連続音声入力を想定したキーワード抽出システム†

中村 修†† 川野邊 正††
雪下充輝†† 小黒雅己††

自然に発声する連続音声入力を想定し，この認識結果から各種の操作指示に必要となるキーワードを抽出するアルゴリズムの提案と，実験システムによる評価結果を示す．ここで提案する抽出アルゴリズムは，正解となるキーワードを誤って棄却しないように，音声認識結果（キーワード抽出システムの入力）の音素記号列から，音素置換確率により，1入力音素当り複数の正解音素候補を生成する．さらに，正解音素候補の組合せから音節候補を生成し，これら音節候補の組合せから，存在しうるすべての単語を，木構造インデックスを用いて高速に抽出する．抽出した単語候補については，操作指示に直接必要とならない敬語表現等の慣用句の除去等により予備選択を行い，次いで，操作指示内容に応じて作成した構文パタンを用いて，構文候補を生成する．構文候補に対しては，構文・単語レベルのスコアリングを行い，スコアの高い構文に含まれる単語を順次キーワードとして抽出する．シミュレーションの結果，辞書規模を約1,000単語，操作指示種別を5種，音素の平均認識率を77～89%とした条件において，正しい操作指示内容を示すキーワードの組合せを，60%以上の的中率で抽出できることが分かった．また，スコアの上位3位までを取れば，95%以上の包含率で正解が得られることが分かった．さらに，実行時間の評価から，最大100並列の構成で発声速度に追随可能であることが分かった．

1. まえがき

連続して発声する音声を用いて，システムへ操作指示を与えるには，操作指示に必須となる単語（以降，キーワードと呼ぶ）を正確に抽出する必要がある．しかし，現状では話者が発声した連続音声（単文等）を，実用的な認識率で認識することは期待できず，認識結果からだけでは正しいキーワードを抽出することは困難である．これまで，誤認識結果を含む単語音声に対して，音素置換確率（コンフュージョン・マトリクス）を用いて音素を訂正する方法が幾つか提案されている[1)-3)]．一方，自然言語処理技術として，入力される文の内容を小世界（タスク）に限定し，文中に出現する単語候補や想定されるべき構文候補の数を削減する方法により文の意味やキーワードを抽出する研究がなされている[4),5)]．しかし，音声認識技術と自然言語処理技術を効果的に統合し，高精度にキーワードを抽出する技術の確立がなされていなかった．

本論文で述べるキーワード抽出アルゴリズムでは，従来と同様に，音声入力文（音素記号列）からコンフュージョン・マトリクスを用いて正解音素を推定し，これらの音素候補の組合せから，音節候補を生成する．次に，音節候補の組合せを検索キーとして辞書引きを行い，単語として成立する候補を選別する．本論文では，上記の処理のうち，演算の実行回数が，各入力音素から推定する音素候補数の積のオーダとなる辞書検索を，連続音声の発声速度に追随して実行するために，木構造インデックス，並列検索の手法を導入した高速単語候補抽出法を提案する．

さらに，演算量削減の観点から，単語候補の品詞に着目し，敬語表現等の慣用句の除去により明らかに操作指示内容の確定には不要であると判断できる単語をあらかじめ振るい落とす．また，タスクに依存しない汎用の構文スコアリング規則を新たに定義し，これにより確度の高いキーワード候補を選択する操作指示キーワード抽出法を提案する．

最後に，操作指示の的中率の点から評価したソフトウェア・シミュレーション結果，および実行時間の評価結果を示す．

2. 単語候補抽出

文を対象とした全単語候補の抽出では，検索キーとする単語読みの始端，終端の位置を変えながらすべての読みに対して，辞書への登録がなされているか否かを調べる必要がある．この場合の検索すべき単語の総数 w は，文の長さ（音節数／1文）を n，1音節当り m の候補があるとすれば，以下の式(1)のように表される．

$$w=\sum_{i=0}^{n-1}(n-i)m^{i+1}. \qquad (1)$$

† Keyword Spotting System for Continuous Speech Input by OSAMU NAKAMURA, TADASHI KAWANOBE, MITUTERU YUKISHITA and MASAMI OGURO (NTT Electrical Communications Laboratories).

†† NTT 電気通信研究所

例えば，$n=100$，$m=2$ の文に対しては，約 5×10^{30} 回もの検索が必要となる．

そこで，本システムでは，木構造インデックスを用いて効果的に先刈り（検索キーとする音節列と一致する部分木のみを検索）を実行するとともに，並列検索の手法を導入した辞書検索装置によって，検索時間の短縮を図る．以下に，単語候補抽出法と合わせて，これを高速に実行するための辞書検索装置構成について述べる．

2.1 単語候補抽出法

全単語候補の抽出を高速に実行するため木構造インデックスを採用する．木構造インデックス内の各ノードは，音節（または音素），終端マーク（単語の品詞コードを含む），単語辞書データへのポインタ，および下位レベルノードへのポインタから構成する．**図1**は，ノードに音節を割り当てた場合の木構造インデックスの構成例を示している．また，同図では，入力音節列の先頭から単語候補の抽出を行う場合の動作例も示している．すなわち，単語候補抽出は，文の先頭からだけではなく，第2音節，第3音節，…，と順次1音節ずつ抽出開始位置をずらし，入力音節列中の各音節と，音節の位置に対応するレベルの木構造インデックスのノードの音節とが一致する場合に，そのノードに付したポインタによって，木構造インデックスの最上位レベルのノードから順次，下位レベルのノードへとトレースする．このトレースの過程においては，終端マークを検出するたびに単語候補を抽出し，一致する音節がなくなるまでトレースを繰り返すことにより，入力音節列中に含まれる単語候補をすべて抽出する．上記の方法によって，1入力音節当り m 個の候補がある n 音節の入力に対して，$O(n\cdot m^{\alpha})$ の検索回数ですべての単語を抽出できる．ここで α は経験的に2〜3である．

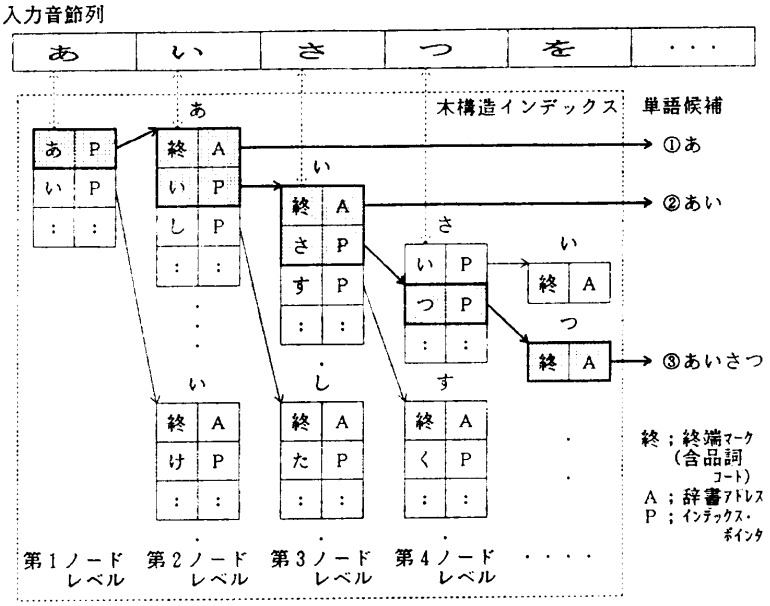

図1 単語候補抽出法
Fig. 1 Word extraction method.

2.2 辞書検索装置構成

2.1節に示した単語候補抽出法の効果を発揮させ，さらに高速化を実現するため，本節では，以下の3手法を導入した辞書検索装置の構成（**図2**）を提案する．

（1） RAM化：インデックスと辞書データをRAMに常駐させることにより，アクセスの高速化を図る．

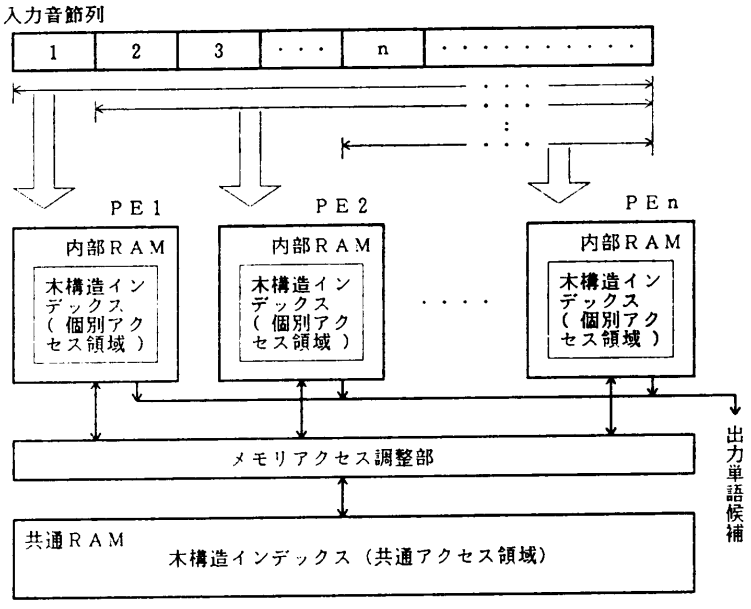

図2 単語検索装置の構成
Fig. 2 Configuration of a word retrieval facility.

(2) 階層化：木構造インデックスにおいて，各ノードからのアークの数（分岐数）が比較的多いノードレベル（音節をノードに割り当てる場合には4ノードレベル前後まで）と分岐数の少ないノードレベルとを別階層とし，それぞれの階層を独立にアクセス可能とすることによって，負荷の分散を図る．

(3) 並列化：上位階層ノードレベルのインデックス・ノードを n 個複写し，その各々に検索機構を割り当てる．すなわち，高アクセス頻度の上位階層レベルのノードに対しては n 並列で検索を行い，低アクセス頻度の下位階層レベルのノードは共通アクセス領域とする．

図2では，n よりも十分に多い音節からなる文を入力データとし，並列検索機構 (PEi) への入力データの割り付けは，PE1 には第1音節から文末の音節まで，PE2 には第2音節から文末の音節まで，…，PEn には第 n 音節から文末の音節まで，のように行う．図2中のメモリアクセス調整部は，共通領域とした下位階層レベルのノードへのアクセス競合を調整する機構である．上記の構成とすることにより，音声の実時間認識に追随した単語候補抽出の実現が期待できる．

3. キーワード抽出法

本章では，単語検索装置によって抽出した単語候補から，より確からしい単語候補を選択する方法について述べる．

単語候補の確からしさは，以下の3項目に関する評価より決定する．

(1) 品詞に基づく選択の優先順位付け，ならびに単語候補の入力音節列中における位置の重複チェック（単語候補の予備選択）．

(2) 動詞句から想定される構文パタンの要素としての妥当性チェック（構文照合）．

(3) 操作指示の対象世界（タスク）との適合性評価（構文スコアリング）．

上記の3項目のうち，(1)は個々の単語候補に着目して評価し，(2)，(3)は単語候補の組合せを対象に評価する．

3.1 単語候補の予備選択

個々の単語候補に着目しての予備選択は，以降の構文レベル（単語候補の組合せ）の選択における処理時間の短縮に与える効果が大きい．しかし，単語候補の絞り込みと，正解単語の誤棄却とは相反する要因であり，また，この段階で正解単語の誤棄却が発生すると，以後復元することができない．したがって，本予備選択では，実験評価から，誤棄却の発生が少ないことが明らかになった以下の二つの方法によって，単語候補の振るい落としを行う．

(1) 構文パタンを推定する基準となる動詞を，最優先に選択し，これと位置が重なる他品詞の単語候補を棄却する．

(2) 操作指示内容の同定に直接必要とならない敬語表現等の慣用句を抽出し，この慣用句自身，ならびに，この慣用句と位置が重なる単語候補を棄却する．

上記のうち，(2)では，あらかじめ慣用句を登録しておく必要があるが，一般的に使用頻度の高い上記慣用句の数は，限られているので登録は容易であると考えられる．

3.2 構文照合

操作指示システムのキーワード抽出向きに簡略化した構文照合の方法を示す．本構文照合の目的は，前節の予備選択によって選択された単語候補のうち，操作指示の種別ごとに定義した単語候補属性の組合せ（以降，構文パタンと呼ぶ）に当てはまるか否かを識別することによって，少数の単語候補に絞り込むことである．**図3**は構文照合処理の流れを示している．以下に，各処理ブロックの機能概要を示す．

①構文パタン要素属性との照合：予備選択によって選択した単語候補を入力とし，それらのうち，動詞句を構成する単語候補を識別する．ここで動詞句とは，動詞語幹＋語尾＋付属語（助詞，助動詞，等）と定義し，付属語については省略可能とする．次に，動詞句によって定まる構文パタンについて，その要素の属性

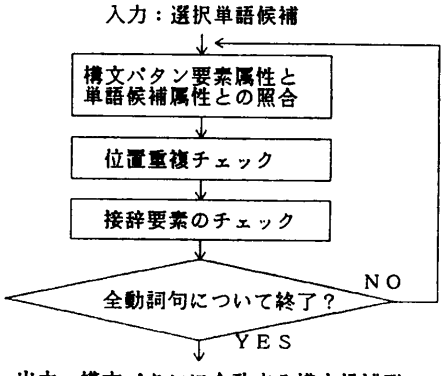

図3 構文照合の処理の流れ
Fig. 3 Sentence structure check process.

と入力された単語候補属性との照合を行い，いずれの構文要素にもなり得ない単語候補を棄却する．また，構文要素となり得る単語候補については，それぞれにどの構文パタンの要素であるかを識別する情報を付加して出力する．

②位置重複チェック：構文候補内において，位置的関係が矛盾する（重なる）単語候補について，異なる構文パタンの要素として識別し，構文候補を生成する．

③構文要素チェック：操作指示語以外に，単独では構文要素となり得ない接辞等の単語候補だけで構成されている構文候補を棄却する．

上記①～③の処理を，すべての動詞句について繰り返すことにより，個々の単語候補に着目した絞り込みを行う．

3.3 構文スコアリング

前節の構文照合処理によって絞り込んだ単語候補から生成する構文候補は，以下に示す条件によって，操作指示として妥当であるか否かを評価する（以降，スコアリングと呼ぶ）．

（1） 構文としての充足条件

（2） 構文構成要素間の局所的な充足条件

上記の条件をより詳細に展開したスコアリングの規則を**表1**に示す．また，表1には，実験により求めたスコアリングの採点基準の例も合わせて示している．このスコアリング規則を用いて，実際に構文スコアリング処理を行う流れを**図4**に示す．

表1 構文スコアリング規則
Table 1 Scoring rules.

スコアリング規則		スコアリング基準例
構文レベル	システム状態との整合性	整合＝5点，不整合＝0点
	構文パタン内要素の充足度	(照合がとれた単語候補個数)÷(構文パタン内要素個数)
	動詞句の位置	文末＝1点，文頭＝0点
単語レベル	属性別単語長	属性別の最長単語含有個数
	単語間の意味的接続関係	接続可＝2点，接続不可＝0点
	位置別単語長	異なる位置ごとの最長単語個数*

* 構文候補内のすべての構文要素が最長単語となる場合は，さらに2点を与える

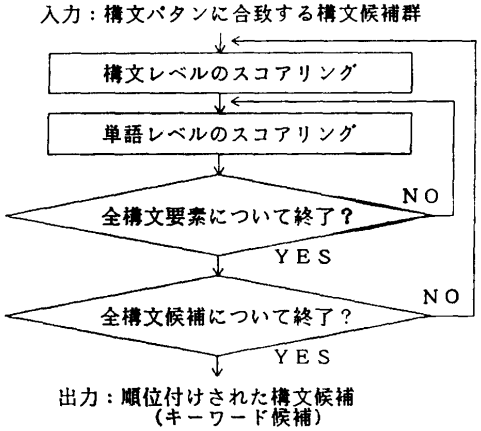

図4 構文スコアリング処理の流れ
Fig. 4 Scoring process for a sentence.

4. 実験評価

前章までに示したキーワード抽出システムについて，その有効性を明らかにするため，具体的に適用環境等を設定し，ソフトウェア・シミュレーションによって，操作指示の的中率，ならびに実行時間の評価を行った．以下，評価実験の方法，結果について述べる．

4.1 タスクの設定

具体的な操作指示システムの例として，音声による電話交換サービスを取りあげる．**表2**は，電話交換サービスとこれらに対応する操作指示語，その他の辞書に登録した語彙，およびこれらの登録形式を示している．なお，表2に示した語彙のうち，慣用句を除く語彙をキーワードとして扱う．また，**図5**は，表2の

表2 実験用辞書の内容
Table 2 Contents of dictionary for the evaluation.

語彙分類		語数（規模別） 250	500	1004	登録形式
操作指示語*	接続	17			単語・句
	転送	10			単語・句
	伝言	3			単語
	割込み	5			単語・句
	呼返し	5			単語・句
名詞*	氏名(姓)	83	167	620	単語
	氏名(名)	0	166		単語
	場所	2			単語
	役職	16			単語
	所属	60		111	単語
副詞*	時間	7			句
慣用句		42			単語・句

* キーワードとして登録する

操作分類	No	構文パタン
接続・転送	①	操作語+時間+電話番号
	②	操作語+時間+所属+役職
	③	操作語+時間+場所+所属
	④	操作語+時間+氏名+所属+役職
割り込み・伝言		操作語
呼び返し	①	操作語+時間
	②	操作語+時間+電話番号
	③	操作語+時間+場所+所属
	④	操作語+時間+氏名+所属

図5 構文パタン例
Fig. 5 Examples of sentence structure.

操作指示語に対応する構文パタンを示す．ここで想定した電話交換サービスの概要は，以下のとおりである．

（1） 接続：発信者の希望する相手に接続する．

（2） 転送：着信した呼を着信者の希望する相手に転送する．

（3） 割込み：話中の相手への割込みの起動を交換システムに伝える．

（4） 伝言：発信者の希望する相手への伝言依頼を交換システムに伝える．

（5） 呼返し：発信者への呼返し依頼を交換システムに伝える．

実験評価には，実際の会話文表現の35例文を用いた．

4.2 音声認識シミュレーション法

文単位の音声を入力として，音節列の出力を行う音声認識装置が入手できないため，また，認識率をパラメータとして扱うため，本実験評価では，コンフュージョン・マトリクス[6]を用いた音声認識のシミュレーション結果を用いることとした．実際の音声認識過程では，誤認識として，置換，付加，および挿入が発生すると予想されるが，ここでは置換のみを対象として，以下の手順でシミュレーションを行った．付加，挿入に関しては別途報告する予定である．

① 正しい音節列(文)から音節を切り出す．

② 切り出した音節列を音素に分解する．

③ 分解した各音素について，コンフュージョン・マトリクスを参照し，乱数によって置換音素を得る．

④ 置換音素を組み合わせて音節を生成する（結果として，後の実験のために，1入力音節当り4個までの置換音節を重複なく生成することとした）．

4.3 音素ラティスの生成

前節で述べた置換音節から，同じくコンフュージョン・マトリクスを用いて，正解音節候補の推定を行う．この推定では，置換音節の生成の逆の処理を行う．具体的な手順は以下のとおりである．

① 置換音節を含む入力音節列を音素に分解する．

② 分解した各音素について，コンフュージョン・マトリクスを参照し，置換確率の高い音素から順に正解音素候補とする．

③ 正解音素候補を組合わせて，正解音節候補を生成する．

上記②において，正解音素候補数の設定に当っては，実際の正解音素がほぼ4位までに含まれることから，最大4個とした．

4.4 操作指示の的中率・正解包含率

操作指示の的中率（スコアの第1位が正解である率）ならびに正解包含率（スコアの第 n 位までに正解が含まれる率）を評価尺度として，本キーワード抽出システムの効果を評価した．的中率（正解包含率）H は，式(2)のように定義する．

$$H=(C/T)\times 100\%. \qquad (2)$$

H：的中率（正解包含率），

C：正しくキーワードが抽出された例文の数（個），

T：例文の総数（個）．

また，主な評価パラメータ値を以下のとおり設定した．

- 辞書規模：表2．
- 音素認識率：文献6)より引用した100都市名の認識実験結果（54%）を基準として，誤認識音素の置換確率を 1/2 (77%), 1/4 (89%) とした．
- 認識結果：音声認識シミュレーション時に出力する音節候補の数を4音節とした．
- 誤認識音素種別：音声認識シミュレーション時の誤認識音素の種別を子音のみ(C)と，子音・母音とも(CV) の2通りを設定した．
- 正解音素候補数：100% の正解音素復元率を確保するための正解音素候補数を，各音素認識率に応じて設定した．

図6～9に的中率ならびに正解包含率を示す．なお，これらの図においては，3位までに正解が含まれていた場合を正解包含率として示している．これらの結果から，主要な評価パラメータと的中率・正解包含率には，以下の関係があることが分かった．

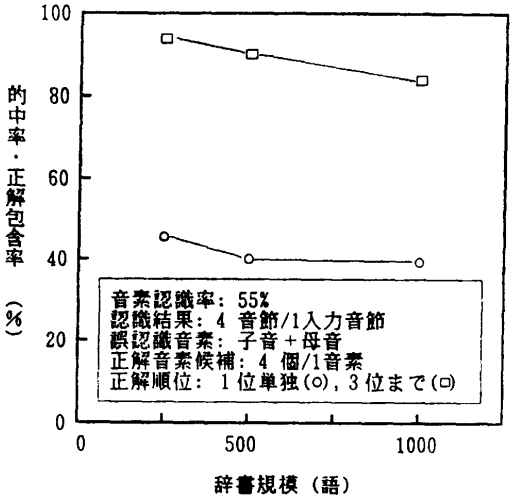

図 6 的中率 (1/4)
Fig. 6 Correct interpretation ratio (1/4).

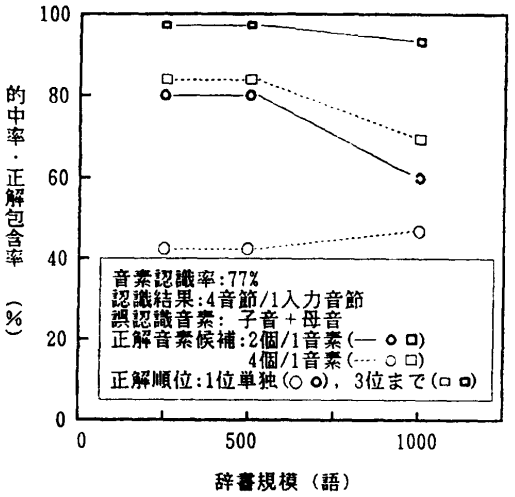

図 7 的中率 (2/4)
Fig. 7 Correct interpretation ratio (2/4).

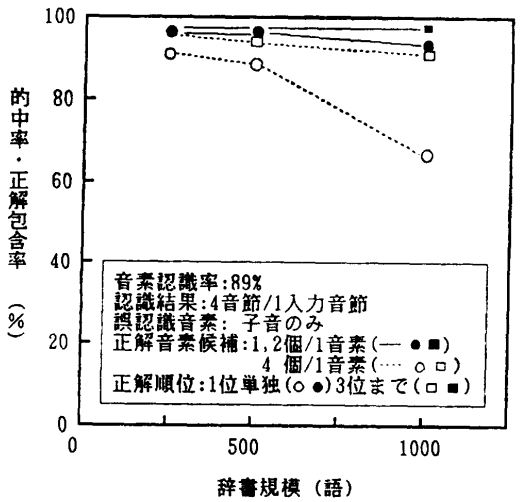

図 8 的中率 (3/4)
Fig. 8 Correct interpretation ratio (3/4).

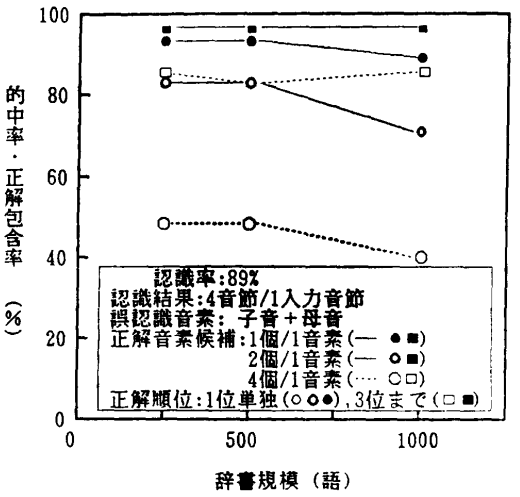

図 9 的中率 (4/4)
Fig. 9 Correct interpretation ratio (4/4).

(1) 辞書規模：本実験評価で設定した 1,000 語までの範囲においては，500 語の増加につき，最大 20%，的中率が低下するが，スコアリングの順位を 3 位まで許容する場合の正解包含率は，たかだか10%前後の低下で済む．

(2) 誤認識音素の種別：子音，母音ともに誤認識する可能性がある場合には，子音のみが誤認識する場合に比べて，最大 40%，的中率が低下するが，(1)と同様にスコアリングの順位を 3 位まで許容する場合の正解包含率は，たかだか 10% 前後の低下で済む．

(3) 音素認識率：50%前後では，平均音素認識率が的中率に与える影響が大きいが，77%以上の領域では，音素認識率が，的中率の支配的要因にはならない．

(4) 正解音素候補数：本実験評価で設定した範囲では，正解音素候補数を 2 とした場合が最も的中率が高い．また，音素認識率が高くなるにつれて，正解音素候補数の 2 以上への増加は，かえって的中率の低下を招く．

(5) 正解が含まれるスコアリング順位：平均音素認識率が 77% 以上，正解音素候補数を 2，スコアリング順位を 3 位まで許容することによって，ほぼ 95%以上の率で正しいキーワードの組合せを抽出できる．

4.5 実行時間の評価

本稿で提案したキーワード抽出システムの，発声速度への追随性を明らかにすることを目的に，実行時間の評価を行った．ここでは，1秒に相当する入力音節列（平均12音節/秒）に対して，正解音素候補生成，単語候補抽出，および構文照合・スコアリングにおける各演算量を算出し，これらを基に実行時間を予測した．上記の各処理ブロックの演算量は，式(3)～(5)で表される．これらの式より，実行時間の支配項は単語候補抽出処理であることが分かる．

- 正解音素候補生成処理の実行時間（T_1）

$$T_1=800\times a\times L\times C. \quad (3)$$

- 単語候補抽出処理の実行時間（T_2）

$$T_2=200\times L\times A^5\times C. \quad (4)$$

- 構文照合・スコアリング処理の実行時間（T_3）

$$T_3=270\times S\times W\times C. \quad (5)$$

ただし，

C：1クロック時間（秒），
a：1発声音素当りの音素認識候補個数（音素），
L：1秒当りに入力される音節列長（音節），
A：1発声音節当りの，正解音素候補から生成する音節候補個数（音節），
S：構文照合処理から出力される構文候補個数（構文），
W：単語候補抽出処理から出力される単語候補個数（単語），

であり，式中の定数は，単語辞書規模を約1,000単語とした条件において，各処理プログラムの分析から推定した値である．

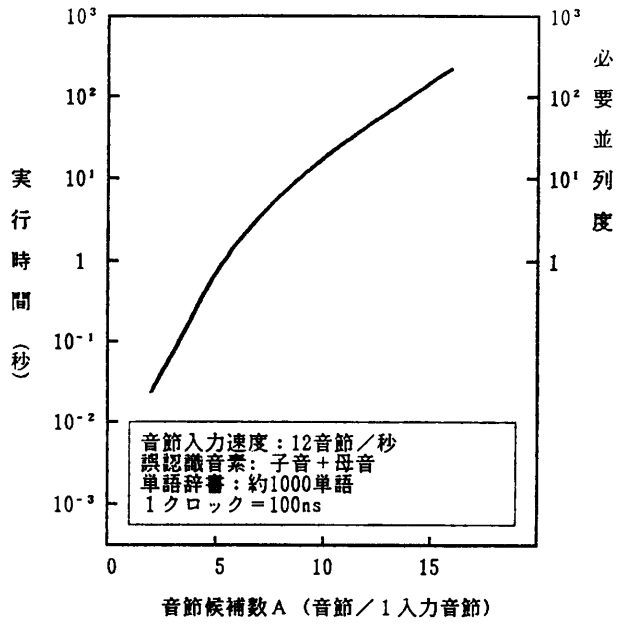

図 10 実行時間
Fig. 10 Execution time.

図10は，式(3)～(5)において，$C=100$ ns，$L=12$，とし，S，W については実測値を用いた場合の，実行時間，および単語候補抽出における必要並列度を示している．図10より，音節候補数 A が15までの範囲（実用的な的中率の達成範囲）においては，最大100並列の単語候補抽出処理によって，発声速度に追随可能であることが分かる．

5. むすび

音声による操作指示の実現をねらいとしたキーワード抽出システムについて，高速単語候補抽出法，単語候補の予備選択，構文パタン照合，および構文スコアリングによるキーワード抽出アルゴリズムを提案した．また，シミュレーションによる実験評価を行い，77%～89%の範囲の平均音素認識率，1,000単語の辞書規模，1発声音節当り4音節の認識結果出力，および，1入力音素当り2個の正解音素候補の推定を行うこととした，システム条件において，正しい操作指示内容を示すキーワードの組合せを，60%以上の的中率で抽出できること，スコアの上位3位までを取れば95%以上の包含率で正解候補が得られることを明らかにした．さらに，本キーワード抽出処理に要する実行時間を評価し，発声速度に追随可能であること，その場合の必要並列度が最大100であることを明らかにした．上記の結果から，ここで提案したアルゴリズムの有効性を確認することができた．

今後は，実際に音声認識装置からの出力を対象として，本アルゴリズムの効果検証を進めるとともに，音素の付加，脱落にも対処可能なシステム構築の検討を進める予定である．

謝辞 本論文をまとめるに当り御指導いただいたNTT情報通信処理研究所・認識処理研究室・篠岡 信室長に深謝いたします．

参考文献

1) 牧野，鈴木，城戸：推移確率の利用による音素系列の訂正，日本音響学会音声研究会資料，S 74-2, pp. 1-10 (1974).
2) 新津，三輪，牧野，城戸：単語音声自動認識における言語情報の一利用法，電子通信学

会論文誌D, Vol. J62-D, No. 1, pp. 24-31(1979).
3) 中川，義永：誤りを含んだ音素系列からの候補単語の検索，計量国語学，Vol. 14, No. 8, pp. 327-334 (1985).
4) 鹿野，好田：会話音声の機械認識における言語処理，電子通信学会論文誌D, Vol. J61-D, No. 4, pp. 253-260 (1979).
5) 浮田，石川，中川，坂井：音声による対話システムにおける発話の確認方法，情報処理学会論文誌，Vol. 22, No. 6, pp. 589-595 (1981).
6) 相川，杉山，鹿野：トップダウン処理による音韻認識，研究実用化報告，Vol. 32, No. 11, pp. 2281-2292 (1983).

(昭和61年5月15日受付)
(昭和61年12月10日採録)

中村　修（正会員）

昭和25年生．昭和49年青山学院大学理工学部電気電子工学科卒業．同年日本電信電話公社武蔵野電気通信研究所入所．以来，ファイル記憶系構成および並列処理技術の研究に従事．現在，日本電信電話(株) NTT情報通信処理研究所主任研究員．電子情報通信学会会員．

川野邊　正（正会員）

昭和22年生．昭和47年茨城大学工学部電子工学科卒業．同年日本電信電話公社武蔵野電気通信研究所入所．以来，電子交換機中央制御装置の実用化，並列処理応用技術の研究に従事．現在，日本電信電話(株) NTT情報通信処理研究所主幹研究員．電子情報通信学会会員．

雷下　充輝（正会員）

昭和34年生．昭和56年東北大学工学部電気工学科卒業．同年日本電信電話公社武蔵野電気通信研究所入所．以来，LSIの階層設計および並列処理技術の研究に従事．現在日本電信電話(株) NTT情報通信処理研究所研究主任．

小黒　雅己（正会員）

昭和38年生．昭和60年熊本大学工学部電子工学科卒業．同年日本電信電話(株)武蔵野電気通信研究所入所．以来，並列処理技術，知識ベースマシンの研究に従事．現在，NTT情報通信処理研究所勤務．