

携帯端末上で動作する日中通訳システムの開発

花沢 健 赤峯 享 石川 開 磯谷 亮輔

山端 潔 服部 浩明 奥村 明俊

NECメディア情報研究所

e-mail: k-hanazawa@cq.jp.nec.com

1. はじめに

我々は、常に携帯して何時でも何処でも手軽に使える通訳システムの実現を目指しており、これまで日英での通訳システムを開発してきた[1]。今回、携帯端末単体で動作する日中双方の旅行会話通訳システムを開発したので報告する。このシステムは、日本人の発声した音声を中国語に翻訳して音声で出力、あるいはその逆を行うものであり、以下の方針で開発した。

- ・ ホテルの予約やレストランでの注文など、旅行での様々な場面で使用する。そこでの数万語規模の大規模語彙に対応する。
- ・ 一般的な話者が通常利用する話し言葉に対応する。中国語としては、標準語（公用語）とされている普通話(Putonghua)を対象とする。
- ・ 携帯して手軽に使える。携帯情報端末(PDA)単体での動作を実現する。

本稿では、本システムの構成、今回新規開発し、日中通訳実現の鍵となった中国語音声認識、およびPDA上での実装結果について述べる。

2. システム構成

システムの構成としては、図1に示すように音声認識部、機械翻訳部、音声合成部の各モジュールを言語ごとに用意し、それらを統合部にて密接に連携・制御することにより高精度な通訳を実現している。

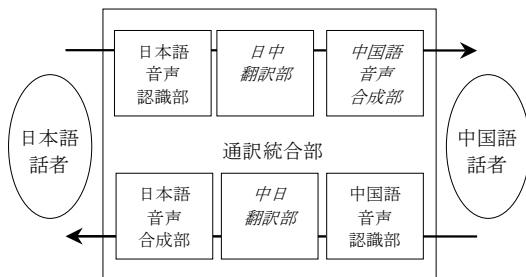


図1. システム構成

表1. 地域分類

ID	Area
N1	華北地区
N2	中西部(山西 陝西)
SE1	上海周辺
SW1	華中地区(武漢周辺)
SW2	重慶周辺
S1	福建
S2	広東
S3	江西 湖南地区

今回開発した日中通訳システムは、従来開発してきた日英通訳システムの英語部分を中国語に差し替えることで対応した。機械翻訳部と中国語音声合成部は既存の製品を統合した。

3. 中国語音声認識部の開発

3.1. 音声コーパス

中国語音声認識部を新規開発するにあたり、課題として幅広いアクセントへの対応と、それを考慮した大規模コーパスの構築が挙げられる。普通話は公用語として定義・教育されているもののその歴史は必ずしも長くはなく、広大な中国全土には様々なアクセントが存在する。今回は、中国全土を大きく8の地域に区分し、それぞれの地域から一定量以上の音声データを収集し、さらにアクセントの無い北京出身話者の音声データも含めて学習コーパスを構築することで、この課題への対処を試みた。今回用いた地域分類を表1に示す。

3.2. モデル構築

一般に音声認識には、音響的確からしさを与える音響モデルと言語的確からしさを与える言語モデルとが必要である。我々の音声認識エンジン自体は言語非依存であるが、これらのモデルは各言語依存で構築しなければならない。

音響モデルは、前節で述べた音声コーパスを用いて状態共有 triphone モデルを学習した。不特

定話者の性別依存モデルとした。言語モデルは、総文数 173k、総単語数 860k の旅行会話テキストコーパスを構築し、単語 Ngram モデルを学習した。認識辞書は、テキストコーパスに出現するものをベースに約 36k 語彙のセットを作成した。発音(pinyin)は普通話をベースに付与した。

3.3. 評価

中国語音声認識部のシミュレーションによる評価を行った。評価データは旅行会話の発声を表 1 と同様の各地域分類別に収集したものをを用いた。男性 101 名分のデータである。音響モデルも男性モデルを用いた。各地域別、および全体の認識率とその分散値を図 2 に示す。認識率は文字正解精度(C.A.)で評価したが、これは多くの単語が 1 文字から成るためである。結果として、各地域別に多少ばらつきはあるものの、いずれも 8 割以上の単語正解精度が得られた。

4. 携帯端末上での実装

新規開発した音声認識部と、機械翻訳部、音声合成部とを統合して日中通訳システムとし、携帯端末に実装した。実機としては市販の PDA(CPU は 400MHz、メモリは 64MB)を用い、通訳システム自体は WindowsCE 上のソフトウェアとして実現した。

中日方向の動作確認を行ったところ、処理速度はリアルタイムにやや届かないものの、数秒程度の発声に対して発声終了から 1 秒程度で合成音声の出力が確認できた。日中方向はほぼリアルタイムで動作することを確認した。通訳画面イメージを図 3 に示す。図 3 では、上段に中国語音声認識結果が、下段にその中日翻訳結果が示されている。なおここでは、音響モデルの自動選択 [2]により性別の指定を不要化している。

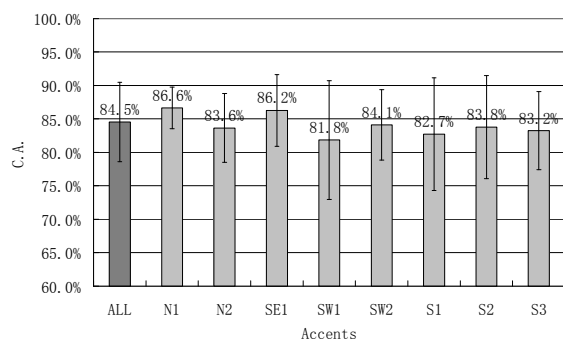


図 2. 評価結果

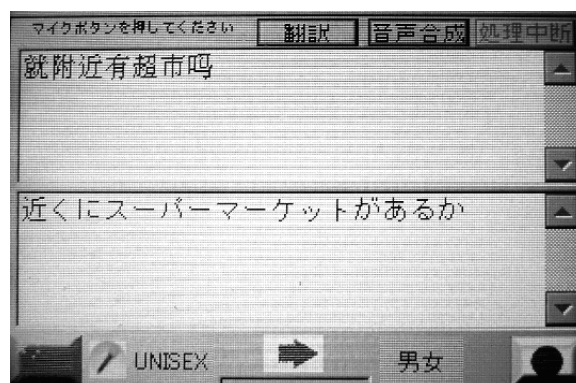


図 3. 通訳画面例

システム全体としての定量的な評価は現在未実施であり、今後の課題である。

5. まとめ

中国語音声認識部を新規開発し、既存の製品・モジュールと統合することにより、携帯端末で動作する日中双方向の旅行会話通訳システムを試作した。中国語音声認識部を開発するにあたり、中国全土をアクセント別に 8 地域に区分し、各地域から一定量以上の音声データを収集することで、地方出身者のアクセントを含む大規模コーパスを構築した。音声認識部のみを評価したところ、地方出身者の発声も含めて 85%程度の認識率が得られた。

PDA 上にシステムを実装し評価したところ、ほぼリアルタイムで動作することを確認した。システム全体としての定量的な評価、および翻訳部も含めた更なる精度向上が今後の課題である。

謝辞

本研究を行うにあたっては、NECEL China (北京)の Li Weishan 氏、Cheng Lifang 氏、Liu Min 氏にご協力をいただいた。

参考文献

- [1] 山端他:「PDA で動作する旅行会話向け日英双方向音声翻訳システム」情処研報, NL-150, 2002.
- [2] 花沢他:「先読みモデル選択による性別非依存音声認識」音講論集, 1-P-18, 2004.