5W-10

子音の自動認識のための音響／音声変換について

山崎　岳人　　野々山　秀文　　北澤　茂良
静岡大学

１．はじめに

我々は、話者に依存しない、大語彙、連続音声を認識する際に、音響的な特徴から音素を認識する音響／音声変換（Acoustic Phonetic Decoding)について研究している。[1]

本研究では、破裂子音/p,t,k,b,d,g/ および鼻子音/m,n/を対象とした調音事象抽出と音素認識のための機構について考察する。

２．音響／音声変換

図１に提案する音響／音声変換の手順を示す。

音声信号
|
素特徴抽出部
|
音素認識部

図１　音響／音声変換の処理手順

音響／音声変換とは、観測した音声波から特徴を抽出することによって、音素・音節を認識する一連の手順全体を指す。提案する音響／音声変換は、調音事象を抽出する素特徴抽出部とその結果に基づいて音素を認識する音素認識部に大別できる。調音事象とは、音素の調音に伴う音響的事象を指す。また調音事象が生起した時点を調音時点と呼ぶ。素特徴抽出部は、閉鎖と破裂（開放）に関する調音時点を求める。[2]このための特徴パラメタとして音声パワーの推移を用いた。また音素認識部において、素特徴抽出部からの出力である調音時点について疑問が生じた場合、調音時点を求めなおす後戻りが必要である。この動作はPrologのバックトラックを用いて実現した。[3]

２．１　素特徴抽出部

この処理部は、入力された音声信号より音声パワーの推移を調べ、調音時点を抽出する。ハミング窓中心の音声パワーを、窓長16msのハミング窓を用いて切り出した音声信号より求めた。この窓を各標本点毎に移動して音声パワーの時系列を抽出した。さらに、ある区間における音声パワーの傾きを表す回帰係数を求める。回帰係数は６つの閾値で分割し、増加（３段階）、減少（３段階）、定常を表す記号を用いて表現した。図２にその記号と意味を示す。

I(Increase):増加中	Ih > Im > Il
D(Decrease):減少中	Dh > Dm > Dl
C(Constant):一定	C

図２　音声パワー推移の記号と意味

調音時点を唯一に決定することは一般に困難である。そこで時間方向に探索を行い、複数の調音時点を候補として挙げる。本研究では、これら複数の調音時点をリスト形式をもちいて調音時点リストとして表す。また、調音事象を示すリストを調音事象リストと呼ぶ。調音時点の抽出が終了した時点で、調音時点リストの見直しを行う。例えば、ある隣り合った２つの調音時点において、時間的に前方の調音事象が破裂（開放）であり後方の調音事象が閉鎖の場合、後方の閉鎖点を調音時点リストおよび調音事象リストから取り除く。

２．２　音素認識部

この処理部は、調音時点リストより順に１組の調音時点（閉鎖点と破裂点）を取り出し音素認識を行う。取り出す順序は、調音時点リストが

$$[\alpha 0, \alpha 1, \cdots\cdots, \alpha n-1, \alpha n]$$

とすると、

$$\begin{aligned}(\text{閉鎖点}, \text{破裂点}) = &(\alpha 0, \alpha n)\\ &(\alpha 0, \alpha n-1)\\ &\cdots\cdots\\ &(\alpha 0, \alpha 0)\end{aligned}$$

の順であり、取り出したある１組において音素が決定できたならば、その時点で音素認識

Acoustic Phonetic Decoding for Automatic Recognition of Consonant
Takahito Yamazaki, Hidefumi Nonoyama, Shigeyoshi Kitazawa
Shizuoka University

が終了する。もし、全ての組について音素が決定できなければ、調音時点（閉鎖点）の抽出のやり直しが起きる。

音素認識はまず閉鎖点と破裂点（開放点）の間のフレーム数と音節長とを比較し、以下の規則によって音素の候補を求める。

1) 破裂点から閉鎖点までの長さが音節長の1/7以下ならば
(/p/,/t/,/k/)である。 ［規則１］

2) 破裂点から閉鎖点までの長さが音節長の1/7より長いならば
(/b/,/d,g/,/m,n/)である。［規則２］

音素候補が求まった後、さらに詳しい規則に従って音素の決定を行う。

２．３ プログラム間の特徴パラメタの受け渡し

音声信号のバイナリ・データをPrologで扱うには問題が多い。したがって直接音声信号を扱う部分にはＣ言語を用い、求めた特徴パラメタをＣ言語プログラムからPrologプログラムへ受け渡した。その様子を図３に示す。

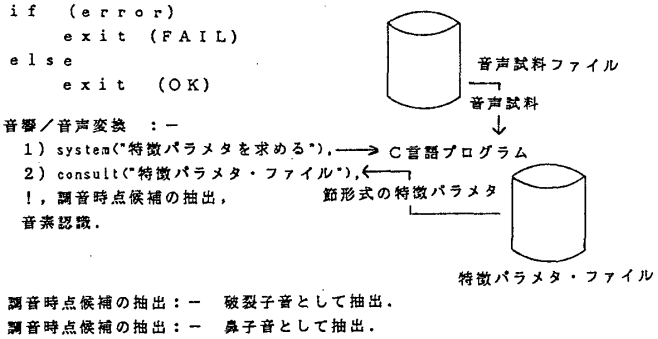

図３ 特徴パラメタの受け渡し

1)関数systemによってPrologよりＣ言語プログラムを呼び出す。Ｃプログラムは音声試料を読み込み、特徴パラメタである音声パワーの推移を求め、それをPrologによって扱えるよう節形式で特徴パラメタ・ファイルに格納する。特徴パラメタを求めることができた時には正常終了（exit(OK)）し、求められなかった時には異常終了(exit(FAIL))する。2)正常終了の場合は関数consultに進み、調音時点の抽出、音素認識のための手続を呼出す。

３． 認識実験の結果

フランス語のひとりの話者について調音事象の抽出および認識の規則を調整した後に、別の話者について認識実験を行った。この結果から素特徴抽出部と音声認識部との関係を調べた。使用した音声試料は、フランス語と日本語の２種類である。フランス語はＣＶ音節88種類、日本語はＣＶ音節30種類である。フランス語、日本語ともそれぞれ話者は成人男性５人である。１つのＣＶ音節について調音時点候補は、平均5.2個求まったが、このうちのひとつが視察による調音時点から1.25ms(回帰係数を求めた区間長の２倍の誤差)以内の場合、正しく調音点が抽出されたものとした。認識実験の結果を表１に示す。

（フランス語）

音素	/p//t//k//b//d,g//m,n/
認識率	.63.74.78.51 .53 .84
調音時点的中率	.76.67.87.56 .61 .78

（日本語）

音素	/p//t//k//b//d,g/
認識率	.48.68.76.60 .64
調音時点的中率	.84.76.92.80 .76

表１ 認識実験結果

ここでの認識率は、正しい調音時点を用いて正しく認識を行った場合と誤った調音時点を用いて正しく認識してしまった場合がふくまれる。調音時点的中率とは、各音素を決定した時に用いた調音時点のうち、それが正しい調音時点であった割合である。

４． おわりに

特徴パラメタとして用いた音声パワーの推移は確かに話者独立性の強い特徴であることが確認できた。しかし音声パワーの推移だけでは、音素認識を行うことはできないので、他のパラメタを組み入れることが必要である。

また本研究では、Prologを用いて認識を行った。それぞれのプログラム言語はそれぞれの特徴を有し、目的にあわせて用いるべきである。その意味で本研究は、Prologがこれから音声認識の分野に取り入れられていく言語であるか検証する研究でもあった。その答えは、現時点で早急に得ることはできないが、知識の表現形式、推論の導入および知識の追加、削除の容易さの点においては有効であった。

謝辞 日頃ご支援とご討論をいただくATRと『音声言語』の皆様に感謝致します。

参考文献

(1)野々山、北澤：”音響／音声変換を用いた子音の自動認識”,音学講論1-2-8 (1988.3)

(2)北澤茂良：”フランス語の頭位の破裂子音および鼻子音の統計的判別”,信学技報 sp87-16 (1987.6)

(3)H.Meroni,R.Bulot:"A PROLOG II ENVIRONMENT FOR PROCESSING ACOUSTIC AND PHONETIC KNOWLEDGE"