最近の音声認識の動向

——単語および文章入力——

渡 辺 貞 一[†],新 田 恒 雄[†]

最近の音声認識の技術動向を、単語入力と文章入力の観点から整理して述べた。前者の単語入力では、最近実用化が始まっている電話音声認識、連続単語認識および、研究の進んでいる大語彙認識の状況を詳述し、後者の文章入力では、注目を集めている音声ワードプロセッサおよび今後重要となる連続音声認識の動向について概説した。

1. ま え が き

音声を用いた情報伝達は、人間との整合性が良く自然なため、最も優れたマンマシンインターフェイスとして活発な研究開発が進められてきた。

音声入力の持つ利点はいろいろあるが、要約すると、

- (1) 目と手が自由になり、人間の操作性が増す。
- (2) 特別な訓練を必要とせず、誰でも簡単に使用できる。
- (3) 人間との整合性が良く、伝達速度が速い。
- (4) 電話および電話網が利用できるので利用範囲が広がる。

ことである。このため、古くから欧米および日本を中心に盛んに開発が進められてきたが、近年、計算機を

中心とした情報処理技術と集積回路技術の進歩、および各分野からの強いニーズに支えられて急速な発展を見せてきた。

特定話者を対象にした単語認識では、すでに小形で廉価な認識装置が発売され、荷物の自動区分や工場の製品検査に広く使用されており、また最近では、電話を通した不特定話者の声を判別できる単語認識装置や、連続した数単語を認識する装置も商品化されるようになった。さらに音韻や音節を認識し、1000単語を越える大語彙を認識する研究や、文章入力を目的とした音声タイプライターや音声ワードプロセッサの試作も発表されている。またさらには、言語情報や各種の知識情報を用いた本格的な連続音声認識の研究も地道に続けられている。

表1に、音声認識の研究開発の現状と応用分野を一覧にして示す。本論文では、実用化の進んだ単語認識の最近の状況と、文章入力を目的とした音声認識の動向について概観する。

表1 音声認識装置の現状と応用分野

発声 対象 話者	離 散 発 声		連 続 発 声	
	単 語	音 節	単 語	文 章
特 定 話 者	実用段階 ・荷物仕分 ・製品検査 ・対話形端末 (コンピューター入力)	開発段階 ・固有名詞入力 ・文章入力 ・音声タイプライター ・音声ワードプロセッサ	一部実用段階 (数単語連続) ・荷物仕分 ・製品検査 ・対話形端末	研究段階 ・文章入力 ・音声タイプライター ・音声ワードプロセッサ
不特定話者	一部実用段階 ・電話を利用した情報サービス	研究段階 ・電話番号の照会 ・情報検索	研究段階 ・電話を利用した情報サービス	研究段階 ・文章入力 ・音声タイプライター ・音声ワードプロセッサ

[†] 東京芝浦電気株式会社 総合研究所
 "Recent Trends on Speech Recognition" by Sadakazu Watanabe
 and Tsuneo Nitta (Toshiba Research and Development Center
 Laboratory, Kawasaki)

2. 単 語 入 力

音声認識の本格的な研究は1960年代の初めに開始されているが、最初に実用化されたのは、1973年に米国で発売された特定話者を対象に離散的に発声する30単語を認識する装置(Threshold Technology 社, VIP-100)であった。これは話し手があらかじめ設定された単語を1～数回発声して“辞書”として登録しておき、同一の話し手の声だけを認識するものである。この種の装置は、その後10社以上から発売され、現在でも音声入力装置の主力製品になっている。技術的には声を登録することで音声の話者変動を回避し、時間的な変動はダイナミックプログラミング手法(動的計画法, DP法と略)を中心とした非線形マッチングにより解決している。機能的には、極めて制約の強いものであるが、現在900単語程度まで認識でき、物流自動区分業務や工場の検査工程などの、音声入力の特質がうまく生かせる業務には広く利用されている。初期のものは、ミニコンピューターを用いた大がかりな装置であったが、現在ではマイクロコンピューター等を利用して小形化され、廉価な1ボードのものやLSIも使用できるようになった。

その後、この初期のもつ制約を徐々に取除く努力がなされているが、その方向は、

(1) 不特定話者を対象にした認識技術、特に電話を用いた不特定話者の単語認識技術の開発

(2) 連続単語認識技術の開発

(3) 1000語以上の大語彙単語の認識技術の開発である。このうち数十単語を対象にした電話音声認識と連続数単語の音声認識技術が実用段階に入り、さらに語彙拡大と不特定話者による本格的な連続音声認識を目ざした研究開発が進められている。

この章では、新しい動向として、実用的にも重要な電話音声認識、連続単語認識、大語彙認識について詳述する。

2.1 電話音声認識

不特定話者を対象にした単語認識では4～128単語のものが発表されているが、利用形態としては、電話を使用した16単語程度の音声認識・応答システムが代表的である。特に最近、銀行の振込通知や残高照会などの電話サービスシステムに使用され、注目を集めている。

不特定話者の電話音声は、話者の違いによる音声パターンのばらつきの他に、電話機や回線路に起因する音声パターンの変動や劣化がある。このため、これらの差異を充分吸収できる方式の確立が必要となる。

認識方式は大別すると、

(a) 音素系列による認識

(b) 単語単位の認識

に分けられる。

前者は、単語を構成する音素を識別したのち、音素系列で書かれた単語辞書と照合する方式で、音素の選択と識別が適切であれば単語数が増大しても対応できる利点を持つ。代表的なものに、音素系列のDPマッチング法^{1)~3)}(横須賀通研, 東北大)がある。

後者は、単語ごとに音声の標準パターンを登録しておき、入力パターンと照合する方式で、辞書となる標準パターンの作成に工夫を要するが安定した成績が得られるため、実用化されているものが多い。ハイブリッド構造マッチング法⁴⁾(東芝)、識別関数法⁵⁾⁶⁾(日電)、荷重マッチング法⁷⁾(沖電気)などが提案されている。

以下では、このうち電話音声認識・応答システムに実用化されている幾つかを紹介する。

(1) 音素系列のDPマッチング法

横須賀通研で開発された方法では、数字を含む16単語の認識に対して、約40種類強の音素パターン(最大64種まで)を16組用意し、15msのフレーム周期で音声の類似度を計算する。音素パターンの変動は、クラスタリング手法により、16組にまとめている。類似度計算のための距離尺度としては、標準音素パターンの最尤スペクトルパラメーター A_i 、入力パターンの自己相関係数 γ_i 、および予測残差 E を用いて表される距離、すなわち尤度に基づく距離 L 、

$$L = \log \sum_i A_i \gamma_i - \log E \quad (1)$$

が使用されている。

次いで、音素系列で表された単語辞書と入力音声との照合がDP法を用いて行われ、最も類似した値の単語を答としている。音素系列は、1単語当たり平均2通り用意されている。実用化されたシステムでは、最大32単語まで認識できるよう構成されている。

(2) 識別関数法

この方法は、訓練用の学習パターンを用いて、線形計画法により作成される区分的識別関数を使用して判定を行う。音声パターンは、16msのフレーム周期でケプストラム分析を行い、10次の分析パラメーターを抽出する。時間方向には、音声の振幅情報を利用して時間圧縮を行い9個の代表点を求める。得られた90次元の標本パターンを、カテゴリーごとに用意された識別関数に代入して、その値から判定を行っている。認識率は、数字と制御語“取消し”の11語を対象

に、1000 名が 1 回発声したサンプルで、平均エラー率 4.5%，上位 80% の話者に対して平均エラー率 1.5 % であり、男女差、年齢差は少ないと報告されている。

なお、単語の数が数十から百程度に増加するとパターンの分離が悪くなり、区分的識別関数が複雑になるので、パターンベクトルの凸面上への写像といった前処理が工夫されている。

(3) ハイブリッド構造マッチング法

この方法は、パターンマッチング法(複合類似度法)と構造解析的手法(遷移ネットワーク)とを 2 段階に組合せて認識を行うもので、単語単位の認識に音素系列による認識を相補的に組合せた方式となっている。パターンの変形や雑音に強い複合類似度法を用いて、単語全体の類似性から候補単語を選択し、その後で音響・音韻的特徴系列を用いた遷移ネットワークによって単語音声の内部構造を分析し、類似パターンの判別と、認識対象単語以外の入力や雑音をリジェクトする構成である。図 1 に認識アルゴリズムの流れを示す。入力音声を AD 変換した後、10 ms ごとに 16 チャンネルの帯域フィルター群で分析し、周波数スペクトルを得る。音声のエネルギーを用いて単語の音声区間を検出して、その区間の 16 等分点を求め、その最近傍フレームをリサンプル点として 16×16 点の標本パターンを得る。このデータと単語辞書データとの類似度(複合類似度) S_k を次式で計算する。

$$S_k = \frac{\sum_{i=1}^n \lambda_j (g\phi_j)^2}{\lambda_1 \|g\|^2} \quad (2)$$

図 1 電話音声単語認識の手順 (ハイブリッド構造マッチング法)

写真 1 電話音声認識・応答システム

ここで、 g は未知入力パターン、 ϕ_j と λ_j は単語 k の標準パターンである。 ϕ_j と λ_j は、KL 変換で得られる固有ベクトルおよび固有値であり、その数 n はパターンの変形を考慮して十数個用意されている。

複合類似度で一意に決定できない時には、類似度の高い m 位までの候補単語に対して、単語の音韻・音響的特徴系列を用いた遷移ネットワークで判定を行う。音韻 (5 母音と鼻音) と音響特徴 (無音性、摩擦性、破裂性など 12 種) を各フレームごとに判別し、ラベル系列を求める。これを各単語ごとに用意された遷移ネットワークに順次入力し、受理されるか否かを判定する。最終判定は、最大類似度 $S_{\max}$ が小さければ、雑音とみなしリジェクトする。 $S_{\max}$ が十分に大きく、他に近い値を持つ候補がなければ、その第 1 位カテゴリーを答とし、他に候補があれば、遷移ネットワークで受理されたものの中から最大の類似度をもつものを答としている。また、受理される候補がなければリジェクトする。

この認識方式は、電話サービス用音声認識・応答システム⁸⁾として実用化されている。認識率は、数字を含む 16 単語で 96% 以上である。

写真 1 に、代表的な応用例として、銀行の残高照会や振り込み通知などの電話サービス業務に使用されている例を示す。

2.2 連続単語認識

不特定話者単語認識の開発と並んで、連続的に発声された数単語を認識する研究開発も活発に進められている。離散発声の単語認識では、単語の前後にある程度長い無音区間が存在するので、音声信号のパワーを調べるだけで単語部分の検出切出しが容易に行えるが、連続発声では、単語の境界が事前にはっきりつかめないため、境界を探索しながら認識しなければならない。また単語中には、無声破裂音 ($[k]$, $[t]$, $[p]$ など)

の前には 150 ms 程度の比較的長い無音区間があるため、単語境界との区別をつけながら判定する必要がある。このため、セグメンテーションを陽に行わない方向の 2 段 DP マッチング法⁹⁾(日電)、逆 DP マッチング法¹⁰⁾(武蔵野通研)、連続 DP マッチング法¹¹⁾¹²⁾(電総研、日立)や、また連続組合マッチング法¹³⁾(東芝)などが提案されている。

2 段 DP マッチング法は、連続発声された音声信号に対し、局所的な 1 単語内の時間軸の変動と、大局的な単語間の分離のすべての組合せの中から、2 段階の DP によって最適な組合せを求めるものである。現在、この方法が DP-100 (日電) およびその後継機の DP-200 として製品化され、5 単語までの連続入力を可能としている。この他、実用化されているものとしては、完全な連続単語認識ではないが、Threshold-680 (Threshold Technology 社)では、50 ms 以上連続した無音区間を単語の切れ目の候補としてあげ、すべての組合せの中から DP により最適な分割を求めて、現実的に数単語の連続入力を可能にしている。

これまでの方法は、いずれも DP マッチングを用いて特定話者を対象に連続単語認識を行うものであるが、不特定話者を対象に、単語のマッチング領域を変えながらすべての組合せの類似度を求め、連続単語認識を行う新しい試み¹³⁾も報告されている。

現在、特定話者の連続数単語の認識が実用化されているが、さらに語数の多い連続単語や文章入力を目的とした開発が進められている。単語単位の認識では、Syntax directed DP 法 (ペル研) や、オートマトン制御 2 段 DP 法 (日電) が比較的良好な成績を示している。また音韻、音節を単位とした連続音声認識の試みも数多く報告されている。文章入力を目的とした連続音声認識については次章で詳述する。

2.3 大語彙認識

1000 単語以上の大語彙を認識する研究も盛んである。現在市販されている音声認識装置は、オプションで 500~900 単語まで拡張可能といわれているが、音声登録の問題や認識率低下の問題などがあり、ほとんどが 200 単語以下で使用されている。

大語彙単語認識の研究では、語数の増加に伴う処理時間や辞書量の増大、話者登録の効果的な解決が必要で、これらを含めた認識方式の開発が主要である。

特定話者を対象とした方式では、SPLIT 法¹⁴⁾(武蔵野通研)、ルート限定マッチング法¹⁵⁾(富士通研)、単音節を認識単位とする方法(京大、武蔵野通研、電総研)などが報告されており、また、不特定話者を対象にした研究には、音声の時間情報を生かした東北大

図 2 SPLIT 法による単語音声認識システム

の方式³⁾が新しいところである。

SPLIT 法は、大語彙で問題となる演算量の低減と単語辞書のメモリ量削減を目的として提案された方法で、単語辞書を擬音韻標準パターンの系列で表現するところに特色がある。擬音韻標準パターンは、学習用の単語から得られるスペクトルパターンをクラスタリングして求めている。

図 2 に SPLIT 法による単語音声認識システムの構成を示す。入力音声は、図 2 の (a) でスペクトル分析され、フレームごとに (b) の擬音韻標準パターンとのスペクトル距離計算が行われ、距離行列が生成される (c)。この距離行列と単語辞書 (d) を用いて、DP マッチング (e) を行い、距離が最小となる単語を出力する。

フレーム周期 16 ms、擬音韻標準パターン数 256、距離尺度としては LPC ケプストラム距離および重み付 LPC スペクトルマッチング尺度を用いて、日本の 641 の都市名を対象に男性 4 名による予備実験が行われ、平均 97.9% の成績が報告されている。この方法は、認識率の向上、演算量の低減には効果があったが、単語辞書の作成法 (現状では発声が必要) については課題を残している。

この他、文節単位の入力を目ざしたルート限定マッチング法では、3000 単語を対象にした認識実験が進められている。この方法は、入力音声の時間軸をあらかじめ決められた 9 種類の時間伸縮パスにそって圧縮し、複数の圧縮パターン (マッチングユニット) を作

図3 不特定話者多数単語認識の手順

成する。この圧縮パターンと、線形に圧縮した辞書パターンとの間で距離計算を行い、その最小の距離を入力パターンと辞書パターンとの距離とし判定するものである。

時間伸縮パスは、1000 単語のデータから傾斜制限付 DP マッチングで求めたマッチングパスを、クラスタリングにより 9 種類にまとめている。マッチングパスを制限することで、通常の DP マッチングよりも計算処理量を 1 桁以下に削減できている。

話者 2 名による 3002 単語の実験では、認識率が 89.8%，また、10 位までの累積認識率は 98% であり、さらに 2 次判定の必要性が検討されている。

不特定話者を対象にした多数単語認識の研究は数が少ないが、東北大学では図 3 に示すような認識手順のシステムを開発している。この方法は、音素認識と音素系列による単語ごとの尤度計算が基本となっているが、混同しやすい音素の識別に時間-周波数パターンを用いて成績の向上を図っている点に特徴がある。現在、成人 20 名による 212 単語の実験で、音素認識率 75%，また単語認識率 92% が報告されている。

3. 文 章 入 力

表 2 に、文章入力を対象とした認識方式の分類を示す。このうち、単音節単位に区切って発声した文、および少数 (100 語程度) の語彙からなる連続音声入力が現在、実用段階にある。文節単位の文入力および中規模 (500～数千語) 以上の語彙からなる連続音声の認識は、まだ基礎研究の段階である。

この章では、初めに単音節単位の文章入力に触れる。単音節認識は、まず 68 単音節を認識できる装置

表 2 文章入力単位による分類

認識対象	語彙数	用 途
単音節単位	大	データ&テキストエントリー
文 節 単 位	大	同 上
文 章 単 位	小	対話形端末
	中	情報検索 データエントリー
	大	テキストエントリー

(注) 語彙数
10 100 1,000 10,000
— 小 —→— 中 —→— 大 —

が各社で開発され¹⁶⁾¹⁷⁾、これと日本語ワードプロセッサで構成した 68 単音節入力音声ワードプロセッサも試作もしくは製品化された¹⁸⁾。現在、より自然な発声で文章を入力するための試みが続けられているが、次節では、最近開発された日本語のすべての単音節を入力対象とする日本語音声ワードプロセッサ¹⁹⁾について概要を述べる。

比較的少数の語彙を対象とする文章入力では、オートマトン制御を組み込んだシステムが高い認識性能を得ている (3.2 節)。また、1000 単語程度を対象とした連続音声の認識システムでは、これまで“ネットワーク形モデル”を採用したシステムが比較的良い成績を示している (3.3 節)。以下では、これらについて紹介する。

3.1 日本語音声ワードプロセッサ¹⁹⁾²⁰⁾

日本語は、漢字・ひらがな・カタカナ・英数字・記号を含む数千字種の文字を入力する必要があるため専用のオペレーターが必要となり、欧米諸国のように英文タイプライターを利用する場合と比較すると、一般ホワイトカラーの文書作成能力は数分の 1 程度になるといわれている。

音声入力による文書作成は、オペレーターをカナキーボードの操作から解放すること、初期訓練が少ないことなどから、誰でも楽な姿勢で気軽に利用できる理想的な方式といえるが、口述された音声、内容を限定せず自動的に認識し文書化する装置が、近い将来実現される見込みは少ないと考えられている。

一方、日本語のように比較的少ない種類の音節が 1 文字ずつに対応し、しかもこれを 1 拍として発音できる言語では、区切って話された音節の系列から文章を構成することができる。これは、例えば英語の音節の数が約 3200 種もあり、また 1 拍で発音できる音は 8 万種以上といわれていることと比較すると、日本語の大きな特長である。また認識単位が孤立単音節となるため、セグメンテーションが簡単 (子音-母音間の分離

図 4 日本語入力方式比較

写真 2 音声ワードプロセッサ

のみ)になるとともに、調音結合の影響が少ないため、連続音声の場合と比較して高い音韻識別率が期待できる。

日本語の音声ワードプロセッサ開発は、単音節単位入力→文節単位入力→口述文章入力と段階的に進むことになると考えられる。図 4 に、主な入力方式の入力速度と習熟に必要な訓練の度合を示した。単音節単位入力は、60 字～80 字/分 (120 音節/分) 程度の速度が限度になるため、タッチ打鍵 (ブラインドタッチ) を習得していない非熟練者向けの入力手段といえる。また文節単位での入力が可能になると、わずかな訓練で専門のオペレーターと同程度のスピードで文書が作成できるようになる。

音声ワードプロセッサの外観を写真 2 に示す。

音声ワードプロセッサでは、オペレーターの負担を軽減するうえで次のような機能が大切となる。

- (1) 拗音を含む 101 単音節を識別できること。これにより、小文字キーを使わずにすみ、入力速度が向上する。また促音付き音節 (行った [イッタ]) など) や長母音音節 (情報 [ジョー・ホー]) など)

図 5 音声ワードプロセッサの構成

もそのまま入力できることが望ましい。

- (2) 助詞「は、へ、を」のほか、長音についても話し言葉-書き言葉の変換 (正書法変換) を自動的に行えること (～でしょう [～・デ・ショー] など)。

- (3) 単音節音声の認識結果は、ほとんどの場合に正解が上位 1～3 位にくる。したがって、言語処理部では候補文字から自動的に補正できること。

音声ワードプロセッサのシステム構成をも図 5 に示す。

単音節認識部では、単音節音声を 16 チャンネルのデジタルフィルターで 8 ms ごとに分析した後、母音-子音の順に認識する²⁰⁾。認識結果は上位 2 位までを言語処理部へ転送する。また促音、長母音は付加や脱落を考慮して (つ、#) のように脱落を示す記号 (#) を候補として使用している。

単語認識部は、文章中使用する句読点や数字、および「挿入、選択、…」などのファンクションを指定する語、計 32 単語を認識する。単語音声は、線形に時間伸縮した後、あらかじめ不特定多数の話者の発声した音声から作成した標準パターンセットとの複合類似度を計算しており、認識率は 99% を得ている²¹⁾。

言語処理部は、文節単位 (文節の指示はリモートキーから行っている) に入力したかな文字列を、単語辞書と照合しながら漢字混じり文に変換する。この際、音声認識部から送られた候補文字列を単語辞書と文法規則を用いて自動的に補正している。また、発音と表記が異なる助詞や長音も正書法表記へ自動的に変換する。

3.2 連続単語音声認識による文章入力

特定話者の連続単語を認識する方法としては、これ

図 6 テキストエディターの有限オートマトン

図 7 文音声処理の流れ

までに 2 段 DP マッチング法⁹⁾を初めとして、計算時間を節約した LB (Level Building) 法²²⁾、クロック同期伝播形 DP (CWDP) 法²³⁾などが提案されている。

2 段 DP 法では、入力音声の部分区間 (l, m) に対して、まず各単語レベルの距離 $D(l, m, n)$ を求める。次に、これを用いて各時刻 m に対して次式、

$$T(m) = \min_{l, n} [T(l) + D(l, m, n)] \quad (3)$$

を満足する単語系列 (文) を認識結果として与えている。この方法は、連続数字音声の認識に対して高い識別率が得られていたが、単語数を増やし精度を向上させるためにオートマトン制御が導入された。この場合、文レベルの判定には次式が用いられている。

$$T(m, q) = \min_{l, p, n} [T(l, p) + D(l, m, n)] \quad (4)$$

ここで、 p, q はオートマトンの状態である。図 6 は、テキストエディター制御文の入力に使用された 26 個の状態からなる有限オートマトンの例を示す。この方式は、対象ごとに遷移図を作成し、登録単語の学習をする必要があるため、比較的小規模のタスクに向いている。

3.3 連続音声認識

文声を正確に認識し、文字列に置き換えることは非常に難しい。文章を認識対象とする研究では、認識システムが発話者の意図を理解し、それに対して正しい応答を行うことを目標に研究が進められてきた。この場合でも、文章中の単語をいかに正しく認識するかが重要であることに変わりない。そして、このためには音声波形を分析して得られる情報以外に、多くの知識源 (knowledge sources) を積極的に利用する必要がある。構文情報 (syntax)、意味情報

(semantics)、一般常識 (pragmatics) などを動員して、単語列に対する“仮説-検証”を行う手法が多く提案されている。

図 7 に文音声の一般的な処理の流れを示した。音声波は、音響分析部で BPE 分析もしくは LPC 分析された後、音韻辞書中の標本パターンと照合され複数候補からなる音韻系列がその類似度と共に単語照合部へ送られる。単語照合部では、この音韻系列から単語を識別するが、同時に知識源を利用して検索する単語を絞ったり、単語列が妥当か否かの検証を行うことになる。

これまでに試作された音声理解システムでは、すべての知識源をネットワークの中に埋め込んだ“ネットワーク形モデル”が成功を納めている。図 8 は、

図 8 Harpy システムの処理の流れ

図 9 部分遷移ネットワークの例

ARPA (米国防省高等研究計画局) の援助で進められた音声理解研究の中で試作された CMU (カーネギー・メロン大) の Harpy システムの処理の流れを示す²⁴⁾。このシステムでは、すべての可能なネットワークがあらかじめ Knowledge Compiler により生成される。1011 単語から成る情報検索では、15,000 の状態を持つネットワークが生成されている。図 9 に、部分ネットワークの例を示した。

入力音声は音韻認識の後、まずネットワークの先頭の音韻との尤度 (音韻認識時の対数尤度比と各音韻の継続時間テーブルから与えられる遷移確率を用いて計算される) が求められる。ビームサーチ法では、この尤度が一定の閾値を越えるパスを枝刈りしながら終端まで進み、累積尤度から最適単語列を決定している。

ARPA 計画以後、意欲的に研究が進められた IBM のシステムでは、Harpy の場合、音韻の継続時間という先験的な遷移確率で状態が遷移したのに対して、あらかじめ膨大な訓練用データから音韻ネットワークおよび単語ネットワークの状態遷移確率を推定する方法が採られている²⁵⁾²⁶⁾。このシステムの性能は、上述の Harpy と同一タスクで 99.9% と高い単語認識率を得ており、また語彙数 1000 単語のレーザー特許文に対しても 91% の単語認識率を得ている。しかし、このシステムではネットワークの遷移確率を求めるために、1 名の話者が長時間 (約 2 時間) 発声する必要があり、またタスクが異なる新たな訓練が必要となる。

4. む す び

最近の音声認識技術を、単語入力と文章入力の 2 つの観点から整理して、その現状と動向を概観した。

単語入力では、特定話者-限定単語認識は完全に実用段階にあり、ここ数年の新しい動きとして、電話音声認識、連続単語認識、および大語彙認識が、実用化

と関連して活発に開発が進められていることを紹介した。また文章入力では、注目を集めている音声ワードプロセッサの開発状況、および今後重要となる連続音声認識の動向を概説した。

残された課題としては、不特定話者大語彙認識、連続音声認識、それに音声理解といった大きな問題があるが、これを支えるパターン認識技術の開発も重要である。実用化されているものも含めて、認識性能の今一段と向上と安定化が必要と考えられる。

(昭和 58 年 3 月 10 日受付)

【参 考 文 献】

- 1) 長島, 中津, 小島, 石井: “不特定話者音声認識法”, 信学会情報・システム部門全大論文集, pp. 1-107 (1981)
- 2) 石井, 中津, 木村, 長島: “電話入力用音声認識装置の方式構成”, 信学会情報・システム部門全大論文集, pp. 1-110 (1981)
- 3) 牧野, 本間: “不特定男子話者を対象とした単語音声認識システム”, 第 8 回東北大応用情報学研究センターシンポジウム予稿集, pp. 43-93 (1983)
- 4) 麻田, 坂田, 高橋, 竹林, 平井, 篠田, 新田, 渡辺: “ハイブリッド構造マッチング法による電話音声の認識”, 日本音響学会講演論文集, 1-4-16 (1982)
- 5) 渡辺, 亘理, 千葉, 高橋, 大淵, 上野山, 保屋, 坂井: “不特定話者用音声認識装置 SR-1000 シリーズ”, 日本音響学会講演論文集, 3-1-24 (1981)
- 6) 亘理, 渡辺, 千葉: “不特定話者用音声認識装置の試作”, 日本音響学会講演論文集, 3-1-23 (1981)
- 7) 能勢, 水野, 飯塚, 田部井, 広田, 山田: “不特定話者単語認識”, 日本音響学会講演論文集, 2-1-20 (1981)
- 8) 篠田ほか: “電話音声認識装置の開発とその応用”, 日本音響学会講演論文集, 1-4-17 (1982)
- 9) H. Sakoe: “Two Level DP-Matching—A Dynamic Programming based Pattern Matching Algorithm for Connected Word Recognition”, IEEE Trans., 27, 6, pp. 588-595 (1979)
- 10) 嵯峨山, 中津: “逆 DP マッチング法による連続音声認識”, 日本音響学会講演論文集 (1979)
- 11) 岡隆一: “連続 DP を用いた連続単語認識”, 日本音響学会音声研究会資料, S 78-20 (1978)
- 12) 市川, 小松, 浅川, 松坂, 目瀬: “連続音韻認識”, 日本音響学会講演論文集, 2-1-14 (1981)
- 13) 浮田, 新田, 渡辺: “不特定話者連続数字認識方式について”, 日本音響学会講演論文集, 2-2-12 (1983)
- 14) 菅村昇: “SPLIT 法による単語音声認識”, 第 8 回東北大応用情報学研究センターシンポジウム予稿集, pp. 21-31 (1983)
- 15) 木村, 岩田, 奈良, 木島, 小林, 佐々木: “簡易時間伸縮パターンを用いた多数単語認識実験”, 日本音響学会講演論文集, 1-4-9 (1982)
- 16) 吉田, 迫江, 千葉: “日本語単音節認識実験”, 日本音響学会講演論文集, pp. 607-608 (1979)
- 17) 新田, 近藤, 村田, 坪井, 後藤, 渡辺: “音声タイプライターの試作”, 信学総合全大, 5-378 (1980)
- 18) 新田, 村田, 坪井, 河田, 武田, 渡辺: “音声ワードプロセッサ”, 東芝レビュー, 36, 7, pp. 631-636 (July, 1981)
- 19) 新田, 村田, 竹林, 坪井, 松浦, 田沢, 河田, 武田, 渡辺: “単語節入力音声ワードプロセッサの開発”, 信学全大 (1983)
- 20) T. Nitta, T. Murata, H. Tsuboi, K. Takeda, T. Kawada, S. Watanabe: “Development of Japanese Voice-activat-

ed Word Processor using Isolated Monosyllable Recognition", Proc. ICASSP'82, pp. 871-874 (1982)

- 21) 齊藤, 村田, 松浦, 新田: "不特定話者単語音声認識に関する一考察", 日本音響学会講演論文集, 2-2-14 (1983)
- 22) C. S. Myers, L. R. Rabiner: "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition", IEEE Trans., ASSP-29, 2, pp. 284-297 (Apr., 1981)
- 23) 迫江, 亘理: "クロック同期伝播 DP 法による連続音声認識の検討", 日本音響学会音声研究会資料, S 81-65 (Dec., 1981)

- 24) B. Lowerre, R. Reddy: "The Harpy Speech Understanding System", In Trends in Speech Recognition, W. A. Lea (Ed.) Prentice Hall, pp. 340-360 (1980)
- 25) F. Jelinek: "Continuous speech recognition by statistical method", Proc. IEEE, 64, 4 (1976)
- 26) A. Nadas, et al: "Continuous speech recognition with automatically selected acoustic prototypes obtained by either bootstrapping or clustering", Proc. ICASSP '81, pp. 1153-1155 (1981)

昭和 58 年度テレビジョン学会誌編集委員会の構成

編集長 和久井孝太郎(N H K)
編集理事 大越 孝敬(東 大)
 日下 秀夫(N H K)
編集幹事 高田 恒佳(東京放送)
 (企画担当) 竹村 裕夫(東 芝)
 村松 珊吾(N H K)
 本橋 昭二(日 電)

論文部門委員会

幹 事 石田 順一(N H K)
 竹本一八男(日 立)
特別委員 佐藤 和弘(日 立)
 笹野 晃(日 立)
 若菜 忠(電 電 公 社)
 永野 俊(電 総 研)
 原島 博(東 大)
 栗原 忠正(日 電)
 三田 誠一(日 立)
 杉浦 幸雄(N H K)
 土屋 裕(浜松ホトニクス)
 鮎田 和夫(東京放送)
委 員 市原 卓(東 芝)
 加藤 勇(早 大)
 河原 哲夫(東 海 大)
 木田満洲男(日 電)
 黒川 湛(日電ホームエレ
 クトロニクス)
 小泉 利雄(コロムビア)
 佐藤 栄一(N H K)

齊藤 利夫(東京放送)
 斎藤 隆弘(神奈川大)
 笹野 晃(日 立)
 鈴木 忠二(シャープ)
 関 祥行(フジテレビ)
 瀧山 龍三(九州芸工大)
 竹歳 和久(N H K)
 千原 洋行(テレビ東京)
 津田 俊隆(富士通研)
 富井 薫(松下技研)
 中嶋 正之(東 工 大)
 中田 孝憲(電 電 公 社)
 永井 信夫(北 大)
 長澤 庸二(東 北 大)
 野口 浩(日立電子)
 橋本 秀雄(電 電 公 社)
 長谷部 望(日 大)
 畑中 義式(静岡 大)
 浜崎 健二(N H K)
 早川 穆典(ソ ニ ー)
 平井 有三(筑 波 大)
 平沢 一紘(筑 波 大)
 松浦 重雄(日 立)
 三浦 秀規(ソ ニ ー)
 三谷 政昭(電 機 大)
 三橋 哲雄(N H K)
 森 克己(電 電 公 社)
 森永 規彦(阪 大)
 吉田 興夫(東 芝)

渡辺 苞(富士フィルム)

海外文献部門委員会

幹 事 二宮 佑一(N H K)
 村上 仁己(K D D)
委 員 小倉 健司(電 電 公 社)
 大熊 直彦(東 大)
 岡村 智之(フジテレビ)
 貝島 誠(東 芝)
 春日井真一郎(東京放送)
 下世古幸雄(郵 政 省)
 下元 泰治(日 立)
 諏訪 昭夫(電 通 大)
 内記 一晃(日 電)
 中川 正雄(慶 大)
 牧田 通宣(テレビ朝日)

ニュース・特許部門委員会

幹 事 木村 曉(N H K)
 荒木 洋哉(日本テレビ)
委 員 大井 一成(東 芝)
 佐々木 寛(K D D)
 白神 俊明(ゼネラル)
 田家 邦宏(沖 電 気)
 平野 真司(ビク タ ー)
 丸山 優徳(日 立)
 吉室 誠(郵 政 省)
 寄田 和男(日 電)