



音波波形情報処理及び認識に関する研究

神内, 教博

(Degree)

博士 (学術)

(Date of Degree)

1983-03-31

(Date of Publication)

2015-11-04

(Resource Type)

doctoral thesis

(Report Number)

甲0397

(URL)

<https://hdl.handle.net/20.500.14094/D1000397>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



音波波形情報処理及び認識に関する研究

昭和57年12月

神 内 教 博

内 容 梗 概

本論文は、著者が神戸大学大学院自然科学研究科システム科学専攻博士課程在学中に行なった「音波波形情報処理及び認識に関する研究」の成果をまとめたもので、本文は次の7章より成っている。

第1章は緒論で、音波波形情報処理、特に音声認識の研究と音楽情報処理の研究の意義と現状について述べるとともに本研究の目的および位置づけを明らかにしている。

第2章では、従来の単語認識・音素認識の手法とその問題点について述べ、それを克服するために、音声波の鼻腔出力と口腔出力を分離ピックアップする方式を提案し、音声波二重入力システムの構成を示している。

第3章では、二重入力システムを用い、計算量の少ない時間領域の計算だけで、孤立発声CV音節を無声破裂音や鼻音等の音素群に大分類する手法と、子音・母音区間を高精度

に分離する手法を示している。

第4章では、摩擦音・破擦音の特徴を表わす各種のパラメータについて検討を加え、その結果、雑音除去差分零交叉数を新しい特徴パラメータとして用いることを提案している。そして、このパラメータも二重入力システムと組み合わせることによって、摩擦音・破擦音が高精度に識別できることを述べている。

第5章では、第3章・第4章で扱った音韻を統一的に識別するためのアルゴリズムを示し、遮蔽板を用いた識別実験と遮蔽板を用いない識別実験を行ない、このアルゴリズムの有効性を確かめている。又、二重入力システムを音声タイプライターとして用いることの有効性と実用性について述べている。

第6章では、オルゴール生産工場における製品の音の検査や曲目別箱詰め工程の自動化を目的として、オルゴール音の分析と符号化、そして曲目の認識実験を行なった結果を述べている。曲目自動認識システムと符号マッチ

ング法を考案し，多数のサンプルを用いた認識実験によってこのシステムが生産工程に導入可能であることを確かめている。

第7章は結論で，二重入力システムによる音声認識およびオルゴールの曲目自動認識システムの有効性・実用性について述べている。又，今後の検討課題にも言及している。

目 次

第1章	緒 論	-----	1
第2章	音声認識	-----	10
2.1	はじめに	-----	10
2.2	単語認識	-----	11
2.3	音素認識	-----	13
2.3.1	破裂音の識別	-----	14
2.3.2	摩擦音の識別	-----	16
2.3.3	鼻音の識別	-----	18
2.4	音声波二重入力システムによる 音声認識	-----	19
2.5	むすび	-----	22
第3章	音声波二重入力システムによる 子音の特徴抽出	-----	23
3.1	はじめに	-----	23
3.2	鼻腔・口腔出力エネルギーの 時間変化	-----	24

3.3	$E_n(t)/E_0(t)$ 時間変化曲線	29
3.4	識別実験	38
3.4.1	遮蔽板を用いた識別実験	38
3.4.2	遮蔽板を用いない識別実験	40
3.5	考察	48
3.6	むすび	50

第4章	口腔出力と鼻腔出力の分離ピック アップ方式による摩擦音・破擦音 の識別	52
4.1	はじめに	52
4.2	摩擦音・破擦音の特徴	54
4.2.1	波形とエネルギー曲線	54
4.2.2	自己相関係数と零交叉数	60
4.2.3	差分零交叉数と 雑音除去差分零交叉数	65
4.3	識別実験	69
4.3.1	遮蔽板を用いた識別実験	71
4.3.2	遮蔽板を用いない識別実験	73
4.4	考察	77

4.5	むすび	80
第5章 音声波二重入力システムによる		
	音韻認識アルゴリズム	82
5.1	はじめに	82
5.2	音韻認識アルゴリズム	83
5.3	識別実験	86
5.3.1	遮蔽板を用いた識別実験	87
5.3.2	遮蔽板を用いない識別実験	88
5.4	実用性についての考察	90
5.5	むすび	93
第6章 オルゴール生産工程における		
	曲目の自動認識	94
6.1	はじめに	94
6.2	曲目自動認識システムの構成	95
6.3	オルゴール音波の特徴	101
6.4	照合と識別実験	106
6.4.1	データの入力法	107
6.4.2	スペクトルマッチング法	109

6.4.3	符号マッチング法	111
6.5	むすび	120
第7章	結 論	122
謝 辞		126
参 考 文 献		128

第1章 緒 論

人間は耳から多くの情報を得ている。人から人へ直接に意志を伝える音声や気分を和らげる音楽は最も代表的なものである。日常生活ではエンジンの音で車の調子を判断したり、スイカをたたいて内部の状態を推測している。これらと同様の理屈で医師は聴診器や打診によって患者の体内の状態を調べている。このように音は情報の伝達と入手の手段として非常に重要なものであり、このゆえに音を測定し分析処理する多くの装置や数学的方法が開発されてきた。マイクロホンは人間の耳で聞きとれない高い周波数の音も測定でき、又、フーリエ変換によって人間の耳以上の周波数分解能が実現されるようになった。しかし、人間は音楽が流れる中で人の声を理解したり、音声中の言語としての意味だけでなく話者の感情や話者が誰であるかも即座に理解する等の高度な処理能力をもっている。機械に人間

の耳をみたせるといふ長年の夢は機械にこれらの処理能力をみたせることができた時に実現される。この夢の実現を目指してあらゆる角度から研究が進められてきているが、この中でも音声認識の研究と音楽情報処理の研究が近年活発に行なわれている。歴史上、本格的な音声の研究は Dudley による vocoder⁽¹⁾ [1939年] と Potter による visible speech⁽²⁾ [1947年] の研究に始まり、周波数スペクトルのエネルギーが集中しているホルマントの重要性が明らかにされた。次に、Fant による音響的音声生成理論⁽³⁾ [1960年] と Stevens らによる analysis-by-synthesis^{(4),(5)} [1961年] が発表され音声生成理論の基礎が確立し、それに基づいた音声合成の研究が着実に進歩した。デジタル計算機の急速な発展により音声の研究もソフトウェアを中心に移り、Cooley と Tukey による FFT⁽⁶⁾ [1965年] はスペクトルの計算に活用されている。これらの成果の帰結として板倉・斉藤による最尤スペクトル推定法⁽⁷⁾ [1966年] と Atal らによる線

形予測法⁽⁶⁾ [1967年] が発表され、更に板倉・
斎藤は PARCOR 音声分析合成法⁽⁷⁾ [1969年] を開
発した。これらの手法と DP マッチング⁽¹⁰⁾ は
音声合成と音声認識の研究を飛躍的に進歩さ
せ現在の研究の主流となった。このような各
種の手法を用いて音声認識の研究が進められ
たが、音素の認識は調音結合の影響等によっ
て困難であることがわかり、比較的長い発
声で情報量が多い単語を認識単位とした研究
が始められ、音素認識の研究より一歩進んだ
状態となった。単語認識システムは数多く試
作され、一部では実用化が進められている。
しかし、単語認識システムでは単語数が増加
するにつれパターンマッチングの回数が増加
するため認識時間の増加や認識率の低下が問
題となり、その上標準パターンを格納するた
めの大容量のメモリーが必要となってくる。
この問題を克服するためには音素単位の認識
方法を用いることが不可欠であり、この方面
の研究が活発に行なわれ始めた。音素認識の

研究にはスペクトルの極の起点周波数と極の時間変化と特徴パラメータに用いる方法や極零周波数を用いる方法等がある。又、声道断面積関数と音声波スペクトルとの対応についても研究がなされている。これら周波数領域での認識方法では音声波の少なくとも数個の部分においてスペクトルを求め計算が必要であり膨大な計算をしなければならぬ。そして、計算量が多いためかかわらず、特定話者の1000語の単語認識・2~3種の音素認識ともに90%程度の認識率にとどまっている現状である。一方、音楽情報処理の研究分野では楽器音の分析を目的とした研究や楽音の自動採譜を目的とした研究等が活発に行なわれている。これらの研究では音の周波数とその時間変化とFFTによって分析する方法を多く用いている。楽器音の分析は種々の楽器について行なわれ電子楽器の製造等に應用されている。又、楽音の自動採譜の試みも幾つか行なわれている。しかし、現在の自動採譜の研究

においては和音が存在する楽音の基本周波数の検出が困難であり，いまだ実用的な自動採譜システムは試作されていない。

本研究は，このような音声認識と音楽情報処理の研究の現状において計算量が少なく，しかも高精度の新しいパターン認識アルゴリズムを開発することも目標として行なわれたものである。本論文では，2個の小型マイクロホンによって音声波の鼻腔出力と口腔出力を別々に，しかも同時に計算機に入力するシステムを用いた音素認識実験と，オルゴール生産工場における製品の検査と曲目別箱詰め工程を自動化することを目的とした曲目認識実験の2つの新しいパターン認識の手法について述べる。第2章では従来の単語認識・音素認識の手法とその問題点について述べ，それを克服するために考案された音声波の鼻腔出力と口腔出力を分離ピックアップするためのシステム（二重入力システム）の構成を示す。従来の研究では無声破裂音・鼻音等のグ

ループ内での認識実験が多く行われてきたが、日本語68種の単音節をそれらのグループに大分類する研究は極めて少なく、この方面の研究が必要とされている。そこで、第3章では二重入力システムを用いて計算量の少ない時間領域の計算だけで、孤立発声CV音節を無声破裂音・鼻音等のグループに大分類する手法を述べ、遮蔽板を用いた識別実験と遮蔽板を用いない識別実験を行ない、その有効性を確かめる。次に、従来の研究でいまだ多くの困難な問題が残されている子音区間の検出法に対して、孤立発声CV音節の子音・母音区間を高精度に分離する新しい手法を示す。第4章では、鼻腔出力と口腔出力の分離ピックアップ方式による摩擦音・破擦音の新しい認識手法について述べる。従来、摩擦音・破擦音は波形の零交叉数やスペクトルの重心周波数・ピーク周波数・谷周波数を特徴パラメータとして認識実験がなされてきた。しかし、計算量が膨大であるにもかかわらず安定で高

い認識率を得ることが困難であることがわかってきた。この章では、計算量が少なく、しかも高認識率を得るための手法として、鼻腔出力エネルギーの立上り時刻と口腔出力乱気流の程度を表わす新しいパラメータである雑音除去差分零交差数 σ の立上り時刻との時間差を第1の特徴パラメータに用い、又、雑音除去差分零交叉数の最大値を第2の特徴パラメータに用いた摩擦音・破擦音の認識手法を提案する。そして、遮蔽板を用いた識別実験と遮蔽板を用いない識別実験を行ない、その有効性を確かめる。第5章では第3章・第4章で扱った音韻を統一的に識別するためのアルゴリズムについて述べる。そして、後続母音に /a/ をもつ16種の音韻および /N/ を9群に分離識別する実験を遮蔽板を用いる場合と遮蔽板を用いない場合について行ない、その有効性を確かめる。又、二重入カシステムを音声タイプライターとして用いることの有効性と実用性について述べる。第6章では、オル

ゴール生産工場における製品の音の検査や曲目別箱詰め工程の自動化を目的として、オルゴール音の分析と符号化、そして曲目の認識実験も行なった結果について述べる。これは近年活発に行なわれている計算機による音楽情報処理の一研究である。オルゴール音はぜんまいの力の変化によって曲の演奏速度が変わるといった特異な性質ももっている。この問題を解決するため、ドラムの回転角を測定する回路をもつシステムを考案し時間軸方向の正規化を行なった。次に、このシステムを用いた曲目認識のためのマッチング法として符号マッチング法を考案した。現在まで曲目認識実験を行なった研究は極めて少なく、この符号マッチング法はオルゴール音のみならず一般の曲目認識の一手法として重要であると考えられる。最後に18種の曲目と10個の不良品を用いた認識実験を行ない、このシステムが生産工程に導入可能であることを確かめる。第7章では二重入力システムによる音声認識

とオルゴールの自動曲目認識システムの有効性と実用性について述べる。又、今後の検討課題に言及し、結論とする。

第2章 音声認識

2.1 はじめに

音声認識の研究は計算機の飛躍的發展にもないソフトウェアを中心に行なわれている。音声認識は音素認識・単語認識・連続音声認識に分類される。音素認識は調音結合の影響のため高認識率を得ることが困難であることがわかり、実用化は単語認識システムから始められた。連続音声認識は原理的には音素認識や単語認識に構文・意味情報を加えたものであり、音響分析は音素や単語レベルで行なわれている。そこで、2.2節では最初の実用化が行なわれた単語認識システムの構成とその問題点について述べ、次に2.3節において、記憶容量や認識時間の点で実用化が望まれている音素認識システムの音響分析の手法について述べる。最後に2.4節において、従来の音素認識の問題点を克服するために、鼻腔出力と口腔出力の分離ピックアップ方式による

音素認識方法を採用することの意義を述べ、
これを実現するためのシステムの構成を示す。

2.2 単語認識

音声認識の研究は音声タイプライターのような一般的なものではなく、単語認識システムという極めて限定された形で最初に実用化の段階を迎えた。実用化されている単語認識システムでは限られた数十単語を特定話者が単語ごとに区切って発声する必要があるが、例えば荷物の仕分けを音声で命令する作業では十分に実用的である。図2-1は単語標準パターンを用いる単語認識システム^{(11),(12)}の構成を示している。このシステムでは入力単語音声は音響分析部で10~20 msec毎に短時間スレ

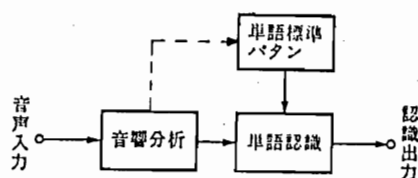


図2-1 単語標準パターンを用いる単語認識

クトルもしくはそれに等価な自己相関関数などに変換され特徴パラメータの時系列を生成する。一方、単語標準パターンもあらかじめ同様の音響分析を行ない特徴パラメータの時系列としてメモリーに格納されている。単語認識部では未知入力単語音声と単語標準パターンの特徴パラメータの時系列間のマッチングを行ない距離を計算する。そして、最短距離の単語を認識結果として出力する。図2-2は音素標準パターンを用いる単語認識システム^{(13), (14)}の構成を示している。このシステムでは音響分析部の出力から音素記号の時系列を作成し、単語辞書における音素系列との類似度を計算する。そして、類似度が最大である単語を認識結果として出力する。

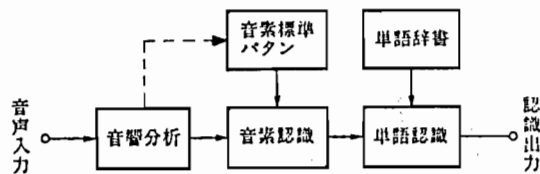


図2-2 音素標準パターンを用いる単語認識

図 2-1 のシステムでは単語数が増加するにつれパターンマッチングの回数が増加するため認識時間の増加や認識率の低下が問題となり、その上標準パターンを格納するための大容量のメモリーが必要となる。一方、図 2-2 のシステムでは単語数が増えても音素の数には上限があり、この問題は解決される。しかし、単語中の音素の認識は調音結合の影響のため非常に困難であり、音素認識部での精度が悪く全体の認識率も不十分なものにしている。このような理由で、現在市販されている音声認識装置の大部分は図 2-1 の構成をとっている。

2.3 音素認識

前節で述べたように単語認識システムは認識率・認識時間・記憶容量の点で数十単語が限度であり、音声タイプライターの実用化のためには音素認識システムの開発が不可欠である。現在、破裂音・摩擦音・鼻音等の音素

認識の研究が活発に行なわれているが、ほとんどすべての研究では有声破裂音・無声破裂音・無声摩擦音・鼻音等の各音素群内での識別に限定した実験を行なっており、他群との識別を試みた研究は極めて少ない。又、識別率についても各群内の2-3種の音素の識別で90%程度にとどまっておられ、実用化には不十分な結果となっている現状である。本節では従来の音素認識の手法を破裂音・摩擦音・鼻音の各々について述べる。

2.3.1 破裂音の識別

破裂音の識別の研究は1950年代から現在に到るまで盛んに行なわれている。破裂音は唇等が急激に開放する際のステップ状圧力変化によって生ずる。有声破裂音 /g/, /d/, /b/ の自然音声から切り出した部分的音声の聞き取り実験によって、有声破裂音の個々の情報は破裂時点から20msecの短い区間内に多く含まれていることが確かめられている⁽¹⁵⁾。そこで、

正確な識別のためには破裂時点の検出を高い精度で行なう必要があるが、エネルギー・零交叉数・自己相関係数等の値の急激な変化点を破裂時点として検出する方法⁽¹⁶⁾が提案されている。周波数領域の研

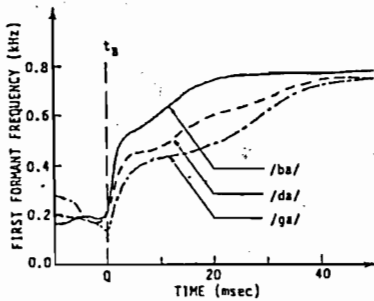


図2-3 有声破裂音の第1ホルマント周波数の遷移パターン

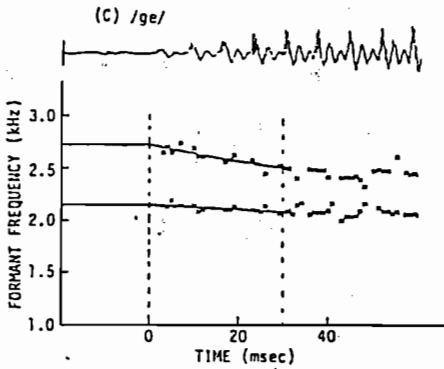
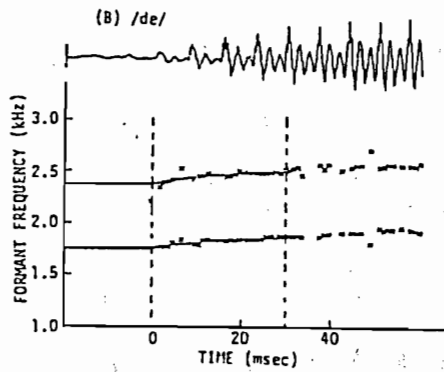
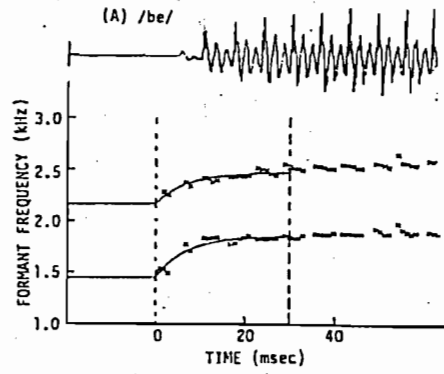


図2-4 有声破裂音の第2・3ホルマント周波数の起点と遷移パターン

究では第1ホルマント周波数の遷移パターンは図2-3のような特徴をもち第2・第3ホルマント周波数の起点と遷移パターンは図2-4のような特徴をもちことが確かめられている⁽¹⁷⁾。一方、無声破裂音 /p/, /t/, /k/ については破裂部スペクトルの中心周波数により識別できることが合成音の知覚実験によって確かめられている⁽¹⁸⁾。無声破裂音の識別の特徴パラメータとしてスペクトルモーメントとその帯域幅を用いるもの⁽¹⁹⁾、3極線形予測法による極の周波数と帯域幅を用いるもの⁽²⁰⁾、スペクトルの極零周波数を用いるもの⁽²¹⁾、調音パラメータを用いるもの⁽²²⁾等が提案されている。

2.3.2 摩擦音の識別

無声摩擦音は声帯の非振動時に口腔内を通過する気流が歯や舌、唇の狭めによって乱気流となることにより発生する。このため、音声波形は不規則な細かい振動をす。従って、基本的には零交叉数が大きく、波形の周期性

の強弱を表わすパラメータである自己相関係数は小さくなり、このことにより無声摩擦音と他の音素群との判別が可能であると考えられる。しかし、/h/の中には必ずしもこの特徴を示さないものがあり、一般的な有声/無声の判定が難しい現状である。一方、有声摩擦音は声帯振動波が歯や舌、唇の狭めを通過するときに発生する。このため、声帯振動波に細かい振動が重ね合わされた波形となり、零交叉数や自己相関係数では他群との判別が困難である。現在、摩擦音と他群との判別について定量的実験を行なった研究は極めて少なく、スペクトルピークのパワーの増加傾向から無声摩擦音と無声破裂音・破擦音の判別を行なった研究⁽²³⁾のように限定された条件下の例があるだけである。無声摩擦音 /s/, /h/, /ʃ/ の群内での識別実験がスペクトルの重心周波数・ピーク周波数・谷周波数を用いて行なわれているが、/s/ と /ʃ/ の判別が難しいという問題が残されている。

2.3.3 鼻音の識別

鼻音は /m/, /n/, /ŋ/ からなる鼻子音と母音が鼻音化した鼻音化母音に分類される。鼻音の発声時には口蓋帆が下がり、咽頭から鼻腔への通路が生じ鼻孔から音声波が出力される。鼻音の特徴抽出の研究は周波数領域中心に進められ、特徴パラメータとしてスペクトルピークの起点周波数と遷移パターンを用いるもの⁽²⁴⁾や極零周波数を用いるもの⁽²⁵⁾等がある。又、鼻腔に対応する分岐管を考慮した声道断面積関数も音声波から推定し、調音器管の位置や形状から鼻音を識別する試みもなされている⁽²⁶⁾。この中には鼻腔・口腔出力を分離録音した音声波のスペクトルも分析することによって鼻腔の形状が鼻音に及ぼす影響を調べた研究⁽²⁷⁾もある。しかし、これらの方法は一般に計算量が多いにもかかわらず十分な精度の識別率が得られていない。

2.4 音声波二重入力システムによる音声認識

従来の単語認識・音素認識の手法とその問題点について述べてきたが、認識率・認識時間・記憶容量と ρ 限界に達し不満足な結果に終わっている。そこで、これらを克服する方法として音声波の鼻腔出力と口腔出力を分離ピックアップすることによって音素を認識することも考える。有声音発声時には声帯振動波は口蓋帆によって鼻腔と口腔に分割され、鼻腔・口腔それぞれの伝達特性によって修飾を受け、鼻孔と唇から出力される。この時、鼻腔形状は変化しないが口蓋帆は動き、又、口腔形状は歯や舌等によって変形し唇は開閉する。一方、無声音発声時には肺からの気流が口腔内で破裂気流や乱気流となって唇から出力される。このような音声の生成過程を考慮するならば音声波の鼻腔出力と口腔出力を分離して観測することは音素の認識を極めて容易にすると予想できる。そこで、これを実現するためのシステムを音声波二重入力シス

テムと呼び、本節でその構成を示す。

図2-5に示すように鼻と口の間には遮蔽板 (shield plate) を置き、鼻腔 (nasal cavity) 出力をマイクロホン (ECM-150, SONY製, 以下マイクと略す) Nに、口腔 (oral cavity) 出力をマイク Mに入カする。マルチプレクサは20kHzでスイッチを切り換え、N・Mの音声波はA/Dコンバータを通して8bitで計算機 (LSI-11) に入カされ、ソフトウェアによって両出力音声波を分離する。遮蔽板は厚さ1cm、幅85cm、長さ70cmの合成樹脂の板を用い、N・Mは直

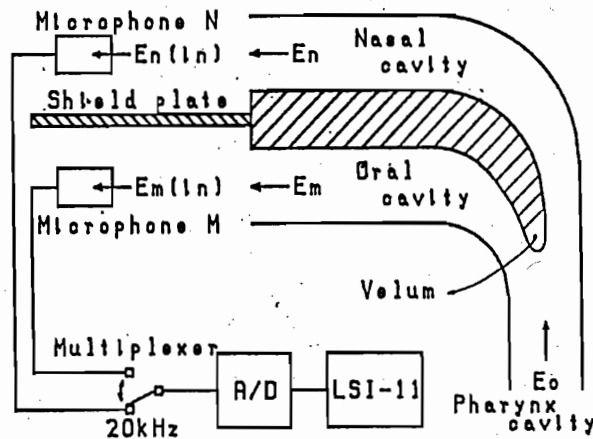


図2-5 二重入力システムの構成

径 8.5 mm, 長さ 15 mm のコンデンサーマイクを使用した。N は鼻から 1.5 cm, M は口から 3 cm の位置に遮蔽板の振動がマイクに伝わらないように固定した。この二重入カシシステムは声帯振動時の音声波のみならず, /p/, /t/, /k/ の破裂気流や /h/ の乱気流をも非常に感度良く検出することによって音素を識別することを目的としており, 従って N・M は鼻と口の直前に配置している。次に, この装置を用いて遮蔽の程度を調べる実験を行なった。先ず, 口を閉じて /N/ を発声し, 発声中に N・M が観測する音声波エネルギーの比を計算する。逆に鼻を閉ざして /a/ を発声し, 同様の計算をする。これらの値は各々鼻側から口側へのエネルギーの減衰と, 口側から鼻側への減衰の程度を表わすと考えられる。各 20 回の発声で実験を行なった結果, 鼻側から口側への減衰は -23 dB 以下, 口側から鼻側への減衰は -20 dB 以下であった。以上のことから以下の実験では音声波の漏れを無視する。音声波は計算

機室内で入力し，特別に大きな騒音源が存在しない平靜な状態において孤立発声されたC V音節の音声試料を集めた。

2.5 む す び

この章では，現在最も研究が進み一部で実用化されている単語認識システムの構成とその限界について述べ，次にそれを解決するために実用化が望まれている音素認識システムの研究の現状と問題点を述べた。それらの問題を解決する方法として，鼻腔出力と口腔出力の分離ピックアップ方式を採用することの意義を述べ，二重入力システムの構成を示した。

第3章 音声波二重入力システムによる 子音の特徴抽出^{(28), (29)}

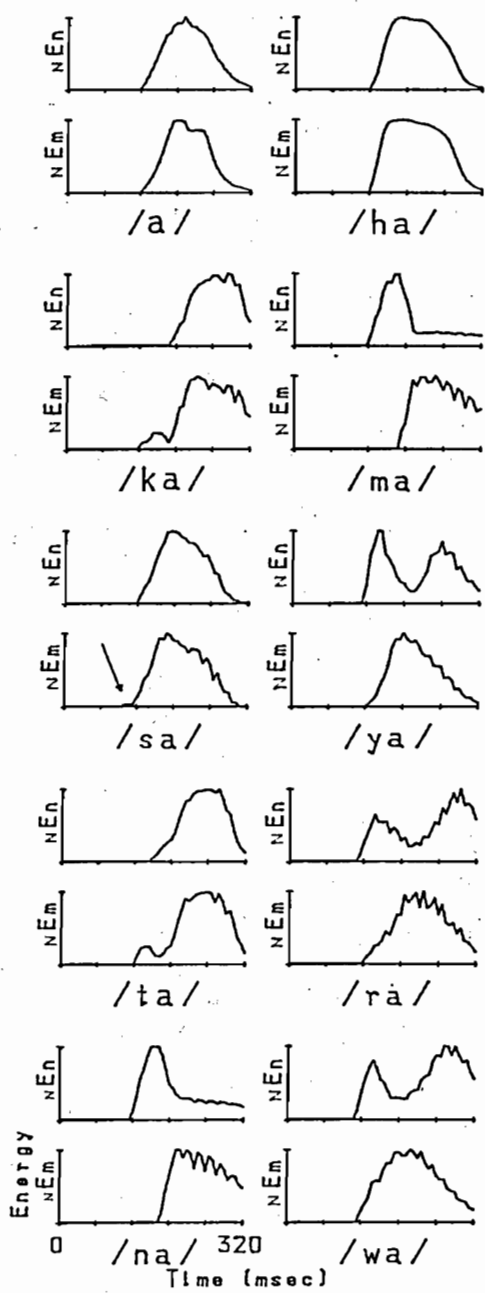
3.1 はじめに

音声認識において音素の認識は困難なものの一つである。音素認識の研究は周波数領域におけるパラメータを中心に進められている。ところで、周波数領域での識別方法では音声波の少なくとも数カ所の部分でスペクトルを求める計算が必要であり、高精度の結果を得るためには膨大な計算をしなければならない。そこで、周波数領域を問題とする前に、計算量の少ない時間領域において、子音部と母音部を分離するための情報や子音を大まかに分類するための情報が得られれば計算量を減らすことができ、その上、それらの情報が高信頼性ののであればスペクトルマッチングにおける参照の数を減らすことができ、認識率の向上が期待できる。この章では二重入力システムを用いて、時間領域の計算だけで子音

の特徴を抽出する新しい手法を示す。後続母音に /a/ をもつ子音ではこの特徴が特に顕著であるので、この章ではこの手法をこれらの子音11種と撥音 /N/ に適用した結果について述べる。最初の遮蔽板を用いた実験ではこれら12種の音韻を6群に分類する方法を示す。次に遮蔽板を用いない場合の補正法を検討し識別実験を行なう。その結果から二重入カシシステムの実用性が裏付けされる。

3.2 鼻腔・口腔出力エネルギーの時間変化

図3-1は、1名の男性が孤立発声した、後続母音に /a/ をもつ子音11種と /N/ についてマイク N・M が観測した音声波エネルギーを各々最大値で正規化したエネルギーの時間変化曲線である。以下ではこれをエネルギー曲線、又は、Normalized energy の意味で ${}_nE_n(t)$ 、 ${}_nE_m(t)$ と呼ぶが、特に必要のない場合には (t) を省略する。又、このエネルギー曲線は破裂音における破裂気流の観測値も含んでいる。



とこで、音声波エネルギーを計算する場合、分析区間が短い程ピッチの影響によってエネルギー曲線にリップルが現れ易くなる。逆に分析区間が長い程エネルギーの細かい変化の情報が失われ、この点を考慮して、フレーム周期

図3-1 鼻腔・口腔出力エネルギーの正規化時間変化曲線

6.4 msec, 分析区間 25.6 msec でエネルギーを計算した。従って, l をフレーム番号として $t = l \times 6.4$ (msec) の関係がある。

次に, 図3-1のエネルギー曲線の特徴について述べる。/a/, /ha/ では ${}_{N}E_n$ と ${}_{N}E_m$ が同時に立上がり, 発声中の ${}_{N}E_n$ と ${}_{N}E_m$ は同じ変化をしている。/ka/, /ta/, /pa/ では ${}_{N}E_m$ の立上り時に破裂気流によるピークが現われ, ${}_{N}E_m$ は ${}_{N}E_n$ より早く立上がっている。/sa/ では /s/ の区間で ${}_{N}E_m$ に小さな値(矢印)が現われている。/na/, /ma/ では ${}_{N}E_n$ が ${}_{N}E_m$ より早く立上がり, ${}_{N}E_m$ が増加を始めると同時に ${}_{N}E_n$ が減少を始める。/ya/, /ra/, /wa/ では ${}_{N}E_n$ と ${}_{N}E_m$ がほぼ同時に立上がるが立上り時の傾斜は ${}_{N}E_n$ が ${}_{N}E_m$ より大きい。/N/ では口腔出力が極めて小さく, ${}_{N}E_m$ には室内騒音のエネルギー曲線が現われている。

以上で, 各音韻のエネルギー曲線の特徴を述べたが, この特徴を表わすパラメータを式(3-1)(3-2)で定義する。

Delay time

$$D = t_{no} - t_{mo} \quad (3-1)$$

S_n / S_m

$$S = \frac{N E_n(t_{m3}) - N E_n(t_{m0})}{N E_m(t_{m3}) - N E_m(t_{m0})} \quad (3-2)$$

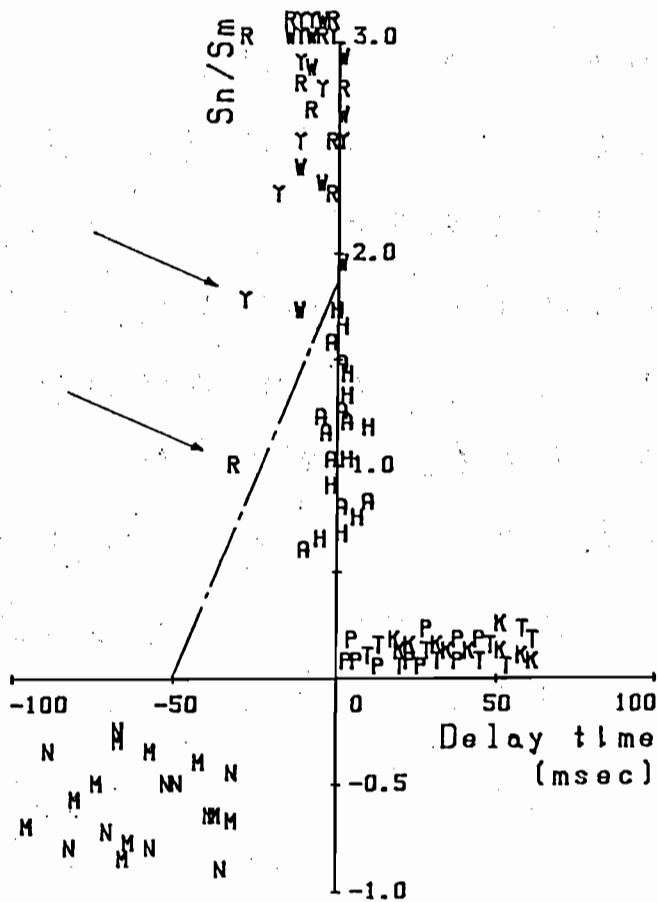


図 3-2 10名の男性が孤立発声した単音節の D-S平面上での分布 (遮蔽板使用)

但し, t_{no} , t_{mo} は ${}_N E_n$, ${}_N E_m$ が各々 5% 点を初めて越えた時刻, t_{m3} は t_{mo} から 3 フレーム後の時刻である。式 (3-1) は ${}_N E_n$ と ${}_N E_m$ の立上り時間差を, 式 (3-2) は ${}_N E_m$ の立上り時における ${}_N E_n$ と ${}_N E_m$ の傾斜の比を表わしている。

図 3-2 は /sa/ と /N/ を除く 10 種の単音節をこの 2 つのパラメータが張る D-S 平面上にプロットしたもので, 音声試料は 10 名の男性が孤立発声したものを用了。/sa/ では /s/ の区間で ${}_N E_m$ の値のばらつきが大きく t_{mo} の検出が不安定となり, また /N/ では口腔出力が極めて小さいため t_{mo} が決定できないという理由で除外した。図 3-2 によれば /a/, /ha/ の D は小さく, S は 1.0 を中心に分布する。/ka/, /ta/, /pa/ では $D > 0$ で S は小さい。又, /na/, /ma/ では $D < 0$, $S < 0$ であり, /ya/, /ra/, /wa/ では D が小さく, S は他の音韻群より大きい。図 3-2 では $S > 3.0$ の場合, $S = 3.0$ の位置にプロットしている。/ya/, /ra/, /wa/ では, ${}_N E_m$ が ${}_N E_n$ の立上りよりややおくれる

($D < 0$) 音声試料があることが図3-2に示されている。この場合, ${}_N E_n$ と ${}_N E_m$ の概形は図3-1と同じであるが, 図3-1における /ya/, /ra/, /wa/ の最初のピークの頂上付近で ${}_N E_n$ の傾きを計算するため, S がやや小さくなることがある(図3-2矢印)。

3.3 $E_n(t)/E_o(t)$ 時間変化曲線

図2-5に示すように咽頭腔(pharynx cavity)通過エネルギーを E_o , 鼻腔・口腔出力エネルギーを各々 E_n, E_m , 又, マイク $N \cdot M$ が観測するエネルギーを $E_n(in), E_m(in)$ とした時, 観測値 $E_n(in), E_m(in)$ から $E_n(t)/E_o(t)$ 時間変化曲線を推定する問題を考える。エネルギーは声道内で無損失であると仮定すると式(3-3)が成り立つ。

$$E_o(t) = E_n(t) + E_m(t) \quad (3-3)$$

又, C_n, C_m を放射エネルギーのうち各マイクに入る比率として $E_n(in), E_m(in)$ は式(3-4)で表わせれ, 式(3-3)(3-4)より式(3-5)を得る。

$$\left. \begin{aligned} E_{n(in)}(t) &= C_n E_n(t) \\ E_{m(in)}(t) &= C_m E_m(t) \end{aligned} \right\} (3-4)$$

$$C_n E_o(t) = E_{n(in)}(t) + (C_n/C_m) E_{m(in)}(t) \quad (3-5)$$

ここで式(3-5)における C_n/C_m を決定する方法について述べるが，この実験では1名の男性が孤立発声した音声試料を用い，又，各エネルギー曲線の立上り時刻を時間軸の原点として議論を進める。

先ず，直径8cm，長さ50cmの円筒の一端に1個のマイクを配置し，他の一端で鼻と口を覆い /ma/ を発声する。これは E_o が体内で E_n と E_m に分割されたものを再び円筒内で合成し， $E_o(t)$ の曲線形を推定する実験である。このマイクに観測されるエネルギーは E_o に比例しており，正規化した場合には両方の曲線形は同一であると考えられる。そこで，/ma/ の10発声についてマイクで観測したエネルギーを正規化したものを ${}_n E_o(t)^i$ ($i=1\sim 10$)，それらの平均曲線を $\overline{{}_n E_o(t)}$ として式(3-6)を計算し，その結果を図3-3(a)に示す。

$$\overline{N E_0(t)} = \frac{1}{10} \sum_{i=1}^{10} N E_0(t)^i \quad (3-6)$$

又、各 $N E_0(t)^i$ の曲線形の類似度を調べるために $\overline{N E_0(t)}$ を中心とした $N E_0(t)^i$ の分散(式(3-7))を計算し、図(e)にこの分散の変化を相対値で

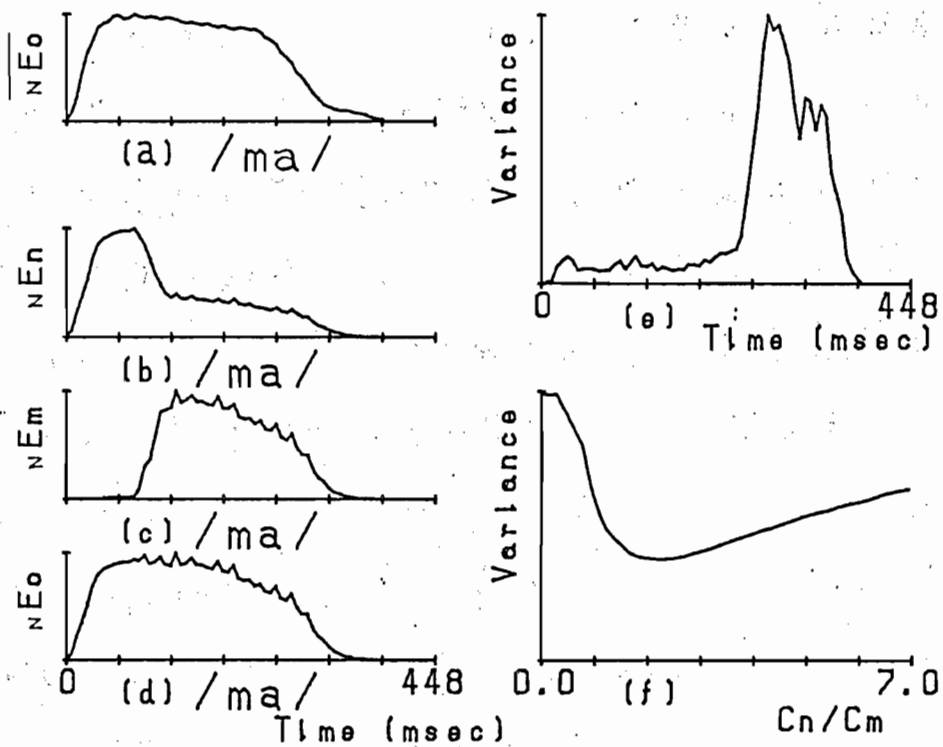


図3-3 (a)~(d) $/ma/$ のエネルギー曲線

(e) $N E_0(t)^i$ ($i = 1 \sim 10$) と $\overline{N E_0(t)}$ の類似度

(f) $N E_0(t, C_n/C_m)^j$ ($j = 1 \sim 10$) と $\overline{N E_0(t)}$ の類似度

表わす。

$$\text{Variance}(t) = \sum_{i=1}^{10} \left({}_N E_0(t)^i - \overline{{}_N E_0(t)} \right)^2 \quad (3-7)$$

図(e)より /ma/ の発声の前半 (0~192 msec, 0~30フレーム) では各 ${}_N E_0(t)^i$ の曲線形が似ているが, 後半では発声継続時間長の相違のため各曲線形が異なっていることが確かめられる。そこで, 立上りから 192 msec までの区間で, 図(a)の曲線を /ma/ の正規化 $E_0(t)$ の代表とする。

次に, 2個のマイクによる実験を行なったが実験中の C_n, C_m の値が一定であるようにマイクを固定した。図(b)(c)は /ma/ の ${}_N E_n, {}_N E_m$ であるが, これによると /ma/ の発声中には口蓋帆 (velum) や舌, 唇等の各器管が時間的に動きながら E_0 を E_n と E_m に分割していることが推測できる。そこで C_n/C_m を変数として式 (3-5) から得られる正規化 $C_n E_0(t)$ の曲線形と図(a)の曲線形が最も似るように C_n/C_m の値を定める。すなわち, 新たに発声した 10 個の /ma/ について各々の観測値 $E_n(i_n), E_m(i_n)$ を式 (3-5) に代入

し、 C_n/C_m を変数として得られた $C_n E_0(t)$ を正規化したものを ${}_N E_0(t, C_n/C_m)^j$ ($j=1 \sim 10$) とする。次に、 $\overline{{}_N E_0(t)}$ の曲線形と ${}_N E_0(t, C_n/C_m)^j$ の曲線形の類似度を分散 (式 (3-8)) として、 $l = 0 \sim 30$ フレームの範囲で計算する。

$$\text{Variance}(C_n/C_m) = \sum_{j=1}^{10} \sum_{l=0}^{30} \left({}_N E_0(l, C_n/C_m)^j - \overline{{}_N E_0(l)} \right)^2 \quad (3-8)$$

図(f)はこの分散の変化を相対値で表わしたものである。極小点における C_n/C_m の値が求める値である。この値はマイク $N \cdot M$ の配置と増幅器の利得によって定まる。図(b)(c)は $/ma/$ の一音声試料の $E_n(in)$ 、 $E_m(in)$ を各々正規化したものであるが、式 (3-5) によってこの $E_n(in)$ と $E_m(in)$ を上で得た値 C_n/C_m の比率で加え正規化した結果を図(d)に示す。0 ~ 192 msec の範囲で図(d)は図(a)に非常に似た曲線形であることが確かめられる。

ここで、子音の特徴を表わすパラメータとして $E_n(t)/E_0(t)$ を導入し、上で得た C_n/C_m の

値と観測値 $E_{n(in)}(t)$, $E_{m(in)}(t)$ によって式 (3-9) を計算する。

$$E_n/E_0 = \frac{C_n E_n}{C_n E_0} = \frac{E_{n(in)}}{E_{n(in)} + (C_n/C_m) E_{m(in)}} \quad (3-9)$$

但し、式 (3-9) では (t) を省略している。又、分母の E_0 が小さい区間では E_n/E_0 が不安定な変動をすするため、経験的に、 E_0 が最大値の 15% 以上の区間で式 (3-9) を計算することにした。図 3-4 は図 3-1 と同一の音声試料について E_n/E_0 を計算したものである。図 3-4 によると、 $/a/$ では E_n/E_0 の値が小さく変化も小さい。これは $/a/$ の発声中には口腔の通路が広く開いた状態で各調音器管が静止しているためであると考えられる。 $/ka/$, $/sa/$, $/ta/$, $/ha/$ では無声子音区間で E_0 が小さく、図 3-4 には表わされていないため $/a/$ と同様の曲線となっている。 $/pa/$ では曲線の左半分の区間で破裂気流の影響が残り、 $E_n/E_0 \approx 0$ となっている。鼻音 $/na/$, $/ma/$ では最初 E_n/E_0 が 1.0 に近く、次に急激に値が小さくなる。これは子音

区間では口蓋帆や舌，唇等の調音器管が口腔出力を阻止しているが，次に各調音器管が急激に動いて母音/a/の位置になり，口腔の通路を開放するためであると考えられる。又，

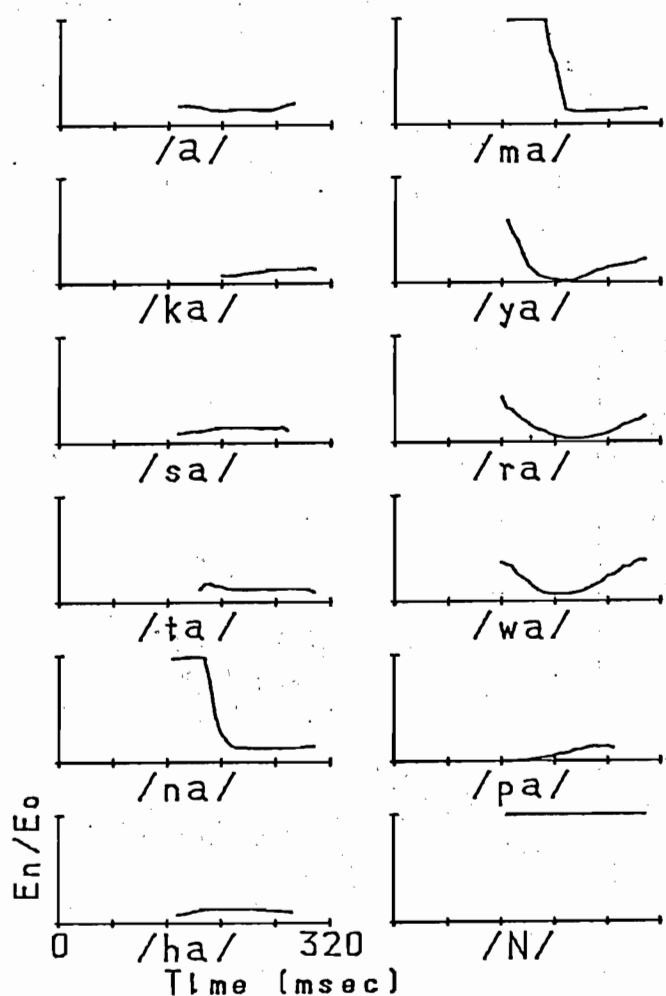


図3-4 特徴パラメータ E_n/E_0 の時間変化曲線

但し, t_0 は E_0 が最大値の 15% 点を初めて越えた時刻, t_1 は時刻 t_0 から 37 レーム後の時刻である。式 (3-10) は E_n/E_0 曲線の左端の値を, また式 (3-11) は曲線の最大傾斜を表わしている。図 3-5 は 10 名の男性が孤立発声した 12 種の単音節 (図 3-2 と同一音声試料) をこの二つのパラメータが張る平面 (これを $O-\Delta$ 平面と呼ぶ) にプロットしたものである。(図 3-5, 3-7, 3-9, 3-10 では図を見やすくするため 5 個の音声試料のみプロットした音韻があるが, 他の 5 個も同様の分布をしている。又, /i/, /u/ は後の考察で参考にするため描いている。) 図 3-5 では各話者別に上述の方法で C_n/C_m を定め式 (3-10) (3-11) を計算した。この結果, 12 種の単音節が 97.5% (117/120) の識別率で (/a/, /ka/, /sa/, /ta/, /ha/, /pa/), (/na/, /ma/), (/ya/, /ra/, /wa/), (/N/) の 4 群に分類された。

3.4 識別実験

3.4.1 遮蔽板を用いた識別実験

識別フローチャートを図3-6に示す。第1段階では12種の単音節を図3-5のO-Δ平面上で4群に分類する。第2段階では(/a/, /ka/, /sa/, /ta/, /ha/, /pa/)の中から零交叉分析法を用いて /sa/ を分離する。ここで、零交叉数はフレーム周期 6.4 msec, 分析区間 25.6 msec で

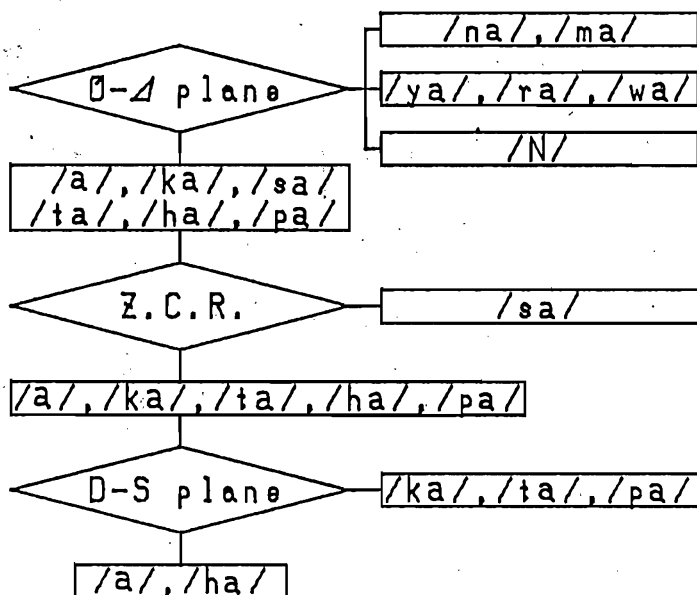


図3-6 12種の単音節を分類するためのフローチャート

式(3-12)によって計算し，第 j フレームの零交叉数を $Z(j)$ として式(3-13)によって $Z(j)$ の中の最大値 $Z.C.R.$ を求める。

口腔出力音声信号を $\{x_i\}$ として， $x_i x_{i+1} < 0$ ならば区間 $[i, i+1]$ で零交叉が1回あった。
 (3-12)

$$Z.C.R. = \text{Max}[Z(j)] \quad (3-13)$$

図3-7は，図3-5と同一の音声試料について $Z.C.R.$ を示したものであるが(/a/, /ka/, /sa/, /ta/, /ha/, /pa/)の中から /sa/ が完全に分離されていることがわかる。第3段階では(/a/, /ka/, /ta/, /ha/, /pa/)を図3-2のD-S平面上で

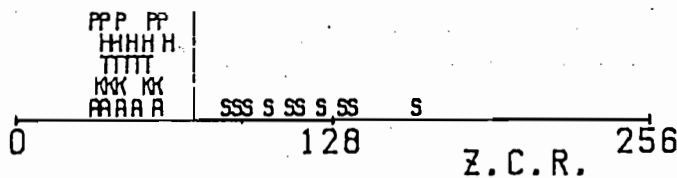


図3-7 各単音節の最大零交叉数
 (遮蔽板使用)

(/a/, /ha/) と (/ka/, /ta/, /pa/) に分類する。後述の理由のため D-S 平面上で /na/, /ma/, /ya/, /ra/, /wa/ の識別は行わない。

以上で識別アルゴリズムを示したが図 3-2, 3-5, 3-7 に共通して用いた 120 個の音声試料にこのアルゴリズムを適用する。但し, 各図の識別境界は事後的に設定したものである。誤識別は第 1 段階の図 3-5 で 3 個の /wa/ にみられるだけであり, 全体として 97.5% (117 / 120) の識別率が得られた。次にこれらの識別境界と同じ境界を事前に設定し, 同じ 10 名が同様の条件で発声した別の音声試料を試験用データとして実験を行ない, 4 個の /wa/ の誤識別で 96.7% (116 / 120) の識別率を得た。

3.4.2 遮蔽板を用いない識別実験

遮蔽板を取り除き, マイク N・M をそれぞれ鼻の直前および口から 3 cm の位置に配置して実験を行なった。ところで, この実験では遮蔽板を用いないため鼻腔・口腔出力が N・

Mに混入することが予想される。そこで、第2章2.4節で述べた遮蔽の程度を調べる実験と同様の実験を行なった結果、鼻側から口側への減衰は -11dB 以下、口側から鼻側への減

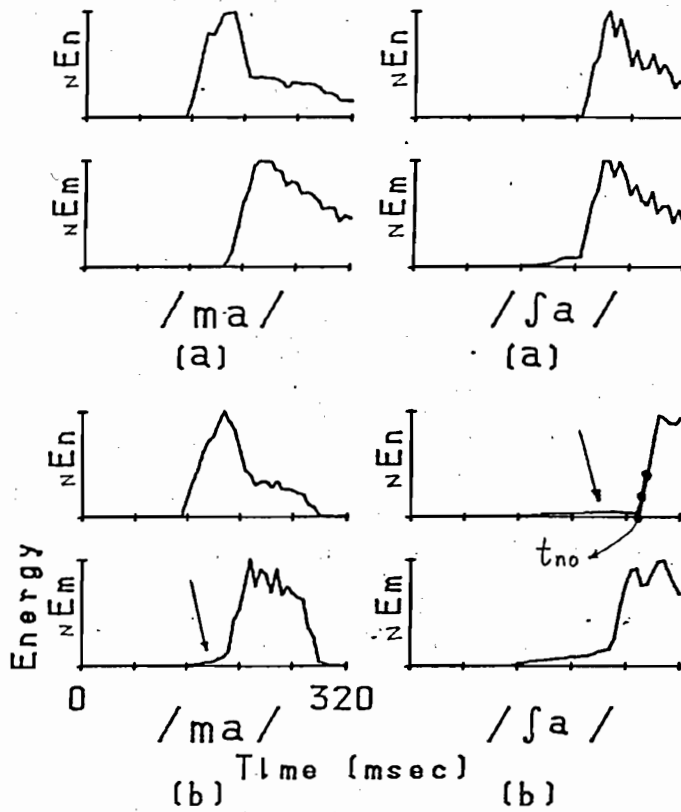


図3-8 鼻腔・口腔出力エネルギーの正規化時間変化曲線

- (a) 遮蔽板使用の場合
- (b) 遮蔽板を使用しない場合

衰は -10dB 以下であった。この混入の様子が最も顕著に現われる発声 $/ma/$ と $/ja/$ の例を図 3-8 に示す。図 (a) は遮蔽板を用いた場合、図 (b) は用いない場合である。 $/ma/$ では図 (a) に示されるように鼻腔出力が口腔出力より早く立上がるため図 (b) では実際に口腔出力が存在しないはずの部分 (矢印) にエネルギーが観測されてしまう。逆に、 $/ja/$ では図 (a) に示されるように口腔出力が鼻腔出力より早く立上がるため図 (b) では実際に鼻腔出力が存在しないはずの部分 (矢印) にエネルギーが観測される。この混入の程度は予想より小さい値であったが、これは遮蔽板を用いた場合には遮蔽板の厚さのためにマイクの位置や角度に制限があったが、遮蔽板を取り除いたことにより、自由に適正なマイクの配置が可能になり、入力が効率良く行なわれたためである。

次に、鼻腔・口腔出力の混入が前述の特徴パラメータに及ぼす影響について検討する。口腔出力が鼻腔出力より早く立上がる発声 (

例えば図3-8 /sa/ (b))では $N E_n$ の立上り時刻 t_{no} の検出が不正確となり，特に式(3-1)に大きな誤差を生じさせる。そこで図3-8 /sa/ (b)に示すように $N E_n$ の35%点と15%点を直線で結び，時間軸との交点を t_{no} として推定することにした。但し，この35%，15%の値は経験的に定めた。又，式(3-2) (3-10) (3-11) (3-12) (3-13)では t_{no} を直接には用いていない。このように口腔出力が鼻腔出力より早く立上がる発声に t_{no} の補正法を用いれば，相互の混入の程度が小さいため十分な精度の結果が期待できる。一方，鼻腔出力が口腔出力より早く立上がる発声(例えば図3-8 /ma/ (b))では $N E_m$ の立上り時刻 t_{mo} の検出が不正確となり，特に式(3-1) (3-2)に大きな誤差を生じさせる。遮蔽板を用いない場合，(/na/ , /ma/)と(/ya/ , /ra/ , /wa/)は $N E_m$ の立上り付近の曲線形が似るため， t_{mo} に上述の補正法を用いることは適当でない。そこで式(3-1) (3-2)のパラメータによるD-S平面上では(

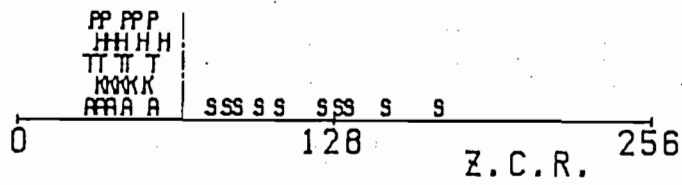


図 3-10 各単音節の最大零交叉数
(遮蔽板なし, 設計用データ)

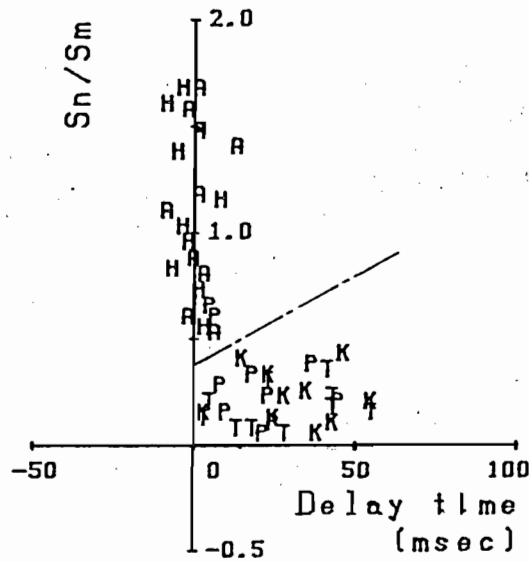


図 3-11 10名の男性が孤立発声した単音節の
D-S平面上での分布
(遮蔽板なし, 設計用データ)

いても相互の混入の程度が小さいため十分な精度の結果が期待できる。

図3-9, 3-10, 3-11は遮蔽板を用いないで集められた音声試料に t_{no} の補正を施した時の各パラメータの分布を表わしている。音声試料は10名の男性によって孤立発声されたもので

表3-1 設計用データの識別結果
(遮蔽板なし, 識別率94.2%)

出力 入力	/a/ /ha/	/ka/ /ta/ /pa/	/sa/	/na/ /ma/	/ya/ /ra/ /wa/	/N/
/a/	20					
/ka/ /ta/ /pa/	2	28				
/sa/			10			
/na/ /ma/				20		
/ya/ /ra/ /wa/	5				25	
/N/						10

ある。これらによると t_{no} の補正が適切であり，混入の影響も小さいことがわかる。そして，これらを設計用データとして図3-6のアルゴリズムによって識別実験を行なった結果を表3-1に示す。次に，設計用データと同じ識別境界を事前に設定し，同じ10名の話者が

表 3-2 試験用データの識別実験
(遮蔽板なし，識別率91.7%)

出力 入力	/a/ /ha/	/ka/ /ta/ /pa/	/sa/	/na/ /ma/	/ya/ /ra/ /wa/	/N/
/a/ /ha/	20					
/ka/ /ta/ /pa/	4	26				
/sa/			10			
/na/ /ma/				20		
/ya/ /ra/ /wa/	6				24	
/N/						10

同様の条件で発声した別の音声試料を試験用データとして識別実験を行なった結果を表3-2に示す。表3-1, 3-2によれば, 設計用・試験用データに対して各々94.2% (113/120), 91.7% (110/120)の識別率が得られている。

3.5 考察

(1) /ya/, /wa/における nE_m の立上り時の傾斜と E_n/E_0 の特徴について考察する。図3-5には母音/i/, /u/の分布を示しているが, これによれば/i/, /u/は/a/と同様に $\Delta(E_n/E_0)$ の値は小さいが, 一方 $(E_n/E_0)_0$ の値が大きいことがわかる。すなわち, /i/, /u/の E_n/E_0 曲線は/a/の曲線を上方に平行移動した平坦な曲線である。このことは2個のマイクのモニター用オシロスコープによって, /i/, /u/の鼻腔出力が/a/より大きく, 逆に, 口腔出力は/a/より小さいことが容易に観察されることから説明できる。/y/, /w/の各調音器管の形状は各々/i/, /u/のそれに似ているこ

とから， $/ya/$, $/wa/$ は $/i/$, $/u/$ と $/a/$ を連結した発声に類似していると考えられる。このため，子音区間 $/y/$, $/w/$ では鼻腔出力が優勢であり，母音区間 $/a/$ では口腔出力が優勢となり， ${}_N E_m$ の値は最初は小さく，次に大きくなる。又， E_n/E_0 の値は最初は大きく，次に小さくなると考えられる。

(2) セグメンテーションの問題について考察する。実験によると遮蔽板を用いた場合，マイク N に鼻腔出力が観測されるのは口を閉じた呼吸時における肺からの排気流と声帯振動時の音声波のえっただけであり，無声音の発声時には鼻腔出力は観測されない。(例えば図 3-1 $/sa/$ の $/s/$ の区間)。従って，孤立発声した単音節については鼻腔出力は声帯の振動を直接反映しており， ${}_N E_n > 0$ の区間では声帯が振動していると考えられる。このため無声子音では ${}_N E_n$ の立上り時刻以前を子音区間，以後を母音区間として，又，有声子音では E_n/E_0 曲線が平坦で，かつ小さい値の区間

を母音区間，それ以前を子音区間としてセグメンテーションが容易に行なえる。遮蔽板を用いない場合にも鼻腔出力の立上り時刻 t_{no} の補正によって上述のことが同様に行なえ，母音・子音区間の境界が高い精度で検出できる。従って，この章で分類された6群の音韻をさらに細かく分類するために周波数領域で分析を行なう時，スペクトルを計算すべき区間が正確に検出され，このため，計算量の圧縮と識別率の向上が期待できる。

(3) この章では，後続母音に /a/ をもつ子音の特徴について述べたが，後続母音が /e/, /o/ の場合も同様の結果が得られている。しかし，(1) で述べたように /i/, /u/ では E_n/E_0 の値が大きくなり，このため，特に /mi/, /mu/, /yu/ 等の E_n/E_0 の特徴が不明確となり，今後の検討を要する。

3.6 む す び

鼻腔・口腔各々から出る音声波エネルギー

の時間変化に着目して、12種の単音節を6群に分類する実験を行なった。子音によって鼻腔・口腔出力の立ち上がり時間差と立ち上がり特性が異なり、又、 E_n/E_0 曲線には4つのタイプがあることも確かめた。音声波エネルギーに関するこれらの特徴量を用いて識別実験を行ない、二重入力システムの有効性を確かめた。

第4章 口腔出力と鼻腔出力の分離ピックアップ方式による摩擦音・破擦音の識別⁽³⁰⁾

4.1 はじめに

従来、摩擦音・破擦音は原波形の零交叉数やスペクトルの重心周波数・ピーク周波数・谷周波数を特徴パラメータに用いて認識実験がなされてきた。この章では二重入力システムによって得られた音声波の鼻腔出力と口腔出力を時間領域の中だけで処理し、日本語の68種の単音節の中から無声摩擦音 /s/, /h/, /ʃ/ と有声摩擦音 /z/, 破擦音 /tʃ/, /tsu/ を自動識別する新しい手法を示す。最初に二重入力システムを用いて得られた音声波の波形の特徴と、鼻腔・口腔出力エネルギーの時間変化の特徴を述べ、音声生成過程に基づいた解釈を行なう。そして、自己相関係数と零交叉数による分析法を検討し、その結果から雑音除去差分零交叉数による分析法を提案する。この雑音

除去差分零交叉数の最大値を特徴パラメータの一つに用い，又，鼻腔出力エネルギーの立上り時刻と雑音除去差分零交叉数の立上り時刻との時間差をもう一つの特徴パラメータに用いることによって，/s/，/h/，/ʃ/，/z/，/tʃi/，/tsu/ を高い精度で識別できることを示す。

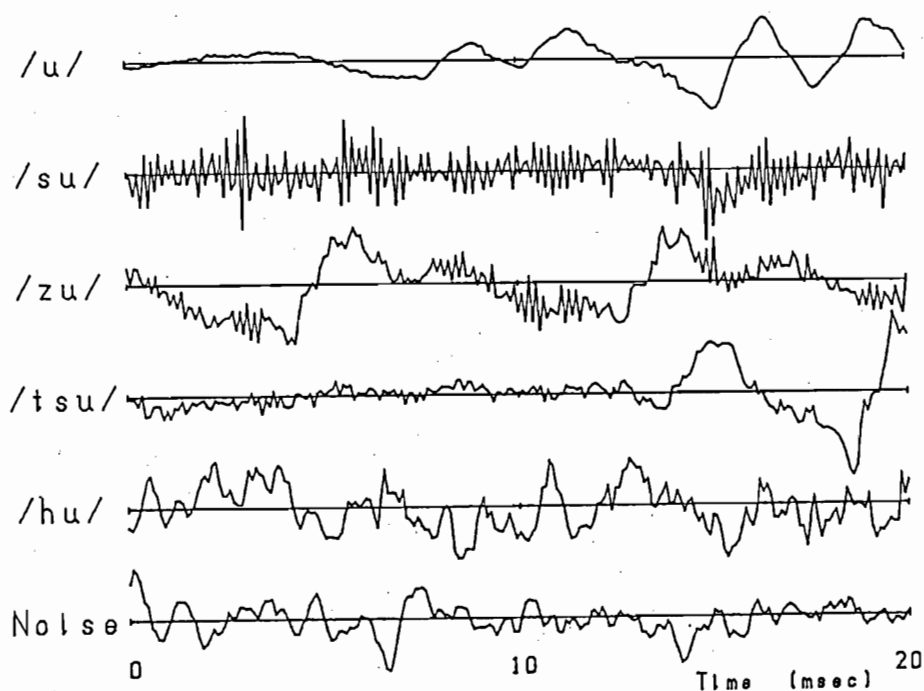


図 4-1 1 名の男性によって発声された単音節の発声開始区間における口腔出力音声波形と室内騒音波形

最後に遮蔽板を使用しないで行なった識別実験の結果を示し、このシステムの有効性と実用性を確かめる。

4.2 摩擦音・破擦音の特徴

本節では後続母音に /u/ をもつ孤立発声された CV 音節について議論を進める。これは5つの後続母音の中でも後述の特徴が顕著であることによる。

4.2.1 波形とエネルギー曲線

図4-1はマイクMで観測した母音 /u/ の立上り時の波形と /su/, /zu/, /tsu/, /hu/ の子音区間の波形である。これらは一例であるが、一般に /u/ の波形は滑らかであるのに対して /su/, /tsu/ の子音区間の波形は零レベルを中心に細かい振動をもつ。又、/hu/ の子音区間の波形は /su/, /tsu/ とは多少異なった不規則性をもち、/zu/ の子音区間については口腔の伝達特性によって修飾された声帯振動波の上に細かい振

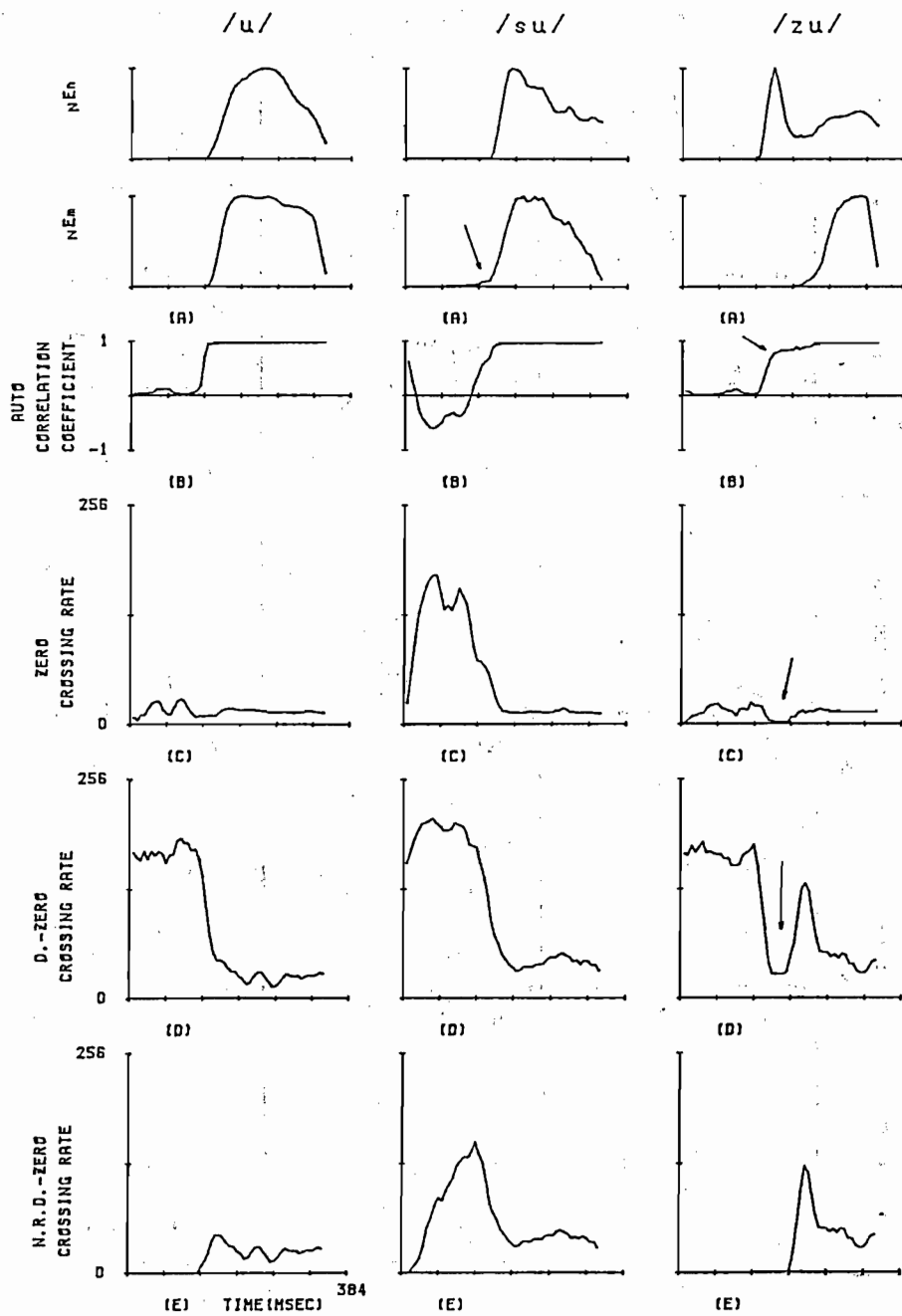


図 4-2 (A) 鼻腔・口腔出力エネルギーの正規化時間変化曲線
 (B) 口腔出力の自己相関係数の時間変化曲線
 (C) 口腔出力の零交叉数の時間変化曲線
 (D) 口腔出力の差分零交叉数の時間変化曲線
 (E) 口腔出力の雑音除去差分零交叉数の時間変化曲線

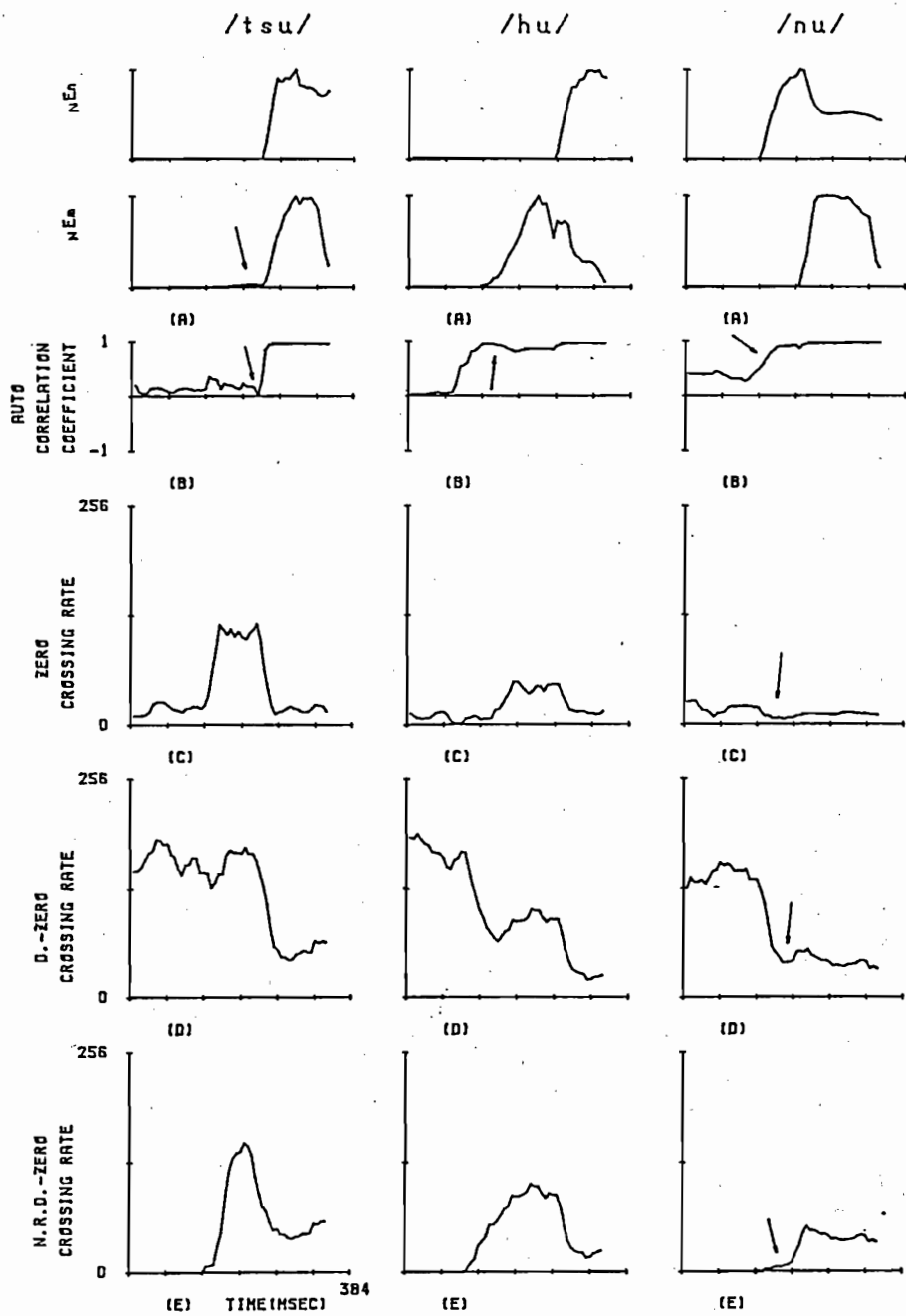


図 4-3 (A) 鼻腔・口腔出力エネルギーの正規化時間変化曲線
 (B) 口腔出力の自己相関係数の時間変化曲線
 (C) 口腔出力の零交叉数の時間変化曲線
 (D) 口腔出力の差分零交叉数の時間変化曲線
 (E) 口腔出力の雑音除去差分零交叉数の時間変化曲線

動が重なった波形が現われる。一方、図4-2 (A), 図4-3 (A)は6種の単音節についてマイク N・Mが観測したエネルギー曲線 ${}_N E_n, {}_N E_m$ を示している。このエネルギー曲線は無声子音区間における肺からの気流の観測値も含んで

表4-1 単音節の子音区間における
各パラメータの値

t_{no} : ${}_N E_n$ の立上り時刻

t_{mo} : ${}_N E_m$ の立上り時刻

	t_{no}, t_{mo}	自己相 関係数 ρ_i	零交叉数	差分 零交叉数	雑音除去 差分 零交叉数
/u/	=	大	小	小	小
/su/	>	負	大	大	大
/zu/	<	大	小	大	大
/tsu/	>	小	大	大	大
/hu/	>	大	中	中	中
/nu/	<	大	≈ 0	小	小
Noise		小	小	大	≈ 0

いる。これらの図に注目すると /u/ では N_{En} と N_{Em} の立上りが同時であるのに対して /su/, /tsu/, /hu/ では N_{Em} の立上りが N_{En} よりも早く、逆に /zu/, /nu/ では N_{En} の立上りが N_{Em} よりも早い。これらの結果を表 4-1 にまとめた。(図 4-2, 4-3 (A)~(E) は同一音声試料のものである。)

以上に述べた波形の特徴とエネルギー曲線の特徴について音声生成過程に基づいた解釈を述べる。実験によるとマイク N に鼻腔出力が観測されるのは口を閉じた呼吸時における肺からの排気流と声帯振動時の音声波の二つの場合だけであり、無声音 /s/ や /h/ の発声時には鼻腔出力は観測されない。従って、孤立発声した CV 音節については鼻腔出力は声帯の振動を直接反映しており、 N_{En} の立上り時刻は声帯の振動開始時刻に等しいと考えられる。母音 /u/ の発声時には声帯振動波が鼻腔・口腔各々の伝達特性により修飾されて鼻腔と口腔から同時に出力される。無声摩擦音

/su/, /hu/ 破擦音 /tsu/ の子音区間では鼻腔出力は観測されない，一方口腔通過気流は歯や唇の狭めの間を通過する時に乱気流となり図4-1のような細かい振動をもつ不規則な波形を生じる。そして，その直後に声帯が振動を開始するため図4-2(A)，図4-3(A)に示されるように nE_n の立上りが nE_m の立上りより遅れる。但し，図4-2(A)，図4-3(A)にはこの nE_n と nE_m の立上り時間差が顕著な例を用いているが，一般には他の後続母音の場合も含めて，無声摩擦・破擦音の子音区間の nE_m は後続母音区間の nE_m に較べて極めて小さく(図4-2(A) /su/ 矢印，図4-3(A) /tsu/ 矢印)，立上り時間差を特徴パラメータに用いることは有効でない。有声摩擦音 /zu/ や鼻音 /nu/ では声帯振動開始時点において各調音器管が口腔出力を阻止しており，このため nE_m の立上りより nE_n の立上りが早くなる。声帯振動開始後，/nu/ では母音区間が始まり口腔出力が観測されるようになるが，/zu/ では口腔伝達特性により

修飾された声帯振動波が歯の間を通過することによって生ずる子音区間 /z/ が母音区間の直前に存在し、この区間での波形は図4-1に示されるように口腔伝達特性により修飾された声帯振動波の上に乱気流が重なったものとなる。

4.2.2 自己相関係数と零交叉数

以上で述べた摩擦音・破擦音の波形の特徴を表わすパラメータとして自己相関係数 (Auto correlation coefficient) と零交叉数 (Zero crossing rate) を用いた分析法を検討する。この章では時間領域での波形情報処理を目的としており、これら二つのパラメータを検討することは重要である。一般に、自己相関係数は基本周波数の検出に用いられ、又、この基本周期に相当する付近での自己相関係数の最大値は有声音と無声音の判定に用いられる。ところで、10kHz でサンプリングした図4-1 /su/ の波形 $\{x_i\}$ が x_i と x_{i+1} で符号が交互に入れ替る

細かい振動をしており，/zu/にもこれと同一周期の振動がみられる。一方，/hu/では/su/，/zu/とは異なる不規則な振動が現われている。これらの事実から，ラグタイムとサンプリング周期 0.1 msec に等しく設定した自己相関係数 ρ_1 と 0.1 msec ごとのサンプルの零交叉数をパラメータとして用いることによって摩擦音・破擦音の識別の可能性が期待できる。本節では特にこの ρ_1 と零交叉数の時間変化について実験を行ない，その結果を検討し，他のラグタイムの自己相関係数や他のサンプリング周期の零交叉数も含めて摩擦音・破擦音のパラメータとしてのこれらの有効性を論ずる。図4-2(B)(C)，図4-3(B)(C)は6種の単音節の口腔出力音声波をフレーム周期 6.4 msec ，分析区間 25.6 msec ，ラグタイム 0.1 msec で計算した自己相関係数 ρ_1 (式(4-1))の時間変化曲線と，同一フレーム周期・同一分析区間で計算した零交叉数(式(4-2))の時間変化曲線である。フレーム周期と分析区間はエネルギー

一曲線と同様な理由によって決定した。

口腔出力音声信号を $\{x_i\}$ とするとき

$$\rho_1 = \frac{\sum_{i=1}^{255} x_i x_{i+1}}{\sqrt{\sum_{i=1}^{255} x_i^2 \cdot \sum_{i=1}^{255} x_{i+1}^2}} \quad (4-1)$$

$x_i x_{i+1} < 0$ ならば区間 $[i, i+1]$ で

零交叉が1回あった。 (4-2)

図4-2(B)(c), 図4-3(B)(c)に示されているように6種の発声に共通して図の中央の子音区間を除いて無音区間の ρ_1 は零に近い値をとり, 母音区間の ρ_1 は1に近い値となる。又, 零交叉数は無音・母音区間共に小さい値となる。従って, /u/ の ρ_1 は零付近から1付近の値へ急激な変化をしますが, 零交叉数は小さな値のまま変化が小さい。これに対し, /su/ の子音区間では ρ_1 は負であり, 零交叉数は増大する。/zu/ では鼻腔出力開始時点から ρ_1 が増加し(図(B)矢印)零交叉数は口腔出力開始時点まで減少を続ける(図(c)矢印)。一般に鼻子

音の零交叉数は小さく，従ってこれは遮蔽板を突抜けてきた極めて振幅の小さい鼻腔出力によってこの現象が起こると考えられる。又，/z/ の区間では図4-1の波形から推測できるように零交叉数は小さい。/tsu/の子音区間では P_1 は小さく（図(B)矢印），零交叉数は増加している。/hu/の子音区間では P_1 は大きく（図(B)矢印），零交叉数はやや大きい。/nu/では P_1 ・零交叉数とも/zu/と同様な傾向がみられる（図(B)(c)矢印）。以上ではサンプリング周期0.1 msec，ラグタイム0.1 msecの P_1 ・零交叉数と各単音節との関係を述べたが，それらを表4-1にまとめた。

ところで，これら6種の単音節の例において無音区間での P_1 に0~0.5の間で不安定な変動がみられるが，これは発声が計算機室内で行なわれたため室内騒音（計算機のファン等）の影響を受けたと考えられる。参考のため，図4-1に増幅器の利得を発声時よりも大きくして室内騒音を計算機に入力した波形を示し

た。この室内騒音はたとえ極めて振幅の小さいものであっても図4-2 /zu/ (B) 矢印，図4-3 (B) /nu/ 矢印と同様に ρ_1 の値を大きく変動させる影響を及ぼすものである。図4-2 (B)，図4-3 (B) には ρ_1 が顕著な値をとる例を示したが，実験によると室内騒音の影響によって /s/ の ρ_1 の値が負から零付近に移る音声試料が多く，全般に ρ_1 の不安定な変動がみられた。ラグタイムを変化させた自己相関係数についても室内騒音の影響は大きく，この不定定さのため自己相関係数を摩擦音や破擦音の特徴パラメータとして用いることは適当でないと考えられる。一方，零交叉数は自己相関係数ほど室内騒音の影響を受けず比較的安定で，特に無声摩擦音 /su/，破擦音 /tsu/ を他の音素から識別するための有効なパラメータであると考えられるが，有声摩擦音 /zu/ では声帯振動波の影響で零交叉数が小さな値となり母音等の他の音素から分離することが困難である。この点についてはサンプリング周期を変化させ

ても改善を期待できない。

4.2.3 差分零交叉数と雑音除去差分零交叉数

本論文ではサンプリング周波数に 10 kHz を用いているため以下ではこのサンプリング周波数に限定した議論を展開するが、この議論は他のサンプリング周波数にも容易に拡張できるものである。前述の検討の結果、零交叉数は無声摩擦音 /su/、破擦音 /tsu/ に有効であるが、有声摩擦音 /zu/ には有効でないことが示されたが、再び図 4-1 の波形を観察してみると、/su/、/tsu/、/zu/ の子音区間に共通な特徴は波形が 0.2 msec 周期の細かい振動をもちこたである。そこでこの細かい振動を表現するパラメータとして式 (4-3) で示す差分零交叉数 (Differential zero crossing rate) を導入する。

$(x_{i+1} - x_i)(x_i - x_{i-1}) < 0$ ならばサンプル点 i において差分零交叉が 1 回あった。(4-3)

エネルギー曲線と同一フレーム周期・同一分析区間・同一音声試料で式(4-3)を計算し、その時間変化曲線を求めた結果を図4-2(D)、図4-3(D)に示す。同図によれば無音区間と摩擦音・破擦音の子音区間の両方で差分零交叉数が多いことがわかる。これは図4-1の波形から明らかのように摩擦音・破擦音だけでなく室内騒音の波形にも細かい振動が存在することによる。又、/zu/, /nu/では図4-2(C)、図4-3(C)の矢印に対応して(D)の矢印の部分でも差分零交叉数の減少がみられる。

次に、無音区間での差分零交叉数を除去するため、式(4-4)で示される雑音除去差分零交叉数(Noise rejected differential zero crossing rate)を導入する。

$$(x_{i+1} - x_i)(x_i - x_{i-1}) < 0 \quad \text{and}$$

$$(|x_{i+1}| > \text{threshold} \text{ or } |x_i| > \text{thr.} \text{ or } |x_{i-1}| > \text{thr.})$$

ならばサンプル点*i*において雑音除去差分零交叉が1回あった。 (4-4)

これは無音区間での口腔出力信号 $\{x_i\}$ の振幅が小さいことに注目して差分零交叉数の定義を修正したものである。雑音除去差分零交叉数の時間変化曲線を図4-2(E), 図4-3(E)に示す。式(4-4)における threshold は無音区間の雑音を観測して適当な値に決めてある。図4-2(E), 図4-3(E)によれば図4-2(D), 図4-3(D)における無音区間の差分零交叉数が除去されている。その上, 図4-2(C), 図4-3(C)よりも摩擦音・破擦音の特徴が強調され, 特に /zu/ では著しいことがわかる。/nu/ では図(C)矢印の部分と同様に, 遮蔽板を突抜けてきた鼻腔出力の影響によって図(E)矢印の部分に小さな値が現われる。以上で述べた差分零交叉数と雑音除去差分零交叉数の各単音節との関係を表4-1にまとめた。

一方, 各単音節^①について雑音除去差分零交叉数の立上り時刻と nE_n の立上り時刻の時間差に注目してみる。この時間差は声帯振動開始時刻に対する口腔出力開始時刻のずれを表

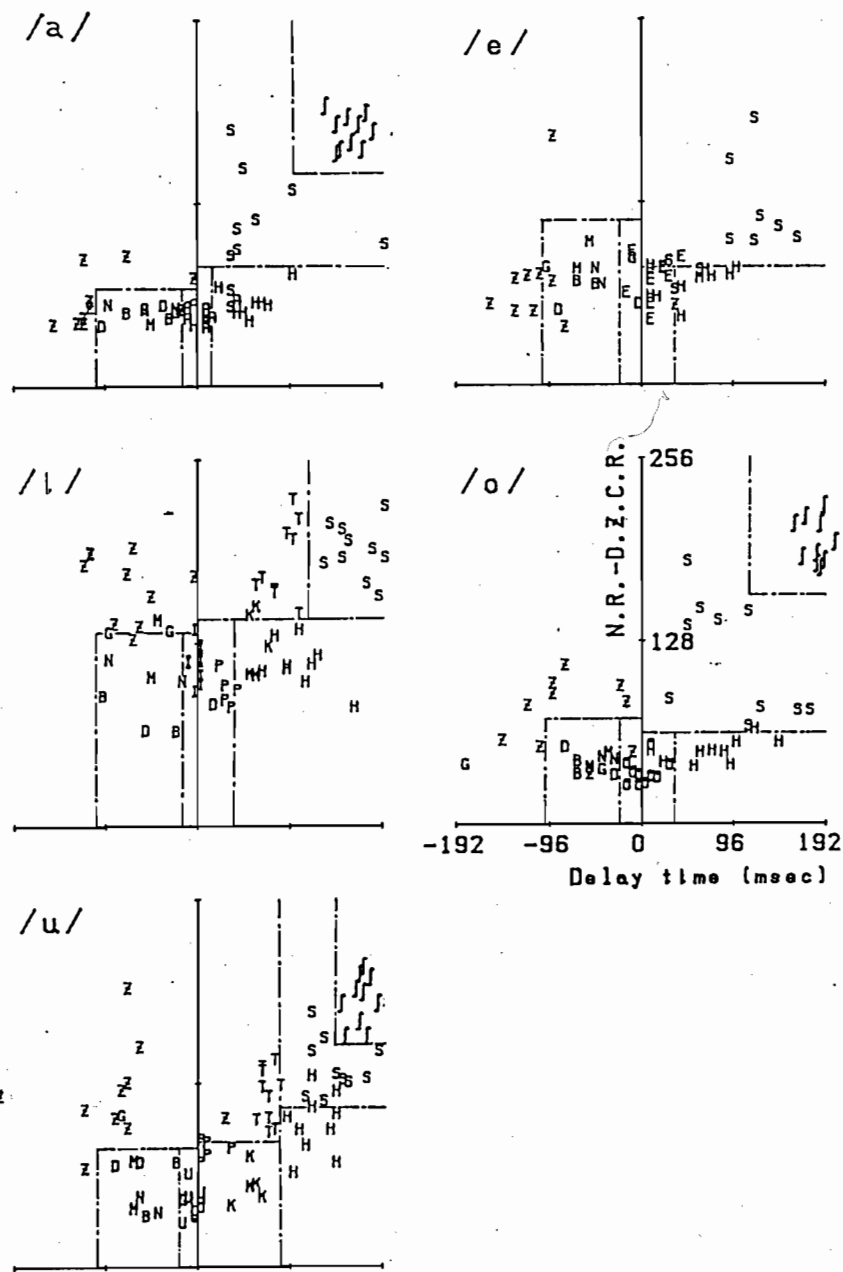


図 4-4 N-D 平面上の各音韻の分布
 パラメータは 10 名の男性の孤立発声
 C V 音節から得られた。
 (遮蔽板使用, 設計用データ)

れし，図4-2(A)(E)，図4-3(A)(E)より /u/ では時間差がほぼ零であるのに対し，/su/，/tsu/，/hu/ では(E)の立上りが(A)よりも早く，/zu/，/nu/ では遅い。以上に述べたことから雑音除去差分零交叉数の最大値と，そして雑音除去差分零交叉数と NE_n の立上り時間差は摩擦音・破擦音を識別するための有効なパラメータであると考えられる。

4.3 識別実験

図4-4に示すように，前節で述べた雑音除去差分零交叉数(N.R.-D.Z.C.R)と縦軸に，式(4-5)に示す雑音除去差分零交叉数と NE_n の立上り時間差(Delay time)と横軸にとった平面(N-D平面)に，後続母音別に各単音節をプロットすることによって摩擦音・破擦音の識別実験を行なった。

NE_n ， NE_m の立上り時刻を各々 t_{no} ， t_{mo} ，雑音除去差分零交叉数の立上りを η 回を初めて越えた時刻 τ として

$$\text{Delay time} = t_{no} - z_0$$

但し, $\text{Delay time} < 0$ のとき

$$\text{Delay time} = t_{no} - t_{mo} \quad (4-5)$$

式(4-5)において $\text{Delay time} < 0$ の場合に t_{mo} を用いる理由は, 鼻腔出力が口腔出力よりも早い子音 /n/, /z/ 等では図4-3(E)矢印の部分のように雑音除去差分零交叉数が完全に零にならない区間が存在し, Delay time が不安定になることによる。母音, /s/, /h/, /ʃ/, /z/, /tʃi/, /tsu/ の音声試料は10名の男性話者が各単音節を1回ずつ発声したものの合計10個用いており, /n/, /m/, /g/, /d/, /b/ については各2個の音声試料, 合計10個を各人に割り当てた。但し, /di/, /du/ は日本語 /ディ/, /トゥ/ に対応する。本節では母音, /s/, /h/, /z/, (/sa/, /su/, /so/), (/tʃi/, /tsu/), (/n/, /m/, /g/, /d/, /b/) の7つのグループの識別実験を後続母音別に行なった結果について述べる。

4.3.1 遮蔽板を用いた識別実験

設計用データとして7グループ各10個の音声試料を遮蔽板を用いた入力によって集め、式(4-4) N.R.-D.Z.C.R. と式(4-5) Delay time を計算した。その結果を図4-4のN-D平面上にプロットし、7つの領域に区分した。又、後の考察で参考とするため /p/, /k/ の音声試料

表4-2 10名の男性が孤立発声したCV音節と母音の識別結果
(遮蔽板使用, 設計用データ)

	/a/	/i/	/u/	/e/	/o/
Vowel	9	10	10	9	10
/s/	8	10	10	8	10
/h/	8	10	8	7	8
/ʃ/	10	/	10	/	10
/z/	9	9	9	7	8
/tʃi/ /tsu/	/	10	10	/	/
/n/, /m/ /g/, /d/, /b/	10	8	8	8	9

と合わせて書いた。設計用データの識別結果を表4-2に示す。表4-2は10発声中の正解数を示している。全体の識別率として90.0% (270/300) を得た。

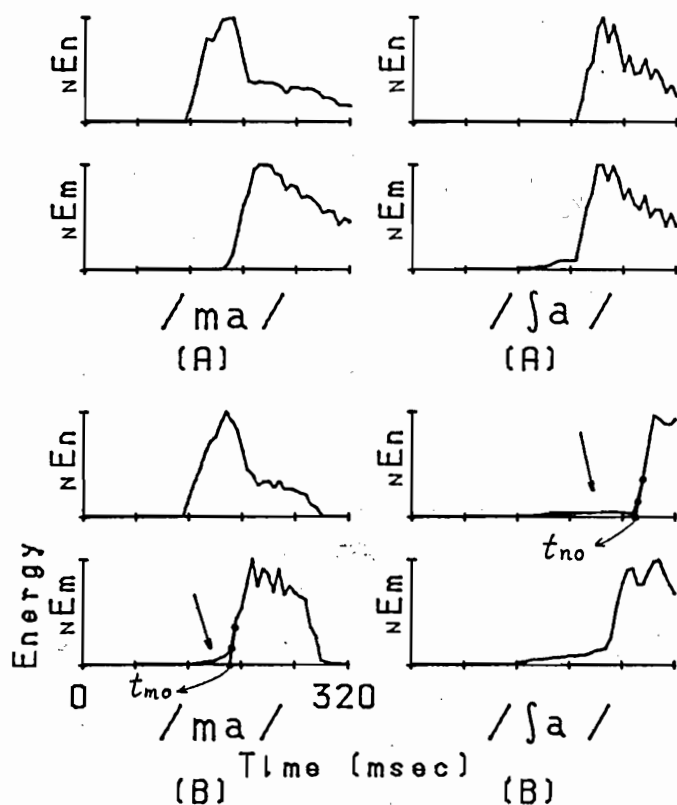


図4-5 鼻腔・口腔出力エネルギーの正規化時間変化曲線

- (A) 遮蔽板使用の場合
- (B) 遮蔽板を使用しない場合

4.3.2 遮蔽板を用いない識別実験

試験用データの識別実験では遮蔽板を取り除き、2個のマイクを各々鼻の直前と口から3cmの位置に配置して7グループ各10個の音声を入力し、図4-4に示す領域内に正しくプロットされたものを正解とした。発声者は設計用データと同じ10名の男性である。ここで、マイクNは N_{En} の立上り検出の目的だけに使われているため、設計用データの識別実験の場合よりも鼻に近く配置しても実験の条件を変えたことにはならないと考えられる。ところで、この実験では遮蔽板を使用しないため鼻腔・口腔出力がマイクN・Mに混入することが予想されるが、混入の程度は3.4.2節で述べたように小さな値であった。混入の様子を図3-8と同様に図4-5に示す。

次に、鼻腔・口腔出力の混入が2つの特徴パラメータに及ぼす影響について検討する。鼻腔出力が口腔出力より早く立上がる発声(例えば図4-5 /ma/ (B))では鼻腔出力がマイク

Mに混入して式(4-4)雑音除去差分零交叉数の値に影響を及ぼし、図4-3 /nu/ (E)矢印の部分と同様な曲線が現われ易くなることが予想される。特に /zu/ では鼻腔出力の滑らかな波形の音声波が口腔出力音声波に混入することにより、式(4-4)の値が多少小さくなることが予想されるが口腔出力に対する鼻腔出力の混入の割合が小さいことから識別には大きな影響はないと考えられる。又、式(4-5) Delay time については、Delay time < 0 の場合に相当するため nE_n と nE_m の立上り検出誤差の増加は識別結果に悪い影響を及ぼす。そこで、図4-5 /ma/ (B) に示すように nE_m の35%点と15%点を直線で結び時間軸との交点を立上り時刻 t_{no} として推定することにした。一方、口腔出力が鼻腔出力より早く立上がる発声(例えば図4-5 /ja/ (B)) では式(4-4)は子音区間で最大となるため混入の影響を受けず、従って、口腔出力がマイクNに混入して式(4-5)における t_{no} へのみ影響が及ぼされるが、これも

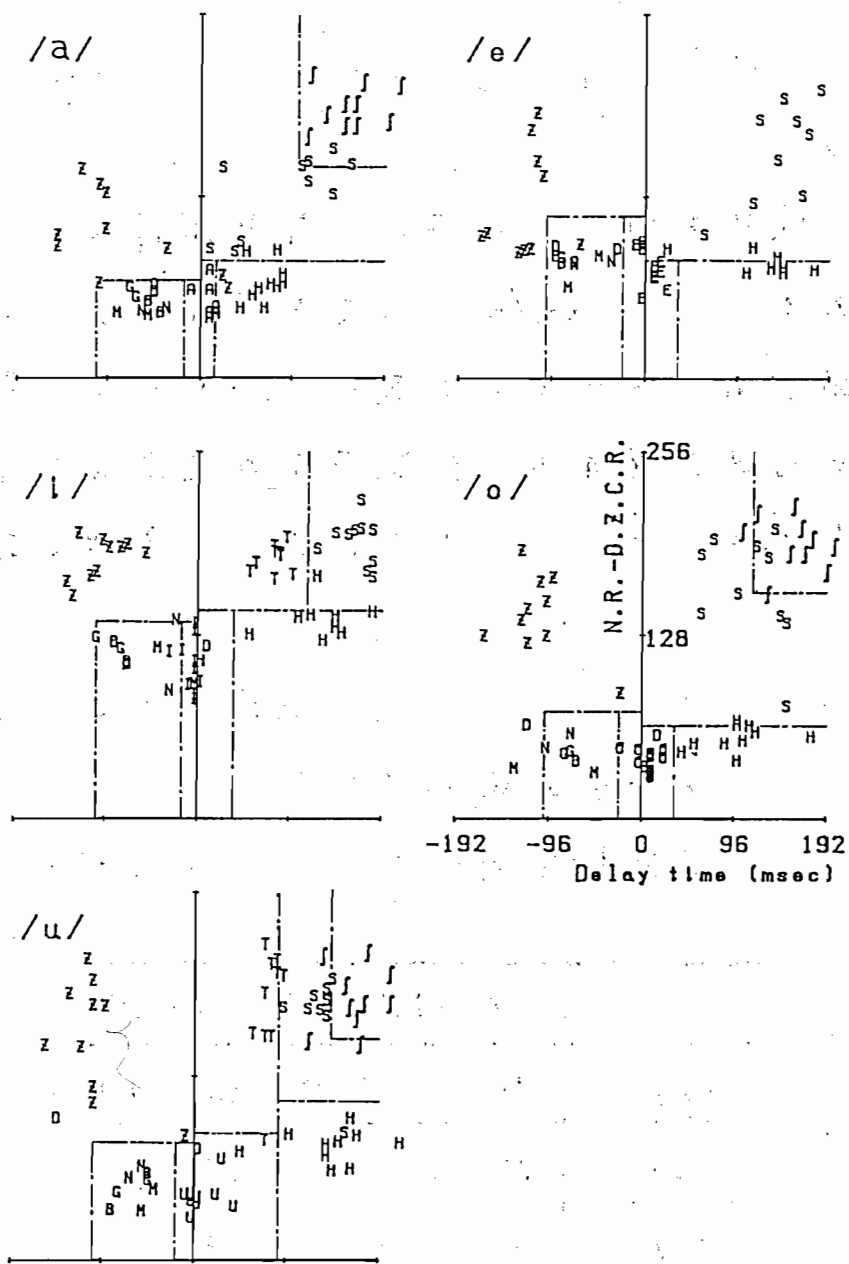


図4-6 N-D平面上的各音韻の分布

パラメータは10名の男性の孤立発声
C.V音節から得られた。

(遮蔽板を用いない, 試験用データ)

t_{no} の推定を行なうことにした(図4-5 /ʃa/ (B))。

遮蔽板を用いない場合の音声試料について以上で述べた t_{no}, t_{mo} の推定を行ない, それらと試験用データとして識別実験を行なった。その結果を図4-6と10発声中の正解数を表わす表4-3に示す。これらによると全体の識別

表4-3 10名の男性が孤立発声したCV音節と母音の識別結果
(遮蔽板を用いない, 試験用データ)

	/a/	/i/	/u/	/e/	/o/
Vowel	10	9	10	10	10
/s/	7	10	9	10	7
/h/	8	8	10	8	9
/ʃ/	10	/	7	/	8
/z/	7	10	10	9	10
/tʃi/ /tsu/	/	10	8	/	/
/n/, /m/ /g/, /d/, /b/	10	8	8	10	6

率として 88.7% (266/300) が得られており、補正が適切であったことが示されている。

4.4 考察

(1) 2つの特徴パラメータと各単音節の関係について考察する。雑音除去差分零交叉数は摩擦音・破擦音の歯や唇の狭めによる乱気流の状態を測るパラメータであり、図4-4、図4-6によれば無声摩擦音・破擦音の中では /s/、(/s/, /tʃi/, /tsu/), /h/ の順に大きい傾向がみられる。又、図4-4によれば有声摩擦音 /z/ について /zi/, /zu/ の雑音除去差分零交叉数は大きい。一方 /za/, /ze/, /zo/ のそれが小さい音声試料が多くみられる。これは /zi/, /zu/ は子音部 /z/ から後続母音への変化の過程で歯の狭めの形状がほとんど変化しないため乱気流の発生が顕著である一方、/za/, /ze/, /zo/ では子音部 /z/ から口を開いた後続母音部へ移るため乱気流の発生が時間的に短かく、かつ十分でないことによると考えられる。そこで試験用

データでは発声者に /z/ を強調して発声する
ように指示も与えたところ，雑音除去差分零
交叉数は著しく増加し識別率の向上がみられ
た（図 4-6，表 4-3）。

次に，Delay time と各単音節の関係につい
て考察する。口腔出力が鼻腔出力より早く立
上がる無声摩擦音・破擦音では，Delay time
は子音区間の継続時間長を表わすパラメータ
であり，/s/，(/s/，/h/)，(/tʃi/，/tsu/) の順に長
い傾向がみられる。有声子音の中では，鼻音
/m/，/n/ の Delay time は子音の継続時間長を表
わしていると言えすが，他の子音については
不適當な表現である。又，/za/，/ze/，/zo/ では
Delay time の絶対値が他の有声子音よりも大
きい傾向がみられる（図 4-4，図 4-6）。図
4-4，図 4-6 のデータはマイク M を唇から 3
cm の距離に配置して入力しているが，ここで，
マイク M と唇の距離が Delay time にどのような影
響を与えるかについて検討する。図 4-4，
図 4-6 には /ha/ の Delay time が一般に小さく，

/hu/ に較べて母音との分離がよくないことが示されている。これは /ha/ は /hu/ よりも唇を大きく開いて発声されるため子音部 /h/ における肺からの気流の速さが小さく、マイク M が気流を観測しにくいことによる。そのため (/hi/, /hu/), /ho/, (/ha/, /he/) の順に Delay time が大きい傾向がみられる。そこで、マイク M を唇の直前に接近させれば図 4-4, 図 4-6 における右半平面に分布する子音は右方向へ移動し、母音と /h/ の識別率が改善されることが予想できる。しかし、左半平面に分布する子音については改善されない。

(2) この章で扱わなかった音素、すなわち、無声破裂音 /p/, /t/, /k/, 半母音 /y/, /r/, /w/, 撥音 /N/ の N-D 平面上での分布について述べる。図 4-4 に /pi/, /ki/, /pu/, /ku/ を各々 5 個ずつプロットしている。他の無声破裂音もほぼ同位置に分布することを実験で確かめたが、図を見やすくするために省略した。前述したように図 4-4, 図 4-6 の平面上での /ha/ の分

布は母音の分布に接近しており，しかも無声破裂音の分布がそれらと重なるため識別が不可能となってしまう。半母音についても母音の分布と重なって分布し，又，口腔出力が存在しない /N/ では平面上にプロットすることができない。そこで，無声破裂音や半母音・撥音はこの章で述べた処理以前に他の処理によって分離しておく必要がある。このことについては第5章で詳しく述べる。

4.5 むすび

この章では，二重入力システムによる摩擦音・破擦音の識別について述べた。第1の特徴パラメータとして摩擦音・破擦音の歯や唇の狭めによる乱気流の状態を表わす雑音除去差分零交叉数を用い，第2の特徴パラメータとして雑音除去差分零交叉数の立上り時刻（または口腔出力エネルギーの立上り時刻）と鼻腔出力エネルギーの立上り時刻の時間差を用いた。そして，これら二つの特徴パラメー

夕が張る平面上に摩擦音・破擦音・母音等、
7つの音素群をプロットすることにより識別
実験を行なった。その結果、遮蔽板を用いた
設計用データについて90.0% (270/300)の識別
率を得た。次に遮蔽板を取り除いた場合、2
個のマイクに鼻腔出力と口腔出力が混入する
問題について述べその補正法を検討した。試
験用データは遮蔽板を使用しないで計算機に
入力し、この補正を施し識別実験を行なった。
その結果、88.7% (266/300)の識別率を得て、
これら2つの特徴パラメータと補正法が有効
であることを確認した。

第5章 音声波二重入力システムによる 音韻認識アルゴリズム

5.1 はじめに

第3章では D-S 平面と O-Δ 平面を用いて、後続母音に /a/エ め つ 16種の音韻および /N/を6群に分類するアルゴリズムを示した。又、第4章では N-D 平面を用いて摩擦音・破擦音を識別するアルゴリズムを示した。この章ではこれらも統一的に扱うアルゴリズムについて述べる。最初に、5.2節において後続母音に /a/エ め つ 16種の音韻と /N/を識別するアルゴリズムを示し、次に、5.3節においてこのアルゴリズムによる遮蔽板を用いた識別実験と遮蔽板を用いない識別実験を行ない、この識別アルゴリズムの有効性を確かめる。最後に、5.4節において二重入力システムを音声タイプライターとして用いることの有効性と実用性について述べる。

5.2 音韻認識アルゴリズム

本節では D-S 平面・O-Δ 平面・N-D 平面を用いて 17 種の音韻 /a/, /ka/, /sa/, /ta/, /na/, /ha/, /ma/, /ya/, /ra/, /wa/, /pa/, /ʃa/, /za/, /ga/, /da/, /ba/, /N/ を識別するアルゴリズムを示す。第 3 章の O-Δ 平面上では /ʃa/, /za/, /ga/, /da/, /ba/ を扱わず、第 4 章の N-D 平面上では /ka/, /ta/, /pa/, /ya/, /ra/, /wa/, /N/ を扱っていない。この理由はこれらの音韻が次のような分布をすることが実験で確かめられているからである。O-Δ 平面上での /ʃa/ の分布は /sa/ と同じであり、/za/, /ga/, /da/, /ba/ の分布は /na/ に類似しているが分布は広い。N-D 平面上での /ka/, /ta/, /pa/ の分布は /a/ と /ha/ の間にあり、/ya/, /ra/, /wa/ の分布は /a/ と同じである。従って、O-Δ 平面で /ʃa/, /za/, /ga/, /da/, /ba/ を扱うこと、及び N-D 平面で /ka/, /ta/, /pa/, /ya/, /ra/, /wa/ を扱うことは識別結果を悪くする。又、/N/ は口腔出力が存在しないため N-D 平面上にプロットすることができない。以上の理由で、各平面

/a/, /ka/, /sa/, /ta/, /na/, /ha/, /ma/, /ya/, /ra/, /wa/
 /pa/, /ʃa/, /za/, /ga/, /da/, /ba/, /N/

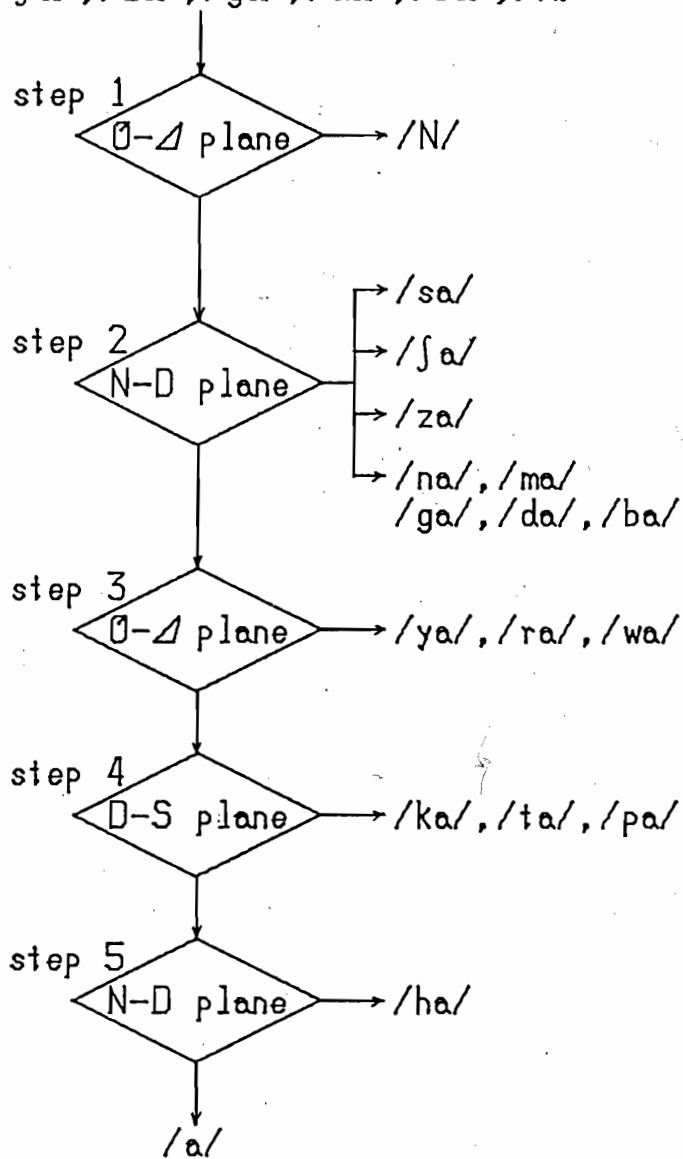


図 5-1 17種の音韻を識別するための
 フローチャート

上では限定した音韻のみを扱う。この制約を補うため以下に述べるアルゴリズムでは各平面での識別を複数回行なう。

図5-1が識別フローチャートである。step 1ではO-Δ平面上で/N/を分離識別する。これはO-Δ平面上での/N/の分布が非常に顕著であり、最初に他の音韻から分離しておくことが適切であることによる。step 2ではN-D平面上で/sa/, /ʃa/, /za/, (/na/, /ma/, /ga/, /da/, /ba/)の4群を分離識別する。step 3では再びO-Δ平面上で(/ya/, /ra/, /wa/)の1群を分離識別する。step 4ではD-S平面上で(/a/, /ha/)と(/ka/, /ta/, /pa/)の分離を行ない、step 5では再びN-D平面上で/a/と/ha/の識別を行なう。以上の5段階の処理によって17種の音韻を9群に分類する。このように5段階の構成をとる理由は各平面上で他の音韻群から顕著に分離している音韻を先に分離識別する方法を採用していることによる。一方、後続母音が/e/, /o/の場合も原則的に図5-1のフロ

一チャートで識別できる。しかし、後続母音が /i/, /u/ の場合には破擦音 /tʃi/, /tsu/ を step 2 で分離する。又、3.5 節 (3) で述べたように 0- Δ 平面上での /mi/, /mu/, /yu/ 等の特徴が不明確であり、後続母音が /i/, /u/ の場合については今後の検討を要する。

5.3 識別実験

第3章・第4章では遮蔽板を用いた場合と遮蔽板を用いない場合のそれぞれについて、10名の男性が孤立発声した音声試料を用いて識別実験を行なった。本節ではこれらと同一の音声試料に図5-1の識別アルゴリズムを適用し識別実験を行なうが、第3章と第4章を統一するため、遮蔽板を用いた音声試料を設計用データ、遮蔽板を用いない音声試料を試験用データとする。音声試料は各音韻について各々10個を用いるが、図を見やすくするため音韻によっては図中に10個すべてをプロットしていないものがある。

5.3.1 遮蔽板を用いた識別実験

設計用データとして図3-2, 図3-5, 図4-4の遮蔽板を用いた音声試料を採用する。図5-1における各stepでの誤識別も次に示す。

表5-1 設計用データの識別結果

(識別率 94.7%)

出力 入力	/a/	/ka/ /ta/ /pa/	/sa/	/na/ /ga/ /da/ /ma/ /ba/	/ha/	/ya/ /ra/ /wa/	/ʃa/	/za/	/N/
/a/	9				1				
/ka/ /ta/ /pa/		30							
/sa/			8		2				
/na/ /ga/ /da/ /ma/ /ba/				50					
/ha/	2				8				
/ya/ /ra/ /wa/	3					27			
/ʃa/							10		
/za/				1				9	
/N/									10

step 1	図 3-5	誤識別なし
step 2	図 4-4	/sa/ の 2 試料 /za/ の 1 試料
step 3	図 3-5	/wa/ の 3 試料
step 4	図 3-2	誤識別なし
step 5	図 4-4	/a/ の 1 試料 /ha/ の 2 試料

従って，170個の音声試料に対して9個の誤識別があり，94.7%の識別率が得られた。このコンフュージョンマトリックスを表5-1に示す。

5.3.2 遮蔽板を用いない識別実験

試験用データとして図3-9, 図3-11, 図4-6の遮蔽板を用いない音声試料を採用する。識別境界は各々図3-5, 図3-2, 図4-4と同一の境界を事前に設定する。図5-1における各stepでの誤識別も次に示す。

step 1 回 3-9 誤識別なし

step 2 回 4-6 /sa/ の 3 試料

/za/ の 3 試料

表 5-2 試験用データの識別結果

(識別率 90.6%)

出力 入力	/a/	/ka/ /ta/ /pa/	/sa/	/na/ /ga/ /da/ /ma/ /ba/	/ha/	/ya/ /ra/ /wa/	/ʃa/	/za/	/N/
/a/	10								
/ka/ /ta/ /pa/	2	27				1			
/sa/			7				3		
/na/ /ga/ /da/ /ma/ /ba/				50					
/ha/			2		8				
/ya/ /ra/ /wa/	5					25			
/ʃa/							10		
/za/				1	2			7	
/N/									10

step 3	図 3-9	/ka/ の 1 試料
		/ra/ の 2 試料
		/wa/ の 3 試料
step 4	図 3-11	/pa/ の 2 試料
step 5	図 4-6	/ha/ の 2 試料

従って、170 個の音声試料に対して 16 個の誤識別があり、90.6% の識別率が得られた。このコンフュージョンマトリックスを表 5-2 に示す。

5.4 実用性についての考察

二重入力システムは入力方法が制約される欠点をもつが、計算機と高速に情報の入出力を必要のあるオペレータが音声によって計算機と対話し、また大量の文章を高速に作成する新聞記者が音声タイプライターとして用いる等の特殊な用途ではこの二重入力システムは実用的であると考えられる。二重入力システムでは上記の欠点をもつ反面、2 個の

マイクの音声波形およびそれら相互の関係から得られる情報量は従来の1個のマイクから得られる情報量の2倍以上であるので、計算量の減少・メモリーの節約・認識時間の短縮・認識率の向上という長所をもっている。

次に、マイクの位置ずれの影響について述べる。解剖学的構造上、鼻腔出力は下方に、口腔出力は体の前方に最も大きい音声波エネルギーが流れるためマイクNを上方に向けて配置すれば1~2cm程度の位置ずれは識別結果にはほとんど影響を及ぼさないことが実験によって確かめられている。マイクMは/p/, /h/等の無声音の発声時に口腔から出る気流を十分に観測できる配置を必要とする。口の前で手のひらを広げた時、この気流を感じる領域内にあれば、口の前方に距離が2~3cm増加しても同時に鼻からの距離も増加するため鼻腔出力の混入の程度は小さい。又、最適位置からのずれを防ぐ方法としては頭部固定方式として装着することが考えられる。

本論文では直径 8.5 mm, 長さ 15 mm の小型マイクを用いて実験を行なったが, 最近ではマイクの小型化が進み, 直径 6 mm, 長さ 10 mm 程度のマイクも出現しており, 十分に実用的なシステムが実現できると考えられる。

従来の研究においては調音結合の影響のため連続音声の中の音素認識は非常に困難であるとされてきた。本論文では孤立発声 CV 音節に限定して識別実験を行なったが, 二重入力システムは連続音声の中の音素認識にも有効性を発揮することが期待できる。それは 2 個のマイクのモニター用オシロスコープによって 2 つの波形を観察しただけでも各音韻の特徴が十分に把握できることによる。この点については今後に残された研究課題である。このように音声波の鼻腔出力と口腔出力を分離ピックアップして処理する方式の研究は始められたばかりであり, 今後, 多くの課題を解決しなければならぬが実用的な音声タイプライターへの一番の近道であると考えられる。

5.5 む す び

この章では D-S 平面・O-Δ 平面・N-D 平面を用いて後続母音に /a/ε/ 16 種の音韻および /N/ε/ 9 群に分類するアルゴリズムを示し、遮蔽板を用いた音声試料と遮蔽板を用いない音声試料を各々設計用データ・試験用データとして識別実験を行なった。その結果、設計用・試験用の各々について 94.7% (161/170), 90.6% (154/170) の識別率を得て、その有効性を確かめた。最後に、二重入力システムを音声タイプライターとして用いることの有効性と実用性について述べた。

この二重入力システムでは日本語 68 種の単音節を、最初に計算量の少ない時間領域において大分類し、次に周波数領域で最終的な識別をする方法を採用しており、これによって計算量の圧縮と識別率の向上を目指している。この章では時間領域での分類について述べたが、周波数領域での識別については今後の課題として残されている。

第6章 オルゴール生産工程における 曲目の自動認識⁽³¹⁾

6.1 はじめに

近年、計算機による音楽情報処理の研究が活発に行なわれ、楽器音の分析を目的とした研究^{(32),(33)}や楽音の自動採譜を目的とした研究^{(34)~(36)}等がなされてきた。これらの中の多くの研究では音の周波数とその時間変化をFFTによって分析する方法を用いている。本研究はオルゴール生産工場における製品の音の検査や曲目別箱詰め工程の自動化を目的として、オルゴール音の分析と符号化、そして曲目の認識実験を行なったためのである。現在、オルゴール生産工場では異なる曲のドラムと振動板の組合せによる不良品出荷や箱詰め時の曲目誤まりの出荷を防ぐため、最終工程で人間の耳による検査を行なっている。オルゴールは数千種の曲目が生産され、また新曲も現われるため検査は熟練を要し、作業能率も悪い。

このためこの工程を自動化することが望まれている。

この章では、異なる4種の曲目の製品が混在する生産ライン上で不良品の分離と曲目別箱詰め判断をするシステム（入力処理時間5秒以内・不良品出荷率0%・曲目誤り出荷率0%）を目標として行なった基礎実験の結果を報告する。

6.2 曲目自動認識システムの構成

図6-1にオルゴールの構造を示す。オルゴールはドラム上の突起（pin）が振動弁（comb

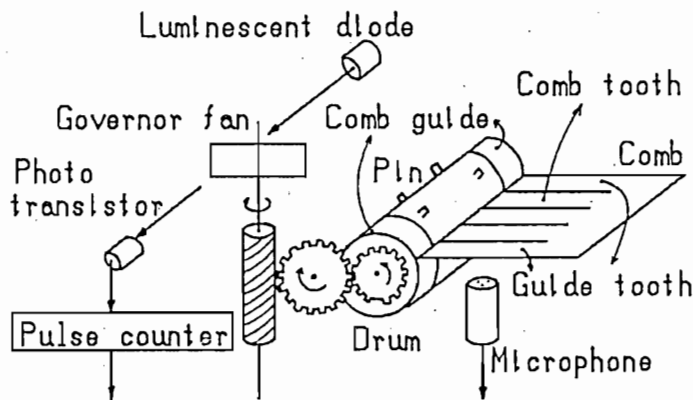


図6-1 オルゴールの構造

tooth)をはじめくことによつて音波を発生する。この音波を高いSN比で分析するため小型マイクロホン (ECM-150, SONY製) を振動板 (comb) の下方2cmの位置に固定する。ドラムは約12~26秒で1回転し, 同時に鳴る音を一音として1回転で40~50音程度鳴る。又, オルゴールは風切り (governor fan) と呼ばれる回転板をもち, 標準タイプのオルゴールではドラム1回転につき風切りは1412回転する。本研究では風切りの回転数を測定することによりドラムの回転角度を知る方法を採用した。図のように発光ダイオードとホトトランジスタ間の

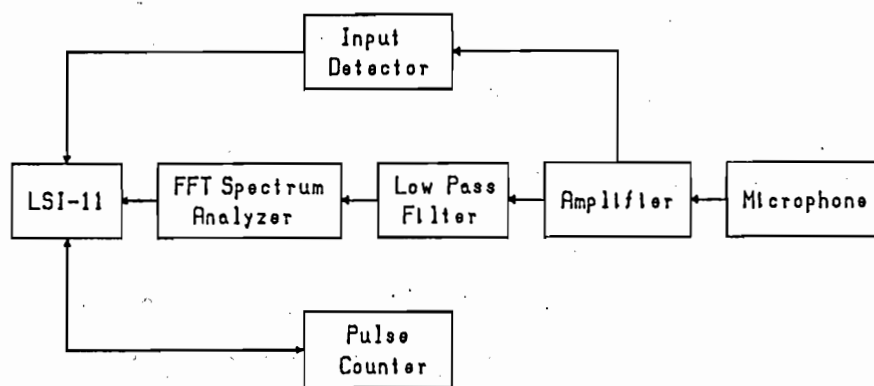


図6-2 曲目自動認識システムの構成

風切りによる遮蔽を検出し，パルスとして出
 力すれば風切り1回転につき2個のパルスを
 出すので風切りの回転数はパルスカウンタで
 計数して知る事ができる。図6-2にマイク

ロホンとパルスカウ
 ンタからの情報を処
 理するシステムの構

成を示す。マイク
 ロホンから入力した
 音波は増幅器とロー
 パスフィルタ（遮断
 周波数5kHz）を通り
 FFTリアルタイム・
 スペクトルアナライ
 ザ（model 3112,日本科
 学工業製，以下FFT
 アナライザと略す）
 にフルスケール±3.16
 Vで入力される。オ
 ルゴールの音は表6-1

表6-1 音程番号と周波数の関係

Sound step number	Frequency (Hz)	Note
1	311.1	D3#
2	329.6	E3
3	349.2	F3
4	370.0	F3#
5	392.0	G3
6	415.3	G3#
7	440.0	A3
8	466.2	A3#
9	493.9	B3
10	523.3	C4
11	554.4	C4#
12	587.3	D4
13	622.3	D4#
14	659.3	E4
15	698.5	F4
16	740.0	F4#
17	784.0	G4
18	830.6	G4#
19	880.0	A4
20	932.3	A4#
21	987.8	B4
22	1046.5	C5
23	1108.7	C5#
24	1174.7	D5
25	1244.5	D5#
26	1318.5	E5
27	1396.9	F5
28	1480.0	F5#
29	1568.0	G5
30	1661.2	G5#
31	1760.0	A5
32	1864.7	A5#
33	1975.5	B5
34	2093.0	C6
35	2217.5	C6#
36	2349.3	D6
37	2489.0	D6#
38	2637.0	E6
39	2793.8	F6
40	2960.0	F6#

に示すように 311~2960 Hz の範囲に分布する
ため FFT アナライザは 5kHz のレンジを使用す
る。従って、FFT アナライザ内では音波を 10
kHz でサンプリングし、1024 個の波形データ
をフーリエ変換した後、512 個のマグニチュー
ードスペクトルが 12 bit で出力される。マグニ
チューードスペクトルはスペクトル成分の実部
a と虚部 b によって $\sqrt{a^2+b^2}$ で表わされる。
512 個の出力のうち 0 番地は直流成分、511
番地は 5kHz に対応し、周波数分解能は 9.8 Hz
である。本研究では 30 番地 293.5 Hz ~ 310 番地
3033.3 Hz のマグニチューードスペクトルをマイ
クロコンピュータ (LSI-11) に入力する。検査
工程においてはすべてのオルゴールを曲の最
初から鳴らせて計算機に入力する。その時、
曲の最初の空白区間の入力によるメモリの
浪費を防ぐため、曲の最初の音を検出しその
音から分析を始める構成をとっている。入力
検出回路として +1.4 V 以上の増幅器出力に対
して単安定マルチバイブレーションが 1 秒間 +5 V

を出力する回路を用い、計算機はこのフラグ
 が上がった後、FFTアナライザからデータ読
 み取りを開始する。以上の様に計算機はパ
 ルスカウンタ・FFTアナライザ・入力検出回
 路から情報を読み取り、またパルスカウンタ
 へリセット信号を送る。図6-3にFFTアナ
 ライザからデータを読み取るタイムチャート
 を示す。オルゴールのドラムは曲の最初の空白
 部分から回転を始め曲を演奏する。(1)計算
 機は曲の最初の音を検出した後、(2)パルス

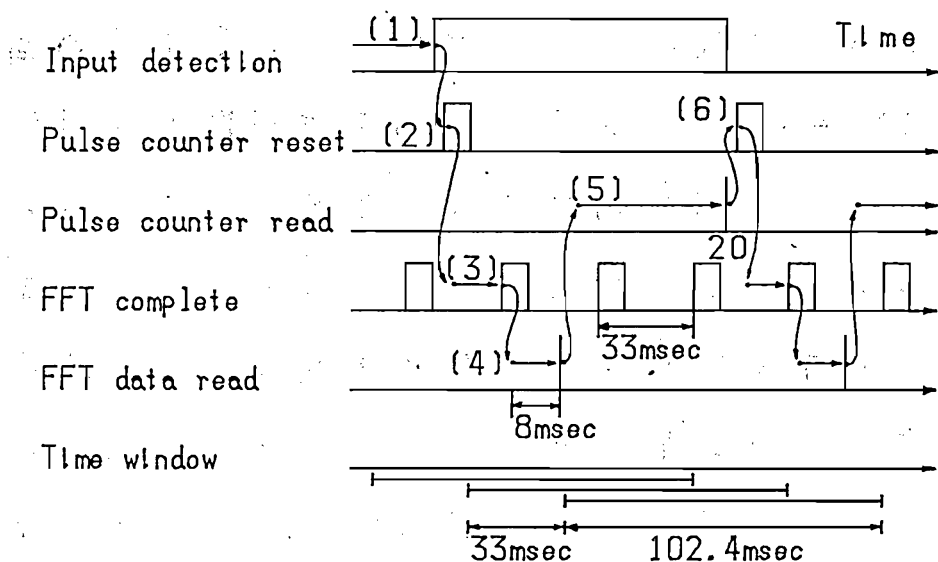


図6-3 FFTデータ読み取りのタイムチャート

カウンタリセット信号を出し，(3) FFT完了信号を待つ。完了信号検出後，(4) FFTデータを読み取る。ここで，変化するドラムの回転速度に影響されず一定の回転角度ごとにFFTデータを読み取るためにパルスカウンタの値を利用する。この値として，計算機の記憶容量やFFT変換時間，ドラムの回転速度を考慮して経験的に20カウント（回転角2.5度）を採用する。(5) パルスカウンタが20を計数するのも待つ。計数の後，(6) パルスカウンタへリセット信号を出す。なお，1回のFFT計算について波形のサンプリングに要する時間は102.4msecであるためFFTは時間窓を33msecごとに周期的にずらして連続して行われる。又，データ読み取りには8msecを要する。以下(2)~(6)のサイクルを繰り返し，1回のデータ読み取りを1フレームと呼ぶことにする。フレーム間の時間間隔はドラムの回転速度が遅い程長くなるが，実験では90~300msecの値になった。最後に，生産工場の騒音

については生産ライン上で簡易無響室を作る
ことが出来るため，本研究では通常の計算機
室内で実験も行なった。

6.3 オルゴール音波の特徴

オルゴール音波は実験の結果，以下のよう
な特徴をもっていることが分った。

(1) 個々の振動弁は正弦波減衰振動をし，分
析する周波数帯 (293.5 ~ 3033.3 Hz) で高調波成
分は小さい。

図 6-4 にオルゴール音波の波形の一例を示
す。矢印の区間の左側は先に鳴った減衰振動
音の効果が残っている部分である。矢印の区

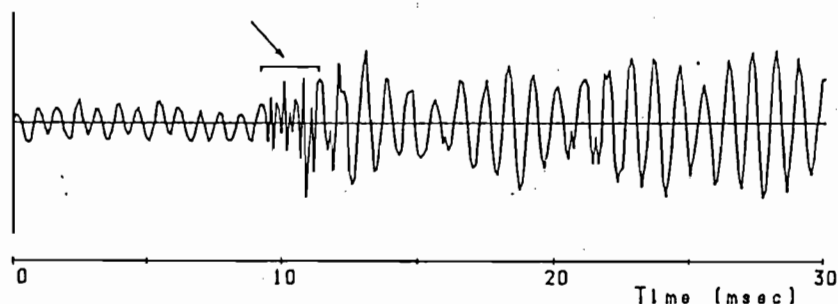


図 6-4 オルゴール音波の波形

間では次の音を鳴らすためドラム上のピンが振動弁に接触してきしみ音を出している。この区間の右側は2つの音が同時に鳴り、2つの正弦波減衰振動が重ね合わされている場合である。矢印の部分を含む区間でスペクトル分析した時、図6-5のように鋭い形のマグニチュードスペクトル(以下スペクトルと略す)(b)~(f)とすその広い形のスペクトル(a)

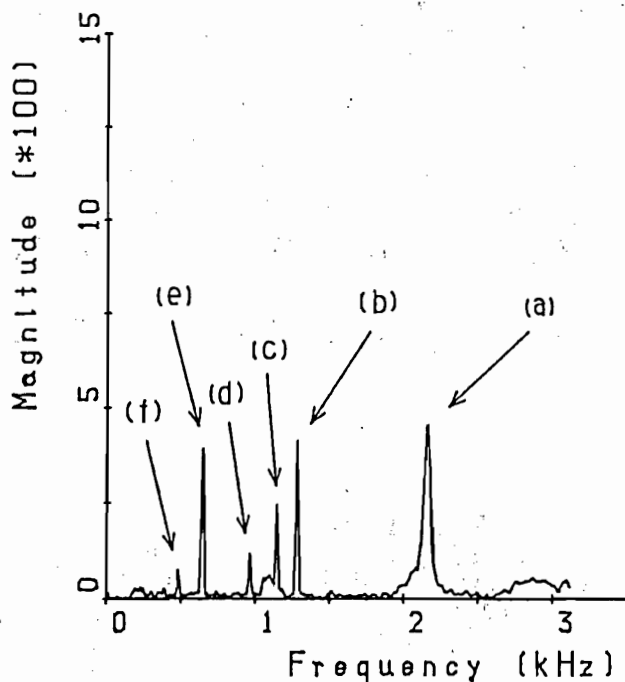


図 6-5 オルゴール音波のマグニチュード
スペクトル

の2種類が現われる。(a)はきしみ音を出して
いる振動弁による音のスペクトルに対応し、
(b)~(f)はすでにはじかれた振動弁による音の
スペクトルに対応する。

(2) スペクトルの時間変化は立上りが鋭く、
その後は緩やかに減衰する。

特定の周波数でのスペクトル強度の時間変
化の例を図6-6、図6-7に示す。図6-6は

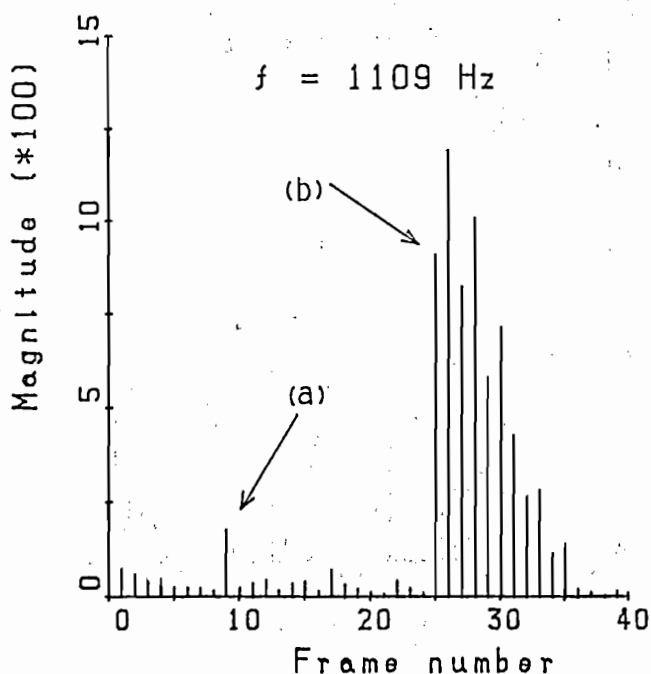


図6-6 中高音域のマグニチュード
スペクトルの時間変化

1109 Hz , 図 6-7 は 659 Hz の場合である。孤立音のスペクトル時間変化は図 6-6 矢印 (b) のように鋭く立上がり緩やかに減衰する。同一周波数の音が連続して鳴った場合, 図 6-7 のように前の音の減衰中に次の音が立上がる。矢印 (a) ~ (e) は音の立上り部分である。

(3) 図 6-6 , 図 6-7 の比較からわかるように一般に, 低音域 (300 ~ 900 Hz) のスペクトル

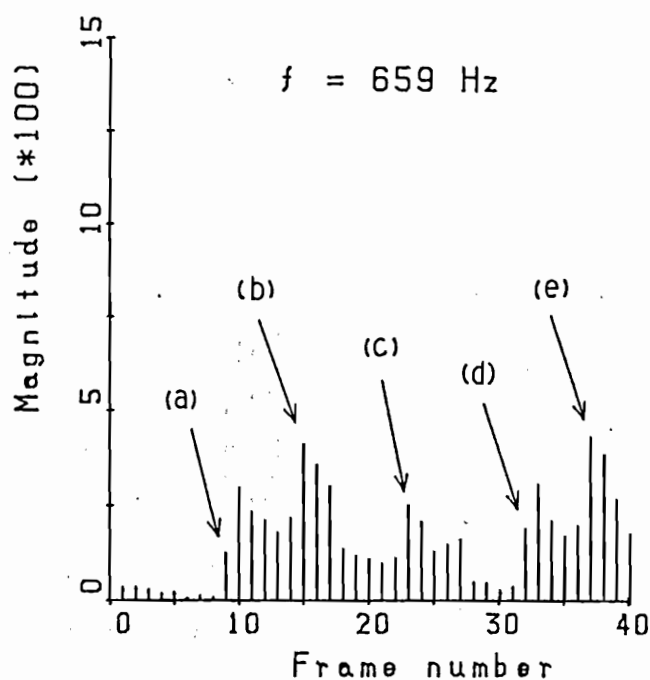


図 6-7 低音域のマグニチュード
-スペクトルの時間変化

の大きさは中高音域のそれに較べて小さい。

(4) 1つの周波数に注目した時(図6-7), 楽譜上の四分音符も八分音符もスペクトルでは同一の立上り特性と減衰特性をもつ。

(5) 1つの周波数に注目した時(図6-7), 連続音の場合, 前の音の減衰分が重なるがスペクトルの立上り時の増分は孤立音とほぼ同じである。

ピアノでは1音程に唯一つの弦が対応しているが, オルゴールでは振動弁が振動中にピンに接触することは音質を悪くするため同一音程の音を出す複数の振動弁が存在し, 同一音程の連続音を鳴らす場合にはこれらの振動弁を連続してはじく。

(6) 前項(5)で述べた同一音程の複数の振動弁が多少ずれた振動数をもつことがある。

(7) 図6-6矢印(a)のように孤立し, しかもマクニチュード700程度の大きな雑音が現われる。

原因として, 立上り時のきしみ音・風切り

や歯車の回転音・図6-1におけるガイド弁 (guide tooth) と振動板ガイド (comb guide) の摩擦音・室内騒音等が考えられる。以下ではこの雑音を孤立雑音と呼ぶ。

6.4 照合と識別実験

高速かつ高精度なシステムを作ることも目標とした照合・識別実験^(注)の方式を、データ入力法・スペクトルマッチング法・符号マッチング法の順に説明し、その結果を合わせて述べる。

(注) 本章では曲目照合と曲目識別の実験を行なうが、照合とはあらかじめ登録されている曲目かどうかを判定するものであり、識別とはあらかじめ登録されている曲目のうちのどれであるかを判定するものである。照合と識別を合わせて認識という。

6.4.1 データの入力法

FFTアナライザから計算機に入力されたデータ系列を

$$\{w_{ij}\} \quad (i=1\sim 320, j=1\sim L) \quad (6-1)$$

とする。Lは入力データ長(フレーム数)である。本研究では入力時間(2秒程度)・処理時間(3秒以内)・記憶容量・鳴る音の数および予備認識実験の結果によりL=20とする。

オルゴールは表6-1に示す40種の音程より構成されている。これらの音程に表6-1のよ
うに番号をつけ、これを音程番号(sound step number)と呼ぶ。オルゴールの場合、音の周波数は表6-1に示したもののから±15セント⁽²⁶⁾の誤差が許されている。これは人間がこの誤差内の音を聞き分けることができないためである。従って、オルゴールの高音では±26Hzの誤差が生じることもある。そこで、(中心周波数)±(許容誤差)の範囲内の音を1つの音程に対応させなければならない。本研究で

はFFTアナライザの出力番地30~310を40区間に分割し、この区間番号を改めて*i*とする。第*j*フレーム第*i*区間内のスペクトルの最大値をその区間の代表値 W_{ij} とする。40区間の分割境界は経験的に決定した。すなわち、

$$W_{ij} = \max_k w_{kj} \quad (i=1\sim 40, j=1\sim L)$$

(*k*は第*i*区間内の番地)

(6-2)

である。

ところで、ドラムの回転角に同期してデータを読み込むため、音の立上りは回転速度によらず時間軸方向で正規化される。しかし、各振動弁の振動の減衰時間はドラムの回転速度によらず一定であるため、回転が速い程見かけ上、 W_{ij} は緩やかに減衰するように見える。従って、 W_{ij} の減衰曲線の形状には重要な情報は含まれず、また各フレームにおけるスペクトル分布の形状も回転速度によって変わる。

本研究で用いたサンプルは通常の計算機室で入力した。内訳は同一曲目のオルゴール5

個(曲目 R とする)を各10回,異なる17曲(曲目 $D_1 \sim D_{17}$ とする)を各3回,異なる曲のドラムと振動板の組合せによる不良品10個を各1回入力し,計111サンプルを用意した。標準パターンは曲目 R の50サンプル中,最初に入力したものである。これらの曲については低音から高音まで音の分布が広い曲,この分布が偏った曲,三連符を含むテンポの速い曲,テンポの遅い曲など多様なものを選んだ。

6.4.2 スペクトルマッチング法

前節で述べた理由により,スペクトルの形状に着目したマッチング法には欠点が考えられるが,ここで簡単で高速なスペクトルマッチング法を試みる。先ず,式(6-2)の W_{ij} を式(6-3)により正規化して ${}_N W_{ij}$ を得る。

$${}_N W_{ij} = \frac{W_{ij}}{\max_{i,j} W_{ij}} \quad (i=1 \sim 40, j=1 \sim L)$$

(6-3)

正規化した標準パターンを $\{N R_{ij}\}$, 未知パターンを $\{N X_{ij}\}$ とし, 単位フレーム当りの距離 Distance を式 (6-4) で定義する。

$$\text{Distance} = \sum_{i=1}^{40} \sum_{j=1}^L |N R_{ij} - N X_{ij}| / L \quad (6-4)$$

これを 111 個のサンプルに適用し照合実験を行なった結果を図 6-8 に示す。曲目 R のうち 3 サンプルが曲目 D₁~D₁₇ と同程度の距離になっているが, これらは雑音の混入により入力を誤検出したためであった。図 6-8 は分離度

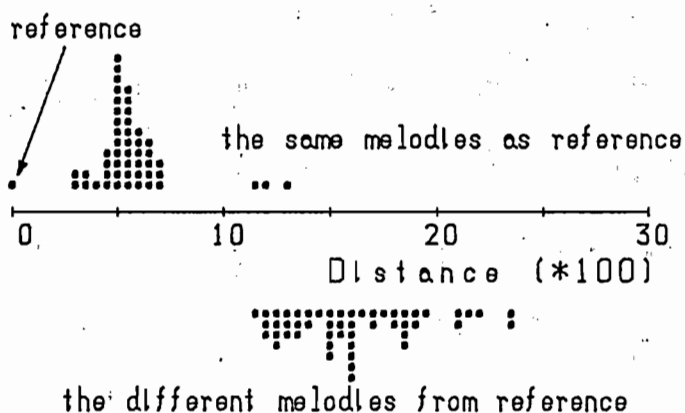


図 6-8 スペクトルマッチングによって得られたパラメータの分布

の良い例であるが，このスペクトルマッチング法では曲によって分離度が変わり，結果が不安定になる。この理由は曲目によって第1フレームまでに鳴っている音の数はばらつきがあり，しで割るだけでは正規化できないことによる。以上で述べたように，スペクトルマッチング法では満足できる結果は得られず，これらの欠点を補うために次節で符号マッチング法について述べる。

6.4.3 符号マッチング法

入力データ系列 $\{w_{ij}\}$ を鳴った音の音程番号と直上り時のフレーム番号の2つのパラメータを用いて符号化する。その結果を用いてマッチングを行なう方法を符号マッチング法と呼ぶことにする。

符号化フローチャートを図6-9に，符号マッチングフローチャートを図6-10に示す。図6-9(a)(b)は6.4.1節で述べた入力部である。(c)は孤立雑音の除去部であるが，これについ

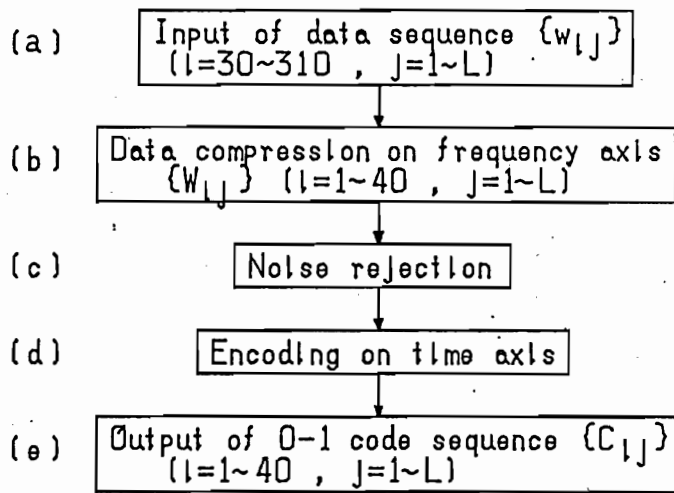


図 6-9 インカデータ系列の符号化フローチャート

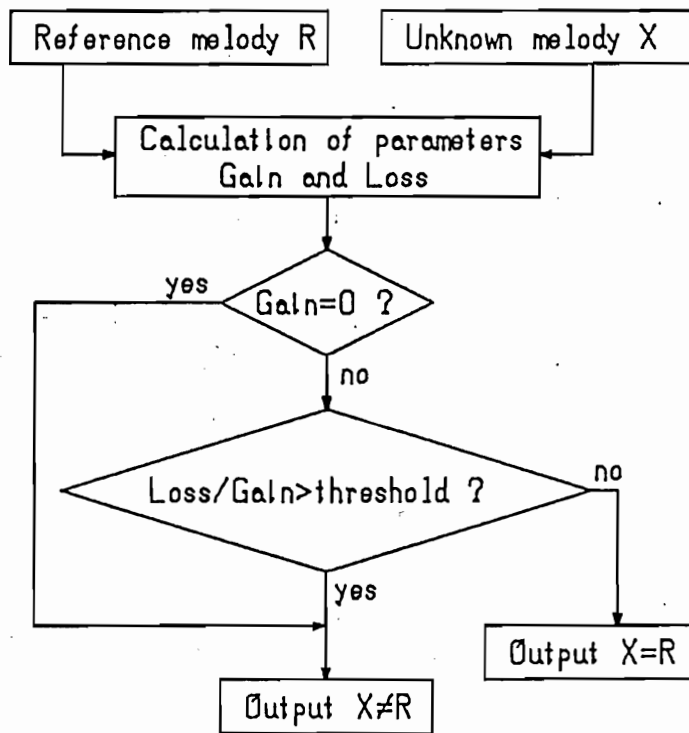


図 6-10 符号マッチングフローチャート

ては後述する。(d)は符号化部であるが、ここの処理は音程番号ごとに行なう。すなわち、図6-6、図6-7のようなデータ $\{W_{ij}\}$ に対して処理を行なう。オルゴール音波のスペクトルは鋭い立上りと緩やかな減衰特性をもち、差分値 δW_{ij}

$$\delta W_{ij} = W_{i,j+1} - W_{ij} \quad (6-5)$$

の値により立上りを検出することができ。すなわち、立上りの位置で式(6-5)の値が大きくなる。又、このようにして得た差分値のうち

$$\delta W_{ij} < \frac{\max_{j=1 \sim L-1} (\delta W_{ij})}{3} \quad (6-6)$$

を満たさぬのは雑音として除去する。これはオルゴールの場合、同一音程番号の音のレベルに大差がないことと利用している。更に、図6-7矢印(a)のように2段階で立上がる場合、差分値は連続して大きな値をとるが、フレーム番号の小さい方のみ有効とする。

このようにして音の立上り位置を検出すると孤立雑音も同様に符号化してしまう。そこで、図6-9(c)に示した雑音除去を符号化以前に行なう必要がある。先ず、大きな孤立雑音は孤立性を考慮して

$$\left. \begin{array}{l} W_{ij} > 5 \times W_{i,j-1} \\ W_{ij} > 5 \times W_{i,j+1} \end{array} \right\} \quad (6-7)$$

の両条件を満たすものを除去する。次に、移動平均値 $\overline{W_{ij}}$ が

$$\overline{W_{ij}} = (W_{i,j-1} + W_{ij} + W_{i,j+1}) / 3 \quad (6-8)$$

$$\overline{W_{ij}} < \theta \quad (6-9)$$

を満たす場合、雑音区間とみなし差分値の計算を行わない。但し、 θ は式(6-7)の条件で除去できない小さな孤立雑音およびその他の雑音を除去するために経験的に決定したしきい値である。

このような雑音処理を行なった後に符号化すれば、図6-11に示すような良好な結果が得

られる。図6-11の●と○が符号化の結果得られたもので、●は正しく符号化され、○は1フレームずれて符号化されたものである。又、×は本来符号化されるべき位置で符号化されていないそれを示している。符号化されなかったものは4か所あり、また雑音は完全に除

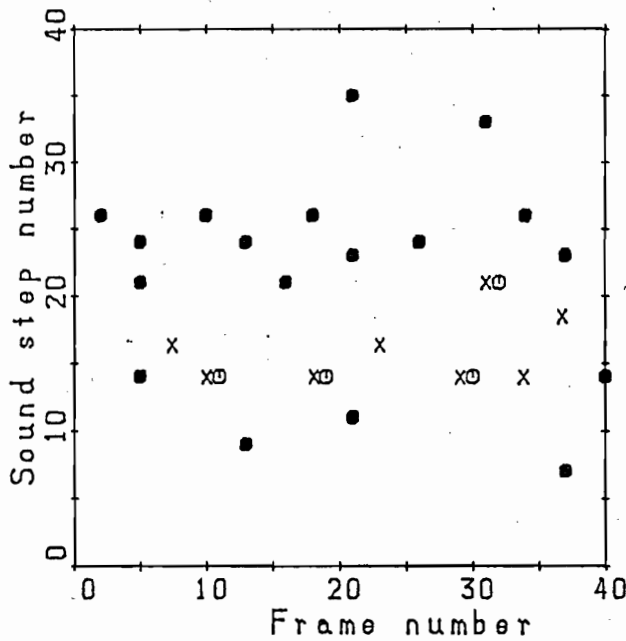


図6-11 符号化の一例

- , ○印はこのシステムにおいて符号化されたもの
- 印は正しい結果
 - 印は誤まった結果
- ×印は符号化されるべき位置で符号化されなかったもの

去できた。図6-9(e)では，図6-11の●と○に
対応する位置に1，他の位置には0を与え符
号パターンを出力する。

次に，図6-10に示している符号マッチング
法について述べる。オルゴール音は振動板の
ずれ，ドラムの回転速度によって，あるいは
符号化の処理の方法によって1フレームずれ
て符号化されることが多い。図6-11において
も4か所でその例がみられる。そこで，マッ
チングを行なう時，符号標準パターンと符号
未知パターン上の1の位置が1フレームずれ
ていても同じ位置にあるものとみなすことに
する。事実，このような処理を行なっても，
本来違った位置にある1を同一のものとみな
すことはなかった。ここで，

Gain : 符号標準パターンと符号未知パタ
ー
ン上の同じ位置にある1の数
(1フレームずれたものも含める)

Loss : 符号標準パターンと符号未知パタ
ー
ン上の異なる位置にある1の数

を定義する。

パラメータ $Loss/Gain$ を前述のサンプルに適用した結果を図 6-12 に示す。图中、右端の点は $Gain=0$ のため $Loss/Gain = \infty$ となったもので、曲目 R のうち 3 個にみられるがこれらはスペクトルマッチングで述べたものと同じく、入力を誤検出したためのものである。しかし、入力を誤検出した場合の符号パターンは、最初の数フレームに 1 が現れられないので容易に判別できる。 $Loss/Gain \leq 2$ のとき曲目 R であるとした場合の判定結果を表 6-2 に示す。この実験に用い

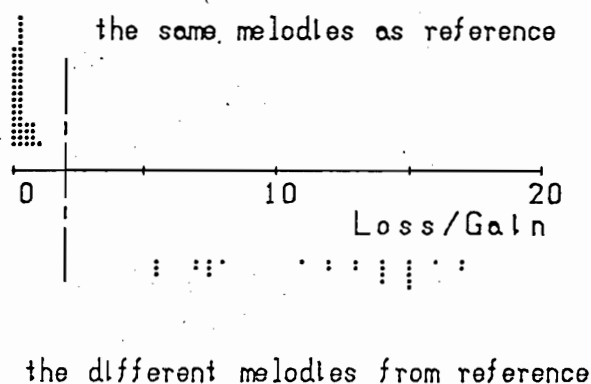


図 6-12 符号マッチングによって得られた
パラメータの分布

た符号標準パターンは、楽譜との照合により
 0を1と誤って符号化したものが2か所、1
 を0と誤って符号化したものが2か所含まれ
 ていることが確かめられた。標準パターンと
 して楽譜とおりのものを使用すれば、より良
 い結果が得られることも確認した。

次に、識別実験について述べる。これは識
 別する曲それぞれに対し前述の照合を行なう
 ことによつて実行できる。各識別境界は前述
 の照合実験と同じものを用い、識別曲目によ
 らず一定の値を使用した。又、識別実験では
 入力データを符号化した後、フレーム番号1
 ~5の区間に1が

現われないものは
 雑音による入力誤
 検出が起つたサン
 プルとして再入力
 した。(再入力し
 たものは176サン
 プル中8サンプル

表6-2 符号マッチングによる
 照合実験の結果

output melody input melody	R	rejected
R	47	3
D ₁ ~ D ₁₇		51
inferior quality		10

であった。) 識別実験に使用したサンプルは、4つの識別曲目 R_1, R_2, R_3, R_4 について各5個、その他の曲目 $D_1 \sim D_{14}$ は各1個、不良品サンプルとして10個、計44個のオルゴールを使用し、それぞれ4回入力した。標準パターンは曲目 $R_1 \sim R_4$ のオルゴール各5個の中から任意に1個ずつ選び、別に入力して作成した。

実験結果を表6-3に示す。識別結果は不良品出荷率・曲目誤り出荷率とも0%を得て、符号マッチング法の有効性を確かめた。

表6-3 符号マッチングによる識別実験の結果

input melody \ output melody	R_1	R_2	R_3	R_4	rejected
R_1	20				
R_2		20			
R_3			20		
R_4				20	
$D_1 \sim D_{14}$					56
inferior quality					40

6.5 むすび

計算機による音楽情報処理の一研究としてオルゴール音の分析と符号化および曲目の識別実験を行なった。分析の結果、オルゴールの個々の音は正弦減衰振動波であるが、立上り時のきしみ音・風切りや歯車の回転音・ガイド弁と振動板がイドの摩擦音・室内騒音等が混入し、符号化に大きな影響を及ぼすことが確かめられた。これらの雑音を除去し、オルゴール音を楽譜に最も一致するように符号化するアルゴリズムを考え、次に、これによって得られる符号を用いた符号マッチング法を考案し、新しいパラメータ Loss/Gain による曲目の識別実験を行なった。その結果、18種の曲目と、異なる曲のドラムと振動板の組合せをもつ10個の不良品を含む176サンプルについて、不良品はすべて正しく棄却され、また良品の4.5%が棄却されたがこれは入力状態が悪いサンプルであり再入力によって良品は100%正しく曲目を識別された。処理時

間については、音の入力から判定の出力まで
5秒以内の結果を得てこのシステムの有効性
を確かめた。

この章では、オルゴール生産工場における
最終工程の自動化を目的として行なった基礎
実験について述べた。以上の結果から十分に
実用性のある事が確かめられた。

第7章 結 論

本論文では音波波形情報処理の分野における新しい工夫の研究について述べた。第1の研究では2個の小型マイクロホンによって、音声波の鼻腔出力と口腔出力を別々に、しかも同時に計算機に入力するシステム(二重入力システム)を用いた音素認識の手法を示した。第2の研究では、オルゴール生産工場における製品の音の検査や曲目別箱詰め工程の自動化を目的とした曲目認識の手法を示した。

第1の研究では従来の音声認識の手法が認識率・認識時間・計算機の記憶容量等の面で限界に達していることを踏まえ、二重入力システムによってこれらの問題を解決することを目指した。最初に、二重入力システムの構成を示し、次に計算量の少ない時間領域の計算だけで、孤立発声CV音節を無声破裂音や鼻音等の音素群に大分類する手法と、子音・母音区間を高精度に分離する手法について述

べた。又、摩擦音・破擦音の特徴を表わす時間領域の種々のパラメータを検討し、その帰結として雑音除去差分零交叉数を導入した。そして、口腔出力の雑音除去差分零交叉数の立上り時刻と鼻腔出力エネルギーの立上り時刻との時間差を特徴パラメータの一つに用い、また雑音除去差分零交叉数の最大値をもう一つのパラメータとして摩擦音・破擦音を識別する手法を示した。これらの手法を統一的に用いて17種の音韻を9群に分類および識別するアルゴリズムを示し、10名の男性話者による遮蔽板を用いた音声試料と遮蔽板を用いない音声試料をそれぞれ設計用データ・試験用データとして識別実験を行ない、その有効性と実用性を確かめた。

この二重入力システムでは日本語68種の単音節を、最初に計算量の少ない時間領域において大分類し、次に周波数領域で最終的な識別をする方法を採用しており、これによって計算量の圧縮と認識率の向上を目指している。

本論文では時間領域での分類および識別について述べたが，周波数領域での識別については今後の課題である。又，連続音声中の音素認識についても今後の研究課題として残されている。

第2の研究では，音楽情報処理の一研究として，オルゴール音の分析と符号化および曲目の認識実験を行なった。オルゴール音はぜんまいの力の変化によって曲の演奏速度が変わるという特異な性質をもつことを述べた。この問題を解決するためにドラムの回転角を測定する回路をもつシステムを考案し，時間軸での正規化を行なった。次に，このシステムを用い，オルゴール音を楽譜に対応して符号化するアルゴリズムについて述べ，この符号を用いた符号マッチング法を示した。最後に，18種の曲目と10個の不良品を用いた認識実験を行ない，このシステムが生産工程に導入可能であることを確かめた。

本論文における第1の研究である二重入力

システムによる音声認識の手法は従来の1個のマイクによる音声認識の多くの問題を解決することができるため、音声タイプライターの実用化への一番の近道であると考えられる。又、第2の研究であるオルゴール音の自動曲目認識システムはFA（ファクトリー・オートメーション）の一例として今後の活躍が期待される。

謝 辞

本研究の全過程を通じて直接理解あるご指導，ご鞭撻を賜わった神戸大学大学院自然科学研究科システム科学専攻 松本治弥教授（工学部計測工学科）に心から感謝の意を表す。

大学院博士課程において，筆者の履習指導委員としてご指導，ご教示いただいた 松本治弥教授（計測工学科），米持政忠教授（計測工学科），西尾真喜子教授（数学科）に心から御礼申し上げます。また熱心なご指導をいただいた 小川枝郎教授（工学部共通講座），石堂正弘助教授（計測工学科），前川禎男教授（システム工学科）に厚く御礼申し上げます。

本研究を行なうにあたって熱心に御討論いただき，種々の面でお世話になった 芳井隆技官（電子工学科），三柳満隆氏（神戸製鋼）に厚く御礼申し上げます。また，実験の面でご協力いただいた 端本安彦氏（三協精機製

作所), 赤羽康彦氏(三協精機製作所)に心
から御礼申し上げます。ハードウェア, ソフ
トウェアの面でご協力いただいた自然科学研
究科および工学部電子工学科第四講座の大学
院生ならびに卒業生の諸氏に深く感謝する。

参 考 文 献

- (1) Dudley, H.: The vocoder, Bell Labs. Record, 17 (1939b), pp. 122~126
- (2) Potter, R.K. et al.: Visible speech, (1947), D. van Nostrand Co.
- (3) Fant, G.: Acoustic theory of speech production, (1960), p 328, Mouton Co.
- (4) Bell, C.G. & Fujisaki, H. et al.: Reduction of speech spectra by Analysis-by-Synthesis techniques, J. Acoust. Soc. Amer., 33 (1961), pp. 1725~1736
- (5) Stevens, K.N.: Toward a model for speech recognition, J. Acoust. Soc. Amer., 32 (1960), pp. 47~55
- (6) J.W. Cooley and J.W. Tukey: An algorithm for the machine calculation of complex Fourier series, Mathematics of Computation, 19, p297 (April 1965)
- (7) 板倉文忠・齋藤収三, “最尤スペクトル推定法による音声分解合成伝送方式”, 音講論集, 231 (1967)
- (8) Atal, B.S., Schroeder, M.R.: Predictive Coding of Speech Signals. Proc. 1967 Conf. Commun. and Process., 360-361

(1967)

- (9) 板倉文忠・齋藤収三, “偏自己相関係数による音声分析合成系”,
音講論集, 199 (1969, 10)
- (10) 迫江博昭・千葉成美, “動的計画法を利用した音声の時間軸
正規化に基づく連続単語認識”, 音響誌 27, 483-490

(1971)

- (11) 鶴田・迫江・千葉, “連続単語音声認識システムの評価実験”,
音講論集, 605 (1978)
- (12) 藤崎・広瀬・金沢, “単語音声認識装置における話者適応の
一方式”, 音響学会音声研資 S79-26 (1979)
- (13) 川端・牧野・三輪・城戸, “音素の信頼度を利用した単語音声の
認識”, 音響学会音声研資 S80-18 (1980)
- (14) 中川聖一・坂井利之, “個人差の種々の学習機能をもつ実時間
単語音声識別システム”, 信学会論誌 J61-D, 395-402 (1978)
- (15) 中島・角所・藤田・鈴木・西山, “切り出し有声破裂音の知覚”,
信学技報 EA80-59 (1980)
- (16) 矢野・溝口・柳田・角所, “有声破裂音の破裂時点検出に
関する考察”, 信学技報 EA80-60 (1980)
- (17) 細谷克美・藤崎博也, “日本語有声破裂音の分析”, 信学技報
EA80-61 (1980)

- (18) 中田和男, “日本語まさつ音およびはいつ音の合成”,
電波研季報, 6, 27, pp. 364-382 (昭35-11)
- (19) 国崎・藤崎, “無声破裂音の分析と特徴抽出”, 信学技報
EA76-54 (1977)
- (20) 三上直樹・大場良次, “3極線形予測法による無声破裂音の
特徴パラメータ抽出”, 信学論J64-A, 12, pp. 1000-1006 (1981)
- (21) 森川・藤崎, “極-零モデルにもとづく無声破裂音の特徴抽出”,
音講論集, pp. 583-584 (昭53-05)
- (22) 北里・白井, “調音パラメータを用いた子音の識別”,
音講論集, pp. 557-558 (昭55-05)
- (23) 田中和世, “日本語無声摩擦子音の分析と自動識別”,
音響誌38, 6, pp. 330-338 (1982)
- (24) 千葉 滋, “スペクトルピークを用いた鼻子音の識別”, 信学技報
PRL78-52 (1978)
- (25) 三原・森田・白江, “複素ケプストラムを用いた鼻子音のスペクトル
分析”, 信学会全国大会講論集, 5, 1282 (昭54-03)
- (26) 白井・松井, “音声波からの鼻音調音状態の推定”, 信学論(A),
63-A, 2, 74-81 (1980)
- (27) 竹内・粕谷・城戸, “鼻音のスペクトルに及ぼす鼻副鼻腔の影響”
音響誌33, 4, pp. 163-172 (1977)

- (28) 神内教博・松本治弥, “音声波二重入カシステムによる口蓋帆の動きの推定”, 昭56 関西連大, G15-5
- (29) 神内教博・松本治弥, “音声波二重入カシステムによる子音の特徴抽出”, 電子通信学会誌 投稿中
- (30) 神内教博・芳井 隆, “口腔出力と鼻腔出力の分離ピックアップ方式による摩擦音・破擦音の識別”, 音響学会誌 投稿中
- (31) 三柳・神内・松本・赤羽・端本, “オルゴール生産工程における曲目の自動認識”, 計測自動制御学会 投稿中
- (32) 中村 勲, “ピアノ音の特性—ピアノの音響学的研究第4報—”, 音響学会誌 38, 8 (1982)
- (33) 山口・安藤, “デジタル信号処理による自然楽器音の分析”, 音響学会誌 33, 5 (1977)
- (34) 北垣・清水・末武, “コンピュータによる楽音認識の試み”, 音響学会誌 33, 9 (1977)
- (35) 上野・中村・井口, “音楽情報の入力法—採譜—”, 信学技報 PRL78-21 (1978)
- (36) 井口 征士, “音楽情報の処理—電算機を用いた自動採譜—”, 計測と制御 19, 3 (1980)