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
Title(English)	
著者(和文)	赤木正人, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J69-A, No. 10, pp. 1277-1285
Citation(English)	, Vol. J69-A, No. 10, pp. 1277-1285
発行日 / Pub. date	1986,
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1986 Institute of Electronics, Information and Communication Engineers.

論　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

正員　赤木　正人†　　正員　古井　貞熙†

Modeling of Vowel Target Prediction Mechanism in Speech Perception

Masato AKAGI† *and* Sadaoki FURUI†, *Members*

あらまし　「人間は，連続音声を聞く場合なまけ音，わたり音などと呼ばれる調音が完全でない区間の音声をある種の補正機構によって補正し，知覚している」との仮説にそってモデル化を検討した．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．本論文では，聴覚に関する心理学的知見，また聴覚末梢系に関する生理学的知見を考慮しつつ，補正機構の1つである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，本モデルの検証を行った．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を用いることなく，ターゲットの予測値を短時間区間（50 ms）で推定できるモデルとなっていること，さらに，本モデルは，わたり音区間の減少特性，単音節における母音調音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．

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，調音結合のために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），たとえば，連続母音 /aia/ を発声したとき，分析された物理的特徴は［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が，人間は［aia］と知覚する．また，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（わたり音の減少），たとえば，連続音声 /ai/ を発声したとき，/a/ から /i/ への調音結合部分で /e/ の物理的特徴［e］は現われていても，人間はわたり音［e］をほとんど知覚しておらず，［ai］と知覚する．

これらの現象は，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，と考えられる[1]～[4],[8]．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の機械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[5]～[7]，藤崎他[8]などが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．

本論文では，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，ターゲット予測機構のモデル化を行った．モデル化にあたっては，聴覚末梢系に関する生理学的知見および音声知覚に関する心理学的知見両方を考慮したが，聴覚末梢系の生理学的構造，また脳中枢の構造を忠実にモデル化するのではなく，母音ターゲットの予測機構という1つの工学モデルを，末梢から中枢までの処理を含めた全体的な機能のモデルとして構成しようと考えた．

この結果，本モデルは音声の物理的特徴の変化を臨界制御2次系モデルで近似した場合，臨界制動2次系を1次系に変換することにより，遷移開始時点の情報を用いることなくターゲットの予測値を短時間区間（50 ms）で推定できるモデルとなった．また，聴覚

†NTT電気通信研究所，武蔵野市
NTT Electrical Communications Laboratories, Musashino-shi, 180 Japan

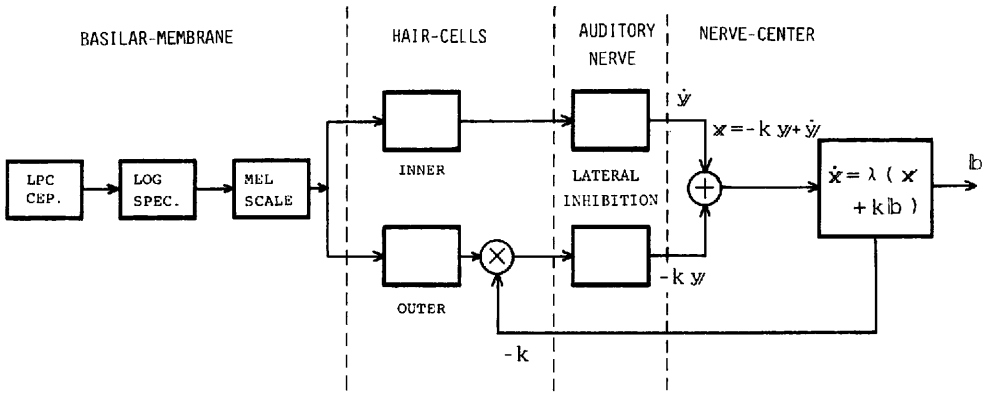


図1　母音ターゲット予測機構のモデル
Fig.1　A peripheral auditory model having vowel target prediction mechanism.

心理実験結果との比較により本モデルの評価を行った結果，本モデルはわたり音区間の減少特性，単音節における母音調音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．

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[9]が提唱した聴覚の工学的機能のうち，

(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

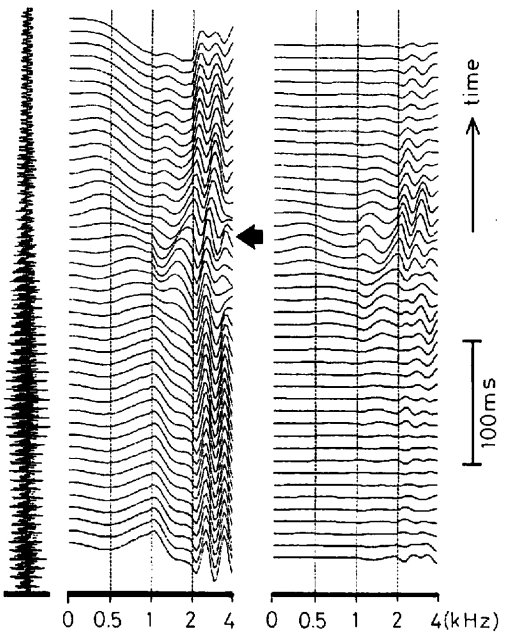
を用いた．

(1), (2)については，8 kHzでサンプリングされた音声波形をフレーム長32 ms，フレーム周期1 ms，次数10次でLPC分析し，LPCケプストラムを求めた後，逆フーリエ変換を行って対数スペクトル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[10]のうち順方向モデルを採用し，

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

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

を離散系で表現することにより実現している．ここで $f(\cdot)$ は側抑制モデルへの入力スペクトル，$G(\cdot)$ は重み関数である．



(a) wave form　(b) spectrum sequence of the outer hair-cells model, y　(c) differential spectrum sequence of the the inner hair-cells model, y

図2　音声 /ai/ に対する聴覚末梢系モデルからの出力スペクトル系列
Fig.2　Output spectrum sequences associated with the peripheral auditory model for a diphthong /ai/.

一方Dallos等[11]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，遠心性線維の両方と結合し，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モデルに組み込み，図1左部分に示すような聴覚末梢系のモデルを構成した．以後本モデル化で使用するスペクトル$\boldsymbol{y}$は，図1の側抑制モデルからの出力とする．

$\dot{\boldsymbol{y}}$は本来スペクトル$\boldsymbol{y}$の時間方向の微分値を表すものであるが，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／が側抑制モデルを通過した場合の$\boldsymbol{y}$，$\dot{\boldsymbol{y}}$を，図2に示す．

3. 予測モデル

3.1 予測モデルの基本原理

予測モデルを作るにあたって，次の点を考慮した．

(1) 人間の聴覚はスペクトル全体の情報を受容する能力があるので，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) スペクトル変化が極大となる時点が前出音と後続音の知覚を分ける重要な時点である[12]．

(3) 変化量と変化速度には知覚的な相補的関係がある[3]．

(4) 音声の物理的特徴量の変化は，臨界制動2次系で近似することができる[13,14]．

(5) 人間の音韻知覚能力を考慮して，予測には短い時間幅の情報だけを用いる[15,17]．

上記(1)～(4)の点を踏まえ，図1の側抑制モデルの出力スペクトル$\boldsymbol{y}$，$\dot{\boldsymbol{y}}$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．

$\boldsymbol{y}_n$を時刻nのスペクトル，$\dot{\boldsymbol{y}}_n$をスペクトル$\boldsymbol{y}_n$の時間方向の微分値，kを外有毛細胞へのフィードバック係数として，$\boldsymbol{y}_n$，$\dot{\boldsymbol{y}}_n$と神経中枢への入力スペクトル$\boldsymbol{x}_n$との間には，

$$\boldsymbol{x}_n = -k\boldsymbol{y}_n + \dot{\boldsymbol{y}}_n \tag{1}$$

の関係が成り立つとする．また，スペクトル$\boldsymbol{x}_n$は，神経中枢においてλを変化時定数の逆数として，微分方程式

$$\dot{\boldsymbol{x}}_n = \lambda(\boldsymbol{x}_n + k\boldsymbol{b}) \tag{2}$$

を満たすとする．ここで$\boldsymbol{b}$は，到達点すなわちターゲット予測値を示す．これらの式およびパラメータの関係を図1右部分に示す．

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$\boldsymbol{b}$の値が異なり，範疇判断[18]を示す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代わる．

(b) 変化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．上記の結果は，考慮した点(1)～(3)を満足する．

式(1)において$k=\lambda$とおき，式(2)に代入すれば，

$$\ddot{\boldsymbol{y}}_n - 2\lambda\dot{\boldsymbol{y}}_n + \lambda^2\boldsymbol{y}_n = \lambda^2\boldsymbol{b} \tag{3}$$

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，

$$\boldsymbol{y}_n = \boldsymbol{a}(1-\lambda n)\exp(\lambda n) + \boldsymbol{b} \qquad n \geq 0 \tag{4}$$

となる．上記の結果は，考慮した点(4)を満足する．

音韻知覚区間長に関する知見[15]によれば，知覚時間区間は50～70 msと言われており，筆者等の検討[17]においても母音知覚時間区間は約50 msという結果が得られている．もし，50 msで$\boldsymbol{b}$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$\boldsymbol{b}$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

前節では生理学的知見，心理学的知見を考慮した結果として臨界制動2次系を導いたが，本節では生理学的および心理学的知見をはなれて，臨界制動2次系の微分方程式の性質からパラメータ推定問題を考察する．

一般に臨界制動2次系を記述する微分方程式は式(3)にも示したように，

$$\left(\frac{\partial^2}{\partial t^2} - 2\lambda\frac{\partial}{\partial t} + \lambda^2\right)\boldsymbol{y} = \lambda^2\boldsymbol{b} \tag{5}$$

と書ける．この解は，

$$\boldsymbol{y} = \boldsymbol{a}(1-\lambda t)\exp(\lambda t) + \boldsymbol{b} \tag{6}$$

となる．

式(5)を変形すれば，

$$\left(\frac{\partial}{\partial t} - \lambda\right)\left\{\left(\frac{\partial}{\partial t} - \lambda\right)\boldsymbol{y}\right\} = \lambda^2\boldsymbol{b} \tag{7}$$

となる．ここで，

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

とすれば，式(7)に代入して，

$$\left(\frac{\partial}{\partial t} - \lambda\right)\boldsymbol{x} = \lambda^2\boldsymbol{b} \tag{9}$$

となる．また，式(6)を式(8)に代入すれば

$$\boldsymbol{x} = -\boldsymbol{a}\lambda\exp(\lambda t) - \lambda\boldsymbol{b} \tag{10}$$

となり，式(10)は式(9)を満たす．

ここで式(9)は1次系の式であるから，式(8)に示す$\boldsymbol{y}$

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 をある値に固定し，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)，(10)を考慮して1次系の式

$$\dot{\hat{x}} = -\lambda \hat{a} \exp(\lambda n) - \lambda \hat{b} \tag{11}$$

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 a，b に関して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 $e(\lambda)$ が求まる．そこで，この誤差 $e(\lambda)$ 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，λ を推定する．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う．

誤差 $e(\lambda)$ が最小となったとき，式(6)，(10)，(11)から，最適化された λ は臨界制動2次系の時定数の逆数，ここで求まる $\hat{b}$ はターゲット予測値 b 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，

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

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，と言える．

臨界制動2次系のパラメータ推定に関する従来の方法[13],[14]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，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，本論文における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 さえ求められればよい．本方式は上で示したように，予測値 b を求めるときに指数関数のパラメータ推定を行っているので，遷移開始時点を特定する必要がない．このため，遷移開始時点を含む長い区間をパラメータ推定に用いる必要がなくなり，短時間区間でのパラメータ推定が可能となる．

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次系関数を表す．また，一点鎖線は $-x/\lambda$，太い実線は b を表す．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．

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，図2に示した二連母音 /ai/ の分析結果であり，図2(b)と図4(c)の矢印の部分を比較すれば，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．

臨界制動2次系の短時間区間でのパラメータ推定は，微分方程式を式(8)，(9)のように分解することにより良好な結果が得られることがシミュレーション実験からも明らかとなった．ところが，**3.2** でパラメータ推定のために導いた式(8)，(9)と **3.1** で生理学的・心理学的知見を導入するために用いた式(1)，(2)を比較すれば，これらは同じ形となっている．このことは，工学的に短時間区間でのパラメータ推定を行うことと，予測モデルを構成するために考慮した点(1)～(4)を実現することが式の上で全く矛盾なく結合されるということを表

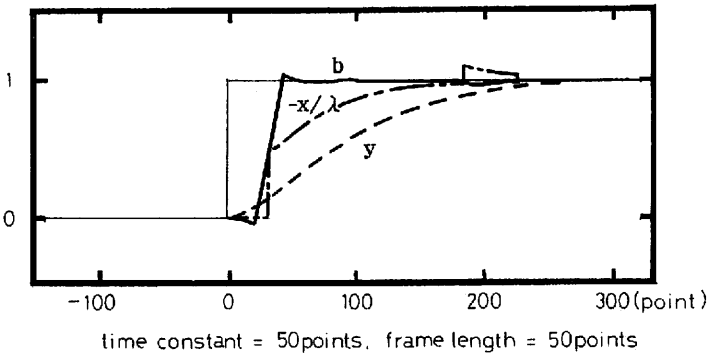


図3　臨界制動2次系モデルによる予測シミュレーション結果

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

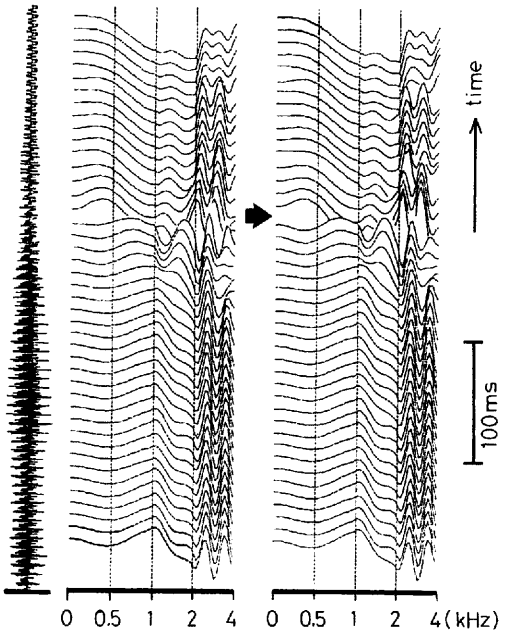


図4　音声/ai/に臨界制動2次系モデルを適用したときの予測出力スペクトル系列
Fig.4　Output spectrum sequences associated with the 2nd-order critical damping model for a diphthong /ai/.

している．さらに言えば，このことは，図1で示したモデルが短時間区間でのターゲット予測値 b の推定に有効なモデルとなっていることを表している．

4. 聴覚心理実験

人間が音声を聞く場合，どの時点まで前出音を聞いていて，どの時点から後続音を聞き始めているのかを知るために，次のような種々の刺激を被験者（通話試験員）に呈示し，明瞭度試験を行った．

4.1 刺激音

刺激音として次の2種類を用意した．

◆Data 1 ：女性アナウンサー2名が発声した100音節を10msごとに話尾切断したデータ．なお，切断部分には10msの線形な傾斜を適用している．

◆Data 2 ：表1に示す対称形三連母音を含む単語（20種類）を男性3人，女性1人に発声してもらい原データとした．刺激音は前後の音韻の影響を取り除くために，フレーム長60ms（立上り20ms，立下り20ms）の台形窓を用いて原データから連続して40個切り出して作成した．なお，窓の移動周期は10ms，刺激音総数は発声者1人について40個×20単語=800個である．

表1　実験に用いた対称型3連母音

/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/

4.2 実験方法

◆Data 1 ：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て4～5回呈示．

◆Data 2 ：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．

受聴は両データとも聞き易いレベルとし，雑音は加えていない．試験員に対しては，聞こえた音声をそのまま音韻記号で記述してもらうこととした．得られたデータから，Data 1については後続音の明瞭度，Data 2については後続音，前出音双方の明瞭度を求め，文献[12]と同様に明瞭度が80～90%を切る時点を聞こえの限界点として，前出音が聞こえなくなる限界点および後続音が聞こ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，Data 1については切断部の最後部，Data 2の前出音については切り出し窓の最前部，後続音につ

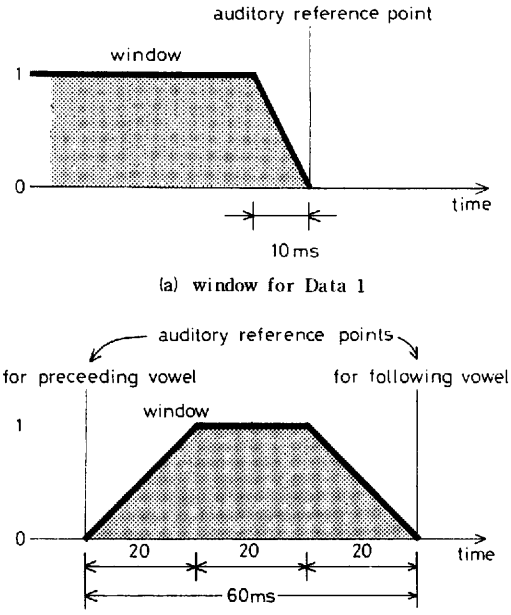


図5 聴覚心理実験用切り出し窓と聴覚的基準点
Fig.5 Truncation windows used for the listening experiments and auditory reference points used for evaluation of the proposed model.

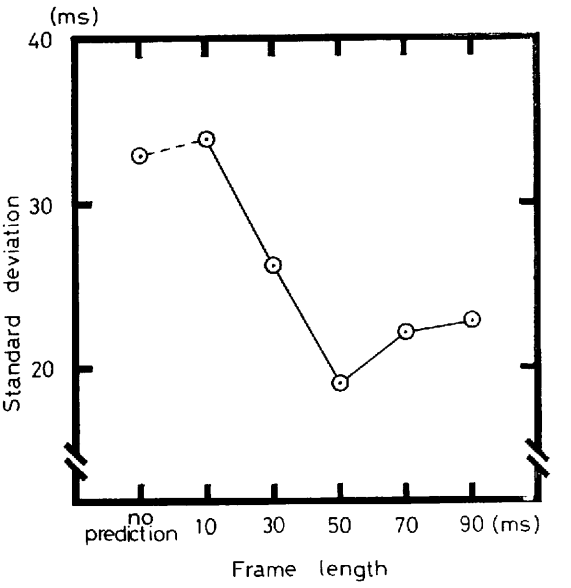


図6 フレーム長を変化させた場合の聴覚的基準点と物理的評価点との差の標準偏差
Fig.6 Standard deviation of the difference between the auditory reference point and its physically estimated point obtained based on the proposed model under five conditions of the frame length.

いては窓の最後部とする．図5にそれぞれのデータについて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，Data 2の前出音について16.0 ms，後続音について16.3 msである．

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ターゲット予測値 b と発声者ごとに求めた5母音のカテゴリ中心とのユークリッド距離を測り，目的の音が他の4母音との距離より初めて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した．この時点を物理的評価点（physically estimated point）と呼ぶことにする．本モデルの評価には，評価点と基準点との時間差の平均値と標準偏差を用いる．ここで，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こととする．

なお，本モデルは上述のようにスペクトルの変化を臨界制動2次系で近似しているが，比較のために，1次系で近似する場合[18]についても実験を行った．

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，Data 1中の発声者1人の100音節データを用いて，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6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50 msのときの標準偏差が最も小さい．この結果から，実験には50 msの区間長を用いることにする．なお，3.で示した図4(b)，(c)は，区間長を50 msとして計算したものである．

区間長50 msは，音韻知覚区間長に関する知見[15],[17]と良く対応している．このことは，本モデルが心理的な聞こ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

図7にData 2中の単語 /kiai/ を分析し，5母音のカテゴリ中心とユークリッド距離の系列を計算した結果および距離の最も小さいカテゴリの時間変化パター

ンを示す．図7(a)は本モデルによる予測値 $\boldsymbol{b}$ との距離系列および最近接カテゴリの時間変化パターン，(b)はスペクトル $\boldsymbol{y}$ との距離系列および最近接カテゴリの時間変化パターン，(c)は1次系モデルによる予測値 $\boldsymbol{b}$ との距離系列および最近接カテゴリの時間変化パターンである．また図中実線 A が後続音／a/に対する評価点であり，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．ここで，図(a)と(c)を比べれば，(c)すなわち1次系モデルを用いた方は，音韻が入れ換わる時点で距離系列が大きく変動しており，予測が不可能となる区間もある．しかしながら，(a)すなわち本方式の方が距離は距離系列が切れめなく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ためである．

1次系モデルを用いた方式では，微分方程式

$$\dot{\boldsymbol{y}}_n = \lambda(\boldsymbol{y}_n + \boldsymbol{b})$$

を解くことになるが，λ が0に近づく時点，すなわちスペクトル $\boldsymbol{y}$ 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7(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 $-\boldsymbol{x}/\lambda$ に関する図でも明らかなように，式(8)によってスペクトルの変化を強調[8],[9]した上で微分方程式(9)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7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10～20 ms の区間でしか知覚されていなかった．一方，スペクトル $\boldsymbol{y}$ と5母音のカテゴリ中心との距離系列を示した図7(b)では，わたり音[e]が広い範囲（60～70 ms）に存在する．ところが本モデルによる結果の図7(a)では，わたり音[e]はほとんど存在しない．

Data 2 全データ中のわたり音が認められる区間についてわたり音長の平均 μ とその標準偏差 σ を求め，わたり音の減少傾向を見ると，図8のようになった．図中の μ の値を見ると，予測を用いないものは62 ms，1次系モデルによる予測値を用いたものは28 ms，また本モデルによる予測値を用いたものは27 ms であり，明らかにわたり音区間の減少傾向が認められる．

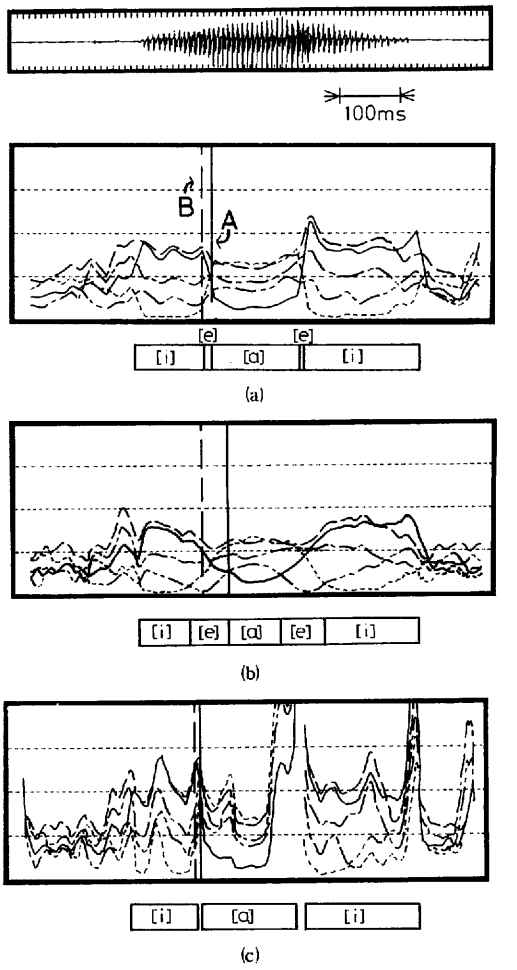


図7 単語/kiai/を発声したときの，(a) 臨界制動2次系モデルによる予測値 $\boldsymbol{b}$，(b) スペクトル $\boldsymbol{y}$，および(c) 1次系モデルによる予測値 $\boldsymbol{b}$ と5母音中心との距離の変化．音韻記号は最も距離の小さい母音を示す．

Fig.7 Distance sequences between each vowel category center and (a) estimated value $\boldsymbol{b}$ based on the critical damping model, (b) spectrum $\boldsymbol{y}$, and (c) estimated value $\boldsymbol{b}$ based on the 1st-order differential equation model for a word utterance /kiai/. Phonetic symbols indicate the nearest vowel category center.

これらの結果は，本モデルがわたり音区間を減少させ，「音声の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」という人間の心理的な聞こえに良く対応できるモデルであることを示している．

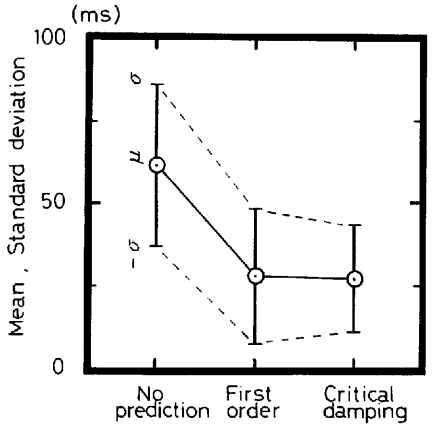


図8 わたり音区間長の平均 (μ) およびその標準偏差 (σ)

Fig.8 Mean (μ) and standard deviation (σ) of the length of the spurious vowel intervals.

5.5 基準点と評価点の差の平均および標準偏差の減少

Data 1 およびData 2を用いて，基準点と評価点の差の平均μと標準偏差σにより本モデルの評価を行った．図9に結果を示す．図(a)はData 2を用いた結果，図(b)はData 1を用いた結果である．図から明らかなように，σの値が，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．たとえば，予測を用いないものと本モデルとの結果を比較すれば，図(a)の後続母音については36 msから20 ms，図(b)の単音節全体については35 msから20 msへと，標準偏差がいずれも2/3～1/2と本モデルの方が小さくなっている．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．

図(b)では，拗音/j/を含む音節で基準点と評価点の差の平均μの値がだんだんと小さくなっている．たとえば，予測を用いないものでは42 msであるが，本モデルを用いれば15 msとなっている．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．ところが本モデルを用いることによってより速くターゲットに近づけることができるようになり，μ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，特に拗音を含む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，「音声の物理的特徴が後続母音に到達していないにもかかわらず，人間はその母音が発声されたかのように知覚する」という人間の聴覚特性を良く表すモデルであることを示している．

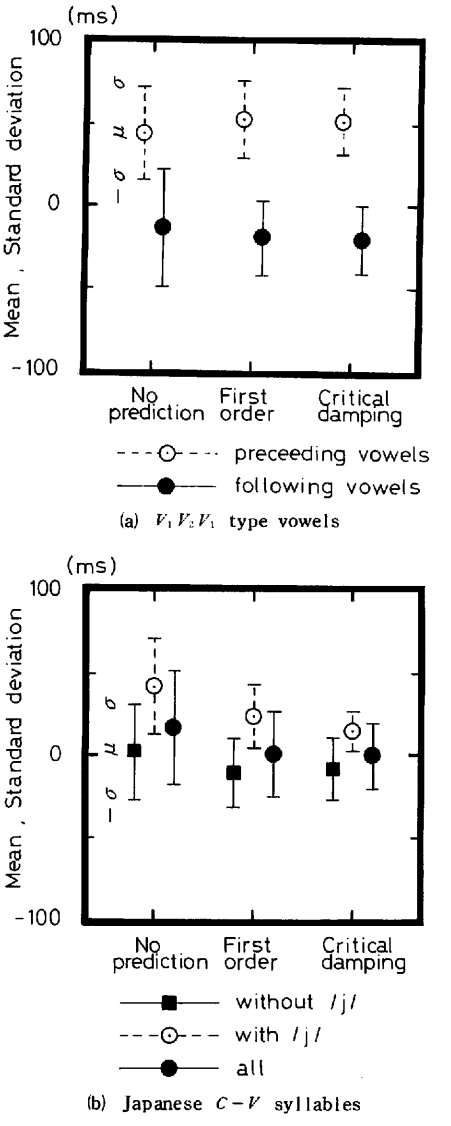


図9 聴覚的基準点と物理的評価点との差の平均 (μ) およびその標準偏差 (σ)

Fig.9 Mean (μ) and standard deviation (σ) of the difference between the auditory reference point and its physically estimated point.

6. むすび

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．本論文では，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．

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視聴覚機構研究グループ・筧グループリーダをはじめ，視聴覚機構研究グループおよび第四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．

文　献

(1) P. T. Brady et. al.：“Perception of Sounds Characterized by a Rapidly Changing Resonant Frequency”, J. of Acoust. Soc. Am., 33, 10, pp.1357-1362 (1961).
(2) B. E. F. Lindblom：“On the Role of Formant Transition in Vowel Recognition”, J. of Acoust. Soc. Am., 42, 4, pp.830-843 (1967).
(3) 鈴木久喜：“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).
(4) 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).
(5) 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).
(6) 桑原尚夫：“連続音声中の母音知覚と特徴抽出の一試み”，信学論(A)，J61-A, 3, pp.247-253 (昭53-03).
(7) 桑原尚夫：“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).
(8) 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).
(9) D. H. Klatt：“Speech Processing Strategies based on Auditory Models”, The presentation of speech in the peripheral auditory system, R. Carlson and Granstrom eds. pp.181-196 Elsevier Biomedical press (1982).
(10) 塚原仲晃：“脳の情報処理”，朝倉書店 (1984).
(11) P. Dallos et al.：“Cochlear Inner and Outer Hair Cells：“Functional Differences”, SCIENCE, 177, pp.356-359 (1972).
(12) 古井貞熙：“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).
(13) 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).
(14) 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).
(15) 難波編：“聴覚ハンドブック”，第7章，ナカニシヤ出版 (1984).
(16) 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).
(17) 古井貞熙：“音韻知覚を決定づける音声区間の特性について”，60年度音響学会秋季講論 1-3-15 (1985).
(18) 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”，60年度音響学会春季講論 3-5-16 (1985).
(19) 古井貞熙：“スペクトル変化強調による単語音声認識”，音響学会音声研資，S85-77 (1986).

（昭和61年2月26日受付，6月2日再受付）

赤木　正人

昭54名工大・工・電子卒．昭59東工大大学院博士課程終了．同年日本電信電話公社（現ＮＴＴ）武蔵野電気通信研究所入社．以来，人間の音声知覚機構に関する研究に従事．現在，ＮＴＴ基礎研究所・情報通信基礎研究部・視聴覚機構研究グループ・研究主任．工博．ＩＥＥＥ，日本音響学会，日本認知科学会各会員．

古井　貞熙

昭43東大・工・計数工学科卒，昭45同大学院修士課程修了．同年日本電信電話公社（現ＮＴＴ）電気通信研究所入社．以後音声認識，話者認識，音声知覚などの基礎研究に従事．昭53～54米国ベル研究所客員研究員．現在ＮＴＴ基礎研究所・情報通信基礎研究部・第四研究室・室長．工博．昭49年度本会米沢賞受賞．著書「ディジタル音声処理」（東海大学出版会）．ＩＥＥＥ，日本音響学会会員．