論　文

文脈効果のモデル化とそれを用いたワードスポッティング

米沢　裕司†　　赤木　正人†

Modeling of Contextual Effects and Its Application to Word Spotting

Yuji YONEZAWA† and Masato AKAGI†

あらまし　人間は聴覚の補正機構の働きにより，調音結合のためになまけた音韻であっても正しく認識することができる．本論文では，聴覚の補正現象の一つとして考えられる文脈効果をモデル化し，モデルをワードスポッティングの前処理として用いる．モデルは近接したスペクトルピークの影響で人間の知覚するスペクトルが本来のスペクトルに比べ変化することを記述したものであり，報告されている心理実験の結果をもとに，最小分類誤り学習によりモデル化を行った．次に，モデルの能力を検証するため，モデル適用によるホルマント軌跡の変化を調べた．その結果，モデルには調音結合のため連続音声中に現れる「なまけ」を回復する働きがあることがわかった．また，モデルを前処理として用いた母音認識実験を行い，その結果，モデルを用いることにより有意な認識率の向上が見られた．更に，モデルを前処理として用いたワードスポッティング実験を行った．その結果，単語検出率が1-6ポイント程度向上し，モデルが前処理として有効に機能することを示した．

キーワード　文脈効果，調音結合，ワードスポッティング，なまけ

1. まえがき

連続発話音声中には，調音器官の制約のため，音声の物理的特徴が単独発声されたときの物理的特徴に到達する前に次の音韻に移る「なまけ」や，音韻から音韻への過渡状態である「わたり」といった不完全な音が現れる．機械において音声認識を行う際に，これら不完全な音が誤認識の原因の一つとなっている．

一方，人間はこれら不完全な音が含まれた音声でも正しく知覚することができる．このことから，人間は音声を知覚する際にこれら不完全な音をある種の補正機構により補正した後に知覚すると考えられる．そこで，人間のもつ補正機構を模擬した機構を音声認識に取り入れられれば，認識能力を向上させることができると考えられる．

補正現象としては，文脈効果が知られている[1]．文脈効果は，ある刺激（対象刺激）の知覚が時間的に近接している刺激（文脈刺激）の影響を受ける現象である．このうち，対象刺激を文脈刺激と同じように知覚する現象を同化効果と呼び，逆に，対象刺激を文脈刺激と異なるように知覚する現象を対比効果と呼んでいる（図1）．そして，同化効果は同一音韻を知覚する際，知覚パターンの変動を軽減することに寄与し，対比効果は異なる音韻を連続して知覚する際，その差異を強調し，音韻間の知覚パターンの分離に寄与すると考えられる．そこで，この文脈効果をモデル化すれば，

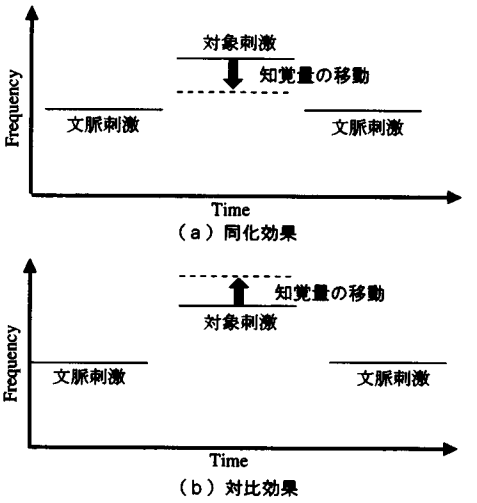

図1　文脈効果，(a) 同化効果 (b) 対比効果
Fig. 1　Contextual effect, (a) assimilation effect (b) contrast effect.

† 北陸先端科学技術大学院大学，石川県
Japan Advanced Institute of Science and Technology, Hokuriku, 1-1 Asahidai, Tatsunokuchi, Nomi, Ishikawa-ken, 923-12 Japan

モデルによりなまけやわたりを補正すること，更には連続音声認識の能力を向上させることが期待できる．

そのため，さまざまな心理学的な研究[2]～[4]のほか，工学的なモデル化に関する研究が行われている．例えば，藤崎ら[5]は2連母音刺激を用いた心理実験の結果から，非定常音の知覚の定量的なモデルを提案している．また桑原[6]は，隣接した母音の影響を考慮することで感覚的なホルマントの位置を求める手法を提案し，それが母音認識率の向上に有効であることを示している．これらは，文脈効果をモデル化することにより調音結合による影響を解消する手法として興味深いものであるが，いずれもホルマントを特徴量として取り扱ったものであるため応用上有効であるとはいいがたい．

そこで本論文では，スペクトルを対象とした工学的な文脈効果モデルを提案する．モデルではスペクトルを特徴量として取り扱っているため，ホルマントを特徴量として扱っているモデルに比べ応用上の利点がある．またモデルは，心理実験の結果を反映しているほか，モデルを用いた際の認識率が向上するよう最小分類誤り学習[7]を用いて設計されているため，音声認識への応用に有効である．以降では，モデル化の方法について述べると共に認識実験を行い，モデルを前処理として用いることにより調音結合によるなまけを回復させ，ワードスポッティング能力の向上を図る．

2. スペクトルを対象とした音響的文脈効果モデル[8]

2.1 従来のモデル

これまでに，赤木は「音響レベルの文脈効果はスペクトルピーク対の相互作用の和としてモデル化できる」ことを仮定した文脈効果モデルを提案している[1]．このモデルは，スペクトルピーク間の相互作用の結果スペクトルピークの知覚が変化することをモデル化したものであり，対象刺激として5ホルマント音，文脈刺激として単ホルマント音を用いた心理物理実験を行い，その結果をもとにモデル化を行っている．また，モデルを実音声に適用した結果，モデルがなまけの回復に有効であることを示している．しかし，モデルが対象としている物理量がスペクトルピークに限定されているなど，モデルを音声認識の前処理として用いるには必ずしも十分であるとは言えなかった．今回，このモデルをもとに，音声認識へ応用可能な工学的な文脈効果モデルを提案する．

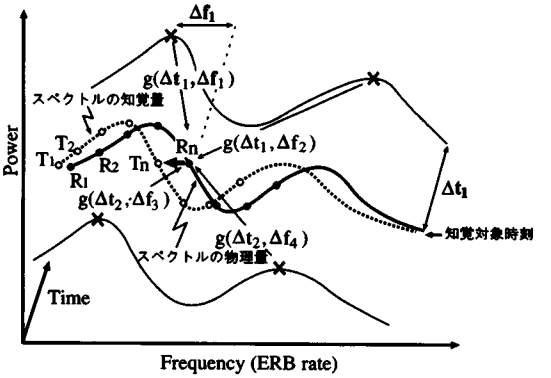

図2 文脈効果モデル
Fig. 2 Contextual effect model.

2.2 モデルの高度化

今回提案するモデルは従来のモデルを拡張したものであり，近接したスペクトルピークの影響で人間の知覚するスペクトルが本来のスペクトルに比べ変化することを記述している．モデルが対象としている物理量がスペクトルであるので，容易に音声認識の前処理として用いることができる．

図2にモデルの概要を示す．ある時刻のスペクトル中の点 R_n（図中●）を知覚する際，近接しているスペクトルピーク（図中×）からそれぞれ $g(\Delta t, \Delta f)$ だけ影響を受ける．ここで，Δf は R_n とスペクトルピークとの周波数差で，スペクトルピークの周波数が R_n に比べ大きいときに正の値をとる．なお，周波数軸としては人間の聴覚特性を考慮した ERB rate $= 24.7\log(4.37f\ [\mathrm{kHz}]+1)$ で表される ERB rate [9] 軸を用いており，Δf の単位は ERB である．また，Δt は R_n とスペクトルピークとの時間差であり，スペクトルピークの時刻が R_n に比べ遅いときに正の値をとる（単位は msec）．つまり，スペクトルピークから受ける影響の大きさは知覚対象時刻のスペクトル中の点と近接したスペクトルピークとの周波数差 Δf および時間差 Δt により一意に定まるとしている．そして，それぞれのスペクトルピークからの影響量 g の総和分，R_n に比べ周波数方向に移動した T_n（図中○）を知覚する．すなわち，

$$T_n = R_n + \sum_{m}^{M}\sum_{n}^{N} g(\Delta t_m, \Delta f_n) \qquad (1)$$

である．但し，M は文脈効果の影響を考慮するフレームの個数，N は第 m フレームのスペクトルにおける

スペクトルピークの個数である．式(1)を用いてスペクトル中の各点(R_1, R_2, ···)に対応する知覚量(T_1, T_2, ···)を求めることにより，本来のスペクトルから人間が知覚するスペクトルを推定することができる．

2.3 知覚的影響量 g の決定

文脈効果モデルは，人間の聴覚がもつ補正機構を模擬したものであるから，聴覚特性をよく反映しているものでなければならない．またそれに加え，モデルの音声認識への応用を考えると，モデルを音声認識に適用することにより認識能力が改善されるようなモデルでなければならない．従って，これまでの心理実験の結果と適合し，かつモデルを用いたときに認識率の改善が見られるような知覚的影響量 g をモデルで使用する必要がある．そこで，報告されている心理実験の結果を参考に知覚的影響量 g を近似式で表し，近似式で用いられているパラメータを最小分類誤り学習[7]により決定する．

2.3.1 心理実験の結果を考慮した知覚的影響量 g の近似形式

以下に述べる心理学的知見などを参考に，知覚的影響量 g はパラメータ A–G を用いて次式のように近似することとした．

$$g(\Delta t, \Delta f) = e^{A|\Delta f|}\sin(B\Delta f) \cdot \left(Ce^{-\frac{\Delta t^2}{D}} - Ee^{-\frac{\Delta t^2}{F}}\right)|\Delta t|^G \quad (2)$$

但し，$|\Delta f| > \frac{\pi}{B}$では $g = 0$

これまでの心理実験の結果[1]では，Δf がある値で知覚的影響量は極大となり，また $\Delta f = 0$ 付近では知覚的影響はほとんど見られなかった．つまり，ある周波数だけ離れたスペクトルピークから最も文脈効果を強く受け，そこから離れるに従い徐々に文脈効果の影響量は減衰していくと考えられる．このことを考慮し，Δf による知覚的影響量の変化をパラメータ A, B と sin 関数を用いて記述している．

また，重野によると，文脈効果には二つの短期記憶が関与しており，一方の短期記憶は同化効果に，もう一方は対比効果に関与している[3]．また，同化効果と対比効果は，それぞれ $\Delta t = 0$ を中心として徐々にその影響量が減衰していくと思われる．これらを考慮して，スペクトルピークからの知覚的影響量 g を同化効果成分と対比効果成分に分け，それぞれの影響量をパラメータ C, D を用いたガウス分布とパラメータ E, F を用いたガウス分布で記述している．但し，$\Delta t = 0$ 付近のスペクトルピークからは文脈効果はほとんど受けないものとして，$|\Delta t|^G$ をそれぞれの成分に掛けている．

表1 パラメータ A–G の意味
Table 1 Meaning of parameter A–G.

パラメータ	意味
A	Δf に対する文脈効果の範囲を示す
B	文脈効果が最大となる Δf の値を示す
C	同化効果の大きさを示す
D	Δt に対する同化効果の範囲を示す
E	対比効果の大きさを示す
F	Δt に対する対比効果の範囲を示す
G	$\Delta t = 0$ 付近での文脈効果の減衰の度合を示す

表2 学習に使用した音声データ
Table 2 Speech data for learning.

ATR音声データベース[10]，男性話者 mht サンプリング周波数 20 kHz	
入力パターン	標準パターン
国際会議予約に関する自由発話文章データ 25 文章（タスクコード scl)	単独発話母音データ（タスクコード sy)
40 次 FFT ケプストラム平滑化スペクトル（フレーム長 30 msec，フレームシフト 10 msec) を使用 周波数軸には ERB rate 軸を使用 スペクトルの平均パワーは 0 に正規化	

パラメータ A–G の意味合いは，表1のように，文脈効果（同化効果と対比効果）が最大となる周波数差 Δf や，対比効果，同化効果それぞれの範囲などを記述している．次項ではこのパラメータ A–G の値を最小分類誤り学習により決定する．

2.3.2 最小分類誤り学習によるパラメータの決定

最小分類誤り学習は分類誤りが最小になるように識別関数のパラメータを決定する学習方法であり，片桐らにより提案されている[7]．本研究では最小分類誤り学習を用いて，「モデルを前処理として用いた場合に入力パターンの母音中心の認識率の向上に最も効果のあるようなパラメータ」を求めた．使用したデータに関する詳細は表2のとおりである．

以下の手順でパラメータ A–G を求めている．式(2)のパラメータ集合を Λ，入力パターンに文脈効果モデルを適用した結果得られたスペクトルを $\boldsymbol{x}$ とする．但し，$\boldsymbol{x}$ は自由発話文章中の母音中心のスペクトル（母音の特徴が顕著なスペクトル）であり，その時間的位置および音韻はデータベース中のラベルに従っている．また，連続発話音声の 16 次の LPC スペクトルのスペクトルピークのうち，24 ERB rate（約 2800 Hz）以下のピークを抽出し，このスペクトルピー

クから知覚的影響を受けるものとして文脈効果モデルを適用し，$\boldsymbol{x}$ を求めている．次に，$\boldsymbol{x}$ とカテゴリー C_j $(j=1,2,\cdots,K$．ここで K は母音の種類の数である 5) の標準パターンとのスペクトル距離により識別関数 $g_j(\boldsymbol{x};\Lambda)$ を定義し，誤分類尺度 d_k を

$$d_k(\boldsymbol{x};\Lambda)=g_k(\boldsymbol{x};\Lambda)-g_i(\boldsymbol{x};\Lambda) \tag{3}$$

により求める．ここで，C_k は入力パターンの属するカテゴリー，$g_i(\boldsymbol{x};\Lambda)$ は $g_k(\boldsymbol{x};\Lambda)$ 以外の識別関数のうち最小の識別関数である．そして，コスト関数 l_k を

$$l_k(\boldsymbol{x};\Lambda)=\frac{1}{1+e^{-\xi d_k(\boldsymbol{x};\Lambda)}} \tag{4}$$

により求める．但し，ξ は正の値である．次に，入力 $\boldsymbol{x}$ に対するコストを

$$l(\boldsymbol{x};\Lambda)=\sum_{k=1}^{K} l_k(\boldsymbol{x};\Lambda)1(\boldsymbol{x}\in C_k) \tag{5}$$

により求める．また，1() は indicator 関数で，

$$1(v)=\begin{cases}1, & \text{if } v \text{ is true}\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

である．コスト $l(\boldsymbol{x};\Lambda)$ は分類結果を正と誤の 2 値ではなく連続値で表しているものであり，こう配探索が可能である．そこで，コスト $l(\boldsymbol{x};\Lambda)$ が小さくなるようにパラメータ Λ を更新する．

以上の手順を繰り返すことにより学習データ全体の分類誤り数が最小になるようなパラメータが求まり，最終的に式 (7) および図 3 に示す知覚的影響量 $g(\Delta t,\Delta f)$ が得られた．

$$\begin{aligned} g(\Delta t,\Delta f) &= e^{-0.031|\Delta f|}\sin(0.27\Delta f)\\ &\quad\cdot\left(0.023e^{-\frac{\Delta t^2}{775}}-0.0036e^{-\frac{\Delta t^2}{5075}}\right)\\ &\quad\cdot|\Delta t|^{1.16} \end{aligned} \tag{7}$$

但し，$|\Delta f|>\dfrac{\pi}{0.27}$ では $g=0$

2.3.3 知覚影響量 $g(\Delta t,\Delta f)$ の特性

得られた知覚影響量 $g(\Delta t,\Delta f)$ を $\Delta f=4$ ERB 一定として図示すると，図 4 となる．g が正であるということは，周波数の高い方向にスペクトルの知覚が移動するということを示し，g が負であるということは，周波数の低い方向に知覚が移動するということを示している．また，この場合 Δf が正であるので，g の値が正であれば同化効果を受けることを示し，負であれば対比効果を受けることを示す．すなわち，学習結果は図 5 に示すように，

- $|\Delta t|<50$ msec のスペクトルピークからは同化効果
- $|\Delta t|>50$ msec のスペクトルピークからは対比効果

を受けることを示している．これは，これまでの心理実験の結果 [1] とほぼ一致する．

また，$\Delta t=30$ msec 一定として図示すると，図 6 となる．g の絶対値は文脈効果の大きさを示すものである．従って，図 6 は文脈効果の大きさが $\Delta f=\pm5$ ERB で極大になることを示している．これも，これまでの心理実験の結果 [1] とよく一致する．

このように，学習の結果得られた知覚影響量 $g(\Delta t,\Delta f)$ とこれまでの心理実験の結果は同じような特徴をもっており，今回提案するモデルは心理実験の結果からも支持されるものである．

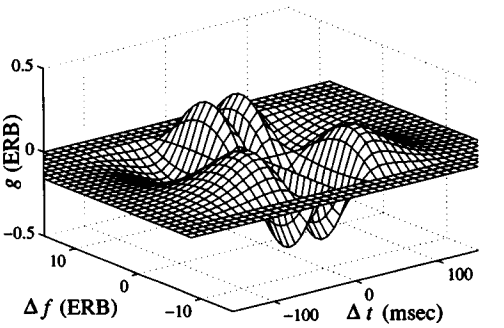

図 3 g の学習結果
Fig. 3 Learning result of g.

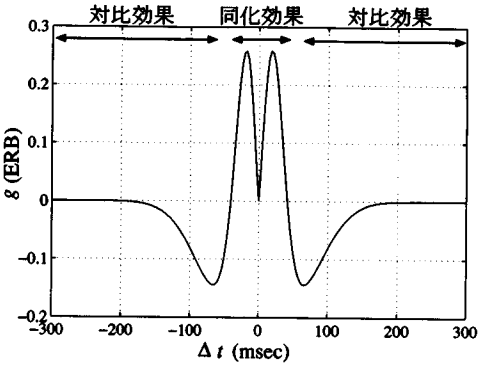

図 4 Δt と g の関係 ($\Delta f=4$ ERB)
Fig. 4 Δt–g relationship at $\Delta f=4$ ERB.

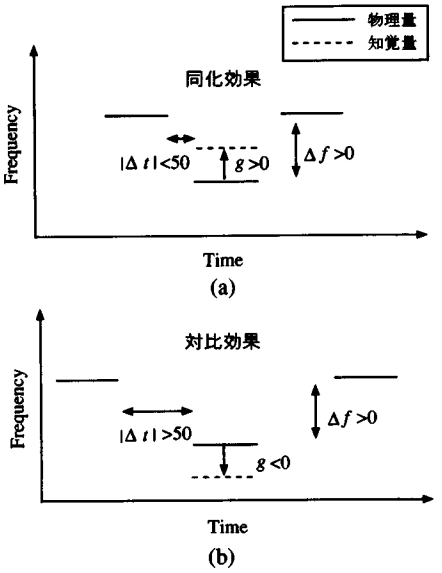

図5 同化効果と対比効果
Fig. 5 Assimilation (a) and contrast (b) effect.

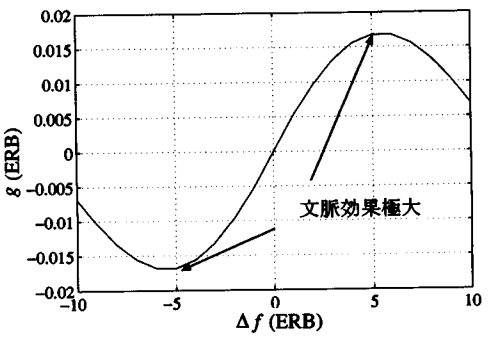

図6 Δf と g の関係 ($\Delta t = 30$ msec)
Fig. 6 Δf-g relationship at $\Delta t = 30$ msec.

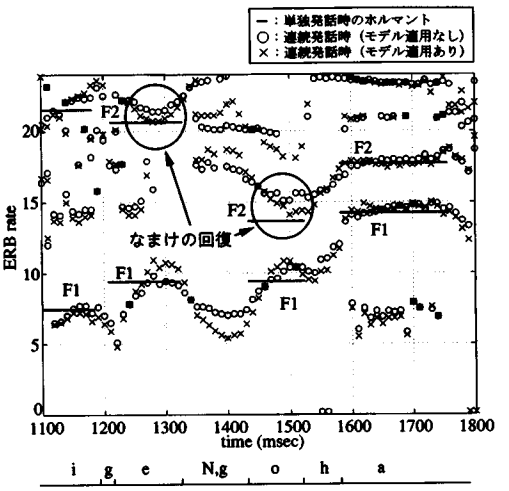

図7 スペクトルピーク軌跡の例
Fig. 7 Example of spectral peak trajectories.

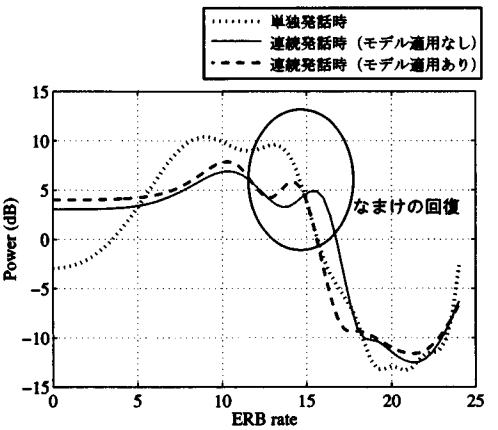

図8 /o/ のスペクトル
Fig. 8 Spectra of /o/.

3. モデルによるなまけの回復

連続音声には単独発話時の特徴量に達しきらない「なまけ」が見られ，これが音声認識の妨げとなる．もし，モデルにより連続音声中に現れるなまけを回復できるのであれば，モデルを前処理として用いることにより，認識能力の向上が期待できる．そこで，モデルによるなまけの回復能力について考察する．

例として，文章データ（ATR 音声データベース:タスクコード mhtsc322）の一部分のスペクトルピーク軌跡を図 7 に，そのうち/o/の母音中心におけるスペクトルを図 8 に示す．図中で，モデルを適用しない場合かなりのなまけが見られ，連続発話時のホルマント位置は単独発話時とはかなり異なっていることがわかる．一方，モデルを適用するとなまけが回復し，単独発話時の特徴量に近づいていることがわかる．モデルによるスペクトルの補正は周波数方向に限られたものではあるが，例えば図 8 では，モデル適用により F2 の位置が単独発話時の位置に近づき，単独発話時と連続発話時とのスペクトル間距離は約 7%小さくなっている．

一方，図 7 の/e/の F1 の箇所のように，あまりなまけていない箇所に対してモデルを適用すると過度の補正を加えて単独発話時の位置から離れてしまう場合が

あり，今後改善する必要がある．しかし，全体としては，文脈効果モデルはなまけたホルマント軌跡をオーバシュートさせ，なまけを回復させるのに有効に機能していると言える．

4. モデルを前処理として用いた母音中心の認識

文脈効果モデルは母音中心の認識率が向上することをねらって設計されている．そこで，その能力を検証するため，連続音声中の母音中心の認識実験を行った．音声データは，$g(\Delta t, \Delta f)$ の学習に使用したデータ（表 2）に加え，テストデータとして自由発話文章データ 50 文章（タスクコード mhtsc2，mhtsc3）を用いた．分析条件は表 2 と同一である．

実験では，連続発話文章中の母音中心のスペクトルを抽出し，モデルによりスペクトルを補正した．認識に際しては，単母音のスペクトルとのスペクトル間距離を比較し，距離が最小の母音を認識結果とした．また，比較のため，モデルによる補正を行わない場合も同様の方法で認識を行った．

結果を図 9 に示す．いずれのデータセットにおいても，モデルを適用してスペクトルを補正することにより，認識率が向上している．$g(\Delta t, \Delta f)$ の学習に使用した mhtsc1 における認識率の向上は当然であるが，それ以外のデータセットにおいてもモデルを適用することにより母音中心の認識率の向上が見られた．また，全体の認識率の向上は，χ^2 検定[11]の結果有意なものであった（$\chi^2 = 5.05$，有意水準 5%）．この結果から，モデルが母音認識の前処理として有効に機能していることがわかる．

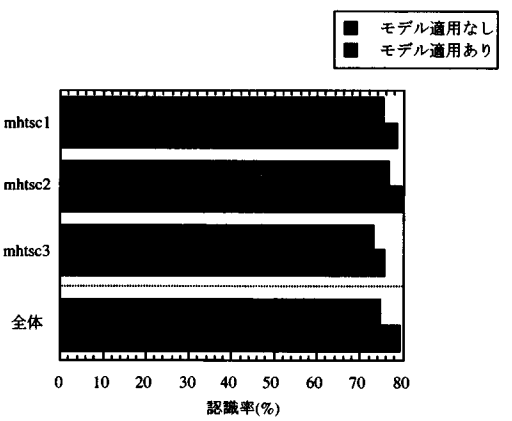

図 9　母音中心の認識率の比較
Fig. 9　Comparison of vowel recognition accuracy.

5. モデルを前処理として用いたワードスポッティング[12]

連続発話文章データをテストデータ，単語データを標準データとして用いるワードスポッティングでは，発話形態の違いによるスペクトル形状の違いが誤認識の原因の一つとなる．つまり，テストデータはなまけの程度が大きく，標準データはなまけの程度が小さいため，このなまけの程度の違いが認識の妨げとなる．従って，テストデータに文脈効果モデルを適用してなまけを回復させれば，ワードスポッティング能力が向上することが期待できる．

そこで，ワードスポッティングのテストデータとして用いる連続発話文章データのスペクトルを文脈効果モデルを用いて補正し，モデルの能力を検証する．

5.1 音声データ

これまでと同様に ATR 音声データベース中の男性話者 mht のデータを使用し，テストデータ，標準データとして次のものを用いた．

[テストデータ]
国際会議予約に関する自由発話文章データ 75 文章（タスクコード sc1，sc2，sc3）．文節数は 695 個
[標準データ]
次の 13 の単独発話単語データ（タスクコード 1）．但し，括弧内はテストデータでの出現回数
会議（23），通訳（12），発表（11），会場（9），資料（8），英語（7），翻訳（7），言語（7），論文（6），締切（4），原稿（3），要約（3），説明（2）

5.2 認識方法

図 10 にワードスポッティングの流れを示す．テストデータ，標準データから 40 次 FFT ケプストラム平滑化スペクトルを求め，テストデータのスペクトルに対し文脈効果モデルを適用して補正を加えた．なお，分析条件は表 2 と同様である．次に，テストデータと標準データとの間で連続 DP [13] によりスコアの計算を行った．スコアの計算の際にはスペクトル間の距離を用いている．つまり，テストデータと標準データのスペクトルの差異が少ないほどスコアは小さな値をとる．そして，しきい値よりも小さく，かつ最も小さいスコアである単語を認識出力とした．

5.3 認識結果

文脈効果モデルを用いてスペクトルを補正した場合と，比較のため文脈効果モデルを用いなかった場合

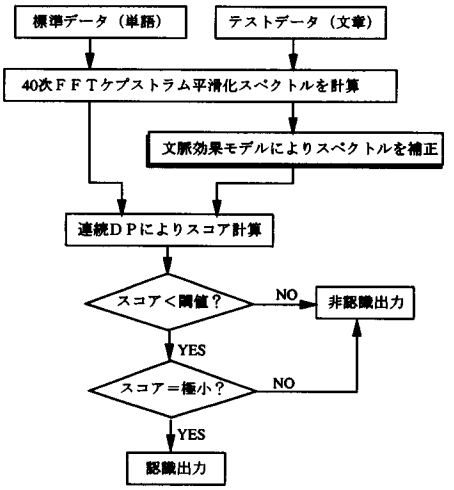

図10 モデルを用いたワードスポッティングのブロック図
Fig. 10 Block-diagram of the word spotting with the model.

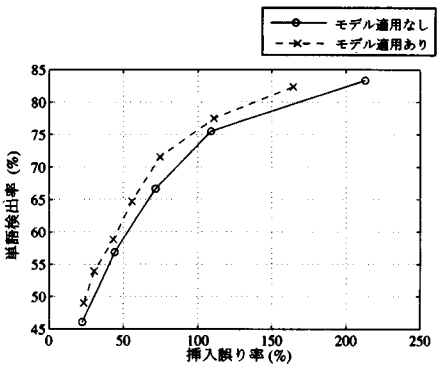

図11 ワードスポッティング能力の比較
Fig. 11 Comparison of word spotting performance.

の単語検出率を図11に示す．図11において横軸の挿入誤り率は挿入誤り数/認識対象単語数としている．図11から，モデルを用いることにより挿入誤り率が小さく，また単語検出率が大きくなっていることがわかり，ワードスポッティング能力が向上していることがわかる．これはモデルによりなまけが回復し，誤認識が減ったためと考えられる．

一方，モデルは母音を対象に設計されているため，モデルにより母音のなまけが回復するとはいうものの，子音に対する有効性は保証されていない．しかし，このようなワードスポッティング能力の向上が見られたことは，モデルが子音に対して，少なくとも重大な不整合をもつものではないことを示している．ワードスポッティング能力の更なる改善には子音の特性を考慮したモデルの設計が必要ではあるが，母音のみを対象に設計された現在のモデルでも，有効に機能していると言える．

6. むすび

スペクトルを対象として文脈効果のモデル化を行った．モデル中のスペクトルピークによる知覚的影響量 g は連続音声の母音認識率の向上をねらって最小分類誤り学習により求めたが，学習結果の特徴はこれまでの心理実験の結果と適合するものであった．

また，モデルを前処理として用いてスペクトルを補正し母音認識およびワードスポッティングを行った．モデルにはホルマント軌跡をオーバシュートさせ，なまけを回復させる働きがある．その結果，連続音声の母音認識率が向上したほか，ワードスポッティング能力の向上が見られた．これらは，モデルが認識の前処理として有効に機能し，単独発話時と連続発話時のスペクトルの違いを吸収しているためである．

一方，今回の文脈効果モデルは周波数方向の知覚の移動のみを考慮している．そのため，図8に見られるように，周波数方向のなまけはモデルにより緩和されるが，パワー方向の差異は解消されない．よりいっそう調音結合の影響の解消を図るには，この点についても考察する必要がある．また，全体としてはモデルを用いることによりなまけが回復されるが，個々について見ると図7の/e/の箇所のようにあまりなまけていない箇所に対して過度の補正を加えてしまうことがある．そして，モデルは母音を対象に設計されているため，子音に対しては有効性は保証されておらず適切に機能していない可能性も否定できない．今回はモデルの単純化のため知覚的影響量 g はスペクトルピークとの時間差と周波数差（Δt と Δf）のみによって定まるとしているが，最近の研究[14]からは，スペクトルのなまけの程度や音韻性が文脈効果の大きさに関与していることが明らかになっている．今後，これらの要素などをモデルに組み込むことにより，これらの問題を解決し更にモデルの能力を高める必要がある．

謝辞 本研究の一部は，立石科学技術振興財団の助成（931001）によるものである．

文　献

[1] M. Akagi, "Modeling of contextual effects based on spectral peak interaction," J. Acoust. Soc. Am. vol.93, no.2, pp.1076-1086, Feb. 1993.

[2] B.E.F. Lindblom, "On the role of formant transitions

in vowel recognition," J. Acoust. Soc. Am. vol.42, no.4, pp.830–843, 1967.
[3] S. Sigeno, "Assimilation and contrast in the phonetic perception of vowels," J. Acoust. Soc. Am. vol.90, no.1, pp.103–111, July 1991.
[4] R.J.J.H. van Son and Louis C.W. Pols, "The influence of local context on the identification of vowels and consonants," Eurospeech'95, Madrid, Spain, no.WEam2B.1 pp.967–970, Sept. 1995.
[5] 藤崎博也，樋口宜男，二見　徹，重野　純，"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, Oct. 1981.
[6] 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, Jan. 1985.
[7] B.-H. Juang and S. Katagiri, "Discriminative learning for minumum error classification," IEEE Trans. SP, vol.40, no.12, pp.3043–3054, Dec. 1992.
[8] 米沢裕司，赤木正人，"最小分類誤り学習による文脈効果モデルの定式化，" 信学技報，SP94-114, March 1995.
[9] B.R. Glasberg and B.C.J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hearing Research, pp.103–138, 1990.
[10] 武田一哉，匂坂芳典，片桐　茂，阿部匡伸，桑原尚夫，"研究用日本語音声データベース利用解説書，" ATR Technical Report, TR-I-0028, May 1988.
[11] P.G. ホーエル，"入門数理統計学，" 培風館，1978.
[12] 米沢裕司，赤木正人，"文脈効果モデルを用いたワードスポッティング，" 信学技報，SP95-108, Dec. 1995.
[13] 速水　悟，岡　隆一，"連続DPによる連続単語認識実験とその考察，" 信学論（D），vol.J67-D, no.6, pp.677–684, June 1984.
[14] 萩原　力，米沢裕司，赤木正人，"文脈効果の大きさと音韻性の関係について，" 信学技報，SP95-139, March 1996.

（平成8年2月16日受付，8月15日再受付）

米沢　裕司　（学生員）

平5阪府大・工・電気卒．平7北陸先端大博士前期課程了．現在，同大博士後期課程在学中．聴覚機構のモデル化の研究に従事．日本音響学会会員．

赤木　正人　（正員）

昭54名工大・工・電子卒．昭59東工大大学院博士課程情報工学専攻了．工博．同年電電公社（現NTT）研究所入社．以来，ATR視聴覚機構研究所，NTT基礎研究所を経て，現在，北陸先端科学技術大学院大学情報科学研究科助教授．この間，昭63米国MIT客員研究員，平5オランダアムステルダム大学客員研究員．音声信号処理，聴覚機構のモデル化の研究に従事．昭63年度本会論文賞受賞．日本音響学会，IEEE，ASA，ESCA各会員．