

An Automatic Speech Translation System for Travel Conversation

Akitoshi Okumura, Ken-ichi Iso, Shin-ichi Doi, Kiyoshi Yamabana,
Ken Hanazawa, Takao Watanabe

Multimedia Research Laboratories, NEC

4-1-1 Miyazaki, Miyamae-ku, Kawasaki 216-8555, Japan

{okumura, iso, s-doi, yamabana, hanazawa, watanabe}@ccm.cl.nec.co.jp

ABSTRACT

We present a speech-to-speech translation system for notebook PC's that helps oral communication between Japanese and English speakers in the various situations in the travel abroad. Due to the high accuracy of the compact continuous speech recognition engine and our lexicalized grammar approach to machine translation that utilizes corpus but is oriented to model general linguistic phenomena as well as word-specific ones, a versatile speech-to-speech translation system for travel abroad is achieved, which is with much larger vocabulary (50,000 Japanese words and 10,000 English words) and capable of translating spoken conversations in more situations compared to previous work [1].

Keywords

automatic interpretation system, speech recognition, machine translation.

1. INTRODUCTION

We present a speech-to-speech translation system for notebook PC's that helps oral communication between Japanese and English speakers in the various situations in the travel abroad. Due to the high accuracy of the compact continuous speech recognition engine and our lexicalized grammar approach to machine translation that utilizes corpus but is oriented to model general linguistic phenomena as well as word-specific ones, a versatile speech-to-speech translation system for travel abroad is achieved, which is with much larger vocabulary (50,000 Japanese words and 10,000 English words) and capable of translating spoken conversations in more situations compared to previous work [1].

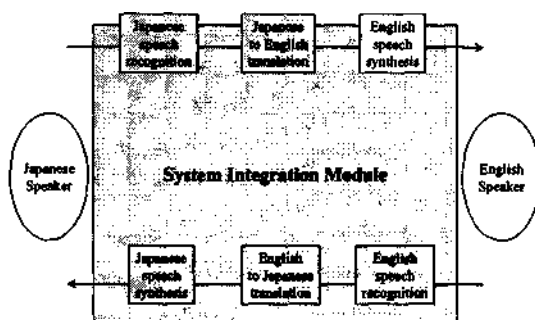


Figure 1. Configuration of the automatic interpretation system.

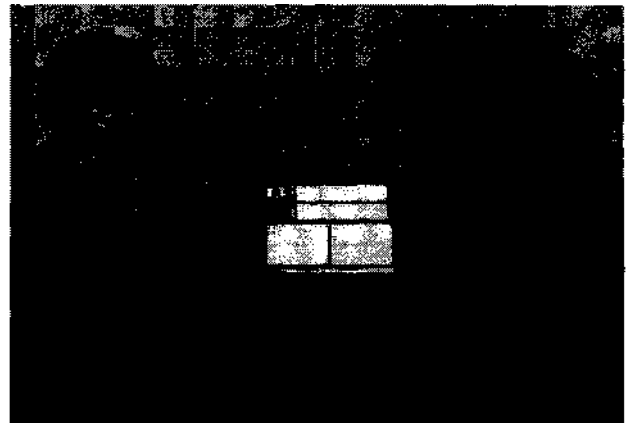


Figure 2. System demonstration.

2. SYSTEM OVERVIEW

The system consists of four modules: speech recognition, translation, speech synthesis and system integration as is shown in Figure 1. Users can have their speech simultaneously translated in real-time by a mobile PC from either Japanese or English to the other, as shown in Figure 2, 3.

To reduce misunderstanding in the conversation between users talking with each other through the system and to avoid the halt in the conversation, the system accepts input of any utterance unit other than a sentence, namely a fragment of a sentence, a phrase, or a word. For each utterance, translated result is obtained in real-time. System requirements for the software include a mobile PC with a Pentium II-class processor (400MHz) running either Windows 98/NT/2000/Me and 128MB of RAM.

3. MODULES

3.1 Speech Recognition

Speech recognition module performs speaker-independent large-vocabulary continuous speech recognition of conversational Japanese and English. The module consists of an acoustic model, a language model, a word dictionary and a search engine. For domain independent acoustic model, triphone HMM was adopted and trained with a large speech corpus. The language model was designed for travel conversation. It contains a bigram and a trigram. The search engine performs two-stage processing. On the first stage, Viterbi beam search is performed to decode input speech to generate a word candidate graph using the acoustic model and the bigram language model. On the second stage, the

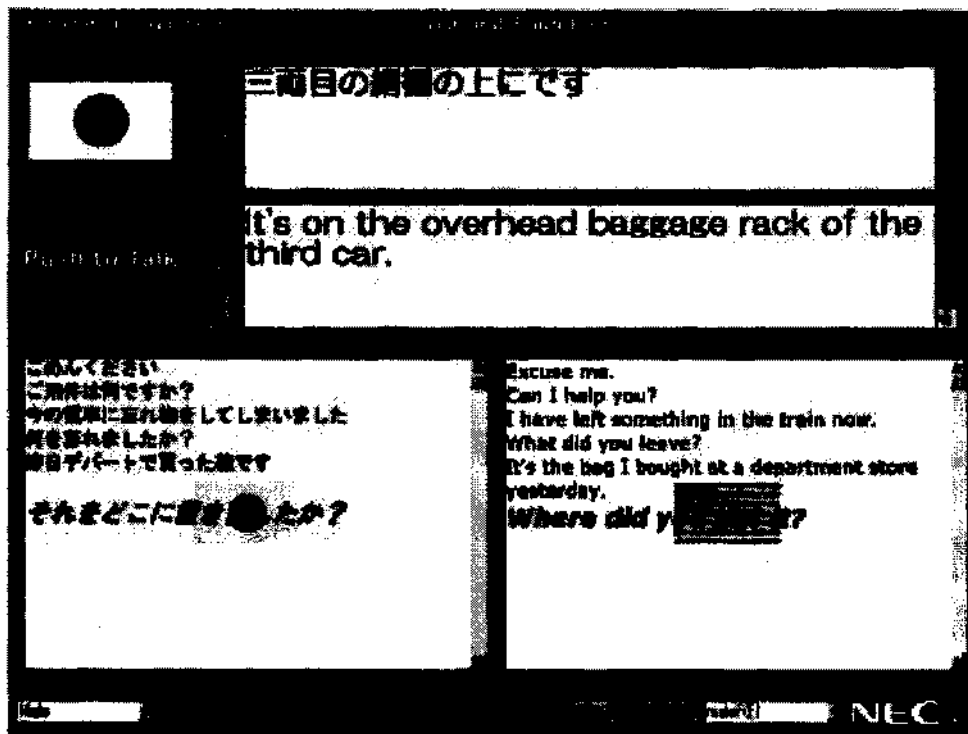


Figure 3. Example of a display

engine performs a search to find the optimal word sequence using the trigram language model. The recognition module also has a speaker adaptation capability. It is possible to adapt the acoustic models efficiently to the speaker just using as few as five utterances.

3.2 Translation

In translation of conversations, a translation module is required to cope with highly word-specific phenomena, including various colloquial and idiomatic expressions. Handling idiosyncratic word behavior is also important to improve the translation quality for the target domain. In addition, translation module is required to cover a wide range of input sentences.

To achieve both broad coverage for general input and high quality for the target domain, we employed a rule-based method that allows writing of both general abstract rules and example-like concrete patterns in a unified framework. Precisely we adopted the Lexicalized Tree Automata-based Grammar (LTAMG)[3]. This method is in line with strong-lexicalization approach to the grammar [2], where each grammar rule (tree) is associated with at least one word, making all the rules lexical. An advantage of the LTAMG to other strongly-lexicalized grammars is use of a simple bottom-up chart-parsing algorithm, which is a straightforward extension of the context-free grammar case.

4. EVALUATION

For the preliminary evaluation for relative short and clean sentences, a word accuracy of 85 - 96% for speech recognition and a sentence understanding rate of 83 - 87% for translation were obtained as a result. We expect that the performance will be further improved by expanding the grammar with regard to domain-specific or colloquial expressions, which were not yet described in the grammar and caused the incorrect translation. We will also evaluate the usability of the system as aids for cross lingual communication.

5. REFERENCES

- [1] Watanabe, T. et al., An experimental automatic interpretation system: INTERTALKER, Proc. of Acoustic Soc. of Japan, (Spring 1992), 101-102 (In Japanese).
- [2] Schabes, Y. et al., Parsing Strategies with 'Lexicalized' Grammars, COLING'88, (1988) 578-583.
- [3] Yamabana, K. et al. Lexicalized Tree Automata-based Grammars for Translating Conversational Texts, COLING 2000, (2000) 926-932.