音声のスペクトルのローカルピークによる認識

東北工業大学
環境情報工学科
古賀秀昭

Ｉ．単母音のスペクトルのローカルピークと傾斜を用いた話者認識

１．１ はじめに

ンるしてる化犯てフ術コ使実　てがルいも　て識定　０あ
コあとしれルタしル技とにを　しどトてり　れ者判　１で
ピしを述クよこ、で法とさバーにアの間利会　となクいよ　。さ話を　て別　
ルと数で一にる し声用ド表ーュかや識人便社　量数ペ用量る録るか　せ識　
カ眼波後ピ識べ出の利一代ロピい字認、り報　徴係スを徴あ登すう　わ者　
一主周はル認調抽誰のワにグンを数者はよ情　特Ｃのク特でめ定ど　あ話　
ロに割てカるか をし識ストのコ報の話とをな　のＰ音一のらじ判か　、た　
と特分い一よの徴析認パッそて情来、こンか　声Ｌ母ピ他かかをる　分し　
斜でのつロにる特解者のネはし人従てるョ豊　音、単ルがるらかあ　人と　
傾験斜にと用れ声徴。個タ会償。るらら一全うで関は一一てがあ者　各対　５象
３実傾細識併さ話人一でと個。べれシで。は数、カクいあるで　
のの３詳認の善手のあはイ化のきでと用ニ、だ研相究、ルら音者人。男タ　女を
ル回なのの方改しそでのに報さ大題号でュてるの己研とカ知力の本るは一
ト今適数み両にク、最波の、うペがに周斜しよスる識割傾較のはい認分３比どて①（②をが究い、と、識率。研用はこと認識る本をのる）の認あ

話てとる特情利が問暗面ミしき識自本斜一が入誰があでデはっこれ。の便性なる場コにで認や、傾ロとはら声が験のとよるらる代、能要よなののが者ムが３、こにか音合実音識にすえあ現い可重にうタもと話ラたたはい識中力照た母認タ別考で、なくはトよ一いこのトれめれ強認の入者っ単者一判に用にも招かッのュする来スら求こに者声、話行の話ュを主応うとをるべこピやす従プいら。音話るとる回分。ピかてのよに罪守ァがンい現ケ用かる騒あ別す今人る。

１．２ 実験方法

１－２－１．実験の流れ

１－２－２．音声資料

実験に使用した音声データについて述べる。本実験では単母音を対象としているので、
日本語の5母音/a/,/o/,/u/,/i/,/e/を実験資料とする。被験者（日本人男女）に防音室で、各母音をランダムに10回ずつ発声してもらい録音する。録音にはDATを用いた。録音した音声を4.5kHzのローパスフィルタに通した後、サンプリング周波数12kHz、分解能12bitでAD変換した。これを視察で1536点切り出した。さらに512点を1フレームとして3フレームに重複なしで分割した。従って、被験者一人あたりのデータの数は
5（5母音）×10（10回発声）×3（フレーム）=150個
である。（表1－1）

1-2-3.特徴抽出

前述した音声資料を1フレーム（512点）毎に以下の処理を行う。（図1－1）

1. 時間窓をかける。（ハミング窓）
2. 14次線形予測係数（LPC係数）を求める。
3. LPC係数を高速フーリエ変換（FFT）し、LPCスペクトルを求める。
4. スペクトルに最小二乗近似直線を引いてスペクトルの3傾斜を求める。
5. LPCスペクトルからローカルピークを抽出し1/6 0ct.帯域幅のバンドパスフィルタに割り当て、25chのローカルピーク系列に圧縮する。

この処理を傾斜の分割周波数を500Hzから1/30ct.毎に変化させ、4000Hzまで繰り返した。分割周波数を固定させずに変化させたのは、傾斜の引き方によって認識にどのような影響があるのか、どこに最適な分割周波数が存在するのかどうかを調べるためである。

1-2-4.傾斜の引き方

LPCスペクトルの全帯域（235Hzから4.5kHzまで）の最小二乗近似直線の傾斜（AT:Total 図1－2）、235Hzから分割周波数（500Hzから4.0kHzまで変化させる）までの傾斜（AL:Low 図1－3）、分割周波数から4.5kHzまでの傾斜（AH:High 図1－3）とした。また、分割周波数は500Hzから1/30ct.毎に変化させたので

$$f_n = 500 \cdot 2^{\frac{(n-1)}{3}}$$

より求まる周波数f_nをその分割周波数とした。

ただし、（n=1，2，・・・，10）である。

1－2－5．認識方法

LPCスペクトルの3傾斜によるBAYES判定を行い、その対数尤度（SL）を求める。SLは次のように求める。入力データをXとする。話者をカテゴリとするkを用いて

$$SL_k = -\frac{1}{2}\{n \cdot \log(2\pi) + \log\left|\sum_k\right| + (X-\mu k)^t \sum_k^{-1}(X-\mu k)$$

として、このSL_kが最大であるカテゴリkをBAYES判定の結果とする。ただし、μは平均値、$\sum$は共分散行列、nは特徴の次元である。この場合は3傾斜なのでn=3である。

続いてローカルピーク系列{0，1}で、2ch同時確率の対数尤度（LP）を求める。LPは次のように求める。認識用の入力音声データ系列をX_{in}とする。

$$X_{in} = (x_1, x_2, \cdots, x_i, \cdots, x_j, \cdots, x_n)$$

x_i, x_jは1または0であり、それぞれピークの有無を示す。話者をカテゴリとするkを用い、$x_i = x_j = 1$となるときの確率をP_{ijk}とする。もし、$x_i = x_j = 1$となった場合に標準パターンP_{ijk}として、そうでない場合は$(1-P_{ijk})$より次式を計算し対数尤度LP_κを求める。

$$LP_\kappa = \log\left(\prod_{j=i}^{n}\prod_{i=1}^{n} P_{ijk}^{\,x_i x_j} \cdot (1-P_{ijk})^{1-x_i x_j}\right)$$

$$= \sum_{j=i}^{n}\sum_{i=1}^{n}\log(1-P_{ijk}) + \sum_{j=i}^{n}\sum_{i=1}^{n} x_i x_j \left\{\log P_{ijk} - \log(1-P_{ijk})\right\}$$

この対数尤度LP_kが最大であるカテゴリkが2ch同時確率の認識結果となる。

以上のように求められたSLとLPの線形結合NLを

$$NL = \alpha \cdot SL + (1.0 - \alpha) \cdot LP$$

で計算する。このときのNLが最大となるものを認識結果とする。ここでαは線形結合係数でα=0.0を0.1刻みで1.0まで変化させて認識率の変化を見る。
従ってα=0.0ではローカルピークのみの、α=1.0では3傾斜のみの認識となる。

1-3.認識結果

/a/の場合では線形結合によって認識率が最大となったのは分割周波数4kHz、α=0.6のときで97.00%を得た。/o/の場合では線形結合によって認識率が最大となったのは分割周波数2519.84Hz、α=0.7のときで94.33%を得た。/u/の場合では線形結合によって認識率が最大となったのは分割周波数2kHz、α=0.5のときで98.67%を得た。/i/の場合では線形結合によって認識率が最大となったのは分割周波数2kHz、α=0.6のときで95.33%を得た。/e/の場合で線形結合によって認識率が最大となったのは分割周波数2kHz、α=0.6のときで97.67%を得た。最後に全単母音の認識では線形結合によって認識率が最大となったのは分割周波数4kHz、α=0.5のときで80.20%を得た。以上の結果を(表1－2)に簡単にまとめた。これは各単母音の認識率が最大をとる分割周波数と線形結合係数αと、その認識率である。結果として/u/の認識率98.67%が最も高い認識率であった。

1-4.あとがき

今回の実験では、傾斜とローカルピークの線形結合をすることにより認識率が相対的に改善されたことがわかった。

また本研究に引き続き、音声に雑音を加えた場合のロバストネスの検討、および伝達特性の変化へのロバストネスについても検討してみる必要がある。

Ⅱ.言語障害者が発声する単母音の音声認識

2・1.まえがき

情報技術が身近になってきた現在、インターネットや電子メール、携帯電話などの便利な情報技術を音声の生成,受聴器官に障害を持つ者が活用しようとする場合は、健常者とは異なった人間一機械間のインターフェースが必要となる。

音声の受聴系の障害に対しては、補聴器や人工内耳などの研究が行われている。また、喉頭がんなどで喉頭の摘出を行った音声の生成系の障害に対しては、人工喉頭や合成音声などの研究が古くから行われている。

いろいろな原因で、言葉をうまく発音ができない音声生成系に障害を持つ障害者は、意思伝達の手段として、文字盤やトーキングエイドなどを用いているのが現状である。

音声受聴系には障害がなく、音声生成系の調音器官に障害を持つことで、発音が不明瞭な言語障害者の音声を機械的に認識できれば、現在行われている方法より、簡単にコミュニケーションをとることが可能になると考えられる。

しかし、現在の音声認識の研究は不特定話者、連続音声の認識が中心であり、対象者は健常者である。いろいろな原因で、発音が不明瞭な言語障害者を対象とした音声認識については、ほとんど研究がなされていないのが現状である。

我々は、音韻性が音声スペクトルそのものやローカルピークの振幅にあるのではなくローカルピークの周波数にあると考えて、音声認識をLPCスペクトルのローカルピークの周波数を用いて行っている。また、ローカルピークの周波数を用いた音声認識が、人間の聴覚系の処理機構に対応していることを報告している。

今回は、言語障害者の発声した音声の認識実験を、線形予測分析に基づくローカルピークの周波数を用いた音声認識法で行った。対象とした言語障害者は、音声受聴系には障害がなく、音声生成系の調音器官に障害を持つことで、発音が不明瞭な発声障害者である。

実験では、健常者の単母音を標準パターンとした言語障害者の発声する単母音の音声認識と、言

語障害者自身の単母音を標準パターンとした言語障害者自身の単母音の音声認識を行い、その認識率を比較した。

さらに、言語障害者の単母音の聴取実験を行ったので、その結果も合わせて報告する。

2-2. 音声資料

実験で使用した単母音の音声資料は、表2-1に示すように、言語障害者は、男4名、女2名の合計6名に、また、健常者は、男60名、女60名の合計120名に、/a/, /o/, /u/, /i/, /e/の5母音をランダムな順序で、各10回ずつ発声してもらい、その音声をDATに録音した。表2-2に、発声被験者の言語障害者の原因と障害の種類を示す。言語障害の種類は、マヒ性構音障害である。構音障害の程度は、多少聞き取りやすい人から聞き取るのが困難な人まで、様々である。

録音した音声資料から認識用の音声データへの加工は、サンプリング定理を満足させるように、遮断周波数4.5kHz,遮断特性－80db/octのローパスフィルターを通し、サンプリング周波数12kHz、分解能12bitでAD変換した。AD変換した音声波形から母音中心付近の1536点を視察で切り出したレーム512点として3フレーム用いた。したがって、母音数は、各人150個となる。

2-3. 認識実験

音声認識は、まず初めに、男女別に60人の健常者の単母音を標準パターンとして、言語障害者6名のの単母音を一人ずつ認識した。したがって、これは各人のオープン認識実験となる。次に、言語障害者6名の単母音を標準パターンとして、言語障害者の単母音を一人ずつ認識する。これは、クローズ認識実験となる。

2-3-1. 健常者の単母音を標準パターンとした認識結果

図2-1は、男女別に60人の健常者の単母音を標準パターンとして、言語障害者6名の単母音を一人ずつ認識し、各母音についてどの母音として認識したかを示すグラフである。健常者の単母音

を、認識した場合、我々の実験では、各母音について、約90%前後の高い認識率を得ているが、今回の言語障害者の実験では、16%から43%と全体的に低い認識率となり、各母音にわたって分散して認識されていることがわかった。

このことは、言語障害者の発声の仕方が、健常者と大きく違うためであると考えられる。

2-3-2. 言語障害者各人の単母音を標準パターンとした認識結果

図2-2は、言語障害者各人の単母音を標準パターンとして、言語障害者6名の単母音を一人ずつ認識し、各母音についてどの母音として認識したかを示すグラフである。認識率は各母音で、84%から94%台と高く、安定して発声することが難しい/u/と/i/についても、84.4%と85.0%の高い認識率となった。この結果から、我々には聴き取りにくい音声でも、発声者自身は、個人の中で確実に発声の区別をしていることがわかった。

2-3-3. 各発声被験者の認識結果

表2-3は、発声被験者ごとに、オープンとクローズの2つの認識実験で得られた認識率を比較したものである。この結果から、どの被験者もオープン認識率が18%から43%台と低いのに対して、クローズの認識率が75.0%から97.3%と高いことがわかる。

特に、被験者4と被験者6のオープンの認識率が18%台と低く、健常者とは、発声法が大きく違うと考えられるにもかかわらず、クローズでは、79.3%と90.0%の高い認識率を示している。このことからも、各被験者は個人の中では安定に発声の区別をしていることがわかる。

2-4. 聴取実験

聴取実験では、認識実験で得られた認識率と聞き取りによる単母音の認識率を比較した。

聴取実験用の音声は、認識実験に用いた各人の単母音音声データを、1母音につき10個、合計50個を持続時間128msとしてDA変換し、DATに録音した。この際、音声信号の立ち上がりや立ち下りでのクリック音などを除去するために、データの両側に10msのテーパを付けた。

聴取実験の方法は、ＤＡＴに録音した単母音を、学部学生５名と障害者の治療にあたって障害者の声を聞きなれている介護職員５名で、聞き取りによる音声の聴取実験をおこなった。聞き取る際には、ヘッドホンを使用した。

2-4-1. 各母音の聴取実験結果

言語障害者６名の単母音を一人ずつ聞き取り、各母音についてどの母音として認識したかをグラフで示す。

図２－３のグラフは、学部学生５名の認識結果で、認識率は５５.０％から８５.３％である。

図２－４のグラフは、障害者の治療にあたって障害者の声を聞きなれている介護職員５名の認識結果である。認識率は５６.０％から７７.０％である。

どちらのグループも、単母音のオープン認識の認識率よりは高い認識結果であるが、学部学生の認識率には、ばらつきがあるのに対して、介護職員の方は、ばらつきが少ない。全体的には、特徴的な差は見られなかった。

2-4-2. 各発声被験者の聴取実験結果

表２-４は、発声被験者ごとに、聴取実験の結果を比較したものである。

学部学生５名の認識率は、６７.４％であり、聴取被験者が日常的に接している介護職員５名の認識率は、６５.５％である。日常的に接している介護職員の認識率のほうが高いのではないかと予想したが、予想に反して、この２つのグループの認識率に大きな差はなかった。この原因としては、聴取実験をか単母音に限って行ったためだと考えられる。しかし、日常的に接している介護職員は、動作や前後の音声から言葉を類推してコミュニケーションをとっていると思われる。

2-5. まとめ

今回は、言語障害者の発声した音声の認識実験を、線形予測分析に基づくローカルピークの周波数を用いた音声認識法で、単母音に限って行った。対象とした言語障害者は、音声受聴系には障害がなく、音声生成系の調音器官に障害を持つことで、発音が不明瞭な発声障害者である。

認識実験の結果、言語障害者の音声は聴き取り

に く い が 言 語 障 害 者 自 身 は 安 定 な 発 声 を し て お り 、
個 人 の 中 で は 発 声 の 区 別 を し て い る こ と が わ か っ
た 。
こ の こ と か ら 、 言 語 障 害 者 個 人 の 特 性 に 合 わ せ
た 音 声 認 識 装 置 の 開 発 に 向 け 、 さ ら に 実 験 を 進 め
る 予 定 で あ る 。

謝 辞

本 研 究 を 進 め る に あ た り 、 音 声 収 集 に ご 協 力 頂
い た 宮 城 県 拓 桃 医 療 療 育 セ ン タ ー の 関 係 者 の 皆 様
に 深 く 感 謝 し 、 心 か ら お 礼 を 申 し 上 げ ま す 。
ま た 、 音 声 デ ー タ の 収 集 と デ ー タ 処 理 に 協 力 し
て 頂 い た 本 学 卒 業 生 （ 平 成 12 年 度 ） の 高 田 真 一 君
と 若 生 悟 君 に 心 か ら 感 謝 い た し ま す 。

全 体 の 謝 辞

これらの研究を進めるにあたり、貴財団からの助成により多くの成果を挙げることができました。ここに記して厚く御礼を申し上げます。ありがとうございました。

単母音の数	5個
遮断周波数	4.5kHz
サンプリング周波数	12kHz
分解能	12bit
切り出した点の数	512点×3=1536点
一人当たりのデータの総数	5×10×3=150個

表１－１話者認識用音声資料

単母音	分割周波数Hz	線形結合係数	認識率
/a/	4000.00	0.6	97.00
/o/	2519.84	0.7	94.33
/u/	2000.00	0.5	**98.67**
/i/	2000.00	0.6	95.33
/e/	2000.00	0.6	97.67
全単母音	4000.00	0.5	80.20

表１－２１０人の話者認識率

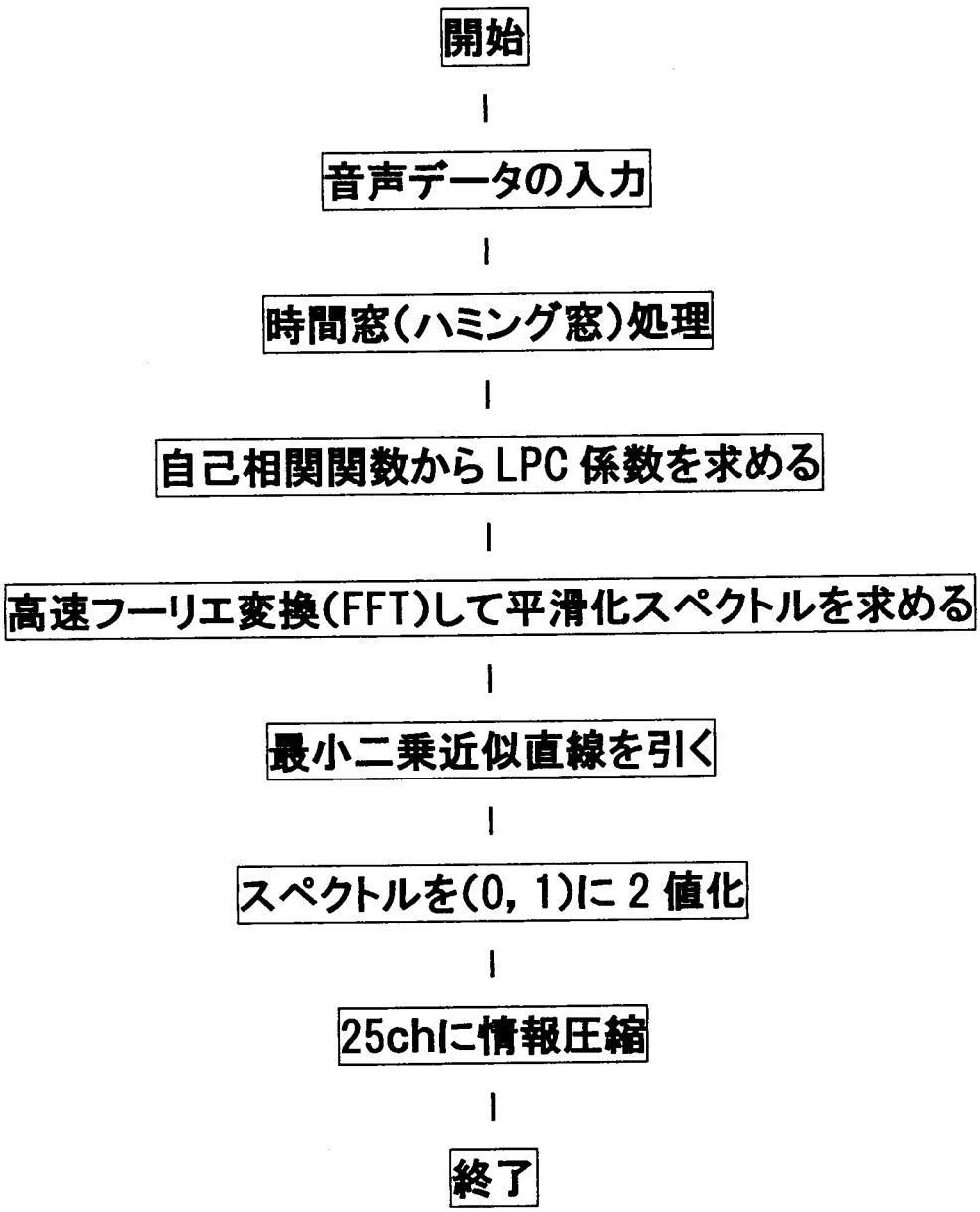

図１－１ローカルピークの抽出法

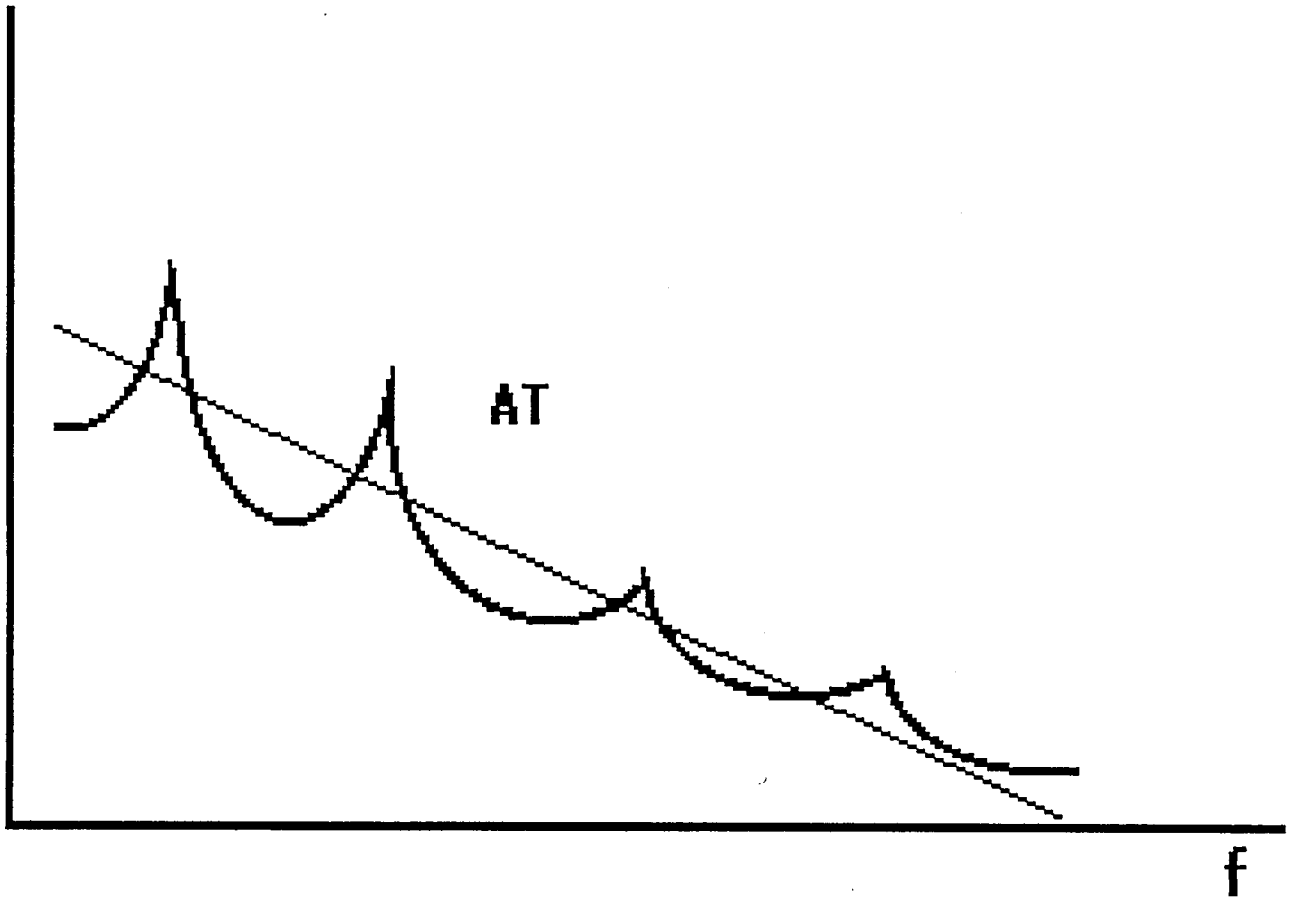

図１－２ＡＴの抽出法

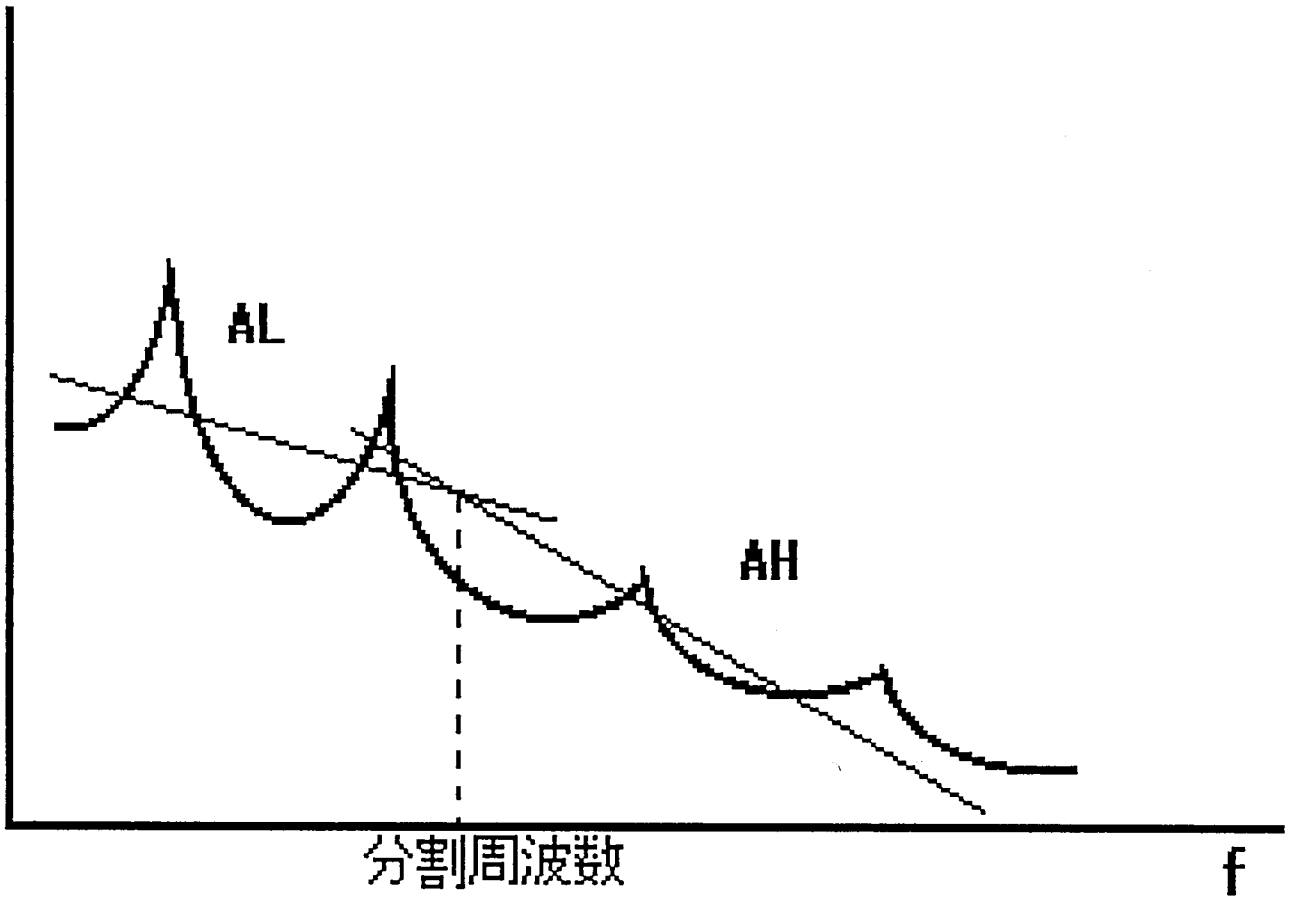

図１－３AL、AH の抽出法

	言語障害者	健常者
話者数	男　4名 女　2名 計　6名	男　60名 女　60名 計 120名
障害の種類	マヒ性構音障害	—
ＬＰＦ	遮断周波数 4.5kHz，遮断特性-80dB/oct	
サンプリング周波数	12 kHz	
分解能	12 bit	
音声波形の切り出し	1,536 点	
１フレーム	512 点（データ重複なし）	
母音数	各人 150 個	

表２－１単母音音声資料

発声被験者	診断名	言語障害の種類
被験者 1(女)	多発奇形	口蓋裂構音障害
被験者 2(女)	脳性マヒ	マヒ性構音障害
被験者 3(男)	頚直型両手マヒ	マヒ性構音障害
被験者 4(男)	頚性四肢マヒ	マヒ性構音障害
被験者 5(男)	脳性マヒ	マヒ性構音障害
被験者 6(男)	頚性四肢マヒ	マヒ性構音障害

表２－２言語障害の原因と障害の種類

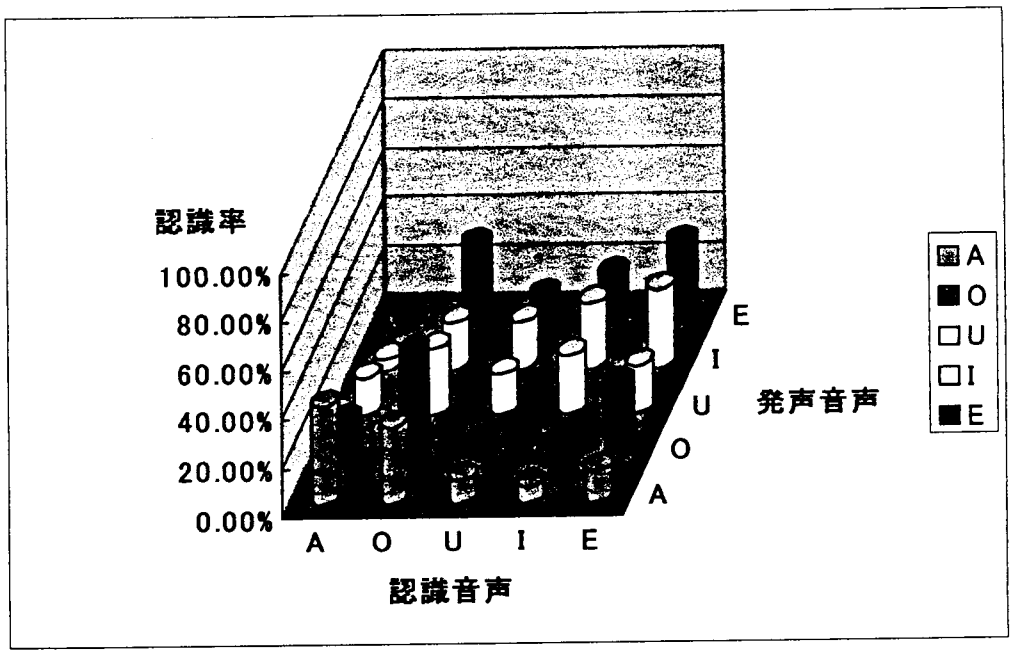

図２－１健常者の単母音を標準パターンとした
言語障害者の認識結果

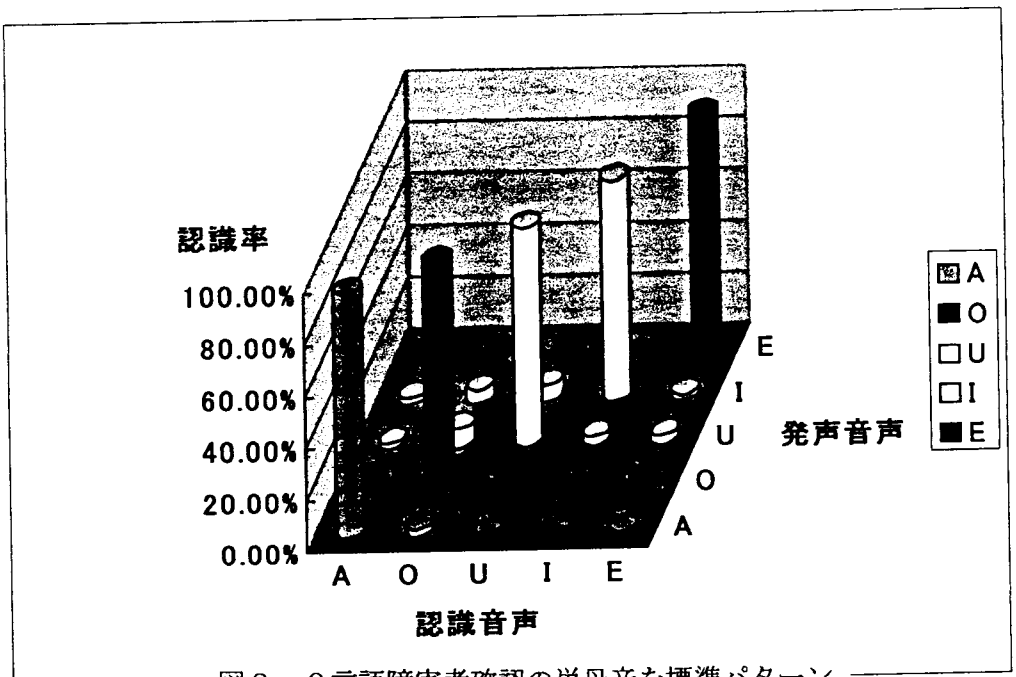

図２－２言語障害者確認の単母音を標準パターンとした認識結果

発声被験者	(1) オープン	(2) クローズ
被験者 1(女)	37.3%	97.3%
被験者 2(女)	32.7%	75.3%
被験者 3(男)	39.3%	95.3%
被験者 4(男)	18.0%	79.3%
被験者 5(男)	43.3%	96.0%
被験者 6(男)	18.7%	90.0%
平均値	31.6%	88.9%

表２－３認識実験結果

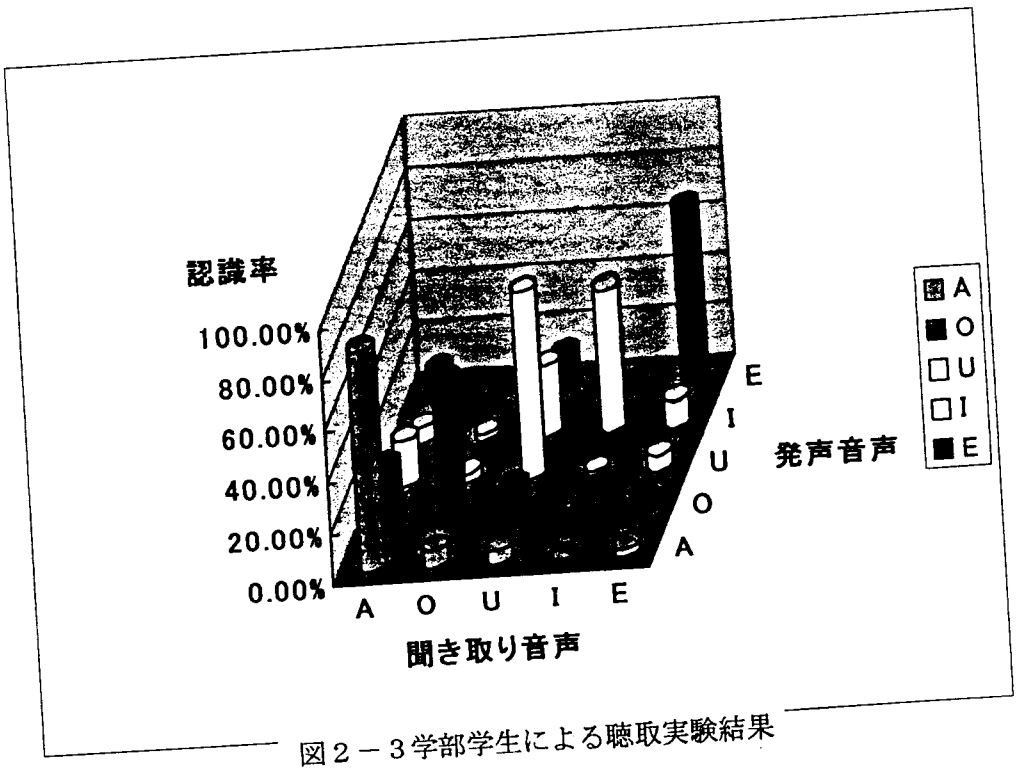

図２－３学部学生による聴取実験結果

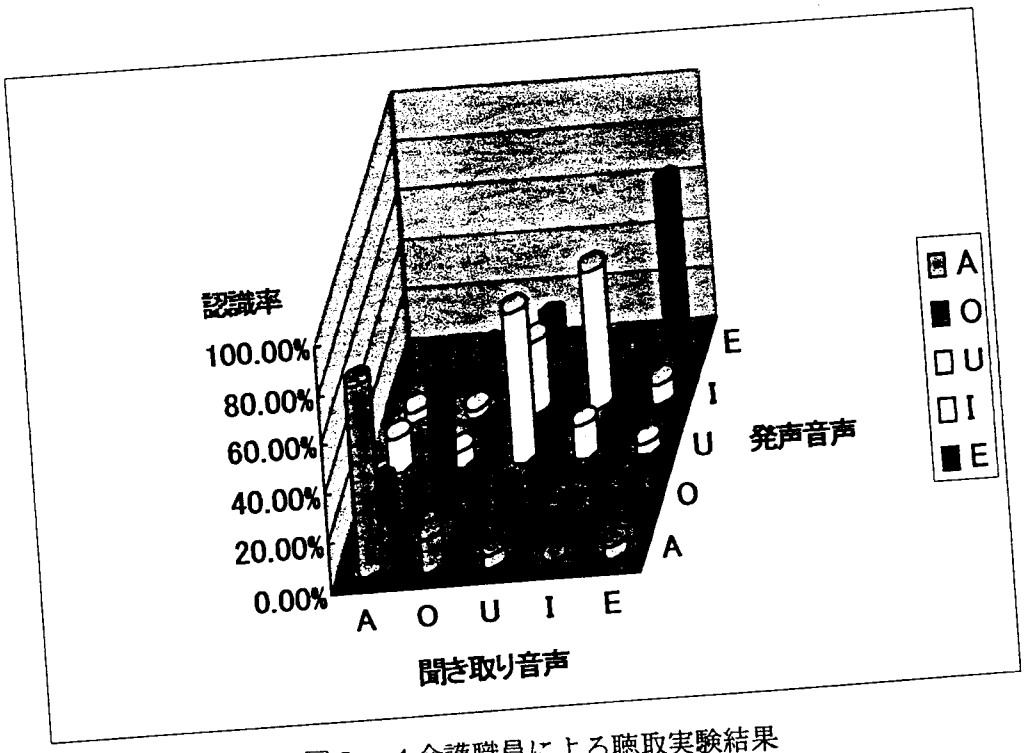

図２－４介護職員による聴取実験結果

発声被験者	学部学生	介護職員
被験者 1(女)	76.0%	63.6%
被験者 2(女)	41.6%	38.8%
被験者 3(男)	79.2%	76.0%
被験者 4(男)	63.2%	67.2%
被験者 5(男)	96.4%	90.8%
被験者 6(男)	48.0%	56.4%
平均値	67.4%	65.5%

表２－４聴取実験結果