

定型文認識を用いた携帯電話向け音声通訳システム

大淵康成, 北原義典, 小泉敦子
日立製作所中央研究所
obuchi@crl.hitachi.co.jp

海外旅行者が自分の携帯電話から気軽に利用できる, 音声通訳システムを開発した。あらかじめ登録してある定形文のみに対する通訳機能に限定することにより, 高い通訳性能を達成し, 実用レベルのサービスシステムを実現した。これまでも任意文を対象とした自動通訳システムの開発は行なわれていたが, 認識や翻訳の精度の問題から, 実用化までは時間がかかると考えられていた。我々のシステムは, 機能限定ながらも実環境での一般ユーザーの使用に耐えうる安定性を得られたことから, 一般に対する公開実験を行ない, フィールドデータ収集とユーザーニーズ調査を行なった。一般公開したシステムは, 日本語から韓国語への片方向通訳システムで, 東京都心に設置されたサービスセンターに回線交換で接続して使用するものである。これまでに約4,000件のアクセスがあり, 現在その解析を進めている。

Voice Interpretation System for Mobile Phones Using Fixed Sentence Speech Recognition

Yasunari OBUCHI, Yoshinori KITAHARA, and Atsuko KOIZUMI
Central Research Laboratory, Hitachi, Ltd.
obuchi@crl.hitachi.co.jp

We have developed the voice interpretation system, which the travellers can use easily through their own mobile phones. We achieved high performance of interpretation by focusing on the recognition of preliminarily registered sentences. Previous research has been focusing on the recognition of free sentences, therefore, it has been said that they need more time to start the commercial service. Since we made a stable system for the public use, even the function is limited, we decided to make an open test for the public, in order to collect the field data and to investigate the needs of the users. The test system interprets Japanese to Korean only. The user calls the service center in the downtown Tokyo through the circuit switched network. The test system have received about 4,000 calls until now, and we are in the process of analyzing these data.

1. はじめに

「言葉の通じない外国人と, 通訳を介さず自由に会話をしたい」というのは, 音声・言語研究の究極の目標の一つである。未来技術に関するアンケート結果[1]を見ても, 「手のひらサイズの音声自動

通訳機」に多くの人が期待を寄せている。このような機械の実現のために, 音声認識や機械翻訳の分野で多くの研究が為され, そのいくつかは実用化へ向けて公開されるようになってきている。

これまで発表された自動通訳システム[2]は, このように高い志を掲げながら

も、そのタスクの困難さ故に、実際に使ってみるとまだまだ実用レベルにはほど遠く感じられる。これは、通訳というタスクが人間の言語活動全般を対象としており、言語知識はもとより一般常識の理解までもを必要とするためだと言えよう。人間の聴覚の場合を考えても、固有名詞の正確な聞き取りが困難なことは、前後の文脈による内容の推定が不可欠であることを表しているし、映画専門の通訳者や、スポーツ専門の通訳者などが存在するということは、通訳という仕事に言語知識以外の知識が重要であることを表している。ところが、人工知能の分野では、「コンピューターに一般常識を持たせることができるか」という問いは古くから非常に困難な課題だと考えられており、自動通訳の研究も現状ではこの障壁の前で停滞している状況にある。

一方で、海外渡航者数の増大や語学学習熱の高まりに伴い、手軽に使える外国語会話補助機具が数多く市場に現れてきている。これらの機器は、完璧な通訳の代わりに、対訳例文の表示や辞書機能など、限定された機能を手軽かつ安価に提供することによって成功している。

我々は、このような機能限定の通訳装置に、我々自身が開発した音声認識技術を導入し、新たなコンセプトを持つ製品の開発を行なった。その結果、汎用RISCプロセッサ(Hitachi Super-H RISC Engine Series)上の音声認識ミドルウェア[3]を応用した携帯型音声通訳機の試作機を完成させるに至った[4]。この通訳機は、入力を単語に限定し、キーワードからの例文検索というインタフェースを採用することで、マイコン向け音声認識の性能を最大限に活用し、実用性にすぐれた携帯機器となっている。

しかしながら、この通訳機の開発において、専用機という方式を選んだが故の問題点も数多く指摘された。その中でも、文章入力ができないことや、製品としての専用通訳機を購入する際のインシヤルコストがかかることなどについては、別のアプローチからの解決手段を見

つけることによって、さらに大きなマーケットが開拓できると期待される。

そこで我々は、端末として携帯電話を用いることによりインシヤルコストの低減を図り、ユーザーがより気軽に使うことのできる通訳システムの開発を行なうことにした。電話を用いる場合には、通信費用がかかることや、電波状態などにより使用環境が制限されることといったデメリットが存在する。しかし、初期費用がかからないというメリットの他に、センターに設置したサーバーの計算資源をふんだんに使用できることや、コンテンツの入れ換えが容易であるといったメリットがある。また、専用装置の大量生産が不要なことから、簡単に実サービスを始めることができるという利点もある。

我々は、機能限定であっても近い将来にサービスとして実現可能なシステムの開発を目標としている。そのため、現段階で十分な認識精度が期待できる定形文認識をベースとしたシステムを開発し、実際に一般ユーザーに公開することにした。このような通訳サービスの一般公開はこれまでに例がないが、この公開実験を通じて、音声認識データの収集と、ユーザーニーズの探査を行なう予定である。

本報告の構成は以下の通りである。まず次章で以前に開発した専用機タイプの携帯型音声通訳機について簡単に述べる。次に今回開発した携帯電話向け音声通訳システムの技術的な概略を説明した後、このシステムを使ったサービスの公開実験について紹介し、最後にまとめを述べる。

2. 携帯型音声通訳機

日常生活で音声認識が必要とされる状況としては、手が使えない、キーボード操作が不得手、高速な入力が必要など、様々なものが考えられる。たとえばカーナビゲーション装置では、安全のために運転以外のことに手を使えないことが音

声認識の必要性を生んでいる。最近のパソコン向けディクテーションソフトなどでは、キーボード操作が不得手でも入力ができるというのが一つの売り文句になっている。

携帯機器の小さな筐体におけるキーボードの代替手段というのも、音声認識の優位性が大きく現れる状況の一つと言えよう。特に、我々が考えている通訳装置の場合には、海外旅行時の街歩きなどにも手軽に持ち運びできることが必要となることから、音声認識の需要は非常に大きいと言える。

図1は、我々が以前に開発した携帯型音声通訳機の外観写真である。6cm×15cm×3cmの筐体に、CPUとしてはSH-3

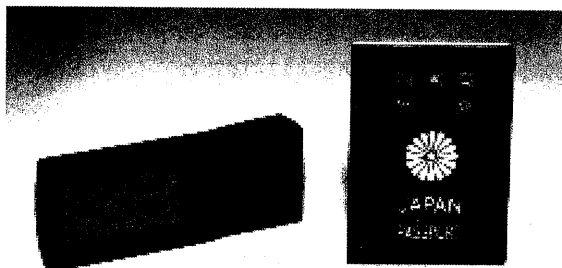


図1 携帯型音声通訳機の外観写真

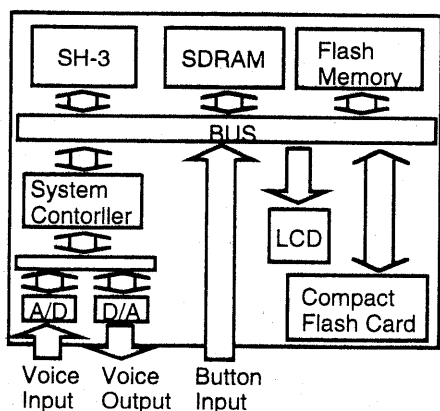


図2 携帯型音声通訳機のシステム構成

(動作周波数60MHz)を搭載し、8メガバイトのSDRAMと16メガバイトのフラッシュメモリ、液晶画面等を備えている。図2は本装置のシステム構成である。この携帯型音声通訳機は、液晶画面での確認操作とスクロールによる参照ができるというメリットを活かすために、単語音声認識によるキーワード例文検索という方式を取っている。単語辞書としては約30,000語を収録しており、その中の約2,000語がキーワードとして例文にリンクされている。収録されている例文の数は約3,000である。また、単語認識の際には、プロセッサの処理能力と認識精度の問題から、約2,000の重要単語を優先的に照合対象とする認識方式が取られている。

認識された単語から例文が検索されると、その対訳を液晶画面で見ることができ、それに対する音声スピーカーから出力させることもできる。出力音声は、あらかじめナレーターによる発声を録音し、ADPCM圧縮したデータをフラッシュメモリに保存してある。また、コンパクトフラッシュカード対応のスロットが用意しており、各国語対応のカードを差し換えることによって様々な言語に対応した通訳装置として用いることができる。

3. 携帯電話向け音声通訳システムの概要

前節で述べた携帯型音声通訳機は、旅行先に手軽に持っていくことができるという長所を持つ反面、購入時に初期投資が必要であり、ユーザーに対する敷居を高くしてしまう。そこで我々は、多くの人が持っている携帯電話を端末装置として、同等のサービスを提供することを考えた。この場合、通信費というランニングコストがかかるという欠点はあるものの、初期投資は不要である。また、小型化したとはいえ15cmの大きさのものを旅行先に持ち歩くことに比べて、もともと持っている携帯電話を使うことによる身軽さというメリットも見のがせない。更

に、センター方式を取ることににより、処理速度に関する制限が大幅に緩和されることに加えて、コンテンツのきめ細かい更新をすることが可能になり、様々な複合サービスへの発展も期待できる。

図3は、携帯電話向け音声通訳システムの構成を表している。サーバーには、動作周波数700MHzの Pentium III をCPUに持つAT互換機を用い、Brooktrout社の電話回線ボード RDSP/9432-I を通して電話音声の入出力を行なっている。

実際にサービスを使用する時には、ユーザーは通常回線交換によってサーバーに接続する。回線交換を用いているため、使用する端末は特定の機種に限定されず、すべての携帯電話に加えて、PHSや通常の有線電話などからもこのサービスを利用することができる。サーバーに接続すると、ユーザーは電話を通じて日本語音声を入力する。入力された音声は、電話音声の学習データベースを用いて作成した電話音声向け不特定話者HMMを使って認識される。この際、様々な種類の電話に対応できるように、CMN[5]による回線適応技術が用いられている。

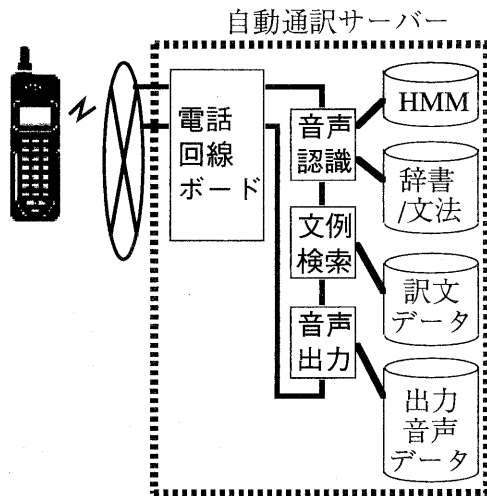


図3 音声通訳サービスのシステム構成

このシステムは回線交換を通じてのサービスであることから、携帯電話の液晶画面による確認・選択等のインターフェースを用いることはできない。そこで、前節で述べた携帯型音声通訳機とは異なり、音声による文例の直接入力という方式を取った。表1に、音声認識に用いる辞書の構成を示す。旅行者向けに作成された1672の文例のうち、150文を基本文例として抽出し、残りの文については挨拶・食事など20種類の場面に分類してある。認識対象が定形文なので、基本的には単語認識の枠組みで対応できるが、文頭の「えーと」「あのー」といった不要語や、文末の表現方法の揺れなどに対応できるように、ネットワーク文法による連続単語認識の枠組みを用いている。さらに、使用方法の説明や場面の選択なども音声を使って行えるよう、コマンド辞書に対する音声認識も行なっている。

音声認識が行なわれると、その結果に対応する訳文データが検索される。定形文に対する通訳であるので、機械翻訳を用いる必要はなく、あらかじめ対訳テーブルの形で用意しておけばよい。更に、出力音声データについても事前に作成しておくことにより、すみやかに音声出力を行なうことができる。

表1 音声認識辞書の構成

総文例数	1672
文例集の構成	基本文例 150 + 挨拶、食事など20場面
文法	ネットワーク文法による 文頭の不要語および文末 変化対応
文頭の不要語	「えーと」「あのー」等 8単語
文末変化	「～して頂けますか」と 「～して下さい」など
コマンド辞書	ガイダンス、場面選択、 終了、動作確認

このようにして検索された外国語の音声データは、電話回線ボードを通じてユーザーの携帯電話に直接出力される。ユーザーは、電話機のスピーカーホン機能などを使って、この音声を対話相手などに聞かせることによって意思の伝達をすることができる。もちろん、音声出力を別の電話に返すことも可能であるが、現在のシステムの仕様と性能を考えた場合に、ユーザーの状況が目に見えない遠隔地通話で用いるよりも、直接目の前にいる相手との対話に用いる方が有効性が高いと考えられることから、このようなインターフェースを取ることにした。

今回、我々は最初の試作として日本語から韓国語への通訳を実現した。しかし、以上に述べたような出力データの構造を考えた場合に、他の言語への拡張は訳文データおよび出力音声データの作成だけで済むことから、英語・中国語をはじめとする多国語版の開発を早急に行なう予定である。

4. 音声通訳サービス公開実験

前節で述べた日韓自動通訳システムの性能評価のため、これを用いたサービスを一般に公開して実験を行なった。公開期間は2000年11月14日から2001年1月13日までの2ヶ月間で、電話番号を一般に告知し、用意した4回線を無料公開（ただし通話料はユーザー負担）して自由に使ってもらった。サーバーは東京23区内に設置したので、都心部からの使用であれば市内通話料金で利用可能であるが、例えば旅行者が韓国から使用してみる場合などには、国際通話料金を支払う必要がある。

図4に、音声通訳サービスの詳しい使用方法を示す。基本的には日本語の音声入力に対して日本語の確認音声および韓国語の対訳音声が出力されるだけの仕組みであるが、韓国語や日本語の音声をもう一度聞いてみたい場合には、ボタンにより操作する。また、周囲雑音による誤動作防止のため、2回目以降の入力に進

むためにはボタン押下が必要な方式を取っている。ただし、この方式を知らないため、2回目の入力のためには電話を掛け直す必要があると誤解しているユーザーが多いことから、この部分については改良の余地があると思われる。また、ボタン操作等についてのガイダンス音声を聞くためのボタンも用意してある。更に、「場面の変更」と発声することにより認識対象辞書を切り替えることもできる。その他、「通訳サービス終了」という発声によりサービスを終了したり（実際にはそのまま電話を切るだけでも可）、システムの応答が悪い場合には「もしもし?」という音に反応して動作を初期化する機構も備わっている。

今回の公開実験を通じて、ユーザーの使用履歴および入力音声データはログとして保存されており、現在その解析作業を進めている。サービス公開後1ヶ月間で、約4,000件のアクセスがあったが、今回は初めての試みということもあり、実際の旅行場面での使用というよりは、サービスの概要に興味をもったユーザーが

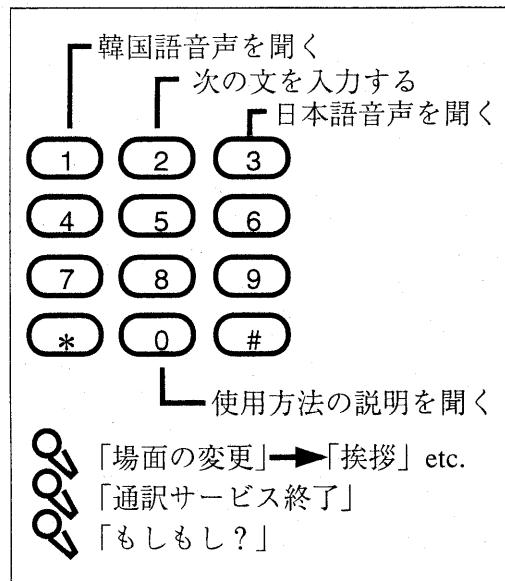


図4 音声通訳サービスの使用方法

試してみたという色合いが強い。そのため、細かい場面の使い分けをするケースは少なく、大半のユーザーは基本文例のみの使用となっている。

サービスの使用方法等については、ウェブサイト[6]を設けて告知を試みたが、インターネットへのアクセスができないユーザーによる使用も多く、そのため使用方法の間違いによる誤認識も数多く発生している。音声認識エンジンの性能による誤認識率と、インターフェースに起因する操作ミスとの割合を明確にするため、現在音声データの書き起こし作業による解析を進めているところである。

5. まとめ

一般の携帯電話を端末として用いることのできる音声自動通訳システムを開発した。あらかじめ登録してある定形文のみに対する通訳機能に限定することにより、高い通訳性能を達成し、実用レベルのサービスシステムを実現した。基本となる枠組みは定形文認識であるが、ネットワーク文法を使った不要語処理や文末処理などにより、ユーザーの自然な音声入力に対応できるようになっている。

このシステムを用いて、一般に対する公開実験を行ない、フィールドデータ収集とユーザーニーズ調査を行なった。一般公開したシステムは、日本語から韓国語への片方向通訳システムで、東京都心に設置されたサービスセンターに回線交換で接続して使用するものである。これまでに約4,000件のアクセスがあり、現在その解析を進めている。

今後は、公開実験の結果を活かして文例集の品質を上げ、多様な表現の揺れにも対応できるシステムの開発を目指す。また、多国語対応や文例追加などの実用面での改善を行なうと同時に、雑音対応や話者適応などの認識精度改善方式も導入していきたい。

参考文献

- [1] 日本経済新聞 2000年6月19日付「ゆめテク・21世紀夢の技術展」関連記事
- [2] T. Takizawa, T. Morimoto, Y. Sagisaka, N. Campbell, H. Iida, F. Sugaya, A. Yokoo, and S. Yamamoto: "A Japanese-to-English Speech Translation System : ATR-MATRIX," Proc. ICSLP '98, pp.2779-2782, Sydney (1998)
- [3] Hataoka, N., Kokubo, H., Obuchi, Y., and Amano, A, "Development of Robust Speech Recognition Middleware on Microprocessor," Proc. of ICASSP '98, pp.837-840 (1998)
- [4] 大淵康成, 北原義典, 小泉敦子, 松田純一, 畑岡信夫: "マイコン向け音声認識技術を用いた携帯型音声通訳機," 電子情報通信学会論文誌, Vol.J83-D-II, No.11, pp.2309-2317 (2000)
- [5] 鯨井俊宏, 小高俊之, 天野明雄: "パワーによるクラスタリングに基づくケプストラム平均正規化手法," 音響学会講演論文集, pp.95-96, (1998.3)
- [6] <http://koigakubo.hitachi.co.jp/mobilingual/>