論文 / 著書情報
Article / Book Information

論題(和文)	音声知覚における母音ターゲット予測機構のモデル化の検討
Title	
著者(和文)	赤木 正人, 古井 貞熙
Author	SADAOKI FURUI
出典(和文)	聴覚研究会資料(H-85-35), , , pp. 263-270
Journal/Book name	, , , pp. 263-270
発行日 / Issue date	1985, 9

日本音響学会音声研究会資料 S85-35(September 19, 1985) ／聴覚研究会資料 H-85-35

音声知覚における母音ターゲット予測機構のモデル化の検討

赤木 正人　古井 貞熙
（NTT武蔵野電気通信研究所）

内容梗概

本報告では、心理学的知見、生理学的知見を考慮しつつ、母音ターゲット予測機構のモデル化を行い、聴覚心理実験結果との比較により、本モデルの評価を行った。この結果、本モデルは音声の物理的特徴の変化を臨界制動2次系モデルで近似した場合、遷移開始時点の情報は予測にとって不要であり、ターゲットの予測値を短時間区間（50ms）で推定できるモデルとなっている。さらに、本モデルはわたり音区間の減少特性、単音節における後続母音（特に拗音/j/の後の母音/u/）のなまけの回復特性に優れており、聴覚心理実験結果と良く対応するモデルであることが明らかとなった。

Transactions of the Committee on Speech Research／Hearing Research
The Acoustical Society of Japan, S85-35(September 19, 1985) ／H-85-35

A Study on Modeling of Vowel Target Prediction Mechanism in Speech Perception

Masato AKAGI and Sadaoki FURUI
(Musashino Electrical Communication Laboratories, NTT)

Abstract In this paper, a model of a vowel target prediction mechanism based on psychological and physiological knowledge is proposed and evaluated by comparing the predicted values based on the model with the results of auditory psychological experiments. When the trajectories of physical characteristics of speech are approximated by a 2nd-order critical damping model, the proposed prediction model can estimate the vowel target values based on a speech wave of a short period (50 ms) without the knowledge of transition onset position. Additionally, this model decreases the length of transitional periods which produce spurious vowels and recovers vowel characteristics in Japanese C-V syllables neutralized by the coarticulation; especially for vowel /u/ in syllables including /j/. Consequently, it can be concluded that this model is a good approximation of the target prediction mechanism in speech perception.

1．まえがき

連続音声を分析しその物理的特徴を抽出した場合、音声の調音器官の制約から、いわゆる「なまけ音」、「わたり音」などと呼ばれる調音が完全でない音声区間が出現する。しかしながら、人間が連続音声を聞く場合、音声の物理的特徴が個々の音韻の物理的特徴に到達していないにもかかわらず、人間はあたかもその音が発声されたかのように知覚する（なまけの回復）。逆に、調音結合部分において、物理的特徴がわたり音の特徴に接近するにもかかわらず、人間はその音をほとんど知覚していない。たとえば、連続母音/aia/を発声したとき[aia]と聞えていても、物理的特徴は[aea]であったり、[i]が存在したとしてもきわめて短い時間であったりする。また、/a/から/i/への調音結合部分で/e/の物理的特徴は現われていても、わたり音/e/は知覚されない。この現象は、聴覚処理系内にある種の補正機構が存在し、補正された特徴が知覚されているために生じる、と考えられている[1][2][6]。

この補正機構のモデル化が実現されれば、調音結合の影響が大きい連続音声認識を容易とするなど、応用範囲は大きなものとなる可能性があるが、現時点では、桑原[3]～[5]、藤崎他[6]などがモデル化を試みてはいるものの、上記の現象を良く説明し応用上有効なモデルは未だに得られていない。

本報告では、新たに、この補正は「聴覚系内に音韻ターゲットの予測機構が存在し、この予測値を人間は知覚している」ために生じると仮定し、心理学的知見のみならず聴覚末梢系に関する生理学的知見をも考慮することにより、ターゲット予測機構のモデル化を行う。

2．聴覚末梢系の生理学的モデル

聴覚末梢系をモデル化するために、Klatt[7]が提唱した聴覚の工学的モデルのうち、

(1) 有毛細胞をモデル化するための半波整流器、

(2) ラウドネス尺度近似のための対数変換、

(3) 基底膜をモデル化するためのメル尺度、

(4) 側抑制をモデル化するための周波数領域での重み付け、

を用いた。

(1)(2)については、8kHzでサンプリングされた音声波形を10次でＬＰＣ分析し、ＬＰＣケプストラムを求めた後逆フーリェ変換を行って、対数スペクトルyを求めることにより実現している。(3)については、周波数軸上の単位区間ごとに線形伸縮を行い、メル尺度に近似した特性をスペクトルyに持たせることにより実現している。(4)については、側抑制のモデル[8]

$$z(x) = y(x) - \int_{-\infty}^{\infty} G(x - x')\, y(x')dx'$$

$$G(x) = \frac{G}{2a} \exp(-\frac{|x|}{a})$$

を離散系で表現することにより実現している。

一方Dallos等[9]によれば、外有毛細胞は基底膜の変位に対する感覚器であり、内有毛細胞は変位の速度に対する感覚器であることが明らかとなっている。また前者は求心性線維、遠心性線維両方と結合し、後者は求心性線維のみと結合していることが知られている。そこで、この生理学的知見も上記の知見に組み込み、図1に示すような聴覚末梢系のモデルを構成した。本モデルは、基底膜の変位をメル周波数軸上に射影した対数スペクトルyに対応させている。

$\dot{y}$は本来スペクトルyの時間方向の微分値を表わすものであるが、計算の安定化のために、計算区間内で1次の回帰モデルを当てはめ、その回帰係数を微分値として扱っている。例として、二連母音/ai/の場合の末梢系モデルからの出力を図2に示す。

3．予測モデル

3．1　予測モデルの基本原理

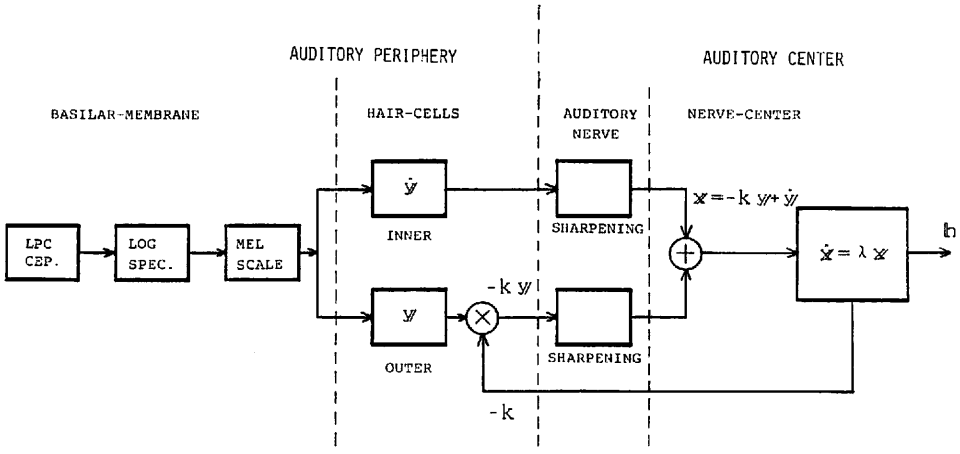

Fig.1-Modeling of vowel target prediction mechanism.

予測モデルを作るにあたって、次の点を考慮した。

(1) ホルマントの変化だけでなく、スペクトル全体の変化を扱う。

(2) スペクトル変化極大点が前出音と後続音の知覚を分ける重要な時点である[10]。

(3) 変化量と変化速度には相補的関係がある[1]。

(4) 音声の物理的特徴量の変化は、臨界制動2次系で近似できる[11][12]。

(5) 予測には短い時間幅の情報だけを用いる[13]。

上記(1)～(4)の点を踏まえ、図1の聴覚末梢系モデルの出力スペクトルy、$\dot{y}$を用いて、次のようなスペクトル変化の関係式を導出した。

y_nを時刻nのスペクトル、$\dot{y}_n$をスペクトル変化速度、kをフィードバックの係数、λを変化時定数として、

$$x_n = -ky_n + \dot{y}_n \qquad (1)$$

$$\dot{x}_n = \lambda x_n \qquad (2)$$

の関係が成り立つとする。式(1)は、内外有毛細胞の機能差を考慮した式となっている。また式(2)は、

(a) $\lambda > 0$、$\lambda < 0$で到達方向と到達点が異なり、範疇判断[14]が行われるモデルとなっている。なお、λの符号はスペクトル変化極大点で入れ換わる。

(b) 時定数が大となれば変化速度小あるいは変化量大、時定数が小となれば逆の関係となり、相補的関係が記述できる。

という特徴を有する。

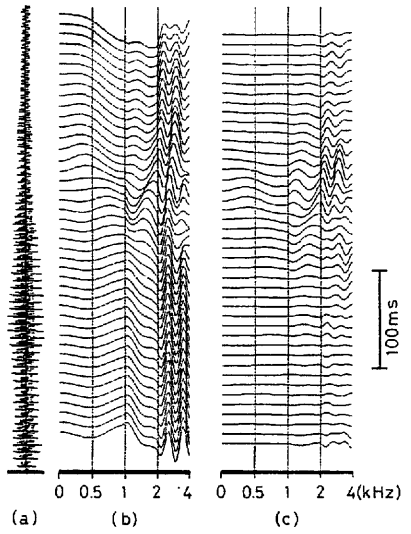

Fig.2-Output spectrum sequences associated with the peripheral auditory model for a diphthong /ai/. (a) *wave form,* (b) *spectrum sequence of the outer hair-cells model,* **y**, (c) *differential spectrum sequence of the inner hair-cells model,* **y**.

式(1)においてk＝λとおき、式(2)に代入すれば、

$$\ddot{y}_n - 2\lambda\dot{y}_n + \lambda^2 y_n = 0 \qquad (3)$$

となり、臨界制動2次系を扱うモデルとなる。式(3)を解けば、

$$y_n = a(1 - \lambda n)\exp(\lambda n) + b,\ n \geq 0 \qquad (4)$$

となる。ここでbがターゲットの予測値を表わす。

音韻知覚区間長に関する知見[13]によれば、知覚時間区間は50～70msと言われており、筆者等の検討[15]においても同様の結果が得られている。もし、50～70msでbの推定が可能であれば、予測モデルを構成するにあたって考慮した点(1)～(5)をすべて満足する予測モデルが構成できる。次節では、予測値bの短時間区間推定法について述べる。

3．2 ターゲット予測値の短時間区間推定法

臨界制動2次系を記述する微分方程式は、

$$\left(\frac{\partial^2}{\partial t^2} - 2\lambda\frac{\partial}{\partial t} + \lambda^2\right) y = 0 \qquad (5)$$

である。上式を解けば、

$$y = a(1 - \lambda t)\exp(\lambda t) + b \qquad (6)$$

となる。

式(5)を変形すれば、

$$\left(\frac{\partial}{\partial t} - \lambda\right)\left\{\left(\frac{\partial}{\partial t} - \lambda\right) y\right\} = 0 \qquad (7)$$

となる。ここで、

$$x = \left(\frac{\partial}{\partial t} - \lambda\right) y \qquad (8)$$

とすれば、式(7)に代入して、

$$\left(\frac{\partial}{\partial t} - \lambda\right) x = 0 \qquad (9)$$

となる。また、式(6)を式(8)に代入すれば、

$$x = -a\lambda\exp(\lambda t) - \lambda b \qquad (10)$$

となり、式(10)は式(9)を満たす。

ここで式(9)に注目すれば、式(9)は明らかに1次系の式である。式(8)に示すyからxへの変換が可能であるならば、臨界制動2次系を1次系に変換してパラメータ推定を行うことができる。

今、仮にλがある値に固定されているとしよう。

スペクトル系列y_nに対して式(8)を適用すれば、スペクトル系列x_nが求まる。このx_nに対して、式(9)を考慮して1次系の式

$$\hat{x} = \lambda\hat{a}\exp(\lambda n) + \lambda\hat{b} \qquad (11)$$

を当てはめる。

このとき、

$$e(\lambda) = \sum_{n=-N/2}^{N/2} | x_n - \hat{x}_n |^2$$

を極小とするように最小自乗法を用いれば，固定されたλに対して最適な$\hat{a}$、$\hat{b}$とそのときの誤差e(λ)が求まる。

λの決定は、この誤差e(λ)を最小化するように、λについての非線形最適化問題を解き、推定している。なお本問題では、単峰性が保証されていないため、λを微小区間に分けてその中で最適化を行っている。

誤差が最小となったとき、式(6)(10)の関係から、λは臨界制動2次系の時定数の逆数、$\hat{b}$はターゲット予測値bとーλの積となる。また$\hat{a}$は、式(6)の遷移開始時点と推定計算のための区間の中心との差をt_0とした場合、

$$\hat{a} = a \exp(\lambda t_0) \qquad (12)$$

の関係がある。

本計算法は、臨界制動2次系のパラメータλ、a、bの推定を指数関数のパラメータ推定問題に置き換えて解く方式である。

臨界制動2次系のパラメータ推定に関する従来の方法[11][12]では、式(6)をそのまま用いてすべてのパラメータを直接推定しようとしていたために、遷移開始時点を決定する必要があった。このため，遷移開始時点を含む長い区間でのパラメータ推定しか行えず、50～70msの短時間区間でのパラメータ推定は不可能であった。

しかしながら、パラメータ推定の本来の目的を考えてみれば、ターゲット予測値bさえ求められれば良い。本方式は上で示したように、予測値bを求めるときに遷移開始時点を特定する必要がない。このため、短時間区間でのパラメータ推定が可能となり、特に5．2節で述べるように区間長50msの分析で聴覚心理実験との対応が最も良くなることが示せる。

本手法の性能を調べるために、シミュレーション実験を行った。図3に結果を示す。図中の細い実線はステップ関数、破線は臨界制動2次系関数を表わす。また、一点鎖線はx、太い実線はbを表わす。図から分かるように、bはステップ関数と良い一致を示しており、良好に予測が行われていることが分かる。なお、この図で1pointを1msと考えれば、音声の物理的特徴の変化を良く近似するモデルとなっている。

実際のスペクトル系列に本手法を適用した例を図4に示す。図は後述するData1中の二連母音/ai/の分析結果であり、図2(b)と図4(c)を比較すれば、図4(c)の予測値の系列が音韻が変化するあたりで急激に変化していることが分かる。

4．聴覚心理実験

人間が音声を聞く場合、どの時点から後続音を聞き始めているのか、あるいはどの時点まで前出音を聞いているのかを知るために、次のような条件の下で実験を行った。

刺激音

刺激音として次の3種類を用意した。

◆Data1：男性話者1名が発声した二連母音を、音素境界を含む25時点（間隔は8ms）で語尾切断したデータ計500個。なお、切断部分は8msの線形な傾斜を用いている。Data1は、実験用の様々なパラメータを決定するための予備実験に主に用いる。

◆Data2：女性アナウンサー2名が発声した100音節を10msごとに語尾切断したデータ。なお、切断部分は10msの線形な傾斜を用いている。

◆Data3：表1に示す対称形三連母音を含む単語を男性3人、女性1人に発声してもらい原データとした。刺激音はContextの影響を取り除くために、フレーム長60ms（立上り20ms、立下

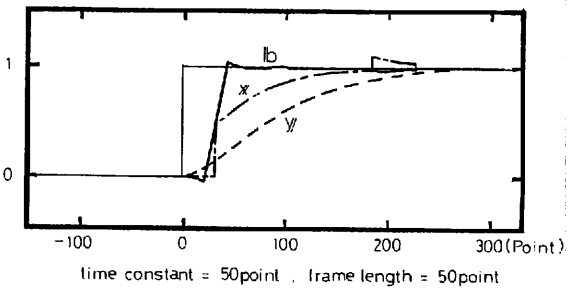

Fig.3-Simulated results of the proposed prediction method based on the 2nd-order critical damping model.

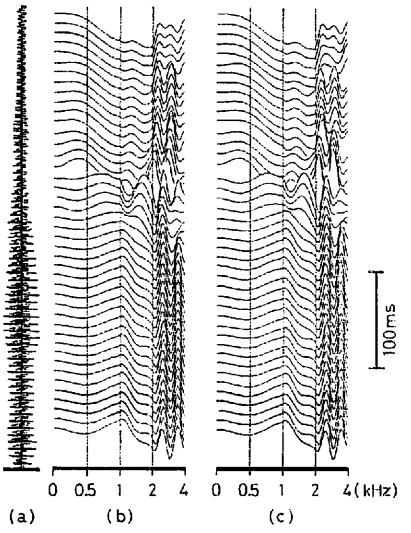

Fig.4-Output spectrum sequences associated with the 2nd-order critical damping model for a diphthong /ai/. (a) *wave form,* (b) *spectrum sequence,* **x**, (c) *estimated spectrum sequence,* **b**.

り２０ms）の台形窓を用いて原データから切り出している。なお、フレーム周期は１０ms、刺激音総数は発声者１人について８００個である。

実験方法

◆Data１：聴力が正常な成人男性１２名に対して、５００個のデータをランダムに呈示。

◆Data２：通話試験員４名に対して、発声者別に種々の刺激のうちほぼ同程度の明瞭度となる１００音節を組として、その中をランダムにならべ４～５回呈示。

◆Data３：通話試験員４名に対して、発声者１人について４回、１回８００個の刺激音を呈示。

受聴は３データとも聞き易いレベルとし、雑音は加えていない。被験者及び試験員に対しては、聞えた音声をそのまま記述してもらうこととした。得られたデータから、Data１、Data２については後続音の明瞭度、Data３については後続音、前出音双方の明瞭度を求め、前出音が聞えなくなる限界点及び後続音が聞え始める限界点を特定した。ここではこの時点を聴覚的基準点（auditory reference point）と呼ぶことにする。なお、基準点を表示する音声波形上の位置は、Data１、Data２については切断部の最後部、Data３については切り出しフレームの中心としている。図５にそれぞれのデータについての切り出し用窓と基準点の関係を示す。実験から求めた基準点の標準偏差は、Data１では１８．５ms、Data３の前出音について１６．０ms、後続音について１６．３msである。

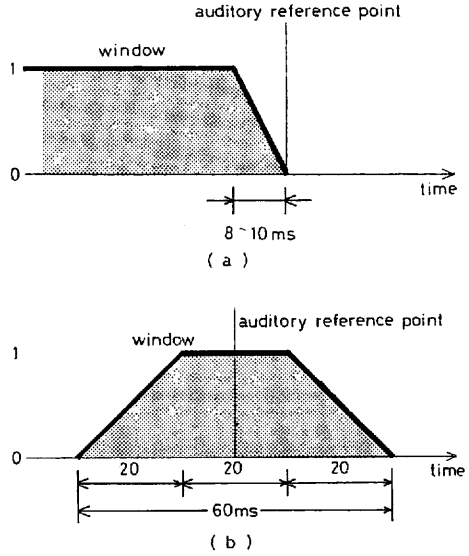

Fig.5-Truncate window used for the auditory experiments and auditory reference point used for evaluation of the proposed model. (a) *for Data1 and Data2,* (b) *for Data3.*

Table 1-Short utterances including $V_1V_2V_1$ type vowel concatinations used for the experiments.

/a/	気合い	/iai/
	（駅で）よく合う（人）	/uau/
	（その仕事は）請け合える	/eae/
	ドアを（あける）	/oao/
/i/	大安	/aia/
	初初しい	/uiu/
	永遠	/eie/
	恋を（する）	/oio/
/u/	買うあて（がある）	/aua/
	言う意志（がある）	/iui/
	目上え	/eue/
	糸魚	/ouo/
/e/	栄えある（賞）	/aea/
	消え入る	/iei/
	杖打つ（人）	/ueu/
	声を（出す）	/oeo/
/o/	顔合わせ	/aoa/
	におい（ぶくろ）	/ioi/
	犬追う（人）	/uou/
	（荷物を）背負え	/eoe/

5．モデルの評価

5．1　モデルの評価基準

モデルから得られた予測値bと発声者ごとに求めた５母音のカテゴリ中心との距離を測り、目的の音が他の４母音との距離より小さくなる時点を本モデルにおける前出音・後続音に対する限界点とする。この時点を物理的評価点（physical estimation point）と呼ぶ。図６(a)にData３中の単語/kiai/を本モデルにより分析し、５母音との距離を計算した結果を示す。図中実線Aが後続音/a/に対する評価点である。また図中破線Bが聴覚心理実験から求めた限界点すなわち基準点である。本モデルの評価には、評価点と基準点との差A－Bの平均値と標準偏差を用いる。なお、差が正の場合は評価点の方が基準点より時間的に後方にあることを示す。

5．2　予測値算出のための区間長

ターゲット予測値bを算出するための短時間区間推定法を３．２節で示したが、最適な短区間長はどれくらいの長さであるかは明らかでない。そこで、Data１を用いて、区間長を変化させながら聴覚心理実験結果との対応関係を調べた。結果を図７に示す。図は横軸に区間長、縦軸に評価点と基準点の差の標準偏差を目盛ってある。図から明らかなように、区間長５０msのときの標準偏差が最も小さい。この結果から、実験には５０msの区間長を用いることにする。図４(b)(c)及び図６(a)(c)は、区間長を５０msとして計算したものである。

区間長50msは、音韻知覚区間長に関する知見[13][15]と良く対応している。このことは、本モデルが心理的な聞えに良く対応できるモデルであることの一端を示している。

5．3　ターゲットの安定性

5．1節で示した図6(a)は5母音のカテゴリ中心と本モデルによる予測値bとの距離系列、(b)はスペクトルyとの距離系列、(c)は1次系モデル[16]による予測値bとの距離系列を示してある。ここで、図(a)と(c)を比べれば、(a)すなわち本方式の方が距離が安定に計算されている。

これは、スペクトルの変化自体が臨界制動2次系で近似できることにも関係があるが、予測値の計算方法に対しても、次のような関係がある。

1次系モデルを用いた方式では、微分方程式

$$\dot{y}_n = \lambda y_n$$

を解こうとしているが、λが0に近づく時点、すなわちスペクトルyの時間変化がなくなる時点あるいは変化極大点で微分方程式が不安定となり、このため図6(c)でも距離系列が不安定となっている。一方臨界制動2次系モデルでは、図3、図4のλに関する図でも明らかなように、式(1)によってスペクトルの変化を強調した上で微分方程式(2)を解いているため、λの絶対値が1次系モデルより大きくなり、0に近づかなくなる。この結果、計算結果が安定となっている。

5．4　わたり音区間の減少

聴覚心理実験結果を見ると、わたり音はほとんど知覚されていない。たとえば図6に示した単語/kiai/について、/i/から/a/に変化する時点あるいは/a/から/i/に変化する時点で、聴覚実験ではわたり音/e/は1フレームないし2フレーム（1フレーム10ms）しか知覚されていなかった。一方、スペクトルyと5母音のカテゴリ中心との距離系列を示した図6(b)では、わたり音/e/が広い範囲（6～7フレーム）に存在する。ところが本モデルによる結果図6(a)では、わたり音/e/はほとんど存在しない。

Data3全データ（160個）中のわたり音が認められるデータに対して、わたり音長の平均とその標準偏差を求め、わたり音の減少傾向を見ると、表2のようになった。表2の最上段は、160個のデータ中でわたり音が認められた数である。1次系モデルの方が臨界制動2次系モデルよりもわたり音数が見かけ上少なくなっているのは、5．3節で述べたように、1次系モデルの方が、スペクトル変化極大点でターゲットが安定に計算されないために、この時点で音韻が判別できないことによる。中段、下段はそれぞれわたり音長の平均と標準偏差である。表中の平均の項を見ると、明らかにわたり音区間の減少傾向が認められる。また、標準偏差の項を見ると、1次系モデルよりも本モデルの方が安定性に優れていることが、この結果からもうかがえる。

これらの結果は、本モデルがわたり音区間を減少させ、心理的

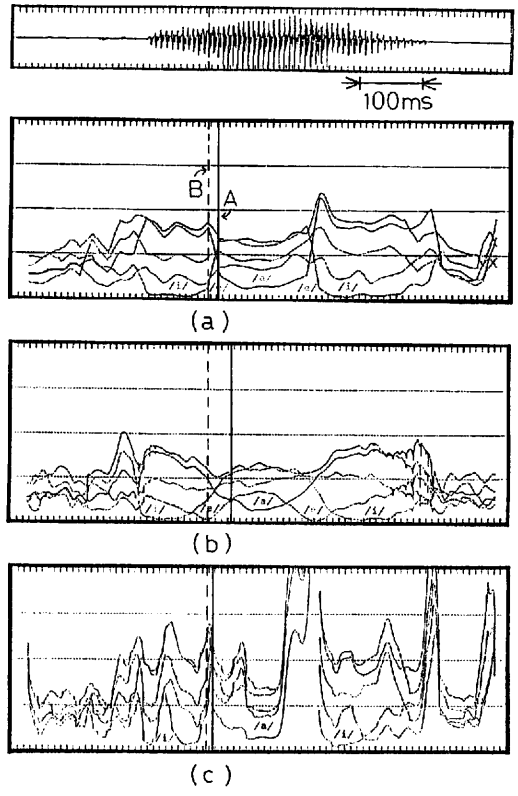

Fig.6-Distance sequences between each vowel category center and (a) estimated value **b** *based on the critical damping model, (b) spectrum* **y**, *and (c) estimated value* **b** *based on the 1st-order differential equation model for a word utterance /kiai/.*

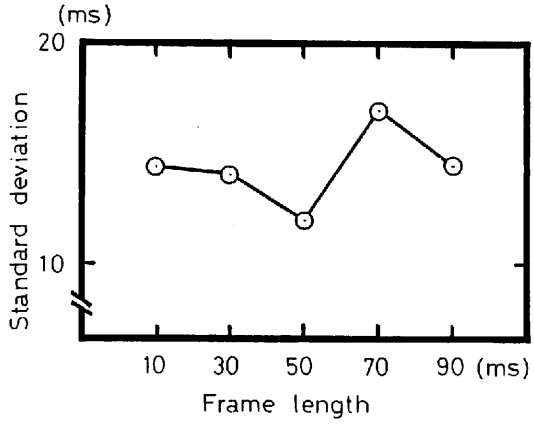

Fig.7-Standard deviation of the difference between the auditory critical point and its estimation based on the proposed model for five conditions of the frame length.

な聞えに良く対応できるモデルであることを示している。

5．5 基準点と評価点の差の平均及び標準偏差の減少

Data2およびData3を用いて，基準点と評価点の差の平均と標準偏差により本モデルの評価を行った。表3に結果を示す。表(a)はData3を用いた結果，表(b)はData2を用いた結果である。表から明らかなように，標準偏差が，予測を用いないものが最も大きく，次に1次系モデルを用いたもの，そして本モデルが一番小さくなっている。これは，ターゲット予測値の安定性，わたり音区間の減少の両特性が有効に働き，基準点と評価点の関係が一定に保たれるようになったためであり，本モデルがより聴覚特性に近づいた結果である，と言える。

表(b)では，差の平均の値がだんだんと小さくなっている。これは特に拗音/j/を含む音節で顕著である。拗音/j/を含む音節では，調音結合のなまけ現象により，後続母音の物理的特徴がその母音のターゲットに到達しないことがあるが，本モデルを用いることにより，より速くターゲットに到達できるようになり，差の平均の値が小さくなったと考えられる。この結果は，本モデルが単音節における後続母音のなまけの回復特性に優れており，調音結合によって後続母音に到達しない物理的特徴を本モデルにより聴覚特性に近づけることができることを示している。

6．あとがき

母音ターゲットの予測機構については，様々な文献で指摘されており，その存在も明らかとなりつつある。本報告では，この予測機構を生理学的・心理学的知見を考慮した工学モデル上に構成し，臨界制動2次系モデルで記述されたターゲットの予測値を短時間区間で推定できる計算方式を示した。また，聴覚心理実験から得られた結果との比較によりモデルの評価を行った。この結果，本モデルがわたり音区間の減少特性，なまけ音の回復特性に優れており，聴覚心理実験結果と良く対応するモデルであることが明らかとなった。

しかしながら本モデルは，生理学的・心理学的知見の一部を取り入れているとはいえ，完全とは言いがたい。今後さらに新しい知見を導入し，予測機構のみならず他の機構，たとえば，同化効果，対比効果などについてもモデル化を進め，心理実験による検証を加えて行きたい。

謝辞

日頃ご指導いただく情報通信基礎研究部第四研究室・篭室長をはじめ，研究室の諸氏ならびに実験を担当された通話試験員諸嬢に感謝いたします。

参考文献

[1] 鈴木：" 母音・半母音・有声破裂音の知覚におけるホルマント遷移の変化量と変化速度との間の相補性およびその識別機構"，音響学会誌，30，3（1974）

[2] 桑原・境：" 動的合成母音による音韻境界の知覚的検討"，音響学会誌，31，1（1975）

[3] 桑原・境：" 連続音声中の母音連鎖における調音結合効果の正規化"，音響学会誌，29，2（1973）

[4] 桑原：" 連続音声中の母音知覚と特徴抽出の一試み"，信学会論文誌A，Vol.61-3（1978）

[5] 桑原：" 聴覚特性による連続音声中の母音の特徴表現と認識"，音響学会音声研資，S84-79（1985）

[6] 藤崎・樋口・二見・重野：" 非定常音刺激の知覚とそのモデル"，音響学会音声研資，S81-35（1981）

[7] D.H.Klatt：" Speech Processing Strategies based on Auditory Models"，Speech Communication Group Working Paper II（1983）

[8] 塚原：" 脳の情報処理"，朝倉書店（1984）

[9] P.Dallos et.al.：" Cochlear Inner and Outer Hair Cells

Table 2-Number of transitional periods which produce spurious vowels, and mean and standard deviation of the length of the spurious vowel intervals. Total number of transitional periods is 160.

	PREDICTION		
	NO	FIRST-ORDER	CRITICAL-DAMPING
NUMBER OF SPURIOUS VOWELS	64	51	55
MEAN LENGTH	61.6 ms	27.9	27.2
STANDARD DEVIATION	24.6 ms	20.3	16.3

Table 3-Mean (μ) and standard deviation (σ) of the difference between the auditory critical point and its estimation.

(a) $V_1V_2V_1$ *type vowels.*

	PREDICTION		
	NO	FIRST-ORDER	CRITICAL-DAMPING
PRECEEDING VOWELS	μ = 14.2 ms σ = 28.2 ms	22.8 27.8	22.0 20.4
FOLLOWING VOWELS	16.7 36.0	11.2 22.2	9.8 20.5

(b) *Japanese C-V syllables.*

	PREDICTION		
	NO	FIRST-ORDER	CRITICAL-DAMPING
SYLLABLES WITHOUT /j/	μ = 2.43 ms σ = 28.95 ms	-10.20 20.34	-7.97 18.22
SYLLABLES WITH /j/	42.02 29.14	23.98 13.64	15.18 12.20
ALL	17.05 34.74	1.25 25.78	-0.38 20.95

: Functional Differences",SCIENCE, Vol.177, pp..356-359 (1972)
[10] 古井：" 音声知覚におけるスペクトル変化情報の役割り"、音響学会聴覚研資、H85-6（1985）
[11] 藤崎・吉田・佐藤・田辺：" ホルマント周波数領域における調音結合のモデルとその母音連続音声の認識への応用"、音響学会音声研資、S73-03（1973）
[12] 板橋・横山：" 線形2次系モデルによるホルマント軌跡の記述とセグメンテーション"、電総研彙報、40,6（1976）
[13] 難波：" 聴覚ハンドブック"、第7章、ナカニシヤ出版（1984）
[14] 藤崎・川島：" 合成音声の弁別と言語音知覚機構のモデル"、音響学会誌、27、9（1971）
[15] 古井：" 音韻知覚を決定づける音声区間の特性について"、音響学会秋季講論（1985）
[16] 赤木・古井：" 音声知覚における母音ターゲット予測機構のモデル化"、音響学会春季講論 3-5-16（1985）