4E－7

ATR-MATRIX: A Speech Translation System Between English and Japanese

リーブスベン,西野敦士,シンガーハラルド,藤澤謙,山田節夫,菅谷史昭,竹澤寿幸,横尾昭男,山本誠一
Ben Reaves, Atsushi Nishino, Harald Singer, Ken Fujisawa, Setsuo Yamada,
Fumiaki Sugaya, Toshiyuki Takezawa, Akio Yokoo, Seiichi Yamamoto
ATR Interpreting Telecommunications Research Laboratories, Kyoto 619-0288

Introduction

The ATR-MATRIX speech translation system[1] translates both ways between English and Japanese, quickly enough to hold a realistic conversation in the hotel reservation task domain. This paper explains its software design, features, limitations, and some future directions.

This paper focuses on features that are new[2] in the past year, including hands-free operation, bi-directional operation, and support for the multinational C-Star II translation project.

Design

ATR-MATRIX integrates these software subsystems: SPREC[3] for recognition, TDMT[4] for translation, and CHATR[5] for synthesis.

Figure 1 shows a block diagram of ATR-MATRIX, in the bi-directional configuration. Communication between controllers is in a packet format (thick lines). The Main Controller is essentially a switch that controls the basic data flow of the system. Each subsystem's controller transforms the data from received packets to the data format and control signals necessary for each subsystem (triple lines).

The host computer for the Japanese side runs the Controllers and a Japanese recognizer, a Japanese-to-English translator, and a Japanese synthesizer. The English result is displayed and sent by the Connect Controller to the English host. From the English host, Japanese translation results arrive at the Japanese host's Connect Controller and are sent to the Japanese synthesizer. The English host runs a complementary set of controllers and subsystems.

For operation in the C-star II consortium, the Connect Controller transforms between packets and the protocol for the C-star II Communication Server, which connects to a network of translation systems in various countries. In another configuration of ATR-MATRIX, German, Korean, and Chinese are output in addition to English. (These configurations are not shown in Figure 1).

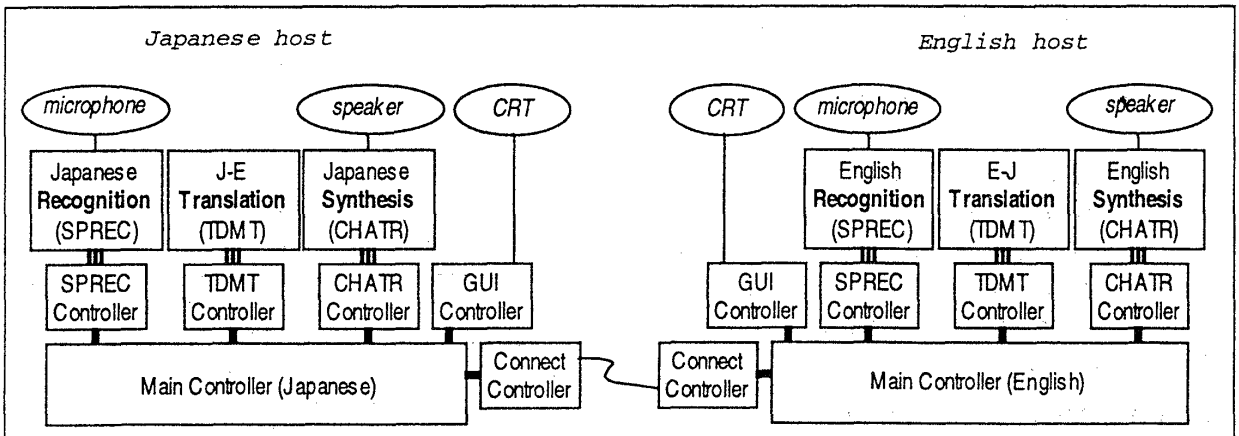

Figure 1: Basic Design for Bi-directional Configuration

ATR-MATRIX:日英双方向音声翻訳システム

ben@itl.atr.co.jp, anishino,singer,fujisawa,syamada,sugaya,takezawa,ayokoo,s-yama

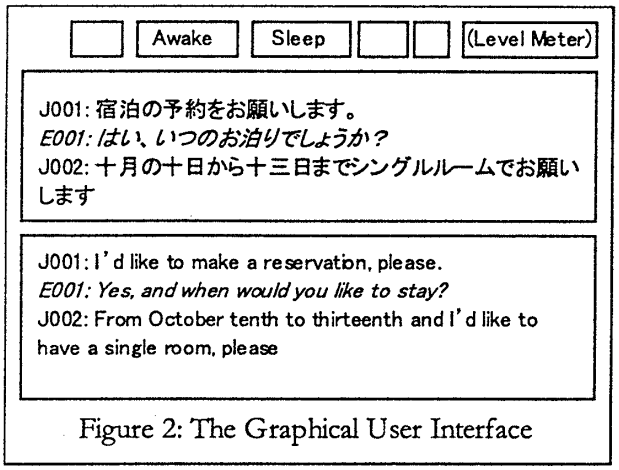

Figure 2: The Graphical User Interface

Figure 2 shows the Graphical User Interface (GUI). At the start of the conversation, the user pushes the Awake button to turn the microphone on. The start and end of speech are determined automatically, and the recognition result is displayed as J001 in the top window. The English translation result is displayed as J001 in the bottom window, and is sent to the English host. When his reply is translated to Japanese, it appears as E001 in the top window. Thus each window shows one full conversation in one language, with each side (party) identified by J or E (also in a different color on the screen, shown here in Italics).

Two elements of prosody are used in ATR-MATRIX: inflection and pause length. Interrogative inflection is detected by the pitch contour of the last vowel in each sentence, for example, "wakarimashita?" must be translated differently from "wakarimashita." Any pause longer than about 0.7 seconds is detected by SPREC's front end Speech Detector. But a pause of only 0.2 seconds occurring where the Language Model expects an end-of-sentence can be detected by a module of SPREC that has knowledge of the Language Model. The result is a much faster response by SPREC, and ATR-MATRIX as a whole.

New Features and Problems

The primary new feature of ATR-MATRIX is the addition of an English recognizer, enabling bi-directional JE/EJ translation. This involved training the recognizer for English, and modifying the Sentence Splitting[6] for English.

The greatest improvement in the user interface is hands-free operation. But if noise is detected as speech, or the user wants to sneeze or talk to someone else, he must use his hand to click the "Sleep" button to abort the recognition.

Recently we started supporting the Slackware Linux Operating System to take advantage of the newest and fastest Personal Computers. But we had to solve some problems with the audio hardware and drivers.

Future Work

The performance of ATR-MATRIX is currently being evaluated by asking naïve users to complete a hotel reservation task. This test brings to light many practical user interface issues. For example, when the user sees that the recognition result is mistaken, he wants to stop ATR-MATRIX from sending it to the other side, or to communicate that it should be ignored. Results of the evaluation will be presented in the near future, and will cast light on new research directions enabling practical use of speech translation systems.

---

[1] T. Takezawa, T. Morimoto, Y. Sagisaka, N. Campbell, H. Iida. F. Sugaya, A. Yokoo, S. Yamamoto, *A Japanese-to-English speech translation system: ATR-MATRIX* Proc. ICSLP 1998, pp. 2779-2782.

[2] B. Reaves, A. Nishino, T. Takezawa, *ATR-MATRIX: Implementation of a Speech Translation System,* Proceedings of the Acoustical Society of Japan, Spring 1998.

[3] 山本博史、シンガーハラルド、リーブスベン、匂坂芳典、"日英音声翻訳システム「ATR-MATRIX」における音声認識部分の構成と制御方法"、日本音響学会平成10年度春季研究発表会講演論文集、I, pp. 161-162 (1998).

[4] H. Mima, O. Furuse, Y. Wakita, H. Iida, *Multilingual Spoken Dialog Translation System Using Transfer Driven Machine Translation,* Proceedings of Machine Translation Summit VI, pp. 148-155 (1997).

[5] N. Campbell, *CHATR: A high-definition speech re-sequencing system,* Proceedings of ASA/ASJ Joint Meeting, pp. 1223-1228 (1996).

[6] 竹澤寿幸、森元逞、"発話単位の分割または接合による言語処理単位への変換手法"、自然言語処理、Vol. 6, No. 2, pp. 83-95 (1999).