

# Recognition of Voiced Plosives Considering Coarticulation

|       |                                                                                             |
|-------|---------------------------------------------------------------------------------------------|
| メタデータ | 言語: jpn<br>出版者:<br>公開日: 2017-10-03<br>キーワード (Ja):<br>キーワード (En):<br>作成者:<br>メールアドレス:<br>所属: |
| URL   | <a href="http://hdl.handle.net/2297/342">http://hdl.handle.net/2297/342</a>                 |

# 調音結合を考慮した有声破裂音認識

三 好 義 昭

## Recognition of Voiced Plosives Considering Coarticulation

Yoshiaki MIYOSHI

あらまし 音声を用いたヒューマンインタフェースの高度化のためには、音素単位での認識に基づく連続音声認識を実用化レベルで実現する必要がある。その為には、個々の音素の音響的な特徴の正確な把握や、音素相互の的確な識別が重要な課題となる。しかし現段階においては、その音響的特徴が明らかになっている音素相互の識別でさえ、困難であるのが現状である。本論文では、音素単位での認識を行なう場合、中でも困難とされている有声破裂音(/b/, /d/, /g/)の認識手法について述べる。日本語では一般に子音のみを単独で発音することがなく、子音+母音型の音節として発音されるため、子音部から抽出した特徴量は調音結合等による後続母音部の影響を受けていると言える。本論では、子音部から抽出した特徴量から後続母音の影響を除去する手法について述べ、実際に有声破裂音認識に適用することにより、その有効性が示されている。

### 1. まえがき

近年のデジタル信号処理技術の飛躍的な進歩による高度情報化社会への移行に伴い、計算機を主体としたいわゆる”知的情報システム”と人間とのコミュニケーション手段の高度化にはますます社会的なニーズが高まってきている。この人間と知的情報システムとの対話媒体として、現在では主に文字や図形が広く用いられているが、もう一つの重要な対話媒体と言える「音声」は人間相互の間の情報伝達の手段と

して、人類の進化過程で発達してきたものであり、人間にとって最も自然で、かつ最も根源的な情報伝達手段であることから、音声を用いたヒューマンインタフェースの高度化は、人間と知的情報システムとの対話を最も自然な形で実現するものとして期待を集めている。この音声を用いたヒューマンインタフェースの高度化のためには、知的情報システムに具備すべき音声認識機能の高度化が必要不可欠である。

音声認識には認識対象別に区分すると、単語音声認識と連続音声認識に大きく分かれ、またそれぞれは話者を限定した特定話者型と誰の音声でも認識できる不特定話者型に区分される。もちろん、音声認識における究極の目標は不特定話者による連続音声認識であるが、これを実用化レベルで実現するためには、単語単位での認識に基づく手法<sup>(1)~(3)</sup>では限界があり、音素単位での認識に基づく手法<sup>(4)~(10)</sup>を確立することが必要と言える。

音素単位の認識を行なう為には、個々の音素の音響的な特徴の正確な把握や、音素相互の的確な識別が重要な課題となる。しかし現段階においては、その音響的特徴が明らかになっている音素相互の識別でさえ、困難であるのが現状である。

本論文では、音素単位での認識を行なう場合、中でも困難とされている有声破裂音(/b/, /d/, /g/)の認識手法について述べる。日本語では一般に子音のみを単独で発音することがなく、子音+母音型の音節として発音されるため、

子音部から抽出した特徴量は調音結合（声道の形が急には変化できないため、音響的性質が前後の音素の影響を受けて連続的に推移する現象）等による後続母音部の影響を受けていると言える。したがって、子音部から抽出した特徴量から後続母音の影響を除去することにより認識率の向上が期待される。

以下、2.において、有声破裂音の特徴について述べ、3.において、子音部に及ぼす後続母音の影響を除去する手法を示し、4.では、本手法を実際の自然有声破裂音認識に適用して、その有効性を示す。

## 2. 有声破裂音の特徴

破裂音は、舌や口唇で声道を遮断することによって呼気を一時的に止め、呼気がその後方にたまって圧力が十分高まったところで、これを急激に開放して衝撃波的な音源を生成することにより発声される。破裂音の生成は声帯の振動の有無とは独立して行なわれるが、破裂が起こると同時期、あるいはそれに先行して声帯音源による励振が行なわれる破裂音を有声破裂音と言い、日本語には /b/, /d/, /g/ の3種類がある。

成人男性が発声した有声破裂音 /b a/ の波形の例を図1に示す。

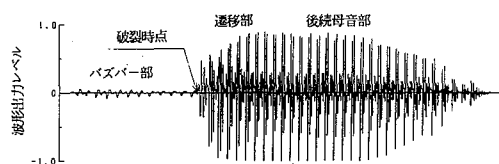


図1 有声破裂音波形の例

以上のような生成過程で発声された有声破裂音は、図1に示すように以下に述べる4つの部分に分けることができる。

(1) バズバー部：破裂以前の声帯による声道の励振部。比較的低い周波数の単一減衰波正弦の繰り返しと見なせる。

(2) 破裂時点：呼気が急激に開放された時点

で、破裂に伴う波形の急変が見られる。

(3) 遷移部：声道の形状がせばめを持った状態から母音の形状へと変化する部分。

(4) 後続母音部：子音部に比べて、スペクトルが定常でありパワーが大きい。

以上の特徴より、有声破裂音において知覚的に重要な部分は破裂点以降のホルマント（音声を発声する時の音響的共鳴系としての声道の共振、またはその周波数）の遷移部であり、この部分のホルマントの遷移パターンが有声破裂音を相互に識別するのに有効な特徴となる<sup>(11)</sup>。

## 3. 後続母音の影響を考慮した子音認識

日本語は一般に子音のみを単独に発音することがなく、子音+母音すなわちCV型音節として発音される。したがって、子音部から抽出した特徴量は調音結合ならびに分析窓長の有限性などによって母音部の影響を受けていると言える。つまり子音部から抽出した特徴量から、母音部分の影響量を取り除くことによって子音の認識率が向上するものと期待される。

今、子音部および母音部の分析から求めた特徴量をそれぞれ  $C_c$ ,  $C_v$  とした場合、子音本来の特徴量  $C_o$  が、後続母音の影響によって図2のように線形に変移して  $C_c$  になっていると仮定する。

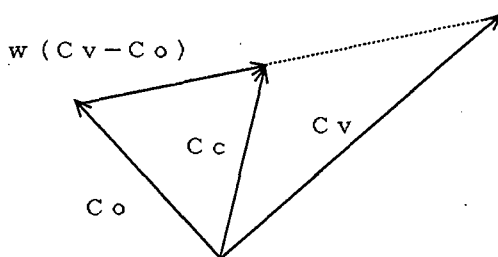


図2 子音部の特徴量  $C_c$  と母音部の特徴量  $C_v$  の関係

図2より、後続母音が子音に及ぼす影響の度合いを  $w$  とすれば、ベクトルの式(1)が成立する。

$$C_o = C_c - w(C_v - C_o) \quad (1)$$

したがって、

$$C_o = \frac{1}{1-w} (C_c - w C_v) \quad (w \neq 1) \quad (2)$$

となる。

すなわち、分析により得られる子音部の特徴量  $C_c$  及び母音部の特徴量  $C_v$  から新たな特徴量  $C_o$  を式 (2) により求め、この特徴量を用いて子音認識を行なうことにより、子音認識率の向上が期待される。なお、 $w=0$  のときは式 (2) より、 $C_o=C_c$  となり、母音部の影響を考慮しない従来の手法と同じとなる。以後、式 (2) の係数  $w$  を影響量係数と称する。以下、本手法を有声破裂子音認識に適用する事により、その有効性を検証する。

#### 4. 有声破裂音認識

##### 4. 1 認識条件

(1) 音声資料 有声破裂音 /b/, /d/, /g/ と母音 /a/ の組み合わせによる CV 型単音節 3 種類を成人男性 37 名が各一回発声した計 111 個を用いた。

(2) 特徴量 一般にホルマントが急激に変化する場合、その時間的変化を正確に追尾するのは困難であるため、ホルマント情報を間接的に担っていると言える LPC ケプストラム係数<sup>(12)</sup>を用いた。分析条件は、分析次数 12、子音部の分析窓長 20 ms、母音部の分析窓長 40 ms である。なお、子音部の特徴量 (すなわち、式 (2) の  $C_c$ ) は視察による破裂時点を時間原点として、分析位置  $T_c$  を -15 ms から 55 ms まで 5 ms 間隔で分析し、各分析位置での LPC ケプストラム係数を用いることにより、子音分析位置の検討も合わせて行った。また、母音部の特徴量 (すなわち、式 (2) の  $C_v$ ) は、視察による母音定常部の中心及びその前後 10 ms を分析して得られる計 3 フレームの

LPC ケプストラム係数の平均値を用いた。

(3) 認識方法 式 (2) により得られる特徴量  $C_o$  を Fisher 空間<sup>(13)</sup>に写像し、この空間上での各クラス (/b/, /d/, /g/ の 3 クラス) の重心からの距離に基づき、/b/, /d/, /g/ の 3 子音を識別した。そして、Fisher 空間における各クラスの類間分散と類内分散の比として定義される Fisher 比<sup>(13)</sup>により /b/, /d/, /g/ の分布の良さを評価した。すなわち、Fisher 比が大きい程 /b/, /d/, /g/ の認識をより安定に行うことができると言える。

##### 4. 2 認識結果

認識率及び Fisher 比の影響量係数  $w$  依存性を図 3 に示す。但し、子音分析位置  $T_c = 20$  ms とした場合の結果である。

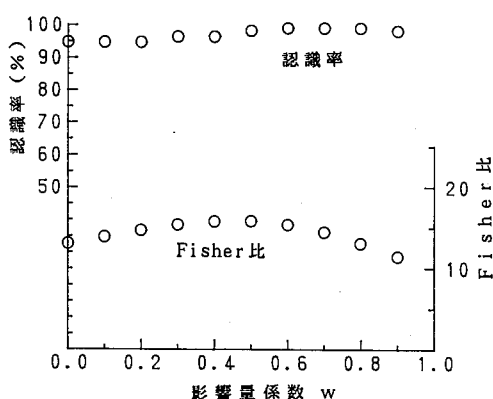


図3 認識率及び Fisher 比の影響量係数  $w$  依存性  
後続母音 /a/ (子音分析位置  $T_c=20$  ms)

図 3 より、認識率及び Fisher 比は影響量係数  $w$  の値によって滑らかに変化し、今の場合、 $w=0.0$  (後続母音の影響を考慮しない従来の方法) では、認識率 94.6%、Fisher 比 13.0 であったのが、 $w=0.5$  において認識率 98.2%、Fisher 比 15.8 にいずれも向上し、影響量係数  $w$  の設定の有効性が示されていると言える。

認識率及び Fisher 比の子音分析位置  $T_c$  依存性を図 4 に示す。但し、図中○印は  $w=0.5$ 、

$\Delta$ 印は $w=0.0$ （後続母音の影響を考慮しない従来の方法）とした結果である。

図4より、LPCケプストラム係数を特徴パラメータとしたFisherの写像空間で有声破裂音認識を行う場合、最適な子音分析位置が存在し、 $T_c$ によるFisher比の変化特性から、 $T_c=20\text{ ms}$ においてFisher比が最大となっていることより、破裂時点から $20\text{ ms}$ 後の位置が最適な子音分析位置であると言える。そして、 $w=0.5$ とすることにより、この付近（ $5\text{ ms} \leq T_c \leq 30\text{ ms}$ ）のFisher比が $w=0.0$ とした場合よりいずれも向上し、後続母音の影響を考慮した

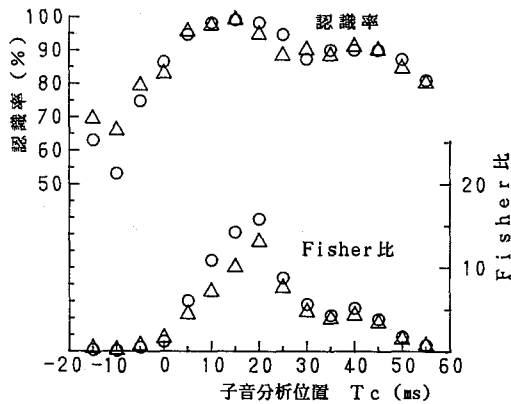
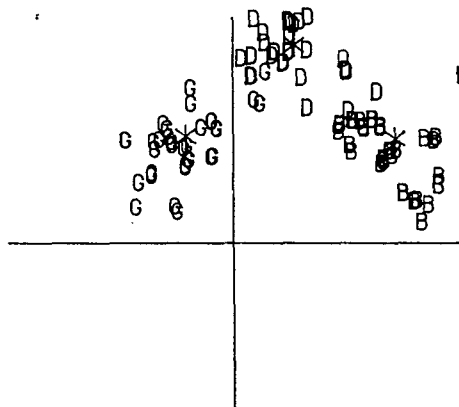


図4 認識率及びFisher比の子音分析位置 $T_c$ 依存性  
後続母音/a/(○印： $w=0.5$ ,  $\Delta$ 印： $w=0.0$ )



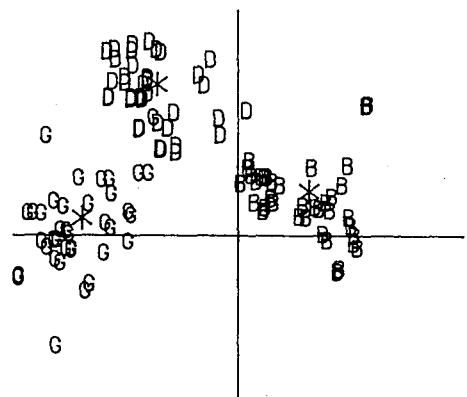
(a) 従来の方法 ( $w=0.0$ ,  $T_c=20.0\text{ ms}$ )

本手法の有効性が示されていると言える。

分析位置 $T_c=20\text{ ms}$ における、Fisher空間での各音声資料の分布の様子を図5に示す。但し、図5(a)は $w=0.0$ 、図5(b)は $w=0.5$ とした場合の分布図であり、図中のB、D、Gおよび\*印はそれぞれ /b/, /d/, /g/ および各有声破裂音の重心位置を示す。

図5より、後続母音の影響を考慮しない従来の方法では、特に/b/と/d/の分布がオーバーラップしているのに対し、本方法では /b/, /d/, /g/ の類内分散が従来の方法よりもやや大きくなっているが、各クラスターの重心間の距離が大きくなっているためFisher比が13.0から15.8に約1.2倍向上し、認識により適した分布状態が得られていると言える。/b/, /d/, /g/ 相互を各重心からのユークリッド距離により認識した結果を表1に示す。

表1より、従来の方法でも、/b/は37個とも正しく認識されているが、3個の/d/及び/g/がそれぞれ/b/と/d/に誤認識されているのに対して、本方法では2個の/g/が/d/に誤認識されているが、他はいずれも正しく認識されており、認識率が平均して94.6%から98.2%に改善する。



(b) 本方法 ( $w=0.5$ ,  $T_c=20.0\text{ ms}$ )

図5 Fisher空間における有声破裂音（後続母音/a/）の分布

表1 Fisher 空間における有声破裂音認識

(a) 後続母音の影響を考慮しない場合  
( $W=0.0$ ,  $T_c=20.0\text{ms}$ )

|              | /b/ | /d/ | /g/ | 認識率     |
|--------------|-----|-----|-----|---------|
| /b/          | 37  |     |     | 100.0 % |
| /d/          | 3   | 34  |     | 91.9 %  |
| /g/          |     | 3   | 34  | 91.9 %  |
| Fisher比 13.0 | 平均  |     |     | 94.6 %  |

(b) 後続母音の影響を考慮した場合  
( $W=0.5$ ,  $T_c=20.0\text{ms}$ )

|              | /b/ | /d/ | /g/ | 認識率     |
|--------------|-----|-----|-----|---------|
| /b/          | 37  |     |     | 100.0 % |
| /d/          |     | 37  |     | 100.0 % |
| /g/          |     | 2   | 35  | 94.6 %  |
| Fisher比 15.8 | 平均  |     |     | 98.2 %  |

## 5. むすび

有声破裂音を、LPCケプストラム係数を特徴パラメータとしてFisherの写像空間において認識する際に、分析により得られる子音部の特徴量及び母音部の特徴量から式(2)に基づき算出した新たな特徴量を用いることにより後続母音の影響を軽減する手法を提案した。本手法を有声破裂音/b/, /d/, /g/と母音/a/の組み合わせによるCV型単音節3種類を成人男性37名が各一回発声した計111個の有声破裂音認識に適用した結果、平均認識率が94.6%から98.2%に改善し、かつFisher比も13.0から15.8に向上するとの結果が得られ、本方法の有効性が明かとなった。

なお、本論文では影響量係数を音声資料に関わらず一定としたが、後続母音の子音部に及ぼ

す影響度合いには個人差があることから、影響量係数を音声資料ごとに適応的に設定する手法の検討が今後の課題と言える。

## 文 献

- (1) C.S. Myers, L.R. Rabiner: "Connected digit recognition using a level-building DTW algorithm", IEEE Trans., Acoust., Speech & Signal Process., ASSP-29, pp.351-363(1981).
- (2) 中川: "パターンマッチング法による連続単語および連続音節の音声認識アルゴリズム", 信学論(D), J66-D, 6, pp.637-644(1983).
- (3) 菅村, 鹿野, 好田: "SPLIT, 単語マルチテンプレート法による不特定話者単語音声認識", 信学論(D), J67-D, 10, pp.1210-1217 (1984).
- (4) B. Aldefeld, et al.: "Automated directory listing retrieval system based on isolated word recognition", Proc. IEEE, 68, pp. 1364-1379(1980).
- (5) 古井: "単音節認識とその大語い単語音声認識への適用", 信学論(A), J65-A, 2, pp.175-182(1982).
- (6) 菅村, 古井: "疑音韻標準ボタンによる大語い単語音声認識", 信学論(D), J65-D, 8, pp.1041-1048(1982).
- (7) 二矢田, 平岡, 森井, 星見: "音素を単位とした小型・高速の不特定話者・多数語用音声認識装置", 音響誌, 43, 4, pp.247-254(1987).
- (8) 中川, 中西: "語中のCV音節を標準パターンとする不特定話者の大語彙単語音声認識", 信学論(D), J70-D, 12, pp.2460-2468(1987).
- (9) 古市, 谷口, 今井: "コンテキスト独立な音素認識により得られた信頼度付き音素ラティスをを用いる単語音声認識", 信学論(D-II), J75-D-II, 3, pp.449-458(1992).
- (10) 渡辺, 磯谷, 塚田: "半音節を単位とするHMMを用いた不特定話者音声認識", 信学論(D-II), J75-D-II, 8, pp.1281-1289(1992).
- (11) 大泉, 藤村: "音声科学", 東京大学出版会(1972).
- (12) J. Makoul: "Linear prediction: a tutorial review", IEEE Proc., 63, pp.561-580(1975).
- (13) R.O. Duda and P.E. Hart: "Pattern Classification and Scene Analysis", John Wiley(1973).