Vol. 31 No. 3 情報処理学会論文誌 Mar. 1990

HMM 音韻認識と拡張 LR 構文解析法を用いた連続音声認識†

北 研二†† 川端 豪†† 斎藤博昭††

高精度の連続音声認識システムを構築するためには，言語情報の利用が不可欠であり，これまでにも，統計的言語モデル，正規文法，文脈自由文法等を用いて音声認識システムの認識率を向上させる方法が提案されている．本論文では，これらとは異なる新しい方法 HMM-LR 法を提案する．HMM-LR 法は，拡張 LR 構文解析法で用いられる構文解析動作表から入力音声データ中の音韻を予測し，予測された音韻の尤度を HMM 音韻照合で調べることにより，音声認識と言語処理を同時進行させる．この方式では，音声認識と言語処理の間に音韻ラティス等の中間的なデータを介する必要がなく，高精度のかつ効率的な認識処理系を構成することができる．また，HMM-LR 法に基づく日本語の文節認識システムを作成し，評価を行った．評価には，日本語の一般的な文節構造を扱うことのできる一般的文法（語彙数約 1,000 語）と認識対象となるタスクに現れる現象のみを扱うタスク向き文法（語彙数約 270 語）の 2 種類の文法を用いた．一般的文法に対する第 1 位での正答率は 72.0%，第 5 位までで 95.3% の正答率を達成した．タスク向き文法に対しては，それぞれ 79.9%，98.6%の正答率を達成した．

1. はじめに

高精度の連続音声認識システムを構築するためには，言語情報の利用が不可欠である．これまでにも，単語の 2 つ組（bigram）や 3 つ組（trigram）等の統計的言語モデル，正規文法，文脈自由文法等を用いて，音声認識システムの認識率を向上させる方法が提案されている[1)~7)]．

本論文では，拡張 LR 構文解析法[8)]で用いられる構文解析動作表から入力された音声データ中の音韻を予測し，予測された音韻の尤度を HMM (Hidden Markov Model)[9)] 音韻照合で調べることにより，音声認識と言語処理を同時進行させる方式（HMM-LR 法）を提案する．HMM-LR 法は，言語の構文的な情報を用いて音声データを直接解析する方式であり，この方式を用いることにより，高精度のかつ効率的な大語彙連続音声認識システムを構築することができる．

HMM 法に基づく音声認識手法は，確率モデルの中に音声の多様なバリエーションを吸収することができ，DTW (Dynamic Time Warping) のマルチテンプレートに匹敵する能力を持っている．また，認識単位を音韻に設定することにより，任意の単語モデルを音韻モデルから合成することができ，大語彙の音声認識システムに適している．

一方，拡張 LR 構文解析法は，プログラミング言語処理系の分野でよく知られている LR 構文解析法[10)]を拡張したものであり，一般の文脈自由文法を取り扱うことができる．LR 構文解析法は表駆動型の構文解析アルゴリズムであり，パーザの動作を記述してある動作表を参照しながら，入力に対するバックトラックなしに構文解析処理を進めることができ，CYK (Cocke-Younger-Kasami) 法[11)]や Earley 法[12)]等をはじめとする従来の構文解析アルゴリズムよりも高速な解析が可能である．

この拡張 LR 構文解析アルゴリズムの音声認識への応用として，Tomita[13)] は単語ラティスを解析する方式を提案している．また，Saito[14)] は誤りを含んだ音韻列を解析する方式を提案している．しかし，これらの方式では，音声データをいったん音韻記号列に変換しているために情報の欠落が生じ，十分な性能を得るのは困難である．これに対し，HMM-LR 法では音声認識と言語処理の間に音韻ラティスや単語ラティス等の中間的なデータを介する必要がなく，高精度のかつ効率的な認識処理系を構成することができる．

本論文では，以下の構成に従って述べる．まず 2 章で LR 構文解析法および拡張 LR 構文解析法の概要を説明する．3 章では HMM による音声認識について簡単に述べる．次に 4 章において HMM-LR 法を提案し，5 章で HMM-LR 法の性能評価のために行った実験について述べる．最後に 6 章で結論を述べる．

2. LR 構文解析法

2.1 LR 構文解析法

LR 構文解析法[10)]は，プログラミング言語処理系の分野でよく知られた構文解析法であり，文脈自由文法の広いクラスに対して有効である．この構文解析法

† HMM Continuous Speech Recognition Using Generalized LR Parsing by KENJI KITA, TAKESHI KAWABATA and HIROAKI SAITO (ATR Interpreting Telephony Research Laboratories).

†† ATR 自動翻訳電話研究所

は，入力記号を左から右に一方向に読みながら，バックトラックなしに決定的に解析を進めることができる．

LR 構文解析法に基づくパーザは，動作表と行先表という2つの表を参照しながら解析を進めていく．動作表は，パーザの現在の状態 s と入力記号 a から，パーザが次に取るべき動作 $ACTION(s, a)$ を決定するのに用いられる．パーザの動作には

(1) 移動 (shift)
(2) 還元 (reduce)
(3) 受理 (accept)
(4) 誤り (error)

の4種類がある．移動はパーザの状態記号をスタックに積む動作であり，還元はスタック上の記号を文法規則によりまとめあげていく動作である．受理は入力された記号列が解析されたことを意味し，誤りは入力記号列が受け入れられなかったことを意味する．また，行先表はパーザの現在の状態 s と文法記号（終端記号および非終端記号）A からパーザの次の状態 $GOTO(s, A)$ を決定するのに用いられる．

パーザは，入力された記号列がどこまで処理されたかという入力ポインタと状態スタックを持っている．スタック最上段の状態がパーザの現在の状態である．パーザは以下のようにして入力記号列を処理する．以下の操作では解析木を陽に作成しないが，作成するように変更するのは容易である．

(1) 初期化．入力ポインタを入力記号列の先頭に位置づける．スタックに初期状態0をプッシュする．
(2) 現在の状態 s と入力ポインタの指す記号 a に対し，$ACTION(s, a)$ を調べる．
(3) $ACTION(s, a)$="shift" ならば，$GOTO(s, a)$ を状態スタックにプッシュし，入力ポインタを1つ進める．
(4) $ACTION(s, a)$="reduce n" ならば，n 番目の文法規則の右辺にある文法記号の数だけ状態スタックから状態をポップする．スタック最上段の状態 s' と n 番目の文法規則の左辺にある文法記号 A とから，次の状態 $GOTO(s', A)$ を求めスタックにプッシュする．
(5) $ACTION(s, a)$="accept" ならば，解析は終了する．
(6) $ACTION(s, a)$="error" ならば，解析は失敗する．
(7) (2)に戻る．

なお，動作表と行先表は文法規則から機械的に構成することができる．

2.2 拡張 LR 構文解析法

従来の LR 構文解析法では，文法規則が曖昧性を持つ場合には対処できなかったが，Tomita[8] によって曖昧性を持つ文法規則を LR 構文解析法で取り扱う手法が確立された．この手法は，LR 構文解析法で用いる動作表の各欄に複数の動作を記述することを許し，複数の動作記述が指定されているときには，各動作を並列的に行い，同時にいくつかの解析を並行して進めるというものである．この手法を用いることにより，自然言語のように曖昧な入力を持つ言語のすべての解析結果を breadth-first で求めることが可能となる．

図1 に文法規則の例を，**図2** に図1の文法規則を拡張 LR 構文解析法で用いる表（動作表および行先表）に変換した例を示す．図2では，終端記号に対する行先は動作表に示されている．例えば，"$s2$" は移動動作を実行したあとに状態2になることを示している．また，"$r1$" は1番目の文法規則による還元動作を実行することを示す．"*acc*" は受理を意味し，空欄は誤りであることを意味している．入力記号欄の "$" は入力の終りを示している．

3. HMM 法による音韻認識

HMM (Hidden Markov Model) 法[9]は近年注目をあびている音声認識の手法である．HMM は音韻のゆらぎを統計的に表現できるという特徴があり，発声状況やコンテキスト等の違いによる音韻変動に対して robust なモデルを構成することができる．また，認

(1)	S	→	JP V
(2)	S	→	V
(3)	JP	→	NP
(4)	JP	→	NP P
(5)	NP	→	N
(6)	NP	→	S NP
(7)	N	→	k a r e
(8)	N	→	k a n e
(9)	V	→	k u r e
(10)	V	→	k u r e t a
(11)	V	→	o k u r e
(12)	P	→	o
(13)	P	→	g a

図1 文法規則の例
Fig. 1 An example of the context-free grammar.

state	a	e	o	u	g	n	k	r	t	$	S	JP	NP	N	V	P
0			s3				s4				1	5	6	7	2	
1			s3				s4			acc	9	5	8	7	2	
2			r2				r2			r2						
3							s10									
4	s11			s12												
5			s3				s14								13	
6			s16,r3		s17		r3									15
7			r5		r5		r5									
8			s16,r3,r6		s17,r6		r3,r6									15
9			s3				s4				9	5	8	7	2	
10				s18												
11						s20		s19								
12								s21								
13			r1				r1			r1						
14				s12												
15			r4				r4									
16			r12				r12									
17	s22															
18								s23								
19		s24														
20		s25														
21		s26														
22			r13				r13									
23		s27														
24			r7		r7		r7									
25			r8		r8		r8									
26			r9				r9		s28	r9						
27			r11				r11			r11						
28	s29															
29			r10				r10			r10						

図 2 動作表および行先表
Fig. 2 An example of the action and goto tables.

識単位を音韻に設定しておけばこれを基に任意の単語モデルを合成することができるため，大語彙の音声認識システムに適している．HMM-LR 法では，LR パーザから駆動される音韻照合機構として，この HMM 法に基づく音韻認識を採用している．

以下で，音声の特徴パラメータをベクトル量子化して扱う離散型 HMM について簡単に述べる．

図 3 に典型的な音韻モデルの例を示す．モデルはいくつかの状態と状態間の遷移を表す弧から構成される．ある時点にある状態にあるモデルは弧に沿って次の状態に遷移し，あるコードベクトルを外部に出力する．この遷移およびそれに伴うコードベクトルの出力は確率的に行われ，各弧，例えば状態 i から状態 j への弧には遷移確率 a_{ij} とコードベクトル k に対する出力確率 b_{ijk} の値がパラメータとして与えられている．モデルはこれらの確率値に従って，さまざまな VQ コードの時系列をさまざまな確率で出力する．

HMM 法を用いて音韻照合を行うためには，あらかじめ音韻の種類だけモデルを用意し，各々，学習用

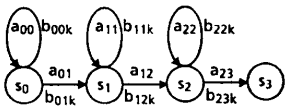


図 3 HMM 音韻モデル
Fig. 3 HMM phone model.

音韻データのコード列を最も高い確率で出力するように，音韻モデルのパラメータを Forward-Backward アルゴリズム[9] を用いて学習しておく．未知音声データのコード列に対し，照合する音韻のモデルからそのコード列が出力される確率を計算し照合結果とする．確率計算は，Viterbi アルゴリズム[9]を用い，以下のようにして行われる．

[記号の定義]

N : 入力音声データに対するコード列の長さ，

M : 照合される音韻モデルの状態数，

O_i : 入力音声データコード列の i 番目のコード，

$a(j_1, j_2)$: 照合される音韻モデルにおいて状態 j_1 と状態 j_2 を結ぶ弧の遷移確率，

$b(j_1, j_2, k)$: 照合される音韻モデルにおいて状態 j_1 と状態 j_2 を結ぶ弧がコード k を出力する確率．

[初期化]

$P(0, 0) = 1.0$,

$P(0, j) = 0.0 \ (j = 1 \ldots M)$,

$P(i, 0) = 0.0 \ (i = 1 \ldots N)$.

[漸化計算] $(i = 1 \ldots N, \ j = 1 \ldots M)$

$$P(i, j) = \max \left(P(i-1, j) \times a(j, j) \times b(j, j, O_i), \right.$$
$$P(i-1, j-1) \times a(j-1, j)$$
$$\left. \times b(j-1, j, O_i) \right).$$

音韻照合の結果は配列 $P(1, M) \ldots P(N, M)$ の中に

求められる．$P(i, M)$ には，入力音声データコード列の時点 i での照合音韻の尤度が入っている．

4. HMM-LR 法

4.1 基本的なメカニズム

図 4 に，HMM-LR 法に基づく連続音声認識システムの構成図を示す．

文法規則は LR テーブル生成系により，あらかじめ LR テーブル（動作表および行先表）に変換しておく．また，それぞれの音韻に対する音韻モデルもあらかじめ用意しておく．HMM-LR パーザは LR テーブルから，発話された音声データ中の音韻を予測し，予測された音韻に対し HMM 音韻照合を駆動することにより，予測された音韻の尤度を調べる．これにより，音声認識と言語処理を同時進行させる．音声認識と言語処理の間に音韻／単語ラティス等の中間的なデータを介さないため，効率的かつ高精度に音声データの処理を行うことができる．

LR テーブルを予測に用いるという点が，従来の LR 構文解析法とは大きく異なる点であり，HMM-LR 法の特徴となっている．

4.2 LR 動作表からの音韻予測

音韻予測の様子を図 2 の動作表を例に説明する．

いま LR パーザが状態 0 であるとする．従来の LR パーザであれば状態 0 の行と入力された音韻記号（例えば /k/）から表引きを行い，次の動作（この場合 s 4）を決定する．これに対し HMM-LR 法では，動作表の状態 0 の横 1 行をすべて調べ，移動動作の指定されている音韻をすべて選び出し音韻照合を行う．図 2 の動作表では，状態 0 で 2 つの音韻 /o/ と /k/ が予測され，これらの音韻に対して音韻照合を行うことになる．これは，文法で規定された制限下で，次の音韻を予測していることになる．このように，動作表は音韻予測に用いられるため，文法規則の終端記号は，単語や品詞名ではなく，音韻となっている．

4.3 アルゴリズム

HMM-LR 法のアルゴリズムを説明するために，まずセル (cell) というデータ構造を導入する．セルは解析に必要な情報を保存しておくためのデータ構造であり，入力音声データに対する認識候補のそれぞれにつき 1 つのセルが用いられる．以下の 2 つの情報がセルに記憶される．

- LR パーザの状態スタック，
- 確率テーブル（確率テーブルは認識された音韻列の時間軸上の各点での尤度を格納しておくための配列である）．

HMM-LR 法では，以下のようにして入力音声データを処理する．なお，記号の定義は 3 章で与えたものを用いている．

（1） 初期化．新しいセル C を 1 つ作り，C の LR 状態スタックに初期状態 0 をプッシュする．また，C の確率テーブル Q を以下により初期化する．

$Q(0)=1.0$,

$Q(i)=0.0$ $(i=1...N)$.

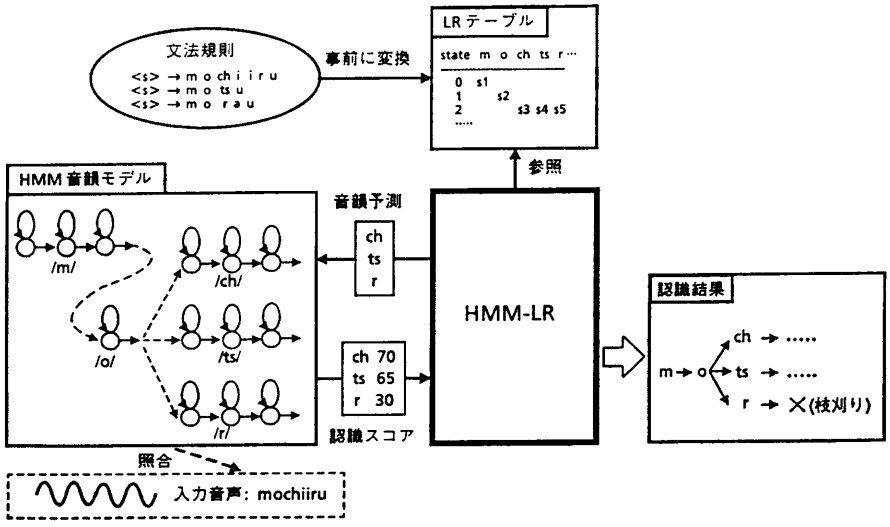


図 4 HMM-LR 連続音声認識システム
Fig. 4 HMM-LR continuous speech recognition system.

(2) セルの分岐．集合 S を以下により作成する．

$S=\{(C,s,a,x)|$
$^{\exists}C,a,x$ (C は受理されていないセル
& s は C の状態スタックの最上段の状態
& a は状態 s において移動または還元動作が指定されている終端記号
& $x=ACTION(s,a)$
& $x\neq$"error"$\}$.

集合 S の各要素 (C,s,a,x) に対し，セル C のコピー C' を作成し，以下の操作を実行する．集合 S が空ならば，解析を終了する．

(3) $x=$"shift" ならば，記号 a を HMM で音韻照合する．このとき，セル C' の確率テーブル Q の値を，下記の式により更新する．

$$P(0,j)=0\ (j=1\ldots M),$$
$$P(i,0)=Q(i)\ (i=0\ldots N),$$
$$P(i,j)=\max\ (P(i-1,j)\times a(j,j)\times b(j,j,O_i),\ P(i-1,j-1)\times a(j-1,j)\times b(j-1,j,O_i))\quad (i=1\ldots N,\ j=1\ldots M),$$
$$Q(i)=P(i,M)\qquad (i=1\ldots N).$$

このとき，$\max Q(i)$ $(1\leqq i\leqq N)$ があらかじめ決められている音韻照合の閾値よりも大きければ，$GOTO(s,a)$ をセル C' の LR 状態スタックにプッシュする．そうでなければ，セル C' をすてる．

(4) $x=$"reduce n" ならば，n 番目の文法規則による還元操作を行う．これは，通常の LR 構文解析法と同じ動作である．

(5) $x=$"accept" であり，セル C' の確率テーブル Q に対して，$Q(N)$ があらかじめ決められている音韻照合の閾値よりも大きければ，セル C' は受理される．そうでなければ，セル C' をすてる．

(6) (2)に戻る．

最終的に残されたセルに認識結果が入っていることになる．一般に，複数のセルが残されるが，確率テーブル $Q(N)$ の値により，これらのセルの間に認識の順位を付けることができる．

4.4 アルゴリズムの改良点

4.3 節で述べたアルゴリズムは最も基本的なものであり，いくつかの改良点が考えられる．これらの改良点は主に計算量の削減に寄与するものであり，実際のシステムを構築する際に有用である．以下に改良点を列挙する．

(a) ビームサーチ法

ビームサーチ法は，多くの音声認識システムで採用されている木 (Tree) の探索方式であり，木のノードに付与されている評価値の高いノードだけを並行して処理し，評価値の低いノードは枝刈りするというものである．

一般にアルゴリズムの(2)の時点で作られる集合 S は比較的大きなものとなる．ビームサーチ法を用い，集合 S の大きさに制限を加えることができる．評価値として，集合 S の要素 (C,s,a,x) に対し，セル C の確率テーブル Q の中の最大値 $\max Q(i)$ $(1\leqq i\leqq N)$ を用いることができる．

また，探索木の各節点における最大分岐数に制限を加える方法も考えられる．4.3 節のアルゴリズムでは，1つのセルなら分岐するセルの数（セルを1つ固定したときの集合 S の大きさ）に制限を加えることになる．

(b) 継続時間長制御

HMM の音韻モデルの継続時間長制御が認識率の改善に寄与することが報告されているが[15]，継続時間長制御はまた HMM-LR 法における計算範囲の縮小にも有効である．例えば，音韻モデルの状態の継続時間に上限，下限を設定するような継続時間長制御を用いる場合には，入力音声データコード列の

$$\sum_{n\in T} D_{\min}(n)\leqq i\leqq \min\ (\sum_{n\in T'} D_{\max}(n),\ N)$$

の部分だけに対して漸化計算を行えばよい．ここで，$D_{\min}$ は状態の継続時間の下限を，$D_{\max}$ は状態の継続時間の上限を表している．また，T はこれまでに照合した状態の集合であり，T' は集合 T にこれから照合する状態を付け加えた集合である．

また，HMM-LR 法のように音韻間の境界を与えずに音韻モデルを連結していく方法にビームサーチを用いる場合，誤った音韻モデルに対しても入力音声とモデルの状態の不適当な対応付けによって確率値が大きくなり，これらの音韻列と競合することにより正解の音韻列が枝刈りされてしまうことがある．この現象を避けるためにも，継続時間長制御は有効である[16]．

5. 認識実験

HMM-LR 法に基づく連続音声認識の手法を評価するために，日本語の文節認識システムを作成した．

「会議に」,「発表するのでは」,
「なくて」,「聴講するだけだと」,
「費用は」,「いくら」,「かかりますか」

図 5 文節発声の例
Fig. 5 An example of a Japanese phrase sequence.

5.1 音声データ

男性アナウンサ1名が文節単位に発声した 25 会話文中の 279 文節をサンプリング周波数 12 kHz で AD 変換後，フレーム周期 3 ms ごとに 256 点ハミング窓で切り出し，12 次 LPC 分析，16 次 PWLR 距離尺度を用いて VQ コード列（コードサイズ 256）に変換したものを認識対象とした．

各文節の構造は，1個の自立語のあとに複数（0個も含む）の付属語が連鎖したものとなっている．認識実験に用いた文節の例を**図 5** に示す．

5.2 音韻認識の条件

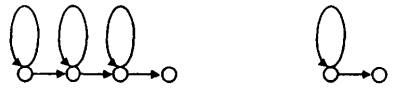
音韻は過渡的なものと定常的なものに大別し，各々4状態3ループ，2状態1ループのモデル構造を設定した．音韻の種類を**表 1** に，モデルの構造を**図 6** に示す．

音韻モデルの学習には，認識対象となる音声データと同一の話者が発声した日本語重要語 5,240 単語を使用した．これらの学習用単語データには訓練されたラベラーにより音韻ラベルと，音響的特徴に対応付けてさらに細分化したイベントラベルが付けられている[17]．

音韻モデルの各状態は音韻を構成する音響的イベントと強く関係付けられており，イベントラベルの継続時間の統計から各状態に対する継続時間の最小値，最大値を設定した．また，学習データ（単語発声）と認識データ（文節発声）の発声速度の変化に応じて HMM の状態の継続時間を補正した[18]．音韻照合の計算には Viterbi アルゴリズムを用いた．

表 1 音韻モデルの種類
Table 1 HMM phone units.

音　韻	ループ数
/b/, /d/, /g/, /p/, /t/, /k/, /m/, /n/, /ng/, /r/, /z/, /ch/, /ts/, /y/, /w/, /gy/, /hy/, /ky/, /my/, /ngy/, /ny/, /py/, /ry/, /zy/	3
/i/, /e/, /a/, /o/, /u/, /N/, /ii/, /ee/, /aa/, /oo/, /uu/, /ei/, /ou/, /s/, /sh/, /h/	1



(a) 3-loop model (b) 1-loop model

図 6 音韻モデルの構造
Fig. 6 HMM phone models.

表 2 文法の大きさおよび複雑性
Table 2 Grammars.

	一般的文法	タスク向き文法
規 則 数	1,461	582
異なり単語数	1,035	275
LR テーブルの状態数	4,359	1,207
タスクエントロピー	17.0	13.2
音韻パープレキシティ	5.9	3.9

5.3 文　法

2種類の文法を用意した．2種類の文法のどちらも日本語の文節構造を規定するもので，文脈自由文法の形式で記述されている．語彙も文法のレベルで記述されている．

2種類の文法のうち1つは一般的な文法（General grammar）であり，きわめて広範囲の日本語の言語現象を取り扱うことができる．もう1つの文法は，タスクに依存した文法（Task-specific grammar）であり，認識対象となるタスクに現れる言語現象のみを取り扱う．タスクに依存した文法は，認識対象である「国際会議参加申し込みに関する問い合せ」に関して ATR が収集した言語データベース[19]中のデータを形態素解析し，単語間の連接の可能性を調べ，それを文脈自由文法の形式で記述したものである．

それぞれの文法の規則数（語彙規則を含む），異なり単語数，LR テーブルに変換した際の状態数，および，文法の複雑性を評価する尺度であるタスクエントロピー[20]，音韻パープレキシティ[20]の値を**表 2** に示す．タスクエントロピーは，一文節あたりの平均情報量を表しており，音韻パープレキシティは音韻あたりの平均分岐数と似た量を表している．

5.4 実験結果および考察

前節で述べた一般的文法およびタスク向き文法を用いて行った文節認識実験の結果を**表 3**，**表 4** に示す．

本方式における認識性能は，探索のビーム幅と探索木の各節点における最大分岐数の2つのパラメータの値によって変わるが，予備的な実験により節点の最大分岐数は 12 に制限した．

表3，表4では，表の縦方向にビーム幅をいろいろな値にしたときの結果を示してある．いずれの場合も

表 3 一般的文法を用いたときの文節認識率
Table 3 Phrase recognition rates (General grammar).

ビーム幅	rank=1	≦2	≦3	≦4	≦5
5	38.4%	44.8	44.8	44.8	44.8
10	56.6	67.0	69.9	70.6	70.6
20	66.0	78.9	82.4	84.2	85.0
50	71.7	86.0	90.3	92.8	93.4
100～	72.0	85.7	92.1	94.3	95.3

表 4 タスク向き文法を用いたときの文節認識率
Table 4 Phrase recognition rates (Task-specific grammar).

ビーム幅	rank=1	≦2	≦3	≦4	≦5
5	72.4%	82.1	83.9	84.2	84.2
10	77.4	88.2	91.4	92.1	92.8
20	80.3	91.8	94.9	96.8	97.5
30～	79.9	92.8	96.1	97.5	98.6

ビーム幅が大きいほど認識率は向上し，ビーム幅がある一定値以上に大きくなると認識率は飽和する．飽和するビーム幅は文法によって異なり，一般的文法に対しては約 100，タスク向き文法に対しては約 30 であった．このことから探索のビーム幅はタスクの複雑さに応じて設定しなければいけないことが判明した．

一般的文法の場合には第1位での正答率 72.0%，第5位までで 95.3% の正答率を達成した．タスク向き文法の場合には第1位での正答率 79.9%，第5位までで 98.6% の正答率が得られた．

認識誤りの例を**表 5** に示す．認識誤りの主なものに，付属語の誤り（助詞の脱落，挿入等）がある．これは日本語の助詞，助動詞の発声が短い上に，それらの間の連接に非常に多くのバリエーションがあるためであると考えられる．いずれの文法を用いた場合にも，認識対象文節総数の約1割は付属語部分の認識を誤っていた．また，発音の似た自立語間での置換誤りも多く，一般的文法を用いたときは認識対象文節総数の約 13%，タスク向き文法を用いたときは認識対象文節総数の約6%が，この種の誤りであった．タスク向き文法を用いた際の自立語部分の誤りが少ないのは，タスク向き文法に含まれる自立語の数が少ないためであると思われる．

表 5 認識誤りの例
Table 5 Examples of recognition errors.

正		誤	
gozyuushoo	(ご住所を)	gozyuusho	(ご住所)
tourokuno	(登録の)	tourokumo	(登録も)
kaigizyoumade	(会議場まで)	kaizyoumae	(会場前)
itadakimasu	(頂きます)	itashimasu	(致します)

表 6 音韻ラティス解析法と HMM-LR 法との文節認識率の比較
Table 6 Phrase recognition rates by phone lattice parsing and HMM-LR.

Recognition order	Recognition rate	
	Phone lattice parsing	HMM-LR
1st candidate	51%	72%
Within top 5	82%	95%
Within top 10	85%	—
Within top 30	87%	—

また，HMM-LR 法の有効性を確認するために，HMM 法と拡張 LR 構文解析法を別個に適用した場合について文節認識実験を行った．この実験では，HMM 音韻認識プログラムを用いて音韻ラティスを作成し，音韻ラティスに拡張 LR 構文解析法を適用して文節ラティスを作成した．誤りを含んだ音韻列を拡張 LR 構文解析法によって解析する方法が Saito[14] に述べられているが，本実験ではその方法を音韻ラティスの解析に用いた．一般的文法を用いたときの文節認識率を**表 6** に示す．表6で Phone lattice parsing と記されたものが，音韻ラティスに拡張 LR 構文解析法を適用したときの認識率である．HMM 法と拡張 LR 構文解析法を別個に適用した場合，30 位までの累積認識率でも 87% であるのに対し，HMM-LR 法では5位までの累積認識率で 95% を得ている．この実験からも HMM-LR 法が連続音声の認識に対し有効な方法であることを示している．

6. おわりに

拡張 LR 構文解析法で用いられる構文解析動作表から入力された音声データ中の音韻を予測し，予測された音韻の尤度を HMM 音韻照合で調べることにより，音声認識と言語処理を同時進行させる方式（HMM-LR 法）を提案した．また，HMM-LR 法を日本語の文節認識実験によって評価し，十分高い認識率が得られることを確認した．実験結果は，HMM-LR 法が連続音声の認識において非常に有効な方式であることを示している．

HMM-LR 法は基本的に言語の構文的な情報を用いて音声データを解析する方式であるが，単一化文法等が用いている選択制限のチェックを還元動作に付随させることによって，より高精度の音声認識／解析システムを構築することができる．これは今後の課題であ

る.

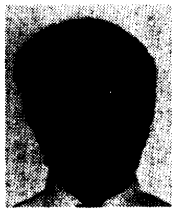
また，本論文で提案したアルゴリズムでは，同じ音韻に対して音韻照合を重複して行うことがあるが，これを避けるために Local ambiguity packing[8] の手法を組み込むことが考えられる.

謝辞 研究の機会を与えて頂いた ATR 自動翻訳電話研究所樽松明社長に深謝いたします．また熱心に討論して頂いた同研究所音声情報処理研究室およびデータ処理研究室の皆様に感謝いたします.

参 考 文 献

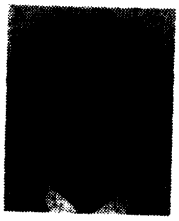
1) Shikano, K.: Improvement of Word Recognition Results by Trigram Model, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-87*, pp. 1261-1264 (1987).
2) Lee, K. F. and Hon, H. W.: Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-88*, pp. 123-126 (1988).
3) Rohlicek, J. R., Chow, Y. L. and Roucos, S.: Statistical Language Modeling Using a Small Corpus from an Application Domain, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-87*, pp. 267-270 (1987).
4) Chow, Y. L., Dunham, M. O., Kimball, O. A., Krasner, M. A., Kubla, G. F., Makhoul, J., Price, P. J., Roucos, S. and Schwartz, R. M.: BYBLOS: The BBN Continuous Speech Recognition System, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-87*, pp. 89-92 (1987).
5) Ney, H.: Dynamic Programming Speech Recognition Using a Context-Free Grammar, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-87*, pp. 69-72 (1987).
6) Nakagawa, S.: Spoken Sentence Recognition by Time-Synchronous Parsing Algorithm of Context-Free Grammar, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-87*, pp. 829-832 (1987).
7) Wright, J. H.: Linguistic Control in Speech Recognition, *Proc. European Conf. Specch Tech.*, Vol. 2, pp. 104-107 (1987).
8) Tomita, M.: *Efficient Parsing for Natural Language: A Fast Algorithm for Practical Systems*, p. 201, Kluwer Academic Publishers (1986).
9) Levinson, S. E., Rabiner, L. R. and Sondhi, M. M.: An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition, *Bell Syst. Tech. J.*, Vol. 62, No. 4, pp. 1035-1074 (1983).
10) Aho, A. V., Sethi, R. and Ullman, J. D.: *Compilers, Principles, Techniques, and Tools*, p. 796, Addison-Wesley (1986).
11) Aho, A. V. and Ullman, J. D.: *The Theory of Parsing, Translation, and Compiling*, Prentice-Hall (1972).
12) Earley, J.: An Efficient Context-Free Parsing Algorithm, *Comm. ACM*, Vol. 13, No. 2, pp. 94-102 (1970).
13) Tomita, M.: An Efficient Word Lattice Parsing Algorithm for Continuous Speech Recognition, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., ICASSP-86*, pp. 1569-1572 (1986).
14) Saito, H. and Tomita, M.: Parsing Noisy Sentences, *Proc. 12th Int. Conf. Comput. Linguist., COLING-88*, pp. 561-566 (1988).
15) 花沢，川端，鹿野：HMM 音韻認識におけるモデル学習の諸検討，電子情報通信学会技術研究報告，SP 88-22, Vol. 88, No. 92, pp. 9-15 (1988).
16) 花沢，北，川端，鹿野：HMM 音韻モデルの文節認識による評価，日本音響学会平成元年度春季研究発表会講演論文集，3-6-6, pp. 81-82 (1989).
17) 武田，匂坂，片桐，桑原：研究用日本語音声データベースの構築，日本音響学会誌，Vol. 44, No. 10, pp. 747-754 (1988).
18) 川端，花沢，鹿野：HMM 音韻認識に基づくワードスポッティング，電子情報通信学会技術研究報告，SP 88-23, Vol. 88, No. 92, pp. 17-22 (1988).
19) 小倉，橋本，森元：言語データベース統合管理システム，情報処理学会自然言語処理研究会報告，No. 69, 88-NL-69 (1988).
20) 川端，鹿野，北：音韻パープレキシティの提案，日本音響学会平成元年度春季研究発表会講演論文集，3-6-12, pp. 93-94 (1989).

(平成元年 6 月 14 日受付)
(平成元年 11 月 14 日採録)



北 研二 (正会員)

昭和 56 年早稲田大学理工学部数学科卒業．昭和 58 年沖電気工業(株)入社．現在，ATR 自動翻訳電話研究所研究員．音声・言語のインタフェースに関する研究に従事．日本音響学会，日本認知科学会，ACL 各会員.



川端　豪

昭和53年東北大学工学部電子工学科卒業．昭和58年同大学院博士課程修了．工学博士．同年日本電信電話公社入社．現在，ATR 自動翻訳電話研究所音声情報処理研究室主任研究員．音声自動認識に関する研究に従事．電子情報通信学会，日本音響学会，IEEE 各会員．

斎藤　博昭（正会員）

昭和58年慶応大学工学部卒業．現在，同大学院博士課程在学中．自然言語処理，音声理解に関する研究に従事．