論文 / 著書情報
Article / Book Information

論題(和文)	音声知覚における短時間区間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)	日本音響学会1985年秋季講演論文集, , , pp. 253-254
Citation(English)	, , , pp. 253-254
発行日 / Pub. date	1985, 9

1—3—16

音声知覚における短時間区間母音ターゲット予測機構のモデル化*

◉赤木　正人　　古井　貞熙　（NTT武蔵野電気通信研究所）

1．まえがき

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素の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。この現象について、筆者等は先に、聴覚系にターゲット予測機構が存在すると仮定し、心理学的、生理学的知見をもとに、この予測機構を1次の微分方程式で記述した[1]。本報告では、有毛細胞に関する生理学的知見を新たに取り入れ、音声の物理的特徴の変化を臨界制動2次系モデルで記述し、ターゲットの予測値を短時間区間で推定できる、母音ターゲットの予測機構モデルを構成する。また、刺激音として対称形三連母音を含む単語を用いた聴覚実験結果と、このデータを本手法で解析した結果の比較も合わせて報告する。

2．予測機構モデル

Dallos等[2]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。そこで新たに、聴覚末梢系のモデルとして、Fig.1に示すモデルを構成する。本モデルでは、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。また、周波数軸上での重み付けを行い、側抑制による Sharpening をモデル化している。ここで、

$$\dot{x} = \lambda x \qquad (1)$$

$$x = -k y + \dot{y} \qquad (2)$$

とすれば、式(1)は、(a)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、(b)変化量と変化速度には相補的関係がある、の2点を考慮した式となっている。また式(2)は、内外有毛細胞の機能差を考慮した式となっている。

式(2)において$k=\lambda$とおき、式(1)に代入すれば、

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

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、

$$y = a(1-\lambda t)\cdot\exp(\lambda t) + b \qquad (4)$$

となる。ここで、bがターゲットの予測値を表わす。

3．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、

$$x = (d/dt - \lambda)\cdot y \qquad (5)$$

$$(d/dt - \lambda)\cdot x = 0 \qquad (6)$$

と展開できる。仮にλが求められているとすれば、式(5)からxが計算され、次に式(6)を考慮して、1次系の式 $\hat{x} = \lambda\hat{a}\cdot\exp(\lambda t) + \lambda b$ をxに当てはめることにより、式(4)中のパラメータbが決定される。λの決定は、$\Sigma\,|x-\hat{x}|^2$ 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。本方式は、臨界制動2次系のパラメータλ、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。

音韻知覚区間長に関する知見[3]によれば、知覚時間区間長は50～70msと言われており、筆者等の検討[4]においても同様の結果が得られている。従来の方法[5][6]は、式(4)から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り、本モデルに適用しようとしても、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。ところが本方式は、最も重要なパラメータである予測値bを推定する方式であるため、遷移開始時点を特定する必要がない。さらに、短時間区間でパラメータ推定が可能となり、特に区間長50msで、後に述べる聴覚実験との対応が最も良くなる。ただし式(4)中のaを求める場合は、遷移開始時点を特定する必要がある。Fig.2に、本手法を単語「気合い」/kiai/に適用した結果を示す。図は、5母音の各母音中心と、本手法を用いた予測

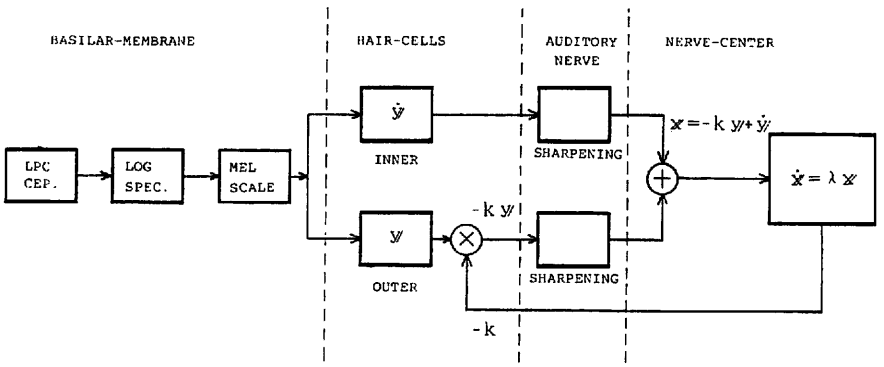


Fig.1-A model of vowel target prediction system.

*Modeling of short-term vowel target prediction mechanism in speech perception
Masato AKAGI and Sadaoki FURUI (Musashino E.C.L., NTT)

値$\hat{b}$との距離(a)、スペクトルyとの距離(b)、1次系（詳細は文献[1]）で求めた予測値$\hat{b}$との距離(c)の時間変化を示してある。(a)と(b)を比較すれば、本方式ではわたり音/e/の出現する区間が短くなっている。全データ中のわたり音が認められるデータに対して、わたり音長とその標準偏差を求め、わたり音の減少傾向を見ると、予測を用いないもの、1次系、臨界制動2次系それぞれ、平均が61．6ms、27．9ms、27．2ms、標準偏差が24．6ms、20．3ms、16．3msであった。これらの結果は、本方式がわたり音区間を減少させ、心理的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。また、(a)と(c)を比較すれば、本方式の方が距離が安定に計算されている。

4．モデルと聴覚実験結果の対応

音声データ：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、刺激音総数は話者1人について800個である。

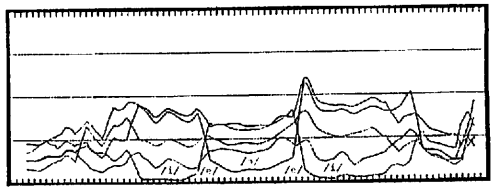
実験：通話試験員（女性4名）に、発声者1人について4回、1回800個の刺激音を聞き易いレベルでランダムに呈示し、聞えた音声をそのまま記述してもらうこととした。呈示はISI1．5秒で行った。聴覚実験結果から、前出音が聞えなくなる限界点及び後続音が聞え始める限界点を求めた。この時点を基準点と呼ぶ。なお、基準点は切り出しフレームの中心を基準としている。実験から求めた基準点の標準偏差は、前出音について16．0ms、後続音について16．3msであった。

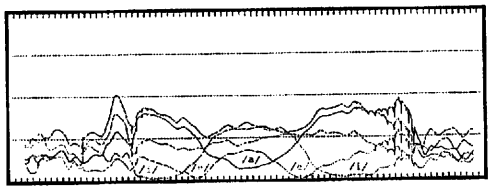
評価：モデルから得られた予測値$\hat{b}$と話者毎に求めた5母音のカテゴリ中心との距離を測り（Fig.2参照）、目的の音が他の4母音との距離より小さくなる時点を、本モデルにおける前出音、後続音に対する限界点とした。この時点を評価点と呼ぶ。本モデルの評価は、評価点と基準点の差の平均と標準偏差で行う。なお差が正の場合は、評価点の方が基準点より時間的に後方にあることを示す。結果をTable 1(a)に示す。また、文献[1]で用いた単音節データに対して本手法を適用した結果を(b)に示す。表から明らかなように、予測を用いないもの、1次系を用いたもの、本手法の順で標準偏差が小さくなっており、本手法がより聴覚特性に近づいたことが分かる。また(b)では、差の平均の値がだんだんと小さくなっている（特に/j/を含む音節）。この結果は、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。

5．あとがき

聴覚末梢系モデルに有毛細胞に関する生理学的知見を取り入れ、臨界制動2次系モデルと結合させた。また、予測値を表わすパラメータを短時間区間で推定できる方法を示した。今後は、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：日頃指導いただく筧第四研究室長をはじめ、研究室の諸氏、諸嬢に感謝いたします。





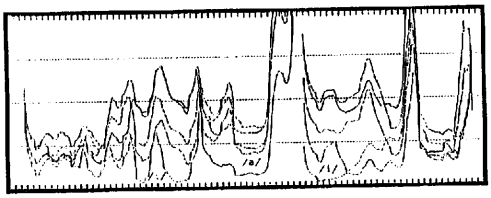


Fig.2-Distance sequences between each vowel category center and estimated value $\hat{b}$ using the critical damping model (a), spectrum y (b), and estimated value $\hat{b}$ using the first-order differential equation model (c) for a word utterance /kiai/.

Table 1-Mean (upper values) and standard deviation (lower values) of the difference between the auditory critical point and its estimation.

(a) V1V2V1 type vowels

PREDICTION	NO	FIRST-ORDER	CRITICAL-DAMPING
PRECEDING VOWEL	14.2 ms	22.8	22.0
	28.17 ms	23.76	20.39
FOLLOWING VOWEL	16.7	11.2	9.8
	25.98	22.26	20.53

(b) Japanese C-V syllables

PREDICTION	NO	FIRST-ORDER	CRITICAL-DAMPING
SYLLABLES WITHOUT /j/	3.93 ms	-10.20	-7.37
	28.95 ms	20.34	18.77
SYLLABLES WITH /j/	42.02	23.98	19.19
	29.14	18.64	15.20
ALL	17.05	1.25	-0.08
	34.74	25.78	20.05

文献

(1) 赤木、古井、音学春季講論3-5-16(昭60)、(2)Dallos et.al., SCIENCE, Vol.177, pp..356-359(1972)、(3)難波他、"聴覚ハンドブック"、ナカニシヤ出版(1984)、(4)古井、"音韻知覚を決定づける音声区間の特性について"、音学秋季講論(昭60)、(5)藤崎他、音学音声研資S73-03(1973)、(6)板橋、横山、電総研彙報、40、6、pp.,530-541(1976)