

## **Towards computer-aided interpreting: state-of-the-art, limitations and future directions**

Dieter Huber

University of Mainz

Faculty of Applied Linguistics and Cultural Studies

Mainz/Germersheim, Germany

Tel: 49-7274-508241; Fax: 49-7274-508449; email: huber@usgi3.fask.uni-mainz.de

Research in computer-aided interpreting (synonymous names sometimes used for the same subject area include machine-aides voice translation, automatic spoken language interpretation, speech or speech-to-speech translation, automatic telephone interpretation and interpreting telephony) aims at the development of new and advanced information processing techniques that will allow speakers of different languages to converse with one another in their own respective mother tongue via a (stationary or mobile) computer system that (semi)automatically translates their speech utterances. The constituent technologies required for such a system include automatic speech recognition (ASR) and speaker adaptation (SpA) at the input side, machine translation (MT) for the actual transfer of meaning between the source and the target language(s), and finally speech synthesis (SS) and voice conversion (VC) to generate the appropriate speech output in the target language(s).

This paper proposes a unified approach to the description and classification of prosodic phenomena in continuous speech, and evaluates its applicability to automatic spoken language interpretation for a limited transfer task between equivalent samples of Japanese and English dialogue. An algorithm is presented which uses the F0-tracings of connected speech dialogue as input and performs speaker independent segmentation into prosodically defined information units. Detailed descriptions of the algorithm and its application to text-to-speech synthesis, automatic speech recognition, spoken language parsing (integrating speech processing and natural language processing techniques), disambiguation, and speaker adaptation have been published earlier (e.g. Huber 1989, 1990, 1991). The present study is aimed to adapt the method to the problem of prosodic transfer in automatic spoken language interpretation between Japanese and English.

The material chosen for this study was selected from the ATR bilingual dialogue database and consists of six recordings of the first of seven simulated Japanese-English telephone dialogues conducted within the applications domain of conference registration. Ten speakers participated in the recording of the material: five native speakers of Standard Japanese (2 female, 3 male) and five native speakers of British (1 male) and American (2 female, 2 male) English. A total of 132 intonation units was established in the six conversations. 75 of these units (56.8 %) pertain to the Japanese recordings, the remaining 57 (43.2 %) to the corresponding English material. Based on the duration, alignment and pausing data derived from this material, a first set of transfer rules for the "translation" of prosodic features from English (source language) to Japanese (target language) is introduced and evaluated.

The research described in this report was initiated during my stay as invited researcher at the ATR Interpreting Telephony Research Laboratories in Kyoto, Japan. I wish to thank Dr. Akira Kurematsu, Dr. Tsuyoshi Morimoto, Shigeki Sagayama and Dr. Yoshinori Sagisaka for their support and valuable comments.