TR-IT-0319

文を発話単位としたHMM-LR
連続音声認識におけるポーズ情報の利用

Using Pause Information in HMM-LR Continuous Speech Recognition Accepting Sentence Utterance

竹沢 寿幸　森元 逞†
Toshiyuki TAKEZAWA　Tsuyoshi MORIMOTO†

2000. 1.11

内容梗概

自然な発話の認識を目指し、文を発話単位とした連続音声認識手法の検討を進めている。HMM-LR連続音声認識手法をベースにして、統語的に適格な文を対象に、文節毎にポーズを置いても置かなくてもよいような発話の受理を目的とした。認識過程で爆発的に増加する計算量を削減するために、ポーズ情報を積極的に利用する2つの手法を検討した。1つは、ポーズが検出できた時に統語的な曖昧性をパッキングするものである。もう1つは、ポーズ情報を音素照合区間の削減に利用するものである。予備実験によりその有効性が確認できたので、その内容について報告する。

ATR音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories
†現在、福岡大学工学部電子情報工学科

目次

1 まえがき

自然な発話の認識を目指し、文を発話単位とした連続音声認識手法の検討を進めている。HMM-LR 連続音声認識手法 [1] をベースにして、統語的に適格な文を対象に、文節毎にポーズを置いても置かなくてもよいような発話の受理を目的とした。認識過程で爆発的に増加する計算量を削減するために、ポーズ情報を積極的に利用する 2 つの手法を検討した。1 つは、ポーズが検出できた時に統語的な曖昧性をパッキングするものである。もう 1 つは、ポーズ情報を音素照合区間の削減に利用するものである。予備実験によりその有効性が確認できたので、その内容について報告する。

2 基本的な考え方

HMM-LR 法を文節発話から文発話に単に拡張すると、認識過程で統語的な曖昧性が爆発し、ビーム幅内を同じような候補が埋め尽くしてしまう (図 1)。

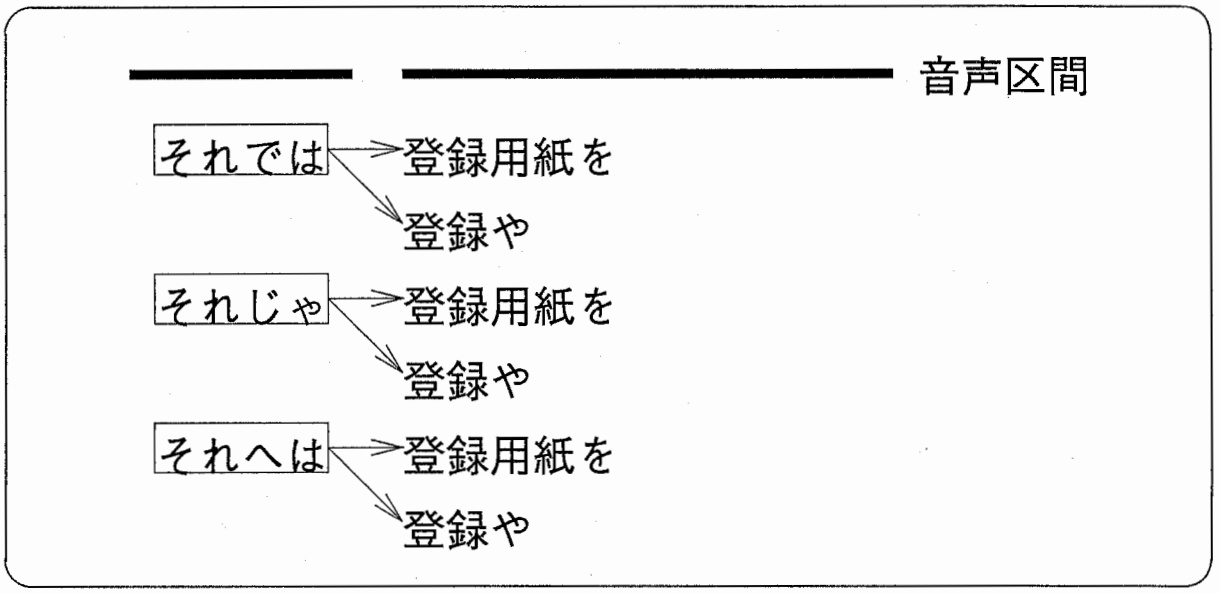


図 1: HMM-LR 法による音声認識過程

音声認識にとって同じような候補をマージしようとしても、音素照合に利用している入力フレームが異なるため、一般にはマージできない。南らは電話番号案内に特化した文法と技法を使って候補のマージを行なっている [2]。

一方、文を発話単位として丁寧に発話された音声であっても、生理的な理由などにより、発話の途中に自然にポーズが挿入されることがある。そのようなポーズが検出できれば、ポーズの時点で入力フレームの同期を取ることができるので、一般的に統語的曖昧性のパッキング (音声認識候補のマージ) をすることが可能である (図 2)。

この処理は Tomita 法 [3] の packed forest を複数文を同時に扱う一般化 LR パーザに拡張したものと解釈することもできる。

また、従来の HMM-LR 法は音素に同期した横型探索を基本としているため、時間が進行するにつれて、照合音素の存在可能範囲が徐々に広くなっていく。計算量削減のために、

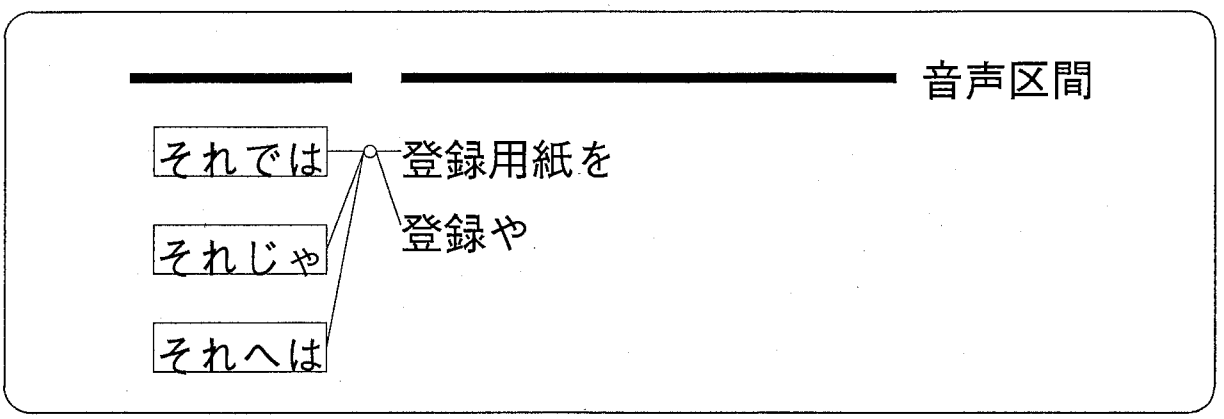


図 2: 統語的曖昧性のパッキング (音声認識候補のマージ)

検出されたポーズ情報を利用して、照合音素の存在可能範囲を狭めることが可能である (図3)。

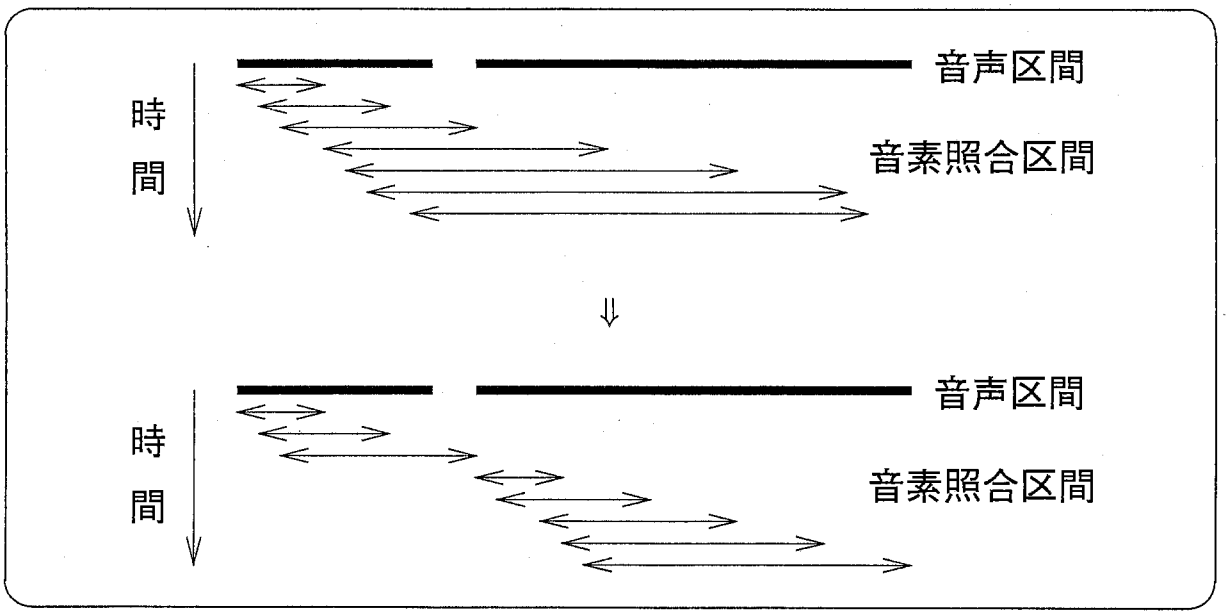


図 3: 音素照合区間の削減

3 セルのマージとスプリット処理を組み込んだ HMM-LR 音声認識アルゴリズム

ポーズは事前に何らかの方法で検出されているとした時の音声認識アルゴリズムの概要を記す。

3.1 セルのデータ構造

従来のHMM-LR法のセル(解析に必要な情報を保持するデータ構造)は、並列解析を可能にするために複数個のセルを単一連結していた。状態スタックの1番上の内容(スタックトップ)が同じセルをマージするための、マージポインタを付加した2重連結構造に、セルのデータ構造を変更する。さらに、ポーズ区間処理のためのセルリスト(ポーズセルリスト)を新たに用意する。

3.2 提案するアルゴリズム

まず、セルのマージ処理と、ポーズの同期処理の要点を記す。

(1) ある枝でポーズが検出(シンボルスタックのトップがポーズ)され、音声の入力フレームがそのポーズ区間の末端まで到達していれば、そのセルをポーズセルリストに登録する。

(2) ビーム探索による枝刈りか、あるいは、統語的に棄却されることで、枝が伸ばせなくなったら、ポーズセルリストに登録されている枝に対し reduce 操作を行なう。「ある統語カテゴリ集合」に属する要素にならない枝をすべて枝刈りする。

(3) 残った枝で状態スタックの1番上の内容が同じものをマージする。音声認識スコアは1番よいもので代表させる。

「ある統語カテゴリ集合」には任意の統語カテゴリを定義することが可能である。今回の評価実験では、音声翻訳システムASURA [4]のために開発された文法規則を利用し、そこで文節カテゴリとして定義されているものを、この統語カテゴリ集合とした。例えば、単語境界にポーズが入るような発話を許容するには、その統語カテゴリ集合をすべての単語区切りに変更すればよい。

次に、セルのスプリット処理の要点を記す。

(4) マージされた位置よりさかのぼって処理をしなければならない時、ポインタを張り換えてセルをスプリットする。音声認識スコアは元の値に戻す。

4 実験と評価

評価実験を行なった。結果を表1に示す。HMM音素モデルは、文献[5]と同じもの(音素環境に依存しない音素モデル)を採用している。ただし、文法規則のバイグラム等の統計的な言語モデルは使っていない。

評価用音声データは、話者1名により、文単位で丁寧に発話された137文である。この実験では、人手によりラベル付けられたポーズ情報のうち、60ms以上のものを検出されたポーズとして採用した。人手によりラベル付けられたポーズ時間の分布を表4に示す。

ローカルビーム幅(1つの枝からの最大分岐数)は12とした。処理時間を計測したマシンはHP9000/755である。

表 1: 文を発話単位とする音声認識実験結果の比較

実験条件	ビーム幅	文音声認識率 (%)		音素照合数／文	処理時間／文 (s)
		1 位	～5 位		
original	50	64.2	70.1	9,813	19.51
HMM-LR	100	65.7	73.7	18,323	37.05
文献 [5]	200	66.4	75.9	33,576	62.85
	400	68.6	80.3	68,217	158.44
ポーズ区間	50	68.6	75.2	9,370	9.21
処理付き	100	68.6	77.4	17,923	18.52
HMM-LR	200	70.1	81.0	33,806	38.21
	400	70.8	82.5	62,663	82.37
ポーズ区間	50	68.6	75.2	9,890	11.42
パッキング	100	70.1	78.8	18,318	22.96
処理付き	200	70.8	81.8	34,946	42.29
HMM-LR	400	70.8	82.5	69,455	134.06

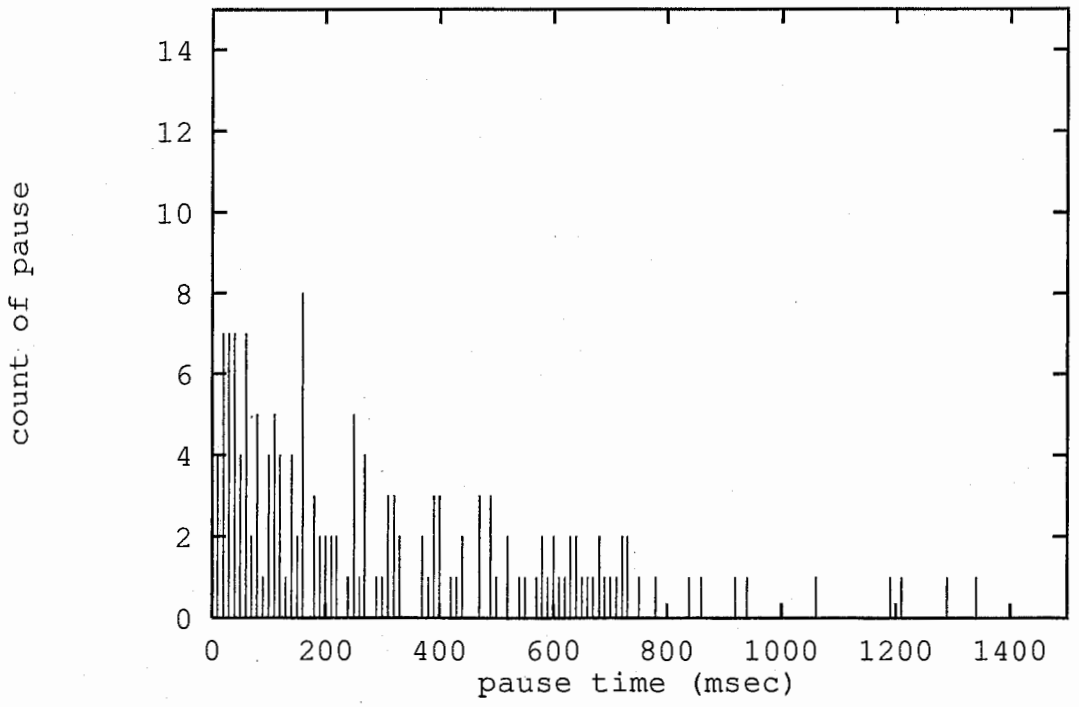


図 4: ポーズ時間の分布

「ポーズ区間処理付き」とは、音素照合区間の削減処理のみを組み込んだものである。「ポーズ区間パッキング処理付き」とは、ここで提案する新しいアルゴリズムによるものである。オリジナルの手法より、音声認識率、処理時間ともに改善されたことがわかる。

5 むすび

自然な発話の受理を目指して、HMM-LR 連続音声認識手法の改良を行なった。文節毎に区切って丁寧に発話しようとしても、文節の判断が揺れることがある。ここで提案する手法は、丁寧に発話された連文節発話を受理するものとして有効である。

今後は、ポーズの自動検出方法の高度化や、音素環境依存 HMM モデル [6] の扱える予測 LR パーザのアルゴリズム、ポーズの確率的な存在を許す予測 LR パーザのアルゴリズム、自然にポーズが置かれやすい箇所を考慮した音声認識のための日本語文法の研究を行なう予定である。また、この手法をベースにして、さらに自然な発話を受理する連続音声認識手法の研究を進めていく。

謝辞

実験を進めるうえで支援していただいた林輝昭氏に感謝いたします。

参考文献

[1] 北研二, 川端豪, 斎藤博昭, "HMM 音韻認識と拡張 LR 構文解析法を用いた連続音声認識," 情報処理学会論文誌, **31**-3, pp. 472–480, 1990.

[2] 南泰浩, 鹿野清宏, 高橋敏, 山田智一, "音韻環境依存 HMM と候補のマージを用いた不特定話者大語彙連続音声認識," 日本音響学会 平成 *5* 年度秋季研究発表会 講演論文集, 2-7-5, pp. 79–80, 1993.

[3] M. Tomita, *"Efficient Parsing for Natural Language – A Fast Algorithm for Practical Systems,"* Kluwer Academic Publishers, 1986.

[4] 竹沢寿幸, 森元逞, 谷戸文廣, 鈴木雅実, 嵯峨山茂樹, 樽松明, "ATR 音声言語翻訳実験システム ASURA", 情報処理学会第 *46* 回全国大会, 6B-5, 1993.

[5] K. Kita, T. Morimoto, K. Ohkura, and S. Sagayama, "Continuously Spoken Sentence Recognition by HMM-LR," *Proc. of ICSLP 92*, pp. 305–308, 1992.

[6] 永井明人, 鷹見淳一, 嵯峨山茂樹, H. Singer, "隠れマルコフ網と一般化 LR 構文解析を統合した連続音声認識," 信学論, **J77-D-II**-1, pp. 9–19, 1994.

[7] H. Tanaka, T. Tokunaga, and M. Aizawa, "Integration of Morphological and Syntactic Analysis Based on LR Parsing Algorithm," *Proc. of 3rd Int. Workshop on Parsing Technologies (IWPT)*, pp. 101–109, 1993.

A 付録: データ構造と動作例

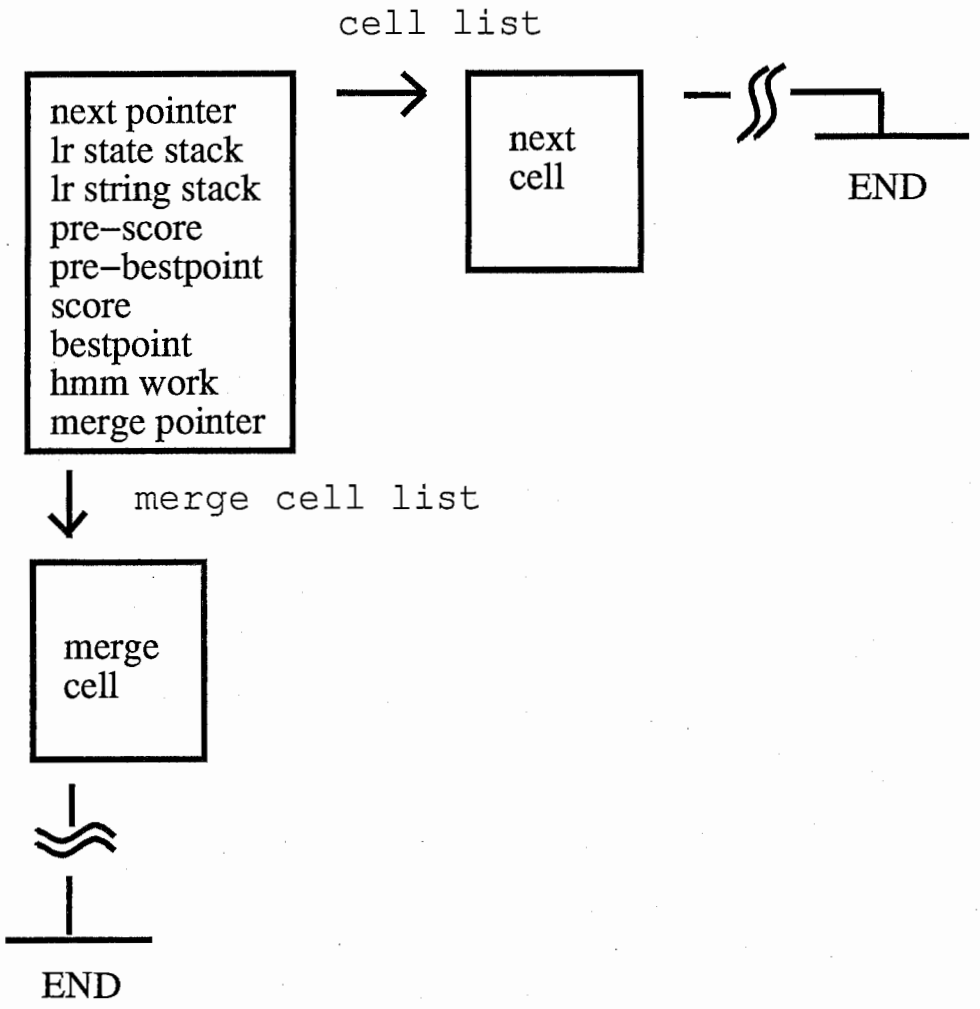
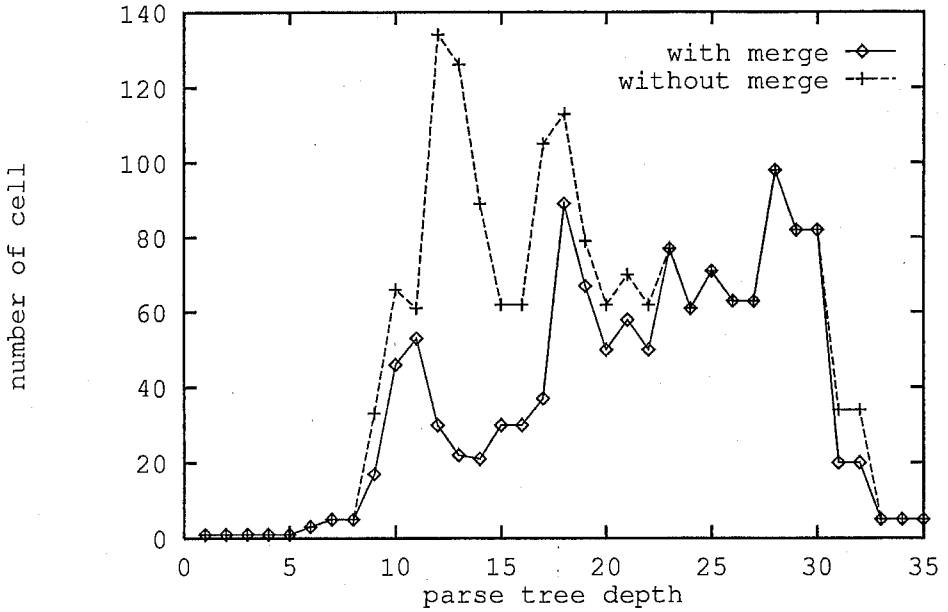


図 5: データ構造



n a m a e w a pau s u z u k i m a y u m i d e s u pau
n a m a e w a pau i i pau s u z u k i m a y u m i d e s u pau
n a m a e n o pau s u z u k i m a y u m i d e s u pau
n a m a e n i pau s u z u k i m a y u m i d e s u pau
n a m a e d e pau i u pau s u z u k i m a y u m i d e s u k a pau
n a m a e w a pau s u z u k i m a y u m i d e s u k a pau
n a m a e w a pau n a m a e w a pau s u z u k i m a y u m i d e s u k a pau
n a m a e w a pau s u z u k i m a y u m i d e s u n e pau
n a m a e w a pau i i pau s u z u k i m a y u m i d e s u n e pau

図 6: 複数文並列処理 LR パーザの動作