論　文

音素文脈依存モデルと高速な探索手法を用いた連続音声認識

非会員　伊藤　克亘[†]　　正　員　速水　悟[††]　　正　員　田中　穂積[†]

Continuous Speech Recognition by Context-Dependent Phonetic HMM and an Efficient Algorithm for Finding N-Best Sentence Hypotheses

Katunobu ITOU[†], *Nonmember*, Satoru HAYAMIZU[††] *and* Hozumi TANAKA[†], *Members*

あらまし　本論文では，複数の候補を出力する連続音声認識の手法について提案する．この手法を用いると，フレーム同期で連続音声のVQコード列から近似的にN-Bestな音韻系列・形態素(単語)系列を自動的に構成することができる．この手法の第1の特徴は，音韻モデルとして，音素文脈依存隠れマルコフモデル(HMM)を用いることである．第2は，音韻レベルでの処理を統合していることである．この手法は，異なる仮説の同じ音韻の部分は，近似的に最大のスコアの仮説の経路を用いる．音韻照合の処理量は，音韻モデルの数にだけ依存し，途中で考慮する候補数や語い数・文法規則数には依存しないので大語いを対象とする場合に，より効果的である．本手法の性能を，学習に用いていない成人男性10名の文音声に対して，113語の語いからなる辞書と文節の平均分岐数が8.2の文法を用いて評価した結果，1,024個のモデルを用いた場合に，文認識率97.3％が得られた．音素文脈依存モデルを使わない場合に比べて，77％の誤りを低減することができ，本論文で提案する手法が連続音声認識に有効であることがわかった．

キーワード　音声認識，連続音声，音声文脈モデル，N-Bestサーチ，HMM

1. まえがき

音声による自然言語を用いたユーザインタフェースを考えるとき，大語いで連続音声が扱えることが必須条件となる．

日本語には，語順が自由であるという特徴がある．そのため，大語いを扱うシステムでは，単語・形態素を認識の単位とすると，次に発話されるものを予測するのは非常に困難なので，常に多くの標準パターンを駆動しなければならない．

従って，日本語による自然言語インタフェースでの音声認識方式としては，認識の単位を音韻などのレベルにとり，それらの組合せで形態素や文節などのより大きい単位を構成していく方式[(1),(2)]が有望であると考えられる．

音韻モデルの推定精度をあげる手法として，音素文脈依存モデルを用いる方法がある[(3)~(6)]．しかし，音素文脈依存モデルを連続音声認識に用いる場合，あらかじめ音素文脈を決められない形態素(単語)の境界での扱いが問題になる．本論文で提案する手法では，あらかじめ音素文脈が決定できる形態素(単語)の内部にある音韻は，あらかじめ音素文脈依存モデルを割り当てておく．一方，形態素(単語)の境界にある音韻は，動的に決まる音素文脈によってモデルを割り当てる．

音韻などの単位を組み合わせて形態素・文節・文などを構成するシステムを，音韻ラティスや単語ラティスなどの中間表現を使わずにフレーム同期で動作させようとすると，処理途中に仮説の数が爆発的に生じてしまう．この場合に，ビームサーチを導入すると，仮説の数をある程度抑えられる．しかし，ビームサーチは，ビーム幅が小さすぎると正解の仮説が処理途中に枝刈りされ，正解が得られなくなる可能性が高まる．従って，高い精度を得るためには，なるべくビーム幅

† 東京工業大学工学部情報工学科，東京都
Faculty of Engineering, Tokyo Institute of Technology, Tokyo, 152 Japan
†† 電子技術総合研究所知能情報部，つくば市
Electrotechnical Laboratory, Tsukuba-shi, 305 Japan

を大きくした方がよい．そこで，システムとしては，ビーム幅を大きくしても処理量がそれほど増えないようにしなければならない．

そのためには，なるべく無駄な再計算を避ける工夫がさまざまなレベルで必要となる．本論文では，そのうち特に処理量が多い，音声のパラメータ系列と音韻モデルを照合して音韻系列の仮説をつくる部分での効率化について考える．

こういった効率化の一つの手法として，文法や辞書のうち，予測される部分だけを動的に展開して，その部分だけで音韻系列の仮説をつくる手法が提案されている[7]．しかし，先に述べたように，単語や形態素のレベルで次に発話されるものがあり得るかあり得ないかという予測をするのは，日本語で未知語を扱おうとしたり，大語いを扱おうとするシステムでは現実的な手法でない．しかし，探索空間を限定する手法をとらない場合，音韻モデルの照合の処理量は，辞書に登録されている語数や文法的な規則の数(ビームサーチを用いる場合には，ビーム幅)に依存して多くなる[8]．

本論文では，これらの問題点を解決するために，異なる仮説に同じ音韻モデルが現れるときには，どれか一つの照合結果を代表させ，他の仮説ではその結果を利用して仮説をつくる方法を提案する．この方法では，あるフレームでの音韻モデルの照合のための処理量は，音韻モデルの数にのみ依存するので，仮説の数が増えても処理量が増えることはない．

我々は，これらの手法を取り入れた日本語による自然言語インタフェースを目指すシステム niNja (Natural language INterface in JApanese) の試作を行っている．本論文ではそのシステムを用いて提案する手法の有効性を評価するための認識実験の結果についても報告する．

はじめに，2.でシステムの全体的な構成について述べる．そして，3.で連続音声認識システムに音素文脈依存モデルを導入する手法について提案し，4.では，音韻レベルの処理の統合について述べる．そして，これらの手法の有効性を評価するために行った実験について5.で述べる．最後に，6.で結論を述べる．

2. システムの構成

本システムでは，各階層の履歴を木構造で保持する[9](図1)．この構造を利用して，それぞれの階層の情報に対する制約の評価は，同じ節点に対して1度行うだけで，それ以降にその履歴をもつ仮説が生成された場合には，過去の結果を利用する．従って，無駄な再計算を避けられる．

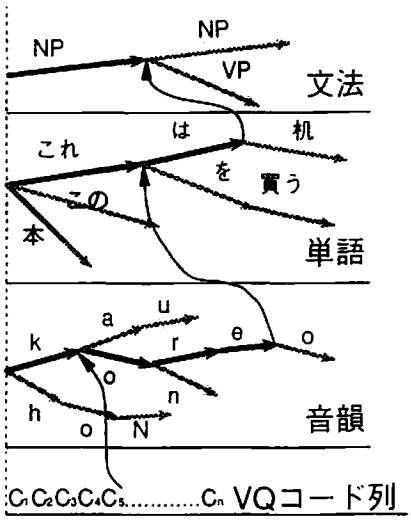

図1 各階層間の関係を保持する木構造
Fig. 1 Tree structures to keep relation between hierarchies.

システムの動作を簡単に説明する．まず，VQコードを音韻モデルと照合する．音韻モデルの照合を終えると，辞書を参照して次に照合する音韻モデルを決める．辞書引きプロセスで単語(形態素)が完成したら，その構文的なカテゴリーを構文解析プロセスに返す．構文解析プロセスは，辞書引きプロセスから受け取った構文カテゴリーに応じて，次の辞書引きプロセスを駆動する．これらの動作を，フレームごとに行う．

このように処理を進めると，例えば，/kore/という発話のはじめの方の処理で，t_i フレームで照合を終了した/ko/と t_{i+1} フレームで照合を終了した/ko/と t_{i+2} フレームで照合を終了した/ko/では，すべて音韻照合のスコアが異なる．本システムでは，これらの仮説をすべて別々に仮説セルとして保持する．仮説セルは，そのフレームまでの音韻照合の結果得られる生起確率の対数をとったものをスコアとして保持する．また，その仮説セルが各階層でどのような履歴をもつかを，各階履の木構造での該当する節点へのリンクとして保持する．

このように，音韻ラティスや単語ラティスなどの中間的な構造を介さず，VQコード列から直接認識結果を作る．従って，情報の欠落が防げ，高精度な処理を行うと期待できる．しかし，このような処理方法を採用すると，処理を進めると共に仮説セルが爆発的に作られるという問題が生じる．

このような問題を避けるため，本システムではビー

ムサーチを導入する．しかし，1.で述べたように，単にビームサーチを導入するだけで，処理量を十分に減らそうとすると非常に精度が悪くなる．本システムでは，以下で述べるように，音韻モデルの精度の向上のために音素文脈依存モデルを導入し，同じビーム幅のときの処理量の削減のために音韻レベルでの処理の統合を行う．

このシステムの特徴は，日本語の構文的な制約を辞書と文法（文節内文法と文節間文法は区別しない）の二つのレベルでとらえている点である．音声認識システムで用いる構文的な制約を二つのレベルでとらえる手法として，2段階構文解析法[10]が提案されている．2段階構文解析法では，文節間文法と文節内文法の二つのレベルをもつ．本手法で，構文的な制約を辞書と文法の二つのレベルに分けた理由を以下に述べる．

一般化LR構文解析法は，基本的に次の終端記号を先読みして動作を決定することで，無駄な処理を行わない特徴をもつ．しかし，例えば，

名詞→h o N（本）

のように，辞書を音韻を終端記号にした規則とみなすと，一般化LR構文解析法でこの規則の辞書引き処理を完了するためには，/N/の次の音韻を先読みしなければならない．しかし，未知語や自由な語順を許した場合，結局，ほとんどすべての音韻を先読みすることになり，先読みが無意味になる．このような理由から，辞書レベルのLR表については，レデュースのときに先読みを行わないようにする[11]．つまり，先の例では，終端記号/N/を処理したところでレデュースすることになる．このように変更した辞書レベルのLR表をLR辞書と呼ぶ．LR辞書を用いると，2段階構文解析法では先読みの扱いについて大きく変更する必要のある音素文脈依存モデルの導入・未知語の扱いが簡単になる利点がある．

3. 音素文脈依存モデルの導入

3.1 音素文脈依存モデル

よく知られているように，同じ音素に相当する音声の音響的特性は，さまざまな要因によって変動する．音響的特性の変動のうち，その音韻のおかれた音素文脈（前後の音素の並び）による変動を音素文脈ごとに異なる音韻モデルによって表すものが，音素文脈依存モデルである．

近年，音声認識における音素文脈依存モデルの有効性が知られるようになってきた[3]~[6]．本論文では，文献(4)の手法を用いて生成した音素文脈依存HMMを用いる．この手法では，音素文脈を先行音素または後続音素の調音特徴に基づいて決定木を用いて分類し，あらかじめ決められたモデル数の音素文脈依存音韻モデルを得ることができる．

しかし，音素文脈依存モデルを連続音声認識に用いる場合，形態素（単語）の境界にある音素では，あらかじめ音素文脈が決まらないので，その扱いについて考えなければならない．本論文では，音素文脈があらかじめ決まっている形態素内の音素については，はじめから音素文脈依存音韻モデルを割り当てておき，形態素の境界の音素については，動的に音素文脈依存音韻モデルを割り当てる手法を用いる．以下，その手法について説明する．

3.2 音素文脈依存モデルの導入

前述したとおり，本システムでは構文的な制約を辞書と文法の二つのレベルでとらえる．LR辞書の終端記号は，（音素文脈独立な）音韻であり，文法のレベルでは音韻が出現しないので，本システムでは，辞書のレベルの処理を変更するだけで，音素文脈依存モデルを扱える．辞書中の，認識前に音素文脈が決まる音韻については，あらかじめ文脈に応じた音韻モデルを割り当て，認識中にしか決まらない音韻については，辞書引きプロセスを拡張して動的に割り当てる．

ここで，永井らのLR法を用いたシステムに音素文脈依存モデルを導入する手法[12]との違いを述べる．文献(12)では，文節内文法に対するLR表を音素文脈依存モデルを終端記号として書き換える手法をとっている．このような場合，文節境界以外でレデュースするときの状態数が非常に大きくなってしまう．それに対して，本手法では，そういった部分はすべて動的に扱うのでLR辞書の状態数が爆発することはない．

3.2.1 辞書の変換

まず，音素文脈独立な音韻表記の辞書を用意する（図2）．次に，音素文脈ごとの音素と音素文脈依存モデルの対応表を用意する（表1）．この表に従って辞書を変換する（図3）．このとき，辞書の項目の最初の音素と最後の音素については，辞書のレベルでは音素文脈が決まらないので，モデルには変換せず，決定している音素文脈だけ記録しておく．/h(*,o)/は，後続の音素が/o/であり，先行の音素が決定していない音韻/h/を表し，/N(o,*)/は先行の音素が/o/であり，後続の音素が決まっていない音韻/N/を表す．

このように辞書を変換したら，あとは，通常と同じ

名詞 → h o N
名詞 → z a q sh i

図2 システムで用いる辞書の例
Fig. 2 Examples of lexicon.

名詞 → h(*,o) o10 N(o,*)
名詞 → z(*,a) a4 q0 sh5 i(sh,*)

図3 音素文脈依存モデルに変換された辞書の例
Fig. 3 Examples of lexicon with context dependent HMM.

方法でLR辞書をつくる(表2). 終端記号の種類が増えるので, LR辞書は大きくなるが, モデルの数を増やしても, 状態数はたかだか辞書項目の音素数の総和になる程度である.

3.2.2 辞書引きプロセスの変更

仮説セルは音素レベルの履歴から, それぞれの音素文脈を参照できる. この音素文脈を利用して, 音素文脈モデルが扱えるように辞書引きプロセスを変更する. ここでは,「ここに本があります.」という発話の「本」の部分を例にとって説明する.

(1) 辞書引きプロセスを駆動するとき

辞書引きプロセスを駆動する仮説セルは, それぞれ次にくる音素を決定している. /kokoni/(ここに)の/i/の場合, モデル数が128個の場合には, /i(n, *)/に対して, 五つのモデルが割り当てられる(表3). これらのモデルはそれぞれ, 表3に示された可能な複数の後続音素をもつ. 従って, /kokoni/には, 最後の/i/に対するモデルが異なる五つの仮説セルがあり, それぞれ, 次にくる音素を決定している.

次に, 駆動するLR辞書の開始状態(状態番号が0の状態)を見る. このとき, /i17/をもつ仮説セルの場合は, 後続音素として/h/を許しているので, LR辞書で示される音素文脈/h(*,o)/を/h(i,o)/とみなして, 表1で示した対応表から/h3/を駆動する. そして, そのモデルの照合が終了したら, LR表に従って状態1に進む. 同様に, /i13/をもつ仮説セルは後続音素として/z/を許しているので, 状態3に進むことができるが, その他の仮説セルはここで捨てられる.

(2) 開始状態以外の状態でのシフトのとき

先読み記号が音素文脈依存モデルなので, そのままそのモデルを駆動して次の状態にすすむ. 例えば, (1)で/h3/を駆動して作られた仮説セルでは, 状態1で/o2/を駆動して状態2に進む.

表1 音素文脈と音素文脈依存モデルの対応表(一部)

音素文脈	モデル
h(i,o)	h3
o(h,N)	o10
a(z,q)	a4
q(a,sh)	q0
sh(q,i)	sh5

表2 辞書のLR表の例(一部)

状態	h(*,o)	z(*,a)	o10	N(o,*)	a4
0	sh1	sh3			
1			sh2		
2				re0	
3					sh4

表3 i(n, *)に対する音素文脈依存モデルの割当て

モデル	後続音素
i5	a, a-, e, e-, i, i-, o, o-, u, u-
i7	ch, f, k, ky, p, py, ry, s, sh, t, ts
i13	b, by, d, dy, j, r, w, y, z
i15	N, g, gy, m, my, n, ny
i17	#, h, hy, q, >

(3) レデュースのとき

レデュースのときは, 次にくる音素が決まらないので, 先行音素だけから決まる可能な音素文脈依存モデルをすべて駆動する. 例えば, 図2の状態2では, /N(o, *)/となっているので, この音素文脈に対して許されるモデルの数に応じて仮説セルを作る((1)参照).

このような手法では, 文献(12)の手法と比べると辞書に登録されている項目数が少ないとき, LR辞書のレデュースのところで, 無駄に駆動される音韻モデル数が多くなってしまう. しかし, 辞書に登録される項目数が増えれば, その差は小さくなる.

4. 音韻レベルの処理の統合

4.1 従来の手法とその問題点

One Pass DP法[13]は, 有限状態オートマトンによる比較的単純な制御構造をもち, 入力音声のフレームに同期して処理が進む特徴をもつ. このため, 多くの連続単語認識システムで, この手法に基づいた手法が用いられている[8],[14],[15].

本システムでは, 文脈自由文法で表現される文法を動的に展開して, 有限状態オートマトンと等価なネットワークとみなして, One Pass DPアルゴリズムを用いるが, 大規模な辞書を構築した場合にはその状態

数が非常に多くなってしまう．有限状態オートマトンで駆動するOne Pass DP法を用いてN-Bestな解を得る方法としては，文献(8)がある．この手法では，各候補について別々に音韻を照合するので，途中の候補数をmとすると，必要な照合量は$O(m)$と，非常に多くなる．

音韻を認識の単位とするシステムでは，DP法を制御するためのネットワークは，音韻レベルでのネットワークになる．本システムの仮説セルは，このネットワークの状態に対応し，一つずつが異なった音韻モデルの系列を表している．この音韻系列には，次の特徴がある．(1)語い数や規則数が増えても，音韻系列に現れる音韻モデル数はある一定数である．(2)異なる音韻系列に同じ音韻モデルが何度も現れる．

本論文で提案する手法では，この二つの特徴に着目して，同じフレームでの同じ音韻モデルに対する照合は，ある仮説セルに対してだけ行い，他の仮説セルには近似的にその結果を使う．この手法では，理論的な最適性は失われることもある[1]が，照合のための処理量は語い数や規則数に依存しない量で，N-Bestな解を求めることができる．従って，従来のN-Bestな解を求める手法よりも処理量が少ない．

単語レベルで照合の処理を統合する手法として文献(16)があるが，認識途中の探索経路を考えると，本論文で提案しているように，音韻レベルの方が同じものがさまざまな探索経路に出現する可能性が高いので，統合の効果も大きく，よりシステムの効率を上げることができる．

以下に，アルゴリズムを示す．

4.2 アルゴリズム

各フレームごとに以下の処理を行う．

(1) そのフレームで，音韻モデル$p_{current}$の最終状態に達している経路については，その音韻での経路の初期状態を見る．各経路の初期状態には，仮説セルの集合が記憶されている．このそれぞれの仮説セルについて，スコアの更新と次に連結する音韻モデルp_{next}(補数のこともある)の決定を行う．

経路のスコアは，(2)で述べるように，初期状態での仮説セルの集合の要素のうち，最もスコアの良い仮説セルのスコアである．初期状態で仮説セルの集合は，最大のスコアをもつ仮説セルとの差を保持している．従って，その分を経路のスコアから引いたものを現在のスコアとする．

次に連結する音韻モデルp_{next}は以下のようにして決める．LR辞書での動作がシフトのときは，表が示す次の状態に遷移する．動作がレデュースのときは，辞書引き結果(構文カテゴリー名を返す)に従って，文法レベルのLR表が動作して次の辞書引き処理を始めて，LR辞書の開始状態に遷移する．

これらの動作のあと，それぞれ遷移した状態で先読みとなっている音韻モデルをp_{next}とする．

(2) それぞれのp_{next}について，(1)の処理を終えた仮説セルの集合を作る．その集合の中で，最大のスコアをもつ仮説セルのスコアを，そのフレームでの初期状態でのスコアとする．他の仮説セルは，最大のものとのスコアの差を保持しておく．

(3) $p_{current}$の最終状態以外の状態については，ビタビサーチを行う．但し，(2)で示した設定方法なので，フレームごとに初期状態に設定される過去の経路が違う．従って，最適でない探索が行われることもある[1]．

各フレームで各音韻モデルを1度しか照合しないので，フレームごとの音韻照合の処理量は，仮説セルの数にかかわらず，音韻モデル数をnとするとたかだか$O(n)$である．しかし，(1)でのスコアの更新のための処理は，そのフレームでの仮説セル数をmとすると$O(m)$なので，仮説セルの数が増えると処理量が多くなる．

従って，音韻モデルの数をある程度の数に抑える必要があるので，ビームサーチを導入する．具体的には以下のようにする．まず，適当な値$\lambda(0<\lambda<1)$を定める．各フレームごとに，そのフレームの仮説セルの最大のスコアから，λの対数をとった値を引いたスコアを，そのフレームでのしきい値とする．このしきい値以下の仮説セルについては枝刈りする．このような手法をとると，仮説セルの数が一定以下になる保証はないが，仮説セルをソートする必要はなく，枝刈りのための処理量もそれほど多くなくて済む．

5. 認識実験

5.1 音声資料

実験に用いたHMMの訓練用音声資料は単語音声と連続音声からなる．単語音声資料の話者は成人男性5名で，発声用テキストは音韻バランス単語集合WD-II(1,542語)[17]である．連続音声資料の話者は成人男性2名で，発声用のテキストはATR音韻バランス文150文である．これらの収録は簡易防音室で行い，標本化周波数15 kHzでA/D変換を行った．分析のフレーム周期は5 msである．14次のメルケプストラム係数と，そ

の時間方向の変化量，パワーの時間方向変化量の合計29個のパラメータを一つのコード帳にベクトル量子化した．コード帳のサイズは1,024である．

HMMのモデル構造は，すべて4状態3ループとし，各状態から出る弧の組をタイドアークとした．認識実験に用いたモデルの学習は，文献(4)の方法で行った．モデルの数は，43個(音素文脈独立)，128個，256個，512個，1,024個である．

実験に用いた音声資料の発声用テキストは11文(平均文節数は3.0)の疑問文などである．このテキストを成人男性の2名分を防音室で，8名分を計算機室で収録した．これらの話者・テキストはHMMを訓練した資料には含まれておらず，不特定話者・語い独立な実験条件とした．

認識性能の評価は，文認識率と文節認識率で行う．文認識率は，最もスコアの高い候補が，正解の文と同じになる割合である．

文節認識率は，次のようにして求める．まず，認識結果の文節系列と正しい文節系列とのDPマッチングを行い，文節の置換・挿入・脱落誤りの個数を求める．これらの個数から，次の式を用いて文節認識率を計算する．

$$\text{文節認識率}=\frac{\text{全文節数}-\text{置換}-\text{挿入}-\text{脱落}}{\text{全文節数}}$$

5.2 認識実験

以下の二つの実験を行った．

(1) 音素文脈依存モデルの効果

音素文脈依存モデルのモデル数と認識率の関係を調べるために，辞書・文法・枝刈りのためのパラメータλは一定で，モデル数だけを変化させて実験を行った．

この実験には，図4に示すような文テンプレートを11含む文法を用いた．この図で，〈〉で囲まれた名前は非終端記号を表す．総単語数は113，文節の平均分岐数は8.2である．なお，文節間には任意の無音区間を許す．

結果を表4に示す．音素文脈独立モデルのときは，文認識率88.2％，文節認識率94.2％だったのが，モデル数が多くなるのにつれて認識率も向上し，モデル数が1,024個の場合には，文認識率97.3％，文節認識率99.1％になった．誤りは，文節認識率で84％，文認識率で77％低減した．

訓練に用いたデータの違いはあるものの，同じタスクの文献(18)の文認識率76.4％と比較すると，本実験の結果は良好であることがわかる．

<代名詞・場所> <に> <本> <が> <何冊> <あります>

<代名詞・場所> = ここ|そこ|あそこ|どこ
<本> = 本|辞典|マンガ|雑誌|アルバム

図4 文テンプレートの例
Fig. 4 Examples of sentence template.

表4 音素文脈依存モデルのモデル数と認識率

モデル数	43	128	256	512	1024
文認識率(%)	88.2	90.0	92.7	94.5	97.3
文節認識率(%)	94.2	96.4	97.6	97.9	99.1

文 -> 文節
文 -> 文節 文
文 -> 文節 ポーズ 文
文節 -> 名詞
文節 -> 名詞 語尾1
文節 -> 動詞 語尾2

図5 文法規則(一部)
Fig. 5 Examples of grammar rule.

(2) 音韻レベルの処理を統合した効果

音韻レベルの処理の統合の度合(どれだけの処理が省けるか)を調べるために，実験(1)に使った文のうち，一つの文を用いて，各フレームごとの仮説セルの数と実際に照合を行った音韻数を枝刈りのためのパラメータλを変えて測定した．この実験では，辞書は，音韻バランス単語集合WD-I[17]に含まれる単語492単語に実験(1)を用いた11文に含まれる単語22語と付属語12語を加えたものを用いる．文法は，辞書に含まれる単語から作られた文節が任意の順序でいくつかつながったものを文とみなすものを用いる．その一部を図5に示す．この文法は，非常に制限が緩い．文節の平均分岐数は650である．HMMは音素文脈独立モデルを用いた．

この条件で，λを1.0×10^{-3}と1.0×10^{-20}と値を変えて認識実験を行った．結果を図6に示す．

枝刈りの条件を緩くする(λを小さくする)と，認識結果として正解が得られている．認識中に生成された仮説セルの数と音韻照合数について，表5に示す．

それぞれ仮説セルの数は音韻照合数の15倍，33倍とかなり大きい数である．枝刈り条件を緩くすることで，途中で生成される仮説セルの数は，2.8倍になった．しかし，総音韻照合数については1.3倍であり，仮説セル

$\lambda = 1.0 \times 10^{-3}$ のとき

< ch a n o y u g a # k u q ky o- w a #
n o q t e i m a s u k a >

(茶の湯が屈強はのっていますか)

$\lambda = 1.0 \times 10^{-20}$ のとき

< sh a sh i N g a # t a k u s a N #
n o q t e i m a s u k a >

(写真がたくさんのっていますか) (正解)

図6　認識実験結果
Fig. 6　Sentence recognition results.

表5　処理量の変化

	λ	1.0×10^{-3}	1.0×10^{-20}
仮説セル数	合計	262668	723415
	最大	696	3614
	平均	443	1221
音韻照合数	合計	16432	21716

の増加分ほどは増えていない．

この結果から，本論文で提案した音韻レベルの処理を統合する手法を用いると，音韻照合の処理量をかなり削減できることがわかる．また，ビームサーチと組み合わせて用いると，音韻照合の処理量をそれほど増すことなく，ビーム幅を大きくして，認識精度を上げられることがわかる．

6. むすび

不特定話者の連続音声認識の手法について報告した．本論文では，連続音声の VQ コード列からフレーム同期で，音韻系列・単語(形態素)系列を自動的に構成するシステムについて提案した．音韻モデルには，音素文脈依存 HMM を用い，連続音声認識にも有効であることを示した．音素文脈依存モデルを用いた場合，モデル数が 1,024 のときに文認識率 97.3 % が得られた．VQ コード列を音韻モデルと照合して，音韻系列を構成する部分については，処理を統合する手法について提案した．この手法は，照合の処理量については，途中に保持される候補の数に依存しない．

なお，音韻・単語・文法やそれ以上のレベルでの言語情報の導入や，今回実験を行ったものより更に複雑なタスクにおいても，本論文で提案した手法が有効であるかを検討することなどが，今後の課題として挙げられる．

謝辞　本論文で使用した連続音声資料は日本音響学会連続音声データベース調査委員会の研究用連続音声データベースの一部であり，関係各位の御尽力に感謝致します．また，本研究をすすめる上で，熱心に御討論・御指導頂いた東京工業大学田中研究室ならびに，電子技術総合研究所音声研究室の皆様に感謝致します．

文　献

(1) 伊藤克亘，速水　悟，田中穂積："拡張 LR 構文解析法を用いた連続音声認識"，信学技報，SP90-74 (1990-12).

(2) 北　研二，川端　豪，斉藤博昭："HMM 音韻認識と拡張 LR 構文解析法を用いた連続音声認識"，情処学論，31，3，pp. 472-480 (1990-03).

(3) 速水　悟，田中和世，太田耕三："音素片ネットワークによる音声の音響的変動の記述と単語認識実験"，信学論(D)，J71-D，2，pp. 265-273 (1988-02).

(4) 速水　悟，田中和世："木構造音韻モデルによる未知音素文脈中の音響的変動の予測と評価"，信学技報，SP90-64 (1990-12).

(5) Lee K. F. : "Automatic Speech Recognition : "The Development of the SPHINX System", Kluwer Academic Publishers (1989).

(6) Schwartz R., Chow Y., Kimball O., Roucos S., Krasner M. and Makhoul J. : "Context dependent modeling for acoustic phonetic recognition of continuous speech", Proc. ICASSP-85, pp. 1205-1208, IEEE (1985).

(7) 岡田美智男："アクティブチャート解析法に基づく One-Pass アルゴリズムの構文制御について"，信学技報，SP90-24 (1990-06).

(8) Schwartz R. and Chow Y. L. : "The N-best Algorithm : An efficient and exact procedure for finding the N most likely sentence hypotheses", Proc. ICASSP-90, pp. 81-84. IEEE (1990).

(9) 小林　豊，表　雅則，遠藤栄憲，新美康永："連続音声認識システム SUSKIT-2 の評価"，信学技報，SP90-21 (1990-06).

(10) 北　研二，竹澤寿幸，保坂順子，江原暉将，森元　逞："2 段階 LR 構文解析法を用いた文認識"，1990 音響学会秋季全大，pp. 127-128 (1990-10).

(11) 堀内靖雄，伊藤克亘，田中穂積："拡張 LR 構文解析アルゴリズムによる未定義語を含む日本語文の構文解析"，1990 情処春季全大，40.

(12) 永井明人，嵯峨山茂樹，北　研二："音素コンテキスト依存型 LR テーブルの生成アルゴリズム"，1991 音響学会春季全大，pp. 91-92 (1991-03).

(13) Bridle J. S., Brown M. D. and Chamberlain R. M. : "An algorithm for connected word recognition. Proc. ICASSP-82, pp. 899-902. IEEE (1982).

(14) 南　泰浩，中川正雄："Trigram モデルを用いた複数候補を求めるフレーム同期型 HMM 連続音声認識"，信学論(D-II)，J73-D-II，9，pp. 1383-1391 (1990-09).

(15) Ney H., Mergel D., Noll A. and Paeseler : "Data-Driven organization of a Dynamic Programming beam search for continuous speech recognition", Proc. ICASSP-87, pp. 833-836. IEEE (1987).

(16) 渡辺隆夫，畑崎香一郎，吉田和永："「Bundle サーチ」による連続音声認識のための高速化手法"，1991 音響学会春

季全大，pp. 125-126 (1991-03).
(17) 速水 悟，田中和世，横山晶一，太田耕三："研究用音声データベースのためのVCV/CVCバランス単語セットの作戦"，電総研彙報，49, 10, pp. 803-834 (1985).
(18) 岡 隆一："音素片スポッティングとベクトル連続DPを用いた連続音声認識．1991音響学会春季全大，pp. 129-130 (1991-03).

(平成3年6月19日受付，4年1月20日再受付)

伊藤 克亘

昭63東工大・工・情報卒．平2同大大学院修士課程了．現在，同大学院博士課程在学中．音声認識，自然言語処理の研究に従事．日本音響学会，人工知能学会各会員．

速水 悟

昭53東大・工・産業機械卒．昭56同大大学院修士課程機械工学専攻了．同年電子技術総合研究所入所．平1～2カーネギーメロン大学客員研究員．現在知能情報部音声研究室主任研究官．音声認識・理解，対話に関する研究に従事．人工知能学会，日本音響学会，日本機械学会，情報処理学会，日本認知科学会，IEEE，ESCA各会員．

田中 穂積

昭39東工大・理工・制御卒．昭41同大大学院修士課程了．同年電気試験所(現，電子技術総合研究所)入所．昭58東京工業大学工学部情報工学科助教授．昭61同大学教授となり現在に至る．工博．人工知能，自然言語処理の研究に従事．認知科学会，日本ソフトウェア科学会，情報処理学会，人工知能学会，計量国語学会各会員．