

## パソコン向け音声認識ソフトウェア

5V-5

畑崎香一郎 磯健一 高木啓三郎 篠田浩一 Farzad EHSANI  
 野口淳 坂井信輔 山田栄子 服部浩明 渡辺隆夫 水野正典†  
 NEC 情報メディア研究所 NEC 技術情報システム開発(株)†

### 1. はじめに

近年、パソコン(PC)のCPU性能が大きく向上し、またA/Dコンバータが標準的に搭載されるようになるにつれて、PC上で特別なハードウェアを使わず動作する音声認識ソフトウェアの実現が可能になってきている。PCに音声認識機能を搭載することによって種々のアプリケーションソフトへの音声入力インタフェースの付加が容易になる。また、ソフトウェアでの実現によって、音声認識機能を低コストで提供できる、CPUの速度向上に従って語彙サイズ・処理速度など認識性能の向上が可能であるという利点が生まれる。

しかしながら、高度な音声認識処理の実現には現在のCPUの処理能力はなお十分ではない。このため、これまで実現されているPC上の音声認識ソフトウェアの多くは、単語単位の標準パターンを用いるものであり、話者の発声による単語登録の負担が大きく、また不特定話者に対して安定した認識精度を得ることが困難であった。

筆者らは、これらの問題を解決するために、半音節認識単位を用いた不特定話者単語音声認識ソフトウェアをPC上に実現したので報告する。

### 2. 特徴

今回作成した音声認識ソフトウェアの特徴を以下に示す。

- 不特定話者認識が可能であり、ユーザの声の事前登録が不要である。
- アプリケーションに依存しない50単語程度の発声を用いてユーザの声の特徴を学習し、認識性能をさらに向上できる。

A speech recognition software for personal computers, by Kaichiro HATAZAKI, Ken-ichi ISO, Keizaburo TAKAGI, Koichi SHINODA, Farzad EHSANI, Jun NOGUCHI, Shinsuke SAKAI, Eiko YAMADA, Hiroaki HATTORI, Takao Watanabe, and Masanori MIZUNO†(Information Technology Research Laboratories, NEC Corp. and NEC Scientific Information System Development, Ltd.†)

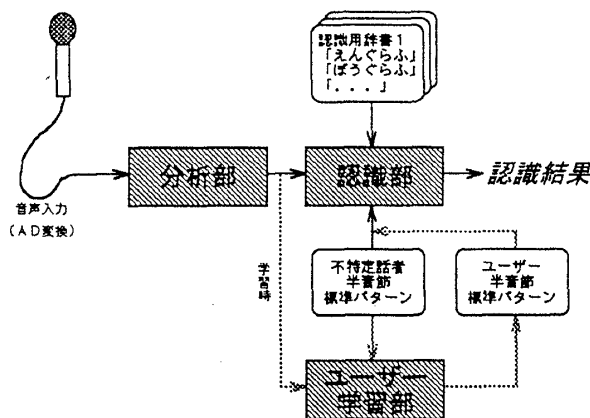


図1: システム構成

- 認識対象単語の読みをかな表記で定義できる。発声による単語登録が不要であり、認識辞書の作成が容易である。
- 同時に識別できる単語数は200単語程度であるが、アプリケーション毎あるいは場面に応じて認識対象を切替えることができ、全体としてより多くの単語を認識対象とすることができる。
- 認識の結果、あらかじめ定義されたキーストロークをアプリケーションに送るキーエミュレーション機能を持つ。これによって既存のアプリケーションを音声で操作することができる。
- 音声始端検出後に音声認識処理を行うことによってCPU負荷を減らしている。さらに、木構造確率分布を用いた認識処理量削減の結果、インテル i486<sup>TM</sup> 程度のCPU上で高速に動作する。

### 3. システム構成

システム構成を図1に示す。

#### 3.1. 分析

PCに入力後、AD変換された音声信号に対し、そのパワー情報を用いて音声検出を行う。音声の始端を検出すると、それ以降の音声信号に対して16ミリ秒フレーム周期のメルケプストラム分析を行う。このとき、

スペクトルサブトラクション処理によって背景雑音の影響を低減している。

### 3.2. 認識

混合ガウス分布型半音節 HMM を用いた不特定話者音声認識 [1] を行う。半音節認識単位は基本的には音節をその母音中心で分割したものであり、これによって音素環境による音声パターンの変動を効率良く表現でき、高精度の認識が可能になった。また、混合ガウス分布型の出力確率分布を用いて話者の違いによるパターンの変動に対処し、不特定話者認識が可能になった。

認識処理は主に、HMM の各状態における特徴ベクトル出力確率の計算と、入力音声と各単語の半音節モデル列との時間軸整合のための漸化式計算であり、これらの計算がフレーム毎に繰り返される。なかでも出力確率計算が認識処理量の大半を占めることから、木構造確率分布 [2] を用いることでその低減を行ない、高速化を図った。この結果、出力確率の計算量を従来に比べて 10 分の 1 以下に低減できた。

漸化式計算では、尤度補正に基づくリジェクション方式 [3] を組み込み、認識対象外単語のリジェクト性能を安定させた。

### 3.3. ユーザ学習

ユーザの音声を用いてスペクトル内挿写像に基づく話者適応化処理 [4] を行なう。すなわち、不特定話者半音節標準パターンのガウス分布のパラメータを適応化し、ユーザ用の標準パターンを作成する。効率的に学習できるように音韻バランス 50 単語からなる学習単語セットを用意した。途中までの単語で学習した後、後日、追加的に学習することも可能である。

### 3.4. 認識用辞書

認識辞書を作成するために、ユーザはかな表記による単語の読みと、認識結果としてアプリケーションに送るべきキーストロークを定義する。認識処理に先だつて、システムは単語の読みを半音節モデル列に変換しておく。

辞書はアプリケーション毎に切替えることができる。さらに単語を複数のグループに分けておくことにより、認識した結果に応じて次に認識対象とするグループを選択することができる。同時に識別できる単語数は 200 単語程度であるが、これによって全体としてより多くの単語を認識対象とすることができる。

表 1: 認識性能 (単語認識率 [%])

	不特定話者標準パターン		ユーザ標準パターン (学習後)
	男女混合	性別	
男性 A	92.8	95.2	97.6
男性 B	92.8	89.2	94.8
男性 C	88.4	93.2	98.4
女性 A	92.8	97.6	99.6
女性 B	90.4	93.2	98.0
女性 C	92.4	95.2	99.2
平均	91.6	93.9	97.9

### 3.5. 標準パターン

不特定話者認識用には、男性 23 名、女性 20 名による音韻バランス 250 単語で学習した男女混合不特定話者標準パターンに加え、男性、女性別々の性別不特定話者標準パターンを用意した。ユーザ標準パターンも含め、各標準パターンの大きさは約 120K バイトである。

### 4. 性能評価

学習に使用した話者とは異なる男女各 3 名による認識性能評価を行なった。学習とは別の音韻バランス 250 単語を認識対象単語とし、各話者がその 250 単語を各 1 回、比較的静かなオフィス環境で発声した。結果を表 1 に示す。不特定話者標準パターンを使用した場合には話者によるばらつきが見られるが、ユーザ学習後は平均認識率 97.9% と安定した性能が得られた。

### 5. まとめ

半音節認識単位を用いた不特定話者単語音声認識方式によるパソコン上の音声認識ソフトウェアを構築した。木構造確率分布を用いた計算量削減の結果、特別なハードウェアを使わずに高速な認識処理が可能になった。認識対象単語はかな表記によって容易に追加、変更でき、パソコンの既存のアプリケーションを音声入力によって操作することが可能になる。

謝辞: システム開発にご支援戴いた古賀真二氏はじめ音声言語研の諸氏に感謝します。

### 参考文献

- [1] 渡辺他. 半音節を単位とする HMM を用いた不特定話者音声認識. 電子情報通信学会論文誌, Vol. J75-D-II, No. 8, pp. 1281-1289, 8 1992.
- [2] 渡辺他. 木構造確率分布を用いた音声認識. 日本音響学会講演論文集, 10 1993.
- [3] 渡辺他. 音節認識を用いたゆ度補正による未知発話のリジェクション. 電子情報通信学会論文誌, Vol. J75-D-II, No. 12, pp. 2002-2009, 12 1992.
- [4] 篠田他. 半音節 HMM による音声認識のための話者適応. 日本音響学会講演論文集, pp. 23-24, 9 1990.