting techniques, such as map-task and information-seeking scenarios, to a specific corpus of controlled interactive speech collected by means of a motion capture system. When motion capture is used it is possible to register facial and head movements with a high degree of precision, so to obtain valuable data useful for the implementation of facial displays in talking heads.

A specific coding scheme has been developed, tested and used to annotate feedback. The annotation has been carried out with the support of different available software packages for audio-visual analysis.

The procedure followed in this thesis involves initial analyses of communicative phenomena in spontaneous human-human dialogues and human-machine interaction, in order to learn about regularities in human communicative behaviour that could be transferred in talking heads, then, for the sake of reproduction in talking heads, the investigation includes more detailed analyses of data collected in a lab environment with a novel acquisition set-up that allows capturing the dynamics of facial and head movements.

Finally the possibilities of transferring human communicative behaviour to a talking face are discussed and some evaluation paradigms are illustrated. The idea of reproducing human behaviour in talking heads is based on the assumption that the reproduction of facial displays related to communicative phenomena such as turn-management, feedback production and expression of emotions in embodied conversational agents, might result in the design of advanced systems capable of effective multi-modal interactions with humans.

© Loredana Cerrato, mars 2007

TRITA-CSC-A 2007:3 ISSN-1653-5723 ISRN-KTH/CSC/A--07/03—SE ISBN 978-91-7178-632-6