多次元尺度法を利用した音声の特徴表示

平野忠男* ・ 山岡清寛** ・ 高田稔浩***

Expression of Characteristic of Voice by means of Multidimensional Scaling Method.

Tadao HIRANO ・ Kiyohiro YAMAOKA ・ Toshihiro TAKADA

This study has been made to clear and indicate the characteristic of voice by multidimensional scaling method(MDS). Speech materials used in our experiment were CV-syllables containing voiceless stop consonants and short-term intonation(STI). Firstly we examined similarity test on the temporary CV-syllable distorted by various method. Secondary similarity test has been made on the three syllables intonation 'aoi' and 'ioa' formed various pattern. Non-similarity matrix were calculated from results of above mentioned test and disposition of characteristic of each speech materials were indicated on 2 or 3 dimensional surface by MDS method.

1 まえがき

音声は人間が言語という情報形式を用いて意志疎通するために、人間の発声、受聴器官の制約のもとで表現した音響的な信号である。また音声は、人間と人間の意志疎通の手段であるばかりでなく、言語の生成と表現とに深い関わりを持つものである。

音声および聴覚に関する研究は、従来計算機の性能と周辺機器の充実が要求されていた。近年になり小型計算機が広く普及し性能が向上するとともに、データをディジタル信号として処理するためのソフトウエアが著しく発展し、小型計算機による音声情報処理が広く行われるようになり、更に音声合成、認識等の研究に関係して信号処理方法も高度化されつつある。しかしこれら研究を更に推進させるにはＣＶ音節の子音部、母音部の音韻情報分布、或はイントネーションの本質等詳細に調べることが必要である。本研究は各種音声試料の聴取実験を通して上記問題を調べ多次元尺度法（以下ＭＤＳ法と略記）を用いてそれらの特徴を視覚的に表現することを試みたものである。

2 音声の概要

音声は声帯を通る断続気流が声道フィルター、即ち、口腔、舌の形状、或は鼻、歯、唇等の影響を受けて放出される音波であり、各種音声が生成する。放出される音波は口腔の形状により　特定周波数に共振し、この共振周波数を一般にフォルマント周波数と言い、各母音（vowel）はこのフォルマント周波数の分布によ

＊経営工学科　＊＊電気工学科　＊＊＊福井工業大学大学院修了（現 福井市役所）

り特長づけられる。子音（consonant）は一般に単独では存在せず、普通の音声は　子音と母音とが結合して形作られる。子音をＣ、母音Ｖをで表わすと、我々が発生し得る最少規模の音声はＶ，ＣＶ、ＶＣ、ＣＶＣの形をとりこれらは音節といわれる。イントネーションは、上記断続気流の断続数即ち基本周波数の影響が最も大きいと言われている。イントネーションによって伝達される情報は多種であり、語の持つ意味、個人性、情緒性あるいは地方色などの差異といったものをかなり的確に伝えることが出来る。ところでイントネーションに似ている概念に「アクセント」あるいは「プロミネンス」といったものが知られている。音声学的にいえば、イントネーションに対して、”音声連続における声の高さの変動あるいは不変動を「イントネーション」と言う”であり、普通にいう文レベルのイントネーションの他に、単語レベルのイントネーションの存在も認められている。またアクセントに対しては、”強め段階の強さあるいは高さに関して一定した社会習慣的な型が存在するとき、このような型を「アクセント」という”としている。従ってアクセントには強さのアクセントと高さのアクセントが存在することになる。

工学的な見地から考察する場合には、”現象”そのものを純粋に指している「イントネーション」に注目することで十分であると考える。本研究においては、単語レベルのイントネーションを文レベルのイントネーションから切り離して考えることにし、これらをＳＴＩ（Short-Term-Intonation）と定義した。　音の物理的な属性とそれに対応する心理量は、

周波数　→　ピッチ

音　圧　→　ラウドネス

波　形　→　音　色

であり、これらが音の心理的な基本的属性である。これら３種の属性は独立であるとはいえ、物理的属性と、心理的属性は必ずしも１対１に固定されたものではない。例えば、ピッチの変化は周波数の変化と対応関係にあるとはいえ、必ずしも周波数のみに依存しているものでないことも報告されている。

３　実験方法

３－１　子音部の音韻情報分布

本報告においては、無声破裂子音からなるＣＶ音節の聴取実験を次のようにして行った。成人女性１名により自然に発生された音声試料はまずデジタル・オーディオ・テープレコーダー（ＤＡＴ）に録音され、音声信号を作成するために以下の手順によって処理された。抽出された音声のサンプルは、標本化周波数48 kHz、量子化精度16bit にて、20kHzのローパスフィルタ（ＬＰＦ）を通してＡ／Ｄ変換し、パソコンＰＣ9800VX（ＮＥＣ）の拡張メモリーに格納した。これらの音声素片は、目的とする音声試料を得るためにサウンド・マスター[4]および音声工房[5]を用いて時間情報処理を行って、別のファイルに格納した。これらはＤ／Ａ変換されて試聴用の磁気テープ上に録音した。磁気テープ上にこうして録音された試料は実験計画に従ってランダムに配列し、ヘッドホンにより再生して片耳提示により主観判断実験に供された。被験者は年齢が約22才であって、いずれも正常な聴覚を有する10人の成人男性であった。

［実験１］　ここで実験に供した音声資料は、原波形の子音部とそれに続く母音部に至るまでの間の区間を３等分し、Fig.１に示すように各区間を独立に除去することによって８種類の刺激音を作成した。ここで作

成した試料は残留部と除去部にそれぞれＯＮ（１）とＯＦＦ（０）を対応させて２進３桁の名称を与えてパタンを区別することにした。刺激音の提示方法は各サンプルの前に１秒の間隔の無声区間をとって原音を置き１組みとして聞かせた。また、次の組みの刺激音との間には３秒の無声区間を置く。削除のなされない原音と刺激音とを組みにして与えて聴取実験を行い、被験者には刺激音が本来の原音とどの程度似ているかを表１に示す５段階評価にて応答することを求めた。

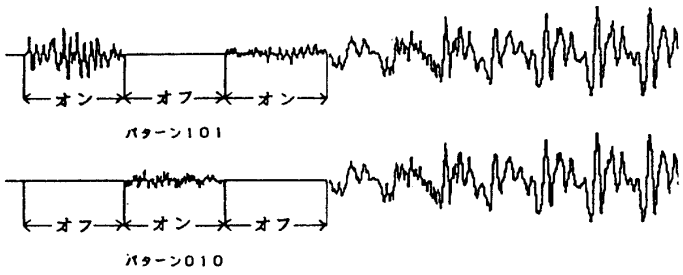


図１．子音部３等分切り取りの例

表１．判断基準の評定表

判断基準	スコア
全く同じように聞こえる	1
似ているように聞こえる	2
中　　間	3
あまり似ていない	4
全く異なって聞こえる	5

［実験２］　子音情報の布置を求める為にここでは実験１で使った音声試料を用い８種類の刺激音の全ての組合せを作り順不同の56組の供試音についてそれらがどの程度異なっているか聴取実験を行った。任意に選んだ２つの刺激音の間の非類似度をすべて求めた。その結果はｋruskalの方法で解析し、各音声試料の間の違いを２次元または３次元の位置関係で表した。

３－２　イントネーションの知覚

実験装置のブロック図を図２に記す。本研究における音声の合成は、成人女性１名によって正常に発生された音声から抽出した音声素片を用い、編集合成する方法をとり、基本周波数を変化させることにより、９通りのイントネーションパターンを持つ３音節合成音を作成した。

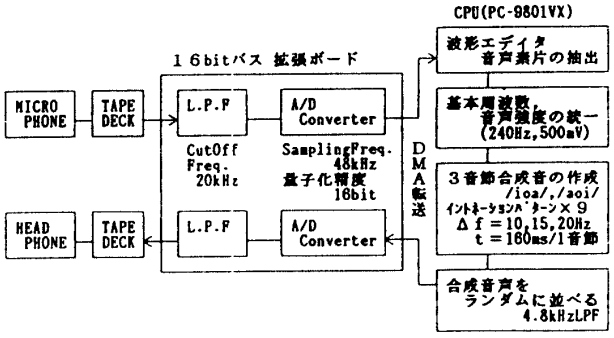


図２．実験装置のブロック図

３－２－１　音声試料の採取

日本語母音の中から３種の母音／ａ／，／ｏ／，／ｉ／を音声試料として選んだ。指定された基本周波数のデータを作成するために、発声者には一定の発声強度にて一定の高さの持続母音を発声するように求めた。発声された音声試料から、基本周波数が安定している部分の１周期分を抽出して、合成するための音声素片とした。合成音の音節ごとの基本周波数及び音声強度を統一するために、基本周波数においてはスプライン補完を行い240Hzとし、音声強度においては実効値を500mVとして振幅変調を行った。

３－２－２　合成音の作成

今回合成した音声は日本語母音の３音節で構成された単語で、有意語の／aoi／と、無意語の／ioa／である。基本周波数の制御は、音声素片の持つ波形の１周期の長さを基準にして、必要とする基本周波数に相当する時間長が得られるようにサンプリングのデータ個数の追加、削除を行った。この音声素片を繰り返すこ

とにより所定の基本周波数を持つ合成音を作成した。本研究で合成した９通りのイントネーションパターンを図３に示す。また、３音節合成音の音声モデルを、図４に示す。また、第１音節の基本周波数をf_0＝240Hzと決め、各音節間の基本周波数の変化幅をΔｆとし、それぞれΔｆ＝10,15,20Hzとした。

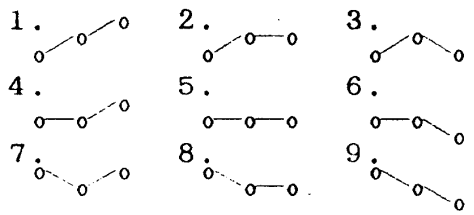


図３．９通りの イントネーションパターン

３－２－３　合成音の聴取

実験Ⅰの目的は９通りの３音節基本周波数パターンの聴取からイントネーション知覚パターンを求めることであり、実験Ⅱの目的は第２、第３音節の基本周波数を変化した場合におけるイントネーションの心理的類似度を求めることである。

作成された３音節合成音をランダムに並べD／A変換し、14人の被験者にヘッドホンを通じ、一定のラウドネスで両耳による聴取実験を行った。なお、被験者は正常耳を有する22歳～24歳の成人男性及び女性である。具体的な実験方法を以下に記す。

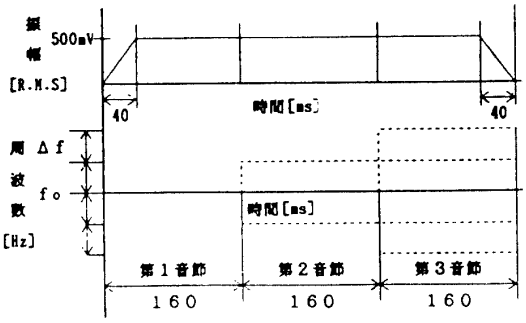


図４．合成音の一般的モデル

（ａ）実験Ⅰ

被験者自身が自覚したイントネーションパターンを図３に示す９通りのパターンから選ぶ方法で応答を求めた。作成された試料間の時間長は、被験者が応答用紙記入のための時間等を考慮し、3.5秒とした。

（ｂ）実験Ⅱ

イントネーションパターンの心理的類似度を、Kruskalの方法による解析が適用可能とするため、３音節合成音を周波数変化幅ごとに、２つずつの組み合わせ、36組みを作った。作成された２組の試料間の時間長は、３音節合成音と同じ時間長とした。また、次の２組の合成音との時間長は、実験Ⅰ同様3,5秒とし、聴取実験を行った。被験者には２組の合成音を比較して、前記に示す５段階の判断基準に従って、スコア１～５で応答するように求めた。

４　実験結果

４－１　子音部の音韻情報分布

用いられた刺激音は３ビット（$b_3b_2b_1$）にて表記した。図２で先頭（左端）がb_3、母音に近い部をb_1とする。子音部の全てが除かれた（０００）から原音に等しい（１１１）までS_0～S_7からなる。聴取実験の結果は、時間切断ひずみを与えた８種類の各刺激音がＣＶ節からなる原音とどの程度に異なるかを判断スコアに従って応答を求め、度数分布を求めた。この結果は子音の種類及び後続母音により傾向が異なり一律な結論を出すことが困難であるが、一例として子音／ｋ／の場合について５母音との結び付きにおける分割区間の非類似度の傾向を示したのが図５である。また、各音素別の子音部の音韻情報分布の値を表２に与えた。この表での平均とは各音

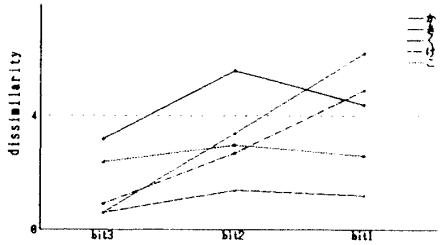


図５．／ｋ／の子音部音韻情報分布．後続母音による影響

素の音韻情報の中心部がどこにあるかを表わすものである。

4－2　子音情報の布置

聴取実験による2音の対を比較することによって8個の刺激音の相互の間の非類似性に関するスコアが求まった。無声破裂子音のうち／ｋａ／および／ｐａ／についての結果を平均値によるスコアにて表3に示す。これらの関係から、Kruskalの方法によって刺激音の間の心理的距離を計算し、それぞれの座標点を2次元空間布置として表示したのが図6，7である表3においての大きい数値は三角形において対応している2つの刺激音の間を結ぶ距離がより長くなることを意味するもので、全体の布置がこれらの数値から定まった。ここで両者のstressを比べたとき、／ｋａ／は2次元では比較的大きな6.90％を示したが3次元布置では、2.00％の値となった。

表２．子音部音韻情報分布例
（各母音による平均）

consonant	bit 3	bit 2	bit 1	average
/k/	17.9%	37.8%	44.4%	1.74
/t/	56.8%	13.8%	29.4%	2.27
/p/	34.1%	43.4%	22.5%	2.12
/s/	15.7%	37.3%	47.1%	1.69

表３．非類似度行列

（ａ）無声破裂子音／ｋａ／

$b_1b_2b_3$	000	001	010	011	100	101	110	111	dissimilarity
000		3.0	3.2	3.9	3.0	4.1	3.6	4.0	3.2
001	3.0		2.2	2.5	2.5	2.1	2.6	2.9	2.4
010	3.2	2.2		2.2	2.2	1.5	1.6	2.2	2.0
011	3.9	2.5	2.2		2.9	1.4	2.3	1.6	2.2
100	3.0	2.5	2.2	2.9		2.1	2.4	3.4	2.4
101	4.1	2.1	1.5	1.4	2.1		1.7	1.8	2.0
110	3.6	2.6	1.6	2.3	2.4	1.7		1.4	2.1
111	4.0	2.9	2.2	1.6	3.4	1.8	1.4		2.3

（ｂ）無声破裂子音／ｐａ／

$b_1b_2b_3$	000	001	010	011	100	101	110	111	dissimilarity
000		2.1	3.6	3.8	2.2	3.6	3.6	3.9	3.0
001	2.1		2.9	3.2	2.2	3.3	3.1	3.3	2.6
010	3.6	2.9		1.8	3.0	1.7	1.6	1.7	2.2
011	3.8	3.2	1.8		2.8	1.6	1.5	1.6	2.2
100	2.2	2.2	3.0	2.8		3.0	2.6	3.2	2.5
101	3.6	3.3	1.7	1.6	3.0		1.6	1.5	2.2
110	3.6	3.1	1.6	1.5	2.6	1.6		2.1	2.1
111	3.9	3.3	1.7	1.6	3.2	1.5	2.1		2.3

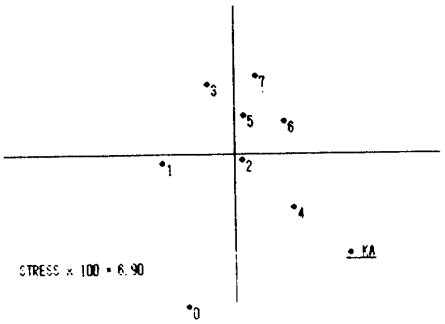
4－3　イントネーションについての実験[3]

（ａ）実験Ⅰ

前記方法により、提示したイントネーションと被験者の知覚したパターンに関して、同定者数を図8に示す。

（ｂ）実験Ⅱ

Kruskalの方法による解析のため。刺激ｉとｊの間の非類似度は、表1のスコアと聴取実験から得られた人数との積を集計することによって求める。すなわち、スコアの値をｓ、そのスコアの応答人数を$n_{ij}(s)$とするとき、



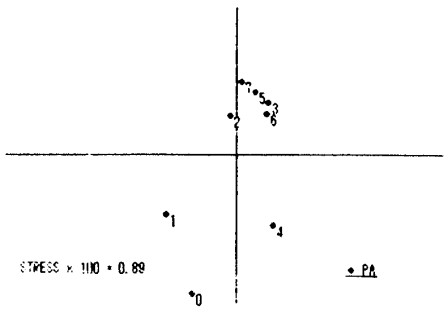


図６．”か”の２次元心理空間布置　図７．”ば”の２次元心理空間布置

$$\delta_{ij}=\sum_{s=1}^{5} s \cdot n_{ij}(s)$$

で与えられ、その最大値は５Ｘ（応答延べ人数）で与えられる。

以上の手法により、５段階のカテゴリーで評定した応答表から、非類似度マトリックスを導くことが出来、いずれも最大値は、70となる。９通りのイントネーションパターンに対する心理空間の布置として２次元を選んだ場合の結果を図９，図10に示す。

5　結果の検討

5－1　子音部の音韻情報分布

各破裂子音について音韻情報分布を求めたところ図5、表2に見るように／k／においてはｂ３＜ｂ２＜ｂ１の傾向を示した。ほかの子音／t／、／p／においては若干傾向が異なっていた。ここで、ｂ３＜ｂ２＜ｂ１の傾向のもっとも強い／s／の場合を加えて考察すると、これら音素の音韻情報の中心分布地点は時間軸上において左側の区間から順に

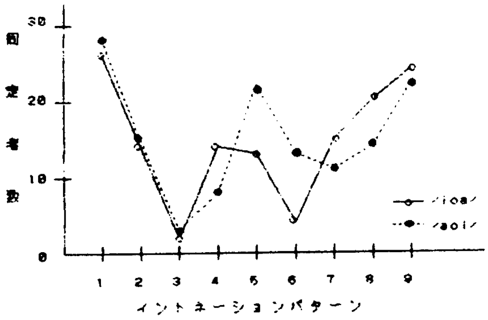


図8．各ｲﾝﾄﾈｰｼｮﾝﾊﾟﾀｰﾝの同定者数（聴取者数４２人）

ｂｉｔ３　　　　ｂｉｔ２　　　　ｂｉｔ１

←－－－－－－－－－－－－－－－－－－－－－－－－→

ｔ　－　ｐ　－　ｋ　－　ｓ

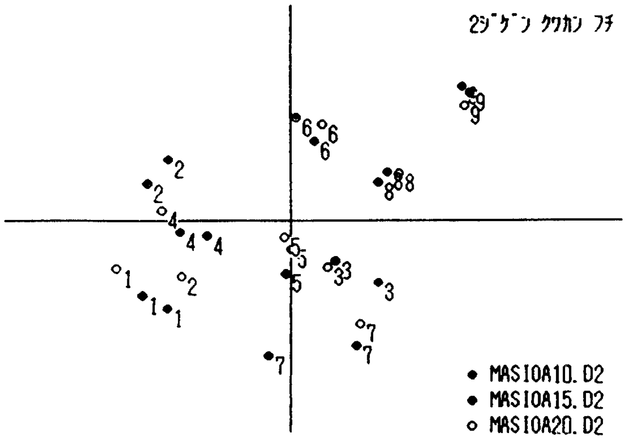


図９．２次元空間布置／ｉｏａ／

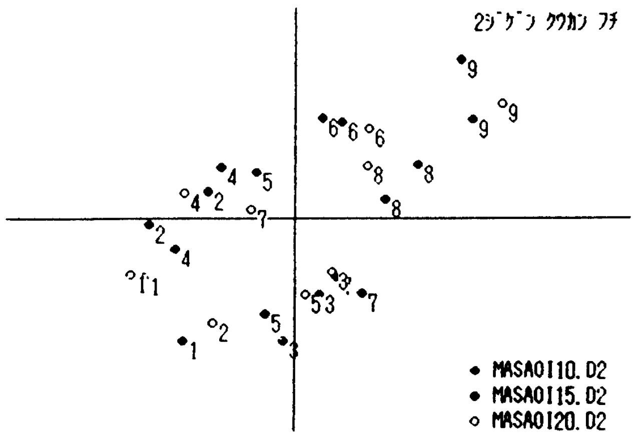


図10．２次元空間布置／ａｏｉ／

と分布していることがわかった。文献[(1)(2)]において、子音情報が後続母音にどの程度影響を及ぼすかについて調べ、その結果後続母音／ｉ／、／ｕ／の場合時間軸上で最も長く継続し、／ａ／の場合減衰が速いことが報告されている。図５では”き”、”く”がｂ１の情報分布が大きく、上記後続母音への影響が説明できるようである。

5－2　子音情報の布置

ＭＤＳについては／ｋａ／、／ｐａ／について考察を加える。これら図６及び図７の二つの布置を眺めるとき、いずれもS_0が他の刺激音から十分に距離が離れている。それに対する位置にあるのがS_7となっている。これらは（０００）と（１１１）に対応しているため子音情報の分布という点から十分に説明できる。次に興味を引くのは（S_3，S_5，S_6）がS_7に接近して布置していることである。この場合の刺激音の共通点はいずれも刺激音の除去区間がただ１つということであり、原波形のS_7に似ていることを意味する。１区間だけからなる刺激音の中ではS_2がそれに続き、／ｐａ／では（S_1，S_4）とは異なった布置を示す。すなわち、中央の区間b_2が子音部全体との間に厳密な関係を持つという例になる。しかし、／ｋａ／の場合にはそこまで言い切ることは出来ない。この意味でも、子音の種類によって布置に違いがあらわれるはずである。表３の右端の列に非類似度の平均を求めてみた。この値はある刺激音のその他の刺激音からの孤立の指標と

なっており、S_0、S_1、S_4にその傾向を見ることが出来る。

5－3　イントネーション

5－3－1知覚パターンによる影響

（1）図8において、各音節の基本周波数変化回数を情報量とみなせばパターン1、3、7、9は情報量は等しいが、同定率は1、9が高く、3、7が低いことよりイントネーション知覚は周波数変化の方向性が重要な意味を持つと思われる。

（2）有意語と無意語についてはパターン4、5、6以外は大体類似の傾向を示す。4、5、6で両者の差が大きいのは第一ホルマント周波数の変化パターンの影響があることも考えられる。

（3）イントネーション知覚は基本周波数の変化パターンが尾高型に知覚され易い。

5－3－2　各パターンの心理的類似度

図9、図10より以下のことが認められる。

（1）布置はパターン1、2、4の群、3、5、7の群、6、8、9の群即ち変化パターンが尾高型、尾低型、その他に大別される。

（2）周波数変化幅に対する布置の変化は有意語の方が無意語より大きい。即ち無意語の方は客観的な判断をしているのに対し、有意語においては心理状態、先入感等の影響が含まれているものと思われる。

（3）座標軸を回転して1、9方向をX軸とすれば原点からの距離は第1、第3音節の間の周波数変化を示している。

6　むすび

子音情報の時間軸上の分布およびイントネーションのパターンによる知覚特性を聴取実験を通して定量化することを試みた。さらに各特徴を視覚的に表わす為、MDS法を並用し、その効果が大きいことを確めた。今後は、試験精度の向上、例えば、子音部の分割を窓をかけて滑らか行うこと、イントネーションについては、供試音の自然性を向上することが必要である。さらに供試音を、他の子音、或は男性話者についても行い、データを集積してゆくことが望まれる。

終りに本研究を実施するに際し供試音の作成、聴取実験、及びデータ整理に協力してくれた福井工業大学の卒研生、特に中小路、永治両君に深く感謝する。

(1) 平野，山岡，桑原，松浦：破裂子音の音韻情報についての基礎調査、
電気関係学会北陸支部連合大会 B-148 (1991)

(2) 松浦，桑原，山岡，平野：無声破裂子音の音韻情報分布、
岐阜大学工学部研究報告　(第42号　1992)

(3) 平野，山岡，高田，松浦：イントネーションの知覚とその特徴表示、
電気関係学会北陸支部連合大会 B-139 (1992)

(4) カノーブス電子（株）：サウンドマスター・ユーザーズマニュアル

(5) NTTアドバンストテクノロジー（株）：音声工房ユーザーズマニュアル

(平成4年12月17日受理)