論　文

フォートラン・プログラムの音声認識システム*

関口　芳廣**　大輪　一**　青木　憲一**　重永　実**

Abstract

Speech recognition system of programs in JIS Fortran 3000 level uttered with a rather longer pause between successive words has been constructed. It consists of an acoustic analyser and a linguistic processor, and has an automatic error correcting function. In linguistic processor two types of lexicon are prepared. One of them is concerning to gross features. The other consists of the phoneme string of each word (total 79 words), which is essentially spelled in Roman letters and also some rewriting rules of phonemes are contained in each word. By syntatic and semantic analyses and first type of lexicon, some possible input words are predicted. The phoneme sequences of predicted words are matched with the input one and the word having the highest reliability is adopted as the output. Each statement of programs is syntactically examined and is automatically corrected whenever there is any syntactic error. If some confusions among English letters are allowed, we can recognize any program.

1. はじめに

音声認識における動向は，単語の認識のみを対象とするもの[1)~6)]を除けば，三つに大別される．いずれも音響的な処理から，構文，意味，プラグマティクス情報の利用まで，多層のレベルを有機的に結合してゆく必要があるが，その一つは，対象を狭い範囲に限定し，かつ構文も簡単ではあるものの，自然言語の単文を対象として，その中に含まれるキー・ワードを抽出，認識，判断して，音声理解のシステムを完成させようとするもので，カーネギーメロン大学を中心とする研究グループで始められた[7),8)]．わが国でも計算機との会話システムを完成させようとするもの[9)]や，列車の予約システムの構成に成果を上げているもの[10)]などで代表される．これらの場合，現在扱われている対象としては単語数 65 個ないし 100 個で，文節ないしは単文を扱っている．第二のものは自然言語の認識を目ざすものであるが，現在のところ一般にはむずかしく，依存文法に従う単文に制約することにより成果を上げているものがある[11)]．

他の一つは，形式言語ではあるが，言語全体を対象とするもので，この方面では，認識音素は母音と /s/ だけで，使用可能変数を X, Y, Z の 3 英字と数字との組み合わせに制限しているものの，連続的に発声した BASIC 言語を対象にし，構文，意味情報を駆使し成果を上げているもの[12)]がある．このように言語全体を対象とする場合には，意味ないしはプラグマティクス情報は文の中だけではなく，文相互間でも作用し合うから，現時点では，他の分野におけるものよりも意味情報等の利用の範囲が広がっている．

何れを対象にしても，音声認識の究極の目標の一つは，人間と同様に言語を認識し，理解し得ることにある．ただ広く任意の言語を扱おうとすると，文法のあいまいさが増し，構文情報だけで次の単語ないしは句を予測，限定することが困難になってくる．その反面，意味情報は増加してくるのが通例である．しかし，意味情報によって，次の単語を予測，推論して，認識対象を狭い範囲に限定することはかなりむずかしいであろう．従って，認識対象単語の数が大変多くなることは避けられず，単語の認識率を上げることがむずかし

* Speech Recognition System for FORTRAN Program by Yoshihiro SEKIGUCHI, Hajime OOWA, Ken'ichi AOKI and Minoru SHIGENAGA (Faculty of Engineering, Yamanashi University)

** 山梨大学工学部計算機科学科

くなってくる．これを救うには結局，極めて困難ではあるが，音響レベルにおける処理の充実によらざるを得ないであろう．

筆者らは認識システムの拡張性を考えて，音響レベルでの処理を重視し，心理学的なモデルを導入しながら，音素単位の認識を試み，入力音声波をローマ字つづりに，一部音響的特性を加味した音素列で表現することを試みてきた[13]．そして認識対象として，フォートラン言語を取り上げることとした[14]．その理由の主なものは(1)使用単語数，百数十個程度で言語全体が扱える．(2)その単語中には EXPONENTIAL のような長いものがある反面，一部の英字のように単音節からなるものもあり，これらが認識できれば，任意の単語を取り扱い得るようになるであろう．(3)言語全体を扱うからプラグマティクスないしは意味情報をより広範囲にわたって使い得る．このことは自然言語を扱う場合に重要になってくるであろう．(4)構文情報が明確である．(5)実用面からは，コンパイラ言語の音声入力ができれば，タイプする必要がなくなり，人間を機械的な仕事から解放してくれるであろう．

現在までのところ，JIS 3000 レベルのフォートラン・プログラム（単語数：英字26文字を含め79個）を，各単語間に幾分長目のポーズをおいて発声した資料を対象にし，誤り自動修正の機能を付加したシステムを作り上げ，簡単なプログラムに対し，単語誤り率3％以下で動作しうるようになっている．

2. システムの概要

認識システムの概要を **Fig. 1** に示す．システムは音響処理部と認識部とに大別される．音響処理部では入力音声を 10 kHz，符号＋10 ビットでサンプリング，量子化し，10 ミリ秒間を1フレームとして各フレームごとに特徴パラメータを抽出する．この特徴パラメータを使用して候補音素を選定し，音声波形の定常性を考慮しながら音素列を作成する[13]．

認識部では，音響処理部から送られた音素列から単語の大局的な特徴を抽出し，候補単語をおおまかに予測する．一方では先行の認識結果より，言語情報を利用して更に候補単語を限定し，単語辞書とマッチングをとり，最も信頼度の高いものを出力する．

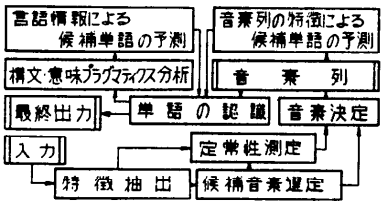

Fig. 1 Block diagram of speech recognition system.

3. 認 識 方 式

次の点を考慮して認識部を構成した．

(1)認識単位は意味のある最少の単位，すなわち単語の単位で行う．(2)単音節からなる単語が存在することや単語数の増加や構文情報の拘束力の減退などに対処し得るようにして，システムの拡張性を計る．そのため音響的処理を重視し，その情報を十分信頼し，活用する．(3)フォートラン言語では FORMAT 文の中などをのぞいては文法が明確であるので，十分な構文解析を行う．(4)単語の増加に備え，単語辞書及び書き換え規則はできるだけ簡単にする．(5)マッチングのための候補単語の数を構文，意味及びプラグマティクス情報の他に，音素列の大局的な特徴を使って，できるだけ少なく制限する．(6)認識単語の最終的決定のための閾値を可変にして，出力単語の信頼性を高める．(7)文の型（Read 文，算術代入文など）を正確に把握するために，出だしの数単語は特に重視する．(8)もし誤っても，できるだけ修正できるよう考慮する．

以上のような考えから **Fig. 2**(次頁参照) に示す方法で認識部を構成している．その概略を述べる．

① まず入力音素列中に5個以上続いている音素と語中の無音部の有無を順序付きで $\{A_i'\}$ として抽出する．これに対しては **Table 1** に示すような特徴音素辞書 $A\{A_i\}$ が作ってあり，入力と同一の特徴音素をもっている単語に点数 a を与える(**Fig. 3** 次頁参照)．

② 入力音素列から雑音部 (UNV)，まさつ音 /s/，無声破裂音（P），語中の無音部(・)の出現場所を抽出して $\{B_i'\}$ とする．これに対して特徴辞書Bが作っ

Table1 Examples of lexicon A.

単 語	(	=
辞 書 C	PA・PO	I・PORU
特徴 \ $\{A_i\}$	A・O	I・OU
i	0000	1000
e	0000	0000
a	1000	0000
o	0010	0010
u	0000	0001
s	0000	0000
・	0100	0100
f	0000	0000

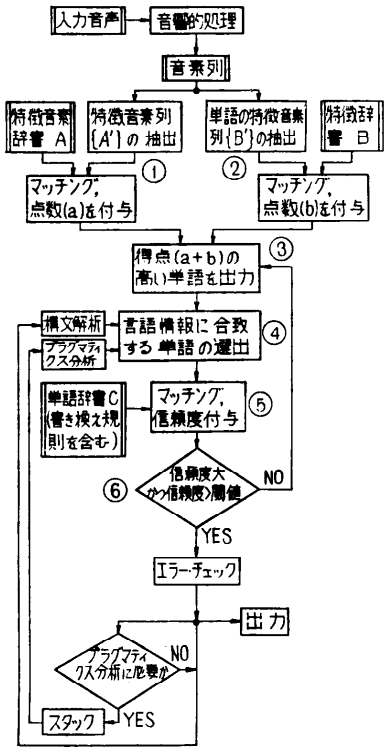

Fig. 2 Flow chart of linguistic processor.

```
UNV     : **                                  **
入力音素列: K==AAAAAAAAAAAAAA..... KK===00000000..
抽出音素列: {A'}=A・0, {B'}=9(P), 1(・), 0(S), 9(UNV), {C'} PA・PO
        P: 無声破裂音, ・: 無音部, =: 過渡部, UNV: 雑音性
```

Fig. 3 Examples of an input phoneme string and its {A'}, {B'} and {C'} for '(:/kakko/'.

Table 2 Examples of lexicon B.

特徴 \ 単語	(	=
	PA・PO	I・PORU
P	9	1
・	1	1
S	0	0
UNV	9	1

(例) PA・PO
↑ ↑
8 + 1 = 9

てあり，これとマッチングを行って，これらの特徴をもっている単語にその一致の数に応じて点数 b を与える．**Table 2** に辞書 B の例を示す．③ 得点$(a+b)$が多い単語を選び，言語情報解析部へ送る．④ 候補単語を言語情報に合致するものだけに制限する．⑤ 入力音素列と単語辞書 C $\{C_i\}$ (C_i：音素) 中の候補の単語とのマッチングを取り，二種類の信頼度を与える．⑥ 上記の信頼度が最大で，かつその値がある閾値を起えているものを最終出力として出す．もし信頼度が閾値より小さい場合には③にもどり，次の得点の単語を選び出し，同様な過程で認識を行う，また，最終出力単語がプラグマティクス情報として必要ならば，スタックに蓄わえ，以後の単語予測に利用する．なお最終出力が適当か否かのチェックも行い，不適当な場合には修正する．

Table 3 Examples of lexicon C.

単語	発音	辞書	書き換え規則
A	エ イ	PEI	P→φ
1	イ チ	I・PSI	I→φ
+	プ ラ ス	PURASU	R→φ, U→φ
FORMAT	ケイシキ	PEIS・PI	P→φ, E→φ
COS	コサイン	POSAEIN	P→φ, E→φ, N→U

注．辞書中のPは無声破裂音，Nは鼻子音（撥音を含む），
・は無音部を示す．

る．なお単語辞書C[14]は **Table 3** に示すように，通常，発音をローマ字式に書き，一部音響的特徴を加味して納めてあるが，まちがい易い音素や現われないことがある音素，あるいは綴にはないが現われる恐れのある音素に対して，各単語ごとにその音素を書き換えてもよい事を示す書き換え規則を付してある．この規則は通常1単語当り 1～2 個ですんでいるが，最高3個までに限定している．このようにすることにより，調音結合や発声上のなまけに対処すると同時に，書き換え規則を共通に作ることによる混乱をさけている．

4. 候補単語の予測

単語数が多くなると，認識に先立って，候補単語を限定することが，マッチングの回数の軽減，認識率の向上のために必要である．筆者らは音響的特徴の大局的な情報を抽出して，この大局的特徴により，まず候補単語を予測し，その中から，言語情報を満足するものを選択するようにしており，認識機構のモデルの構成上，重要な位置を占める部分である．

4.1 音響的特徴による予測

(a) 特徴音素列の抽出とその辞書

入力音声は 10 ミリ秒ごとに音素に変換され，その音素列 $\{X_{m,n}\}$ が認識部へおくられる．いま仮に入力音素列が

$X_{1,1}\cdots X_{1,s1}$, $X_{2,j+1}\cdots X_{2,j+s2}$, $X_{3,k+s3}$, $X_{4,l+1}\cdots$

ただし $j=s1$, $k=j+s2$, $l=k+s3$, $\cdots$

$X_1, X_2, X_3\cdots$ はそれぞれ隣同士は異なった音素

とすると，$\{s_i\}=(s1, s2, s3, \cdots)$ に着目して，特徴音

素列 $\{A_i'\}$ を構成する (Fig. 3 参照). ここで

$$A_i' = \begin{cases} X_i & : s_i \geqq 5 \text{ かつ } X_i \in (i, e, a, o, u, s, f, \cdot) \\ \phi(\text{空}) & : s_i < 5 \text{ または } X_i \notin (i, e, a, o, u, s, f, \cdot) \end{cases}$$

すなわち, A_i' として /i, e, a, o, u, s, f/・(無音) の8種類の音素 (語中の無音部も音素と同様に扱う) を採用している. これらのうち /f/ は英字 F の認識のために, 特別に加えたものである.

このようにして作られた特徴音素列 $\{A_i'\}$ に対して特徴音素辞書Aが用意してある. 現在までのところ特徴音素としては1単語あたり最初からの4個を採用すれば十分であるので, 各単語に対し, 特徴音素ごとに4ビットの辞書を用意している. 特徴音素の数は8個あるので, 1つの単語に対し4バイトの記憶容量が必要である.

いま特徴音素 z の i ビット目を $[a_i]_z$, 特徴音素列を $\{A_i\}$ で表わすと

$$[a_i]_z = \begin{cases} 1 & A_i = z \\ 0 & A_i \neq z \end{cases}$$

ただし $i=1,2,3,4$

となる. たとえば, 単語 ' (', 単語辞書: PA・PO の特徴音素列は $\{A_i\}=\{A \cdot O\}$ となり

$$[a_1]_A=1,\ [a_2]_{\cdot}=1,\ [a_3]_O=1$$

であり, あとはすべて0となる. そこで Table 1 のような辞書ができる.

次に入力の特徴音素列 $\{A_i'\}$ に対応する各特徴も同様に表わし, 辞書Aと較べ, 入力特徴音素列と同じ特徴音素辞書を持つ単語に, 点数 $a=1$ を与える.

(b) 音響的特徴系列の抽出とその辞書

一方入力音素列から雑音部(UNV), まさつ音(/s/), 無声破裂音(Pで表わす), 語中の無音部(・)の出現場所を抽出し, 単語予測に利用している. 前述の特徴音素列 $\{A_i\}$ が比較的定常的な母音などを中心に考えられているのに対し, これらの特徴は人間の認識において, 重要な役割をするといわれている子音に関するものである.

入力音素列 $\{X_{m,n}'\}$ から $\{A_i'\}$ を作ったのと類似の方法で, $\{X_m\}$, $\{u_n'\}$ (UNV の列を縮小したもので, X_n に UNV がついているときのみ $u_n'=1$) から (P, ・, s, UNV) の存在とその位置を調べ, 特徴系列 $\{B_i'\}$ $(i=1,\cdots,4)$ を作る. 各特徴 z に対してそれぞれ4ビットを使って, B_i' を数値 b_z' で表わす. ここに

$$b_z' = 8b_{1z}' + b_{2z}',\quad b_{1z}' = \begin{cases} 1 : X_1 = z \\ 0 : X_1 \neq z \end{cases} \text{ または } \begin{cases} u_1' = 1 \\ u_1' \neq 1 \end{cases}$$

b_{2z}': $X_n' = z$ または $u_n' = 1 (n = 2, 3, \cdots)$ を満足する n の数

すなわち, 特徴が語頭にあるときは $\{b_i'\}_z (i=1,\cdots 4)$ の1ビット目 (MSD) b_1' を1にし, その他の場所にあるときは $\{b_i'\}_z$ に1ずつ加算する (Table 2 参照).

この $\{B_i'\}$ に対する辞書Bも単語辞書Cを参照して $\{B_i'\}$ と同じ形に容易に作りうる. そして両者のマッチングをとって, 各ビットの一致した数に応じて, 各単語に点数 b を与える.

(c) 候補単語の選定

以上のような方法で得られた得点 a, b の和 $a+b$ の大きなものから順に, 音響的特徴による候補単語となる. すなわち, 入力音素列より, まず入力単語の大局的な特徴を取り出して, 候補単語を限定しようと試みている. まず言語情報を使って候補単語を限定してから, この $a+b$ の情報を使うことも考えられるが, 後述の誤り修正の機能を組み込む際, ここで行っている順序にして, 冗長性を持たせることが大変役立つ.

4.2 言語情報による予測

(a) 構 文 情 報

フォートラン言語は自然言語などに比較して, 文の生成規則が明確であるため, 単語の予測に構文情報を利用することは有効な手段となりうる[12]. そこで, フォートラン言語は基本的には有限オートマトンで表現しうるという仮定のもとに, 構文情報を算術代入文をも含めて状態遷移図で表現している[12)~14)]. その一例を **Fig. 4** に示す. フォートラン言語では構文情報が比較的明確であるといっても, その量的, 質的複雑さはかなり大である. そこで構文情報を構成, 使用する際, その表現が簡単であること, 記憶容量が少なくてすむこと, 検索が速いこと, 修正が容易なことなどを考え, 次のような方法をとっている. 認識対象の直前の単語 w, それ以前の先行単語より得られる特性 $\boldsymbol{n}$ より, **Fig. 5** (次頁参照) に示すように次の候補の単語

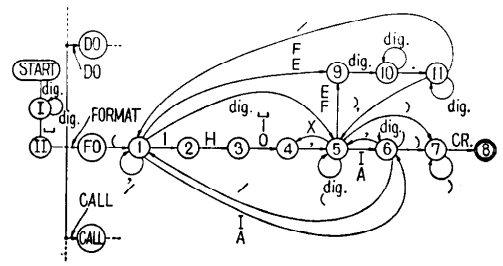

Fig. 4 Example of state transition network for Format statement.

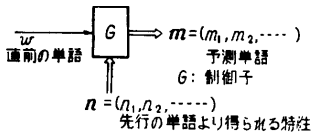

Fig. 5 Standard type of node.

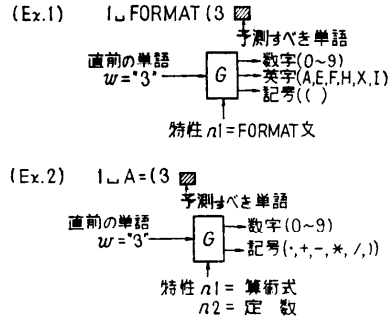

Fig. 6 Examples of standard type of node.

群$\boldsymbol{m}$を予測するという標準型を構成している．Fig. 5で G は数字，英字，関数などのグループの制御子である．このような標準型にすることにより，表現も統一され，簡単で，修正も容易である．Fig. 5 の標準型は，たとえば **Fig. 6** のような単語列が認識されてきた場合，その次の単語の予測は Fig. 6 のように限定される．制御子 G は直前の単語より選択され，特性 $\boldsymbol{n}$ は直前の単語も含めた先行の単語系列より決定されるものである．たとえば Fig. 6 の Ex. 2 では "$A=$" より n_1=算術式，"(3" より n_2=定数という特性が得られる．これは状態遷移図では **Fig. 7** のようになることであり，Fig. 7 における状態ⓐ，ⓑは同じ数字 "3" からの遷移であるが，次に予測する単語は前の経歴でそれぞれ異なる．そこで，それぞれに適したグループ群を生成するルーチンを用意せねばならないが，Fig. 5 のように標準型にすることにより，各グループを生成するルーチンを用意して共通に使用することができる．

（b） プラグマティクス情報

プラグマティクスや意味情報の導入はシステムの拡張性を考える場合，最も重要なことである．しかしコンパイラ言語を扱う場合には，意味情報と言いうる程のものを使用しうることは希で，通常構文的なものか，あるいはプラグマティクスと呼ばれるべきものであろう．本システムでは下記のような事例をプラグマティクス情報の利用ということにし，大きくわけて二つの場合を扱っている．一つは文単位の情報であり，それが WRITE 文とか IF 文であるということを理解することである．他の一つはプログラム全体からわかる情報で，この変数はディメンションとして定義されているとか，すでに数値が格納されているとかいうことである．

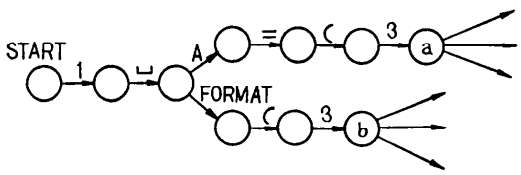

Fig. 7 Example of transition network.

最初の文単位の情報は前述の構文解析における Fig. 5 と密接に関連しており，その文がどのような型の文であるかを知ることは，以後の構文解析を信頼できるものとしてくれる．またプログラム全体からの情報としては，算術代入文の左辺の変数や READ 文の変数などをスタックしておき，後で候補単語の限定に利用する．

5. 単語の認識

音響的特徴や言語情報により予測された候補単語の辞書と，簡単化された入力音素列 $\{X_m\}$ とのマッチングをダイナミック・プログラミング的にとり，各候補単語に対して次の二つの信頼度 S_i, S_l を計算する．

$$S_i=m/n_i,\quad S_l=m/n_l$$

ただし m は入力と辞書との間でマッチングがとれた音素の数で，n_i, n_l はそれぞれ $\{X_m\}$ の音素数と辞書の音素数である．そして，まず S_i の大きい順にならべ，S_i が等しい場合には S_l の大きい順にする．もし S_i, S_l とも等しい場合には書き換え規則適用回数の少ない順にならべる．そして，S_i, S_l が閾値を上まわっておれば，その最大のものを出力として出す．もし候補単語の中に該当するものがなければ，$a+b$ の音響的な特徴の一致を一段ゆるめ，候補単語のわくを広げる．しかしこの場合には閾値を高くし，より正確に単語辞書と一致することを要求している．

6. 誤りの自動修正機能について

以上のような方法で認識を行っても，いろいろな誤りが発生し得る．そして特殊な誤りが生じた場合には，それが以後の構文解析に影響し，認識を狂わせる．この誤りを早期に発見し，誤った部分を修正し，以後の認識に影響を与えないようにする必要がある．そこで第一の方法はまず音響情報のみで "CR"(改行)

の単語が出現するまで認識を行い，その結果の単語列を $\{O_i\}$ とする．("CR" の誤認識はほとんどない)．そして文の最初までもどり，言語情報を使用した認識を行い，$\{O_i\}$ に対応するこの結果を $\{P_i\}$ とする．また最終的な出力単語列を $\{Q_i\}$ とする．

まず仮に $\{Q_i\}=\{O_i\}$ としておき，$\{O_i\}$, $\{P_i\}$ の単語系列より最終的な $\{Q_i\}$ を決定していく．まず判別関数 $g_Q(n)$, $g_P(n)$ を定義する．k を文の単語数とすると，$g_Q(n)$ は単語系列 $\{O_i\}$, $\{Q_i\}$ $(1\leqq i\leqq k)$ に対して

$$g_Q(n)=\sum_{i=1}^{k-1} e_i.\quad \text{ただし } Q_n=O_n,\ (1\leqq n\leqq k)$$

$$e_i=\begin{cases}0: Q_iQ_{i+1} \text{ というならびが許される.}\\ 1: Q_iQ_{i+1} \text{ というならびが許されない.}\end{cases}$$

すなわち，$g_Q(n)$ は $\{Q_i\}$ の n 番目の単語 Q_n を $\{O_i\}$ の n 番目の単語 O_n でおきかえた場合，許されないならびが $\{Q_i\}$ の系列に幾つ存在するかを示している．同様にして，単語系列 $\{P_i\}$, $\{Q_i\}$ から判別関数 $g_P(n)$ を定義する．これらの判別関数を使用して，次のような方法で最終的な $\{Q_i\}$ を決定する．

$$Q_i=\begin{cases}O_i: g_Q(i)\leqq g_P(i)\\ P_i: g_Q(i)>g_P(i)\end{cases}$$

この方法により誤りをどの程度修正できるかを調べた例を **Fig. 8** に示す．$\{P_i\}$ でCを IFIX と誤ったため（雑音が原因）以後混乱をきたしたが，$\{Q_i\}$ ではかなり修正されていることがわかる．

もう一つの方法は，**4.2 (b)** に記した文の型の誤認識を修正せんとするものである．ここでは 18 種類の文型に分類している．各文型はキーワード列 $\{K_i\}$，キーワードの個数 m，その文らしさを強調するプラス情報 P，その文を否定するマイナス情報 M で表現されている．その例を **Table 4** に示す．まず文を音響情報のみにより認識し，その単語列を $\{O_i\}$ とする．この $\{O_i\}$ の中に含まれる各文型 C のキーワードを求め，特徴単語列 $\{F_i\}_c$ を作る．すなわち

$$F_{ic}=\begin{cases}O_i: O_i\in(K_1, K_2, \cdots, K_n)_c\\ \phi(\text{空}): O_i\notin(K_1, K_2, \cdots, K_m)_c\end{cases}$$

$\{F_i\}_c$ と $\{K_i\}_c$ とを較べて，一致した数 r を求める．

$\{O_i\}$: ␣␣X = A*((B IABS K)/G+H**3*IFIX(I)) (CR)
$\{P_i\}$: ␣␣X = A*((2−IFIX(COS(ABS(3.3*IFIX(SIN(0 (CR)
$\{Q_i\}$: ␣␣X = A*((B − K)/ G+H**3*IFIX(I)) (CR)
正答: ␣␣X = A*((B − G)/ G+H**3*FLOAT(I)) (CR)
$\{O_i\}$音響情報のみ使用 $\{P_i\}$言語情報を加味 $\{Q_i\}$最終出力

Fig. 8 Example of result obtained by speech recognition system with automatic error correcting function.

Table 4 Examples of statement type.

文 型	キーワード列	キーワードの個数	プラス情報	マイナス情報
算術代入文	A=	2	+−*/の個数	
DIMENSION 文	DIMENSION A ( )	4		
DO 文	DO n A = ,	5		+−*/の個数
FORMAT 文	n FORMAT ( )	4		
READ 文	READ (n)	4		( )内に文字
WRITE 文	WRITE (n)	4		( )内に文字

(注) n: 数字　A: 文字

また各文型ごとに P, M を求め，文型 C の単語列 $\{O_i\}$ に対する得点 T_c を次式により計算する．

$$T_c=(r+P-M)/m$$

そして T_c の最大の文型 C を $\{O_i\}$ の文型とみなす．

例えば '␣READ (5, 6) A, B…' 中の READ の後半が不明瞭で，B と誤認識されたとすると，A を認識する段階で文法に合わなくなるから，システムは以後暴走することになる．そこで音響情報のみにより最初から認識しなおし，'␣B (5, 6) A, …' なる $\{O_i\}$ を得たとする．すると READ 文と WRITE 文の T が 0.75 の最高点を示し，B は READ または WRITE であるべきことがわかる．なお READ 文か WRITE 文かはプラグマティクス情報により変数 A 以下の性質を調べる必要がある．ただしこの単純な方法では，誤りの個処あるいは誤り方によっては同じ T_c の値のものが複数個現われたり，正しく修正しえない場合があり得るので，修正したあと，文法にあっているかどうかを調べ，もし文法に合わない場合には次の T_c の値の文型に対して試みる必要がある．

以上二つの方法は，本システムでは音響情報のみによって認識しても，単語の誤認識率が，英字の間を除けば，15% を越すことはほとんどないという実験結果により可能となっている．

7. 実験と結果

文法は JIS 3000 に準拠しているが，注釈行，継続行及びフォーマット文中に入れられる任意の文字を除いている．また文の最後には '改行'，カード上の第6桁では 'スペイス' と発声してもらい，構文解析を楽にしている．資料は成人男子1名が静かな部屋で単語間に長目のポーズを入れて発声した，合わせて 22 の文からなり，199 単語（スペイス，改行を入れると 243 単語）が含まれている2つのプログラムについて，誤り修正の機能を働かせないで行った結果を示す．文は

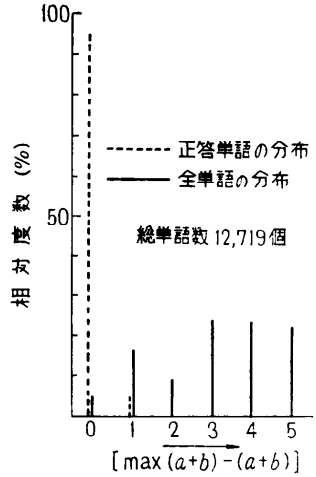

Fig. 9 Distribution of number of words which obtained points $(a+b)$.

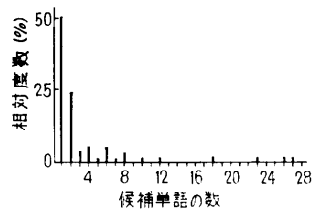

Fig. 10 Distribution of number of final candidate words.

すべて正常に終了した．識別できない P, T; B, D; M, N 間の誤りを除けば，5単語が誤認識された．その内訳は雑音の混入によるもの3，発音の曖昧さに起因するもの 1，前の誤りによるプラグマティクス分析の結果に縛られて誤ったもの1となっていた．

音響的特徴だけで単語の認識を行った場合，正答の単語が得点 $a+b$ のどの範囲に入るかを調べて **Fig. 9** に示す．正答単語の 95% が $a+b$ の最大値をとり，最大値より2点以上低い正答単語はなく，また全単語のうち5% が $a+b$ の最大値をとることがわかる．さらに $a+b$ の得点に言語情報を加味した場合，最終的な候補となった単語の数を調べて **Fig. 10** に示す．最終的に候補として残る単語の数がただ1個だけの場合が認識回数の半数を占めていることがわかる．

8. むすび

本論文では単語単位に発声された JIS 3000 レベルのフォートラン・プログラムを対象にとりあげ，その音声認識システムを構成し，特に単語認識部の手法と，誤り修正の方法について述べると共に，いくつかの例について実験を行い，その結果の例を示した．本システムの運用とその結果より次のことがいえる．①音響処理部で音素単位の認識を行っており，それを簡略化した音素列にまとめ以後の処理に使っている．そのため，発声の速さの変化に対処するための時間の正規化の問題について，余り考えなくてよい．すなわち，ある程度発声の速さを変えても十分認識し得る．②単語の認識に当り候補単語を予測しているが，その際，言語情報と共に大局的な音響的特徴を使用している．特に音響的な情報は最終的なマッチングの回数を十分少なくすると共に，誤りの個所を発見するのに役立っている．③システムの構成に当り言語情報が規格化されており，単語辞書もローマ字式に近い表現で，少ない記憶容量ですむなどのことにより，システムの拡張性は十分あると思われる．④問題点はまだ子音の識別が不十分で，特に単音節からなる単語，たとえば，B, D; M, N; P, T 間などの識別ができないことである．破裂音[15]，鼻子音[16]の各グループ内の識別もかなりできるようになっているが，まだこのシステムに組み込むには至っていない．今後以上の問題点の解決とシステムのレベルの向上に努める予定である．

参考文献

1) 中津，好田：VCV 音節を単位とした連続単語音声の認識，信学研資，PRL 75-44 (1975)
2) 追江：2段 DP マッチングによる連続単語認識，音声研資，S75-28 (1975)
3) 佐藤，藤崎：限定語彙単語音声認識の一方式，音声研資，S75-27 (1975)
4) 三輪，牧野，松岡，城戸：オンライン単語音声自動認識装置，音声研資，S75-26 (1975)
5) 坂井，中川，林：音韻スペクトルの個人差の予備学習による限定語彙単語音声の認識，信学研資 EA 75-61 (1976)
6) G. M. WRITE: Speech Recognition Experiments with Linear Prediction Bandpass Filtering And Dynamic Programming: 第2回日米コンピュータ会議論文集，(1975)
7) A. Newell, et al,; Speech Understanding System, Final Report. (1971)
8) Special Issue on IEEE Symposium on Speech Recognition: IEEE Trans. on ASSP, Vol. ASSP-23, No. 1 (1975)
9) 坂井，中川：タスク内の自動言語音声を理解するシステム，——LITHAN——，信学研資 (1975)
10) 好田他：音声による質問回答システム，信学研資，EA 75-59 (1976)
11) 武谷，田町：依存文法を利用した連続音声認識

の実験，信学研資，PRL 75-43 (1975)
12) 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. 741～749 (1975)
13) W. A. Woods: Transition network grammers for natural language analysis, Comm. ACM, Vol. 13, pp. 591～606 (1970)
14) M. Bates: The Use of Syntax in a Speech Understanding System, IEEE Trans. on ASSP Vol. ASSP-23, pp. 112～117 (1975)
15) 関口，重永：学習と予測を導入した連続音声のセグメンティション，音声研資，S74-26 (1974)
16) 関口，大輪，青木，重永：JIS フォートラン 3000 レベルのプログラムの音声認識，音声研資，S75-29 (1975)
17) 関口，重永：破裂音の識別，音響学会講演論文集，p. 347 (1975年10月)
18) 関口，重永：単語中の鼻子音の識別，音声研資 S75-03 (1975)

(昭和51年2月19日受付)
(昭和51年4月9日再受付)