解　説

単語音声の認識*

千葉　成美**

1. はじめに

音声は人間と人間との最も自然なコミュニケーションの手段であり，それを人間から機械へのコミュニケーションに利用可能とするのが音声認識である．音声認識の究極的な目標は，普通の話し言葉を，話し手(話者)を問わずに，即時に認識できることであるが，この目標が手頃な価格で実現されれば，応用分野はほとんど無限にあると言えよう．しかし，このような一般的な音声認識は，文字認識でいえば，走り書きの漢字かなまじり文の認識かそれ以上に相当する複雑さ，困難さを持っており，近い将来に完全に解決されることは期待できそうもない．一方，最近のコンピュータ利用の一般化と共に，特にその入力がネックとなっており，従来からのキーボード，ペンタッチ，OCR などに代る新しい入力手段として，音声入力へのニーズが高まってきている．音声入力の主な特長として期待されているのは，①人間にとって最も自然な入力手段であるため，特別の訓練なしに誰にでも容易に使える，②マイクロホンを頭部に固定することにより，両手が自由になり，また体を自由に移動できるので，作業能率が向上できる，③既設の一般電話機からの入力が可能になる（これにより端末のコストが無視できる），ことなどであろう．これらの特長は，音声認識の究極的な頂上を極めなくとも，例えばその3合目とか5合目位の中間目標として，ある程度限定された機能が実現されれば，多くの場合，十分生かすことができるのではないかと考えられる．このような中間目標の1つが，ここで述べる単語音声認識であり，具体的には，たかだか数百語程度の限定語彙を認識の対象とするものである．

単語音声の認識は，使用目的から要求される性能によって，いくつかの方式に分けることができる．いづれにしても，実用化にたえるだけの，99% 前後の認識率が得られると同時に，実時間処理が可能で，かつ妥当なコストが実現できることが必要条件である．これらの条件を一応満足するものとして，最も簡単な種類の単語音声認識装置が，すでに 1970 年代の初めに米国で実用化されている[1]．この認識方式は，開発当時のハードウェア技術レベルによる制約から，アナログ回路による特徴抽出に重点を置いているため，性能上の限界があり，適用可能な分野はかなり限定されたものであった．その後，ディジタル演算に重点を置いた処理により，一段と優れた性能を実現した認識装置が国内で開発され[2),3)]，製品化されている．これは，最近の集積回路技術を採り入れることによって複雑なアルゴリズムの実時間処理に成功したものである．今後は，このような方向でより優れた認識方式が実用化されていくものと思われ，これに伴って音声入力が広く使用され，社会的ニーズにこたえていくものと期待される．

本文では，このような単語音声の認識方式の分類と代表的な認識システムの紹介によって単語認識の研究開発の流れを概観し，続いて実用化における問題点と音声入力の応用分野などについて述べることにしたい．

2. 認識方式の分類

パターン認識の基本的な問題は，パターンのいろいろな変動にいかに対処するかにつきるといえる．音声は，物理的には，音源，すなわち声帯の振動波または声道のせばめで生じる乱流による雑音波，が声道内で共鳴を受けたものであり，音源や共鳴の状態が時々刻々移り変っていくことにより音声パターンが形成されている．音声パターンの基本的な単位は音素と呼ばれており，発音記号と対応している．音声パターンの変動の要因としては，①発声速度の変化，②発声器官の

* Spoken Word Recognition by Seibi CHIBA (Communication Research Laboratory, Central Research Laboratories, Nippon Electric Co., Ltd.).

** 日本電気株式会社中央研究所通信研究部

形状やその動かし方に起因する個人差，の他に，③時間的に近い音素同士が互に影響を及ぼしあう調音結合と呼ばれる現象もある．これらの変動要因への対処の仕方に関連して，単語音声認識へのアプローチは，次のような3つのパラメータによって規定することができ，これによっていくつかの方式に分類することができる．

(A) 発声の方法：1（離散発声）；2（連続発声）
(B) 話者への適応：1（全単語学習）；2（学習不要）
(C) 基本識別単位：1（単語）；2（音素）

このうち，(A)と(B)は認識装置の外部仕様に関するもので，適用分野に影響を与える．

(A)の発声の方法は，単語と単語の間に区切りを置くかどうかであり，(A)-1では，単語間に完全に区切りを置いた発声を対象としている．この場合には単語の前後に必ず無音区間があるため，単語の切出しが容易にできるので，時間的な変動の正規化も行いやすく，単語の認識上極めて有利になる．この場合，単語中には無声破裂子音（/k/ など）の前に 150 ms 前後の無音区間が存在するため，単語間に 200 ms 程度以上の無音部を置く必要があり，これによって平均的な情報入力速度は低くおさえられる．これに対して，(A)-2では，数個の単語を連続して一息で発声したものを認識の対象としている．この場合には，連続的な音声パターンを，最終的には単語単位に分離したものとして認識する必要があるため，技術的には格段に困難な問題となる．その代償として入力速度は大幅に向上し[4]，また，発声に対する制約も少なくなるため，使い勝手の上からも有利になる．従来製品化された単語認識装置は全て(A)-1の方式であったが[1),5)]，最近になって(A)-2の方式の装置[2),3)]が実用化された．

(B)の話者への適応は，音声パターンの個人差に対処する方法に関するものである．(B)-1では，認識対象の全単語をそれぞれ1回から10回程度学習のために発声させ，認識装置の内部パラメータをその話者に適応させてから認識が行われる（特定話者の音声認識）．この方法によれば，話者の個人差に起因する音声パターンの変動（主として周波数構造に関連する）は無視できることになり，発声速度による時間軸上の変動のみを正規化すれば，学習サンプルを標準パターンとして用いて，パターンマッチング法により認識を行うことができる．このため，比較的単純な認識処理により高い認識率が得られる．現在製品化されている認識装置はすべてこの方式によっている．この場合，話者の標準パターンを作成するのにある程度の時間がかかるが，業務用として継続的に使用する場合にはほとんど問題にならない．これに対して，(B)-2では，個人差による音声パターンの変動を前以って多人数の音声サンプルを用いて調べておき，このような変動のもとで安定な認識が行われるような識別機構を作っており，話者ごとに別個に適応させることは行わない（不特定話者の音声認識）．この方式では，時間軸の変動に加えて，個人差による複雑な音声パターンの変動に同時に対処しなければならないため，高い認識率を得ることは容易ではない．現在このような方式は研究開発中であるが，最近ではかなり良い結果が得られている[6),7)]．これが実用化されれば，一般電話機を用いて不特定多数からの音声入力が可能になるため，その意義は大きい．なお，これらの中間的なものとして，一部単語のみを用いて学習を行う方法も試みられている[8)]．

(C)の基本識別単位は，認識装置の内部において情報量の多い特徴パターンのレベルからディスクリートな少数のシンボルへの変換，すなわち識別を何を単位として行うかに関するもので，これにより所要メモリ量，処理量などが大きく影響される．(C)-1では，特徴パラメータから直接単語の識別が行われる．これに対して，(C)-2では，その中間において一旦音素（又は音節）の識別が行われ，その結果に基づいて最終的に単語の認識が行われる．この両者を比較すると，(C)-1では認識処理が単純で，中間での情報の損失が少ないため高い認識率が得られる傾向にあるが，反面，単語当りに要するメモリ量，処理量とも大きくなるため，比較的少数語彙の場合に適している．これに対して，(C)-2では，単語の中で互いに調音結合の影響を受けた音素に対して，セグメンテーション及び識別を十分安定に行うことは困難であり，また周囲雑音の影響もより受けやすくなると思われる．しかし，この方式では不特定話者の音声認識の場合には，語彙の変更，拡大が単語辞書の変更のみで容易に実現できる利点がある．現在製品化されているのはすべて(C)-1の方式であり，(C)-2の方式は基礎的な研究の段階である[9)]．また，いわゆる音声理解システムでは，通常音素識別が行われるので，その研究成果が今後単語認識に利用されていくことも考えられる．

以上述べた，発声方法に関する(A)，話者への適応に関する(B)，及び識別単位に関する(C)の3つのパ

表-1 単語音声認識方式の分類

No.	タイプ (A) 1: 離散発声 2: 連続発声	タイプ (B) 1: 特定話者 2: 不特定話者	タイプ (C) 1: 単語単位 2: 音声単位	方式の特徴	適用分野	実用化状況	代表例	世代*1
1	1	1	1	技術的に最も容易	仕分装置制御，検査データ入力など	製品化	Threshold[13),14)] Scope[17)]	第1
2	2	1	1	高速入力が可能	同上	同上	日電[2)]	第2
3	1	2	1	電話入力に適す	予約，問合せなど	研究開発中	日電(大プロ)[6)]	第3
4	2	2	1	(技術的に困難)	同上	—	—	
5	1	1	2	メモリ，処理量が小	No. 1 に同じ	研究開発中	東大[11)]など	
6	2	1	2	(SUS*2 に包含される)		—		
7	1	2	2	語彙の変更，拡大が容易	No. 3 に同じ	研究開発中	東北大[9)]	第4
8	2	2	2	(No. 6 に同じ)		—		

*1 単に実用化の順序（予想も含む）を示す．
*2 Speech Understanding System

ラメータにより，単語音声認識装置は **表-1** に示すように8種類に分類することができる．このうち，技術的に最も容易な (1, 1, 1) タイプは，第1世代システムとも言うべきもので，既に数年前に Threshold Technology 社及び Scope Electronics 社により開発され製品化されている．特に Threshold の製品は，これまでに米国を中心に 200 システム前後が出荷されたと言われており，限られた分野ながら着実に音声入力の地歩を築きつつある．連続発声単語の認識が可能な (2, 1, 1) タイプは最近になって初めて日本電気から発表された．これは第2世代システムとも言うべきもので，これにより入力速度が向上するため，特定話者向認識装置の応用分野が拡大するものと期待される．続いて第3世代システムとして実用化されると思われるのは (1, 2, 1) タイプであり，その後，(1, 2, 2) タイプが第4世代システムとして実用化される可能性がある．それ以外の，(2, 2, 1) タイプは現在までに技術的な可能性は示されておらず，(1, 1, 2) タイプはある程度研究は行われているが[10),11)]，今のところ認識性能に問題があり，(1, 1, 1) タイプに対抗して実用化するのは難しいのではないかと思われる．(2, 1, 2) 又は (2, 2, 2) タイプになると，むしろ音声理解システムの一部として研究が行われているので，ここでは取上げないことにした．特に (2, 2, 2) タイプについては，電総研において長期的な研究プロジェクトが進行中であり[12)]，着実に成果が上りつつある．

3. 認識システム

単語音声認識に関する研究開発は，すでに 1950 年代から行われているが[13)]，ここでは，前節の 表-1 で各々のタイプの代表例として取上げたシステムを中心に，比較的最近に発表されたいくつかの特長ある単語音声認識システムについて簡単に紹介する．

Threshold のシステム　Threshold Technology 社では，その名の示すようにアナログ閾値回路による特徴抽出を用いた認識装置を開発し，1972 年に VIP-100 と名付けて製品化した[1)]．これはミニコンをベースにしたシステムで，仕分装置の制御用など，主として物流関係で使用された．その後でミニコンの部分をマイクロコンピュータ (LSI-11) に置きかえて低価格化を計った Threshold 500 が 1975 年に発表されている[14)]．これは，標準パターンの登録機能などをホストコンピュータに受持たせることにより，認識専用の音声入力ターミナルとして簡素化したものである．基本的な認識処理はこれらのモデルでは同一である．すなわち，19 チャンネルフィルタによる周波数分析結果を対数圧縮した後，1 ms 程度のフレーム周期で 32 の，1, 0 状態をとる特徴系列に変換する．特徴の内容としては，各音素に対応したものが大部分であり，その他にスペクトル形状の大まかな特徴を含んでいる[15)]．得られた 32 次元特徴ベクトルの時系列はプロセッサに読込まれ，線形時間正規化により1単語につき 16 個のベクトルに統合されて，512 ビットの単語パターンが作られる．学習モードでは，10 個程度の訓練サンプルから平均的な標準パターンが作られ，認識モードでは入力パターンと標準パターンとの類似度が計算される．類似度の計算には，時間的に1点づつ前後させた成分を加えるなどの工夫が行われている[16)]．認識語数は基本構成で 32 語で，拡張も可能となっている．

Scope のシステム　Scope Electronics 社でも 1970 年代の初めから，パターンマッチングによる単語音声認識装置を発表している[17)]．音声分析は 16 チャンネルのフィルタにより，10 ms 程度のフレーム周期で行われ，スペクトル情報は，ローカルピーク検出により

1, 0 に2値化している．時間正規化は線形ではなく，サンプル点間のスペクトルの変化量が等しくなるように，1単語につき一定の回数だけサンプリングを行う方式をとっている[18]．このような処理により最近の VDETS-1000 システムでは，1単語を 240 ビットで表わしている．標準パターンは，5回程度の発声から平均的なものが作られる．ハードウェアは，分析部とミニコン (NOVA) から構成されており，入力は4チャンネルまで拡張可能となっている．入力語数としては，数百語までの拡張が可能とされている[19]．なお，このシステムは最近 Interstate Electronics 社から販売されるようになったと言われている．

日本電気のシステム DP-100 音声入力装置として本年3月に発表されたもので，DP マッチング法[20]を改良した2段 DP マッチング法[21]により，連続発声単語の認識を初めて可能にしたシステムである．すなわち，1単語内の時間軸の変動と，連続単語における単語の結合順序に関する最適化を，2段階に DP（動的計画法）を適用することにより効率的に実現している．この方法は単語へのセグメンテーションが不要のため，それに起因する誤認識がないのが特長である．また，標準パターン作成用の登録発声は1回でよいため，実用上有利である．音声分析は 16 チャンネルフィルタで行われ，18 ms のフレーム周期でサンプリングされる．特別のデータ圧縮は行われていないので，単語当りの情報量はかなり大きくなっている．ハードウェアは 図-1 に示すように，音声分析器，高速マイクロプロセッサ，DP プロセッサ，標準パターンメモリ，入出力プロセッサからなるマルチプロセッサ構成となっている[2]．認識語数は全体で 120 語であり，これを2チャンネル入力でシェアして使用することができる．装置の外観は 図-2 に示す通りであり，音声分析を含めて全体をディジタル処理化したためコンパクトになっている．

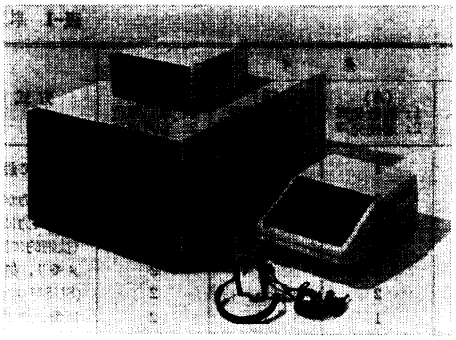

図-2 DP-100 音声入力装置の外観

以上は，これまでに製品として発表されたシステムであり，すべて特定話者向であるが，これらの諸元の比較を表-2(次頁参照) に示す．この他にも Dialog Systems 社，Perception Technology 社，Centigram 社などでも製品を発表したと言われているが詳細は不明である．

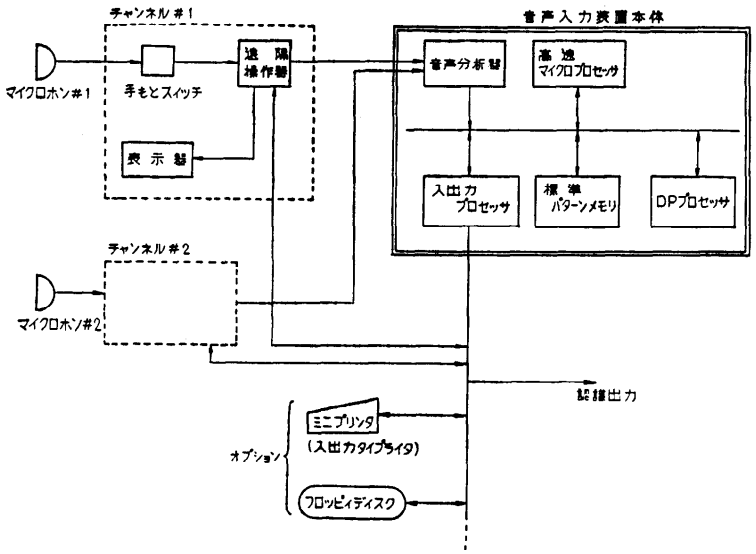

図-1 DP-100 音声入力装置のシステム構成図

不特定話者を対象とした研究開発中のシステムの代表例としては次のようなものをあげることができる．

日本電気（通産大プロ）のシステム 通産省工技院の大型プロジェクト「パターン情報処理システム」の一環として，1977年に日本電気が開発したシステム[6),7)]で，基本的には，不特定話者による音声パターンの変動をカバーするような識別関数を求め，これを用いて認識を行う方式をとっている．認識処理は単語を単位として，図-3(次頁参照) に示すプロセスによって行われる．前処理部では，12 ms

表-2 製品化された単語音声認識装置の比較

メーカ	名称	発表	認識対象	時間正規化	単語情報量	登録発声	入力チャンネル	語数	プロセッサ	形状
Threshold	VIP-100	1972	離散単語	線形	512 bit	～10回	～3	32+	NOVA	卓上形
同上	Threshold 500	1975	同上	同上	同上	同上	1	同上	LSI-11	同上
Scope (Interstate)	VDETS-1000	197?	同上	非線形	240 bit	～5回	～4	25+	NOVA	コンソール形
日電	DP-100	1978	連続単語	最適 (DP)	512 Byte*1	1回	2(1)	60(120)	MMP*2	卓上形

*1 単語長 540 ms の場合
*2 マルチマイクロプロセッサ

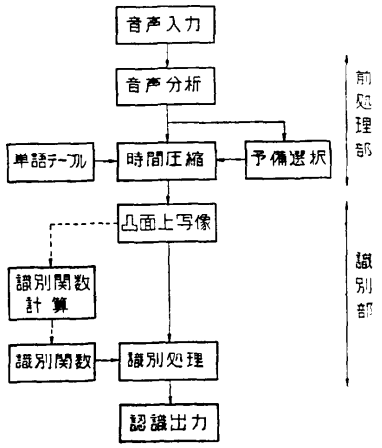

図-3 認識処理のプロセス

フレーム周期での音声分析により，8次元ベクトルの時系列として音声を表わしており，時間軸上のサンプリングにより，1単語につき15点を抽出しているので，1単語は120次元ベクトルで表現される．ここでサンプリング点の決定に当っては，入力単語を，認識対象語彙のそれぞれであると仮定した場合に，その単語の音素構造を最もよく代表するような点を求める方式をとっており，時間軸に関する正規化と情報圧縮が同時に実現されている(時間圧縮)．しかし，このために必要な処理量はかなり多くなるので，あらかじめ大まかな認識を行って，認識対象を数分の1にしぼっている(予備選択)．識別部では，120次元の特徴ベクトルを超放物面上に写像した後，区分的線形識別関数との内積が計算され，その結果を用いて等価的に区分的2次識別関数による識別が行われる．識別関数は，前以って多人数の発声した単語音声サンプルを用いて，線形計画法により計算される[21]．この識別系では，学習サンプルは完全に識別可能であるが，数字，アルファベットの一部を含む100語を対象とした場合，男女を含む40名の発声した1,500個のテストサンプルに対して，99.1%の認識率が得られている．

東北大のシステム 東北大では数年前から音素認識を基礎とした不特定話者の単語認識の研究を進めてきている[9]．分析部では，聴覚系をモデルとしてQの低い1/6オクターブフィルタによる分析を行い，スペクトルのローカルピークの形で低次ホルマントを抽出して，音素を識別している．得られた音素系列は，やはり音素系列で記述された単語辞書と照合され，単語認識が行われる．音素系列との照合にはDPマッチング[20]が用いられており，これによりセグメンテーション誤りなどの種々の変動に対する最適化が行われている．認識率は，166語の場合15名に対して83%程度となっている．

その他のシステム 公社通研では，(1, 1, 1)タイプにおいて，標準パターンを音素の特徴パラメータの系列で記述する方式を検討している[23),24)]．このような方式では標準パターンのメモリ量が低減し，また，新しい話者への適応が部分的な学習でも一応可能となる特長がある．同じく公社通研の板倉氏がベル研で作成したシステム[25]は，やはり(1, 1, 1)タイプであるが，LPC分析とDPマッチング法の組合せにより，電話音声を用いて200語の認識が可能であることを実証している．東大では，音素識別をベースにした単語認識システムの研究を行っている[11]．

4. 実用化における問題点

単語音声認識が実用にたえるためには，認識率，処理速度などの基本認識性能が十分であることは言うまでもないが，それ以外にも考慮すべきいくつかの問題点がある．

周囲雑音については，一般の事務室程度の騒音レベル(約65ホン)では特別の対策は必要ないが，仕分装置の制御用などになると，85ホン以上の騒音下での動作が要求される場合もあり，何らかの対策が不可欠である．このような場合には，単一指向性の接話形マイクロホンを用いてSN比の改善を計ることが有効でよ

く行われており，また標準パターンの登録を同じ高騒音下で行うことも非常に効果的である[4]．

入力レベル変動は，レベルメータを見ながら増幅器のゲイン設定を細く変えることにより一応対処できるが，AGC (自動利得制御) を用いることも有効である．しかし，手動のゲイン調整にたよることは操作性の上からさけなければならず，また高騒音下では AGC の動作上問題が多い．このため，処理系のダイナミックレンジは十分大きくとり，信号レベルの影響をなるべく受けないようにする必要がある．

音声の始端，終端の検出，すなわち音声の切出しは，周囲雑音がなく，信号レベルがほぼ一定の場合には，固定の閾値でレベルを検出することにより容易に行えるが，実際には面倒な問題であり，しかも認識率に与える影響は極めて大きい．そのために最近ではいくつかのパラメータから総合的に判定する方式も検討されている[26]．また，接話形マイクの場合には，単語の発声が終った後に出る息の音を強く拾うために，終端の検出が遅れることがあり，この部分を除くために特別の処理を行う場合もある[27]．

特定話者向システムにおける標準パターンの作成は，使用者の負担を軽くするためになるべく少ない発声回数ですむことが望ましい．Threshold 及び Scope (Interstate) のシステムでは 1 単語当り 10 回又は 5 回程度の発声が必要であるが，日本電気のシステムでは 1 回の発声で十分高い認識率を得ており[3]，これはスペクトル領域での DP マッチング法がこの点でも優れていることを示している．

不特定話者向のシステムは，一般電話機からの音声入力に適しているが，電話系を通した音声は，カーボン送話器の影響で歪が多く，信号帯域も 3.4kHz 程度に制限されている．このため，分析法，識別方式とも電話系に適した方式を検討する必要があり，また，使用する語彙も適当なものを選定する必要がある．

最後に，装置の価格は実用上非常に重要なファクタである．コストを下げるためには，特徴抽出方式の改良によりメモリ及び処理量を減らすのが有効であるが，認識性能を落さずにこれを実現することは簡単ではない．入力を多重化すれば，ある程度チャンネル当りのコストを下げることができるが，システム構成の柔軟性と信頼性の面からは必ずしも望ましい方法ではない．最近進展の著しい集積回路技術の利用によって，いずれハードウェアのコストが問題にならない時代となるのではないかと期待される．

5. 音声入力への応用

単語音声入力システムは 1973 年頃から米国で実用期に入り，様々な用途での音声入力に実際に応用されてきている．現在までのところ，フィールドにおける実績が報告されているのは，ほとんどが Threshold のシステム，すなわち (1, 1, 1) タイプの第 1 世代システムである．音声入力が他の手段に代って用いられるためには，何らかの利点がなければならない．これを，ハードベネフィット（金銭化できる利益，省力効果など）と，ソフトベネフィット（それ以外の利益，便利さなど）に分けた場合，当然ハードベネフィットのある所から導入が始められる．これに相当するのは，両手を使う作業に関連して情報入力を行う場合で，仕分装置の制御，検査データの入力，NC プログラムの作成などである[28]．

仕分装置制御 音声入力が最初に実用されたのは，空港における手荷物の行先別仕分システムであった．このような場合には，両手で荷物を扱いながら行先のコードを入力する必要があるので，従来のキーボード（テンキー）による入力方式では，仕分けのスピードを上げるためにはどうしても荷物の扱いとキーボードの操作を 2 人で行う必要がある．音声入力によれば，1 人で無理なく作業が進められるので，2 対 1 の省力効果があり，大きいハードベネフィットが得られる．また，キーボード入力では，行先の地名を必ず数字コードに変換してから入力する必要があるが，音声では直接地名で入力できるので，オペレータの負担を軽くし，訓練期間を短縮できるなどのソフトベネフィットも期待できる．このような音声入力仕分けシステムは，空港以外でもデパートなどの配送センタ，小包配達サービス会社などで実際に用いられている．このような仕分け作業では，単位時間当りの仕分け個数が問題となるので，平均入力速度を上げるために，連続発声単語の認識が可能な第 2 世代システムが今後用いられていくことになろう．

検査データ入力 工場における製品又は半製品の検査工程で，自動化できない場合には，熟練した検査員が手を使ったり体を動かしたりして検査を行いながらデータを入力しなければならない．このような場合に音声入力を使用すれば検査の能率が大幅に向上することになり，ハードベネフィットが得られる．具体的な応用例としては，カラーテレビ用ブラウン管の品質管理データ入力，自動車のアッセンブリラインにおける

表-3 NC プログラム作成用語彙の例[29]

The Digits 0-9	Grid
Decimal Point (.)	Line
Minus (−)	Trim
Inside	Offset
Outside	Coordinate
Profile	Define Pattern
Holes	Terminate
Increment	Erase
Arc	Cancel
Bolt Hole Circle	Go
Skip	Slope

検査データ入力，受入検査のデータ入力などがある．

NC プログラム作成 数値制御 (NC) のプログラムテープの作成を，簡易なコマンド言語を開発して，音声入力によりインタラクティブに行うシステムが開発されており，すでに数十システムが使われていると言われている．**表-3** にその最も簡単な場合の語彙の例[29]を示す．このようなコマンド系列の音声入力により NC プログラムの作成が非常に容易になり，効率化されるので，ソフトベネフィットは勿論，ハードベネフィットもかなり期待できる．

その他の応用 以上は (1, 1, 1) タイプの第1世代システムの実用例であるが，その他の応用例としては，連続音声認識システムを地図作成用のデータ入力に用いる実験[30]，軍用データシステムへの適用[5] などがある．また国内では身障者用の義手の制御に簡単な音声認識を応用するプロジェクトが進められている[31]．不特定話者の音声認識の重要な応用分野は，電話からの音声入力である．現在，押ボタンダイヤル電話器（プッシュホン）と音声応答とを組合せたデータ入力システムがいくつか稼動しており，特別の端末を必要としないため好評と言われている．しかし，特に国内ではプッシュホンの普及が全国平均で数1% 程度にとどまっているため，利用者が限定されるのが問題となっている．音声入力によれば，一般電話機からの入力が可能になり，利用者を飛躍的に増加できる．第3世代の (1, 2, 1) タイプシステムが実用化されれば，このような利用法が実現することになる．

6. おわりに

単語音声認識装置は第1世代から第2世代の時代に入ろうとしており，さらに第3世代，第4世代システムもやがて実用化されると思われる．このようにして音声入力の適用分野は拡大していく訳であるが，それでも性能的には，人間の能力に比較すれば，はるかに限定されたものであることには変りはない．従って，音声入力を使いこなしていく鍵は，いかにその限界の中で使えるように工夫していくかにかかっている．この意味で，シンタックスを用いて認識対象をしぼる方法[32),33)]は非常に有効であり，できるかぎり利用すべきである．また，手書 OCR の分野で行われているように，ある程度のユーザ教育を行うことによっても，認識率をかなり改善できると思われる．単語セットの選び方に注意することも重要である．例えば，アルファベットの A, B, C を Alpha, Bravo, Charlie などの単語におきかえることによって[5]，普通の単語認識装置でアルファベットの認識が確実に行えるようになる．この様な方法は，日本語の場合，カナ文字についても同様に適用できる．

認識装置の側からも，それぞれのタイプの中でアルゴリズムを改良して認識性能を向上させていくと同時に，汎用又はカスタム LSI の導入により，一段とハードウェアコストを下げて行かなければならない．このためには，LSI 化に適した認識方式を研究開発していくことが必要である．

コンピュータの利用形態は，いろいろな理由から，集中処理から分散処理に移行しつつあると言われている．これは，コンピュータ，従って端末装置が従来よりも広範囲にばらまかれることを意味し，これを使用するオペレータも，少数の熟練者を期待することはできなくなってくる．このような背景のもとで，非専門家に適した音声入力は，分散処理におけるソースデータ入力用を中心に，今後ますます普及していくものと思われる．

参考文献

1) M. B. Herscher and R. B. Cox: An Adaptive Isolated-Word Speech Recognition System, 1972 Conference on Speech Communication and Processing, C 1 (1972).
2) 鶴田，迫江，千葉，中田：マイクロプロセッサ化多重入力連続単語音声認識システム，昭和52年度電子通信学会情報部門全国大会，No. 221 (1977).
3) 鶴田，迫江，千葉：連続単語音声認識システムの評価実験，日本音響学会講演論文集，4-1-19 (1978. 5).
4) 鶴田，迫江，千葉：DP を用いた連続単語音声認識システム，情報処理学会マンマシン研究会資料 MMS 23-2 (1976).
5) J. J. Kalinowski, et al.: Application of Discrete Word Recognition and Response to Multiuser Tactical Communications: WRS, IEEE Int'l. Conf. ASSP, pp. 222〜225 (1976).

6) 千葉，亘理，渡辺：不特定話者を対象にした単語音声認識システム，昭和 52 年度電子通信学会情報部門全国大会，No. 219 (1977).
7) 千葉，亘理，渡辺：不特定話者を対象にした単語音声認識システム，昭和 53 年度電気学会全国大会，S. 5-7 (1978).
8) M. Kohda and S. Saito: Speech Recognition by Incomplete Learning Samples, 1972 Conference on Speech Communication and Processing, H 10 (1972).
9) 三輪，新津，牧野，城戸；音声スペクトルの概略形とその動特性を利用した単語音声認識システム，日本音響学会誌 Vol. 34, No. 3, pp. 186～193 (1978).
10) G. M. White and R. B. Neely: Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programming, IEEE Trans. Vol. ASSP-24, No. 2 pp. 183～188 (1976).
11) 藤崎，佐藤ほか：限定語彙単語認識の一方式，日本音響学会講演論文集，2-2-11 (1975. 5).
12) 中島，石崎：線型調音モデルによる調音の動的特徴抽出，日本音響学会講演論文集，1-4-21 (1975. 10).
13) 千葉，加藤：限定語彙の機械認識，電子通信学会誌，Vol. 51, No. 11, pp. 1440～1444 (1968).
14) M. B. Herscher and R. B. Cox: Source Data Entry Using Voice Input, IEEE Int'l Conf. ASSP, pp. 190～193, (1976).
15) M. B. Herscher and T. B. Martin: Word Recognition System for Voice Controller, U. S. Patent No. 3, 588, 363 (1971).
16) T. B. Martin, et al.: Word Recognition Apparatus and Method, U. S. Patent No. 4, 069, 393 (1978).
17) J. W. Glenn and M. H. Hitchcock: With a Speech Pattern Classifier, Computer Listens to Its Master's Voice, Electronics, May 10, pp. 84～89 (1971),
18) グレン，ヒッチコック：コンピュータが主人の声を聞いて動く，日経エレクトロニクス，1971 年 10 月 25 日号，pp. 70～77.
19) Scope Electronics Inc.: Voice Data Entry Terminal System VDETS 1000 Series, 製品カタログ.
20) 迫江，千葉：動的計画法を利用した音声の時間正規化に基づく連続単語認識，日本音響学会誌，Vol. 27, No. 9, pp. 483～490 (1971).
21) 迫江：2 段 DP マッチングによる連続単語認識，音声研究会資料 S 75-28 (1975).
22) 渡辺，千葉：線形計画法を用いた区分的非線形識別関数計算システム，電子通信学会パターン認識と学習研究会資料，PRL 76-41 (1976).
23) 好田，橋本，斉藤：数字音声の機械認識系，電子通信学会論文誌 Vol. 55-D, No. 3, pp. 186～193 (1972).
24) 好田，嵯峨山，長島：音韻単位の標準パターンを用いた単語音声認識装置，昭和 53 年度電子通信学会総合全国大会講演論文集，pp. 5-335-5-336 (1978).
25) F. Itakura: Minimum Prediction Residual Principle Applied to Speech Recognition, IEEE Trans. Vol. ASSP-23, No.1, pp. 67～72 (1975).
26) L. R. Rabiner and M. R. Sambur: An Algorithm for Determining the Endpoints of Isolated Utterances, B. S. T. J. Vol. 40, No. 2, pp. 297～315 (1975).
27) T. B. Martin: Applications of Limited Vocabulary Recognition Systems, Speech Recognition, pp. 55～71, ed. by D. R. Reddy, Academic Press (1975).
28) T. B. Martin: Practical Applications of Voice Input to Machines, Proc. IEEE Vol. 64, No. 4, pp. 487～501 (1976).
29) M. B. Herscher and R. B. Cox: Voice Programming of Numerically Controlled Machines, IEEE Int'l. Conf. ASSP, pp. 452～455 (1977).
30) C. Goodman, et al.: An Application of Connected Speech to the Cartography Task, IEEE Int'l Conf. ASSP, pp. 811～814 (1977).
31) 舟久保：全腕式電動義手開発の意義，自動化技術 Vol. 8, No. 1 (1976).
32) R. Alter: Utilization of Contextual Constraints in Automatic Speech Recognition, IEEE Trans. Vol. AU-16, No. 1 pp. 6～11 (1968).
33) 迫江，千葉：FORTRAN 語入力用音声認識システム，情報処理学会第 12 回大会，176, pp. 351～352 (1971).

(昭和 53 年 5 月 17 日受付)