

アクセント、ポーズ位置情報を用いた日本語音節文候補の絞り込み

荒木 哲郎

池原 悟

福井大学 工学部

NTT コミュニケーション科学研究所

横川 秀人

鈴木 誠

福井大学 工学部

福井大学 工学部

1 はじめに

音声信号は、声道の共振特性である言語的な情報(音韻情報)と、有声音源の繰り返し周期(ピッチ)特性である韻律情報(アクセントやイントネーションで、話者の個人性、情緒性、音声の自然性に寄与するもの)から成り立っていることが知られている。現状ではピッチの抽出が必ずしも容易でないことや、個人差が大きいことなど研究課題も多いが、音声合成への応用や音声認識への応用が期待されている[1][2]。

またこれまでにアクセント句境界やアクセント核の位置が、正しく検出された場合において、アクセント位置に関する情報量を定量的に測定し、これがかかなり大きい情報量であることが示されている。[3] 日本語音声認識システムにおいては、一般に音韻的、形態素的、構文的、意味的な性質が音声認識の曖昧さを解消するのに有用であると一般に考えられており、自然言語処理が音響処理に基づく認識アルゴリズムによって得られた音節認識候補ラティスに適用され、今までに、音節文字の2重マルコフモデルが音節候補の曖昧さ(主に置換誤り)の解決に有効であることが知られている[5]。

本報告では、日本語の連続音声認識に韻律情報を応用するための研究の第一ステップとして、アクセント句境界の位置やアクセント核の位置が正しく検出された場合に、従来の音節マルコフ連鎖確率モデル[5][6]に加えて、さらにアクセント位置についてのマルコフ連鎖情報を用いた音節認識候補の絞り込み方法を提案する。さらに新聞記事を用いた音節マトリックスシミュレーション実験を行い、その有効性を定量的に把握する。

2 諸定義とアクセント位置情報を用いたマルコフ連鎖モデルによる音節認識候補の絞り込み方法

音声認識装置(音響処理)から出力された音節認識候補(音節ラティス)の中から、マルコフモデルを用いて正しい音節列を決定する方法の有効性を評価する為の、日本語音声入力システムのモデルを図1に示す。

日本語文は文節と呼ばれる単位に分割される。ここで文節は、一つ以上の自立語=名詞、動詞、形容詞、形容動詞、副詞、連体詞、接続詞、感動詞、形式名詞と附属語=助詞、助動詞、接辞からなり、一般に前者は必ず文節に含まれるが、後者は必ず含まれるとは限らない。

さらに、音節、およびアクセント、ポーズの情報は、日本語の漢字かな交じり文に対する日本語音声出力システム[4]システムの出力結果として得られることが知られている。最初に日本語音声出力システム[4]によって与えられるポーズ位置情報、及びアクセント核をアクセントの位置情報として、その位置情報を考慮した音節マルコフ連鎖モデルについて述べる。

【定義1】アクセントの情報が付与されていない音節文字の集合を S と表す時、ポーズ位置およびアクセント核が存在する位置の音節文字 $s(s \in S)$ を、それぞれ \hat{s} および \check{s} と表し、その集合をそれぞれ \hat{S} および \check{S} と表す。音節列 $x=s_1s_2\ldots s_n$ において、 s_i が全て S の要素から成っているとき、 x をアクセント無しの音節列とよび、また $s_i \in$

$(S \cup \check{S}), s_i \in (S \cup \check{S})$ および $s_i \in (S \cup \check{S} \cup \check{\check{S}})$ から成り立っていると、 x をそれぞれポーズ付きの音節列、アクセント付きの音節列、ポーズおよびアクセント付きの音節列と呼ぶことにする。このときアクセント無しの音節列、ポーズ付きの音節列、アクセント付きの音節列、ポーズおよびアクセント付きの音節列に対する音節マルコフ連鎖確率値の集合を、それぞれ NM, PM, AM および PAM と呼ぶ。

但し、本報告では、アクセント付きおよびアクセント無しの2重の音節マルコフ連鎖確率を扱うことにする。

【定義2】文節単位に発声された連続音声に対して、音響処理の結果得られる音節認識装置からの曖昧な音節認識候補を表したものを、音節マトリックスと呼ぶ。但し、セグメンテーションは正しく行われ(すなわち音節区間は正しく認識され、脱落・挿入誤りは無く候補の置換誤りだけが存在し)、正解候補は其中に必ず存在するものとする。このとき文節内に存在するアクセント核の位置が正しく与えられた仮定したとき、その位置の音節候補を全て、アクセント付きの音節文字(S の要素)によって表現した音節マトリックスを、特にアクセント付きの音節マトリックスと呼ぶ。

ポーズ位置およびアクセント位置を含む音節マトリックスの例を図2に示す。

以上定義1および2では文節を単位として定義してきたが、次に日本語音声出力システムから出力されるポーズ情報を用いて、ポーズから次のポーズの区間までを一つの単位として組み立て、音節マルコフ連鎖確率、および音節マトリックスを定義1、2と同様に定義する。次にアクセント付きの音節マトリックスからの候補絞り込み法を次に示す。

【音節候補の絞り込み法】ポーズ位置およびアクセント付きの音声マトリックスの音節候補を組み合わせて得られる音節列候補を、アクセント無し及びポーズおよびアクセント付きのマルコフ連鎖確率 NM, PM, AM, PAM を用いて、最尤な音節候補列を選ぶ方法をそれぞれ NM-法, PM-法, AM-法及び PAM-法と呼ぶ。

音節ラテイスの各列は、最大10候補からなり、正しい候補がその中に含まれるものとする。実験では、110音節文字が使われる。各音節ラテイスの候補は、次のような制限を持つ対表的な音声認識から出力される音声認識候補の特性に基づいた、コンフュージョンマトリックス(各音節に対する可能な置換誤りのリスト)を用いて疑似的に生成される。(1)置換誤りは、正しい母音と同じ母音を持つ音節の間でのみ生じる("ア"は、"タ"や"パ"には誤るが、"イ"や"ク"には誤らない)。(2)2重子音"n"や長い母音は誤らない。(3)候補の平均正解率は89%である。

特に、音響処理におけるセグメンテーション過程には誤りがなく、各音節ラテイスには常に正解の音節列が含まれていると仮定する。

3 実験結果

3.1 実験条件

1. マルコフ連鎖確率辞書作成に用いた日本語文：新聞記事 77 日分

- (a) 総音節文字数: 5,261,139 音節
- (b) 総アクセント文字数: 679,125
- (c) 総ポーズ文字数: 558,081
- (d) マルコフ連鎖確率辞書の種類: NM, PM, AM, PAM

2. 入力データの文の種類

- (a) 総文数: 200 文
- (b) 総ポーズ数: 932
- (c) 総アクセント数: 1172
- (d) 1 文当たりの平均長: 54.8 文字

3.2 実験結果

曖昧な音節認識候補の絞り込みにおけるポーズ位置およびアクセント位置情報の効果を、定量的に把握するために、ポーズ位置およびアクセント付きの音節マトリックスによって与えられた音節認識候補を、それぞれ組み合わせて得

られる音節列に対して、ポーズ位置およびアクセント付きの音節マルコフ連鎖確率PM, AM, PAMを用いて評価し、10位内に文節正解候補が入る正解率を求めた。その実験結果を図4に示す。またPAMによる10位内正解音節列候補の例を図5に示す。

[1] アクセント有り/無しの2重マルコフ連鎖確率のエントロピー

アクセント有り/無しの2重マルコフ連鎖確率が、77日間の新聞記事統計データを用いて求められた。音節の統計データを用いたマルコフ連鎖確率の飽和特性を図2に示す。同図から、ルール(0でない2重マルコフ連鎖確率 $p(x_i|x_{i-1}x_{i-2})$ の3組 (x_{i-2}, x_{i-1}, x_i) の集合)の80%は飽和していることがわかる。統計データが増加するに従って、2重マルコフ連鎖確率の飽和度は改善される。マルコフ連鎖確率のエントロピーが小さくなれば、マルコフ連鎖確率を用いた正しい候補の決定がより効果的となる。ポーズおよびアクセント有り/無しのマルコフ連鎖確率NM, PM, AMおよびPAMのエントロピーはそれぞれ3.78, 3.28, 3.33および2.88とであった。

[2] ポーズおよびアクセント有りの2重マルコフ連鎖確率を用いた音節候補の絞り込み効果

ポーズおよびアクセント有り2重マルコフ連鎖確率PM, AM, PAMを用いた音節候補の絞り込み効果を、図4に示す。

同図よりPM, AM, PAMを用いる効果は、NMに比べて第一位正解率でそれぞれ12.5%, 15.5%, および15.6%向上することがわかり、また10位内累積正率でそれぞれ13%, 14.5%および28%の向上が図れることがわかった。

4 おわりに

本報告では、日本語の連続音声認識などに韻律情報を応用するための研究の第一ステップとして、ポーズ位置およびアクセント句境界の位置やアクセント核の位置が正しく検出された場合に、これらのマルコフ連鎖情報を用いて音節認識候補を絞り込む方法を提案し、その有効性を、新聞記事を用いた音節マトリックスシミュレーション実験によって行った。その結果、次の知見を得た。

1. ポーズ位置のみ付加したマルコフ連鎖確率をもちいることにより、用いない場合に比べて第一位正解率は12.5%向上することがわかり、また10位内に正解候補が13%の精度で得られることがわかった。
2. ポーズ位置およびアクセント位置を付加したマルコフ連鎖確率を用いることにより、両方とも用いない場合に比べて第一位正解率が15.6%向上すること、10位内累積正解率が28%となることがわかった。

(参考文献)

- 1) 広瀬、藤崎、河井、山口：基本周波数パターン生成過程モデルに基づく文章の音声の合成、信学論、J72-A,1, pp32-40 (1989)
- (2) 遠藤、小林、白井：韻律情報を用いた構文の推定とその音声認識への応用、信学技法、SP91-103, pp9-14 (1991)
- (3) 村上、荒木、池原：音声におけるアクセント情報の持つ情報量の考察、信学会音声研究会、SP91-50 pp13-19 (1991)
- (4) 宮崎、大山：日本文音声出力のための言語処理方式、情処論、27,11, pp1053-1061 (1986)
- (5) 荒木、村上、池原：2重マルコフモデルによる日本語文節音節認識候補の曖昧さの解消効果、情処論、30,4, pp467-477 (1989)
- (6) 村上、荒木、池原：日本語文節入力に対して2重マルコフ連鎖モデルを用いた漢字かな交じり候補の抽出精度、信学論、D-II, J75-D-II,1, pp11-20 (1992)
- (7) 荒木、池原、横川：アクセント位置情報を用いた日本語音節文候補の絞り込み、情処学会NL研究会、104-5, pp33-38 (1994)
- (8) Araki, Ikehara, Yokokawa: Using Accent Information to Correctly Select Japanese Phrases Made of Strings of Syllables, International Conference on Spoken Language Processing, Vol.4, pp2155-2159 (1994)

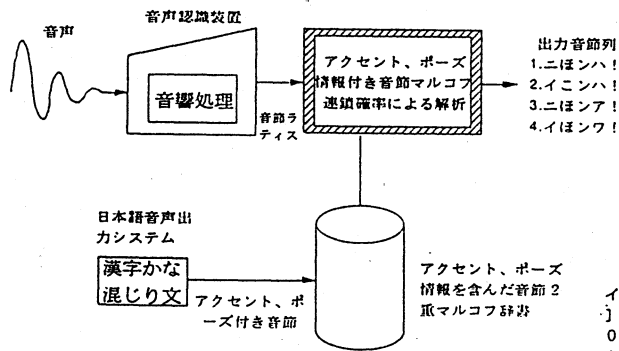


図1 日本語音声入力システムモデル

アクセント位置		ポーズ位置	
↓		↓	
b	b	ニ	ぼ
		ミ	お
		イ	こ
		ビ	ぽ
			ワ!
			ア!
			バ!
			ハ!
			バ!

*注: カタカナ表記はアクセントなし
ひらがな表記はアクセントあり
!付き文字はポーズあり
bは空白

図2 アクセント、ポーズ付き音節マトリックスの例

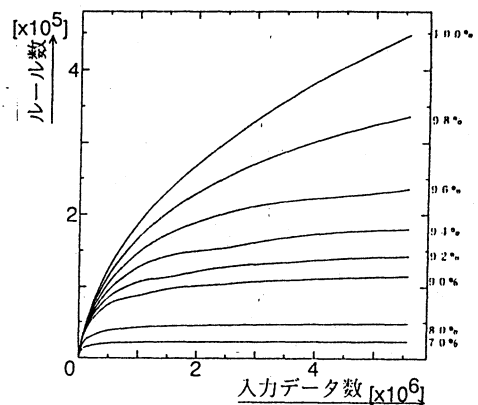


図3 アクセント、ポーズ情報付加文辞書の飽和特性(2重)

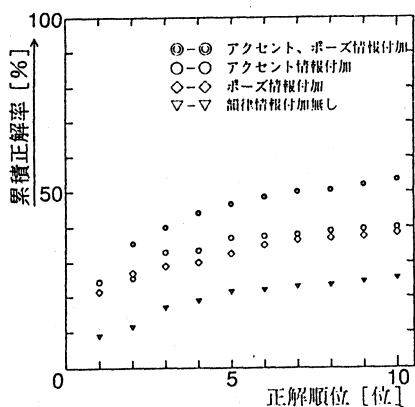


図4 アクセント、ポーズを用いた音節候補絞り込み実験結果(標本内)

キ ケ ンガ! あル!
ヒ ヘ タ! ばブ!
チ ゲ ア! はギ!
ビ テ ラ! たウ!
イ エ カ! プ!
ギ ベ ワ! ヌ!
バ!

1	キケンガ!あル!	27.92306
2	イケンガ!あル!	28.08509
3	イテンガ!あル!	30.04572
4	チケンガ!あル!	30.70501
5	キテンガ!あル!	31.20705
6	キゲンガ!あル!	31.34461
7	チテンガ!あル!	32.69486
8	キケンワ!あル!	32.74089
9	イケンワ!あル!	32.90292
10	チゲンガ!あル!	33.20127

図5 アクセント、ポーズ付加文の絞り込み例