音楽情報科学 24−16
(1998. 2. 20)

音声教育のための中国語有気無気音の識別

田 嘉鵬、三輪 譲二

岩手大学 工学部

〒020-0066 岩手県盛岡市上田4 - 3 - 5

tian@cis.iwate-u.ac.jp, miwa@cis.iwate-u.ac.jp

あらまし 学習者が独習できる中国語音声教育システムの構築を目指して、中国語音声教育における重要な課題の一つである有気無気音の自動識別に、動的低域パワー (DLP,Dynamic Low-passed Power) を音響特徴量とする識別方法を提案する。オープン実験において学習話者以外の中国語母国語話者５人の音声データを識別した結果、有気音は 91.8%、無気音は 99.6% の識別率が得られた。また、７人の日本人学習者の音声データに対して、知覚とシステム識別により得られた正解率が、全体的に同一の傾向を示した。さらに、DLP の２番目のピークの値から得られた事後確率のスコアリングによりを行い、日本人学習者発音の良さを評価でき、実際にシステムを構築する可能性が得られた。

キーワード 中国語音声、有気無気音、自動識別、音声分析、CAI システム、音声教育

A Discrimination of Chinese Aspirated/Unaspirated Sounds for Speech Learning System

Jiapeng Tian and Jouji Miwa

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

4-3-5 Ueda Morioka-shi Iwate-ken, 020-0066 Japan

tian@cis.iwate-u.ac.jp, miwa@cis.iwate-u.ac.jp

Abstract We proposed a method of automatic discrimination of aspirated and unaspirated sounds for a computer assisted instruction system of Chinese speech learning, and a likelihood metric of native speech based on *a posteriori* probability. In the method, a dynamic low-passed power (DLP) is used as the acoustic feature. In an opened discriminant experiment for the consonants uttered five Chinese native speakers, a discriminant score of aspirated consonant is 91.8% and a score of unaspirated is 99.6%. In a experiment for seven Japanese native speakers, the scores are comparable to the perception scores. From the results, we can say that the method is available for Chinese speech learning system.

key words Chinese Speech, Aspirated/Unaspirated, Recognition, Analysis, CAI, Speech Learning

1 はじめに

中国語は世界で英語の次になる話されている人口の多い言語だと言われている。中国は千九百八十年代から改革開放政策を取り入れてから、この二十年間にわたって速いスピードで経済が発展すると同時に、国際化もどんどん進んで来ている。経済、文化などの交流のため、世界各地で中国語を学習する人は年々増えていて、中国語の言語教育が重要になっている。

一方、言語教育は、文法教育と音声教育からなる。文法教育の役割は、学習者を言語の意味を理解させることである。その言語の発音を学習者に身に付けさせるのは、音声教育の働きである。中国語の音声体系は、世界で最も複雑な体系の一つだと言われる [1]。中国語において、音声教育は文法教育より遅れているのだけでなく、教師不足などの理由でまだ十分に行われていない。日本人学習者にとって、中国語の文法を理解するのはそれほど難しくはないが、日本語の発音の影響を受けているせいか、中国語の発音がなかなかうまくできない。従って、中国語教育において、音声教育は非常に深刻な課題である。

また、ここ数年マルチメディア技術の進歩と、インターネットの普及に伴って、パソコンを利用する人が増えつつあっている。パソコンとマルチメディア技術を活かして、簡単に音声分析と音声合成のためのユーザーインタフェースを作ることが可能 [2] になった。そして、コンピュータを利用する言語教育のための CAI システム (コンピュータ援助指導システム) が数多く開発 [3] されている。しかし、これまでの CAI システムの多くは、音声をコンピュータに直接入力して、それを評価した結果を学習者にフィードバックするようなものはあまり見られない。

さらに、中国語音声教育において、特に重要な課題が二つある。一つは中国語の四声であり、もう一つは中国語の有気／無気破裂子音である。これまでの研究は、四声の音響特徴量として基本周波数を用いて、高い認識率が得られたが、有気／無気破裂子音の音響的特徴はまだはっきりされていない。このため、本研究は、音声を入力とする音声教育システムを構築するため、中国語の有気／無気破裂子音の自動識別を取り上げる。

音声認識において、子音は母音と違って、スペクトルの時間的変化によって特徴付けられている。日本語の場合、破裂子音を識別するために従来行われた方法は、次の 2 種類に大別することができる。一つは破裂時点近辺のスペクトル構造によって識別する方法 [4] であり、もう一つは調音結合によるスペクトルの遷移によって識別する方法 [5] である。また、特徴量として、ケプストラム係数 [6][7] と、臨界帯域スペクトル [4][8] はよく利用されている。ほかには、スペクトルのローカル・ピークと傾きを使って日本語の破裂子音の認識 [9] と、破裂部のケプストラムと遷移部のケプストラムを使って中国語の無声無気破裂音と零声母の認識 [10] との研究も報告されている。これらの研究はほとんどネイティブ話者を対象としたが、言語教育のための非ネイティブ話者を対象とする子音認識の研究は、あまり見られない。

そこで、本研究は、動的低域パワー (Dynamic Low-passed Power, 以下 DLP と呼ぶ) を音響特徴量として、子音の遷移部における DLP の特徴を利用して、中国語有気音 p[p‘] , t[t‘] と無気音 b[p] , d[t] を識別する方法を提案する。

2 中国語破裂子音の特徴

子音を発声するとき、調音位置と調音様式はそれぞれ異なり、また清音と濁音、有気と無気などの区別もあり、人間の発音能力を全部利用していると言われる。子音の音響的な特徴も母音よりずっと複雑である。すべての子音は幾つかの音響特徴の組み合わせで特徴付けられている。特徴量の組み合わせモデルはさまざまであり、あまり安定ではない。よって、子音の認識は一般に母音より困難である。

子音の中の破裂子音において、日本語や英語などのような言語には、破裂子音には有声 / 無声の対立があり、閉鎖が解かれるとき、声帯が振動するかどうかによって区別され、例えば [b] と [p] の表記となる。

これに対して、中国語の破裂子音には、有気 / 無気の対立があり、閉鎖が解かれるとき声帯の振動はなく、閉鎖が解かれた後、有気音 (Aspirated sound) は、いったん「息」が流れ出てから声帯が振動を始めるが、無気音 (Unaspirated sound) は、直ちに声帯が振動を始める。このため、例えば [p‘] と [p] の表記となる。

発声上では、有気音の場合、息が流れ出るので、子音の持続時間が無気音より長い。また、「息」が流れ出る間、つまり破裂してから声帯が振動を始めるまでの間に、息がだんだん弱くなるに従って、音声パワーが徐々に下がり、その後、声帯が少しずつ振動を始めるに従って、また徐々に上がるというパターンで変化する。これに対して、無気音の場合、破裂してから息が流れ出ないので、有気音のようにパワーが下がってから上がるという変化パターンが現れない。

破裂子音の破裂性は、パワーの急峻な立ち上がりで特徴付けられる。日本語でも中国語でも破裂時点を検出するには、パワーの時間変化がよく利用 [11] されている。又、有声音と無声音のセグメンテーションにもパワーの時間変化が有効な特徴量 [12] として使われている。本研究は、中国語有気無気音が遷移部における DLP の特徴の違いを利用して、中国語破裂子音の有気性を識別する。

3 識別方法

3.1 音響特徴量 DLP の抽出

音声波形に対して FFT で、フレームごとに短時間パワースペクトルを求める。それから、改良ケプストラム分析 [13] でパワースペクトルの包絡を抽出する。有気音と無気音のパワースペクトルの時間変動の例を、それぞれ図 1と図 2に示す。

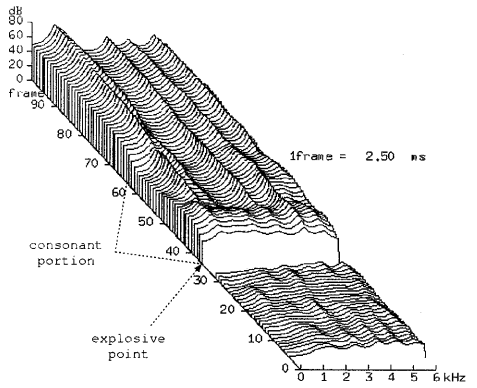

図 1: 有気音パワースペクトルの時間変動の例

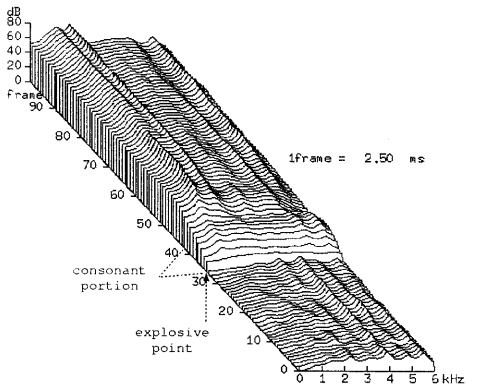

図 2: 無気音パワースペクトルの時間変動の例

図 1と図 2のパワースペクトルの時間変動を較べると、低周波数帯域における変化パターンは、有気音と無気音によって異なることが分かった。有気音の場合は、低域パワーが破裂時点 (35 番フレーム) で一回急峻に立ち上がり、その後少し下がってまた母音に入るところ (63 番フレーム) でもう一回急峻に立ち上がった。これに対して無気音の場合は、低域パワーが破裂時点 (34 番フレーム) で一回急峻に立ち上がり、数フレーム続けて徐々に上がってからすぐ母音に入り、二回目の急峻な立ち上がりが現れなかった。この違いを利用して、有気音と無気音を識別する。

パワースペクトルの包絡から、式 (1) で低周波数帯域の平均パワー $p[i]$ を求める。

$$p[i] = \frac{1}{n_1 - n_0 + 1} \sum_{n=n_0}^{n_1} G_i(\Omega_n), \qquad (1)$$

$$(\Omega_n = \tfrac{2\pi n}{N}, N = 512)$$

ここで、

n_0 : 周波数帯域の始点
n_1 : 周波数帯域の終点
i : フレーム番号
G_i : 対数パワースペクトル包絡

である。

式 (1) で、フレームごとに求めた低域パワーの例を、図 3と図 4に示す。

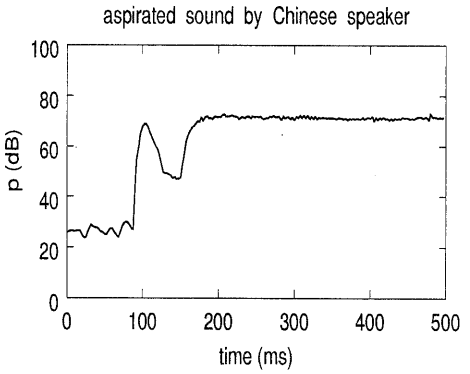

図 3: 有気音の低域パワーの例

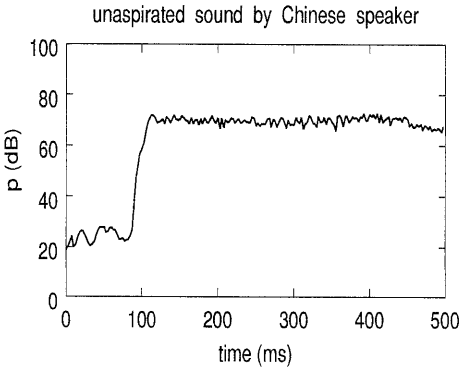

図 4: 無気音の低域パワーの例

さらに、式 (2) で低域パワーを平滑化してから一次差分を求め、これを動的低域パワー DLP とする。

$$dp[i] = \sum_{m=-M}^{M} (m \cdot p[i+m]) \qquad (2)$$

ここで、$2M+1$ は、フィルターの次数である。

式 (2) で求めた DLP の例を、図5と図6に示す。

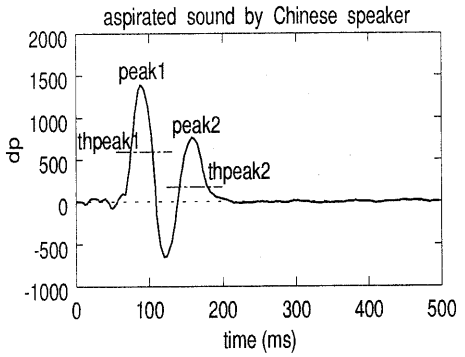

図 5: 有気音の動的低域パワーの例

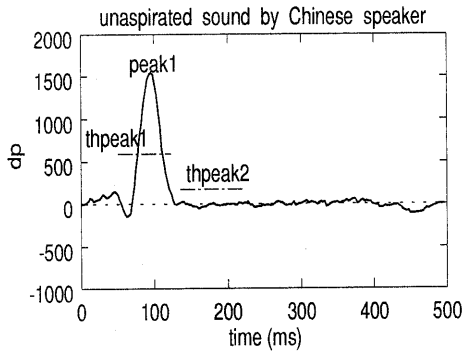

図 6: 無気音の動的低域パワーの例

図5と図6の中において、$peak1$ は、有気音と無気音の破裂部における DLP の有意な最初のピーク (以下、ピーク1と呼ぶ) の値であり、$peak2$ は、有気音の遷移部における DLP の2番目のピーク (以下、ピーク2と呼ぶ) の値である。

3.2 識別条件

遷移部における DLP の特徴、即ちピーク2の有無により有気音と無気音を識別する。ただし、破裂時点からピーク2を検出する範囲が長すぎると、母音区間を超えて、次の音節の DLP のピークが間違って検出される恐れがある。そこで、識別する時に、学習データから得られた有気音の平均持続時間 $range$ を用いる。また、無気音でも雑音の影響で、$thpeak2$ より大きい凸形状が出る可能性があるので、識別する時にピーク2の信頼度 (凸形状と両側の凹形状の間の差) のしきい値 $thconf$ も、識別条件の一つとして用いる。

これらのことを考慮に入れて、識別する時に

1. 最初に $thpeak1$ より大きい凸形状が現れたら、その時点を破裂時点とする。
2. 破裂時点から、有気音の最大平均持続時間の範囲 $range$ の中で、$thpeak2$ より大きい凸形状が現れ、しかもその凸形状の信頼度が $thconf$ より大きい場合、その子音を有気音とする。その他の場合は、無気音とする。

識別処理全体の流れを、図7に示す。

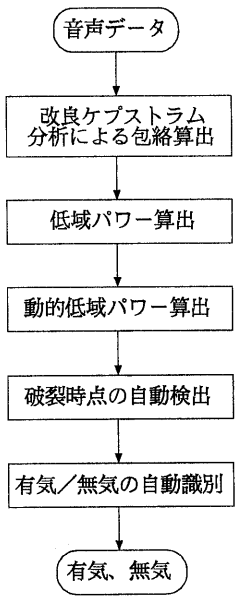

図 7: 有気無気音識別処理全体の流れ

4 識別実験と結果

4.1 音声データ

実験のために、まず音声データを収集した。音声資料は、語頭の有気音 p[p‘] , t[t‘] 又は無気音 b[p] , d[t] がすべての母音と四声と組み合わせ可能な有意味な2音節単語合計 200 個を用いた。収集では、普通の静かな部屋でマイク (ソニー製 ECM-959DT) を用いて、ディジタル・オーディオ・テープレコーダ (DAT) に

録音して、サンプリング周波数 48kHz、量子化ビット数 16 で A-D 変換している。録音したテープから各単語を切り出した時に、視察によって有音部分に前後それぞれ 100ms を加えて切り出してから、12kHz にダウンサンプリングした。発声者は、中国語標準語が話せる中国人話者 10 名と中国語を学んでいる日本人話者 7 名である。

音響特徴量を抽出するため、ダウンサンプリングされた 12kHz の音声データに対して、音声分析を行った。分析条件を、表 1に示す。

表 1: 分析条件

サンプリング条件	12kHz,16bits
分析区間長 N	30ms
フレーム周期 L	2.5ms
フーリエ変換	512 点
フィルタ次数 (2M+1)	17
周波数帯域	0Hz ～ 515Hz

4.2 識別結果

表 1の分析条件で、学習用のデータとする中国人話者 5 名の音声データを分析して、特徴量を抽出した。$peak1 - peak2$ の散布図を図 8に示す。

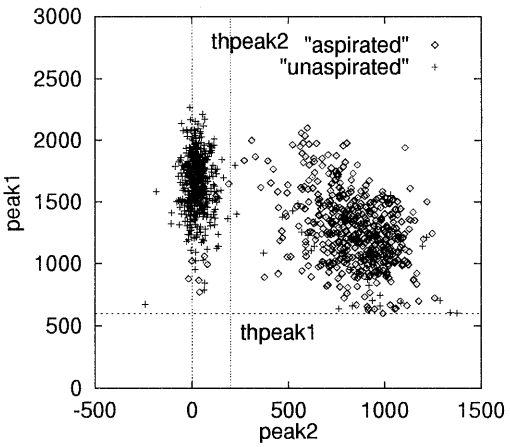

図 8: 中国人話者有気無気音の特徴量の散布図

これより、$peak1$ と $peak2$ の特徴を用いることにより、有気無気音が識別できることが分かる。

また、有気音の持続時間 $range$ とピーク 2 の信頼度 $peak2conf$ の分布を、それぞれ図 9と図 10に示す。ただし、図 10の中に、$peak2conf.left$ はピーク 2 の左側の信頼度であり、$peak2conf.right$ はピーク 2 の右側の信頼度である。

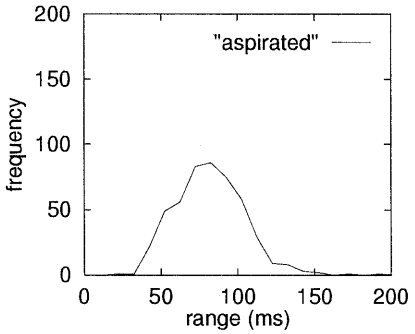

図 9: 中国人話者有気音持続時間の分布

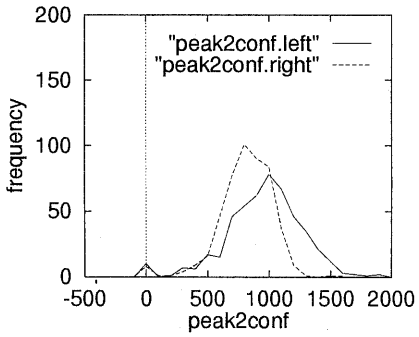

図 10: 中国人話者ピーク 2 信頼度の分布

これらの分布により決定した識別しきい値を、表 2に示す。

表 2: 識別しきい値

しきい値	意　味	値
$thpeak1$	$peak1$ の検出	600
$thpeak2$	$peak2$ の検出	200
$thconf$	$peak2$ の信頼度	250
$range$	破裂時点からの有気音の検出範囲	150ms

この識別条件により、学習用のデータ (中国語話者 5 人) と評価用のデータ (中国語話者他の 5 人) を識別した結果を、表 3に示す。

表 3: 中国語母国語話者のシステムによる識別結果

		有気音	無気音	総数
学習データ	有気音	96.6%	3.4%	476
	無気音	1.0%	99.0%	510
評価データ	有気音	91.8%	8.2%	488
	無気音	0.4%	99.6%	520

評価データに対しても、高い識別率が得られたので、提案法は有気無気音の識別に有効であることが分かった。

さらに、同一の分析条件と識別条件で、日本人学習者のデータを識別した。日本人学習者のデータの中に、正しい発音と間違った発音と曖昧な発音が混ざっていたので、コンピュータによる識別率だけで評価しにくい。そこで、中国語母国語話者による知覚実験も行った。日本人学習者が中国語の有気無気音を正しく発音できる能力を、知覚により判断して、J1 から J7 まで平均知覚率のスコア昇順で番号を振った。知覚による識別の結果は表 4に、システムによる識別の結果は表 5に示す。

表 4: 日本人学習者の知覚による識別結果

話者	単語	正解率 (%)	知覚		数
			有気	無気	
J1	有気	72.4	71	27	98
	無気	98.9	1	94	95
J2	有気	95.7	88	4	92
	無気	67.7	32	67	99
J3	有気	88.7	86	11	97
	無気	90.4	10	94	104
J4	有気	88.7	86	11	97
	無気	92.4	8	97	105
J5	有気	83.2	79	16	95
	無気	98.1	2	102	104
J6	有気	91.7	88	8	96
	無気	100.0	0	105	105
J7	有気	96.9	95	3	98
	無気	100.0	0	104	104
平均	有気	88.1	593	80	673
	無気	92.6	53	663	716

表 5: 日本人学習者のシステムによる識別結果

話者	単語	正解率 (%)	識別		数
			有気	無気	
J1	有気	34.7	34	64	98
	無気	33.7	63	32	95
J2	有気	91.3	84	8	92
	無気	84.8	15	84	99
J3	有気	81.4	79	18	97
	無気	93.3	7	97	104
J4	有気	67.0	65	32	97
	無気	95.2	5	100	105
J5	有気	64.2	61	34	95
	無気	91.3	9	95	104
J6	有気	82.3	79	17	96
	無気	100.0	0	105	105
J7	有気	88.8	87	11	98
	無気	100.0	0	104	104
平均	有気	72.7	489	184	673
	無気	86.2	99	617	716

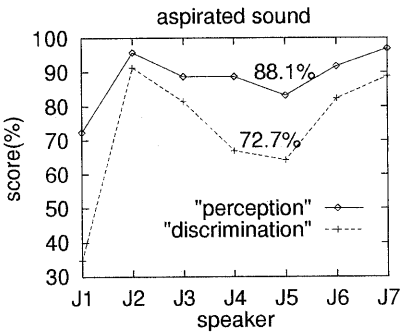

図 11: 日本人学習者有気音の識別結果

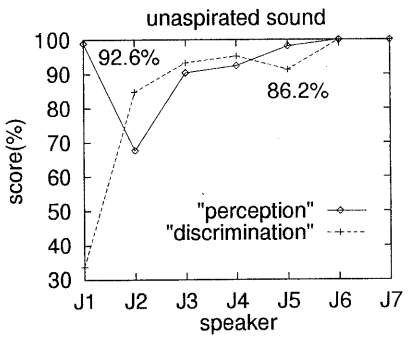

図 12: 日本人学習者無気音の識別結果

表 4と表 5に示す結果を図で表すと、図 11と図 12のようになる。

図において、破線はシステムの識別結果であり、実線は知覚の識別結果である。知覚により得られた発声者の正解率とシステムによる識別を行い得られた正解率が、全体的に同一の傾向を示すことにより、識別方法の有効性が分かった。

ところで、日本人学習者の中で、中国語有気無気音の発音方法をまだ十分に身に付けていない学習者にとって、曖昧な発音が出る可能性が非常に高いと思われる。そのため、知覚による判断の誤りが十分にあると思われ、それは発声者 J1 のように知覚とシステム識別の結果がかなり違うことの原因になっていると考えられ、今後の課題である。

4.3 学習者発声のスコアリング

実際に音声教育システムを構築することを考えたとき、学習者の発音を分析して、ただそれは有気音か無気音かを教えるだけでなく、学習者の発音の良さを評価できれば、発音指導の効率がもっと高まると考え

られる。そこで、中国語母国語話者のデータから得られた特徴量の事後確率を利用し、学習者の発音に対してスコアリングを行った。

まず最初に、中国語母国語話者の音声データから、DLP の $peak2$ の平均と標準偏差を求める。それから、$peak2$ の分布が正規分布と仮定し、有気音と無気音の条件付き確率を求める。ここで、DLP の２番目のピークにおいて、日本人学習者音声データの分散は、中国人話者より大きいので、中国人話者のデータから得られた分散を用いてスコアリングを行うと、スコアの分布は極端になり、学習者発声の良さをスコアで表現しにくい。そこで、1.5 倍した標準偏差値を用いてスコアリング関数を作った。

表 6: スコアリング関数計算値

変数	説　　明	実測値	計算値
μ_1	有気音 $peak2$ の平均値	788	788
σ_1	有気音 $peak2$ の標準偏差	222	333
μ_2	無気音 $peak2$ の平均値	70	70
σ_2	無気音 $peak2$ の標準偏差	200	300

表6の中に示す計算値を用いて、有気音と無気音のピーク２の条件付き確率を求め、図13に示す。

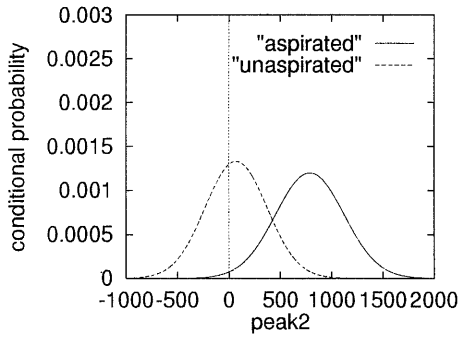

図 13: 条件付き確率

計算値の統計値を用いて条件付き確率を求めた後、さらに、ベイズの定理に基づいて、先験確率が同一と仮定し、式 (3) と (4) で有気音と無気音の $peak2$ の事後確率を求め、これを 100 倍してスコアリング関数値 s_1, s_2 として用いる。中国人話者のデータから求めた $peak2$ の事後確率とスコア値を、図 14に示す。

$$s_1 = \frac{100 \times p(x|\omega_1)}{p(x|\omega_1) + p(x|\omega_2)} \quad (3)$$

$$s_2 = \frac{100 \times p(x|\omega_2)}{p(x|\omega_1) + p(x|\omega_2)} \quad (4)$$

ここで、

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\{-\frac{(x-\mu_1)^2}{2(k\sigma_1)^2}\}$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\{-\frac{(x-\mu_2)^2}{2(k\sigma_2)^2}\}$$

s_1 : 有気音スコア値　　s_2 : 無気音スコア値
ω_1 : 有気音　　ω_2 : 無気音
x : $peak2$ の値　　k : 標準偏差調整倍数

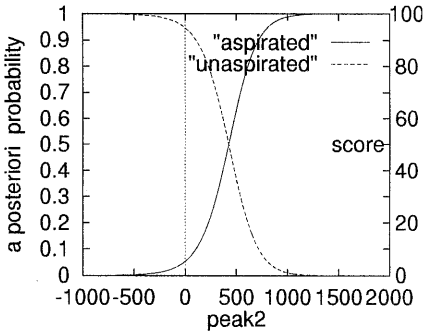

図 14: 事後確率とスコア値

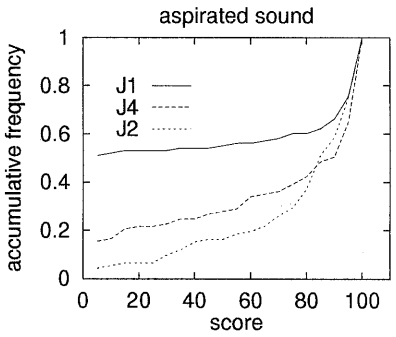

図 15: 有気音発声のスコア値の正規化累積分布の例

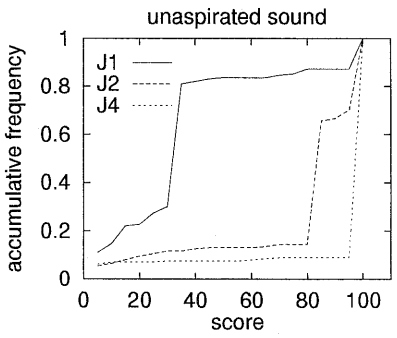

図 16: 無気音発声のスコア値の正規化累積分布の例

日本人学習者の発声に対して、スコアリング関数を用い、DLP の $peak2$ の値によりスコアを求めた。

図15と図16は、学習者J1,J2,J3の発声のスコア値の正規化累積分布である。図15の有気音のスコア値の累積分布は、単調に増加しているから、スコア値はほぼ均等に分布していると判断でき、スコア値は学習者発声の評価値として有効であることが分かる。しかし、無気音に対してまだ検討する必要がある。

これより、このスコアリング法を実際のCAIシステムに利用した場合、学習者が同じ発音を繰り返して発声し、自分にフィードバックしてきたスコアにより自分の発声の良さを判断して、発音方法を調整することができるので、発声訓練の効果が高まることが予想される。

最後に、スコアリングを利用する有気無気音音声教育CAIシステムの構成図は、図17となる。

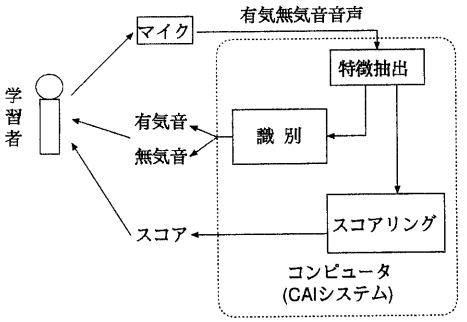

図 17: 音声教育CAIシステムの構成図

5 結論

本研究では、中国語音声教育のための有気音p[p‘], t[t‘]と無気音b[p], d[t]の自動識別に、動的低域パワー(DLP,Dynamic Low-passed Power)を音響特徴量とする識別方法を提案した。この識別方法用いて、中国語母国語話者のオープンデータに対して、有気音は91.8%、無気音は99.6%の識別率が得られた。また、日本人学習者のデータに対して、知覚評価の結果とシステム識別の結果が同一の傾向が得られた。これらのことより、本提案法は中国語有気無気音の識別に有効であることが分かった。

また、DLPの2番目のピーク値から得られた事後確率を利用して、日本人学習者の発声に対して、スコアリングを行い、発声の良さを評価することができた。よって、学習者が独習できるような中国語有気無気音識別システムを構築する可能性が得られた。

今後、他の有気無気音(g[k], k[k‘], z[ts], c[ts‘]等)への拡張を検討する必要がある。

謝辞

本研究の音声データの収集につき、協力して頂いた岩手大学中国人留学生のみなさんと盛岡市中国語クラスの学習者のみなさんに、深く感謝致します。

参考文献

[1] 王理嘉：”語音学教程”， 北京大学出版社, 北京 (1992).

[2] Jouji Miwa：”Interactive Visualization and Auralization of Speech Production Using Variable Vocal and Nasal Area Function”, ASVA97, pp.271-278 (1997).

[3] 三輪譲二、熊谷有香、田嘉鵬、今石元久：”オンデマンド・ネットワーク型日本語音声教育システムの構築”, 信学技報,SP97-17, pp.55-62 (1997).

[4] 北澤茂良、堂下修司：”破裂部スペクトルによる日本語無声破裂子音の識別”， 音響学会誌,40, pp.332-339 (1984).

[5] 井出和之、牧野正三、城戸健一：”時間一周波数パタンを用いた無声破裂子音の認識”， 音響学会誌,39, pp.321-329 (1983).

[6] 花沢利行、川端豪、鹿野清宏：”Hidden Markovモデルによる日本語有声破裂音の認識”， 音響学会誌,45, pp.776-785 (1989).

[7] Horacio Franco：”Recognition of intervocalic stops in continous speech using context-dependent HMMs”, J. Acoust. Soc. Jpn. (E),11,3, pp.131-143 (1990).

[8] 堂下修司、河原達也、水谷陽一、児島宏明、石川雅朗、北澤茂良：”2群対判別法による不特定話者日本語単音節中の子音の識別”， 音響学会誌,45, pp.827-836 (1989).

[9] 廖莉莉、牧野正三、城戸健一：”スペクトルの時間変化、ローカル・ピーク、傾斜を利用した破裂子音の検出と認識の検討”, 音響学会誌,45, pp.499-506 (1989).

[10] 易傑、鈴木久喜：”中国語の無声無気破裂音と零声母の認識”, 音響学会誌,44, pp.361-368 (1988).

[11] 胡志平、今井聖：”音素モデルと音節モデルを用いた中国語連続音声認識システムの作成”， 信学論,J75-D-2, 3, pp.459-468 (1992).

[12] Liyou Hu,Satoshi Imai,and Chieko Furuichi：”Phonemic segmentation for continuous Mandarin speech recognition”, J. Acoust. Soc. Jpn. (E),18,1, pp.1-8 (1997).

[13] 今井、阿部：”改良ケプストラム法によるスペクトル包絡の抽出”， 信学論,J62-A, 4, pp.217-223 (1979).