TR-I-0072

# スペクトログラム・リーディング知識を用いた音韻セグメンテーション・エキスパートシステム

# Phoneme Segmentation Expert System Using Spectrogram Reading Knowledge

畑崎香一郎　小森康弘
Kaichiro Hatazaki, Yasuhiro Komori

1989. 3

## 概要

専門家はスペクトログラム・リーディングによって音韻境界を非常に正確に見つけることができる。ここで専門家が用いている手法および知識を用いて、連続音声中の音韻セグメンテーションを行うエキスパートシステムを構築した。このシステムはルール化された専門家の知識、戦略に従って入力音声中の音韻を検出し、その境界を決定する。音韻環境に依存するあいまいな知識を表現するために確信度を用いた仮説推論を行い、音韻環境の仮説のもとで音響特徴を抽出する。この枠組みによって調音結合などによる音韻の変形に対処でき、またセグメンテーションに必要なスペクトログラム上の特徴を正確に抽出できる。日本語子音のセグメンテーション知識を記述して実験を行った結果、90.8%の音韻が正しく検出でき、特に音韻境界誤差は5.8msecと、視察による場合と同程度に正確であった。

ATR 自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

# 目次

# 1. はじめに

スペクトログラム・リーディングは、音声スペクトログラム上に視覚的に現れている音声の音響特徴(例えば有声/無声の別、パワーの強さと変化、破裂の位置と強さ、摩擦の強さ、ホルマントの位置など、以下これらをスペクトログラム特徴を呼ぶ)を参照して、音韻の位置と種類とを読み取る技術である。専門家はスペクトログラム・リーディングを通じて音韻の音響特徴に関する知識を経験的に獲得し、この結果、連続音声中の音韻を80%以上の正確さで読み取ることができるようになる[Zue 79]。獲得された知識は、調音結合などの影響による音韻の変形に関するものが多く、連続音声認識を行う際には非常に有用である。このため、スペクトログラム・リーディング知識を用いた特徴ベースの音韻認識の試みがこれまでにいくつか行われており[Carbonell 84][Zue 86][Stern 86][Connolly 86][Mizoguchi 87]、スペクトログラム・リーディング知識の音韻識別に対する有効性が示された。

一方、専門家のスペクトログラム・リーディングでは音韻の識別だけでなく音韻位置の検出、すなわち音韻セグメンテーションも同時に行われる。このときにも専門家は音韻の音響特徴や調音結合に関する知識を持ち、音韻環境に応じた戦略やスペクトログラム特徴を用いている。これによって、調音結合などによる音韻の種々の変形にもかかわらず音韻境界を非常に正確に見つけることができる。

筆者らはこのスペクトログラム・リーディングの音韻セグメンテーションに対する有効性に着目し、スペクトログラム・リーディングの知識、戦略によって音韻セグメンテーションを行うエキスパートシステムを構築した。このシステムは、ルール化された専門家の知識、戦略に従って入力音声のスペクトログラムを読み、入力音声中の音韻とその左右の音韻境界を決定する。本システムは、汎用エキスパートシステム構築ツールであるART[ART]を用いて構築した。

スペクトログラム・リーディングで用いられる専門家の知識と戦略は音韻環境に依存し、かつあいまいである。また種々の局所的、大域的なスペクトログラム特徴を手掛りにしている。スペクトログラム・リーディングを計算機上で実現するためには、このような知識、戦略をエキスパートシステムのルールとして自然にかつ容易に表現できる枠組みが必要となる。

上記の従来のシステムではこれを単純なif-thenルールと確信度の枠組みで表現していた。本システムでは、仮説推論によって音韻環境に依存する知識と非決定的な戦略を表現し、確信度およびファジイ集合の考えによって知識の不確実性とファジイ性を表現する。これらの枠組みにより、専門家の知識を容易にかつ自然に表現できる。

スペクトログラム特徴に関しては、スペクトログラムからの抽出が困難であることから、従来は限られた特徴をあらかじめ視察や手続き的処理の前処理で求めたものを記号化し、これにルールを適用したものが多い。スペクトログラム特徴を前処理で求める場合には、音韻の特徴や調音結合に関する知識などのトップダウン情報が使えないために微妙な特徴がうまく抽出できない、スペクトログラム上に現れている特徴のすべてを前もって求めておくわけにはいかないために使用できるスペクトログラム特徴が限られるなどの問題点が生じる。画像処理のアプローチで自動的に特徴を抽出することも試みられている[Leung 86]。しかしながら、スペクトログラム特徴には大局的な特徴もあれば、局所的で微妙な特徴もあり、その自動的な抽出は一般に容易ではない。

これに対し、筆者らの方法では、スペクトログラム特徴は、ルールによって参照されるときに音韻環境に応じた適切な抽出方法およびパラメータを用いてスペクトログラムから抽出する。これによってセグメンテーションに必要な特徴を正確かつ容易に得ることができる。

本報告書では、専門家のスペクトログラム・リーディング知識を自然にかつ容易に表現するための枠組み、その枠組みの上での音韻セグメンテーションの方法を述べたのち、日本語子音セグメンテーション実験の結果を述べる。

# 2. 知識表現の枠組み

## 2.1 音韻環境に依存する知識と非決定性の表現

連続音声中の音韻は前後の音韻との調音結合の影響を受け、さまざまに変形する。これに対し専門家は、いろいろな音韻環境あるいは音韻変形を仮説し、それぞれの仮説応じた適切な戦略あるいはスペクトログラム特徴によってスペクトログラムを読み、より確かな音韻境界を得る。

本システムでは仮説推論を用いて音韻環境や音韻変形に依存する知識、非決定的な戦略を扱う[1]。仮説推論では、いくつかの事象を仮説した世界(仮説世界)が作られ、それぞれの仮説世界の中で独立に推論が進められる。これによって複数個の解の可能性を並行して調べることができる。また、複数個の仮説の組合せが前提となる場合にはそれらの仮説が同時に成り立つ仮説世界が自動的に作られる。相矛盾する仮説の組合せの禁止など、仮説世界の整合性はエキスパートシステム構築ツールの持つATMS(Assumption-based truth maintenance system)[Kleer 87]によって保たれる。

音韻環境に依存する知識は音韻環境を前提条件とするルールとして記述され、その音韻環境が仮説された仮説世界の中で適用される。いくつかの音韻環境が考えられるときには、それぞれの音韻環境の仮説のもとで独立にかつ並行して音韻境界の検出が試みられる。

加えて、それぞれの仮説に対してはそれが正しいかどうかの評価が行われる。この結果、正しい仮説のもとでは一連のルールが矛盾なく適用され、音韻境界を得ることができる。一方、間違った仮説のもとでは、ルールが前提としている音韻環境が実際のものとは異なっているため、ルールの適用途中で矛盾が生じ音韻境界を得るに至らないか、もしくは信頼度の低い結果しか得られない。いくつかの仮説のもとで得られた結果のうち、信頼度のもっとも高いものが最終結果となる。

仮説推論を用いることで専門家の持つ音韻環境、変形に依存する知識は、そのままその音韻環境、変形を前提とするルールとして自然な形で記述することができる。また、それぞれのルールの記述の際には、そのルールが適用される音韻環境、変形だけを念頭に置けばよいので、ルールの記述は非常に容易になる。

## 2.2 不確実性の表現

スペクトログラム・リーディングにおいて専門家は、いく通りかの音韻環境、変形の仮説を立てるとともに、その仮説の真偽の判定を、仮説を肯定あるいは否定する証拠となるスペクトログラム特徴を検証することによって行う。しかしながら、これによって仮説の真偽は明確に定まるわけではなく、それぞれの仮説の確からしさが明らかになるにすぎない。また、その証拠自身についてもスペクトログラム特徴の存在の有無は普通明確ではなく、「その特徴はかなりの程度認められる」とか「どちらかというとその特徴がある」といったように、証拠の確からしさとして捕らえることが多い。つまり仮説の確からしさは証拠の重要度、信頼度、およびその数などを総合して判断される。このように、スペクトログラム・リーディング知識およびリーディング結果のすべては、その真偽は明確ではなく、「確からしさ」の尺度で表される。

不確実な知識を扱う問題では一般に、仮説と証拠の間にAND、OR、およびCOMB(combination)の各関係がある。AND、OR関係にある証拠はそれぞれ仮説の必要条件、十分条件となり、COMB関係にある証拠は仮説を独立に支持あるいは反証する[Ishizuka 85]。「0-500Hzパワーが大きい」ことは母音であることを肯定する証拠であると同時にその必要条件でもあ

1 ARTの仮説型ビューポイント機構を用いる[Clayton 84]。

る。一方、/k/以外の破裂はダブルバーストを持たないから、「ダブルバーストを持つ」ことは破裂音であることを支持する証拠であるがその必要条件ではない。さらにスペクトログラム-リーディングでは、これらの証拠が独立にかつ順不同に検証される。

不確実な知識を扱う方法としては、Bayes確率、MYCINの確信度、主観的Bayesの方法、Dempster-Shaferの確率理論、ファジイ論理などがあり、それぞれに長所、短所がある。本システムでは、無知量を扱えること、COMB関係の演算ができること、証拠の確からしさや重要さを直感的に与えやすいこと、確信度の演算が簡易で理解しやすいことの理由からMYCINで用いられた確信度計算のモデル[Buchanan 85]を基本として用い、さらに、複数の証拠が順不同に検証されることに対処した。

MYCINの確信度計算のモデルでは、確信度$CF$は[-1、1]間の値をとり、$CF$=-1はその仮説の完全な否定、$CF$=1は完全な肯定を意味する。$CF$=0はその仮説が正しいとも間違っているとも言えない状態である。なんの証拠も得られていない状態の仮説の確信度は$CF$=0である。また、2つの証拠$x$、$y$の確からしさ$CFx$、$CFy$は以下の式で統合される。

(1)x、yがAND関係のとき、

$$and(CFx, CFy) = min(CFx, CFy)$$

(2)x、yがOR関係のとき、

$$or(CFx\ CFy) = max(CFx\ CFy)$$

(3)x、yがCOMB関係のとき、

$$combine(CFx、\ CFy) = \begin{cases} CFx + (1 - CFx)CFy & CFx, CFy\ both > 0 \\ \dfrac{CFx + CFy}{1 - min(|CFx|, |CFy|)} & one\ of\ CFx, CFy < 0 \\ -combine(-CFx、\ -CFy) & CFx, CFy\ both < 0 \end{cases}$$

一般に、N個の証拠の確からしさは、上記の式を順次適用することによって統合できる。

上記の確信度演算は、*and*、*or*、*combine*のそれぞれの演算では結合律が成立する[2]。すなわち、いずれか1種類の演算で3個以上の証拠の確からしさを統合する場合には統合の順序に依らず結果は同じである。しかしながら、これらの演算を混在させると統合の順序によって結果が異なってしまい[3]、このままではAND、OR、COMB関係が混在するいくつもの証拠を順不同に検証することができない。

そこで本システムでは、推論途中の仮説の確信度$CFh$を確からしさの3つ組で表すことでこれに対処する。すなわち、

$$CFh = \{CFand,\ CFor,\ CFcomb\}$$

とする。$CFand$、$CFor$、$CFcomb$はそれぞれAND、OR、COMB関係にある証拠の確からしさだけを統合した結果である。仮説の確信度の初期値、すなわち証拠が一つも得られていないときの値は、

2 上記の*combine*の式はEMYCINで導入されたものである。初期のMYCINで用いられた*combine*では結合律は成り立たなかった。

3 例えば、$combine(\ and(\ 0.6, 0.2), 0.8) = combine(\ 0.2, 0.8) = 0.84$ に対し
$and(\ combine(\ 0.6, 0.8), 0.2) = and(\ 0.92, 0.2) = 0.2$ となる。

$$CFinitial = \{1, -1, 0\}$$

である[4]。順次検証される証拠の確からしさは、仮説との関係に応じて*and*、*or*、*combine*のいずれかの演算で*CFand*、*CFor*、*CFcomb*のそれぞれの値に統合し[5]、全ての証拠を検証し終わった時点で次式によってスカラ値に変換する[6]。

$$CFh = max(min(CFcomb, CFand), CFor)$$

なおCOMB関係にある証拠の確からしさを統合するときには、仮説に対する重要さを表す重み[7]を確からしさの値に掛ける。

以上の方法では、仮説の検証途中では*and*、*or*、*combine*の演算を独立に行っておりそれぞれの演算では結合律が成立している[8]。従って、それらをスカラ値に統合した結果も検証順序に依存しない。すなわち、AND、OR、COMB関係の混在したいくつもの証拠の検証結果を順不同に統合することができる。

## 2.3 ファジイ性の表現

スペクトログラム特徴に関する専門家の知識にはファジイ性がある、言い換えれば定性的である。例えば「母音の低域パワーは強いが、有声摩擦音ではそれほど強くない」といったように、物理量が「かなり強い」、「強い」、「それほど強くない」、「少し弱い」、「弱い」など境界のはっきりしない定性的な単語で表現される。しかも、専門家にとって物理量とこれらの表現との対応関係の感じ方は、音韻やその環境および基準の取り方によって異なる。例えば-50dBの低域パワーは母音であれば「少し弱い」が、有声摩擦音にとっては「ちょうどよい大きさ」であり、一方-20dBは母音の低域パワーとしては「十分強い」が、有声摩擦音であれば「強すぎる」と感じる。

このような境界が曖昧な量を扱うための枠組みとしてファジイ集合論[Zadeh 65]があり、メンバーシップ関数を用いることで定性表現を量的に特性づける。本システムでもこの考え方を用いて知識のファジイ性を取り扱う。

スペクトログラム・リーディングでは上で述べたように「強い」、「弱い」などの定性表現と物理量との対応は一定ではないので、定性表現のそれぞれにただ一つのメンバーシップ関数を与えることはできない。そこで、それぞれのスペクトログラム特徴に対しそれぞれの音韻環境でのメンバーシップ関数を定義する。例えば図1は語中の有声摩擦音の0-500Hzパワーに対するメンバーシップ関数である[9]。この関数は、ある音韻環境でのスペクトログラム特徴のもっともらしさを表している。従って、抽出された物理量に対する関数の値は、その特徴が存在することの確からしさを表していると言える。そこで関数の値の範囲を確信度と同じ[-1、1]とする

4 確信度の初期値は関数*CFunknown()*で与える。

5 仮説の確信度*CFh*に証拠の確信度を*and, or, combine*演算で加える。すなわち、第1引数に仮説の確信度、第2引数以降に証拠の確信度をとる。*combine*演算を行うときには、証拠の確信度に証拠の重要度に従った重み係数を掛ける。

6 スカラ値への変化については、$CFh = min(max(CFcomb, CFor), CFand)$ も考えられるが、セグメンテーション・ルールでは十分条件(OR条件)はほとんど現れないので、現れる場合はその存在を尊重し、必要条件(AND条件)よりも優先するために本文の式を用いた。

7 ルール中ではstrong-evidence, evidence, weak-evidenceのいずれかで表現し、それぞれ0.8、0.5、0.3の重みが掛けられる。この他にno-evidenceには0.0の重みが掛けられる。

8 ある仮説のすべての証拠を検証し終わっていないうちに、確信度によって仮説の枝刈りをする場合がある。このときには、その段階でスカラー値を求めるので、当然それまでに検証された証拠によって確信度が決まり、証拠の検証順序によって確信度が異なってしまう。しかしながら、通常枝刈りは-0.2といった小さい確信度で行うので、結果的にはAND条件のみによって枝刈りされるかどうかが決まり、従って検証順序が変わっても問題はない。

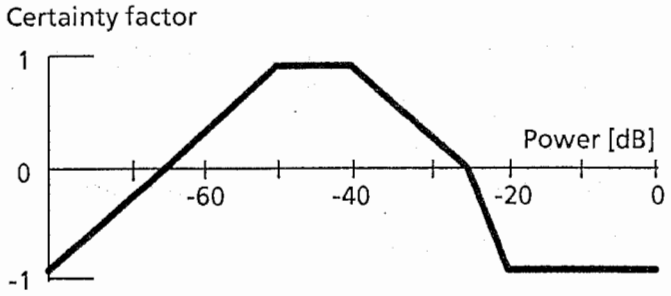


図1　有声摩擦音の0-500Hzパワーのメンバーシップ関数
Fig. 1: Membership function for 0-500Hz power of voiced fricatives.

ことによって、この値をそのまま証拠の確からしさとして、前項で述べた確信度計算を行うことができる[10]。

以上の枠組みによれば、専門家はスペクトログラム特徴の物理量に関する知識を感覚的に表現することができ、閾値を設けることなくスペクトログラム特徴を仮説の確からしさの判断に用いることができる。

## 2.4 スペクトログラム特徴の抽出

スペクトログラム・リーディングにおいて専門家が種々の大域的、局所的なスペクトログラム特徴をうまく捕らえることができるのは、音韻環境、変形の仮説のもとで特徴の存在をあらかじめ予想し、その特徴に応じてスペクトログラム上の注視点、特徴抽出の方法、閾値等を適切に変化させているためである。

本システムにおいても、スペクトログラム特徴の抽出はルールの適用と同時に行う。これによって、抽出すべき特徴および音韻環境の仮説に応じて最も適した特徴抽出方法、パラメータ、例えば周波数域、時間範囲、閾値、平滑化係数などをルール中に記述できる。この結果、専門家が参照する種々の特徴を容易にかつ正確に抽出することができる。

現在、日本語子音のセグメンテーションのために用いている特徴は、帯域パワーの大きさ、帯域パワーがある閾値を越えて変化する時刻、帯域パワー変化率のピークの時刻とその大きさ、帯域パワー比、スペクトル変化量、スペクトラムのピークの周波数とその大きさ、摩擦パワーのカットオフ周波数である。

9　ルール中では関数の各頂点の座標を与え、その間は直線補間される。

10　この関数では純粋にスペクトログラム特徴のもっともらしさを与え、証拠の重要さは上で述べたCOMB条件の重みで与える。これによって、この関数を定義するときには証拠としての重要さを考えなくてもよい。

# 3. 音韻セグメンテーションの方法

現在日本語子音のセグメンテーションのために約150個のルールを記述している[Hatazaki 89]。ルールはほぼ音韻クラス毎に用意されており、図2に示すように(1)音韻候補の検出、(2)音韻環境の仮説、(3)音韻境界の検出とその確からしさの評価、(4)音韻境界の選択の順序で音韻の検出を行う。以下、これらの処理について述べる。

## 3.1 音韻候補の検出

音韻クラス毎に、各音韻クラスを特徴付けるスペクトログラム特徴を参照することによって音韻のおおよその位置を捜し、音韻候補として仮説する。同時に、検出に用いた特徴とさらにいくつかのスペクトログラム特徴からその音韻クラスであることの確信度を計算し、音韻候

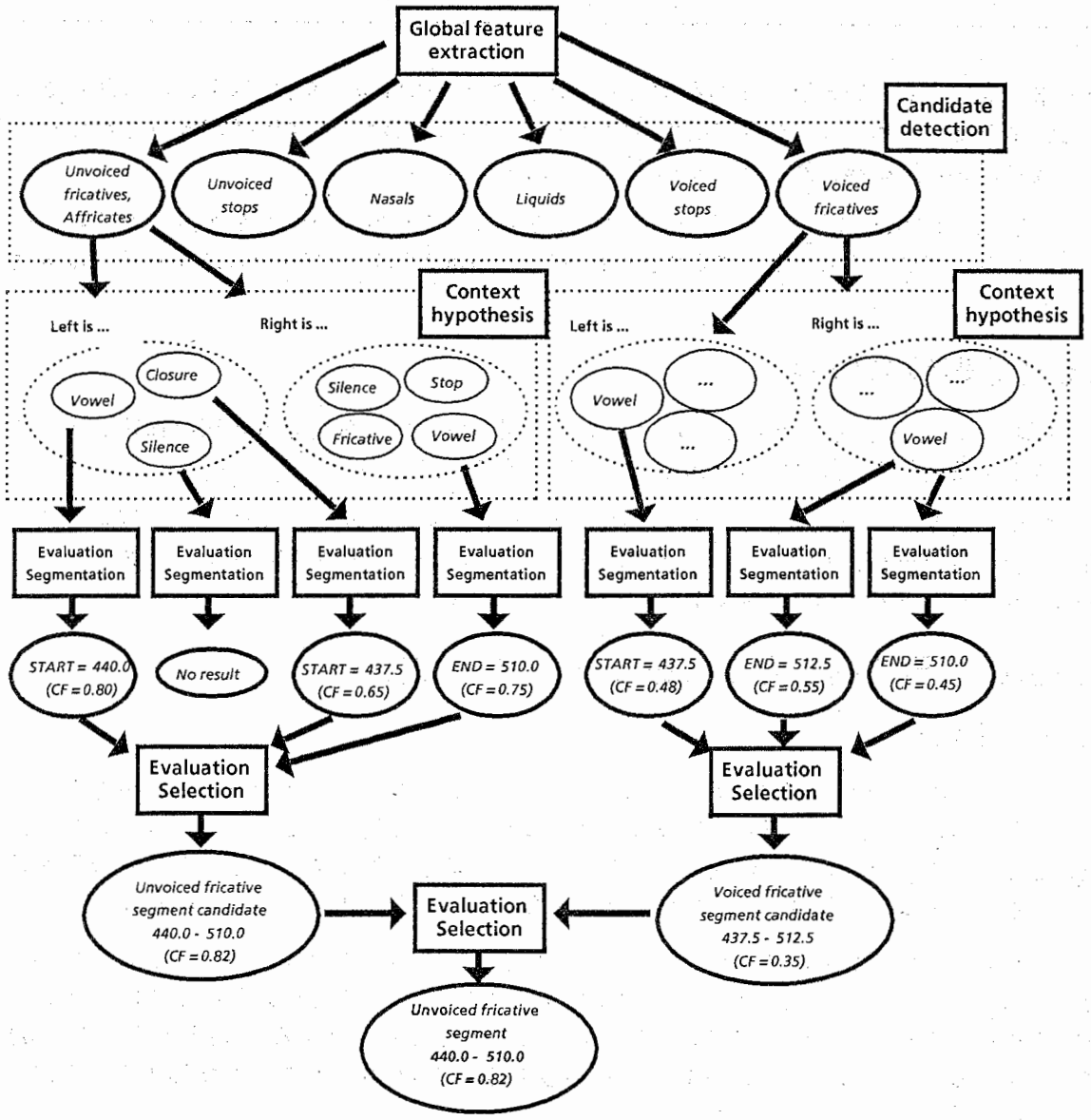


図2 セグメンテーションの戦略

Fig. 2 *Segmentation strategy.*

表1 音韻クラスと候補検出のための特徴

Fig. 1 *Phoneme classes and spectrogram features for extracting the candidates.*

| 音韻クラス | 音韻 | 音韻候補検出のための特徴 |
|---|---|---|
| 無声破裂音 | $p, t, k, ts, ch$ | クロージャ(語中)、破裂(語頭) |
| 有声破裂音 | $b, d, g$ | バズのあるクロージャ、破裂(バズの弱い語頭の破裂音) |
| 無声摩擦音 | $s, sh$ | 摩擦パワー |
| 気音 | $h$ | 低域パワーのディップ |
| 有声摩擦音 | $z$ | 摩擦パワーと弱いバズ |
| 鼻音 | $m, n$ | 高域パワーのディップ |
| 流音 | $r$ | 中高域パワーのディップ |

補の確信度とする。音韻クラスの種類と候補検出のための特徴を表1に示す。この段階では非常に大まかなスペクトログラム特徴での検出を行うため、余計な音韻候補が仮説されたり、同じ位置に複数の異なる音韻クラスの音韻候補が仮説される。

例えば、5000-6000Hzパワーがあってかつ0-500Hzパワーのない区間を無声摩擦音の候補とし、それらのパワーの大きさおよび0-500Hzパワーと1000-2000Hzパワーの比からその確信度を計算する。

## 3.2 音韻環境の仮説

仮説された音韻候補のそれぞれに対して、音韻環境すなわち左右の音韻の種類と典型的な音韻変形を仮説する。仮説する音韻環境と音韻変形を表2に示す。

例えば、無声破裂音に対しては、その左側に母音、無声化母音、語頭を、右側に母音、摩擦音、別の破裂音を仮説し、さらにダブルバーストの有無、バースト部およびアスピレーション部の低域パワーの有無を仮説する。

## 3.3 音韻境界の検出とその確からしさの評価

### 3.3.1 境界検出

それぞれの音韻環境の仮説のもとで、詳細なスペクトログラム特徴を参照して音韻境界の検出を試みる。この結果、正しい仮説のもとでは音韻境界を見つけることができる。誤った仮説のもとでは誤った音韻境界が検出されるか、あるいは推論途中で矛盾を生じ音韻境界を得るに至らない。一つの音韻環境の仮説のもとで複数の音韻境界が得られることもある。

例えば、無声破裂音の右側境界は、母音が後続し、かつアスピレーション部に低域パワーを持たないという仮説のもとでは、図3に示すルールによって次のようにして検出される。(1)0-500Hzパワーが増加する時点を後続母音のおおよその始端とする。(2)母音部分の0-500Hzパワーを得る。(3)そのパワーから計算した閾値を越えて0-500Hzパワーが増加する時点が無声破裂音の終端である。

表2　音韻環境の仮説
Fig.2　*Phoneme context hypotheses*

| 音韻クラス | 左の音韻 | 右の音韻 | 音韻変形 |
|---|---|---|---|
| 無声破裂音 | 母音、<br>無声化母音、<br>語頭 | 母音、<br>無声破裂音、<br>無声摩擦音 | 破裂の有無、<br>破裂時点の低域パワーの有無、<br>アスピレーション部の<br>低域パワーの有無、<br>ダブルバーストの有無、<br>クロージャ部の低域パワーの有無 |
| 有声破裂音 | 母音、無音 | 母音 | 語中か語頭か、<br>バズの有無 |
| 無声摩擦音 | 母音、語頭、<br>無声破裂音 | 母音、<br>無声破裂音 | |
| 気音 | 母音、語頭、<br>無声摩擦音 | 母音、無音、<br>無声摩擦音 | |
| 有声摩擦音 | 母音、語頭 | 母音 | |
| 鼻音 | 母音、語頭 | 母音 | |
| 流音 | 母音 | 母音 | |

### 3.3.2 音韻境界の評価

検出された音韻境界の確からしさを、先に得られている音韻候補仮説の確信度と、音韻環境仮説の確信度の結合として計算する。音韻環境の仮説は明示的に評価される場合と、音韻境界検出の結果として評価される場合とがある。

音韻環境を特徴づけるスペクトログラム特徴がはっきりしている場合には、その特徴を直接検証することで仮説を明示的に評価することができる。例えば、右側の音韻が無声破裂音であるという仮説は、クロージャの存在を検証することによってその確からしさを評価できる。

一方、音韻環境の仮説の証拠となるスペクトログラム特徴を直接検証することが困難である場合には、その仮説のもとでともかく音韻境界を検出する。その後、検出された音韻境界付近のスペクトログラム特徴が仮説されている音韻環境のもとで妥当であるかどうかを評価しこれを音韻境界の確からしさとする。この結果、正しい仮説のもとで検出された音韻境界に対して、高い確信度が与えられることになる。

例えば、無声破裂音がアスピレーション部に低域パワーを持っていないという仮説は、アスピレーション部の低域パワーと後続母音の低域パワーとを区別することができず、直接検証することが困難である。そこで、この仮説のもとで音韻境界を検出するルール(図3)をともかく適用する。この結果、仮説が正しい場合には正しい音韻境界が検出され、かつその境界直後のパワーおよびvoice-onset-time(破裂から後続母音始端までの時間)から計算される確信度は大きな値となる。他方、仮説が間違っている場合、すなわちアスピレーション部に低域パワーがある場合には、ルールはそのパワーの立ち上がりを後続母音の始端すなわち音韻境界と誤認する。しかしながら、その誤った境界の直後のパワーは母音に比べて小さく、またvoice-onset-timeも短く、従って正しい仮説のもとで得られた音韻境界に比べて小さな確信度しか与えられない。この結果、次のステップで正しい音韻境界が選択されることになる。

```
(defrule sg-uvstop-bfr-vowel-3a
        "find a right boundary of an unvoiced stop
        with no aspiration power in low freq,
        and followed by a vowel."
    (declare (salience ?*left-segmentation*))
    (segment-status ?segment segmentation)
    (CF (category ?segment unvoiced-stop)
            ?CFcategory&:mightbe-valid)
?x <- (CF (right-context ?segment vowel)
            ?CFcontext&:mightbe-valid)
    (CF (has-burst ?segment yes) ?CFburst&:mightbe-valid)
    (prop ?segment (burst ?burst-start ?burst-end ?burst-freq))
    (prop ?segment (has-burst-power-in-low-frequency no))
    (prop ?segment (has-aspiration-power-in-low-frequency no))
    (not (applied ?segment sg-uvstop-bfr-vowel-3a))
    (power-increase (?vowel-region&~NONE ?change)
            after = (- ?burst-start 20) 150
            0 500
            ?*normal-smoothing-size* ?*default-min-change*)
    (power-strength
            = ( + ?vowel-region 20) = ( + ?vowel-region 40)
            0 500
            ?voicing-power)
    (CF (vowel 0-500-power ?voicing-power) ?CFvowel)
    (power-start ?vowel-start&~NONE
            before = ( + ?vowel-region 40) 100
            0 500
            = (- ?voicing-power 10))
    (CF (unvoiced-stop voice-onset-time
            = (- ?vowel-start ?burst-start)) ?CFvot)
  =>
    (assert (applied ?segment sg-uvstop-bfr-vowel-3a))
    (retract ?x)
    (assert (prop ?segment (following-vowel-start ?vowel-start)))
    (assert (CF (right-context ?segment vowel)
        = (CFand (CFcomb ?CFcontext
                                (CFweight ?*evidence* ?CFvowel)
                                (CFweight ?*weak-evidence* ?CFvot))
                ?CFvowel
                ?CFvot))))
```

図3　ルールの例

Figure 3. *Rule example.*

## 3.4 音韻境界の選択

音韻境界検出の結果、1つの音韻に対して複数の音韻境界が検出されたり、入力音声の同じ位置に複数の音韻が検出されるので、これらのうちから最も信頼度の高い結果を選び出す。一般にはより大きな確信度が与えられている結果を選ぶ。音韻クラスによっては、音韻継続長などをさらに調べることによって確信度を再計算する。最終的に、音韻クラスとその左右の境界位置が確信度付きで得られる。

# 4. セグメンテーション実験

## 4.1 実験条件

成人男性話者1名が発声した音韻バランス単語セット(216単語)、およびこれとは異なる2,620単語からなる単語セットに対して、これらに含まれる子音のセグメンテーション実験を行った。12kHzサンプリングの音声信号は、2.5msecフレーム周期、5msecハミング窓のFFTによって分析し、ピッチ等による局所的なパワー変動を取り除いた後、パワーの正規化[11]を行う。実験に先立ち、216単語を用いてルールの調整を行った。

## 4.2 結果

セグメンテーション結果を表4に示す。segmentation correctは視察で求めた音韻境界から50msec以内[12]に正しく検出された音韻の割合、alignment errorは正しく検出された音韻境界の誤差の平均値、false alarmsは付加誤りすなわち余分に検出された音韻の割合である。

### 4.2.1 音韻検出率

音韻バランス216単語に対する音韻セグメンテーション率は全子音で96.0%であった。音韻クラスによって82.4%から100%までのばらつきが見られるが、これは一部の音韻についての音響特徴の変形や音韻環境が、ルールにおいていまだ十分に考慮されていないためである。/g/に

11 全帯域パワーについて250フレームのスムージング(移動平均)をした後、最終ピーク時刻より単語終端までを持ち上げ、さらに単語中の全帯域パワーの最大値を-20dB、最小値を-85dBにする。

12 無声破裂音/k/などでは、エキスパート・システムで正しく検出された境界とラベルデータとの差が50msec近くなることもあるので、50msecまでの誤差は正しく検出されたとする。

表4 セグメンテーション結果

Table 4 *Segmentation scores.*

| | Balanced 216 words | | | | 2,620 words | | | |
|---|---|---|---|---|---|---|---|---|
| phoneme category | number of samples | segmentation correct [%] | alignment error [msec] | false alarms [%] | number of samples | segmentation correct [%] | alignment error [msec] | false alarms [%] |
| p | 22 | 95.5 | 4.4 | 16.3 | 28 | 96.4 | 4.2 | 20.8 |
| t | 31 | 96.8 | 2.8 | | 461 | 98.0 | 4.3 | |
| k | 89 | 98.9 | 5.7 | | 1300 | 97.8 | 5.7 | |
| ts | 8 | 100.0 | 5.0 | | 220 | 93.6 | 5.4 | |
| ch | 22 | 90.9 | 6.8 | | 141 | 95.0 | 5.5 | |
| s | 32 | 93.8 | 2.4 | 35.1 | 572 | 89.0 | 3.8 | 39.6 |
| sh | 25 | 100.0 | 2.4 | | 387 | 93.8 | 4.3 | |
| h | 32 | 90.6 | 7.7 | 21.9 | 313 | 88.5 | 8.3 | 17.6 |
| z | 40 | 97.5 | 7.7 | 7.5 | 315 | 87.0 | 9.3 | 9.5 |
| b | 37 | 100.0 | 4.8 | 18.3 | 230 | 92.2 | 5.2 | 21.8 |
| d | 22 | 100.0 | 3.1 | | 177 | 94.4 | 3.7 | |
| g | 34 | 82.4 | 9.2 | | 263 | 70.3 | 8.7 | |
| m | 44 | 100.0 | 6.3 | 52.5 | 485 | 84.9 | 6.8 | 40.8 |
| n | 36 | 100.0 | 5.5 | | 273 | 90.8 | 7.2 | |
| r | 72 | 93.1 | 5.7 | 194.0 | 760 | 84.5 | 6.2 | 120.7 |
| total | 546 | 96.0 | 5.5 | 43.8 | 5925 | 90.8 | 5.8 | 32.8 |

*(MAU, March 1989)*

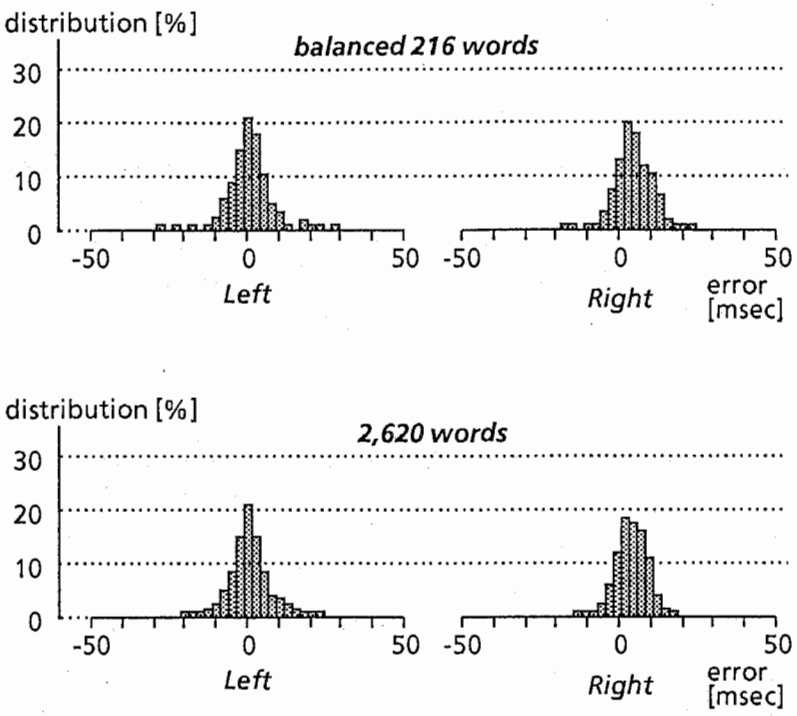


図4 音韻境界誤差の分布

Fig. 4 *Distribution of boundary alignment errors.*

ついては語中の鼻音化したもの(/ng/)の一部[13]が検出できないため、検出率が悪くなっている。これらの音韻を除けば、高いセグメンテーション率を得ることができた。

2,620単語に対するセグメンテーション率は90.8%で、216単語での場合に比べて若干低下している。これはルールの調整に使われた216単語中に現れなかった音韻変形や異音が原因となっている。例えば、無声摩擦音の低域において後続母音の立ち上がりに先行してパワーが現れる、促音のクロージャ前半にバーストに似たパワーが現れる、語頭の無声破裂音のバーストが弱い、などである。

### 4.2.2 音韻境界誤差

視察で求めた音韻境界との誤差の絶対値の平均は5.4ないし5.8msecであった。図4に誤差の分布を示す。誤差の正、負は視察による音韻境界位置に対してそれぞれ右、左にずれていることを表す。両単語セットにおいて、90%以上の音韻境界が視察による境界位置から10msec以内に検出された。視察による音韻境界自身にも8msec以下の誤差があること[Takeda 87]から、本システムで検出された音韻境界は視察によるものと同程度に正確であると言える。

### 4.2.3 付加誤り率

付加誤りの数は、216単語、2,620単語中の全子音数に対してそれぞれ43.8%、32.8%であった。特に流音の付加が多く、母音と撥音あるいは別の母音との間のパワーディップの部分に見

13 /ng/のうち/m/, /n/のようなパワー・ディップのないものが検出できない。

られた。無声破裂音の付加誤りのほとんどは、語頭の母音パワーの急激な立ち上がりをバーストと間違えたものであった。

# 6. むすび

専門家のスペクトログラム·リーディング知識を用いた音韻セグメンテーションの方法と、日本語子音のセグメンテーション実験結果を述べた。

このシステムでは専門家の持つ音韻環境に依存しかつ曖昧な知識と戦略を記述、適用するために、確信度を用いての仮説推論を行い、音韻環境仮説のもとでスペクトログラム特徴の抽出を行う。これによってセグメンテーション知識を自然にかつ容易に記述することができた。

スペクトログラム·リーディングに基づくセグメンテーションの方法の利点は次の通りである。

(1) 音韻環境に依存する方法で境界を見つけるために、調音結合等による音韻の変形に対処できる。

(2) 音韻環境に応じた適切なスペクトログラム特徴を用いることで、セグメンテーションに必要な特徴を効果的に捕らえることができる。

(3) 音韻環境に応じた方法での特徴抽出を行うために、局所的な特徴を正確に抽出でき、この結果、音韻境界を正確に求めることができる。

(4) いくつもの仮説のもとで並列に処理を進め、かつ確信度を用いることで、最終的によりもっともらしい結果を得ることができる。

実験の結果、調音結合等による種々の音響特徴の変形にもかかわらず、また帯域パワーやその変化といった比較的簡単な特徴しか用いていないにもかかわらず、高いセグメンテーション率を達成できた。特に検出された音韻境界は非常に正確であり、これはスペクトログラム·リーディングに基づく方法の一つの優れた点である。

今後は、話者の違いや、発声速度の違いへのルールの対応を検討するとともに、音韻境界の正確さをいかしてニューラルネットワークなどの音韻識別手法との結合による音韻認識システムの構築[Komori 89]を行う予定である。

## 謝辞

研究の機会を与えてくださったATR自動翻訳電話研究所樽松明社長、研究を進める上での助言を頂いた鹿野清宏室長、有意義な討論をして頂いた川端豪主任研究員に感謝します。また、音響処理部の作成に協力いただいた田村震一氏、ルールの作成を手伝っていただいた村山浩一氏、システムのインプリメントに協力いただいた田中孝明氏に感謝します。

## 文献

**[ART]** ART reference Manual. Inference corp. (1987)

**[Buchanan 85]** Bruce G. Buchanan、 Edward H. Shortliffe. Rule-based expert systems. Addison-Wesley Publishing Company (1985)

**[Carbonell 84]** Noelle Carbonell 、 Jean-Paul Damestory 、 Dominique Forh 、 Jean-Paul Haton 、 Francois Lonchamp. APHODEX 、 design and implementation of an acoustic-phonetic decoding expert system. Proc. of *IEEE ICASSP*、 pages 1201-1204 (1986)

**[Clayton 84]** Bruce D. Clayton 、 A first look at viewpoints 、 ART programming tutorial 、 Vol.2、 Inference Corporation (1985)

**[Connolly 86]** J.H. Connolly 、 E.A. Edmonds 、 J.J. Guzy 、 S.R. Johnson 、 A. Woodcock. Automatic speech recognition based on spectrogram reading. Proc. of *IEEE ICASSP*、 pages 611-621 (1986)

**[Hatazaki 89]** 畑崎香一郎、小森康弘. スペクトログラム･リーディングに基づく音韻セグメンテーション知識. ARTテクニカル･レポート、TR-I-00xx、(1989.3)

**[Ishizuka 85]** 石塚満.曖昧な知識の表現と利用.情報処理、Vol.26、No.12、pp.1481-1486 (1985.12)

**[Kleer 87]** J. de Kleer、 An assumption-based TMS、 Artificial Intelligence、 Vol.28、 pp.127-162 (1986)

**[Komori 89]** 小森康弘、畑崎香一郎、田中孝明、川端豪. スペクトログラム･リーディングに基づく音声認識エキスパートシステムの構築. 音響学会講演論文集、pages.91-92 (1989.3)

**[Leung 86]** Hong C. Leung, Victor W. Zue. Visual characterization of speech spectrograms. Proc. of *IEEE ICASSP*, pages. 2751-2754 (1986.4)

**[Mizoguchi 87]** 溝口理一郎、田中康宣、福田尚行、辻野克彦、角所収:"連続音声認識エキスパートシステム-SPREX-"、信学論(D)、J70-D、 6、 pp.1189-1198 (昭62-06)

**[Stern 86]** Paul-Eric Stern 、 Maxine Eskenazi 、 Daniel Memmi. An expert system for speech spectrogram reading. Proc. of *IEEE ICASSP*、 pages 1193-1196 (1986)

**[Takeda 87]** 武田一哉、匂坂芳典、片桐滋、桑原尚夫、音韻ラベルを持つ日本語音声データベースの構築、信学技報SP87-19 (1987)

**[Zadeh]** L.A. Zadeh. Fazzu sets. Inform. Control、 8、 pages 338-353 (1965)

**[Zue 79]** Victor W. Zue 、 Ronald A. Cole. Experiments on spectrogram reading. Proc. of *IEEE ICASSP*、 pages 116-119 (1979)

**[Zue 86]** Victor W. Zue 、 Lori F. Lamel. An expert spectrogram reader: A knowledge-based approach to speech recognition. *IEEE ICASSP*、 pages 1193-1196 (1986)

# 付録

## A. 本研究に関連する発表文献

(1) 畑崎香一郎、田村震一、川端豪、鹿野清宏
連続音声中の音韻認識エキスパートシステムの検討
情処第35回全国大会、pages 1443-1444 (1987.9)

(2) 川端豪、田村震一、畑崎香一郎
日本語スペクトログラム特徴の英語との比較
音響学会講演論文集、pages 109-110 (1987.10)

(3) 川端豪、田村震一、畑崎香一郎
日本語スペクトログラム特徴の英語との比較
信学技報SP87-95 (1987.12)

(4) 畑崎香一郎、田村震一、川端豪、鹿野清宏
スペクトログラム・リーディング知識を用いた音韻認識エキスパートシステム
信学技報SP87-117 (1988.1)

(5) 畑崎香一郎、田村震一、川端豪、鹿野清宏
スペクトログラム・リーディング知識を用いた音韻セグメンテーションの試み
音響学会講演論文集、pages 21-22 (1988.3)

(6) Kaichiro Hatazaki, Shin'ichi Tamura, Takesho Kawabata, Kiyohiro Shikano
Phoneme segmentation by an expert system based on spectrogram reading knowledge
Proc. of *Speech '88, 7th FASE Symposium*, pages 927-934 (19888)

(7) 畑崎香一郎、田村震一、川端豪、鹿野清宏
スペクトログラム・リーディング知識に基づく無声摩擦音の検出
音響学会講演論文集、pages 1-2 (1988.10)

(8) Kaichiro Hatazaki, Yasuhiro Komori, Takesho Kawabata, Kiyohiro Shikano
Phoneme segmentation using spectrogram reading knowledge
Proc. of *IEEE ICASSP 89*, pages 24.S8.2 (1989.5)

(9) 畑崎香一郎、小森康弘
スペクトログラム・リーディング知識による音韻セグメンテーションの評価
音響学会講演論文集、pages 233-234 (1989.3)

(10) 小森康弘、畑崎香一郎、田中孝明
スペクトログラム・リーディング知識に基づく音声認識エキスパートシステムの構築
音響学会講演論文集、pages 91-92 (1989.3)