

アクセントを用いた音声認識精度向上の検討*

堀田波星夫 (鳥取大), 菊井玄一郎, 山本博史 (ATR), 村上仁一

1 はじめに

音声認識におけるアクセント情報の有効性を調査した研究として [4] がある。[4] では、日本語音声のポーズとアクセントの位置の持つ情報量を、定量的に評価している。しかし、実際の音声認識システムにおいて、アクセント情報の有効性は調査されていない。そのため、本研究では、音声のアクセントの情報が認識できたと仮定し、アクセント情報を利用して、文音声認識でどの程度認識精度が向上するかを調査した。アクセント情報は、多くの情報があるが、本研究はアクセント句とアクセント核の位置を利用する。

以下のように、アクセント情報を利用した文音声認識精度の向上を調査した。

1. 音声の発話文に対してアクセント情報付きの形態素解析を行い、発話文のアクセント情報を求める。
2. 100-best で出力された認識結果に付与されている単語のアクセントや品詞等の情報から、アクセント情報を求める。
3. 音声の発話文と文認識結果のアクセント情報を尤度が高い順に比較し、一致する結果を認識結果とする。

2 アクセント情報の決定

本研究では、音声データからアクセント情報を抽出するのではなく、テキストからアクセント情報を得る。また、アクセント情報を、アクセント句とアクセント核の位置と定義する。

評価データのアクセント情報は発話文の形態素解析の結果から得る。また、文認識結果のアクセント情報は単語に付与された形態素の情報から得る。両者の取得は音声合成ツールの XIMERA(Ver1.1)[2] の API を利用する。

形態素解析には茶筌 (Ver2.3.2) と茶筌用の日本語辞書 ipadic(Ver2.6.2)[3] を使用する。形態素解析の結果は、“アクセント型-モーラ数, アクセント結合情報 [2]” の形式の付加情報を含む。形態素解析結果を表 1 に示す。単語”私”に対する、読み、基本形、品詞、活用型、活用形、発音、付加情報であり、存在しない情報は空欄である。“私”のアクセント型が 0 型、モーラ数が 3 であり、アクセント結合情報が C2 であるので、付加情報は”0-3:C2”である。

本研究では、アクセントの情報はアクセント句毎に、“アクセント核の位置-アクセント句のモーラ数/”

Table 1 単語”私”に対する形態素情報

読み	基本形	品詞	活用型
ワタシ	私	名詞-代名詞-一般	
活用形	発音	付加情報	
	ワタシ	0-3:C2	

とする。例えば、“食前酒を下さい”のアクセント句は“食前酒を”と“下さい”に分かれる。“食前酒を”のアクセント型は 3 型であり、モーラ数は 7 であるので、アクセント情報は”3-7/”となる。“下さい”も同様にして、“食前酒を下さい”のアクセント情報は”3-7/3-4/”となる。

3 評価実験

3.1 言語モデルと音響モデル

形態素解析結果の読み、基本形、品詞、活用型、活用形、発音、付加情報の情報を単語に付与し、単語 N -gram モデルを作成して言語モデルとする。

形態素の発音の情報とカナから音素への対応表に基づいて、単語の音素列を得る。決定した音素列に基づいて、音響モデルを作成する。

3.2 実験条件

ATRIUMS(Ver2.3.1)[1] は音声認識研究のための実験環境構築やアプリケーション、プロトタイプシステムを構築することを目的としたツール群である。これを実験に用いる。

ATR の BTEC1 は海外旅行者のために旅行会話で一般的に使われる表現を対訳付きで列挙している。学習データはこの日本文の 16 万 2 千文を N -gram の学習に用いる。評価データは BTEC1 の set01 の 510 文、set02 の 508 文、set03A の 506 文、set03B の 506 文を用いる。

3.3 実験方法

1. ATRIUMS を用いて文認識を行い、評価データ文に対する 100-best を得る。
2. 100-best の尤度が高い順に、評価データの発話文のアクセント情報と比較し、一致した文を認識結果とする。
3. 1-best を認識文とした結果と 2 の結果を比較して認識精度を調べる。

実験方法の例を図 1 に示す。評価データを”いい旅行をなさって”としたときの実験方法である。なお単語の読み、基本、品詞形等の情報は省力している。100-best 中で、1-best はアクセント情報が一致せず、

*Examination of the improvement in speech recognition accuracy using the accent. by HOTTA Haseo(Faculty of Engineering, Tottori University), KIKUI Genichiro, YAMAMOTO Hirofumi (ATR), MURAKAMI Jin'ichi

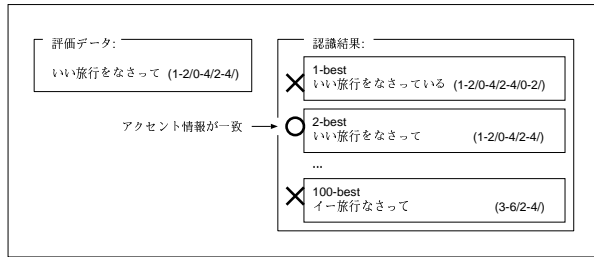


Fig. 1 実験方法の例

2-best でアクセント情報が”1-2/0-4/2-4/” と一致しているので認識結果とする。

3.4 正解の基準

認識結果において、単語、読み、基本形、品詞、活用型、活用形、発音、付加情報の情報が完全に一致する結果を正解とする。また、表記の仮名と漢字の異なりと付加情報の異なりを、同等の意味であると考え正解とする。

4 実験結果

表 2 に実験結果を示す。アクセント情報を用いた文認識結果の内訳を表 3 に示す。なお、比較として、モーラ数のみを用いた実験結果も示している。ここで、モーラ情報はアクセント句でのモーラ数とする。

全ての評価データセットの結果において、アクセント情報を用いた結果、モーラ情報を用いた結果、1-best の結果の順に文認識率が高い。アクセント情報とモーラ情報は、1-best が正解ではなく、100-best 中に正解が存在する結果しか改善できない。そのため、表 3 において、アクセント情報とモーラ情報を用いた結果で、1-best での正解率と 100-best 中に正解が存在しない結果の率は、変わっていない。

アクセント情報を用いた結果、2-best から 100-best 中に正解が存在する 267(174 + 93) の結果のうち約 35% ($\frac{93}{267} \times 100$) が正解となる。モーラ情報を用いた結果では、改善可能な結果のうち、約 28% ($\frac{76}{267} \times 100$) が正解となる。

Table 2 文認識率

	1-best	アクセント情報を使用	モーラ情報を使用
set01	61.6% (314/510)	65.7% (335/510)	65.1% (332/510)
set02	64.0% (335/508)	69.1% (351/508)	67.7% (344/508)
set03A	61.7% (312/506)	66.2% (335/506)	65.6% (332/506)
set03B	64.0% (312/506)	68.6% (347/506)	67.8% (343/506)
合計	62.81% (1275/2030)	67.39% (1368/2030)	66.55% (1351/2030)

5 考察

5.1 アクセント情報の有効性

実験の結果、文のアクセント情報を用いることで、文認識率が向上することを確認した。文認識率に注目

Table 3 文認識結果の内訳

	アクセント情報	モーラ情報
1-best の正解率	62.81% (1275/2030)	62.81% (1275/2030)
改善率	4.58% (93/2030)	3.74% (76/2030)
100-best 中に正解が存在し改善されなかった率	8.57% (174/2030)	9.41% (191/2030)
100-best 中に正解が存在しない率	24.04% (488/2030)	24.04% (488/2030)

すると、1-best の結果から約 5% の向上であり、2-best から 100-best 中に正解がある結果の約 35% が改善された。アクセント情報を用いた文認識率とモーラ情報を用いた文認識率は差が小さく、アクセント情報の中でモーラ情報の情報量が大きいことがわかる。

5.2 発音の誤り

単語の発音と実際の音声と異なっている事例がある。例えば、形態素解析により発音を「何時」を「イツ」、「新色」を「シンイロ」、「コーラの中をください」を「コーラノナカラクダサイ」としている。発音と音声の異なりのために、100-best 中に正解が存在しない認識結果がある。正確な認識率を求めるためには発音を訂正する必要があるが、改善可能な認識結果のうちのアクセント情報で改善される結果の割合はあまり変わらないと考えられる。

6 結論

本研究では、アクセント情報付きの形態素解析の結果からアクセント情報を決定することによって、アクセント情報を認識できたと仮定し、アクセント情報を利用してどの程度文音声認識率が向上するかを調査した。

その結果、改善可能な文認識結果の約 35% を改善することができた。また、ほとんどの結果はモーラ情報によって改善された。

謝辞

本研究は情報通信機構の研究委託「大規模コーパススペース音声対話翻訳技術の研究開発」により実施したものである。

参考文献

- [1] *ATRIUMS(Ver2.3.1)*. ATR 株式会社 国際電気通信基礎技術研究所。
- [2] *XIMERA(Ver1.1) and API Reference*. ATR 株式会社 国際電気通信基礎技術研究所。
- [3] *chasen-2.3.2 and ipadic-2.6.2*. 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座 (松本研究室), 2003.
- [4] 村上, 荒木, 池原. 音声におけるポーズ長およびアクセント位置の情報量の考察. 日本音響学会講演論文集, No. 3-3-11, pp. 89-90, 1988.