

論文 / 著書情報
Article / Book Information

論題(和文)	FMニューロンモデルを用いた音声認識
Title(English)	
著者(和文)	相川 清明, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	昭和63年電子情報通信学会秋季全国大会, , , pp. A-1-113-114
Citation(English)	, , , pp. A-1-113-114
発行日 / Pub. date	1988, 9
URL	http://www.ieice.org/jpn/books/t_g.html
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) 1988 Institute of Electronics, Information and Communication Engineers.

SA-1-3 FM ニューロンモデルを用いた音声認識

Speech Recognition Using FM-neuron Model

相川 清明 古井 貞熙

Kiyoaki AIKAWA and Sadaoki FURUI

NTT ヒューマンインタフェース研究所

NTT Human Interface Laboratories

1. はじめに 聴覚神経系には音の立ち上がりに反応する ON 型ニューロンや、周波数の推移に反応する FM ニューロンなど、特定の機能を持つ神経系がある[1]。ON 型ニューロンモデルを用いると音素境界を精度良く検出できる[2]。また、FM ニューロンモデルを用いるとスペクトルの移動を抽出できる[3,4]。動的な特徴は音素知覚上で重要であるので[5]、スペクトル移動の情報をスペクトル情報と併用することにより、スペクトル情報のみを用いた場合に比べて音声認識率を向上させることができる[4,6]。本報告では、従来の各種スペクトル距離と FM ニューロンモデルの出力から得られるスペクトル移動パタン距離を組み合わせることにによる音声認識率の向上、及び、バンド幅拡大によるスペクトル移動パタンの平滑化の効果について述べる。

2. スペクトル移動パタン スペクトル移動量をその移動方向を区別して求めることができる FM ニューロンのモデルは、数学的には対数スペクトルの時間周波数微分によって表現できることがわかった[3,4]。この演算により抽出された特徴をスペクトル移動パタンと呼び、 $F(\omega, t)$ により記述する。FM ニューロンモデルによるスペクトル移動抽出のメカニズムを図 1 に示す。

スペクトル推定誤差による微小変動は時間周波数微分により強調されるので、これを抑制するために時間周波数平滑化が必要である。微分も平滑化もフィルタリングと考えることができるので両者は同時に行なうことができる。対数スペクトル $S(\omega, t)$ は k 次ケプストラム係数 $c_k(t)$ を用いて

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

と展開できるので、平滑化されたスペクトル移動パタン $F(\omega, t)$ は

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

$$f(k) = \begin{cases} -1 & k \geq L \\ k & k < L \end{cases}$$

により表わすことができる。なお、 ω は周波数、 t は時間を表わす。 $c'_k(t)$ は $c_k(t)$ の 1 次回帰直線の傾き[7]、 L は周波数微分のための重み付けが飽和するケフレンシであり、平滑化のためのカットオフケフレンシにあたる[8]。

LPC スペクトルを求める際に鋭いピークを平滑化する方法としてバンド幅拡大法がある[9]。拡大するバンド幅を B (Hz)、サンプリング間隔を T (sec) とすると k 次の線形予測係数 a_k は

$$a_k = a_k e^{-\pi B T k} \quad (3)$$

により変換され、この係数から平滑化されたスペクトルのケプストラム展開係数が求まる。

3. スペクトル移動パタン距離 ケプストラム距離と同様に、2つのスペクトル移動パタン $F_1(\omega, t)$, $F_2(\omega, t)$ の間の距離を

$$D_{SM}(t) = \frac{1}{\pi} \int_0^{\pi} (F_1(\omega, t) - F_2(\omega, t))^2 d\omega \quad (4)$$

$$= 2 \sum_{k=1}^K (c'_{1k}(t) - c'_{2k}(t))^2 f(k)^2$$

のように定義することができる。これをスペクトル移動パタン距離と呼ぶことにする。式(4)は古井の動的特徴を用いた距離 D_{ACEP} [7] や、東倉の重み付きケプストラム距離 D_{WCEP} [8] に類似するが、 D_{ACEP} はスペクトルの時間微分、 D_{WCEP} はスペクトルの周波数微分をとらえているのに対して、本方法はスペクトルの時間周波数微分をとらえていることが異なる。

パワースペクトルが $P(\omega, t)$ である音声、伝達関数が $H(\omega)$ である伝送路を通った後でのスペ

クトル移動ボタンは

$$\frac{\partial^2 \ln(P(\omega, t) H(\omega))}{\partial \omega \partial t} = \frac{\partial^2 \ln P(\omega, t)}{\partial \omega \partial t} + \frac{\partial^2 \ln H(\omega)}{\partial \omega \partial t} = \frac{\partial^2 \ln P(\omega, t)}{\partial \omega \partial t} \quad (5)$$

となり、FMニューロンモデルには伝送路周波数特性の影響を除去できる特徴があることがわかる。

4. スペクトル距離との併用 音声認識にはスペクトル移動ボタンのみではなく、スペクトルボ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 D_{SPEC} とスペクトル移動ボタン距離 D_{SM} を

$$D(t) = r D_{SM}(t) + (1-r) D_{SPEC}(t) \quad (6)$$

のように組み合わせた距離を用いることにした。 D_{SPEC} としては D_{CEP} (ケプストラム距離)、 $D_{\Delta CEP}$ 、 D_{WCEP} 、 D_{WLR} (Weighted Likelihood Ratio(WLR)距離)、 D_{PWLR} (パワー込みWLR[10] 距離)を用いる。

5. 実験 種々の距離 D を用いる効果を D P マッチングによる50都市名認識実験により調べた。男1名の音声を標準音声と見なし、別の男10名の音声を認識した。以下の結果には10人の認識率の平均を示す。サンプリング周波数は8 kHz、フレーム長は32 ms、フレーム周期は16 msとした。 D_{SPEC} として D_{CEP} を用いて予備実験を行なった結果、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)は5、バンド幅拡大は80Hz、飽和ケフレンシは3が最適であることがわかった。以後の実験では全てこの値を用いている。

各 D_{SPEC} を単独で用いた場合、 D_{SM} と併用した場合、さらに、スペクトル移動ボタンをバンド幅拡大法により平滑化した場合の単語認識誤り率の比較を表1に示す。ただし、結合係数 r が最適な場合のみについて示した。表1から、いずれの距離を用いた場合においても D_{SM} を組み合わせることにより誤りが減少していることがわかる。また、バンド幅拡大による平滑化によりさらに誤りを減少できることがわかる。

6. むすび FMニューロンモデルによるスペクトル移動ボタン抽出法とその性質について述べ、従来のスペクトル距離とスペクトル移動ボタン距離を併用することにより、スペクトル距離のみを用いる場合に比べて単語認識誤りを減少できることを示した。また、バンド幅拡大法によりスペクトルの平滑化を行うと、認識誤りをさらに減少できることを示した。

文献 [1] 大沢ほか、神経回路と生体制御、生物科学講座6、(1979)。[2] 相川、信学論、Vol. J71-A, No. 3, pp. 592-600, (1988-3)。[3] 相川、音学講論、Vol. 1, pp. 99-100, (1987-10)。[4] Aikawa and Furui, ICASSP88 Proceedings, Vol. 5, pp. 223-226, (1988-3)。[5] Furui, J. Acoust. Soc. Am., 80(A), pp. 1016-1025, (1986-10)。[6] 相川、音学講論、Vol. 1, pp. 47-48, (1988-3)。[7] Furui, IEEE Trans., ASSP-34, No. 1, pp. 52-59, (1986-2)。[8] Tohkura, IEEE Trans., ASSP-35, No. 10, pp. 1414-1422, (1987-10)。[9] 東倉、博士論文、東京大学、(1980)。[10] 相川、信学論 Vol. J68-D, No. 3, pp. 316-322, (1985-3)。

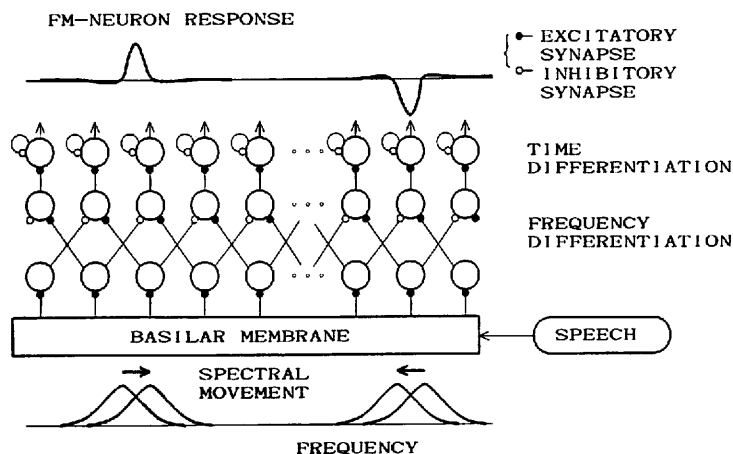


Fig.1 FM-neuron model

表1 各種スペクトル距離とスペクトル移動ボタン距離を組み合わせることによる認識率の改善、及びバンド幅拡大によるスペクトル移動ボタン平滑化の効果

スペクトル距離尺度	単独	D_{SM} との組み合わせ	バンド幅拡大
CEP	9.5 %	3.8	3.5
ΔCEP	5.0	4.2	4.1
WCEP	5.9	3.5	3.1
WLR	3.2	2.1	2.0
PWLR	2.1	1.6	1.6