

## パソコン用連続音声認識クライアント・サーバシステムの実装

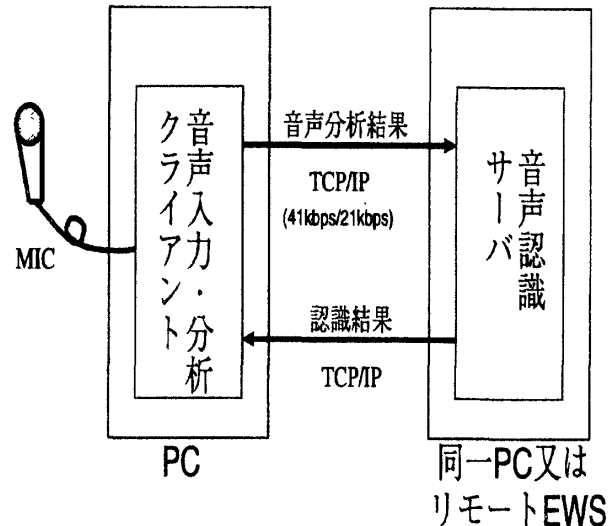
3 J-5

伊藤敏彦 甲斐充彦 山本一公 中川聖一

豊橋技術科学大学 情報工学系

## 1 はじめに

近年、パーソナルコンピュータ (PC) の性能が向上し、音声・動画といった計算パワーが必要なマルチメディア関係のアプリケーションも多く見られるようになってきた。そのため、アプリケーションの入力インターフェイスとしてもキーボード・マウスだけでなく、これまでは計算量の問題から実現が難しかったソフトウェアによる音声認識も使用可能となってきた。パソコン上でソフトウェアのみで動作する音声認識システムはいくつか提案されている [2, 1]。我々はワークステーション上で開発された音声認識システムをベースに、PC 上で動作する音声認識システムを開発した。この音声認識システムは音声入力・分析クライアントと音声認識サーバから構成されておりネットワークを介した文、句などの複数単語の系列 (連続音声) の音声認識が可能である。



## 2 音声認識システムの特徴

開発した音声認識システムの特徴を以下に述べる。

- 話者を限定しない不特定話者認識システム。
- 辞書・文法を変更することによって認識対象を簡単に変更可能。
- 辞書・文法の記述によって単語だけでなく文・句などの複数単語や冗長語の挿入からなる発声が認識可能。
- 汎用のサウンドカードを装備した Window95/NT 上で動作可能。
- 音声入力・分析部と音声認識部が独立したクライアント・サーバ方式で、それぞれ単独でもネットワークを利用した構成で利用可能 (音声認識サーバは WS 上で動作させることも可能)。

## 3 音声認識システムの概要

本システムは汎用サウンドボードを装備したパソコンにおいて Microsoft 社の Windows95 上で動作する。音声認識システムの構成を図 1 に示す。音声入力システムは音声入力・分析クライアントと音声認識サーバから構成され、完全に独立・並列に動作している。相互の情報のやりとりには TCP/IP を用いている。

Client-server-based continuous speech recognition system on PC

Toshihiko Ito, Atsuhiko Kai, Kazumasa Yamamoto and Seiichi Nakagawa

Toyohashi University of Technology, Department of Information and Computer Sciences

図 1: 連続音声認識システムの構成図

## 3.1 音声入力・分析クライアント

音声入力・分析クライアントは汎用サウンドボードのマイク入力からの音声の取り込みとその分析を行なう。マイクを通して PC へ入力された AD 変換後の音声信号に対して、逐次特徴パラメータ系列への変換を行なう。音声の始端と終端の検出は分析されたパワー情報を用いている。音声の始端が検出されると始端検出フラグとそれ以降の特徴パラメータを TCP/IP を用いて音声認識サーバへ送る。分析条件を表 1 に示す。またスペクトルの過渡的な特徴をより良く捉えるための回帰係数パラメータは、各次元毎に前後 4 フレームを含めた 9 フレーム幅セグメントのメルケプストラム係数を用いて算出するが、転送するデータ量の問題からこの回帰係数パラメータは音声認識サーバ側で計算される。

表 1: 音声の分析条件

|           |                           |
|-----------|---------------------------|
| サンプリング周波数 | 11.025 kHz                |
| フレーム長     | 23.22 msec. (256 points)  |
| フレーム周期    | 8.7 msec. (96 points)     |
| プリエンファシス  | $1 - 0.98Z^{-1}$          |
| 分析窓       | Hamming Window            |
| 分析        | 14 次の LPC 分析              |
| 特徴パラメータ   | 10 次メルケプストラム係数 + 10 次回帰係数 |

## 3.2 音声認識サーバ

音声認識サーバでは、TCP/IP を介してフレーム毎に送られてきた特徴パラメータを使用し、回帰係数パ

ラメータの計算や文脈自由文法による言後制約下での仮説の照合・探索を行なう。HMM による音声の照合には、ビームサーチ法に基づく効率的な連続音声認識アルゴリズムを実現するために、Viterbi アルゴリズムを採用している。また、後述するように、未知語処理の機能により、冗長語への対処が可能である。

構文解析処理は、仮説の照合・探索を効率的に行うために、トップダウン型構文解析法の一つである Earley 法を改良している。途中までの部分的な仮説として予測された結果と解析の状態は、終端記号をアークに対応させた動的に拡大する有限状態オートマトン（構文ネットワーク）の形式で保持する [3]。

文脈自由文法による言語解析では、構文的なあいまいさが多数生じることがある。特に、倒置を可能とするような冗長な文法の記述を行う場合、あいまいな解析の状態が多数現れ得る。その結果、あいまいさをもつ同一の部分的な仮説について重複した照合が行われることになり、文法の記述内容によって生じるあいまいさの度合いに応じて処理効率が大きく左右されることになる。本システムでは、このような問題を回避するため、文法の記述をあいまいさを減少するように修正するツールを提供している。また、構文解析部において構文ネットワークへと展開する過程で、同一の終端記号で遷移するときの解析の状態を一つにまとめる方法を採用している [4]。

このようなアルゴリズムを使用して音声入力側から送られてきた音声の終端フレームの処理が終了した時点で最も可能性の高い仮説を認識結果として出力する。

### 3.3 文法・辞書の記述

システムに与える文法・辞書の定義の形式は、文脈自由文法の書換え規則と、ワードクラスとの対応や音節表記を与える単語辞書の 2 種類のものから構成される。以下に示す書換え規則の書式の例は、「都道府県名」認識用システムでの一部の例である。

```
SSSS SENT
SENT ANY_WD prefecture BS.IS
:
ANY_WD any_wd
ANY_WD
BS.IS desu
BS.IS
```

1 カラム目が書換え規則の左辺、2 カラム目以降が右辺に対応する。大文字で表される記号は非終端記号を表し、小文字の記号はワードクラス名（単語集合の名前；前終端記号）を表す。また、左辺が“SENT”の書換え規則は、想定するタスクでの発話文表現を規定するためのもので、文節（例では“BS.”で始まる非終端記号）レベルで組み立てられている。一方、文節内の規則は、一般にワードクラスの単位で組み立てられる。しかし、これらの記述の単位に特に制限はない。

本システムでは、冗長な発話に対処するために、未知語を許す記述が可能となっている。例えば、「えーと東京です。」のような発話を単語認識で許すために、単語の前後に未知語を許す特別な記述（例では、ANY\_WD）をすることにより対処することができる。

なお、各行の書換え規則には、通常文脈自由文法の表現に加えて、“/”とシンボル名からなる解析時のための制御情報を与えることができる。“/”の後には、その行の書換え規則に対応する「構文カテゴリ」名を記述する。この「構文カテゴリ」名を、対話処理部などの外部のプロセス側から随時指定することにより、使用する書換え規則を任意に制御できるようになり、実質的に対話状況に応じた文法の切り替えが可能になる。

単語辞書の記述では、以下の例のように、ワードクラス名、単語名（認識結果で実際に出力される形、例中の“[あの]”などは括弧付で認識結果が出力される。）、単語名の読み、音節表記、の 4 項目からなっている。

|            |      |    |       |
|------------|------|----|-------|
| prefecture | 愛知   | あい | a i   |
| prefecture | 岐阜   | ぎふ | gi fu |
| :          |      |    |       |
| any_wd     | [あの] | #  | a no  |
| any_wd     | [えと] | #  | e to  |
| desu       | です   | #  | de su |
| desu       | だ    | #  | da    |

## 4 都道府県名認識システムの構築例

これまで述べてきた音声認識システムを用いて都道府県名及びその県庁所在地名の発話を認識するデモシステムを構築した。システムの単語数 65 単語で、リアルタイムの 1~2 倍程度で認識できる。また、現在認識結果を JAVA を用いて WWW のブラウザ上で地図表示するシステムを構築中である。このように全て TCP/IP によるコンポーネント間の通信によってシステムが構築されておりアプリケーションをネットワーク型にするだけで音声認識が利用可能となっている。

## 5 まとめ

これまで我々の研究室のワークステーション上で開発されてきた音声認識システムをベースとしてパソコンで動作するクライアント・サーバ型の音声認識システムを開発した。さらにその音声認識システムを利用し、都道府県名を認識できるデモシステムを構築した。

この音声認識システムは全てソフトウェア上で C 言語によって記述されており、プログラムは全て下記 URL にて公開している。

<http://www.slp.tutics.tut.ac.jp/SPOJUS/>

## 参考文献

- [1] 篠田, 坂井, 磯, 畑崎, 渡辺, 水野: パソコンソフト連続音声認識, 情報処理学会第 50 回全国大会, 7R-5, (1995.3)
- [2] 有馬, 福西, 吉谷, 小泉: PC によるソフトウェア音声認識システムの構築, 日本音響学会講演集, 2-Q-27, (1996.9)
- [3] 甲斐, 中川: 冗長語, 言い直し等を含む発話のための未知語処理を用いた音声認識システムの評価, 信学論, Vol.J79-DII(1997)
- [4] 中川, 甲斐: 文脈自由文法制御による one pass 型 HMM 連続音声認識法, 信学論, Vol.J76-DII, No.7, pp.1337-1345(1993)