T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

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)	日本音響学会1985年春季講演論文集, , , pp. 215-216
Citation(English)	, , , pp. 215-216
発行日 / Pub. date	1985, 3

Powered by T2R2 (Science Tokyo Research Repository)

3—5—16　音声知覚における母音ターゲット予測機構のモデル化*

○赤木　正人　　古井　貞熙　(日本電信電話公社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、母音定常部に音声の特徴が到達する前に人間は到達音を予測し、あたかもその音が発声されたかのように知覚する。たとえば、調音結合部分におけるなまりの回復、無声化した母音の回復などがこの現象を良く説明している。また、筆者等の他の研究[1]によれば、スペクトル変化極大点を超えた時点で音節が知覚されていることが明らかとなっている。このような予測機構の存在については、すでに鈴木[2]、桑原[3]、藤崎他[4]なども指摘している。本研究では、この予測機構を聴覚の生理学的、心理学的知見をふまえてモデル化し、聴覚実験との対比からこのモデルを評価する。

2. 聴覚実験

(a)音声データ

DATA1:男性話者1名が発声した二連母音を8msごとに話尾切断したデータ。

DATA2: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。

両データとも切断には幅8～10msの線形な傾斜を用いた。

(b)実験条件

DATA1:男性被験者12名に対してランダムに提示。

DATA2:女性試験員4名に対して4～5回ランダムに提示。

受聴は両データとも聞きやすいレベルとし、雑音は加えていない。被験者およ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後続母音の明瞭度を求め、明瞭度が80%を超える切断位置を後続母音が聞え始める時点(基準点)とした。

3. 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[5]が提唱した聴覚の工学的モデルのうち、(1)基底膜をモデル化するためのメル尺度、(2)有毛細胞をモデル化するための半波整流器、(3)時間軸上での微分作用と自己相関、(4)ラウドネス尺度近似のための対数変換、(5)側抑制のモデル化のための周波数領域での重み付け、を用いた。そして、Fig.1(a)に示す末梢系の各ブロックに対応させて、同図(b)に示すような工学的モデルを作成した。モデルからの出力をFig.2に示す。図は二連母音/ai/の分析結果である。

4. 予測モデル

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[1]。(3)変化量と変化速度には相補的関係がある[2]。(4)予測には短い時間幅の情報だけを用いる。上記の点をふまえ、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，λを変化時定数の逆数とすれば

$$\dot{y}_n = \lambda y_n \qquad \text{—— (1)}$$

の関係が成り立つとする。

上式は、(1) λ>0、λ<0 で到達方向と到達点が異なり、範疇判断が行われるモデルとなっている、(2)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。

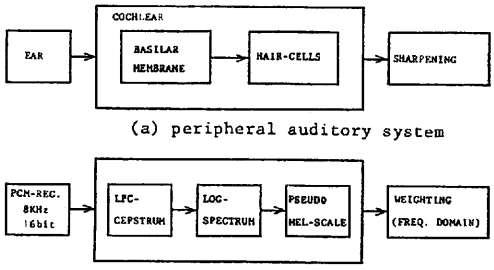
式(1)を解けば、

$$y = a \times \exp(\lambda t) + b \qquad \text{—— (2)}$$

となる。ここでbは到達点を表わす。λ、a、bの計算には短区間での最小自乗法を用いる。

5. モデルの評価基準

bと5母音(a,i,u,e,o)のカテゴリ中心(単音節から抽出)との距離をはかり、後続母音に対応するカ



(a) peripheral auditory system

(b) model

Fig.1 A model of the peripheral auditory system.

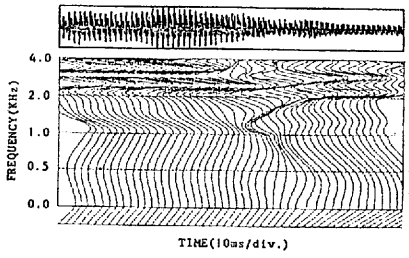


Fig.2 Output spectrum sequences of the peripheral auditory model for the connected vowel utterance /ai/.

* Modeling of Vowel Target Prediction Mechanism in Speech Perception. by Masato AKAGI and Sadaoki FURUI (Musashino E.C.L., N.T.T.)

テゴリの中心との距離が、他の母音のカテゴリ中心との距離に比べて小さくなる時点を、このモデルにおける聞え始めの時点(評価点)とする。Fig.3(a)に/ai/を本モデルにより分析した結果を示す。図中Aが評価点である。また、Bが明瞭度の基準点である。本モデルの評価には、評価点と基準点との差、すなわちA－Bとその標準偏差を用いる。

6. 到達点算出のための区間長

到達点bを算出するための最適短区間長を決定するために、DATA1を用いて予備実験を行った。結果をFig.4に示す。図中のμは、評価点Aと基準点Bの差の平均、σはその標準偏差である。また、μが負の値の時は評価点が基準点より前にあることを示す。図から分かるように、区間長50～70msの時σが小さい。この結果から、実験には50msの区間長を用いることとする。区間長50msは、経時マスキングの影響がほとんどなくなる時間、また知覚時間仮説(6)の時間窓と良く対応している。

7. モデルの評価

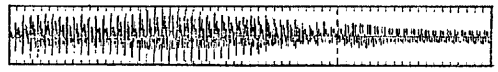
(a)わたり音の減少：DATA1による聴覚実験によれば、/ai/におけるわたり音/e/はほとんど知覚されていない。一方、二連母音/ai/におけるスペクトルyと5母音のカテゴリ中心との距離の系列を示したFig.3(b)では、/a/から/i/に変化する時点で、わたり音/e/が広い範囲に存在する。しかし、本モデルによる結果Fig.3(a)では、わたり音はほとんど存在しない。

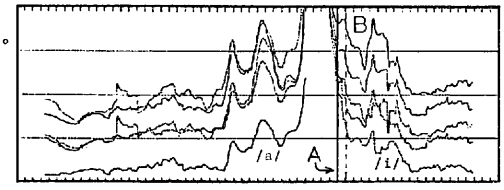
(b)標準偏差の減少：DATA1およびDATA2を用いて、差の平均と標準偏差により本モデルの評価を行った。Table1に結果を示す。表中μ、σ及びその符号はFig.4と同じ意味である。表中(a)(b)のσの値から分かるように、予測機構を付加することにより標準偏差が減少し聴覚実験結果に近づくことが明らかとなった。表中(b)(c)に示すσの値は、話尾切断の間隔がDATA1で8ms、DATA2で10msであることから、許容できる範囲である。

これらの結果から、本モデルは、わたり音を減少あるいは消滅させる働きを有し、聴覚実験結果とも良く対応するモデルである、と言うことができる。またこの結果から、本モデルは、連続音声中のなまけの回復、セグメンテーションなどの機械認識の方式としても有効であると考えられる。

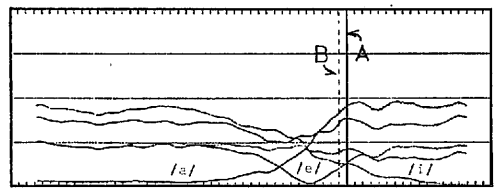
8. あとがき

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。本研究では、この予測機構を生理学・心理学的知見をもとにした工学モデル上に構成し、聴覚実験から得られた結果との対比によりモデルの評価を行った。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、データとして子音なども含めた総合的なモデルを構成するとともに、他の聴覚心理実験による検証を加えて行きたい。





(a) DISTANCE (ESTIMATED VALUE b)



(b) DISTANCE (SPECTRUM y)

Fig.3 Distance sequences between vowel category center and estimated value b (a), or spectrum y (b) for connected vowel utterance /ai/.

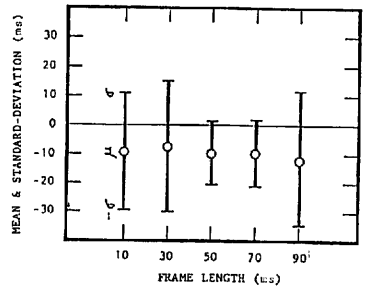


Fig.4 Mean(μ) and standard deviation(σ) of the difference between the auditory critical point and its estimation by the proposed model as a function of frame length.

Table 1 Mean(μ) and standard deviation(σ) of the difference between the auditory critical point and its estimation.

	(a)	(b)	(c)
DATA	DATA1	DATA1	DATA2
PREDICTION	NO	YES	YES
μ (ms)	-7.11	-9.37	-0.38
σ (ms)	21.34	11.15	18.04

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に感謝いたします。

[文献](1)古井、聴覚研資H85-6、(2)鈴木、音響学会誌30-3、(3)桑原、音響学会誌31-1、(4)藤崎他、音声研資S81-35、(5)Klatt, Speech Communication Group Working Paper II (1983) (6)難波他、"聴覚ハンドブック"、ナカニシヤ出版(1984)