

日本語音声教育のための 日本語アクセント型判定法

熊谷 有香、吉田 奏子、三輪 譲二
yuka,miwa@cis.iwate-u.ac.jp

岩手大学 工学部 情報工学科

〒020-8552 岩手県盛岡市上田4-3-5
019-621-6974

あらまし 本論文では、日本語音声教育のための日本語アクセント型判定法を提案する。アクセント型の判定では、相対アクセント位置および対数基本周波数の傾きの2つの特徴を用いる。

アナウンサーの音声について、97%の割合でアクセント型を正しく判定した。よって本判定アルゴリズムは有効であるといえる。また、留学生のアクセントの正答率は69%であったが、間違った発音に対しては異なるアクセント型に判定された。よって日本語アクセント発音学習に役立てることができるといえる。

キーワード 音声言語教育、日本語音声、アクセント

On a Decision Method of Accent Type for Japanese Learning

Yuka Kumagai, Kanako Yoshida and Jouji Miwa
yuka,miwa@cis.iwate-u.ac.jp

Department of Computer and Information Science,
Faculty of Engineering, Iwate University

4-3-5 Ueda Morioka-shi Iwate-ken, 020-8551 Japan
019-621-6974

Abstract We propose a decision method of accent type for computer assisted Japanese language learning. In the method, we use two features such as a relative position of accent and decade of logarithmic fundamental frequency.

In the experiment for the decision uttered by announcers, the correct score for decision of accent type is 97%. So we can say that the method is effective. For the decision uttered by foreign students, the score is 69% which includes wrong accent utterances. So we can say that the method is useful for foreign students to learn Japanese speech.

key words Language Education, Japanese Speech, Accent.

1 序論

語学教育において、音声教育の重要性は認識されているが、音声教材は高価であり、充分であるとは言えない。特に外国の人々が日本語を学習する場合、日本語教師の母語が日本語でない場合があるなど、日本語の音声教育が十分に行われていないのが実情である [1]。また学習する時間や場所についても制限が加えられてしまう。

ところで、ここ最近のマルチメディア技術の進歩、パソコンの性能向上と価格低下によりだれにでもマルチメディアコンテンツを利用できる環境が整いつつある。そこで、このような環境を利用した日本語音声の学習ができるシステムが必要である。だれにでも利用できる日本語音声教育システムとするためには、Windows、Macintosh、Unix といった異なる OS 間でも動作可能なマルチプラットフォーム型のシステムが必要である。そこで Java 言語を用いることで、異なる OS 間はもちろん、ネットワーク端末など Java をサポートしている機器であれば同じプログラムを動作することができる [2][3]。

本研究では、日本語音声のアクセント教育のためにアクセントに関する特徴パラメータの抽出を行い、アクセント型の体系をもとにアクセント型を自動判定するためのアルゴリズムを提案し、留学生音声を用いた評価実験により有効性の検討を行う。

2 単語アクセント型の自動判定法

2.1 日本語アクセントの体系

日本語アクセントは「高」「低」アクセントであり、英語のアクセントは「強」「弱」アクセントである。

日本語のアクセントは、より高い段とより低い段から成り立っており、一つの拍は、「高(High)」か「低(Low)」のいずれかである。「高」でも「低」でもないというものは存在しない。アクセントとは、「赤」(HL)、「垢」(LH) などのように、一つ一つの語句についてきまっている高低の配置のことをいう [4]。

日本語のアクセント型の種類を、表 1 に示す。さらに、図 1 にアクセント型の 2 値 (High, Low) 表現を示す。ここで N は単語中のモーラ数であり、 M は、 $1 < M \leq N$ である。

表 1: 日本語のアクセント型の種類

モーラ数 N	起伏式			平板式	型の数 $N+1$
	1型	2型	3型	0型	
1	H(L) 火(が)			L(H) 日(が)	2
2	HL(L) 空(が)	LH(L) 花(が)		LH(H) 鼻(が)	3
3	HLL(L) 命(が)	LHL(L) 垣根(が)	LHH(L) 男(が)	LHH(H) 桜(が)	4

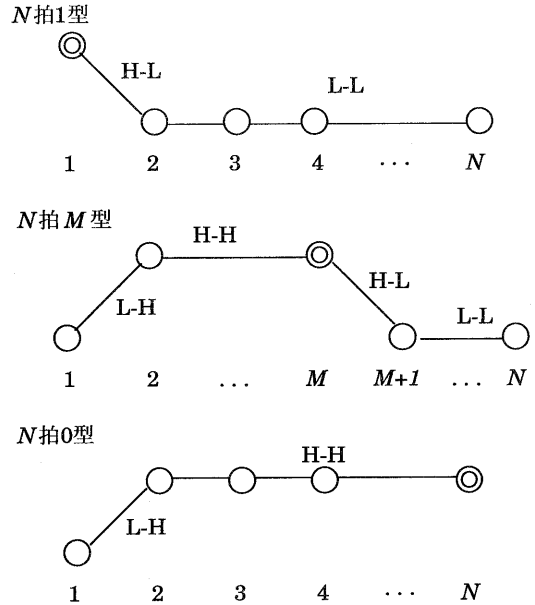


図 1: 各アクセント型の 2 値表現

各アクセント型の実際の基本周波数 (F0) [5] パターンの例を図 2 に示す。特徴としては、1 型は他の型よりもスタートの値が大きい。また逆に 0 型は他の型よりもスタートの値が小さいことが多い。さらに、0 型と他の型の場合は 0 型がなだらかな変化であるのに対し、他の型は変化の割合が大きい。

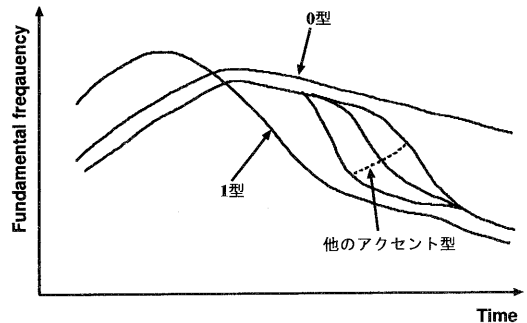


図 2: 各アクセント型の基本周波数パターン

アクセントの例として「赤」と「垢」の基本周波数パターンをそれぞれ図 3、4 に示す。ここで + 印が抽出された基本周波数で、実線が平滑化 F0 パターンである。発音スピードは常に一定ではないので、アクセント位置を単語音声の全体の時間を 1 とした場合の相対位置 [6][7] で示す。「赤」(HL) は第 1 音節に、「垢」(LH) は第 2 音節にアクセントがくる単語であるが、アクセント位置に対応した 0.3 と 0.65 の位置にアクセ

セントがきていることがわかる。

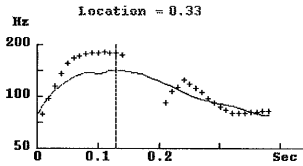


図 3: アクセントの例一赤 (HL)

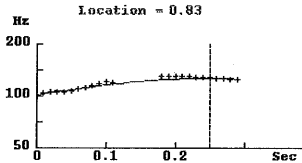


図 4: アクセントの例一垢 (LH)

単語が N 拍である場合、東京方言ではアクセント型は $N+1$ 種類存在する。例えば「あき」という単語の場合「秋(が)HL(L)」(1型)、「飽き(が)LH(L)」

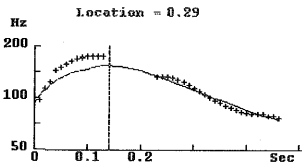


図 5: 1型の実例の例一秋(が)(1型)(HLL)

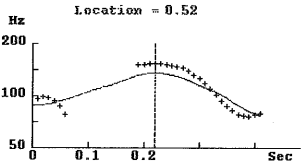


図 6: 2型の実例の例一飽き(が)(2型)(LHL)

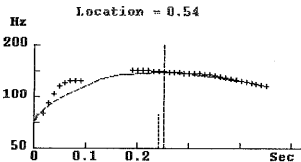


図 7: 0型の実例の例一空き(が)(0型)(LHH)

(2型)、「空き(が)LH(H)」(0型)の3種類のアクセント型が存在する。ここで「飽き」と「空き」は同じ(LH)型である。これは表1に示されているように、日本語のアクセントでは最後の拍が高く終わるアクセントの語には、次に助詞などの他の語が来たとき、その語を低くつかせるものと、高くつかせるものと2種類あるためである。このため、助詞などがつかず単独の場合はこれらのアクセント型は区別できない。

このように助詞が付いた場合でも、相対アクセント位置だけでは区別できないので、基本周波数の傾きの特徴を用いる必要がある。

2.2 アクセント型自動判定アルゴリズム

2.2.1 アクセント型自動判定の手順

アクセント型の体系や各アクセント型の特徴を踏まえ、アクセント型の判定をおこなう。アクセント型判定アルゴリズムの流れ図を図8に示す。すなわち手順は以下の4段階で行われる。

1. 基本周波数の抽出
2. 抽出した基本周波数の内挿
3. 特徴パラメータ (T_a, A_{F0}) の抽出
4. ラベルデータを参照したアクセント型の判定

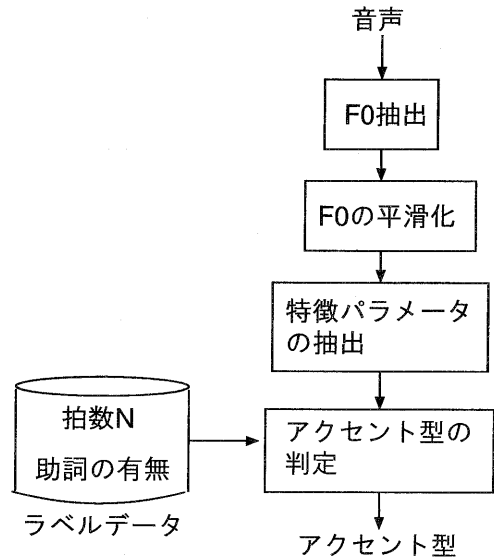


図 8: アクセント型判定アルゴリズム

2.2.2 基本周波数の抽出

基本周波数の抽出方法は、時間領域を用いる方法と周波数領域を用いる方法に大別できる。時間領域を用いる方法は一般的に処理速度が速く、周波数領域を用いる方法はスペクトルなどの情報を利用した他のシステムに取り込みやすい等の特徴があげられる。

本研究では、処理時間的に他より高速抽出が可能な改良自己相関基本周波数抽出法 (IAUTOC:Improved AUTOC)[8]を用いた。

改良自己相関法で用いる自己相関関数は以下の式のように表現される。

$$R_n(k) = \sum_{m=0}^{N-1-k} x(m)x(m+k) \quad (0 \leq k < k_e)$$

$$R_n(k) = \sum_{m=0}^{N'} x(m)x(m+k) \quad (k_e \leq k \leq 2k_e)$$

前半が一般的な自己相関関数、後半が変形自己相関関数である。本研究では、 $N = 2k_e, N' = k_e = 180$ ($F_s = 12\text{kHz}$) であり、連続的に結合している。

図9は、あるフレームでの改良自己相関関数の様子を示している。基本周波数の取り得る範囲 (成人男性の場合:50-200Hz, 成人女性の場合:200-400Hz) から、基本周波数の範囲を求め、1フレームごとに範囲内で相関関数のピーク位置を抽出し、その位置がそのフレームでの基本周期である。基本周波数は、基本周期の逆数を取れば求めることができる。

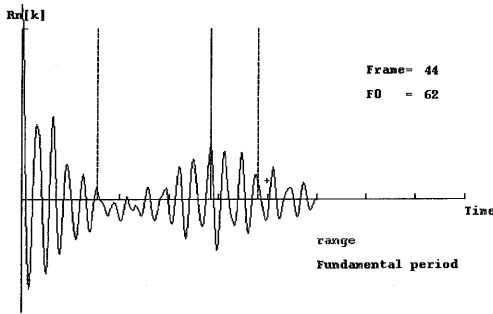


図9: 改良自己相関関数と基本周波数の抽出の例

2.2.3 F0の内挿平滑化

無声音部や無音部では、声帯の振動が停止するために基本周波数が存在せず、単語の途中にもかかわらず基本周波数がとぎれてしまう可能性がある。しかし、単語のアクセントは有声音のみでなく、無声音部にも現れると考えられるため、無声音部に仮の基本周波数

を補間して連続した基本周波数パターンを内挿平滑化する[6]。

補間の方法は、基本周波数データに関して、補間幅が n_i であるとき、 $-n_i$ から n_i の範囲で回帰直線 $y = ax + b$ を引く。この直線式の切片 b の値を新たな基本周波数データとする。図10では \circ は補間前の基本周波数データを示している。

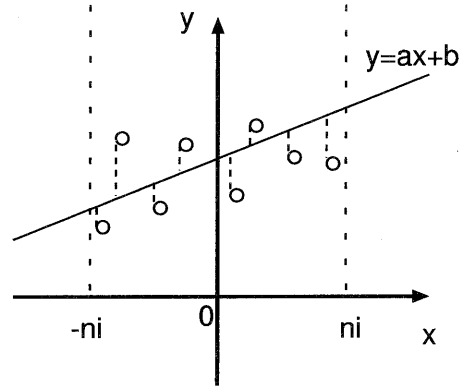


図10: 回帰直線の例

この回帰直線式の誤差エネルギー E は

$$\begin{aligned} E &= \sum_{i=1}^N \{y_i - (ax_i + b)\}^2 \\ &= \sum_{i=1}^N \{y_i^2 - 2y_i(ax_i + b) + ax_i^2 + 2abx_i + b^2\} \\ &= \sum_{i=1}^N y_i^2 - 2a \sum_{i=1}^N x_i y_i - 2b \sum_{i=1}^N y_i + a^2 \sum_{i=1}^N x_i^2 \\ &\quad + 2ab \sum_{i=1}^N x_i + b^2 N \end{aligned}$$

と表される。誤差エネルギーを最小にするためには、この式を a, b それぞれで偏微分したものが0になるように回帰直線式を設定すればよいので、

$$\begin{aligned} \frac{\partial E}{\partial a} &= -2 \sum_{i=1}^N x_i y_i + 2a \sum_{i=1}^N x_i^2 + 2b \sum_{i=1}^N x_i = 0 \\ \frac{\partial E}{\partial b} &= -2 \sum_{i=1}^N y_i + 2a \sum_{i=1}^N x_i + 2bN = 0 \end{aligned}$$

上の2つの連立方程式を解いて、 a, b の値を次式で求める。

$$\begin{aligned} a &= \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i \sum_{i=1}^N x_i} \\ b &= \frac{\sum_{i=1}^N y_i - a \sum_{i=1}^N x_i}{N} \end{aligned}$$

この時の b の値が、補間後の基本周波数データとなる。

1回目の補間は、有声区間の基本周波数データを平滑化するために、前後3フレーム ($n_i=3$)、計7フレーム ($2*3+1$) で回帰直線を引く。2回目の補間は、無

声部分を補間するために、前後 10 フレーム ($n_i=10$)、計 21 フレーム ($2*10+1$) で行う。補間幅を大きくすることで、元々データのない無声音部にも基本周波数を存在させることができる。

有声音部では 1 回目のデータを、無声音部では 2 回目のデータを採用し、さらに 5 点メディアンフィルタを用いて平滑化することで、補間幅の違う有声区間と無声区間を滑らかにつなげ、アクセント検出に使用できる連続的な基本周波数データを作成する。

2.2.4 特徴パラメータの抽出

アクセント型判定のための特徴パラメータの抽出アルゴリズムを図 11 に示す。

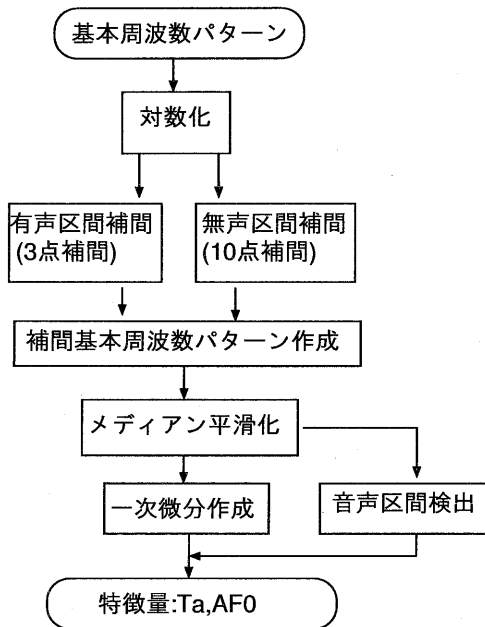


図 11: 特徴パラメータ抽出アルゴリズム

アクセント位置は、基本周波数データが最大値をとる位置にあると考えられるので、図 12 に示すように補間後の基本周波数データを用いてアクセント位置 ($F0_{peak}$) を検出する。発音速度は常に一定ではないので、検出したアクセント位置は単語の始端から終端までを 1 としたときの相対位置に直す。すなわち相対アクセント位置 Ta は、以下の式で表される。

$$Ta = \frac{T_{peak}}{(T_{end} - T_{start})}$$

また、0 型と M 型を判別するために図 12 に示すような基本周波数パターンの概略的な傾きの特徴を調べる。基本周波数の最大値 ($F0_{peak}$) と終端の値 ($F0_{end}$)

の基本周波数パターンの傾きを特徴量として用いる。

ここで傾きは、

$$A_{F0} = \frac{\log_2 F0_{end} - \log_2 F0_{peak}}{(T_{end} - T_{peak})/100}$$

単位 (octave/sec)

とする。

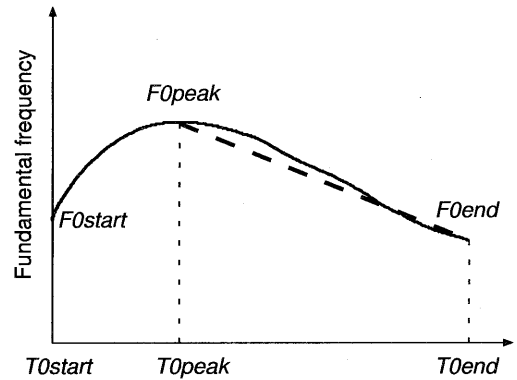


図 12: $F0_{peak}-F0_{end}$ の傾き (A_{F0})

求めた M 型と 0 型の基本周波数パターンの傾き A_{F0} の値を用いて度数分布を求め 0 型と M 型のしきい値を求める。

2.2.5 アクセント型の判定

相対アクセント位置としきい値を用い、アクセント型の判定基準を以下のように考えた。ここで、 N は単語の拍数の値、 Ta は相対アクセント位置、 θ_N は傾きのしきい値である。 ($0 < Ta < 1, 1 < M < N$)

表 2: アクセント型判定条件

型	判定基準
N 拍 1 型	$0 < Ta < \frac{1}{N}$
N 拍 M 型	$\frac{1}{N} < Ta < \frac{M}{N}$ かつ $A_{F0} < \theta_N$
N 拍 0 型	$\frac{1}{N} < Ta$ かつ $A_{F0} > \theta_N$
N 拍 N 型 助詞あり	$\frac{1}{N} < Ta$ かつ $A_{F0} < \theta_N$
N 拍 N 型 助詞なし	$\frac{1}{N} < Ta$ かつ $A_{F0} > \theta_N$

アクセント学習の場合、単語名が既知であるので、ラベルデータとして拍数、助詞の有無を与え判定を行う。単語の後ろに助詞がつかない場合は 0 型と N 型との区別は行わない。すなわち、アクセント位置の正誤についてのみ判定する。

3 判定実験

3.1 音声資料

アクセント型の判定実験に用いた 20 組 42 単語のアクセント教材を表 3 に示す。

表 3: 日本語音声教育アクセント教材リスト

読み	1 型	2 型	3 型	0 型
あか	赤	垢		
あき	秋が	飽きが		空きが
あく	悪			空く
あさ	麻	朝		
あたり	辺り			当たり
あつい		熱い		厚い
あめ	雨			飴
いし	意志	石		
いち	位置	一 (1)		
いらい	依頼			以来
うえる		飢える		植える
うむ	有無			産む
おく	奥			置く
かえる	返る			蛙
かき	牡蛎			柿
かち	価値	勝ち		
くらす	クラス			暮らす
げんきん			現金	厳禁
はし	箸が	橋が		端が
はな		花が		鼻が

これらのアクセント教材を、アナウンサー 2 名 (男性、女性)、外国人の留学生 6 名 (男性) に発音してもらい、アクセント型の判定実験をおこなった。発音をするときには、「赤」「垢」、「秋が」「飽きが」「空きが」というように各アクセント組ごとに 1 型から 0 型へと順番に発音してもらった。サンプリング周波数は 12kHz であり、音声区間の切り出しは手動で行った。またフレーム周期は 10ms である。

留学生に対しては各単語について図 13 に示すような高低の配置を情報として与えて発音してもらった。情報を与えることで全くランダムなアクセントになることがなくなるので、発音しにくいアクセント型がわかると考えられるからである。

判定に用いる基本周波数パターンの傾き値 A_{F0} のしきい値 θ_N は、アナウンサーに発音してもらった音声教師音声として図 14 の度数分布を求め、M 型と 0 型のしきい値は $\theta_N = -1.68$ とした。

3.1.1 アナウンサー音声

アクセント型判定の様子を図 15 に示す。

また、求めたしきい値を用いたアナウンサー音声の判定実験結果を表 4 に示す。

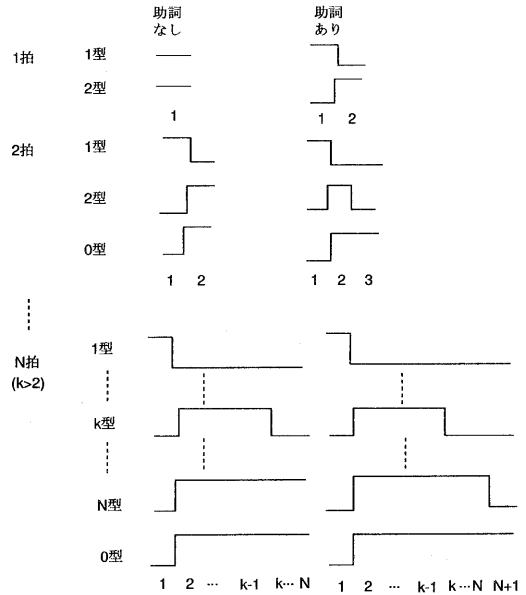


図 13: アクセント型の高低の配置

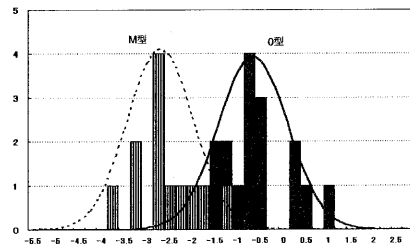


図 14: A_{F0} の度数分布

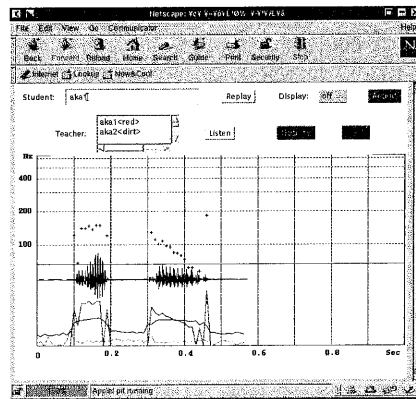


図 15: アクセント型判定プログラムの動作例

表 4: アクセント型の判定結果 (アナウンサー - 男女 2人)

		判定した型				資料数
		1型	2型	3型	0型	
実 際 の 型	1型	97%	0	3%	0	32
	2型	0	95%	0	5%	20
	3型	0	0	100%	0	2
	0型	0	3%	0	97%	30

アナウンサーの正しい発音に対して正しいアクセント型を97%判定できていることから、このアルゴリズムは有効であるといえる。

教師音声で正しく判定できなかったものを図16に示す。これは、倍ピッチによる基本周波数の抽出に問題があったために正しく判定できなかったと思われる。

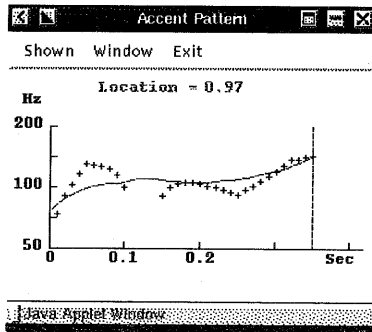


図 16: アナウンサー音声の正しく判定できなかった例-一辺り (1型)

3.1.2 留学生音声

留学生の音声についてアクセント型の判定をおこなった結果を表5に示す。正しいアクセント型と判

表 5: アクセント型の判定結果 (留学生 - 男性 6人)

		判定した型				資料数
		1型	2型	3型	0型	
実 際 の 型	1型	94%	2%	2%	2%	96
	2型	5%	71%	4%	20%	60
	3型	0	17%	33%	50%	6
	0型	0	11%	9%	80%	90

定された割合は全体で69%であった。最初にアクセントがくる1型はわかりやすいため、ほぼ正しく発音されていた。間違えやすい発音としては、2,3型を0型に間違える場合が多かった。なお、留学生の音声に対しても正しい発音に対しては正しいアクセント型に

判定し、間違った発音に対しては間違ったアクセント型に判定した。

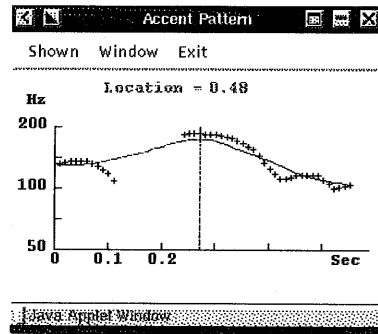


図 17: 留学生の発音の正しい例-熱い (LHL)

図17の「熱い」(2型)と18の「厚い」(0型)という単語を例にとって説明する。

図17の正しい発音例の場合、相対アクセント位置 $T_a = 0.48$, 傾き $A_{F0} = -2.61$ である。これを判定アルゴリズムに適用すると、 $\frac{1}{3} < T_a < \frac{2}{3}$, かつ $A_{F0} < \theta_N$ となり、アクセント型の判定も正しく2型であると判定される。

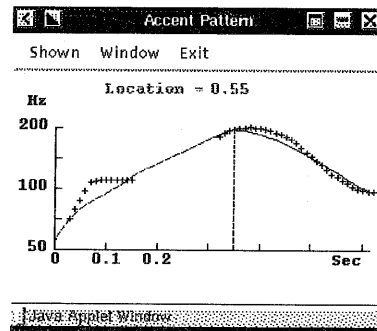


図 18: 留学生の発音の間違った例-厚い (LHH)

図18の間違った発音例の場合、相対アクセント位置 $T_a = 0.55$, 傾き $A_{F0} = -4.82$ である。これを判定アルゴリズムに適用すると、 $\frac{1}{3} < T_a < \frac{2}{3}$, かつ $A_{F0} < \theta_N$ となり、2型であると判定されるので誤った発音であることがわかる。

よって、このアルゴリズムを用いてアクセント型の判定をおこなうことにより、正しいアクセント位置を知ることができるので、日本語アクセント学習に役立てることができるといえる。

4 結論

日本語アクセントの特徴を踏まえて、基本周波数の抽出とアクセント判定用特徴量の抽出をおこない、アクセント判定アルゴリズムを提案し、実際の留学生の音声を用いて判定をおこなった。

アナウンサーの発音の正しい音声について、アクセント型が正しく判定された割合は97%であった。よって本判定アルゴリズムは有効であると言える。

また留学生の音声についてアクセント型の判定をおこなった結果、正しい発音に対しては正しいアクセント型に、間違った発音に対しては異なるアクセント型に判定された。アクセント型全体の正答率は69%であった。よって日本語アクセント学習に役立てることができるといえる。

しかし、教師音声でも正しく判定できないものがあつたため以下について今後検討する必要がある。

- 基本周波数の抽出アルゴリズムについて改良をおこない抽出精度をあげる。
- 基本周波数の内挿法についても改良を加える。
- 資料数を増やし、より精度の高いしきい値を求める。

謝辞 本研究の一部は、文部省科学研究費補助金・基盤研究(B)(09558022)によつた。

参考文献

- [1] Goh Kawai and Keikichi Hirose : "A bilingual speech recognizer for detecting phone-level pronunciation errors in nonnative speech", 日本音響学会 秋季研究発表会 講演論文集 I, 1-2-24, pp.221-222 (Sep. 1998)
- [2] Sun Microsystems, Inc. : "Javaプログラミング講座", アスキー出版局, 東京 (Oct. 1996)
- [3] 大谷 卓史, 武藤 健志 : "はじめてのJava", 技術評論社, 東京 (Mar. 1996)
- [4] 金田一 春彦 : "国語アクセントの史的研究 原理と方法", 塙書房, 東京 (Mar. 1974)
- [5] 三輪 譲二 : "パソコン音声処理" 昭晃堂, 東京 (July 1991)
- [6] 加藤, 三輪 : "移動平均と帯域制限を用いたケプストラム型基本周波数抽出とその応用", 電子情報通信学会技術研報告, SP94-95 (Feb. 1995)
- [7] 三輪, 熊谷, 吉田 : "オンデマンド・ネットワーク型日本語アクセント音声教育システム", 日本音響学会 春季研究発表会 講演論文集 I, 2-P-26, pp.345-346 (Sep. 1998)
- [8] X.Gao, Y.Kikuchi and H.Kasuya : "An Improved Algorithm of Autocorrelation Pitch Detection", Trans.IECE of Japan, Vol.E67, No.5, (May. 1984)