

論文 / 著書情報
Article / Book Information

論題(和文)	パソコンソフト連続音声認識
Title(English)	
著者(和文)	篠田浩一, 坂井信輔, 磯健一, 畑崎香一郎, 渡辺隆夫, 水野正典
Authors(English)	Koichi Shinoda
出典(和文)	情報処理学会第50回(平成7年度前期)全国大会講演論文集, Vol. 2-465, No. , pp.
Citation(English)	, Vol. 2-465, No. , pp.
発行日 / Pub. date	1995, 3
権利情報 / Copyright	<p>ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。</p> <p>The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof.</p>

パソコンソフト連続音声認識

7R-5

篠田浩一 坂井信輔 磯健一 畑崎香一郎 渡辺隆夫 水野正典†

(NEC 情報メディア研究所 †NEC 情報システムズ)

1. はじめに

近年、マルチメディア機能をもつパーソナルコンピュータ(PC)が普及し、音声・動画などを扱うアプリケーションが増加している。このような状況下で、キーボード・マウスの他に音声を入力インターフェースとして用いることができれば、使いやすさがより向上し、新しい形態のアプリケーションが可能になると考えられる。先に、筆者らは、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を開発した[1]。このシステムは単語ごとの発声を入力とするものであったが、今回、さらに、文、句などの複数単語の系列(連続音声)の認識を可能にした。より自然な発声に近い入力形態で音声を入力することができる。また、いくつかの機能を追加し、音声入力インターフェースの改良を図った。

2. 特徴

本システムの特徴を以下に述べる。*印のあるものが今回新たに追加されたものである。

1. 誰の声でも認識可能な不特定話者認識システムである。ユーザーの声の事前登録は不要である。
2. 50単語程度の発声を用いてユーザーの声の特徴を学習し、さらに認識性能を向上させることができる。
3. 新規に登録する認識対象単語の読みをかな表記で定義できる。発声が必要で、認識辞書の作成が容易である。
4. 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の機能をもつ。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。
- 5.*文、句などの複数単語からなる発声を認識することができる(連続音声認識)。各々の単語に対しキーストロークを定義でき、一つの発声で複数の操作の組み合わせを実行することができる。

PC Software-only continuous speech recognition,
by Koichi SHINODA, Shinsuke SAKAI, Ken-ichi ISO, Kaichiro HATAZAKI, Takao WATANABE and Masanori MIZUNO†
(NEC Corporation †NEC Informatec Systems)

6.*連続音声の対象となる文を表形式で表示・編集するグラフィカルなインターフェースが用意されている(図1)。単語を行列の形に並べて書くことにより、認識可能な文をユーザーが容易に定義できる。

7. 同時に認識できる単語数は200単語程度である。アプリケーションごとあるいは場面ごとに認識対象を切替えることができ、全体的により多くの単語を認識対象とすることができる。
8. インテル i486™ 程度の CPU 上で高速に動作する。入力と同期して認識処理を行ない、音声入力の終了と同時に認識結果を出力する。
- 9.*音声認識機能の起動・終了を音声入力を用いて行なう音声スイッチの使用が可能。認識を使用していないときの誤動作を防止する。
- 10.*リジェクト機能をさらに強化した。認識対象外の発声が入力された場合の誤認識を防止する。

3. システムの概要

本システムはサウンド取り込み機能をもつパソコンにおいて、Microsoft®Windows™3.1上で動作する。システム構成を図2に示す。

3.1. 分析部

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なう。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なう。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. 認識部

認識単位として半音節を用いた混合ガウス分布HMMを用いた不特定話者音声認識[2]を用いている。不特定話者の標準パターンは、男女43名による音韻のバランスを考慮した250単語の1回発声を用いて作成されている。連続音声の認識においては、認識対象は有限状態オートマトンで定義され、フレーム同期DPを用いて認識処理を行なう。

連続音声の認識処理は、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

