



# 人間と機械とのコミュニケーション

## ——河原田研究室～知能科学専攻——



河原田 弘 教授

現在、コンピュータは、あらゆる分野で利用されている。しかし、コンピュータの知識がない一般の人々にとって、コンピュータは扱いにくいものだ。だが、コンピュータに入力するときに、人間が情報伝達をするときに使う文字や音声を使うことができるとしたら、抵抗なくコンピュータを使うことができるのではないだろうか。

このように人間と機械とのコミュニケーションを、文字や音声を使って行なうのが「マン・マシンインターフェイス」である。今回訪れた精密工学研究所（総合理工学研究科知能科学専攻担当）の河原田研究室ではこれに関係したパターン認識、特に文字及び音声についての研究を行っている。我々はこの二つの研究についてお話を伺った。



## パターン認識システム

はじめにパターン認識のシステムについて簡単にみてみよう。

認識をするには、まず見本となるパターンをつくり、それと認識しようとする文字や音声を照合し、どの言葉であるかを決定する。この見本となる参照パターンをつくるために認識しようとする対象の特徴を調べる。文字の場合は、特定の見本文字から、音声の場合は多くの音声のサ

ンプルの周波数分析の結果から特徴を取り出す。参照パターンができたなら、いよいよ認識である。認識しようとする文字や、音声から、参照パターンを作る場合と同様の方法で、特徴を取り出す。次に参照パターンと入力した文字や音声を比較し、どれにあてはまるかを調べる。これをパターンマッチングという。このとき近いと思われる候補をいくつか挙げておき、あらかじめ入れておいた単語辞書と比較して単語レベルの認識を行い、さらに文法などの知識をもとに文章レベルでの認識を行って最終的にどの言葉であるかを決定する。また認識実験の結果、誤認識した文字や音声については、参照パターンにおけるその誤認識したものの特徴を強調する。言い換えれば、参照パターンを「学習」によって改善するということである(図1)。それでは文字認識や音声認識における具体的な特徴の取り出し方について見てみよう。

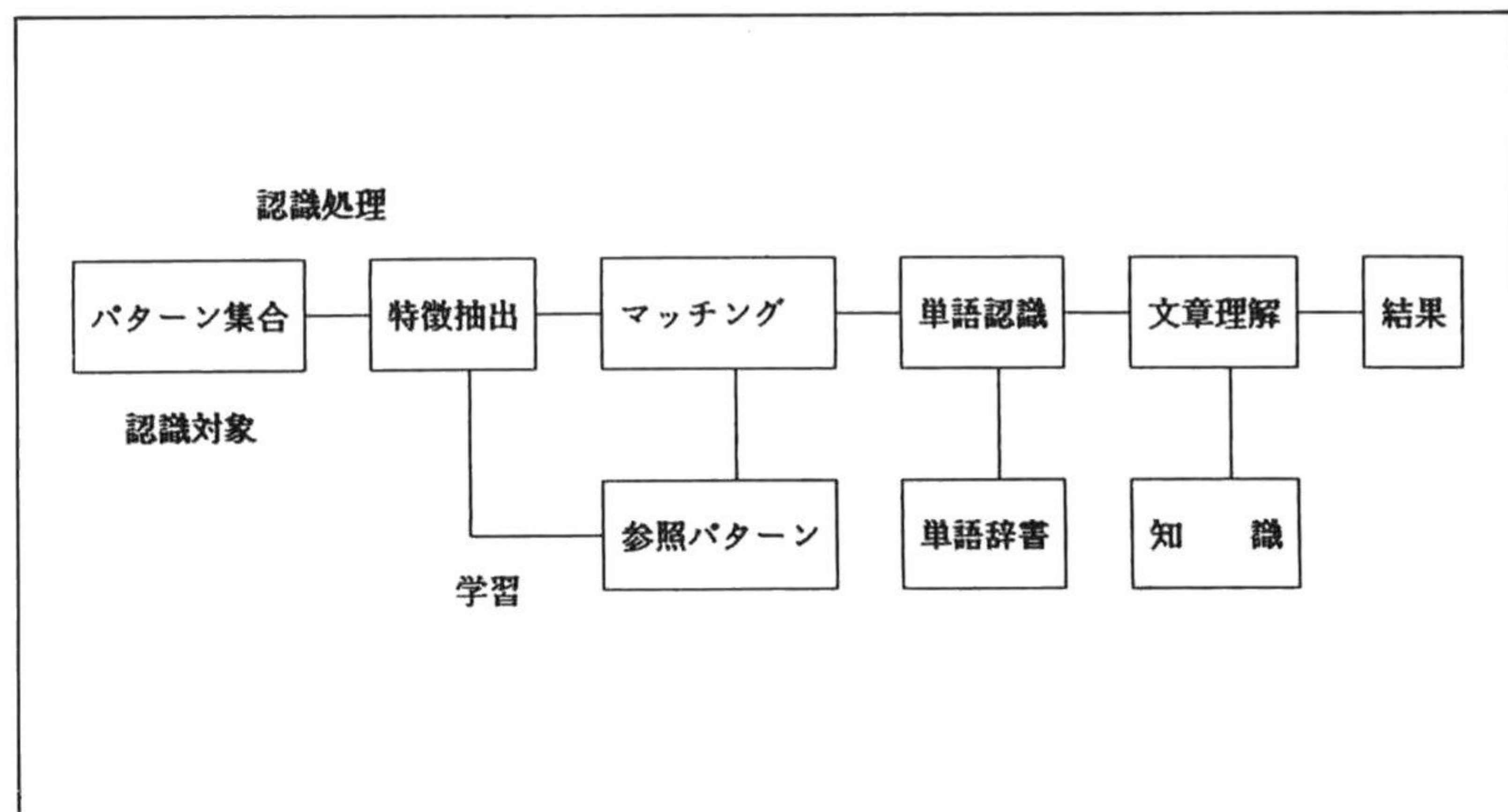


図1 パターン認識システム





## 見本文字から文字の特徴を取り出す

文字認識において文字の特徴を取り出すにはいろいろな方法があるが主な方法を取りあげてみる。

一つは文字の濃度分布を調べてそれをもとに縦、横、斜めの線分の存在場所を調べるものである。この方法では、まず64×64個の画素で文字を表して、そこから、例えば10×10の大きさの枠の分だけ一部を取り出してその内部の黒と白の比を求め、それを濃度とする。この取り出す枠を文字上で少しずつ縦、横、斜めに動かしてみる。例えばこの枠を水平方向に移動させたときに濃度が淡、濃、淡と変化したら、そこには縦の線があるということが分かる。このようにして各方向の線がどの位置に

どの程度あるかを特徴としてとらえるのである。

もう一つは境界線分方向の密度特徴を調べる方法である。これはある枠内に書いた文字を8×8の小領域に分割して、その64の領域それぞれに縦、横、斜めの白黒の微少境界線分線がどの程度あるのかを調べ、その線の数で表すものである。

他に文字上のいくつかの点のまわりのモーメント特徴を調べるものもある。これは、モーメントを調べるときに中心となる点からある点までの距離とその点の線分方向別濃度数との積を、特徴とするものである。

人間が実際に文字を読んで特徴を取り出すときは、各種の特徴を並列

的に調べている。しかし計算機では並列的に調べることができない。そのためいくつかの方法を組み合わせで順番に調べるのだそうだ。

これらの取り出した特徴から参照パターンを作るとき、今まではいろいろな手書き文字の特徴を平均化していたが、これでは各人の癖まで含まれてしまう。そこでこの研究室では見本文字といったものを設定し、そこから文字の特徴を取り出して参照パターンをつくっている。しかし個人差に対応できる参照パターンをつくることは非常に難しいとのことである。



## 視覚的に音声の特徴を取り出す方法

次は音声認識の特徴の取り出し方についてである。

はじめに周波数分析の方法について見てみよう。分析する際には、まず入力音声をローパスフィルタに通し、分析にあまり必要でない高音域部分をカットする。次にこうして取り出したアナログ信号をデジタル信号に変換し、それを周波数分析器に通して、周波数と信号の強さの関係を表すパワースペクトルを、時間を区切って、1.6msまたは0.8msごとに調べる。

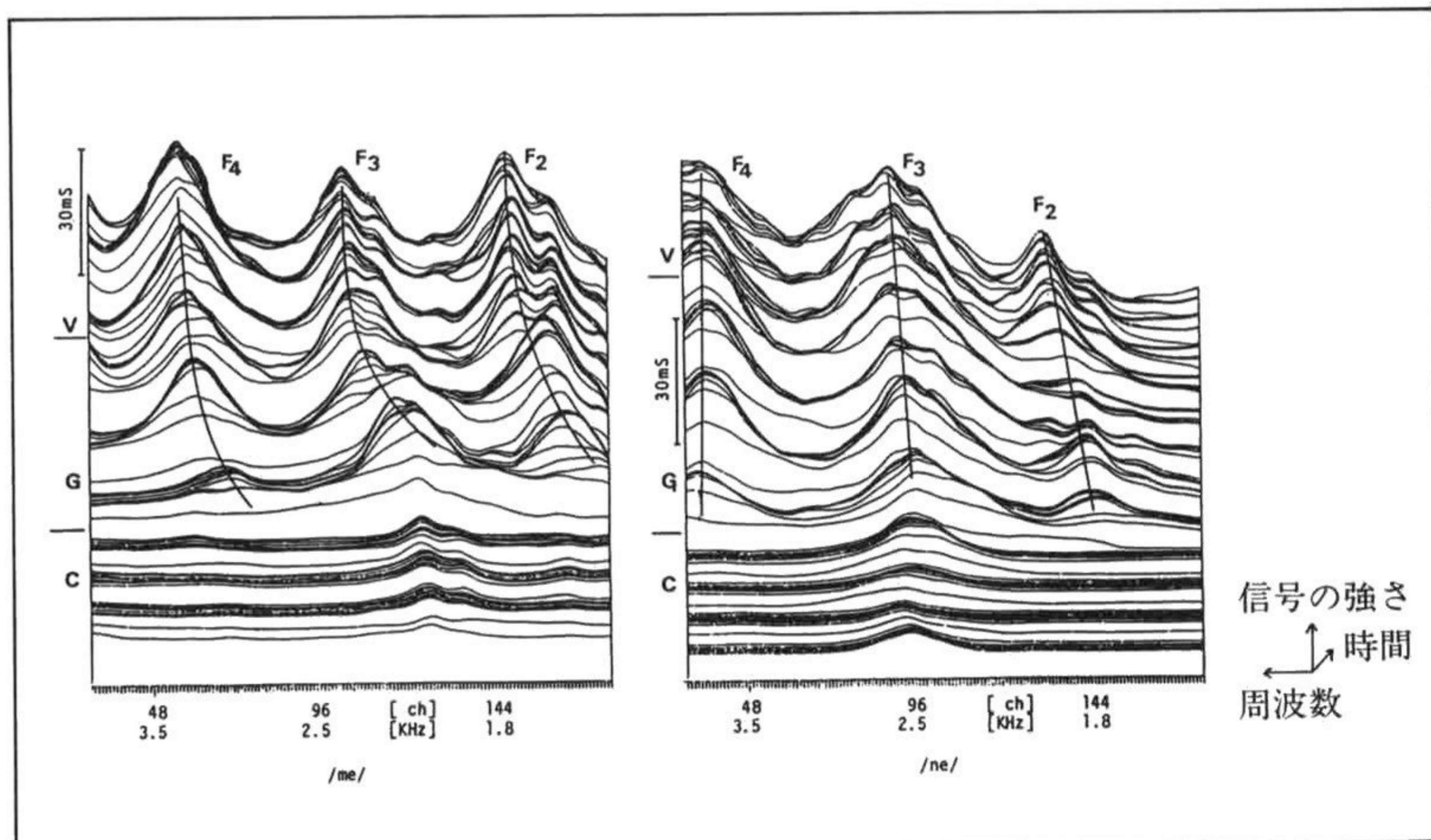
このようにして得られたパワースペクトルの時間変化パターンから特徴を取り出す。音声認識に必要なのはスペクトルの包絡線（スペクトルを滑らかにつないだ曲線）の時間変化である。スペクトル包絡線の山の高くなっている所をホルマントと言い、その周波数は声道（発生時に音波が通る口腔など）の共振周波数である。周波数の低い方から順番に第

1, 2, 3, 4ホルマントと呼ばれ $F_1$ ,  $F_2$ …と表される。普通音声認識には、 $F_1 \sim F_4$ が利用される。

音声には文字のように見本文字といった基本となるものがないため、多くの人から音声データを集めて観察し、特徴がどこにあるのかを調べる。この研究室では視察によって特

徴を取り出した結果から、特徴を表現する方法として、昇次特徴空間法によるものを開発した。ここでは鼻子音である/m/と/n/、特に/me/と/ne/の区別について具体的に説明していただいた。

はじめに視察による特徴の取り出し方から見てみよう。まず図2を見



※C,G,Vはそれぞれ、子音部、わたり、母音部を表す。

図2 /me/と/ne/のスペクトル包絡線の時間変化



てもらいたい。この方法では、スペクトル包絡線の時間変化を調べるのだが、特に子音から母音へと変化する過程である“わたり”の部分のホルマントの時間変化を視覚的に観察する方法である。

先ほど述べたホルマントの時間変

化は、特に  $F_2$ ,  $F_3$ ,  $F_4$  において大きく表れている。 $/me/$  と  $/ne/$  を比べてみると  $/me/$  の場合は、わたりの部分でホルマントが周波数の高い方へ急激に移動しているが、 $/ne/$  の場合はゆるやかにしか移動していないことがわかる。



## 新しい音声特徴の表現の仕方

次に昇次特徴空間法による特徴の表現の仕方について見てみよう。今回は、 $/m/$  と  $/n/$  の区別を  $/me/$  と  $/ne/$  の場合を例に取りあげて説明していただいたが、後続母音 ( $/me/$  の場合は  $/e/$ ) が異なると、ホルマントの移動も異なるため、あらかじめ母音の認識を行った方が認識処理計算が少なく済む。ここでは、まず母音認識への昇次特徴空間法の適応の方法を見てみよう。母音の場合、ホルマントの時間変化はほとんどなく、ホルマント間の相対的な位置関係だけの1次元で表すことができる。各人の個人差を考慮して幅をもたせると二つの母音で重なってしまうところがあるので、そこで昇次特徴空間法を用いてみる。x軸、y軸にまったく同一の母音部のスペクトルをとると2次元的に表され、各母音間の差異

を、はっきりととらえることができる(図3)。このように文字通り次元をあげるのが、この方法の特徴である。このようにして、できたものを参照パターンと呼ぶ。この方法は、子音を含む音にも用いることができる。本来、ある瞬間のスペクトル包絡線は周波数と信号の強さを2本の軸とする、2次元で表すことができる。これをy軸方向にわたりの始点、すなわちホルマントが変化し始める点でのスペクトル包絡線を取り、x軸方向に母音部のスペクトル包絡線をとることによって表すと、その関数はわたりと母音部のスペクトル包絡線の相関行列となり、3次元的に表される。これをわたりにおける参照パターンと呼ぶ。

$/me/$  では  $/ne/$  の場合よりも、わたりにおける  $F_2$ ,  $F_3$ ,  $F_4$  が母音部の対応するホルマントに対して1/6オクターブほど低域側によっている。このように1次元周波数軸上のスペクトル包絡線を2次元周波数平面上のパワー分布として表現する特徴パターンを導入したため、特徴パターンを平均化して参照パターンをつくっても、重要な音の特徴である、ホルマントの周波数の間隔をオクターブ表示で表したときの、ホルマント間の相対的な位置関係が失われないそう。このようにつくられた参照パターンを利用することによって認識実験において100%の認識率が得られたそう(図4-a,b)。

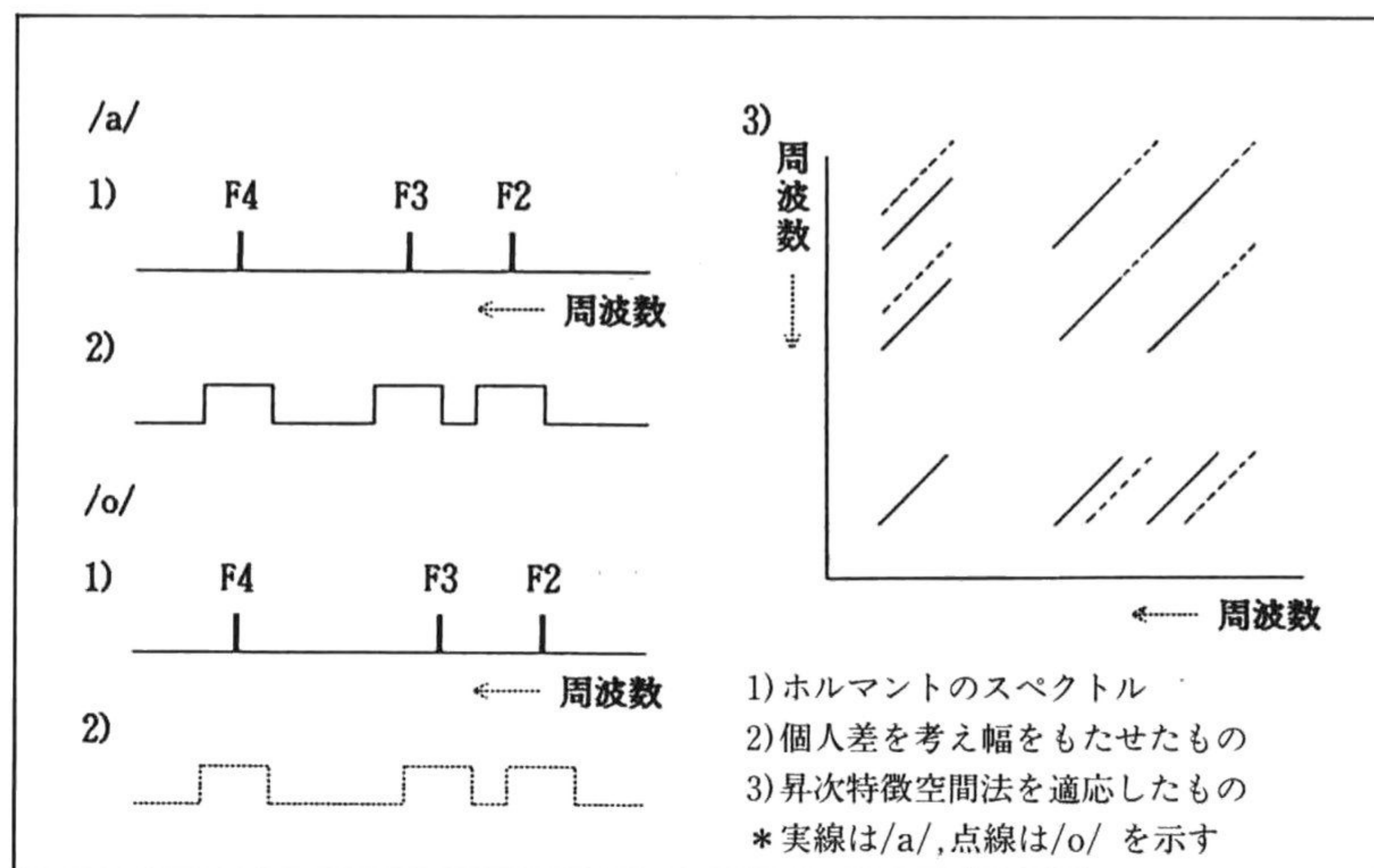


図3 昇次特徴空間法による母音の識別





## 単語の定義と特徴抽出の難点

また認識システムに入れておく連想するための知識の定義の難しさについて「海」という言葉を例にして説明して下さった。

あなただったら「海」をどう定義するだろうか。「地球の表面にあり塩水のたくさんある広い部分」とか定義するとしてみよう。確かにこれは海の定義かもしれないが比喩的な表現として用いるとき、例えば「雲の海」のように「海」という言葉が単に広いところを表している場合は当てはまらない。

このように単語の定義を言語情報で完全に表現することは難しい。ここにも認識システムの難しさがあるとおっしゃっていた。



これまでに、パターン認識システム、特に特徴の取り出し方について見てきたが、ここでそれらの技術的に難しい点を取りあげていただいた。

一つは「融通性」についてである。計算機はもともと計算をするためにつくられたものなので、例えば「斜め」というものを「水平方向と45°の角度をなす方向のもの」とか決めてしまうと、人間は斜めに見えるも

の、大げさに言えば44°や46°のものでも全く違うものとしてはねられてしまう。計算機には、このように融通性がないので、これをカバーするようなソフトを作らなければならないそうだ。

もう一つの難点は「全体を一度に見ることができない」ということである。人間の場合は文字などを全体的な形として、とらえることができるが計算機の場合は一画素ずつ見ていく。これは致命的な欠点なので神経回路網モデルという並列的な処理をする脳の神経細胞のモデルを作って、それで認識処理をしようという研究も進んできているそうだ。

先生はパターン認識システムの実用化について次のように語られた。

「人間が正しく読みとれる文字、正しく聞き取れる音声については認識率100%が目標。しかし、実用化するにあたって、とりあえず認識対象を限定したものができればいいと思っている。例えば、文字認識ならば対象を住所に限定すれば実用化は可能になるだろう。また音声については、地下鉄の切符販売とかに利用できると思う。」

先生はパターン認識のお話の途中で、繰り返し「人間の感覚は素晴らしい」とおっしゃっていた。我々も先生からパターン認識の難しさをお聞きして、改めて人間の能力の素晴らしさを実感した。

最後に先生は学生の学問への取り組み方についてつぎのように語って下さった。

「“百聞は一見にしかず”という言

葉があるが、まったくその通りだと思う。人間の感覚は非常に素晴らしいといったが、その我々の五感すべてを使ってなんでも体験して欲しい。何か問題があったら自分で実験して、自分の感覚を総動員してものを調べる。そうするとものの本質が見えてくる。その感じ方は、百人の人がいれば、百通りの違いがでてくる。それをもとにして何かをやろうとする

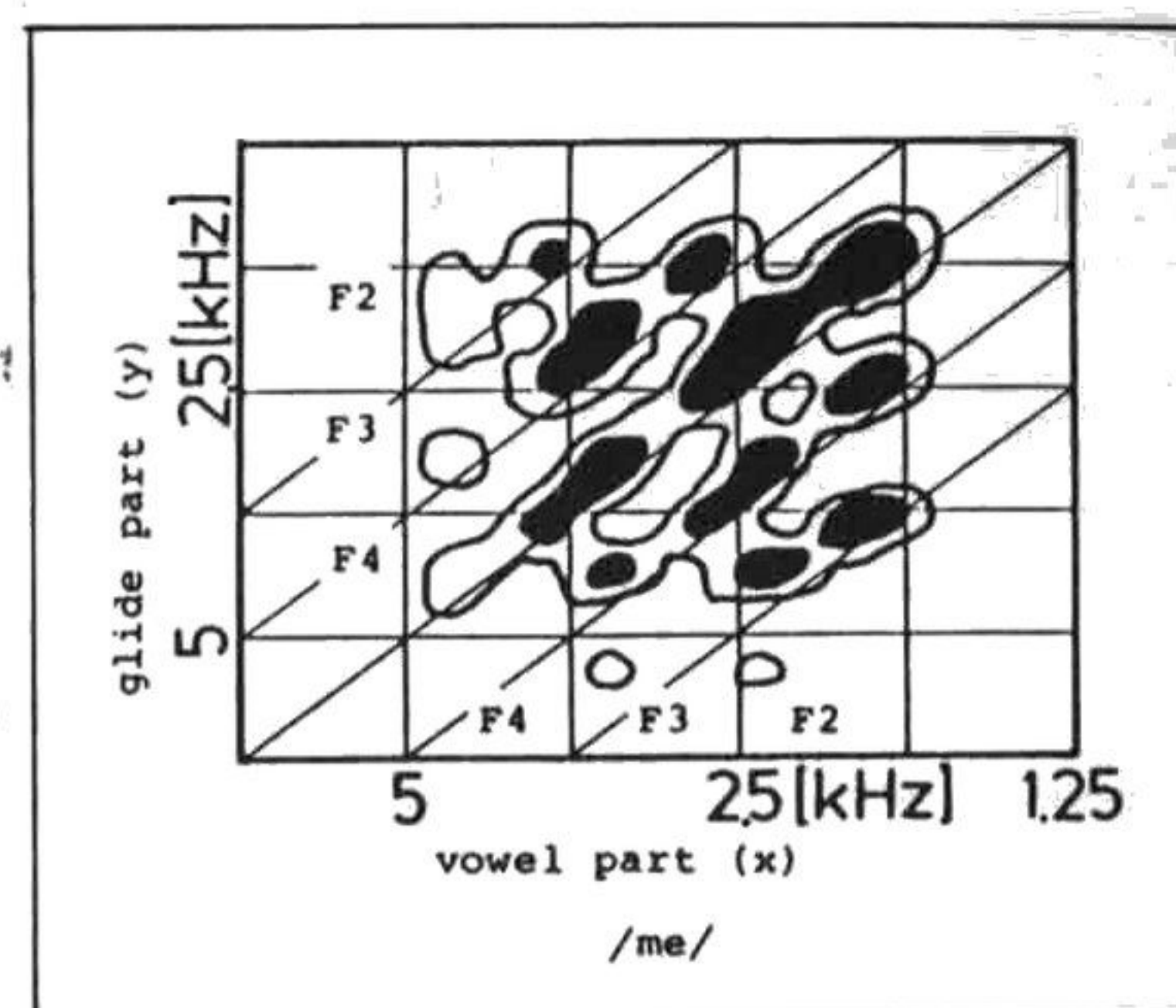


図4-a /me/の参照パターン

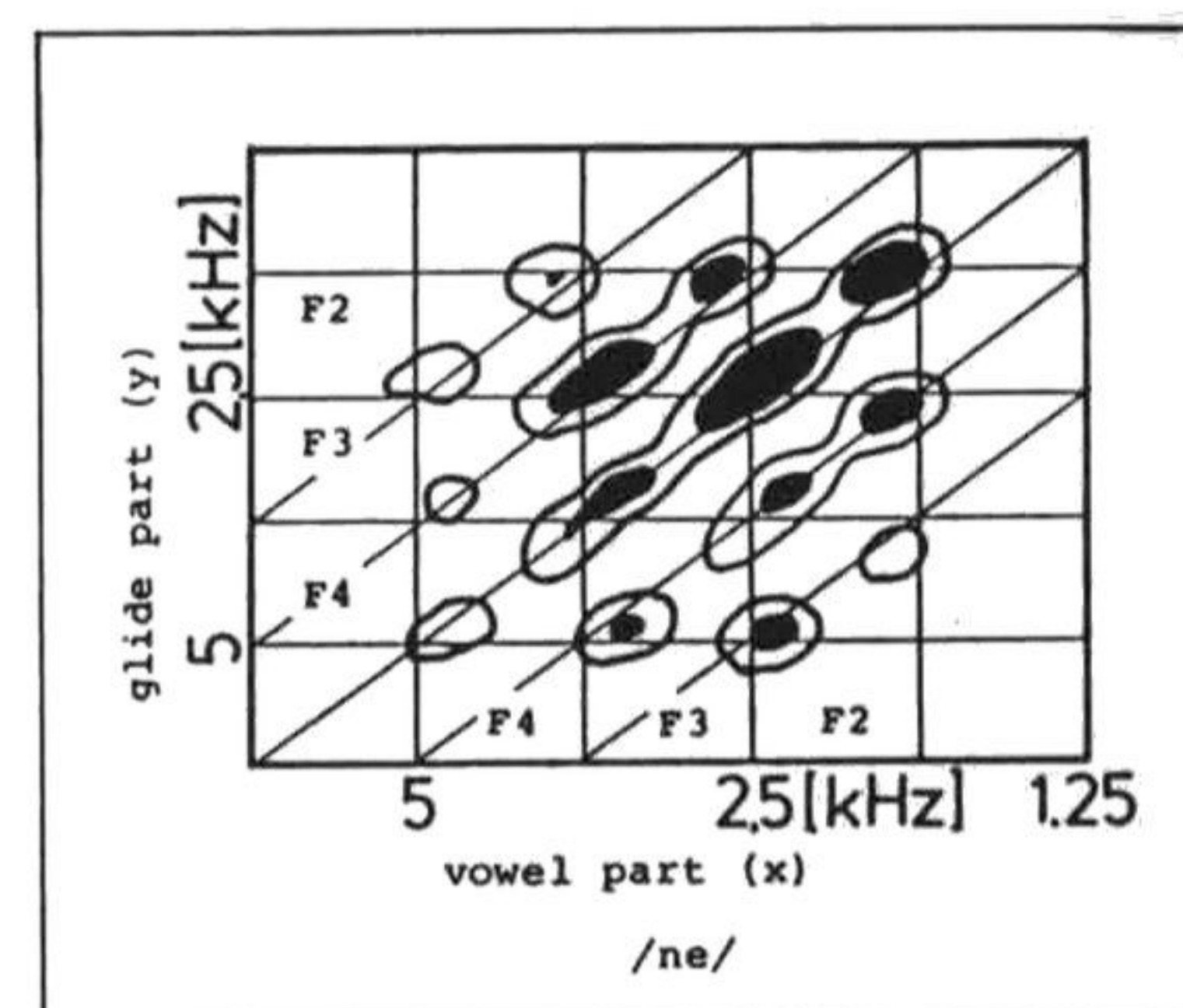


図4-b /ne/の参照パターン

(注) 実線に囲まれた領域は信号の強さの値が最大値の1/4以上、黒く塗りつぶされた領域は1/2以上の部分を表す。

と、そこに独創性が生まれてくる。すでにある理論を学ぶことも必要だが理論をつくることを学んで欲しいですね。」

最後になりましたが、お忙しい中私たちのために時間をさいて取材に協力して下さい河原田教授に深く感謝いたします。

(高橋)