

論文 / 著書情報
Article / Book Information

論題(和文)	半音節単位認識による大語彙音声入力装置の開発
Title(English)	
著者(和文)	古賀真二, 篠田浩一, 高木啓三郎, 渡辺隆夫, 吉田和永, 塚田聡
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成6年度秋季研究発表会講演論文集, Vol. 2-8-7, No. , pp.
Citation(English)	, Vol. 2-8-7, No. , pp.
発行日 / Pub. date	1994,

半音節単位認識による 大語彙音声入力装置の開発 *

◎ 古賀 真二 吉田 和永 塚田 聡 篠田 浩一
高木 啓三郎 藤原 江美子 渡辺 隆夫 (NEC)

1. はじめに

近年、マルチメディアへの期待の高まりに伴い、入力手段として音声認識機能が注目を浴びている。従来実用化されている音声認識装置では、語彙は数百語程度までであった。しかし、マルチメディアの分野では、大規模データベースへのアクセスなどで多量のデータを入力対象とすることがあるので、さらなる大語彙音声認識が望まれる。筆者らは半音節を単位とした認識方式を提案し、大語彙音声認識、不特定話者認識に有効であることを確認している [1][2]。今回、この方式に基づいて大語彙音声入力装置を開発したので報告する。

2. 装置の特徴

今回開発した大語彙音声入力装置は以下の特徴を有する。

- かな文字で表記された単語を入力するだけで認識対象単語を定義できる。発声による単語登録が不要で、語彙の変更等が容易である。
- 同時に識別可能な単語数は 1000 単語であり、さらに認識対象辞書を切り替えることにより数万語以上の単語入力が可能である。
- 連続発声された数字列など、連続単語認識が可能である。
- 不特定話者認識が可能である。
- スペクトル内挿写像に基づく話者適応化処理 [3] により、短時間の発声を用いてユーザの声の特徴を学習し、認識性能を向上させることができる。
- 2 段スペクトルサブトラクション (2SS) [4] を行うことにより、耐雑音性能を向上させている。

* Development of the Large Vocabulary Speech Input System Using Demi-Syllable Units, by Shinji KOGA, Kazunaga YOSHIDA, Satoshi TSUKADA, Koichi SHINODA, Keizaburo TAKAGI, Emiko FUJIWARA and Takao WATANABE (NEC Corporation)

- 尤度補正に基づくリジェクション方式 [5] を用いることにより、不要語のリジェクト性能を向上させている。

3. 方式構成

認識単位として、音節を母音の中央で半分に分割した半音節を用い、各半音節毎に標準ボタンを用意する (半音節標準ボタン)。半音節は、少ない種類 (約 240 種) で全ての音素間の遷移を表現できるので、効率的に標準ボタンの学習が行える。認識は、単語のかな表記をもとに半音節の連鎖として表現された単語標準ボタンと入力ボタンとを対応付けることで行われる。連続単語認識では、単語の接続規則を定義した有限状態ネットワークを用いる。方式構成を図 1 に示す。

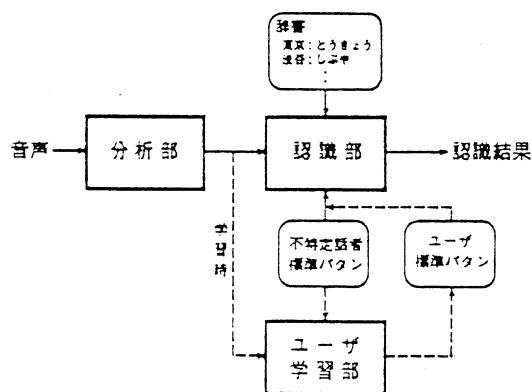


図 1 方式構成

3.1. 分析部

マイクから入力した音声データを 16kHz でサンプリングし、メルケプストラム分析を行う。この時、音声用マイクとは別に用意された周囲雑音用マイクから入力した雑音データを用いて 2SS を行う機能を有する。

3.2. 認識部

認識部では、分析された入力ボタンと単語標準ボタンとの対応付けを、入力ボタンのフレームに同期して行う。フレーム毎に、半音節標準ボタンの各状態に定義された出力確率分布に対する入力ベクトルの局所距離(対数尤度)を求め(局所距離計算処理)、それらの局所距離を用いて単語標準ボタンの各状態における累積距離を漸化式計算にて求める(マッチング処理)。

認識対象外単語のリジェクトのために、尤度補正に基づくリジェクションを行う。

3.3. ユーザー学習部

ユーザ学習部ではスペクトル内挿写像に基づく話者適応化処理を行う。学習用に発声された100～250語の音声データを用いて、不特定話者半音節標準ボタンからユーザ用の半音節標準ボタンを作成する。

4. 装置構成

装置構成を図2に示す。

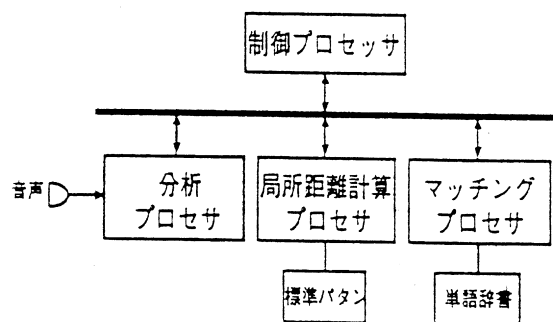


図2 装置構成

分析処理、局所距離計算処理、マッチング処理用の3つのプロセッサが1枚のボードに搭載されている。分析プロセッサ、局所距離計算プロセッサには汎用DSP(16MIPS)を使用し、マッチングプロセッサには処理の高速化のために専用LSIを使用している。局所距離計算プロセッサには半音節標準ボタン用に約512kバイトのメモリを、マッチングプロセッサには単語辞書用に約128kバイトのメモリをそれぞれ用意している。

各プロセッサでの処理は、パイプライン的に並列して行われる。分析プロセッサから局所距離計算プロセッサへは分析データが、局所距離計算プロセッサからマッチングプロセッサへは局所距離データが、それぞれ転送される。

制御プロセッサは認識、学習処理の制御、標準ボタンや単語辞書等のロード、アプリケーションとのインタフェースなどを行う。

5. おわりに

本音声入力装置は、辞書の切替えにより数万語以上の単語入力が可能であり、機器の操作やデータ入力、地名や駅名などの音声入力による情報サービス、大規模データベースの検索などに広く応用できるものである。

参考文献

- [1] 渡辺他, 信学論, J72-D-II, No.3, pp.1264-1269 (1989.8)
- [2] 渡辺他, 信学論, J75-D-II, No.8, pp.1281-1289 (1992.3)
- [3] 篠田他, 音学講論, 1-8-12 (1990.9)
- [4] 高木他, 音学講論, 2-5-3 (1991.3)
- [5] 渡辺他, 信学論, J75-D-II, No.12, pp.2002-2009 (1992.12)

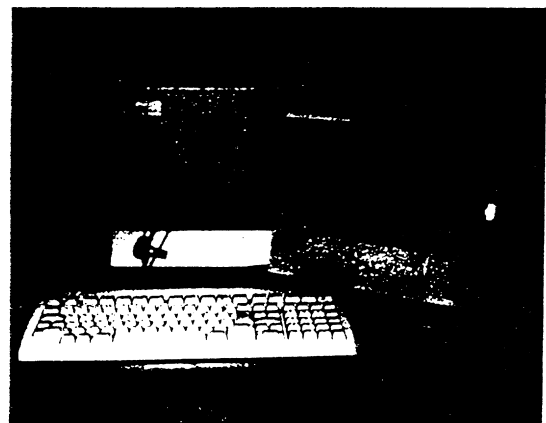


図3 装置概観