Vol. 22 No. 5 情報処理学会論文誌 Sep. 1981

音韻認識能力と単語認識率との関係†

中川 聖一††

大規模語彙や不特定話者の単語音声の認識では，単語単位のパターンマッチングよりも，一旦音韻単位で認識した後，単語を認識する2段階方式の方が有望だと思われる．本論文では，単語音声自動認識システムにおける音韻認識部をシミュレートすることによって，音韻識別率，脱落確率，挿入確率等と単語認識率との関係を調べた．その結果，挿入誤りよりも脱落誤りが単語認識率に与える影響が大きいこと，母音と子音の識別が単語認識率に与える影響に大差がないこと，同一の調音様式をもつ音韻群を同一カテゴリとして識別した場合，単語認識率は数%低下すること，母音と子音の弁別よりも有声音と無声音の弁別の方が重要であること，音響的に安定な継続時間の長い音韻は，複数個の音韻記号列として識別する方がよいことなどがわかった．また，現在の技術では，音韻認識に基づく単語音声認識では，不特定話者に対しては50～100単語，特定話者に対しては100～200単語程度なら，95%以上の単語認識率が得られることがわかった．

1. はじめに

最近になって，限定話者を対象とする単語音声の認識装置が商品化された．限定話者に対しては，個人差への配慮をしなくてよく，単語単位のパターンマッチング法が有効である．わずかに，発声時間長の変動だけが問題となるが，時間軸上での動的計画法を利用した非線形マッチングの導入により解決できた．この方法を，不特定話者用に拡張するためには，個人差の適応化[1),2)]か正規化[3)~6)]を考えなければならない．単語音声に現われる主な個人差には，発声物理器官の差（ハード差）と発声法の差（ソフト差）がある．ソフト差には，各音韻の継続時間長や無声化母音の存否などに個人差がある．これらは，入力音声を音韻レベルで記述できれば問題はない．

単語音声を音韻系列で記述すると各種の言語情報の導入が可能である．たとえば，無声化・有声化・鼻音化されやすいコンテキストの指定と変換規則等の導入が考えられる．また，多数話者の音声から標準パターンを作成する場合を考えると，単語単位の標準パターンよりも音韻単位の標準パターンの方が作成しやすく，学習もしやすい．

以上の考察により，大規模語彙や不特定話者の単語音声の認識では，単語単位のパターンマッチングよりも，一旦音韻単位で認識した後，単語を認識する二段階方式の方が有望だと思われる．しかし，音韻単位で認識する場合には，新たな問題が生じる．1つは，音韻単位へのセグメンテーションが非常に困難なこと，ほかの1つは，調音結合の影響を受けた音声を正しく音韻に変換することが困難なことである．これらは，セグメンテーションの脱落誤り，挿入誤り，音韻の識別誤りで評価することができる．多少の誤りを生じても限定語彙の単語認識では，言語のもつ冗長性を利用することによって正しく単語に変換できる場合が多い．では，一体どの程度の音韻認識部の能力があれば，所望の単語認識率が得られるのであろうか．大変興味あるところである．

単語認識に必要な音素対の種類や音素対の誤認識によって混同の起こる単語対などの統計的調査は数多く行われてきた[7)~13)]．これらの統計的調査の結果からは，鼻子音の識別はできなくても，かなりの精度で単語は認識できる等の知見は得られたが，音韻認識率と単語認識率との関係（たとえば，母音の認識率が80%の場合は，単語の認識率はいかほどか）は明らかにされていない．つまり，これらの調査は，音韻を単位とする単語認識システムの直接的な資料とはならなかった．

音韻認識率と単語認識率との関係は，セグメンテーションが完全な場合，つまり音韻認識部の誤りが識別誤りだけの場合は，解析が可能で文字認識の分野で研究されている[14),15)]．しかし，セグメンテーションが不完全な場合は，解析が困難でシミュレーションで検討せざるを得ない．

本論文では，音韻認識部の動作をシミュレートすることによって，音韻認識部の能力と単語認識率との関係を明らかにし，音韻認識部の研究努力目標を明らかにする．

† Relationship between Phoneme Recognition Performance and Word Recognition Rate by SEIICHI NAKAGAWA (School of Information Engineering, Faculty of Engineering, Toyohashi University of Technology).

†† 豊橋技術科学大学工学部情報工学系

2. シミュレーションの方法

シミュレーションが，実際の音韻認識部とかけ離れていると意味がない．この点に注意を払いながら，しかもシミュレーションの方法があまり複雑にならない程度に簡略化している．

2.1 音韻カテゴリ

音韻認識部の動作をシミュレートする際に，取り扱う音韻のカテゴリは次の通りである．母音 /*a*, *i*, *u*, *e*, *o*/，撥音 /*N*/，半母音 /*y*, *w*/，有声子音 /*m*, *n*, *ŋ*, *b*, *d*, *g*, *r*, *z*/，無声子音 /*s*, *c*, *h*, *p*, *t*, *k*/，促音 /*Q*/，/ʃ/ は /*s*/ に，/*ts*, *ch*/ は /*c*/ に，/*dz*/ は /*z*/ に，拗音 /*j*/ は半母音 /*y*/ に統一した．また長母音は同母音を2個続けて表現する．

2.2 音韻認識における誤り

音韻認識における誤りは，次の3種類とする．

ⅰ）置換誤り……セグメンテーションの後，音韻の識別が行われるときに，誤ってほかの音韻として識別されること．

ⅱ）脱落誤り……セグメンテーションの際に，ある音韻が1つのセグメントとして抽出されずに脱落すること．

ⅲ）挿入誤り……脱落誤りとは逆に，1つのセグメントとして抽出されるべき音韻が2つのセグメントとして抽出されること．3つに分割されないとしている．

挿入誤りを起こした場合は，ある音韻 A があって，A 以外の音韻を x とすれば，2つとも正しく識別されて AA となるか，Ax, xA, xx となるかのいずれかである．（挿入誤りがあると，少なくともどちらか一方が誤ると定義することも考えられるが，断らない限り上述に従うものとする．）

ある音韻が認識誤りを起こす過程は，コンテキストを考慮したマルコフ過程としてとらえるべきであるが[16),17)]，実際の音韻認識部のふるまいをできるだけ忠実にシミュレートするためには，多くのパラメータが必要となり，シミュレーション結果の解釈が難しくなってしまう．そこで，シミュレーションの簡単化のため，これらの誤りは，コンテキストに関して統計的に独立であるとした．

2.3 シミュレーションに用いるパラメータ

シミュレーションを行う際に，パラメータの選択と決定という問題は，どのような場合にも生じる．本論文におけるシミュレーションでは，次のパラメータを用いた．音韻識別率，音韻脱落確率，音韻挿入確率，音韻置換確率．これらはいずれもシミュレーション結果に大きく影響する．音韻置換確率は，ある音韻がどの音韻に識別されやすいかを表わす確率である．音韻 i が音韻 j に識別される確率を $p(j/i)$ とすると，$\sum_j p(j/i)=1$．$p(i/i)$ は音韻 i の識別率を表わす．なお，促音 /*Q*/ に限り，識別率は100%，脱落確率は0%，挿入確率は0%とした．2音韻が連続して脱落や挿入誤りを生じることは可能である*．

2.4 実験方法

将来，単語音声の認識対象語彙数は数百単語になるものと思われるが，シミュレーションによる処理時間の制約から200単語を対象とする．同一カテゴリで数百単語の語彙としては都市名が代表的で実用性も高い．それゆえ，我々が実時間単語音声識別システム[21)]の評価のために使っていた100都市名に新たに100都市名を追加した200都市名を認識対象とした（**表1**）．

シミュレーション用のパラメータの各組に対して，乱数表を使って各都市名に対して擬似識別音韻列を生成した．これらの音韻列は，単語辞書で与えられてい

表1 200都市名

Table 1 Vocabulary of 200 city names.

札幌	青森	秋田	盛岡	仙台	山形	福島	水戸
宇都宮	前橋	浦和	千葉	東京	横浜	新潟	富山
金沢	福井	甲府	長野	岐阜	静岡	名古屋	津
大津	京都	大阪	神戸	奈良	和歌山	鳥取	松江
岡山	広島	山口	徳島	高松	松山	高知	福岡
佐賀	長崎	熊本	大分	宮崎	鹿児島	那覇	旭川
函館	釧路	小樽	帯広	網走	川崎	松本	浜松
堺	姫路	下関	北九州	稚内	根室	室蘭	苫小牧
弘前	釜石	いわき	郡山	会津若松	高崎	大宮	川口
船橋	八王子	横須賀	小田原	長岡	大垣	清水	熱海
豊橋	四日市	彦根	舞鶴	東大阪	豊中	尼崎	西宮
明石	米子	出雲	倉敷	呉	福山	宇部	今治
新居浜	久留米	佐世保	別府	北見	大館	八戸	米沢
鶴岡	酒田	石巻	日立	土浦	足利	小山	栃木
桐生	伊勢崎	太田	川越	熊谷	所沢	秩父	越谷
草加	銚子	市川	松戸	習志野	柏	市原	立川
武蔵野	三鷹	青梅	府中	調布	町田	小金井	小平
日野	東村山	国分寺	平塚	鎌倉	藤沢	茅ケ崎	相模原
厚木	大和	上越	上田	飯田	高岡	小松	一宮
岡崎	豊田	春日井	瀬戸	豊川	蒲郡	稲沢	安城
小牧	刈谷	沼津	三島	富士	焼津	富士宮	伊勢
松阪	鈴鹿	桑名	宇治	岸和田	池田	吹田	高槻
枚方	守口	茨木	八尾	寝屋川	門真	泉	松原
伊丹	加古川	宝塚	芦屋	相生	尾道	津山	三原
岩国	徳山	防府	大牟田	唐津	八代	都城	延岡

* 音韻 i の脱落確率を $p_D(i)$，挿入確率を $p_I(i)$，識別率を $p(i/i)$ とすれば，音韻 i の認識率（正しくセグメンテーションされて且つ正しく識別される確率）は $(1-p_D(i)-p_I(i))\cdot p(i/i)$ である．図1参照．

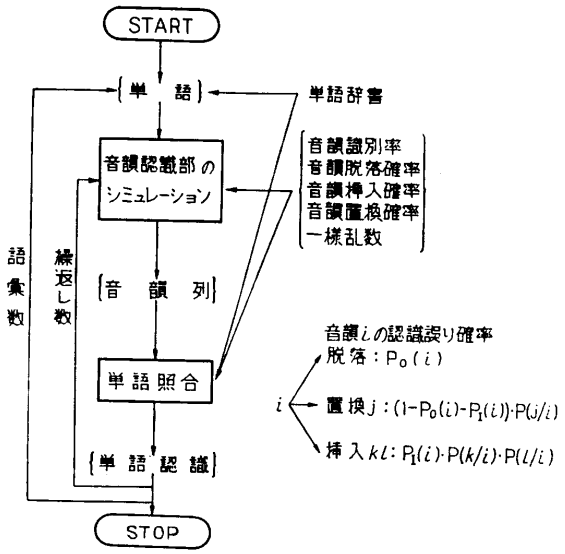

図 1 シミュレーションの概略
Fig. 1 Flowchart of simulation.

る音韻列と次章で述べる方法で比較され，最もよく適合する単語辞書を見い出すことにより単語に変換される．実験方法の概略を **図 1** に示す．繰返し数は，処理時間の制約から2回とした（予備実験の結果，5回としてもあまり有意な差はなかった）．シミュレーションの結果は，語彙数20，50，100，200の場合についてまとめた．たとえば，語彙数50の場合は，200都市名を4グループに分け，各々のグループの認識率の平均を50語彙の結果とした．

3. 単語認識の方法

3.1 誤り確率を利用しない方法

置換・脱落・挿入誤りがあれば，これらを同一の誤り尺度として取り扱う（Levenshtein distance）．たとえば，単語 W_1（音韻列：ABCDE）と単語 W_2（音韻列：ACDFG）の距離 D (W_1, W_2) は3となる．ただし，置換・脱落・挿入の誤りをそれぞれ距離1とする．ここで，単語間の距離は，対応関係の解釈によって種々考えられるが，それらのうち最小の距離と定義する．これは，以下の漸化式によって求めることができる．単語 W_1 ($a_1, a_2, \cdots a_i, \cdots a_I$) と単語 W_2($b_1, b_2, \cdots b_j, \cdots b_J$) の距離 D($W_1$, W_2) は

$$D(i, j)=d(i, j)+\min\begin{Bmatrix}D(i-1, j-2)+1\\ D(i-1, j-1)\\ D(i-2, j-1)+1\end{Bmatrix}$$

$$D(W_1, W_2)=D(I+1, J+1)$$

ここで，

$$D(1, 1)=d(1, 1)$$
$$D(2, 1)=\min\{d(1, 1), d(2, 1)\}+1$$
$$D(3, 1)=\min\{d(1, 1), d(2, 1), d(3, 1)\}+2$$
$$D(1, 2)=\min\{d(1, 1), d(1, 2)\}+1$$
$$D(1, 3)=\min\{d(1, 1), d(1, 2), d(1, 3)\}+2$$
$$d(I+1, J+1)=0$$

$d(i, j)$ は，$a_i=b_j$ なら0，$a_i \neq b_j$ なら1．

シミュレーションによって生成された音韻列と単語辞書によって与えられている音韻列を，上の単語間距離によって比較し，距離最小の単語辞書を認識結果とする．シミュレーション結果を **表 2** に示す（繰返し回数は10回）．ここで，音韻置換確率は，すべての音韻に対して一定とした．つまり，$p(j/i)=p(l/k)$，$p(i/i)=p(j/j)$，ただし $i \neq j$，$k \neq l$．表2より，脱落誤りが単語認識に与える影響が大きいことがわかる．これは，異なった音韻数をもつ単語間の距離は，長い方の単語に脱落誤りがある場合の方が，短い方の単語に挿入誤りがある場合よりも，両者の距離が近づくからであると解釈できる[25]．

この距離尺度に基づく認識方法は，次節で述べる誤り確率を用いる方法と比べると認識能力が低いため，あまり用いられない．ただ，単語の統計的性質を調べる時にはよく用いられる[10)~13)]．この方法で，200都市名の2単語間の距離を調べた結果が **表 3**(a) である．表3(b)は，鼻子音 /m, n, ŋ/，有声破裂音 /b, d, g/，無声摩擦音 /s, c, h/，無声破裂音 /p, t, k/ をそれぞれ同一カテゴリとして扱った場合を示す．表から，同じ調音様式をもつものを同一カテゴリと扱ってもそれ程認識率の低下につながらないことが予想できる（詳細は次章参照）．

表 2 シミュレーション結果
—誤り確率を利用しない場合—
Table 2 Simulation results.
—in the case without the use of error probability—

音韻識別率	脱落確率	挿入確率	20 都市名	100 都市名
60%	0%	0%	86.8%	75.5%
70	0	0	96.5	89.7
80	0	0	100.0	96.5
60	10	0	73.8	56.0
70	10	0	81.9	66.5
80	10	0	90.7	88.4
60	0	10	89.6	74.8
70	0	10	95.0	89.7
80	0	10	99.3	96.4
60	10	10	80.7	62.9
70	10	10	87.5	74.4
80	10	10	93.9	85.2

表 3 200 都市名の 2 単語間の距離分布

Table 3 Distribution of distance between words in 200 city names.

(a) 音韻カテゴリ数 23 個の場合

距離 1 の総数	8	距離 1 の単語対一覧
距離 2 の総数	92	富山—小山　甲府—防府　奈良—那覇
距離 3 の総数	419	熱海—伊丹　出雲—泉　北見—伊丹
距離 4 の総数	1,622	大分—太田　富士—宇治
距離 5 の総数	3,576	
距離 6 以上の総数	14,003	

(b) 音韻カテゴリ数 15 個の場合

距離 0 の総数	0	距離 1 の単語対一覧
距離 1 の総数	22	千葉—飯田　富山—小山　甲府—大津
距離 2 の総数	158	甲府—高知　甲府—防府　大津—防府
距離 3 の総数	662	奈良—那覇　福岡—鶴岡　長崎—尼崎
距離 4 の総数	2,238	大分—太田　川崎—高崎　高崎—高槻
距離 5 の総数	4,410	大垣—小牧　熱海—明石　熱海—伊丹
距離 6 以上の総数	12,410	西宮—一宮　出雲—泉　北見—伊丹
		太田—草加　銚子—調布　富士—宇治
		調布—防府

なお，この距離定義と類似なものに，二音韻列間の一致数を用いる方法[18),19)]も使われるが，上述の方法と大差ないと考えられる．

3.2 誤り確率を利用する方法

前節では，識別誤り，脱落誤り，挿入誤りをすべて同尺度で扱った．ここでは，事前にわかっている誤り確率を積極的に利用する方法を述べる．

入力音韻列 $a_1, a_2 \cdots a_i, \cdots a_I$ とある単語辞書の音韻列 $b_1, b_2 \cdots b_j, \cdots b_J$ の確率的な類似度を考えよう．つまり $p(a_1, \cdots a_I/b_1, \cdots b_J)$ を求める．この確率が最大になるような認識誤りの解釈を見い出せばよい．今，音韻 i の脱落確率を $p_0(i)$，挿入誤りを起こし音韻 j が挿入される確率を $p_1(i, j)$ とすると，コンテキストに独立であるとき，$p(a_1, \cdots a_I/b_1, \cdots b_J)$ は，次の漸化式で求めることができる[20)]．

$$\begin{aligned}&p(a_1, a_2 \cdots a_i/b_1, b_2 \cdots b_j)\\&\quad=p(a_1, a_2 \cdots a_{i-1}/b_1, b_2 \cdots b_j)\cdot p_1(b_j, a_i)\\&\quad+p(a_1, a_2 \cdots a_{i-1}/b_1, b_2 \cdots b_{j-1})\cdot p(a_i/b_j)\\&\quad+p(a_1, a_2 \cdots a_i/b_1, b_2 \cdots b_{j-1})\cdot p_0(b_j)\end{aligned}$$

上式の計算量を軽減するために，右辺の 3 項の最大値で右辺を近似し，両辺の対数をとると次の漸化式が得られる．

$$P(i, j)=\max\begin{Bmatrix}P(i-1, j)+\log p_1(b_j, a_i)\\P(i-1, j-1)+\log p(a_i/b_j)\\P(i, j-1)+\log p_0(b_j)\end{Bmatrix}$$

$$\log p(a_1, a_2 \cdots a_I/b_1, b_2 \cdots b_J)\equiv P(I, J)$$

ここで，

$$P(1, 1)=\log p(a_1/b_1)$$

$$P(1, 2)=\max\begin{Bmatrix}\log p_0(b_1)+\log p(a_1/b_2)\\\log p(a_1/b_1)+\log p_0(b_2)\end{Bmatrix}$$

$$P(1, 3)=\max\begin{Bmatrix}P(1, 2)+\log p_0(b_3)\\\log p_0(b_1)+\log p_0(b_2)+\log p(a_1/b_3)\end{Bmatrix}$$

$p_1(b_j, a_i)$ は，b_j が挿入誤りを起こし，a_i が挿入されることを意味する．本論文では，a_i はあらかじめ与えられている置換確率によって識別されると仮定している．すなわち，$p_s(b_j)$ を b_j の挿入確率として

$$p_1(b_j, a_i)=p_s(b_j)\cdot p(a_i/b_j)$$

しかし，上の漸化式では，a_i は b_j が分離したものと仮定しているが，実際には b_{j+1} が分離したとも考えられる．それゆえ，厳密には漸化式は次式で与えられる．

$$P(i, j)=\max\begin{Bmatrix}P(i-1, j)\\\quad+\max\begin{Bmatrix}\log p_s(b_j)+\log p(a_i/b_j)\\\log p_s(b_{j+1})+\log p(a_i/b_{j+1})\end{Bmatrix}\\P(i-1, j-1)+\log p(a_i/b_j)\\P(i, j-1)+\log p_0(b_j)\end{Bmatrix}$$

シミュレーションによって生成された音韻列 $\boldsymbol{a}$ と単語辞書によって与えられた音韻列 $\boldsymbol{b}^k(k=1\sim200)$ から，上式より $p(\boldsymbol{a}/\boldsymbol{b}^k)$ を求め，最大確率を与える単語辞書 k を認識結果とする．次章では，この方法を用いたシミュレーション結果について述べる．

4. シミュレーション結果

4.1 音韻識別率，脱落確率，挿入確率の影響

シミュレーションでは，実際のシステムの音韻認識部の能力に近いパラメータの値を選ばなければならない．ここでは，各パラメータの値は全音韻一定とし，音韻識別率は 70%，80%，90% の 3 通り，脱落確率は 0%，5%，10% の 3 通り，挿入確率は 0%，5%，10% の 3 通り，合計 27 通りの組合せについてシミュレーションを行った．また，音韻置換確率は **3.1** 節と同じく全音韻等確率とした．結果を **表 4** に示す．（シミュレーション結果の統計的誤差は 1～2%程度と考えられる．）表から次のことがわかる．

(i) 表 2 の結果よりも格段に優れている．

(ii) セグメンテーションが完全な場合は，音韻識別率が 70% でも，200 単語に対して 94.5% の認識率が得られる．

(iii) 脱落誤りは挿入誤りよりも致命的である．実用的には，音韻識別率は 80% 以上，脱落確率は 5% 以内であることが望ましい．

表 4 音韻識別率，脱落確率，挿入確率の単語認識率に与える影響

Table 4 Relationship between word recognition rate and phoneme recognition rate.

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	0%	97.0%	97.0%	96.5%	94.5%
80	0	0	98.0	98.5	98.0	97.0
90	0	0	100.0	100.0	100.0	100.0
70	5	0	93.0	90.5	88.5	86.5
80	5	0	100.0	97.0	96.5	94.5
90	5	0	100.0	99.5	99.0	98.0
70	10	0	94.0	87.5	84.5	82.0
80	10	0	98.0	95.0	91.5	84.0
90	10	0	99.0	98.0	97.5	96.5
70	0	5	99.0	95.0	93.0	90.5
80	0	5	99.0	97.0	95.5	95.0
90	0	5	100.0	99.5	99.5	99.5
70	5	5	94.0	90.0	88.0	83.5
80	5	5	99.0	97.0	95.5	93.0
90	5	5	100.0	99.5	99.0	98.5
70	10	5	98.0	88.0	85.0	81.5
80	10	5	99.0	91.5	89.5	85.5
90	10	5	99.0	95.5	93.0	91.0
70	0	10	99.0	94.5	93.0	89.0
80	0	10	99.0	97.5	97.5	97.0
90	0	10	100.0	99.5	99.5	99.0
70	5	10	98.0	92.5	87.5	83.0
80	5	10	99.0	94.5	93.0	90.0
90	5	10	100.0	99.5	99.5	98.5
70	10	10	92.0	89.0	86.5	82.5
80	10	10	99.0	94.5	92.5	89.0
90	10	10	100.0	97.0	94.0	91.5

(iv) 語彙数が増加すれば，認識率は当然低下するが，それほど顕著ではない．近似的に誤認識率は，語彙数の増加の平方根に比例すると考えてよい．(語彙数が2倍になれば，誤り率は1.4倍になる．) これは文献10)，15) と異なる結果であり，実際には理論上とは異なったふるまいをしていることがわかる．

上述の(iii)は，挿入誤りの定義によって，挿入された音韻が必ずしもほかの音韻として誤識別されないことに起因しているかもしれないので，次のシミュレーションを行った．すなわち，挿入誤りが起これば必ずほかの音韻に誤識別されるとした場合で，結果を**表5**に示す．予想されるように表4の結果より数%認識率は低下しているが，脱落誤りが挿入誤りよりも致命的である事実に変わりはない．通常，セグメンテーションに挿入誤りがあっても，ただ音韻区間が分割されるだけで，必ずしもほかの音韻に誤識別されることはなく，先の挿入誤りの定義は現実的なものであると考えられる．

4.2 音韻グループ別の影響

(a) 母音と子音の比較

表 5 音韻識別率，脱落確率，挿入確率の単語認識率に与える影響
—挿入誤りは必ず誤識別されるとした場合—

Table 5 Relationship between word recognition rate and phoneme recognition rate.
—in the case which an insertion error is always misclassified into an other phoneme—

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	10%	99.0%	92.5%	90.0%	88.0%
80	0	10	98.0	97.0	95.0	94.0
90	0	10	100.0	99.5	99.5	99.0
70	5	10	97.0	89.5	86.5	81.5
80	5	10	98.0	94.5	90.0	88.0
90	5	10	100.0	98.0	97.5	96.0
70	10	10	90.0	81.0	79.0	75.0
80	10	10	96.0	91.0	88.5	84.0
90	10	10	96.0	96.0	93.5	92.5

日本語の場合，母音と子音の出現頻度はほぼ等しく，母音の方が音響的に安定で識別し易いことから，音声自動認識では，母音の認識が重要だと考えられている．ここでは，言語の音韻配列上の観点から，音声自動認識の母音と子音の重要度について考察する．シミュレーションの方法は前節と全く同じで，母音・子音別に音韻識別率，脱落確率，挿入確率を与えた．ただし，母音の脱落誤りや子音の挿入誤りは，現実にはあまり起こらないので，この点を留意した．結果を**表6**に示す．表から，母音と子音の重要度については，子音の方がやや重要であるが大差はないことがわかる．表4との比較から，子音の脱落確率が10%，20%の場合と全音韻の脱落確率が5%，10%の場合と単語認識率に大差はないこと，母音の挿入確率が10%，20%の場合と全音韻の挿入確率が5%，10%の場合と単語認識率に大差はないことがわかる．

(b) 同一調音様式音韻群を同一カテゴリにした場合

音韻認識のうちで最も困難なものは，同一調音様式をもつ音韻間の識別である．そこで，これら相互の識別をしないで同一カテゴリとして扱った場合の単語認識率の影響を調べた．ここでは，鼻子音 /m, n, ŋ/，有声破裂音 /b, d, g/，無声摩擦音 /s, c, h/，無声破裂音 /p, t, k/ をそれぞれ同一カテゴリとしてシミュレーションした．結果を**表7**に示す．表4との比較から，相対的に単語の認識率は低下することがわかる．特に脱落誤りのある場合の認識率低下が著しい．ただし，実際にはこれらの音韻を同一カテゴリとしたことにより，音韻識別率が上昇するので，表4と表7とは直接比較できず，認識率の低下は表7よりも少ないと考え

表 6　母音と子音の認識率の単語認識率に及ぼす影響

Table 6　Relationship between vowel/consonant recognition rate and word recognition rate.

識別率		脱落率		挿入率		20単語	50単語	100単語	200単語
母音	子音	母音	子音	母音	子音				
90%	70%	0%	0%	0%	0%	99.0%	98.0%	98.0%	97.0%
70	90	0	0	0	0	100.0	99.0	98.5	98.5
90	70	5	5	0	0	100.0	98.0	97.0	96.5
70	90	5	5	0	0	100.0	97.5	97.5	96.0
90	70	10	10	0	0	96.0	95.5	94.0	89.5
70	90	10	10	0	0	97.0	96.5	94.5	94.0
90	70	0	0	5	5	100.0	98.0	98.0	96.5
70	90	0	0	5	5	100.0	98.0	98.0	96.5
90	70	5	5	5	5	95.0	94.5	92.5	91.0
70	90	5	5	5	5	95.0	96.0	94.5	93.0
90	70	10	10	5	5	96.0	93.0	92.0	87.5
70	90	10	10	5	5	96.0	94.0	91.0	88.5
90	70	0	0	10	10	100.0	98.0	97.5	96.5
70	90	0	0	10	10	100.0	98.5	98.0	96.5
90	70	5	5	10	10	99.0	97.5	95.0	91.0
70	90	5	5	10	10	96.0	95.0	93.0	89.0
90	70	10	10	10	10	96.0	94.0	90.5	85.0
70	90	10	10	10	10	97.0	93.0	91.0	86.0
70	70	0	10	0	0	93.5	93.5	90.5	89.0
80	80	0	10	0	0	98.0	97.5	95.0	93.0
90	90	0	10	0	0	99.5	99.5	99.5	99.5
70	70	0	20	0	0	93.0	85.0	83.0	80.0
80	80	0	20	0	0	96.0	96.0	94.0	92.0
90	90	0	20	0	0	99.0	99.0	98.5	98.5
70	70	0	0	10	0	97.0	95.5	93.5	91.0
80	80	0	0	10	0	99.0	98.5	98.0	98.0
90	90	0	0	10	0	100.0	100.0	99.0	99.0
70	70	0	0	20	0	96.0	93.5	89.5	89.0
80	80	0	0	20	0	100.0	99.5	99.5	98.0
90	90	0	0	20	0	100.0	100.0	100.0	100.0

表 7　同一調音様式音韻群を同一カテゴリとした場合

Table 7　In the case which phonemes being the same manner of coarticulation are regarded as a same category.

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	0%	98.0%	94.5%	94.0%	90.5%
80	0	0	99.0	97.0	96.0	94.0
90	0	0	100.0	99.5	99.0	97.0
70	5	0	89.0	86.5	83.5	77.0
80	5	0	98.0	93.5	91.5	88.5
90	5	0	100.0	98.5	96.0	94.0
70	0	5	97.0	91.0	90.0	87.0
80	0	5	98.0	97.0	95.0	92.5
90	0	5	100.0	98.0	98.0	97.5
70	5	5	89.0	83.5	79.0	74.5
80	5	5	96.0	93.5	92.5	88.0
90	5	5	97.0	96.0	92.5	91.0
70	10	10	92.0	84.0	80.0	74.0
80	10	10	96.0	89.0	85.5	80.5
90	10	10	98.0	96.0	92.5	90.0

られる．

（c）　母音／有声子音／無声子音の弁別率の影響

これまでは，音韻置換確率は音韻によらず一定としてきたが，実際には，母音は母音相互に，子音は子音相互に誤識別されやすい．このように，音韻置換確率に偏りのある場合の影響を調べたのが **表 8**(a)(b) である．表 8 (a) は母音と子音の弁別が完全にできた場合を，表 8 (b) は母音／有声子音／無声子音の弁別が完全にできた場合を示している．表 8 (a) から，母音と子音間の弁別率に偏りがあると単語認識率が多少低下することがわかる．一方，表 8 (b) から，母音，有声子音，無声子音間の弁別率に偏りがあると単語認識率は多少向上することがわかる．これは，母音――子音の音素対の誤認識はあまり単語間の混同につながらないこと，逆に母音――母音，有声子音――無声子音の音素対の誤認識は単語間の混同を生じることを示唆している[12]．これは，日本語が母音――子音連鎖の言語であることに起因している．

（d）　語頭の有声子音や語中の半母音の認識が不可能である場合

表 8 母音，有声子音，無音子音の弁別率の影響
Table 8 Relationship between word recognition rate and classification rate among vowel/voiced consonant/unvoiced consonant.

(a) 母音と子音の弁別が完全な場合

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	0%	97.0%	96.5%	95.5%	92.5%
80	0	0	99.0	98.5	98.5	97.5
90	0	0	100.0	99.0	99.0	99.0
70	5	5	96.0	90.5	88.0	84.0
80	5	5	99.0	95.0	94.5	91.5
90	5	5	100.0	98.5	96.5	94.0
70	10	10	93.0	90.5	86.0	81.5
80	10	10	99.0	95.0	93.5	89.5
90	10	10	99.0	96.0	93.5	91.0

(b) 母音／有声子音／無声子音の弁別が完全な場合

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	0%	99.0%	97.5%	97.5%	95.5%
80	0	0	100.0	100.0	99.0	98.5
90	0	0	100.0	100.0	99.5	99.0
70	5	5	97.0	96.0	96.0	92.5
80	5	5	100.0	96.5	96.5	93.5
90	5	5	100.0	97.5	97.5	96.0
70	10	10	98.0	93.5	91.5	87.0
80	10	10	99.0	94.5	91.5	89.0
90	10	10	100.0	97.5	94.5	91.5

語頭の音韻の認識は単語音声の認識には重要ではあるが[10]，実際の音声認識では，語頭の有声子音や半母音，無声破裂音の検出や語中の半母音の識別は難しい．上述の語頭の脱落確率を100%，語中の半母音の識別率を0%にした極端な場合をシミュレーションした．その結果，200単語の認識率は3～5%低下することがわかった．しかし，実際には，語頭の有声子音などの検出率は0%ではないので，1～2%の低下にとどまると思われる．

4.3 単語辞書と音韻認識結果の表現方法の影響

前節までは，原則として各音韻は1つの音韻として認識されるものとしてきた．しかし，実際の音声では母音の継続時間長は子音に比べて長く，また音響的に定常であるから，短時間の音響パラメータで認識されることが多い．4.2(a)節の結果からは，母音と子音についての言語の配列上での重要度には差異はなかった．しかし，現実問題として，母音の方が認識しやすいこと，継続時間が長いこと等の理由で，子音よりも重要である．1音韻区間を複数個の音韻記号列で表現することは，1音韻区間を複数個の音韻（第1候補，第2候補，…）で表現する[11]のと同様に，より多くの情報量を保存していると予想される．このことをシミュレートしたのが表9(a)，(b)である．この場合は，単語辞書の母音も複数個で表現した．表4と比較すれば明らかなように，認識率は格段に向上している．また，1母音を2連続母音記号で表現するよりも3連続母音記号で表現する方が認識率が高い．4連続母音記号で表現すれば，さらに認識率が向上するかどうかは明らかでない．しかし，1母音を3連続母音として認識する場合は，現実には，3個とも同じ音韻として認識されることが多く，表10(b)のような認識率の向上は見られないと思われる．

表 9 1母音を複数個の音韻記号で表現した場合
Table 9 In the case which a vowel is represented by some phoneme symbols.

(a) 1母音を2連続母音で表現した場合
〔例〕 /青森/→/aaoomoorii/

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	0%	100.0%	98.0%	98.0%	97.5%
80	0	0	100.0	99.5	99.5	99.5
90	0	0	100.0	100.0	100.0	100.0
70	5	5	97.0	94.5	93.5	92.0
80	5	5	100.0	98.5	98.5	98.0
90	5	5	100.0	99.0	98.5	98.5
70	10	10	96.0	92.0	89.5	87.5
80	10	10	99.0	93.0	91.5	88.5
90	10	10	100.0	98.0	97.0	95.5

(b) 1母音を3連続母音で表現した場合
〔例〕 /青森/→/aaaooomooriii/

識別率	脱落率	挿入率	20単語	50単語	100単語
70%	0%	0%	100.0%	98.0%	98.0%
80	0	0	100.0	100.0	100.0
90	0	0	100.0	100.0	100.0
70	5	5	100.0	97.0	96.0
80	5	5	100.0	99.0	98.0
90	5	5	100.0	99.0	99.0
70	10	10	97.0	96.0	95.0
80	10	10	100.0	99.0	98.0
90	10	10	100.0	99.0	99.0

4.4 現実システムに即したシミュレーション

本節では，現実システムに即したシミュレーションを行う．すなわち，音響的に定常で継続時間の長い母音と無声摩擦音は2連続音韻記号で表現し，同一調音様式をもつ音韻群は同一カテゴリとして扱う．さらに，母音，有声子音，無声子音の音韻識別誤りがあった場合は，その80%がそれぞれ母音，有声子音，無声子音に誤識別されるとした（認識アルゴリズムは前節までと同じ）．結果を表10に示す．現在の音韻認識技術では，不特定話者に対しては音韻認識率は（/*m, n, ŋ*/, /*b, d, g*/, /*s, c, h*/, /*p, t, k*/を同一カテゴリとした場

表 10 現実システムに即したシミュレーション結果
Table 10 Simulation results based on a real system.

識別率	脱落率	挿入率	20単語	50単語	100単語	200単語
70%	0%	0%	99.0%	99.0%	99.0%	97.5%
80	0	0	100.0	99.5	99.0	98.5
90	0	0	100.0	100.0	100.0	100.0
70	5	5	98.0	96.0	94.0	89.5
80	5	5	100.0	96.5	95.5	94.0
90	5	5	100.0	99.0	99.0	99.0
70	10	10	96.0	93.5	90.5	86.5
80	10	10	99.0	96.5	94.5	91.0
90	10	10	100.0	99.0	99.0	97.0
70	20	20	96.0	89.0	86.0	—
80	20	20	96.0	99.0	89.0	—
90	20	20	98.0	96.0	94.0	—

合) 約70～80%, 特定話者に対しては約80～90%, 脱落確率・挿入確率は約5～10%であることを考えると, 不特定話者に対しては50～100単語, 特定話者に対しては100～200単語程度なら, 音韻認識に基づく方法で95%以上の単語認識率が得られるものと期待できる. 我々が先に開発した(本論文の認識アルゴリズムと異なるが)単語音声認識システム[22]や単語音声認識装置[23], 本論文の認識アルゴリズムと類似したほかのシステム[24]は, これらのシミュレーション結果の妥当性を裏付けている.

5. むすび

単語音声自動認識システムにおける音韻認識部をシミュレートすることによって, 音韻識別率, 脱落確率, 挿入確率等と単語認識率との関係を調べた. その結果挿入誤りよりも脱落誤りが単語認識率に与える影響が大きいこと, 母音と子音の識別が単語認識率に与える影響に大差がないこと, 同一の調音様式をもつ音韻群を同一カテゴリとして識別した場合, 単語認識率は数%低下すること, 母音と子音の弁別よりも有声音と無声音の弁別の方が重要であること, 音響的に安定な継続時間の長い音韻は, 複数個の音韻記号列として識別する方がよいことなどがわかった. また, 現在の技術レベルでは, 音韻認識に基づく単語音声認識では, 不特定話者に対しては50～100単語, 特定話者に対しては100～200単語程度なら95%以上の単語認識率が得られることがわかった. (ただし, 本論文で対象とした都市名はランダムに選ばれた名詞よりも多少認識し易いと思われる[11].)これらの誤り率を半分にすることや対象語彙数を2倍にすることは, 精力的に進められている音韻認識研究の成果により近い将来可能になるものと思われる.

本論文では, シミュレーションの簡単化のために, コンフュージョン・マトリックスは原則として偏りのない場合を考察し, さらに音韻認識誤りはコンテキストに独立であるとしたが, 実際にはかなり依存していると思われる. これらのコンテキストに依存する誤りは, 事前知識として利用することができるから, 実際には本論文で得られた結果よりも高い単語認識率が得られるものと考えられる. (しかし, 本論文の相対的な比較による結論は不偏的なものである.) さらに, 実際のシステムでは, 継続時間長や音韻認識結果の第1候補ばかりでなく第2候補や信頼度も利用できるため[21),26)], ここで得られた単語認識率は, 音韻認識をベースとする方法の下限値を示すものと考えられる.

謝辞 本論文は, 筆者が京都大学在職中に坂井利之教授のご指導で行った研究[25]を継続発展させたものである. 日頃からご指導いただく坂井利之京都大学教授に感謝申し上げます.

参考文献

1) Rabiner, L. R. et al.; Speaker-Independent Recognition of Isolated Words Using Clustering Techniques, IEEE Trans. Vol. ASSP-27, No. 4, pp. 336-349 (1979).
2) 新美康永: 限定単語認識システムにおける標準パターンについて, 日本音響学会春季大会, 4-2-18 (1976).
3) 松本, 脇田: Frequency Warping による話者正規化, 日本音響学会音声研資, S 79-25 (1979).
4) 小原, 三輪, 牧野, 城戸: 非線形スペクトルマッチングによる単語音声認識の一方式, 電子通信学会パターン認識と学習技報, PRL 79-46 (1979).
5) 中川, 神谷, 坂井: 音声スペクトルの時間軸・周波数軸・強度軸の同時非線形伸縮に基づく不特定話者の単語音声の認識, 電子通信学会論文誌, Vol. 64-D, No. 2, pp. 116-123 (1981).
6) 中川, 白方, 山尾, 坂井: 不特定話者の音声自動認識のための性別・年齢差による話者分類の考察, 電子通信学会論文誌, Vol. 63-D, No. 12, pp. 1002-1009 (1980).
7) Denes, P. B.: On the Statistics of Spoken English, J. Acous. Soc. Am. Vol. 35, No. 5, pp. 892-904 (1963).
8) 板橋, 城戸: 弁別特徴で表示した単語間の混同について, 日本音響学会春季大会, 2-2-6(1970).
9) 板橋, 鈴木, 城戸: 単語中の幾つかの子音の辞書による識別, 電子通信学会論文誌, Vol. 54-C, No. 1, pp. 10-17 (1971).
10) 牧野, 城戸: 近距離単語間の識別に必要な音素対の性質, 電子通信学会論文誌, Vol. 62-D, No. 8, pp. 507-517 (1979).

11) 牧野，城戸：単語間の識別に必要な音素の性質，電子通信学会電気音響技報，EA 77-49 (1978).
12) 横山晶一：母音の無声化を考慮した単語間の距離について，日本音響学会春季大会，4-1-6 (1978).
13) 横山晶一：国語辞典における最小対立語について，日本音響学会秋季大会，1-1-7 (1980).
14) 上坂，黒木：文字の結合確率の言語認識への利用について，NHK 技術研究，Vol. 17, No. 6, pp. 58-66 (1965).
15) 阿部，泰野，福村：辞書を利用する文字認識系の能力の評価，電子通信学会論文誌，Vol. 52-C, No. 6, pp. 305-312 (1969).
16) 牧野，鈴木，城戸：推移確率の利用による音韻系列の訂正，日本音響学会音声研究資，S 74-02 (1974).
17) Jelinek, F.: Continuous Speech Recognition by Statistical Methods, Proceedings of IEEE, Vol. 64, No. 4, pp. 532-556 (1976).
18) Vives, R. and Gresser, J.-Y.: A Similarity Index Between Strings of Symbols Application to Automatic Words and Language Recogni. tion, Proceedings of the First IJCPR, pp. 308-311 (1973).
19) 関口，重永：グラフを利用して誤りを含む記号列を分類する方法とその音声認識への応用，情報処理，Vol. 19, No. 9, pp. 831-838 (1978).
20) Tappert, C. C.: Experiments with a Tree Search Method for Converting Noisy Phonetic Representation into Standard Orthography, IEEE Symposium on Speech Recognition, CMU (1974).
21) 中川，坂井：音声自動認識に関する情報工学的諸考察，情報処理学会論文誌，Vol. 21, No. 5, pp. 407-417 (1980).
22) 中川，内海，坂井：単語音声の大局・局所両特徴による前照合法と個人差の学習法，日本音響学会音声研資，S 78-23 (1978).
23) 中川，坂井，坪香：音韻認識に基づく不特定話者向き単語音声汎用認識装置，日本音響学会音声研資，S 80-62 (1980).
24) 川端，牧野，三輪，城戸：音素の信頼度を利用した単語音声認識，日本音響学会春季大会，1-7-13 (1980).
25) 中川，水上，坂井：音韻認識率と単語識別率との関係，電子通信学会電気音響技報，EA 79-87 (1980).
26) 中本，中川：単音節単位の入力に基づく単語音声の認識，電子通信学会パターン認識と学習技法，PRL 81-8 (1981).

(昭和55年12月23日受付)
(昭和56年 3 月20日採録)