

周波数解析を用いた日本語母音の特性に関する研究

西田 茂生 宋 剛秀 馬島 了

The characteristics of Japanese vowel using frequency analysis

Shigeki NISHIDA, Takehide SOH and Ryo MASHIMA

言語の起源を探る方法として、基本単語や音韻変化の法則を調べる方法が主流である。本研究では、音そのものの特性を調べることにより日本語の起源を探求しており、その第1段階として母音のもつ特性を周波数解析という手法を用いてアプローチする。母音の音声波形を周波数空間へ変換し、ピークの現れる周波数とそのときの音圧レベルによって母音の特性を考察した。また、各母音の原波形をローパスフィルタ処理した音を用いて、どの母音であるかを認識させる知覚実験を行った。その結果は、周波数解析により得られた特性をよく反映し、解析結果を裏付けるものとなった。すなわち、「あ」は広い周波数域に音を特徴づける要素を持ち、「お」は低周波域に、「え」は高周波域に、「い」「う」は低周波域と高周波域の両方に音を特徴づける要素を持つことがわかった。

1. 緒言

本研究は日本語の起源を探る目的で始めたものである。日本語は、中国より漢字がもたらされる以前にも使用されており、当時それは音声言語として用いられていた。漢字伝来以前の日本語、これを古代日本語として位置づけ、それがどの様に形成されたのか、また古代以前における縄文期の原日本人が使用していた音声言語がどの様なものであったのかを探求していくことにより日本語の根元を探ることを目的としている。そのためには、言語間の音韻および音韻変化の規則を明らかにすることや、基本単語の類似性などを調べることは当然必要なことである。しかし、また言語の音韻そのものの特性を把握することも重要であると考えている。

今回の研究では、言語の音韻特性を明らかにするその第1段階として、日本語の音声的特徴を、音声の周波数解析を行うことにより明らかにする。これまでに行われ

てきた音声の周波数特性は、犯罪捜査やセキュリティ対策に用いられる音声による個人特定である。これは個々の音声間の相違点を抽出し、個人の割り出しを行うものである。本研究ではこれとは全く逆で、同じ音どうしの共通項を探り、その音自身が持つ特徴を抽出するものである。特に本稿では、日本語母音に関する考察を行う。日本語の母音は、若干の変動はあるものの古代より「あ」「い」「う」「え」「お」の5母音で構成されている。これら5つの母音の音声波形からフーリエ変換により周波数スペクトルを求め、各母音の特徴の抽出を行う。

本稿では母音のみの考察であるが、次の段階では子音についての考察を行い、最終的には日本語の音声の周波数特性を把握する。この研究成果の応用範囲は広く、音声合成、話者同定、音声自動翻訳機、聾者のための聴取機械、複雑な音声コマンドに応答する装置、言語病理学、教育プログラムなどが期待される。現在の技術もかなり進歩しており、音声によって文章をタイピングするソフトなどが実用化されている。しかし、このようなタ

イピングソフトは文章に変換するにあたり音声が発した文が文法的あるいは意味的に正確であるという仮定より予想を働かして動いており、文法的な省略や未知語を多く含む会話は変換できない。そしてそういった不規則な会話をリアルタイムで認識変換することが音声解析における大きな課題である。これを実現実用化するには音声の性別、年齢、方言による個人差、雑音の影響にも耐えられる頑強な認識手法などの様々な問題がある。しかし、音声は物理的な音波であり、不特定話者の会話を人が理解できる以上そこには共通する情報が含まれているはずである。本研究での成果により、不特定話者の発話による音声認識も可能となる。

2. 音の3要素

空気中にある物体が振動するとその振動の波が空気を媒質にして聴覚に伝わる。これが音である。音は空気の振動という物理的性質と人に伝わる際の感覚的性質を持っておりこの両性質は密接に関係している。物理的、感覚的にみる音は主に3つの要素が考えられる。

(1) 音の大きさ 物理的には振幅や音圧レベルの大きさであり、大きいほど音は大きく感じられるが単純な比例関係にはならない。音の大きさを「ソーン」というが、これは周波数1kHz、音圧レベル40dBの純音を人が聞いて感じる大きさを1ソーンとする。また音圧レベルの単位dBはある音の圧力を p (Pa)で表すと

$$\text{音声レベル} = 10 \log_{10} \left(\frac{p}{20} \right) \text{ [dB]}$$

となる。音圧レベルが10dB増加すると音の大きさが約2倍に感じられることになるが、これは同じ周波数での比較であり感覚的な大きさは振動周波数に依存する。

(2) 音の高低 音の高さは振動の周波数に依存しているが音の大きさと同様に比例関係にはならない。音の高さは「メル」で表され、周波数1kHz、音圧レベル40dBの純音を聞いて人が感じる高さを1000メルとする感度単位である。また人の可聴周波数域は20~20kHzである。

(3) 音色 音色は大きさと高低を含めた音の個性を表す要素である。音色の違いは大きさと高さと同じでも違う音に聞こえる。異なる楽器で演奏されたドは違う音に聞こえるがそれは音色が違うからである。音色は様々な物理量が多次元に組み合わせられて構成され

ていると考えられているが、今のところ感覚との関係は確立されていない。音は空気の振動であるが、発生させるには元となるべき物体の振動が必要である。音声の場合、音の元になる振動体は気管の入り口にある声帯である。声帯は筋肉と粘膜でできた1センチメートルほどの左右一対の帯であり、その隙間を声門という。普段呼吸をしているとき声門は開いており、声を出すときには声帯が縮んで声門が閉じる。そこに下から無理に空気を送ると、声帯が振動して声門が開閉を繰り返して音を発生するのである。この時空気を押し出すポンプの役割を果たすのは横隔膜と肺による空気の流れである。

3. 発声機構

発声の機構について説明する。図1は人の発声機構を簡単なモデルにした概略図である。人の発声器官は音源である声帯と気管から唇にいたるまでの曲管で構成される。音源である声帯に曲管が接続されることで声道を通して発生される音は、声帯振動と管共鳴による音が合わさったものとなる。これが母音である。なお母音発声時には鼻咽腔は閉じている。鼻咽腔が開放され空気が鼻腔を通ればm音のような鼻音になり、唇や歯を使って狭い空気の流れをつくれれば破裂音や摩擦音になる。

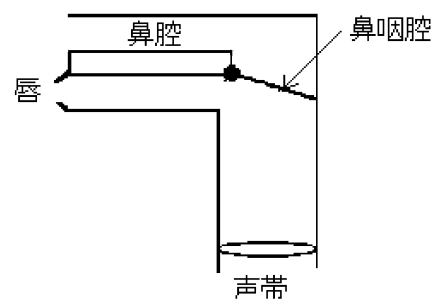


図1 人の発声機構モデル

4. 母音の周波数解析

本実験では音声を解析するために元波形とそのフーリエスペクトルを用いた。以下にその実験環境を示す。音声のデータベースとして18~25歳までの男女20名の母音の音声をパソコンに取り込んだものを使用した。音声を母音に限定したのは、方言のイントネーションの影

響がなく、比較的長時間の録音にもその特性が失われな
いからである。AD変換のサンプリング周波数は基本周
波数の測定では11kHz、母音の特性評価ではCDと同じ
44kHzを用い、量子化ビットは双方とも16bitである。ま
た、フーリエ変換のサンプリング点は2098点である。ま
ず、雑音の影響を調べるために音声を入力しないときの
スペクトルを計測し、雑音の影響を考慮した。

4.1 基本周波数の測定

本研究では母音についての考察を行う。母音は声帯振
動と管共鳴のみで生成されるが、これは管楽器の楽音生
成方法と同じである。そこでまず、ハーモニカの「ド」
音のスペクトルを測定した。図2に測定したスペクトル
を示す。「ド」音は振動体であるリードが131Hzで振動
することによって演奏される。スペクトルの第一番目の
ピークはまさに131Hzであり、一番低い位置にあるピー
クが音源の周波数を表していることがわかる。以下これ
を基本周波数と呼ぶことにする。

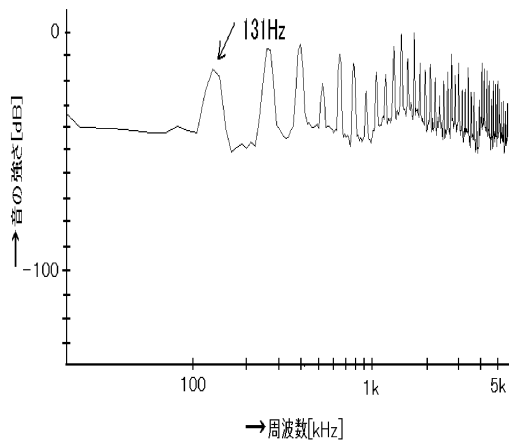


図2 ハーモニカ「ド」音のスペクトル

次に各母音の音声波形をフーリエ変換して被験者の
声帯周波数を測定した。図3にその結果を示す。被験者
に「あ」「い」「う」「え」「お」音をそれぞれ約5秒間発声さ
せ、その波形をフーリエ変換して基本周波数を求めた。
図では「あ」「い」「う」「え」「お」の平均を出力している。横
軸は被験者の番号を表しており、17~20番は女性であ
る。男性の基本周波数は100~150Hz、女性の基本周波
数は200~250Hzであることがわかる。このことは性別
間で声帯の形が大きく異なり、また同性であっても個人
によって声帯の形が微妙にちがうことを表している。

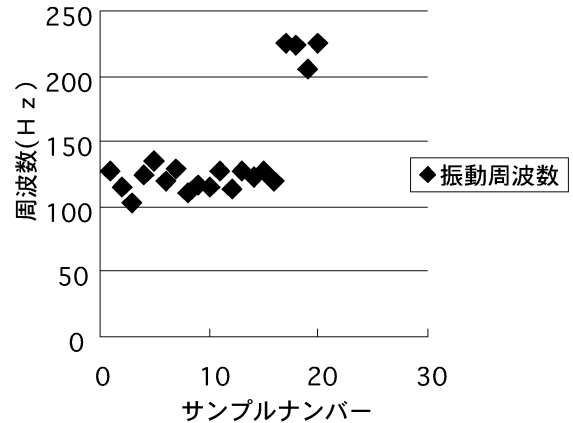


図3 基本周波数

4.2 母音ごとの特性評価

「あ」「い」「う」「え」「お」音の波形をフーリエ変換し被
験者男性16人を対象にそのピーク位置を1.5kHzまで測
定した。図4に測定値の一例の各ピークを示す。

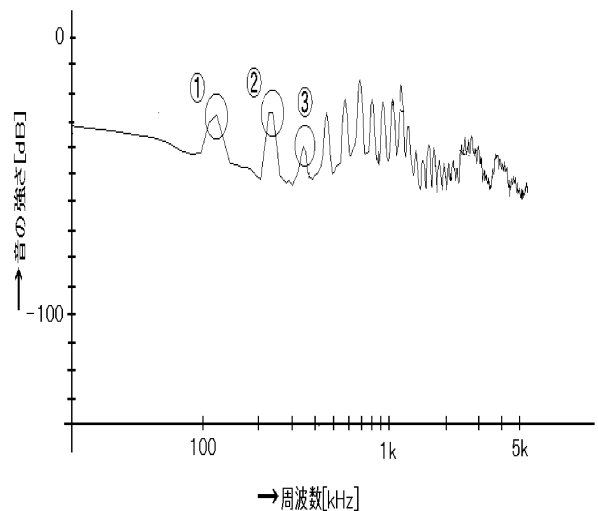


図4 音声の各ピーク値

基本周波数のピークをピーク1とし、順にピーク2、
ピーク3とする。このようにピークとなる周波数とその
ときの音圧レベルを順に調べ、横軸を周波数、縦軸を音
圧レベルとしたグラフにまとめていく。今回の被験者
に対する結果は80%の割合で同じパターンのグラフが得
られた。その結果を図5に示す。

図中のプロットはピークとなる周波数を示しており、
各母音においてピークをとる周波数および、ピークの音
圧レベルが異なり特徴的なパターンを持つことがわか
る。「あ」音のみが他の母音にみられる120Hzのピーク

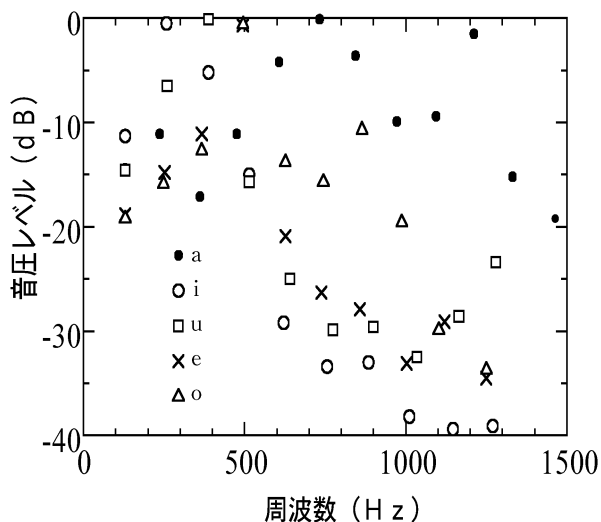


図5 日本語 5 母音の特徴

を持たず、他の母音にみられない1450Hzのピークを持つ。これが「あ」音と他の母音を区別する条件となる。また、各ピークの持つ周波数が母音により差があり、特に高周波域で顕著に見られる。「あ」「い」「う」「え」「お」音の特徴が明確になるよう、それぞれを個別にプロットし直したものが図6から図10である。

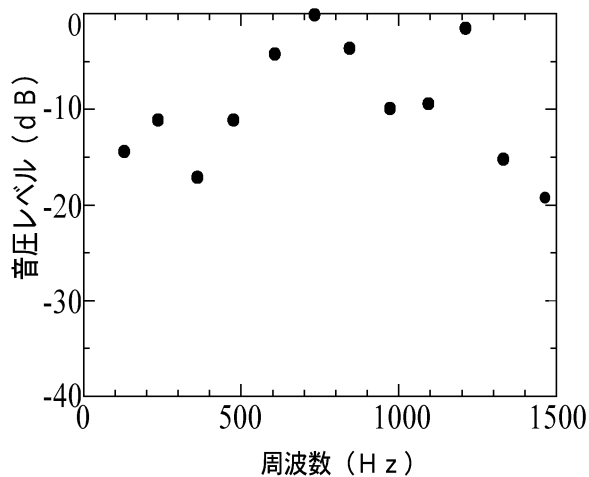


図6 「あ」音の各ピーク周波数と音圧レベルの関係

「あ」音の特徴としてピーク6 (725Hz) が最も音圧レベルが高いことが挙げられる。また、全周波数帯で音圧レベルが平均的に高くなっている傾向が見られる。このことより、「あ」音はピーク6を中心に広い周波数帯に音の特徴づける要素 (以後これを音素ということにする。) を持つと考えられる。

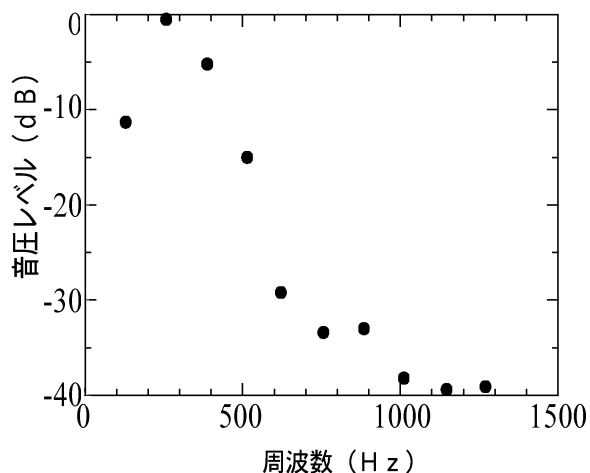


図7 「い」音の各ピーク周波数と音圧レベルの関係

「い」音の特徴としてピーク2 (255Hz) の音圧レベルが最も高く、500Hz以上のピークの音圧レベルが極端に低くなっている。このことから「い」音はピーク2を中心に500Hz以下の周波数帯に音素を持つと考えられる。

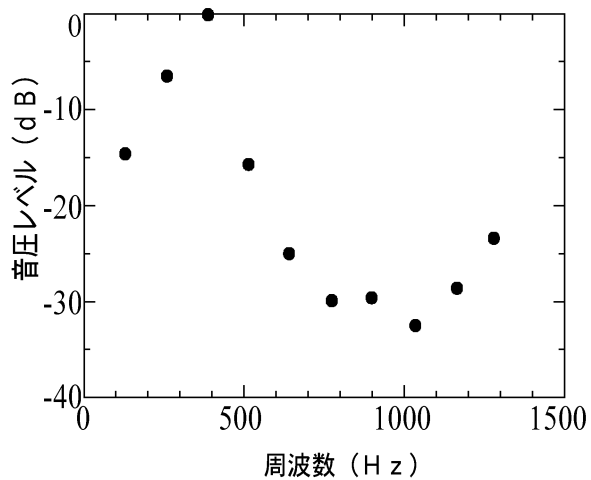


図8 「う」音の各ピーク周波数と音圧レベルの関係

「う」音はピーク3 (385Hz) の音圧レベルが最も高く、500から1kHz間のピークの音圧レベルが低くなり、1kHz以上のピークの音圧レベルがまた高くなるような特徴が見られる。このことから「う」音は中間の周波数帯を除く、低周波数帯と高周波数帯の両側に音素を持つと考えられる。

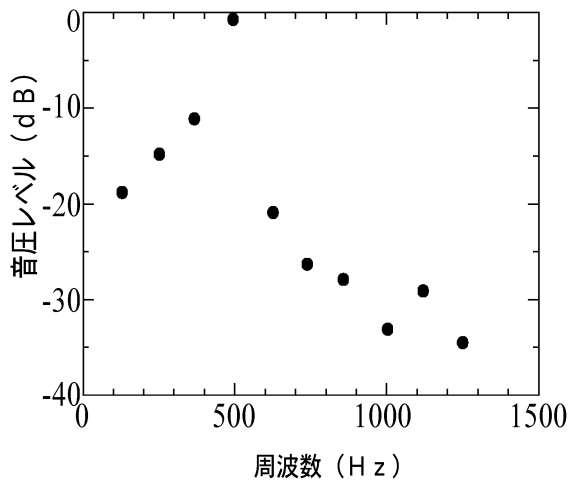


図9 「え」音の各ピーク周波数と音圧レベルの関係

「え」音はピーク4 (490Hz) の音圧レベルが最も高く、次のピーク5の音圧レベルが激しく落ち込み、700Hz以上の高周波数帯ピークで音圧レベルが低くなっている。このことから、「え」音はピーク4を中心に600Hzまでの低周波数帯に音素を持つと考えられる。

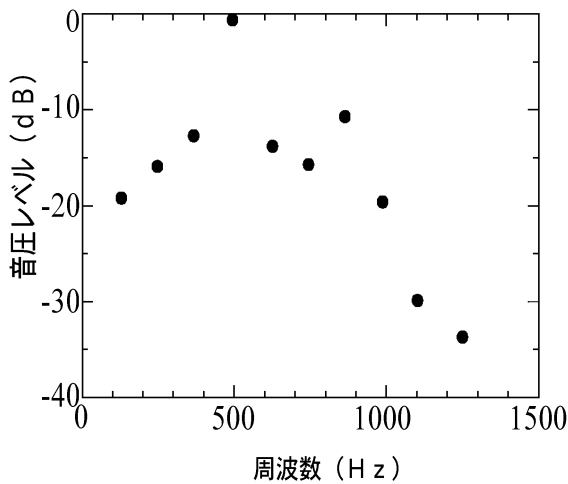


図10 「お」音の各ピーク周波数と音圧レベルの関係

「お」音はピーク4 (500Hz) の音圧レベルが最も高く、ピーク4までは「え」音と特徴が似ている。しかし「え」音と比較しピーク5での音圧レベルの落ち込みは小さくピーク7 (860Hz) の音圧レベルが再び高くなり、「え」音と同様1kHz以上の周波数帯ではピークの音圧レベルは低くなっている。このことから「お」音では、「え」音の音素に700Hzから900Hzの周波数帯にある音素を加えたものが音素となっていると考えられる。

これらの結果をまとめると、以下のことがいえる。

- (1) 700Hzあたりのピーク6が最も強く他の周波数帯で平均して音圧レベルが高いものが「あ」音である。
- (2) 500Hzあたりのピーク2が最も強くこの近辺の低周波数帯に音素を持つものが「い」音である。
- (3) 800Hzあたりのピーク3が最も強く、この付近の周波数帯と1kHz以上の高周波数帯に音素を持つものが「う」音である。
- (4) 「え」音と「お」音は同じように500Hzあたりのピーク4が最も強くなるが、700Hz以上の高周波数帯で音素が存在するものが「お」音、ないものが「え」音である。

4.3 男女差

今回は女声のサンプルがわずかしかとれなかったが、男女による母音の特性についても考察を行った。図11は「あ」音での男女のピークの周波数と音圧レベルの関係を示したものである。

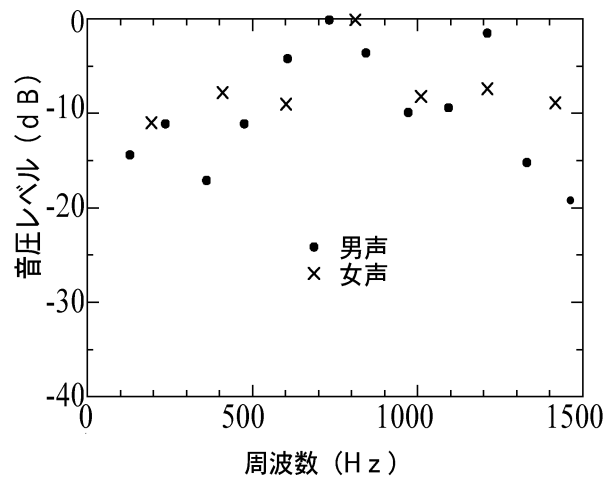


図11 「あ」音の男女差

男女ともよく似た「あ」音の傾向を示すが、最も音圧レベルの高いのは、男声でピーク6、女声でピーク4である。すなわち女声の方が早いピーク値で音圧レベルが最高となる。また、女声の方が男声より高周波数側の音圧レベルが高くなっている。これは女声の方が一般に「声が高く」、男声より高周波側に音声周波数がシフトするためである。

4. 4 母音に関する知覚実験

人がどの周波数帯を聞き取って母音の違いを判断しているのかを調べるために、次のような知覚実験を行った。まず典型的な男性の母音音声を1kHz、800Hzでローパスフィルタにかけ、高周波数域の周波数成分をカットした。処理していない母音、1kHzでカットした母音、800Hzでカットした母音の3種15個をランダムな順序で17人の被験者に聞かせ、それぞれがどの母音であるかを答えさせた。その結果を表1に示す。

表1 母音の知覚実験結果

条件	認識率 (%)				
	あ	い	う	え	お
処理なし	100	71	88	100	82
1kHz z	100	6	41	0	76
800Hz z	65	12	65	0	100

表1より、各母音でフィルタ処理なしと処理後と比較すると、「い」「え」と聞こえる人が極端に少なくなり、かわって「あ」「う」「お」と聞こえる人が多くなっている。また高周波成分をカットするほど「お」と聞こえる人の割合が多くなっている。この結果から、「あ」音では、他の母音と比べて高周波成分をカットしたときの認識率が高くなり、高周波成分をカットした時の「い」「え」の認識率が下がったことから、知覚的にはピークの周波数的位置をもとに母音を判断していると考えられる。「あ」については広い周波数帯に音素が存在する。「お」についてはフィルタ処理を施したあとでも比較的認識率が高いことから、低周波数帯に音素を持つと考えられる。「い」「え」については高周波数帯に音素を持つと考えられ、「う」では低周波帯と高周波帯の両方に音素を持つと考えられる。

知覚実験の結果と周波数解析による結果とを比較検討してみると、「あ」音については、広い周波数帯でピークを持つ周波数解析結果とよく一致している。「お」音については、高周波域にもピークは存在するが、低周波域の方がより音素を持ち、5母音中最も低い周波数域に音素を持つと考えられる。「う」音では高周波域と低周波域に音素を持つことが双方の結果から導くことができる。「え」音については、低周波域にも強いピークが存在するものの高周波域に音素を持つことが考えられる。最後に

「い」音であるが、「う」音と同様に低周波域と高周波域の両方に音素を持つが、高周波側では今回測定した1.5kHzより高い周波数域に音素を持つことが予想される。

最終的に母音の周波数的特性について、ピークをとる周波数位置とそのときの音圧レベルによっておおまかな特徴はつかめたが、機械的に判断できるような明確な定義はできていない。また、「い」音のように今回の測定周波数域では結論づけることができない結果が生じた。

5. 結言

本稿の目的は、日本語5母音の周波数的特徴を明らかにすることであった。結果として日本語5母音の大まかな特性、すなわち各母音がどの周波数域に音を特定するための音素をもつかの目安をつかむことができた。しかし、各母音を客観的に数値で評価するには至らなかった。また、周波数解析とともに知覚実験を行った結果、「え」音を除く他の4母音で結果の一致がみられた。このことは、周波数解析によって母音の特性をつかむことの可能性を示唆している。「え」音に関しても、解析する周波数領域を拡大することによって結果の一致が見込まれる。

参考文献

- 1) 日本音響学会編, 音のなんでも小事典, 講談社, 1996
- 2) Ray D.Kent, C. Read 荒井他監訳
「音声の音響分析」海文堂出版社, 1996
- 3) K. Tanaka “A dynamic processing approach to phoneme recognition. I. Feature extraction”
IEEE Trans. Acoust Speech Signal Process, Vol.27, No.6 (1979) 596-608
- 4) Castelaz P. F. and Niederjohn R. J.
“A comparison of linear prediction, FFT, and zero-crossing analysis techniques applied to synthetic vowel recognition” Proc. Midwest Symp. Circuit Syst. Vol.20th, No.Pt2 (1977) 698-704
- 5) Hess W. A pitch-synchronous digital feature extraction system for phonemic recognition of speech”
IEEE Trans. Acoust Speech Signal Process, Vol.24, No.1 (1976) 14-25