

論文

「SPOKEN BASIC 1」の認識システム*

新美康永** 小林 豊** 浅見俊幸** 三木 豊**

Abstract

This paper describes a highly predictive speech recognition system, developed as a voice-input programming system, in which a modified version of 'BASIC', named 'SPOKEN BASIC 1', is used. The system consists of four major components; acoustic processor, lexical matching procedure, syntactic processor and semantic processor. The acoustic processor transforms incoming speech signals into a sequence of labeled segments. The syntactic processor makes grammatical predictions using a left-to-right parsing scheme and a depth-first or a breadth-first tree search. The semantic processor refines those grammatical predictions and sends predicted words to the lexical matching procedure, which correlates them with the sequence of labeled segments.

142 sentences uttered by four male speakers were processed through the present system. It responded as follows: 116 sentences (81.7%) were correctly recognized, 19 (13.4%) incorrectly recognized and the others rejected.

1. ま え が き

連続音声の自動認識は極めて困難な仕事である。一方我々人間は比較的容易に音声を理解し認識している。人間の聴覚系がどの程度精密な音響分析や特徴抽出を行っているか十分には知られていないが、上位レベルに属する知識、すなわち、文法規則、言葉の意味、話題の内容などの知識を総合的に用いて、発話内容の理解を容易にしているものと考えられている。この種の知識を総称して「言語情報」と呼ぶことにする。言語情報は言語の構成単位の小さな順に、(1)音韻レベルの情報、(2)単語レベルの情報、(3)構文レベルの情報、(4)意味レベルの情報の4つに大別することができる。

このような各種の言語情報をいかにして音声認識システムに組み込むかは古くからの懸案であり、多くの試みがなされてきたが¹⁻⁴⁾、「タスク」を限定すること、すなわち、音声認識の対象——話題、構文、単語な

ど——を限定することによって、シンタックス、セマンティックス、プラグマティックスなどを比較的厳密に記述することが可能となり、限られた範囲の知識を有効に利用する音声認識システムのモデルが確立された⁵⁾。いわゆる「speech understanding system」と呼ばれるものである。

現在実際に総合システムとして動作している音声認識システム⁶⁾⁻¹⁰⁾の規模は、単語数にして20数個から250個程度であり、タスクは種々のデータベースに対する情報検索、計算機に対する指令、問い合わせ、簡単なプログラミング言語、等々多種にわたっている。構文は人工的なもので単純な文脈自由文法で記述されている場合が多いが、近い将来の目標としては1,000個程度の単語で比較的拘束の少ない自然言語の認識理解が挙げられている。

筆者らはプログラミング言語「BASIC」に若干の制限を加えたものをタスクとして選び（これを「SPOKEN BASIC 1」と名付ける）、音声認識システムの研究を行っている¹¹⁾⁻¹⁴⁾。我々のタスクは単語数50で、利用し得る言語情報は構文情報が主であり、それより上位のレベルの情報に少し乏しいのが難点である。上記の諸例と比較してやや小規模であるが、研究の第1

* The Speech Recognition System of 'SPOKEN BASIC 1' by Yasuhisa NIIMI, Yutaka KOBAYASHI, Toshiyuki ASAMI and Yutaka MIKI (Faculty of Industrial Arts, Kyoto Technical University)

** 京都工芸繊維大学工学学部

歩としては典型的な例題といえよう。

本論文では音声認識システムの構成と共に、4名の話者の発声した文章に対する認識結果について述べる。

2. 「SPOKEN BASIC 1」の仕様

我々が認識対象とする「SPOKEN BASIC 1」は会話型プログラミング言語「BASIC」に下記のような修正, 制限を加えて簡単化したものである。

- (1) 変数を構成する英文字をX, Y, Zの3個に限定する。但し, 変数の長さを3文字まで許す。
- (2) 関数の使用は認めない。
- (3) 主として行列演算に関する幾つかのキーワードを除く。

「SPOKEN BASIC 1」で使用できる語彙とその読み方を **Table 1** に示す。ここで構文上同等と考えられる単語をまとめて, syntactic category (以下SCと略記する) と呼んでいる。この欄が空白のものは「単語」の欄と同じ記号を用いるものとする。文の生成規則はこのSCをもとにして記述されるが, それは5.1で述べる。

入力音声は文単位に区切って普通で朗読するものとする。話者も特に限定しないことを目標とする。

「SPOKEN BASIC 1」の認識システムを **Fig. 1** に示す。このシステムは音響処理部, 単語識別部, 構文処理部, 及び意味処理部の4つの要素から成り立っている。後二者をまとめて言語処理部と呼ぶ。以下に各部の詳細を述べる。

3. 音響処理部

3.1 システムで使用する音素記号

文単位で区切って発声された入力音声は, まず音響

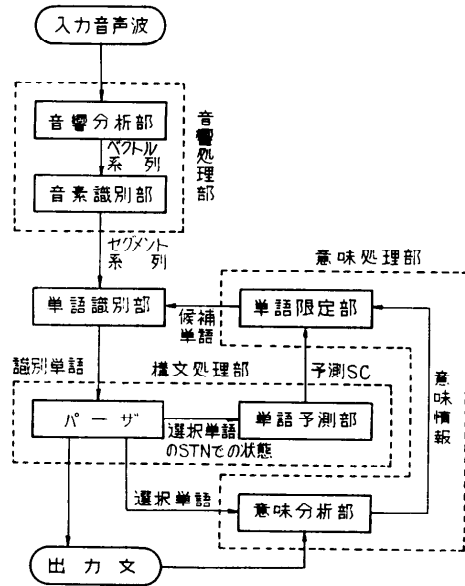


Fig. 1 Speech recognition system of 'SPOKEN BASIC 1'.

分析部において20チャンネル1/4オクターブ帯域フィルタ群で周波数分析され, 10msごとにサンプリングされ, 10ビットの精度でデジタル化されて, 20次元のベクトル(以下セクションと呼ぶ)の系列に変換される。この系列に区切り目を入れて, 各区間(セグメントと呼ぶ)に音素記号を割当てるのが音素識別部の役割である。

現在の技術水準では細かい子音の分類は労多くして実りが少ないので, 本システムでは各セグメントを次の9種類に分類している。

Table 1 Vocabulary of 'SPOKEN BASIC 1'

単語	読み方	SC	単語	読み方	SC	単語	読み方	SC	単語	読み方	SC
0	レイ, ゼロ	d	+	プラス	a	LET	ダイニュー		↑	ベキ	
1	イチ	d	-	マイナス	a	IF	モシ		E	シスウ	
2	ニ	d	*	カケル	m	INPUT	インプット		.	テン	
3	サン	d	/	ワル	m	PRINT	プリント		"	カギ	
4	ヨン	d	<	ショウナリ	r	FOR	トリカエシ		"	カギトジル	
5	ゴ	d	>	ダイナリ	r	NEXT	ツギノ		CR	カイギョウ	
6	ロク	d	キ	ノット	r	TO	カラ		CTRL	トリケン	
7	ナナ	d	GOTO	ジャンプ	g	STEP	ステップ		RUN	ジッコウ	C
8	ハチ	d	GOSUB	ヨビダシ	g	THEN	ナラバ		NEW	プログラム	C
9	キュウ	d	END	プログラム	e	(カッコ		LIST	リスト	C
X	エックス	v	STOP	ストップ	e)	カッコトジル		BYE	オワリ	C
Y	ワイ	v	RETURN	モドレ	e	,	コンマ				
Z	ゼット	v	DIM	ハイレッツ	e	=	イコール				

- 無音部
 - 文と文の切れ目・・・記号B.
 - 文中の呼気段落, または無声閉鎖音の直前の無音部・・・記号P
- 有音部
 - 母音・・・記号A, I, U, E, O
 - 摩擦子音・・・記号S
 - 有声子音・・・記号CO

3.2 母音と摩擦子音の識別

入力系列中の各セクション (X_i) をまずその大きさにより無音部と有音部とに分類する. 連続する無音セクションはこれを1つのセグメントにまとめ, その継続時間長 (セクション数) により記号P, またはBを与える.

有音部に対しては, 5つの母音と摩擦子音に対応する6個の線形識別関数 $f_i(X)$ ($i=1, \dots, 6$) を用いて記号化とセグメンテーションを行う. すなわち, 各 X_i に対して,

$$f_i(X_i) \geq f_j(X_i) \geq f_k(X_i) \quad (k=1, \dots, 6, k \neq i, j)$$

となる番号 i, j , 及びこの i, j に対して識別関数の値の比 $\gamma_{ij} = f_i(X_i)/f_j(X_i)$ を求める. 番号 i , 及び j をそのセクションに対する第1, 及び第2候補音素を表わすものとみなし, 同一の第1候補音素 i をもつ連続したセクションをまとめて1つのセグメントとする. このセグメントの中で最も頻度の高い第2候補音素 j を用いて, このセグメントに記号「 ij 」を割当て. 次いでセグメント中で最小の γ_{ij} を求め, これをセグメントに対する音素判定の確からしさを表わす尺度 (RATIO という) とする.

3.3 有声子音の検出

有声子音の存在を検出する最も単純な方法として振幅包絡線 ($|X_i|$ の時間的变化) の情報を用いる. 振幅包絡線の起伏を調べ, 一定の条件を満たす谷の部分で1つのセグメントとして記号COを割当てる. このセグメントに対しては, 判定の確からしさの尺度として谷の深さ, 及び振幅の絶対値を付しておく. このようにして得られるセグメントCOのすべてが有声子音を表わすわけではないが, 多くの有声子音部においてCOが現われるので有効である. またCO以外のセグメントに対してもその部分の振幅包絡線の状態を属性として与えておく.

かくして音響処理部の出力は, セグメント記号, 継続時間長, 振幅の状態, RATIO の4字組の系列となる. 例を Fig. 2 に示す. なおCOセグメントに対しては谷の深さと振幅値が第2項と第4項に入っている.

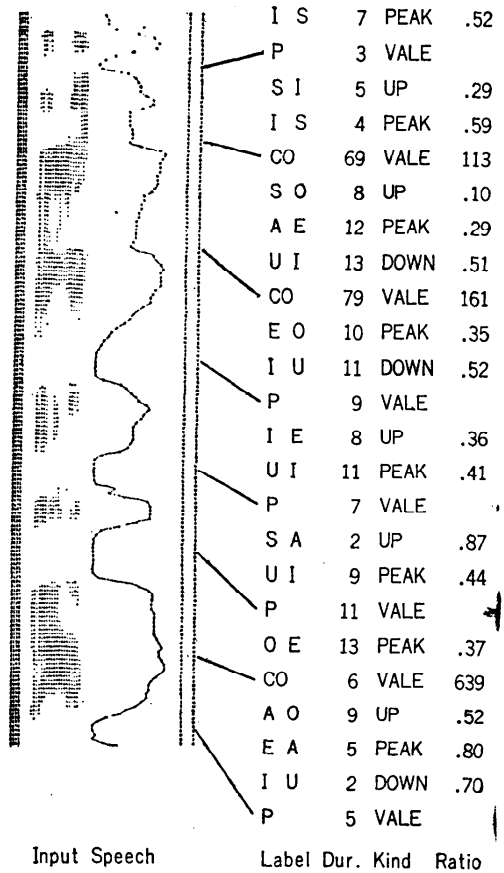


Fig. 2 An Example of Output of the Acoustic Processor.

4. 単語識別部

4.1 単語辞書

辞書は Table 1 に示した各単語に対して1つのエントリーをもっている. それは Fig. 3 に示すようなグラフ構造で記述されている. グラフの節点から出ている枝には各々枝番号が付いており, 更に次の項目が記されている.

(1) ラベルとペナルティ: ラベルはその単語に現われることが期待されている音素を記述するもので,

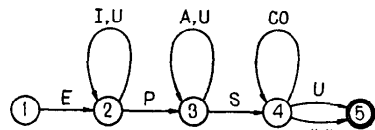


Fig. 3 An example of word graph. This represents word 'X'.

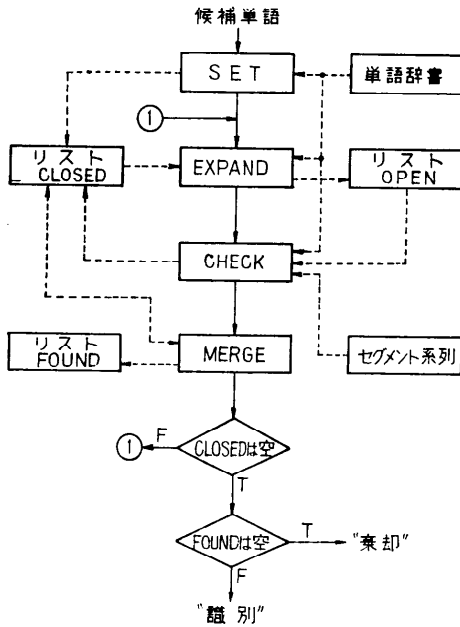


Fig. 4 Flow of word matching.

音響処理部でセグメントに割当てられる音素記号と同一の記号、及び音素の脱落を示す記号「**」が使用される。ペナルティはそのラベルに対する損失値で、音素の脱落、挿入、置換などの変形に対して重み付けを行うものである。1つの枝にこのラベルとペナルティの対を複数個記すことができる。

(2) ラベルの継続時間の許容範囲

(3) 各種の検査項目: ラベルの照合以外に入力セグメント系列に対して行う種々の検査項目を符号化して記す。振幅包絡線の形状、調音結合の可能性の有無、語頭、語尾の特殊性、長母音の途中にみられる弱い谷の存在などを指定している。

(4) 継続節点の番号: 上記の諸条件が満たされたとき遷移すべき節点の番号。

4.2 識別アルゴリズム

単語の識別は単語辞書の各エントリを1種のアクセプタとみて、これをシミュレートすることによって行う。言語処理部の予測した各候補単語に対して以下に述べるアルゴリズムを適用する。その過程を Fig. 4 の概念図に沿って説明する。

(1) SET: パーザが直前に選択した単語 (5.2 参照) の終端のセグメント系列上での位置 (一般にある範囲もっている) は既知であるとする。単語の照合はその次のセグメントから開始する。まず CLOSED

と称するリストに4つの変数の組 (NODE, START, CURRENT, PENALTY) を照合開始点の数に等しいだけおく。ここで NODE=単語グラフの始点の節点番号, START=CURRENT=照合開始点のセグメント番号, PENALTY=ペナルティの累積値=0.

(2) EXPAND: CLOSED の中で CURRENT の値の最少のものを取り出し、節点 NODE から出ている各枝に対して4字組の値を、NODE=枝番号と変更して、これを OPEN と称するリストに入れる。

(3) CHECK: OPEN にある各4字組に対して次のような照合操作を行う。各枝番号に対応するラベルと入力系列との照合を CURRENT で指定されたセグメントから始める。照合がとれない場合はセグメント長の累計がある閾値を越えない限り、時間長に比例するペナルティを加算しながら進んでいく。

ラベルが P, または B の場合はそのまま対応付けを行い、継続時間をも調べる。ラベルが CO の場合は谷の深さの他に指定があれば谷の部分の振幅値も調べる。語頭語尾では CO を P と読み替えることを許している。ラベルが母音、または S の場合は入力セグメントの「RATIO」の値により次の3通りの場合が生ずる。

- (a) $RATIO \geq 0.8$ の場合: 辞書のラベルが入力セグメントの第1, または第2候補音素のいずれかに一致すれば照合がとれたとする。
- (b) $0.8 > RATIO > 0.6$ の場合: ラベルが第1候補音素と一致しない場合は一定のペナルティを付けて第2候補音素と照合する。
- (c) $RATIO \leq 0.6$ の場合: 第1候補音素とのみ照合を行う。

照合は (b) の場合を間にはさんでもよい緩やかな条件と、はさんではいけない強い条件の2通りで行い、両者の終端セグメントが異なる場合は2通りの照合結果があったものとみなす。例えば、Fig. 5 のような場合、前者ではセグメント番号4まで、後者ではセグメント番号2までが照合されることになる。なお上記のいずれの場合にも必要に応じてラベルの継続時間、

照合すべきラベル: A				
セグメント番号	候補音素	RATIO	照合条件	
	1	2	緩	
1	E	A	0.90	○ ○
2	A	E	0.55	○ ○
3	O	A	0.75	○ ×
4	A	U	0.48	○ ×
⋮	⋮	⋮	⋮	

Fig. 5 An example of label matching.

振幅包絡線の属性を調べる。

以上の操作でラベルの照合がとれた4字組は、PENALTYの累積値が閾値以下ならば、NODE=継続節点の番号、CURRENT=照合された終端セグメントの番号、と変更してこれをCLOSEDに加える。1つの4字組に対して2通りの照合結果のある場合は双方をCLOSEDにいれる。

(4) MERGE: CLOSEDの中でNODEとCURRENTの各々が一致する4字組があればPENALTYの大きい方を消去して最適化した後、NODEが最終節点を示すものはCLOSEDよりFOUNDと称するリストに移す。CLOSEDが空ならば考えている単語についてすべての処理を終えたことになり、空でなければEXPANDにもどる。

最後にFOUNDを調べ空でなければ考えている単語が識別されたことになる。そのときのPENALTYの値と言語処理部がその単語に与えた優先順位(5.4参照)とから識別得点を計算し、これをセグメントの終端と共に単語に付して構文処理部に送る。FOUNDの中に4字組がなければその候補単語は棄却されたことになる。言語処理部が予測した候補単語のすべてが棄却された場合は「識別単語なし」という情報を送る。

5. 言語処理部

5.1 「SPOKEN BASIC 1」の構文法

我々は「SPOKEN BASIC 1」の構文法をFig. 6に示すようなstate transition network(以下STNと略記する)の形で与えている。状態間を結ぶ枝に付された記号はSC(2.参照)、またはsubnetwork(以下SBと略記する)を表わす。前者の場合はその枝を通る状態遷移の際にそのSCに属する単語が生成されることを示し、後者の場合はそのSBに遷移することを示している。SBにおける状態遷移は最終状態(図において「〃」を付して示した)だけでなく準最終状態(「'」を付して示した)においても完了できるものとする。

「SPOKEN BASIC 1」の文はmain networkの初期状態「1」から最終状態「37」に至る任意の状態遷移によって生成される。文法全体のSTNは他に譲る^{11)~13)}。

5.2 パーザ

パーザは単語識別部より送られてくる識別単語をパーズングトリー(単語をトリーの節点に対応させる)の形で記憶していくと共に、パーズングトリーの先端

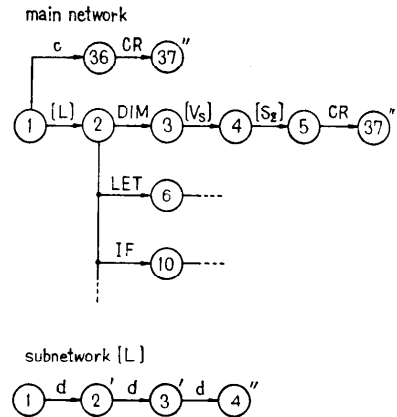


Fig. 6 A part of state transition networks.

の節点を一定の基準で選択して、それをSTNの状態と対応付け、これらの情報を単語予測部、及び意味処理部に送る。いまパーズングトリー上の節点のうち継続節点をもたないものを「開いた節点」ということにする。パーザの動作は次のように述べることができる。

- (1) まだ何も入力されていない最初の状態をスタート節点 n とする。これはもちろん開いた節点である。
- (2) 開いた節点の1つをある基準で選択する。これを節点 n と呼ぶ。この選択基準については後に述べる。
- (3) 節点 n に対応する単語の属するSCによってSTNの状態遷移を行い、STNの状態と節点 n とを対応付ける(n のときはSTNの初期状態に対応させる)。このとき遷移先がFig. 6のSTNの最終状態「37」であれば、 n から n に至る経路上の単語系列を1つの文として認識する。
- (4) (3)で得られた情報を節点 n と共に単語予測部、及び意味処理部へ送る。
- (5) 単語識別部から送られてきた識別単語を節点 n の継続節点としてつなぎ(2)へ行く。このとき単語識別部の出力が「識別単語なし」であれば、節点 n にそれが開いた節点でなくなった旨の印を付けて(2)へ行く。

開いた節点の選択基準を述べるに先立ち、経路のコストと節点の深さについて説明する。節点 n' から継続節点 n に至る枝に、 n に対応する単語の識別得点によって決るコスト $C(n', n)$ を与える。現在使用しているのは識別得点を適当に量子化した絶対コストとそ

の相対順位のみを用いる相対コストの2種類である。スタート節点 n_s から節点 n に至る経路のコスト $g(n)$ を n_s から n に至る経路上の枝に対するコストの和として定義する。すなわち、

$$g(n) = g(n') + C(n', n) \quad \text{但し } g(n_s) = 0$$

である。また節点の深さ $d(n)$ を

$$d(n) = d(n') + 1 \quad \text{但し } d(n_s) = 0$$

と定義する。

パーズングトリー上の開いた節点は g 、及び d の値と関連付けて記憶してある。開いた節点の選択基準は g 、及び d を用いて次のように述べる事ができる。

(1) breadth-first 的な選択を行う場合：最小の g をもつ節点のうちで d が最小のものを選ぶ。

(2) depth-first 的な選択を行う場合：最大の d をもつ節点のうちで g が最小のものを選ぶ。なお純粋な depth-first 型のトリー探索をするときは枝のコストとして相対コストを用いる。

本システムでは上記の典型的な2つの方式の他に、 g 及び d に適当な範囲を与えることにより種々のトリー探索が可能である。

5.3 単語予測部

単語予測部ではパーザから送られてきた節点 n に構文的に続き得る SC を予測する。節点 n には対応する STN の状態が付されているから、その状態から出ている枝を調べれば SC が決定できる。各枝には SC か SB かが付記されている。それが SC であればそれを予測し、SB であればその SB の初期状態に仮の遷移をしてそこから出ている枝を調べる。また現在の状態が SB の準最終状態であれば、その SB 内の枝による予測を終えたのちその SB に遷移してきたもの STN にもどって同様の操作を続ける。

5.4 意味分析と単語の限定

意味処理部は意味分析部と単語限定部の2つから成っている。プログラミング言語の場合、自然言語の場合のような単語ないし文章の「意味」を想定することはできないし、プログラムの内容に立入ってその意味を扱うことも困難である。ここでいう意味分析とは、既に入力されている幾つかのステートメントを調べることによってはじめてわかる規則性とか単語の限定作用とかを明らかにして、そのために必要な情報を蓄積していくことを指している。

意味分析部では、(1)ステートメントに付けられた行番号、(2)キーワードの出現順序、(3)変数の定義などの情報に注目して知識を蓄えていく。

単語限定部では意味分析部に蓄積されている種々の知識を用いて、構文処理部で予測された SC に属する単語をその生起の可能性の度合により4つのクラスに分類する。クラス間には優先順位を付け上位3クラスだけを単語識別部へ送る。この優先順位は単語の識別得点の算定に際してある種のボーナスとして使用される。意味処理部の詳細は他に譲る^{12), 13)}。

6. 認識実験の結果

上に述べた音声認識システムの動作を調べるため、下記のような条件で認識実験を行った。

(1) 音声資料：4名の成人男子が「SPOKEN BASIC 1」で書かれた1つのプログラム(ステートメント数36, 総単語数372, 詳しくは文献14)参照)を1文章ずつ区切って普通の速さで読み、簡易無響室で録音した。

(2) 線形識別関数：音響処理部で使用する線形識別関数の重みベクトルは、上記4名の話者を含む5名の成人男子が、システムで許される50個の単語を単独に発声したものの定常部から、5母音及び摩擦音の各クラスごとに125個のセクションを切り出し、よく知られた学習アルゴリズムによって決定した。

(3) 単語辞書：上記資料中のある1名の話者の音素記号列を参考にして各単語の単語グラフを構成したのち、各話者の音声資料がほぼ一様に認識されるように調整して決定した。従って重みベクトル、単語辞書は各話者に共通である。

(4) トリー探索法：構文処理部におけるトリー探索には depth-first 型の探索法を用いた。

認識実験の結果を Table 2 に示す。ここで話者 NS の資料には録音時の不備により許容し難い雑音を含むものがあったのでこれを除いている。システムに入力された142個の文章のうち、1つの文章として最後までパーズングが行われたものが135個で、他は文章の途中で正しい単語が識別できずに棄却された。最後までパーズングが行われたものうち116個(全体の81.7%)が正しく認識されており、1単語だけ誤ったものは15個(10.6%)であった。

Table 2 Sentence recognition rate

話者	NY	KB	AS	NS	合計 (%)
入力文章数	36	36	36	34	142
正解	32	28	30	26	116 (81.7)
1単語誤り	4	5	3	3	15 (10.6)
2単語以上の誤り	0	2	2	0	4 (2.8)
棄却	0	1	1	5	7 (4.9)

複数人の話者に対して文章認識率80%以上というのは他の代表的な音声理解システム^{10),15)}に比しても遜色はないが、本来の目標であるプログラムの音声入力システムとしては、例えば会話形式の使用形態を想定し1単語程度の誤りは許すとしても、まだ十分な認識率とはいえないであろう。

7. むすび

プログラミング言語「BASIC」を若干修正、制限した「SPOKEN BASIC 1」を「タスク」とする音声認識システムの構成と認識実験の結果とについて述べた。複数人の話者に対して81.7%という文章認識率を得た。これはタスク本来の目標であるプログラムの音声入力システムとしてはまだ十分とはいえないが、その目標に一步近づいたことは確かであろう。今後システムの認識率を更に向上させ、より多くの話者に適応させ、あるいは語彙を増していくために考慮すべき点を以下に列挙しておく。

(1) 音響処理: 連続音声の認識を困難にしている主な原因は話者の個人差、あるいは調音結合による音声の多様性である。本システムでは単語辞書をグラフ構造にすることでこれを吸収しているが、特に個人差に関しては音響処理のレベルで何らかの正規化を行うべきであろう。

(2) 単語識別: 辞書の作成は現在人手によって試行錯誤的に行っているがこれを自動化する必要がある。あるいは識別アルゴリズムをより一層柔軟性のあるものとして単語辞書の構造を簡単化すべきである。

(3) 構文処理: 現用のパーズング方式は文頭から文末へ向って一方向的に進み、1か所で正しい単語が見つからないとそこから先へは進めなくなる。任意の場所から前後両方向へ進むことのできるパーズング方式の採用が必要である。

(4) システム構成: 階層構造をなすパタン認識システムでは、各階層の判断の誤りがシステム全体の判断に大きく影響しないように、可能な限り断定的な判断を避け、強力なトリー探索を併用することによって各階層の不確定さを減少させていくのがよいと思われる。この意味で本システムでも各階層間の情報交換をより密接にしていかなければならない。特に音声の多様性の吸収のためには、音響処理部と単語識別部との間の情報交換が不可欠のようである。

最後に、本研究の遂行に当り音声データのデジタル化に便宜をはかっていただいた京都大学工学部坂井教授並びに同研究室の音声研究グループの方々に感謝します。なお本システムの開発は京都大学大型計算機センターの TSS システムを用いて行った。

参考文献

- 1) D. B. Fry: Theoretical aspects of mechanical speech recognition. J. Brit. IRE, Vol. 19, No. 4, pp. 211~218 (1959).
- 2) P. Denes: The design and operation of the mechanical speech recognition at University College London. J. Brit. IRE, Vol. 19, No. 4, pp. 219~234 (1959).
- 3) 板橋, 城戸: 辞書と音形規則を利用した単語音声の認識, 音響学会誌, Vol. 27, No. 9, pp. 473~482 (1971).
- 4) G. Fant: Automatic recognition and speech research. STL-QPSR, 1/1970, pp. 16~31 (1970).
- 5) A. Newell, et al.: Final report of a study group on speech understanding system. North Holland (1973).
- 6) Proc. of third international joint conference on artificial intelligence, pp. 173~222 (1973).
- 7) Proc. of IEEE symposium on speech recognition, pp. 1~67 (1974).
- 8) 武谷, 河口: 言語構造情報を利用した連続音声認識システムのシミュレーション. 信学論(A), Vol. 56-A, No. 9, pp. 512~520 (1973).
- 9) 好田他: 音声による質問回答システム, 信学技報 EA 75-59 (1976).
- 10) 坂井, 中川: 音声理解システム LITHAN のシステム評価. 音響学会音声研資 S75-30 (1975).
- 11) 新美, 浅見: 音声認識システムにおける言語情報の利用とその効果. 信学論(D), Vol. 58-D, No. 12, pp. 741~747 (1975).
- 12) Y. Niimi, et al.: The speech recognition system of "SPOKEN BASIC". Proc. 2nd USA-JAPAN computer conference, pp. 375~379 (1975).
- 13) 新美, 浅見: 音声認識システムにおける構文分析, 意味分析の効果. 音響学会音声研資 S73-31 (1974).
- 14) 新美, 小林: 音声認識システムにおける構文分析法の比較. 信学技報 PRL 75-56 (1974).
- 15) J. K. Baker: Stochastic modeling as a means of automatic speech recognition. Doctoral thesis of Carnegie-Mellon Univ. (1975).

(昭和51年2月23日受付)

(昭和51年4月24日再受付)