

## 日本語学習支援のためのアクセント認識の検討\*

☆ショート グレグ, 山田武志, 北脇信彦, 牧野昭二 (筑波大)

## 1 はじめに

近年、日本への留学生が急増している[1]。留学生の多くは日本語の学習支援を必要とするものの、中でもアクセントの指導は不十分であるのが現状である。日本語のアクセントは自然性に加えて意味の区別にも多大な影響を及ぼすため、学習者のアクセントの誤りを自動的に検出し、同時にそれを分かり易く指摘するような学習支援が必要である[2]。

従来、日本語のアクセント型を判定するための研究が行われてきた[3]。これらの研究の多くは、入力音声が既定のアクセント型のいずれであるのかを判定するものであった。しかし、学習者は既定のアクセント型には存在しないアクセントパターンを使う可能性がある。よって、このような不特定のアクセントパターンを忠実に認識することが重要である。本稿では、入力音声の基本周波数の変化をモーラ単位の高低パターンとして認識する手法を提案し、その有効性を検証する。

## 2 提案手法

## 2.1 問題設定

学習者は、第一言語干渉により、既定のアクセント型にはないアクセントパターンを使う可能性がある。例えば、米英語を第一言語とする話者が「居酒屋」と発話したとき、基本周波数は Fig. 1 のように変化した。これをモーラ単位の高低パターンとして表すと Fig. 2 のようになる。これは東京方言には存在しないアクセント型である。よって、このような不特定の高低パターンを自動的に認識することが求められる。N モーラの単語の場合、起こりうる高低パターンは2のN乗個である。なお、学習支援が目的であるため、認識対象である音声データの発話内容は既知として扱う。

## 2.2 提案手法の概要

不特定の高低パターンを認識する手法について述べる。提案手法の処理の流れを Fig. 3 に示す。まず、音声認識 (Forced alignment)

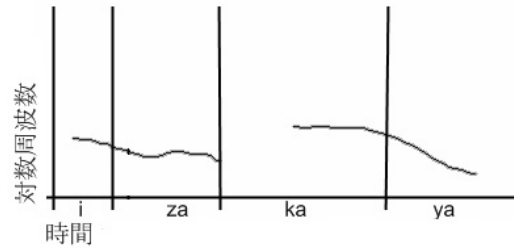


Fig. 1: F0 for non-natives “izakaya”

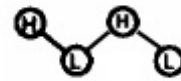


Fig. 2: High/Low pattern for “izakaya”

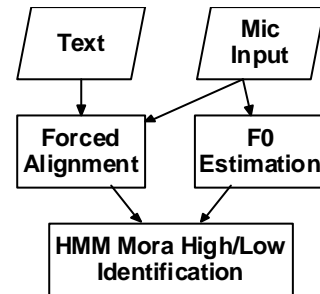


Fig. 3: High/Low pattern recognition flow chart

を用いて入力音声のモーラ境界を検出する。また、入力音声の基本周波数を短時間フレーム毎に抽出する。そして、モーラ間の基本周波数の高低関係を反映する特徴量 (後述) を抽出し、高低パターンを認識する。

高低パターンの認識方法について述べる。人間がアクセントパターンを認識するとき、モーラ間のピッチの高低関係を基本的な手がかりにすると考えられる。このことから、提案法では、認識のためのテンプレートを2つ以上の連続するモーラを単位として用意し、高低パターンの認識を行う。上記の「居酒屋」の例において、2 モーラを単位にすると認識結果は Fig. 4 のようになる。

## 2.3 特徴量

本稿では比較のために、MeanSlope、VMTarget、LQR の3種類の特徴量ベクトルを

\* Lexical accent recognition for Japanese language learning purposes, by SHORT, Greg, YAMADA, Takeshi, KITAWAKI, Nobuhiko, and MAKINO, Shoji (University of Tsukuba).

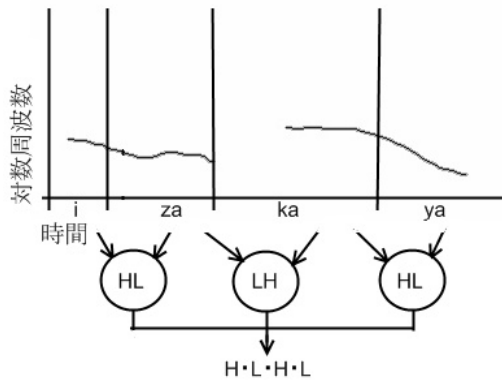


Fig. 4: Recognition result for “izakaya”

抽出した。各特長量ベクトルは、モーラ毎に抽出した特徴量を、対象とするモーラ数だけ並べたものである。以下、モーラ毎に抽出した特徴量について述べる。

□ MeanSlope

対数 F0 の平均値、及び対数 F0 の線形回帰線の傾きを求めた。

□ VMTarget

母音の最後の 2 分の 1 の区間における対数 F0 の平均値、及び対数 F0 の線形回帰線の終点における対数 F0 を求めた。

□ LQM

母音の最後の 4 分の 1 の区間[4]における対数 F0 の平均値を求めた。

上記の傾き以外の特徴量は、各々において求めた対数 F0 の平均の最低値を減算することにより正規化した。

### 3 実験

#### 3.1 実験条件

音声資料として、ATR 音声データベースの男女各 2 名 (mau, mht, fkn, fsu) の単語音声データを用いた。正解の高低パターンは、NHK 日本語発音アクセント辞典をベースにし、実際に音声データを聴取した上で決定した。一部、既定のアクセント型には存在しない高低パターンも含まれていた。本稿では検討の初期段階として、これらの音声資料を学習と評価に共通に用いた。また、認識のためのテンプレートとしては、多次元混合ガウスに基づく確率モデルを用いた。ここで、テンプレートの単位は 2 モーラとした。

#### 3.2 実験結果

以下では、評価尺度としてモーラ毎に正解を判定したときの正解の割合を用いる (以下

Table 1: Overall recognition rate (%)

	MeanSlope	VMTarget	LQM
認識率	85.2	85.0	84.5

Table 2: Recognition Rate for accent types (%)

Type	MeanSlope	VMTarget	LQM
0	85.8	94.1	90.4
1	83.8	61.7	60.5
2	83.7	86.1	88.0
3	90.2	91.8	89.3
4	86.3	89.5	89.2
5	85.8	90.6	91.0

では認識率と呼ぶ)。

まず、評価データ全体に対する総合的な認識率を Table 1 に示す。認識率はいずれの特徴量の場合も 85% 程度であった。次に、アクセント型毎の認識率を Table 2 に示す。Table 2 より、1 型に対する認識率については、VMTarget と LQM は MeanSlope に比べて顕著に低いことが分かる。この原因については今後詳しく調査する必要がある。

## 4 おわりに

本稿では、入力音声の基本周波数の変化をモーラ単位の高低パターンとして認識するという問題を設定した。そして、そのための一手法を提案し、有効性の初期検討を行った。実験の結果、一定の認識性能は得られたものの、絶対的な性能は十分とは言えない。今後は、学習・評価用音声資料を増やす、特徴量の見直しを行うなどして性能改善を図ってきたい。

#### 参考文献

- [1] 日本学生支援機構, “外国人留学生在籍状況調査結果,” 1-11, 2009.
- [2] Kawai, G., Ishi, C.T., “A system for learning the pronunciation of Japanese pitch accent,” Proc Eurospeech’99, Vol. 1, pp. 177-181, 1999.
- [3] 石井カルロス寿憲, 峯松信明, 広瀬啓吉, “ピッチ知覚を考慮した日本語連続音声のアクセント型判定,” 電子情報通信学会技術研究報告, Vol. 101, No. 270, pp. 23-30, 2001.
- [4] Yi Xu, Q. Emily Wang, “Pitch targets and their realization: Evidence from Mandarin Chinese,” Speech Communication, 33, pp. 319-337, 2001.