



Title	時変パラメータを用いた母音および半母音の認識
Author(s)	外山, 淳; Toyama, Jun; 新保, 勝 他
Citation	北海道大學工學部研究報告, 167, 39-45
Issue Date	1994-01-14
Doc URL	https://hdl.handle.net/2115/42412
Type	departmental bulletin paper
File Information	167_39-46.pdf



時変パラメータを用いた母音および半母音の認識

外山 淳 新保 勝
(平成5年8月31日受理)

Recognition of Vowels and Semivowels by Time-Dependent Parameters.

Jun TOYAMA and Masaru SHIMBO
(Received August 31, 1993)

Abstract

Vowels are classified on the plane with coordinate axes of two lower formant frequencies, i.e., F_1-F_2 plane. On the other hand, since semivowels are acoustically similar to vowels, it is hard to discriminate the former from the latter. Based on the fact that the time transitions of peak frequencies and peak powers of semivowels are different from those of vowels, we propose a method to discriminate vowels from semivowels using newly defined time-dependent parameters corresponding to the foregoing time transition characteristics. In an experiment, we discriminate five Japanese vowels and three semivowels with the recognition rate of 84.6% for 4,554 phonemes uttered by three male speakers.

1. ま え が き

音声の認識手法は単語を単位として認識する手法と音素や音節を単位として認識する手法の二つに分類することができる。単語を単位として認識する手法は、音素や音節を単位として識別する方法に比べて実現が容易であるものの、認識すべき単語の数が増加するにつれ、テンプレートを増加させなくてはならないという欠点がある。一方、日本語の音素や音節の数は一定である。例えば東京方言の場合音素は24個であり、音節は約100個である。このため、音素や音節を単位として識別する方法は、カテゴリーへの分類が多少困難であっても、一度認識システムを構成してしまうと、その後テンプレートを変更する必要がないという長所がある。

本論文では後者の立場から半母音と母音の識別・分類を行なうことを試みる。半母音は子音であるため継続時間が短いながらも、母音と同じように明確なホルマントを持ち、それは母音のホルマントと滑らかに繋がっている。このことが母音と半母音の分類、識別を困難としている大きな原因となっている。半母音/y/の識別としては、Schwabらにより後渡りにおける第2ホルマントの周波数の時間変化量に特徴があるという報告があり¹⁾、また松村らにより F_1-F_2 平面上のホルマントの挙動から半母音/y/を認識する方法が報告されているものの、他の半母音については実験されていない²⁾。また、Espy-Wilsonによって全ての半母音に対して識別を行なった例はあるもの

の、ここでは母音との分類を行なっていない³⁾。我々は半母音/y/, /w/の識別を行なうために、前あたりと定常部の第2ホルマントの周波数の時間変化に着目した⁴⁾。しかしながら、この方法では語頭の半母音を抽出することが難しかった。このように、音素を単位とする音声認識の手法において半母音の認識は重要な問題でありながらも、確たる識別方法は確立されていない。

本研究では、半母音がホルマントを持ちながらも継続時間が短いことから、ホルマントの時間変化を捉えた時変パラメータを導入することにより、半母音と母音を識別・分類することを試みる。

2. 音素識別領域の構成

母音はホルマント周波数を用いることによって、分類・識別が可能となることが知られている。ここでは、第1及び第2ホルマントを座標軸とした $F_1 - F_2$ 平面上に母音の識別領域を設定することによって母音の認識を行なう。母音の識別領域は以下のように設定する。母音はこの平面において2次の正規分布に従って分布していることを仮定する。2次の正規密度関数は

$$f(x^1, x^2) = \frac{1}{2\pi|\Sigma|} \exp\left(-\frac{1}{2}(x-\mu)' \Sigma^{-1}(x-\mu)\right), \quad (1)$$

により得られる。ここで、

$$\begin{aligned} \Sigma &= \begin{pmatrix} E[x_k^1 - \bar{x}^1]^2 & E[x_k^1 - \bar{x}^1][x_k^2 - \bar{x}^2] \\ E[x_k^2 - \bar{x}^2][x_k^1 - \bar{x}^1] & E[x_k^2 - \bar{x}^2]^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \\ \mathbf{x} &= \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \bar{x}^1 \\ \bar{x}^2 \end{pmatrix} = \begin{pmatrix} E[x_k^1] \\ E[x_k^2] \end{pmatrix} \end{aligned}$$

である。なお、 x^i は第*i*ホルマントの周波数に対応し $E[x]$ は x の平均を表す。ここで、式(1)の値 $f(x^1, x^2)$ を一定とすると、楕円を表す方程式

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c \quad (c : \text{constant}).$$

を得る。これが、母音の音素識別領域となる。本研究では半母音においても母音と同様に音素が2次の正規分布に従って分布していることを仮定し、同様の方法により識別領域を決定する。図1～3に示した楕円が音素識別領域である。

3. 半母音の特徴パラメータ

半母音は発声方法が母音と似ており、母音のホルマントに滑らかに繋がるホルマントを持つ。そこで、この平面上における母音と半母音の挙動を図1～3に示す。図中の楕円は母音の識別領域を表す。図1は/oya/, /oea/, /uyo/, /uio/の挙動を、図2は/awa/, /iwa/, /iua/の挙動を、図3は/era/, /ora/の挙動を $F_1 - F_2$ 平面上に示している。それぞれの挙動における点の間隔は5 msである。この図から、半母音/y/が分布すると考えられる領域は母音の/i/, /e/の両方の領域にまたがっていることがわかる。また、/oya/と/oea/, /uyo/と/uio/の挙動はほぼ一致している。このように、 $F_1 - F_2$ 平面上の分布領域や挙動だけから半母音/y/と母音を識別分類することは難しい。図2, 3より半母音/w/, /r/についても同様のことが言える。このように半母音を含んだ音素列 $V_1 C V_3$ の挙動はこの平面上において3連母音 $V_1 V_2 V_3$ の挙動と似ており、また半母音の分布領域は必ず

れもどれかの母音の領域と重なっているため、この平面だけを用いて母音と半母音を識別することは困難である。これらのことから半母音を識別するためにはホルマント周波数だけではなく、別のパラメータが必要である。

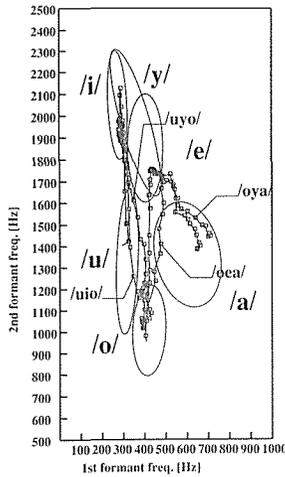


図1 F_1-F_2 上における /uyo/, /uio/, /oya/, /oia/の挙動

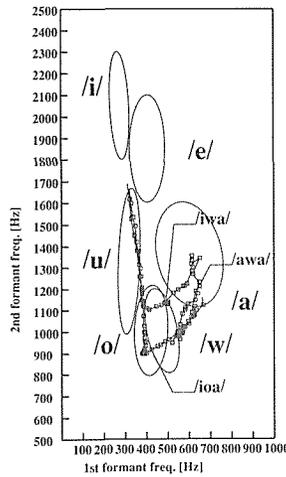


図2 F_1-F_2 上における /awa/, /iwa/, /ioa/の挙動

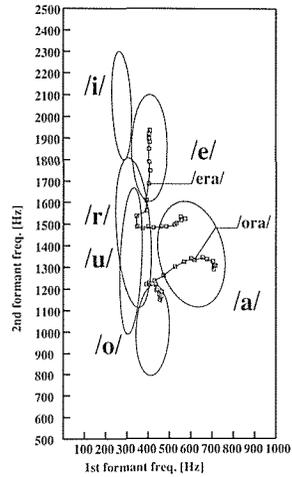


図3 F_1-F_2 上における /era/, /ora/の挙動

半母音は子音的な要素，すなわち発声時間が短いという特徴が示すように，一度の発声に対しても F_1-F_2 平面上において母音のようにその分布が一点に集中することはない。このことは，時間と共にホルマント周波数も変化していることを示している。半母音が母音と同様に顕著なホルマントを持ちながらも発声時間が短く，過渡的な音であることから，ホルマントの時間変化を捉えることが可能なパラメータが識別に有効であると思われる。これを半母音の特徴として捉えるために，新たなパラメータとして従来の F_1-F_2 平面を構成していた座標軸とは独立な時間変化を捉えるパラメータが必要となる。

図4に示すホルマントトラッキングを観察すると，/y/における第2ホルマントの周波数は先行または後続する母音の第2ホルマントの周波数よりも高くなっていることがわかる。また，定常部における時間変化に対して安定した周波数を示す時間は極めて短くなっている。図5に示す/oea/の場合には，先行または後続する母音よりも第2ホルマントの周波数が高くなっている点では同じであるものの，定常部における周波数の時間変化は小さい。同様に，図6に示した/w/の第2ホルマントに着目するとその周波数は/y/とは逆に先行または後続する母音の第2ホルマントの周波数よりも低いことがわかる。また，定常部における安定した周波数を示す時間は極めて短くなっている点は同様である。一方，/r/は図7からわかるように第1ホルマントに特徴があり，先行母音から/r/定常部にかけて第1ホルマントの周波数が先行または後続する母音よりも低くなる。いずれの半母音においても定常部が極めて短いことから，注目しているホルマントの周波数は放物線に似た遷移を示す。

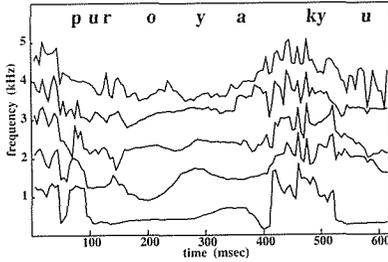


図4 「プロ野球」 (/puroyakyu-/)
のホルマントトラッキング

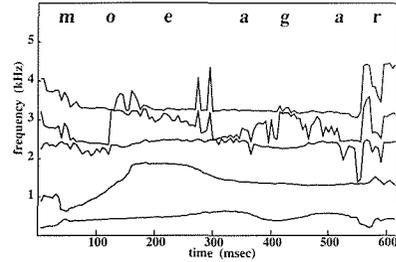


図5 「燃え上がる」 (/moeagaru/)
のホルマントトラッキング

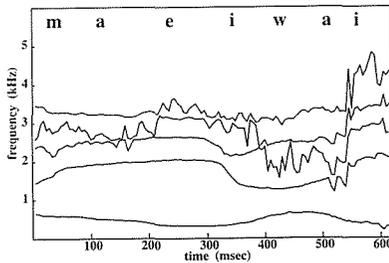


図6 「前祝い」 (/maeiwai/)
のホルマントトラッキング

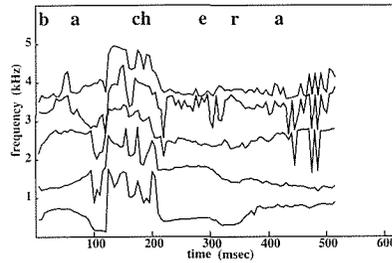


図7 「バチエラー」 (/bachera-/)
のホルマントトラッキング

そこで本研究では、半母音と母音を分類するために、ホルマントを時間の関数と考え、時間で2階微分を施したものをホルマント周波数の時間変化を捉えるパラメータとする。時刻 t における第 j ホルマントのパラメータ $h_j(t)$ を

$$h_j(t) = \frac{\frac{F_{j,t+\Delta t} - F_{j,t}}{\Delta t} - \frac{F_{j,t} - F_{j,t-\Delta t}}{\Delta t}}{\Delta t} \quad (2)$$

として定義する。

ただし、 $F_{j,t}$ は時刻 t における第 j ホルマントの周波数を表す。このパラメータを用いることにより、音素の定常部付近において母音は $h_j(t) = 0$ の値をとり、半母音/w/は $h_2(t) > 0$ に、半母音/y/は $h_2(t) < 0$ 、/r/は $h_1(t) > 0$ の値をとることとなる。しかしながら、各半母音は発話速度の違いによって上記のパラメータ Δt を一定とすることはできない。そこで、パラメータ $h_j(t)$ を放物線の方程式を2階微分した微係数と解釈する。こう解釈した上で、パラメータ h_j は最小2乗法を用いてホルマント周波数を放物線で近似したときの微係数として求められる。このパラメータを導入することによって、 $V_1V_2V_3$ と V_1CV_3 との識別分類だけでなく、多くの場合には半母音と母音を識別分類することが可能となる。

多くの場合はパラメータ $h_j(t)$ を用いることにより半母音と母音を識別分類が可能となるものの、

図8に示すように母音に鼻音が後続した場合には、半母音に母音が後続した場合と似たようなホルマントの遷移を示すことがある。このような場合には式(2)で定義したパラメータだけでは半母音と母音を分類することは難しい。すなわち、ホルマント周波数やその時間変化を考慮したパラメータを用いるだけでは、半母音と母音を分類することは難しい場合がある。そのため新たなパラメータを用いることが必要となる。ここでは、ホルマントの振幅に着目する。図9に単語「プロ野球」(/puroyakyu-/)の各ホルマントの振幅の遷移を示す。この図から、母音から半母音への過渡部においては/r/, /y/のどちらの場合にも第2ホルマントの振幅が小さくなるのがわかる。これは半母音/w/についても同様のことがいえる。このことに着目して、一時的に振幅の小さくなる状態をホルマントの時間変化を捉えるパラメータと同様な形で

$$p(t) = \frac{\frac{P_{2,t+\Delta t} - P_{2,t}}{\Delta t} - \frac{P_{2,t} - P_{2,t-\Delta t}}{\Delta t}}{\Delta t} \tag{3}$$

として表現し、半母音を識別するための第2のパラメータとする。ただし、 $P_{2,t}$ は第2ホルマント

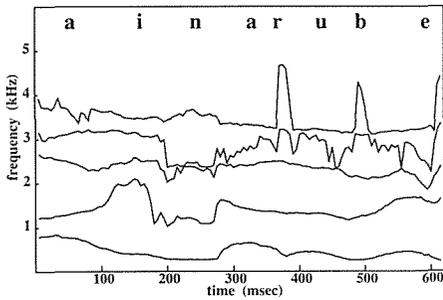


図8 「相成るべくは」 (/ainarubekuwa/)のホルマントトラッキング

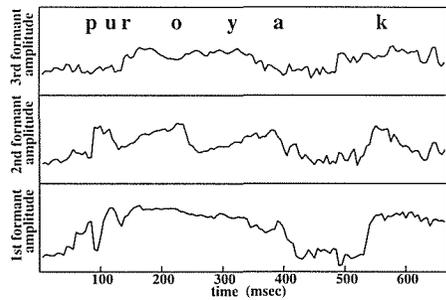


図9 「プロ野球」 (/puroyakyu-/)のホルマント振幅の時間変化

の振幅を示す。このパラメータ $p(t)$ が正のときその音素は半母音の候補となる。このパラメータにおいてもホルマントの時間変化を捉えるパラメータと同様に Δt を一定とすることはできない。これら2つのパラメータを第1, 第2ホルマントに加えて用いることにより、母音と半母音を識別分類することが可能となる。

4. 実験および実験結果

本研究では音声資料として電総研の研究用音声データベースを用いている⁵⁾。この音声データベースは話者1人につき492個の単語が含まれている。各単語はサンプリング周波数15kHz, 14ビットに量子化された原音声とその音声にラベリングを施した音素片ファイルから成り立っている。図10に原音声とラベリングの対応を示す。ラベリングは各音素の定常部と過渡部に分けて付けられており、それぞれの開始フレーム番号と終了フレーム番号が付けられている。ただし、1フレームの間隔は5ms(75点)である。図中では、AA, YYなど、同じ記号が繰返されている部分が定常部であり、他の部分が過渡部となっている。本実験ではこのラベリングの中から過渡部及び定常部の開始フレームとそのフレームが母音または半母音を含んだフレームであるか否かという情報を用いて認識実験を行なう。すなわち、入力される未知の音素は母音か半母音のどちらか

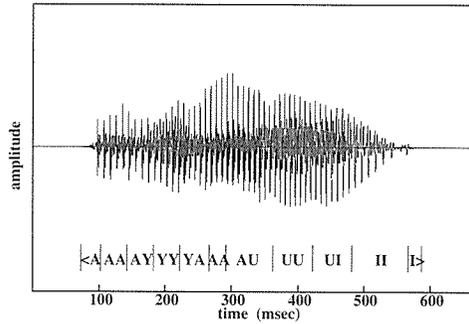


図10 原音声とラベリングの対応例

であることを仮定している。

母音の識別領域は1名の話者が発声した合計492単語の中から母音の定常部を抽出し、その部分のホルマント周波数を得て、その共分散をもとに母音識別領域を F_1 - F_2 平面上に構成した。それぞれの母音の識別楕円の大きさはマハラノビスの距離がほぼ一定となるように設定している。しかしながら、半母音の識別領域については音声資料が少ないため、マハラノビスの距離は母音とは違う値となっている。

認識実験は、話者3名に対して行なった。実験は1フレームを17.1ms (256点)として、フレーム間隔を5 ms (75点)とし、Durbin法⁶⁾を用いて線形予測分析を行い、得られた線形予測係数を係数とする高次方程式を解くことにより、ホルマント周波数およびその帯域幅やパワーを計算している。さらに、式(2)、(3)の二つのパラメータを計算する。これらのパラメータの値から半母音と判断したものに対しては、半母音の識別領域/y/, /w/, /r/のうち、どの識別領域に入るかを調べる。母音に対しても同様のことを行なう。

このような方法により識別実験を行った結果を表1に示す。認識率は84.6%である。誤認識のうち、29.5%は半母音を母音と認識するか、逆に母音を半母音として認識したことによる誤りである。これは本論文で提案したパラメータにより半母音を抽出できなかつたり、誤って抽出してしまった誤りに相当する。しかし、この誤りの多くは余分なホルマントの抽出やホルマントの欠落に起因するものであり、ホルマントの自動抽出についても今後の課題とすべきである。また、誤認識の多くは隣接する識別領域としての誤りであり、領域の設定の方法についても検討する余地がある。

5. ま と め

本研究では音響的に似た性質をもつ母音と半母音を識別・分類するためのパラメータとして、ホルマント周波数やそのパワーの2階微分に相当するパラメータを提案した。このパラメータにより母音と半母音をある程度分類できた。しかしながら、分類できなかった音素に対しては頑健なホルマントの自動抽出を考慮しながら、さらにパラメータを改善する必要がある。また母音、半母音どちらの場合においても、隣接する識別領域の間における誤識別が目立っている。今後、この問題にも対処していく必要がある。

表1 単語中の母音・半母音の認識結果

in \ out	/a/	/i/	/u/	/e/	/o/	/y/	/w/	/r/	del
/a/	1010	0	14	30	24	6	8	8	12
/i/	0	559	45	82	1	47	1	16	1
/u/	1	11	612	16	47	5	4	21	23
/e/	19	7	25	688	4	27	0	12	7
/o/	22	1	45	12	854	0	6	25	27
/y/	1	6	0	1	0	40	0	4	1
/w/	0	0	0	0	1	0	11	3	0
/r/	0	4	8	0	0	8	1	80	0

参考文献

- 1) E. C. Schwab, J. R. Sawusch and H. C. Nusbaum, The Role of Second Formant Transitions in the Stop-Semivowel Distinction. *Percept. Psychophys.*, 29(1981), 121-128.
- 2) 松村雅史・山根浩顕・藤井克彦. スプライン関数によるホルマント周波数軌跡の抽出と連続音声での母音・半母音の認識. *電子情報通信学会論文誌 (A)*. J72-A(1989), 12-22.
- 3) C. Y. Espy-Wilson, APhonetically Based Semivowel Recognition System. *IEEE-IECEJ-ASJ International Conference on Acoustics, Speech and Signal Processing*, 1986, 2775-2778.
- 4) 外山 淳・相良浩之・新保 勝, 時変パラメータを用いた音声空間における半母音の識別. 平成4年度春季研究発表会講演論文集, 1992, 169-170.
- 5) 速水 悟・田中和世・横山昌一・太田耕三, 研究用音声データベースのためのVCV/CVCバランス単語セットの作成. *電総研彙報*, 49(1985), 804-834.
- 6) 安居院猛・中嶋正之, *コンピュータ音声処理*. 産報出版, 1980.