

論文 / 著書情報
Article / Book Information

論題(和文)	ホルマント移動パタン距離尺度
Title(English)	
著者(和文)	相川 清明, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1988年春季講演論文集, , , pp. 47-48
Citation(English)	, , , pp. 47-48
発行日 / Pub. date	1988, 3

1-2-24 ホルマント移動パターン距離尺度*

○相川 清明 古井 貞熙

(NTTヒューマンインタフェース研究所)

1. はじめに 対数スペクトルの時間周波数微分(ホルマント移動関数)によりホルマント移動を方向を区別して抽出できることは既に報告した[1]。本報告では、ホルマント移動関数により抽出された2つのホルマント移動パターン間の距離尺度を導入し、これを従来のスペクトル距離尺度と併用することにより、従来のスペクトル距離尺度のみを用いた場合に比べて音声認識誤りを大巾に減少できることを示す。

2. ホルマント移動パターン距離尺度の導出 ホルマント移動パターン $F(\omega, t)$ は対数スペクトル $S(\omega, t)$ の時間周波数微分により求められる。対数スペクトルは j 次ケプストラム係数 $c_j(t)$ を用いて

$$S(\omega, t) = c_0 + 2 \sum_{k=1}^K c_k(t) \cos k\omega \quad (1)$$

と展開できるので、ホルマント移動パターンは以下のように表わすことができる。

$$F(\omega, t) = \frac{\partial^2 S(\omega, t)}{\partial \omega \partial t} = 2 \sum_{k=1}^K c'_k(t) k \sin k\omega \quad (2)$$

2つの対数スペクトル $S_1(\omega, t), S_2(\omega, t)$ のケプストラム距離

$$D_{CEP}(t) = \frac{1}{\pi} \int_0^\pi (S_1(\omega, t) - S_2(\omega, t))^2 d\omega \quad (3)$$

$$= (c_{10} - c_{20})^2 + 2 \sum_{k=1}^K (c_{1k}(t) - c_{2k}(t))^2 \quad (4)$$

と同様に、2つのホルマント移動パターン $F_1(\omega, t), F_2(\omega, t)$ の間の距離を

$$D_{FM}(t) = \frac{1}{\pi} \int_0^\pi (F_1(\omega, t) - F_2(\omega, t))^2 d\omega \quad (5)$$

$$= 2 \sum_{k=1}^K (c'_{1k}(t) - c'_{2k}(t))^2 \ell^2, \quad \begin{matrix} \ell = L & k \geq L \\ \ell = k & k < L \end{matrix} \quad (6)$$

のように定義することができる。

本来ならば、 ℓ の値は周波数微分に相当する k であり、 $c'_k(t)$ は $c_k(t)$ の時間微分であるが、微分によるスペクトル推定誤差の拡大を避けるため、スペクトルの時間周波数方向の平滑化を微分と同時に進めようことにした。平滑化を含む周波数微分は $\ell = L$ で ℓ を飽和させることにより得られる[2]。また、平滑化を含む時間微分は $c'_k(t)$ として $c_k(t)$ の1次回帰直線の傾きを用いることにより得られる[3]。式(6)は東倉の重み付きケプストラム距離 D_{WCEP} [2] に類似するが、時間微分を含む点が異なる。

* Formant Movement Pattern Distance Measure

By Kiyooki AIKAWA and Sadaoki FURUI (NTT Human Interface Laboratories)

3. スペクトル距離尺度との併用 音声認識にはホルマント移動パタンのみではなく、スペクトルパタンも必要であると考えられるので、従来のスペクトル距離尺度とホルマント移動パタン距離尺度を組み合わせた尺度

$$D_{FMCEP}(t) = r D_{FM}(t) + (1-r) D_{CEP}(t)$$

$$D_{FMWLR}(t) = r D_{FM}(t) + (1-r) D_{WLR}(t)$$

$$D_{FMPWLR}(t) = r D_{FM}(t) + (1-r) D_{PWLR}(t)$$

を用いることにした。 $D_{PWLR}(t)$ はパワー込みWeighted

Likelihood Ratio(WLR)スペクトル距離尺度[4] である。

4. 実験 男10名の音声を試験音声、他の男1名の音声

を標準音声と見なしてDPマッチングによる50都市名認識実験を行なった。サンプリング周波数は8kHz、フレーム長は32ms、フレーム周期は16msとした。フレーム間距離尺度として $D_{FMCEP}(t)$ を用いた場合、1次回帰直線を求めるのに用いるフレーム数(ウィンドウ幅)Wに対して、飽和ケフレンシL、ホルマント移動パタン距離尺度重みr、を変えて求めた単語認識誤り率の最小値を表1に示す。表1より、 $W=5$ のとき最も誤りが少なく3.8%であることがわかる。この値は $D_{CEP}(t)$ のみを用いた場合の認識誤り率9.5%の2/5である。Wが5の場合、L、rによる単語認識誤り率の変化を図1に示す。また、 $W=5$ 、 $L=4$ の条件において、フレーム間距離尺度として $D_{FMWLR}(t)$ を用いた場合の認識誤り率の最小値は2.1%($r=0.6$)、 $D_{FMPWLR}(t)$ を用いた場合には1.7%($r=0.6$)であった。 $D_{WLR}(t)$ 、 $D_{PWLR}(t)$ のみを用いた場合の認識誤り率はそれぞれ3.2%、2.0%であったので、いずれの場合にもホルマント移動パタン距離尺度を併用することにより、認識誤りが減少していることが分る。これらの結果を図2に示す。なお、東倉の D_{WCEP} を用いた場合の認識誤り率は5.9%であった。

5. むすび 従来のスペクトル距離尺度に、スペクトルの動特性を示すホルマント移動関数により抽出されたホルマント移動パタン距離尺度を併用することにより、スペクトル距離尺度のみの場合に比べて単語認識誤りを減少できることを示した。

謝辞 討論して頂いた嵯峨山茂樹氏をはじめとする音声情報研究部の諸氏に感謝する。

文献 [1] 相川、音学講論、Vol.1, pp.99-100, (1987-10). [2] Tohkura, IEEE Trans., ASSP-35, No.10, pp.1414-1422, (1987-10). [3] Furui, IEEE Trans., ASSP-34, No.1, pp.52-59, (1986-2). [4] 相川、信学論 Vol. J68-D, No.3, pp.316-322, (1985-3).

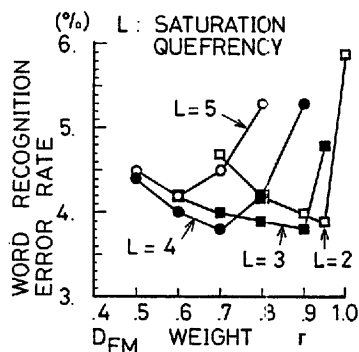


図1 飽和ケフレンシL、ホルマント移動パタン距離尺度重みrによる単語認識誤り率の変化(W=5)

BASE
DISTANCE
MEASURE

D_{CEP}
 D_{WLR}
 D_{PWLR}

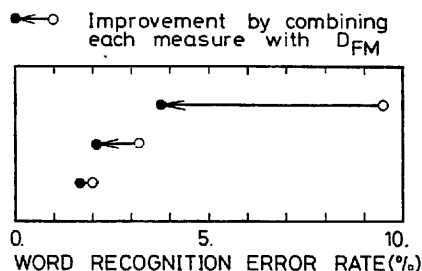


図2 スペクトル距離尺度とホルマント移動パタン距離尺度を併用することによる単語認識誤り率の改善