A Consideration on Recognition of Voiced Plosives

有声破裂音認識に関する一考察

三好　義昭

A Consideration on Recognition of Voiced Plosives

Yoshiaki MIYOSHI

1．まえがき

近年のディ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話者を限定しない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[(1)~(3)]では限界があり，音素単位での認識に基づく手法[(4)~(10)]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ためには，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，子音部に及ぼす後続母音の影響量を軽減することができれば子音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量を軽減する手法を示し，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

平成12年10月2日受理

るのと同時期，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。

以上のような生成過程で発声された有声破裂音は，次の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波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：声道の形状が狭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，周波数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[11]。

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後続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，後続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。本論文ではその一手法として，後続音韻部の特徴パラメータを係数とする波形のフィルタリングによって，後続音韻部の影響量を軽減した波形を求め，この波形に対して通常の認識手法を適用する方式（以下，フィルタリング方式と称する）を提案する。

本方法の処理手順を図１に示めす。まず，後続母音部を線形予測分析することによって後続母音部の線形予測係数を得る。この線形予測係数をフィルター係数として子音部の音声波形をフィルタリングすることにより，後続母音部に対する予測誤差波形を求める。すなわち，このフィルターの係数は認識対象の子音部の線形予測係数ではなく，後続音韻部の線形予測係数を用いるので，このフィルターの出力波形は後続音韻部の線形予測係数で予測可能な波形情報が除去され，この係数では予測できない波形，すなわち調音結合の影響を軽減した子音部本来の波形が抽出できることになる。この後は，通常のＬＰＣケプストラム係数を特徴ベクトルとするＬＰＣケプストラム空間での認識と同じである。すなわち，この予測誤差波形に対して通常の線形予測分析を行って得られる線形予測係数をＬＰＣケプストラム係数に変換し，音声認識の特徴パラメータとして用いる（(注)前節で述べたように，遷移部における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が，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

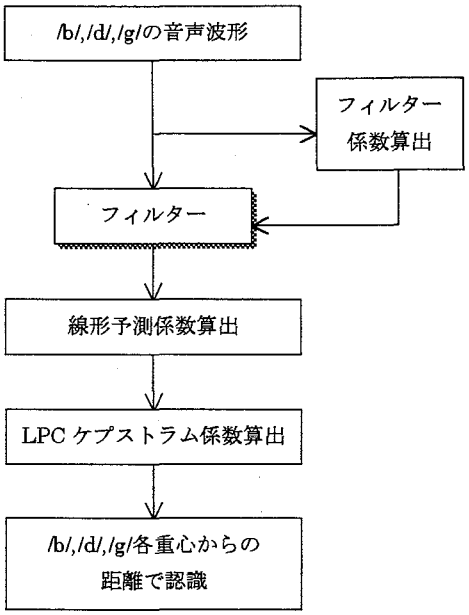

図１　本手法の処理手順

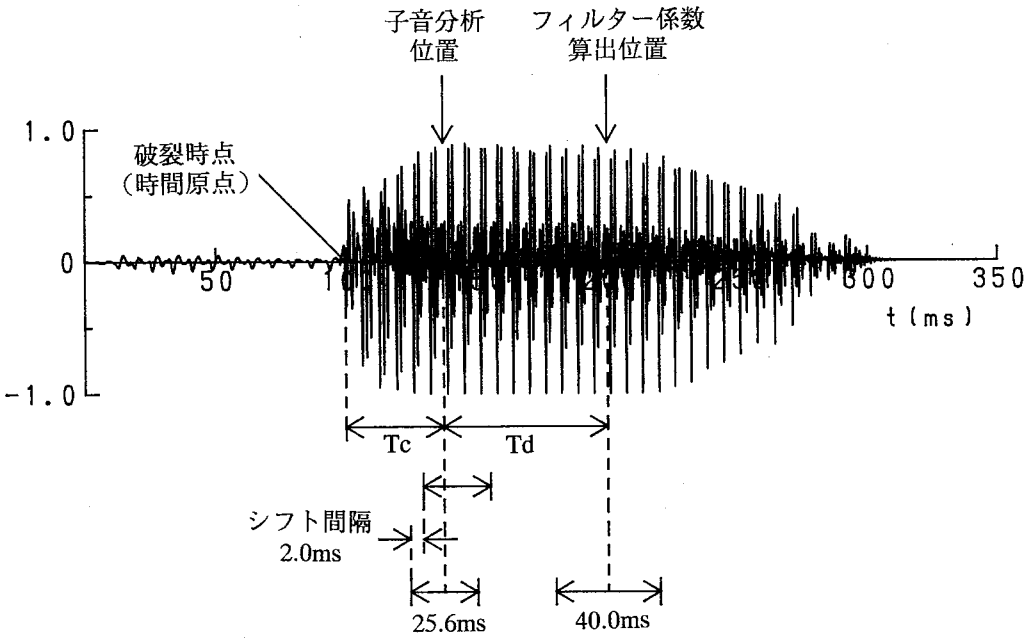

図2　子音分析位置及びフィルター係数算出位置の概略

的に担っていると言えるＬＰＣケプストラム係数[12]を有声破裂音認識の特徴パラメータとして用いた)。このＬＰＣケプストラム係数の各重心からの距離による認識率ならびにＬＰＣケプストラム係数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分散比によって改善度合いを評価することにより，本フィルタリング方式の有効性を検証する。

4．フィルタリング方式による有声破裂音認識

4．1　実験条件

（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45名が各一回発声した計135個を標本化周波数10kHzでサンプリングした音声データを用いた。

（2）分析条件　子音部の分析条件は従来の方法との比較のため，一般的に用いられている分析次数：12次，分析窓長：25.6msとし，本方法のフィルター次数を2次から10次まで1次間隔，フィルター係数算出窓長を10.0msから40.0msまで5.0ms間隔，子音分析位置（Ｔｃ）を視察にる破裂時点を時間原点として，−10.0msから50.0msまで2.0ms間隔，フィルター係数算出位置（Ｔｄ）を子音分析位置を時間原点として,0.0msから50.0msまで2.0ms間隔で分析した。すなわち，音声波形上では破裂時点を時間原点としてＴｃ＋Ｔｄの時点がフィルター係数算出位置となる。図2にＴｄ＝40.0msとした場合の概略を示す。

4．2　認識結果

認識率及び分散比の子音分析位置Tc依存性を図3に示す。但し，図中○印は本方法による結果で，フィルター次数：4次，フィルター係数算出窓長：40.0ms，フィルター係数算出位置Ｔｄ＝18.0ms，とした場合，×印は従来の方法による結果である。

図3より，従来の方法ならびに本方法とも認識率及び分散比はＴｃの値に応じてほぼ滑らか

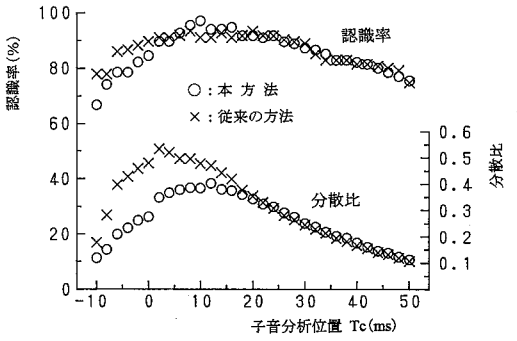

図3　認識率及び分散比の子音分析位置依存症

に変化し，ＬＰＣケプストラム空間上での分布の良さを示す分散比が最大となるのは，従来の方法ではＴｃ＝2.0msにおいて，分散比0.533，本方法では，Ｔｃ＝12.0msにおいて，分散比0.403となり，分散比の最大値は従来の方法の方が良いが，分散比が最大となる子音分析位置Ｔｃが従来の方法ではＴｃの時間原点である破裂時点直後であるのに対し，本方法では破裂時点以降の遷移部に移行すると共に分散比の変化度合いがより緩やかとなり後続音韻の影響軽減の効果が得られており，認識率が従来の方法ではＴｃ＝8.0msにおいて，93.3%であったのが，本方法ではＴｃ＝10.0msにおいて，97.0%に向上している。

認識率及び分散比のフィルター係数算出位置Ｔｄ依存性を図4に示す。但し，フィルター次数：4次，フィルター係数算出窓長：40.0ms，Ｔｃ＝10.0msとした場合の結果で，従来法の認識率及び分散比をそれぞれの座標軸付近に矢印で示す（但し，それぞれ従来法の認識率が最大となるＴｃ＝8.0msでの値。以下同様)。

図4より，分散比はフィルター係数算出位置Ｔｄと共にほぼ単調に増大している傾向がみられ，Ｔｄ＞50.0msにおいて分散比はさらに良くなる可能性があるが，認識率がＴｄ＞40.0msにおいて減少傾向となることからＴｄの範囲を50.0msまでとした。なお，認識率は10.0ms≦Ｔｄ≦40.0msであればさほど変化せず，Ｔｄの最適値は18.0msであると言える((注)Ｔｄ＝

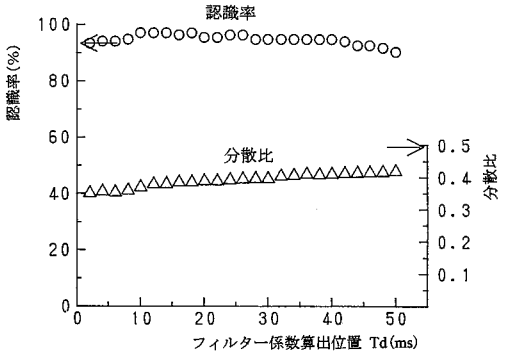

図4　認識率及び分散比の
フィルター係数算出位置依存症

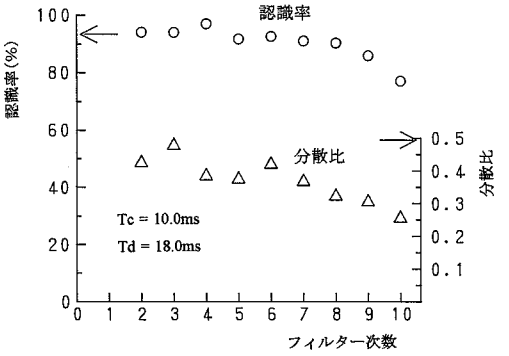

図5　認識率及び分散比の
フィルター次数依存症

10.0ms，12.0ms，14.0ms，18.0msの時いずれも認識率97.0%となるが，同一認識率の場合には分散比が大きい方を最適値とした。以下同様)。

認識率及び分散比のフィルター次数依存性を図5に示す。但し，フィルター係数算出窓長：40.0ms，Ｔｃ＝10.0ms，Ｔｄ＝18.0msとした場合の結果で，従来法の認識率及び分散比をそれぞれの座標軸付近に矢印で示す。

図5より，フィルター次数が7以下において，分散比は多少変動しているが認識率は殆ど変化していないのに対して，フィルター次数が8以上になると認識率ならびに分散比とも減少しており，フィルター次数が4において最大認識率97.0%が得られることが分かる。

認識率及び分散比のフィルター係数算出窓長

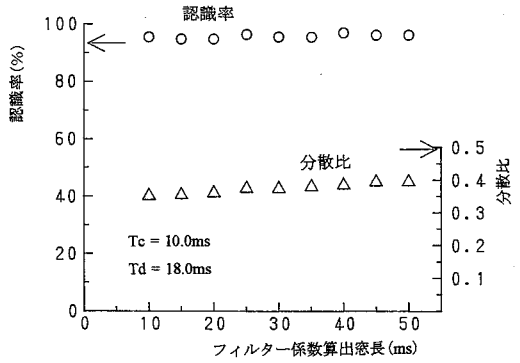

図6　認識率及び分散比の
フィルター係数算出窓長依存症

依存性を図6に示す。但し，フィルター次数：4次，Tc＝10.0ms，Td＝18.0msとした場合の結果で，従来法の認識率及び分散比をそれぞれの座標軸付近に矢印で示す。

図6より，分散比はフィルター係数算出窓長と共に若干ながら単調増大の傾向がみられるが認識率はフィルター係数算出窓長には殆ど依存せず窓長が25.6msならびに40.0msにおいて，最大認識率97.0%が得られ，フィルター係数算出位置Td依存性のところで述べたように，分散比の大きい窓長40.0mをフィルター係数算出窓長の最適値とした。

/b/,/d/,/g/相互間の認識状況を従来の方法と本方法のそれぞれについて表1に示す。

表1より，従来の方法でも，/b/は45個とも正しく認識されており/b/の認識率は100.0%であるが，3個の/d/が/g/に誤認識され/d/の認識率は93.3%，そして2個の/g/が/b/，4個の/g/が/d/にそれぞれ誤認識され/g/の認識率は86.7%にしかならないのに対して，本方法では1個の/d/が/g/に誤認識，1個の/g/が/b/，2個の/g/が/d/にそれぞれ誤認識されているのみで，/b/の認識率100.0%，/d/及び/g/の認識率はそれぞれ97.8%及び93.3%となり，本方法により，特に/g/の認識率が大幅に改善されたと言える。

なお，LPCケプストラム空間における/b/,

表1　有声破裂音認識結果の詳細

(a) 従来の方法による認識結果

	/b/	/d/	/g/	認識率(%)
/b/	45			100.0
/d/		42	3	93.3
/g/	2	4	39	86.7
分散比 0.496		平　均		93.3

(b) 本方法による認識結果

	/b/	/d/	/g/	認識率(%)
/b/	45			100.0
/d/		44	1	97.8
/g/	1	2	42	93.3
分散比 0.391		平　均		97.0

/d/,/g/各クラスの類間分散と類内分散の比である分散比が従来の方法では0.496であるのに対して，本方法では0.391と従来の方法の方が大きいが，これは従来の方法では/b/の認識率が100.0%であるのに，/d/と/g/相互の誤認識が多いことより，従来の方法で得られる/d/と/g/の分布領域が互いに近接しているにもかかわらず，/b/の分布領域がこれらからかなり離れた位置に存在しているため，/b/と/d/ならびに/g/間の類間分散が大きな値となり，結果的に分散比が大きくなっているに過ぎないと言える。すなわち，単純に分散比が大きい方が良いとは言えず，認識システムは第一義的には認識率で評価し，同一認識率となった場合に分散比等を考慮するのが適切と言える。

5．むすび

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後続母音部の影響を受けており，この影響を軽減できれば子音本来の特徴量が抽

出できると言える。本論文ではその一手法として，後続音韻部の特徴パラメータを係数とする波形のフィルタリングによって，後続音韻部の影響量を軽減した波形を求めるフィルタリング方式を提案した。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45名が各一回発声した計135個の有声破裂音認識に適用した結果，平均認識率が93.3%から97.0%に改善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，本論文では本方法のフィルター次数ならびにフィルター係数算出位置を音声資料に関わらず一律に設定したが，後続母音が子音部に及ぼす影響度合いには個人差があることから，これらのパラメータ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), J 66-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), J 67-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), J 65-A, 2, pp.175-182 (1982).

(6) 管村，古井："疑音韻標準パタンによる大語い単語音声認識"，信学論 (D), J 65-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), J 70-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), J 75-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).