

TR-IT-0325

効率的な音素環境依存 LR 表と  
その連続音声認識への応用

An Efficient Phoneme-Context-Dependent LR Table and  
its Applications in Continuous Speech Recognition

竹澤 寿幸

Toshiyuki TAKEZAWA

2000. 1.11

内容梗概

音素環境に依存する音素モデルと LR 構文解析表を組み合わせた、従来の連続音声認識装置では、音素モデルの接続可能性を実行時に調べながら、音声認識の処理を進めなければならず、効率が悪かった。しかも、単語あるいは文節の境界において、あとの還元処理の時点で棄却されるにもかかわらず、音素照合時には接続可能とみなされてしまう異音モデルの数が増大してしまい、処理時間の増大ないし認識率の低下を招いていた。そこで、本稿では、音声認識過程で生ずる音素環境の無駄な接続情報をあらかじめ除去した LR 構文解析表を作成する機能を提案し、効率的で高性能な連続音声認識が実現できることを実験により示す。

ATR 音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

© 株式会社 エイ・ティ・アール音声翻訳通信研究所

© 2000 by ATR Interpreting Telecommunications Research Laboratories

## 目次

1	まえがき	1
2	算法	1
2.1	異音接続行列 . . . . .	1
2.2	語彙規則の変換 . . . . .	1
2.3	異音に基づく LR 表 . . . . .	2
2.4	LR 表の修正 . . . . .	3
3	表のサイズ	4
4	音声認識実験	4
5	むすび	6
	謝辞 . . . . .	6
	参考文献	7

## 1 まえがき

一般化 LR 構文解析器を異音に基づく音声認識過程に統合することにより、連続音声認識においてより良い性能を達成することが期待できる [1]。この場合、一般化 LR 構文解析器は、文脈自由文法から生成される LR 表によって駆動され、異音予測器として用いられる。一般化 LR 構文解析器を異音に基づく音声認識系に統合する際の大きな問題は、いかにして単語連結時の無駄な音素照合を削減するかにある。

本稿では、文脈自由文法の集合と語彙および異音規則から、制約伝播法 (CPM: Constraint Propagation Method) によって、異音に基づく LR 表を生成する新しい手法について述べる。さらに、連続音声認識の実験結果を示す。

## 2 算法

算法は四つの段階からなる。

- (1) 異音接続行列の作成
- (2) 語彙規則の変換
- (3) 異音に基づく LR 表の生成
- (4) 異音に基づく LR 表の修正

### 2.1 異音接続行列

異音接続行列は、二つの隣接異音間の接続可能性を規定する。

例えば、音韻 “i” に対する異音 “i2” と音韻 “d” に対する異音 “d1” の環境集合が次のようであったとしてみよう。

$$i2 : \left\{ \begin{array}{c} \vdots \\ ch \\ \vdots \end{array} \right\} i \left\{ \begin{array}{c} \vdots \\ d \\ \vdots \end{array} \right\}, d1 : \left\{ \begin{array}{c} \vdots \\ i \\ \vdots \end{array} \right\} d \left\{ \begin{array}{c} a \\ \vdots \\ \vdots \end{array} \right\}$$

“d1” は “i2” に後続可能であるといえる。なぜなら、“i2” の右環境が音韻 “d” を含んでおり、“d1” の左環境が音韻 “i” を含んでいるからである。接続行列は 接続 [左異音, 右異音] の配列で表現できる。その値は “1” (二つの異音は接続可能) か “0” (二つの異音は接続不可能) である。

### 2.2 語彙規則の変換

次のような簡単な日本語文脈自由文法と語彙規則を用いて、語彙規則の変換と、異音に基づく LR 表生成に関する制約伝播法の算法を説明する。

- (1)  $S \rightarrow N BE$     (3)  $N \rightarrow ch i ch i$   
 (2)  $N \rightarrow h a h a$     (4)  $BE \rightarrow d a$

異音環境にしたがって、単語内の音韻を異音に変える。すると、語彙規則は次のようになる。

- (2)'  $N \rightarrow h a_1 h_1 a$     (4)'  $BE \rightarrow d a$   
 (3)'  $N \rightarrow ch i_2 ch_2 i$

### 2.3 異音に基づく LR 表

単語連結問題を解決するために、文脈自由文法と語彙規則の集合に異音規則を導入する。異音規則は、音韻と対応する音素をペアにすることで、導出される。

仮に“h”に対して  $\{h_1, h_2\}$ , “a”に対して  $\{a_1, a_2\}$ , “ch”に対して  $\{ch_1, ch_2\}$ , “i”に対して  $\{i_1, i_2\}$ , “d”に対して  $\{d_1, d_2, d_3\}$  があるとすると、次の異音規則の集合が産出できる。

- (5)  $h \rightarrow h_1$     (9)  $ch \rightarrow ch_1$     (13)  $d \rightarrow d_1$   
 (6)  $h \rightarrow h_2$     (10)  $ch \rightarrow ch_2$     (14)  $d \rightarrow d_2$   
 (7)  $a \rightarrow a_1$     (11)  $i \rightarrow i_1$     (15)  $d \rightarrow d_3$   
 (8)  $a \rightarrow a_2$     (12)  $i \rightarrow i_2$

文脈自由文法規則の集合 ((1)), および、変換された語彙規則 ((2)'-(4)'), 異音規則 ((5)-(15)) から、異音に基づく正準 LR 表が生成される。図 1 に、その LR 表の一部を示す。

state	action					
	h2	a1	ch2	i1	d2	d3
0	sh6(c)					
.....						
6*		re6(a)				
7					sh12(c)	sh13(e)
8			sh21			
.....						
18					re2(d)	re2(e)
19					re7(d)	re7(a)
.....						
21				sh23(b)		

図 1: 異音に基づく LR 表の一部

## 2.4 LR 表の修正

図 1 の LR 表は、単語境界における音韻の、いかなる音素環境情報も含んでいない。接続制約を LR 表に組み入れるために、次のようにして、LR 表を修正する (LR 表の修正を説明するために、図 1 を利用する)。

### (1) 接続検査

まず、文献 [2] と同様にして、LR 表の不正な動作を除去するために、接続行列を利用する。

#### (a) 異音規則の不正な還元動作を除去する

例えば、状態 6 の re6 (先読み “a1”) は、もし “h2” (規則 6 の右辺) と “a1” が接続不可能である (接続 [h2, a1]=0) ならば、除去されるべきである。

#### (b) 前が移動動作である不正な移動動作を除去する

例えば、状態 21 の sh23 (先読み “i1”) は、もし “ch2” (直前の移動動作の先読み) と “i1” が接続不可能ならば、除去されるべきである。

## (2) 制約伝播

次に、制約伝播により、他のすべての不正な動作を除去する。

### (c) 空の状態へ至る移動動作を除去する

空の状態は、そのすべての動作が既に除去されてしまった状態として定義される。空の状態に至る移動動作は除去されるべきである。例えば、状態 6 は空の状態であるので、状態 0 の sh6 (先読み “h2”) は除去されるべきである。

### (d) 除去された移動動作に至る還元動作を除去する

状態 19 の re7 (先読み “d2”) について考えてみよう。構文解析器は re7 のあとで状態 18 に遷移する。状態 18 において、次の動作は re2 (先読み “d2”) である。re2 のあとで、構文解析器は状態 7 に遷移する。しかし、この状態の sh12 (先読み “d2”) は段階 (c) により既に除去されている。それゆえ、re7 (状態 19 中) と re2 (状態 18 中) も除去されるべきである。

### (e) 前が行先遷移動作である不正な動作を除去する

状態 18 の re2 (先読み “d3”) について考えてみよう。この動作は、状態 19 の re7 (先読み “d3”) のあとの行先遷移動作から遷移されたものである。re7 (先読み “d3”) は段階 (a) により除去されているので、状態 18 の re2 (先読み “d3”) は除去されるべきである。状態 7 の sh13 (先読み “d3”) についても同様である。

算法は、動作がそれ以上除去されなくなるまで、この制約伝播 (段階 (c)-(e)) を繰り返す。そして、LR 表を圧縮し、表のサイズを小さくする。

## 3 表のサイズ

1225 文脈自由文法規則と 1588 語彙規則をもつ “国際会議の問合わせタスク” のための ATR の文節文法を用いた。テストセットに対する音素パープレキシティは 3.43 である。異音モデルは SSS アルゴリズム [1] により生成される。表 1 に、提案手法により生成された LR 表のサイズを示す。比較のために、26 音韻の場合の LR 表のサイズも示す。音素環境独立 (26 音韻) LR 表と比較すると、1759 異音の場合、異音に基づく LR 表の状態数は、約 35% 増加するのみである。

## 4 音声認識実験

ATR の SSS-LR 音声認識系 [1] に基づく音声認識実験を行なった。ATR の SSS-LR 系では、音素環境依存構文解析器が構文解析器のレベルで実現されている。文法は 3 節で述べた ATR の文節文法を用いた。表 2 に、2 種類の一般化 LR 構文解析器の実現法における音声認識率を示す。一つは、提案手法 (制約伝播法) で生成された異音に基づく LR 表に基づくものであり、もう一つは、ATR の構文解析器のレベルでの実現法である。ビーム幅が 50

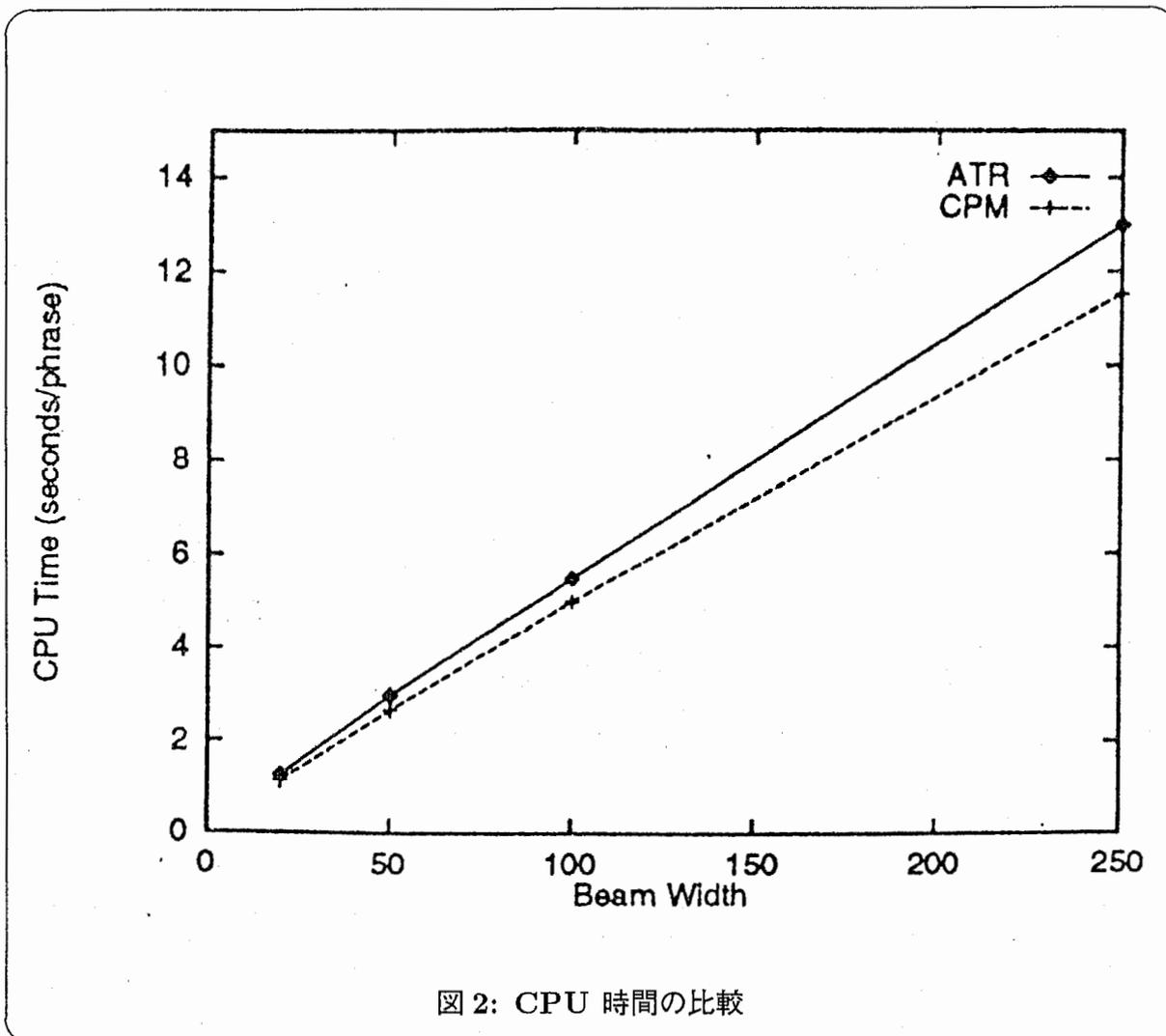
表 1: 表のサイズ

異音数	26*	283	1026	1759
状態数	7411	9004	9628	9994
移動動作	6578	10696	14020	16035
還元動作	15428	21145	23891	24602
行先遷移動作	1463	4013	4013	4013

表 2: 音声認識率

異音数	手法	<i>beam</i> = 50		<i>beam</i> = 250	
		rank 1	rank 5	rank 1	rank 5
26*		82.90	90.43	87.83	97.10
283	従来手法	83.48	90.72	87.83	97.10
	提案手法	84.06	91.30	87.83	97.39
1026	従来手法	88.12	95.07	91.01	99.13
	提案手法	88.70	95.94	91.01	99.13

のとき、音声認識率は提案手法がわずかによいことがわかる。図 2 に、CPU 時間の比較を示す。音素環境依存性があらかじめ LR 表にコンパイルされているため、提案手法で実行時に必要となる CPU 時間は少ない。



## 5 むすび

制約伝播法により異音に基づく LR 表を生成する算法を述べた。さらに、実験結果を示した。今後の課題を記す。

- (1) 文認識
- (2) 異音制約に加えて、形態素制約の LR 表への統合
- (3) 確率文脈自由文法への応用

## 謝辞

実験に協力いただいた Harald Singer, 林輝昭 両氏に感謝いたします。なお、本研究は故李輝 (Li Hui) 氏が東京工業大学工学部情報工学科田中穂積研究室博士後期課程在学中に平成 6 年 4 月 25 日から平成 6 年 6 月 24 日まで ATR 音声翻訳通信研究所に学外実習生として滞在した際の成果である。

## 参考文献

- [1] A. Nagai, et al., "Three Different LR Parsing Algorithm for Phoneme-Context-Dependent HMM Based Continuous Speech Recognition," *IEICE Trans. Inf. & Syst.*, vol. **E76-D**, no. 1, pp. 29-37, 1993.
- [2] H. Tanaka, et al., "Integration of Morphological and Syntactic Analysis Based on LR Parsing Algorithm," *International Workshop on Parsing Technologies*, pp. 101-109, 1993.