

ニューラルネットを用いた擬似ホルマント抽出フィルタリング

高木, 英行
松下電器産業株式会社中央研究所

坪香, 英一
松下電器産業株式会社中央研究所

<https://hdl.handle.net/2324/4479721>

出版情報 : 日本音響学会講演論文集, pp.249-250, 1988-03. The Acoustical Society of Japan
バージョン :
権利関係 :

◎高木英行 坪香英一 (松下電器中研)

1. はじめに 音声認識前処理部として神経回路網モデルでホルマント抽出を行った。聴覚神経系のモデル化は目的とせず、信号処理への応用の可能性を探ることを目的とする。

2. モデルの構成 backpropagation¹⁾で学習した2層の実験モデルを Fig.1~3 に示す。モデル1はフレーム間の時間情報を持たないモデルである。モデル2は周波数の絶対値情報を持たず、スペクトラム全体をスキャン入力する。ホルマントの周波数変動に限りがあることに着目し、モデル3では、上位層のユニットは下位層の近傍ユニットとのみ接続関係を持たせる。従ってユニットの数に比較してモデル規模を小さくできる。

3. 実験方法 [3.1 入力データ] フレーム長10ms・フレーム周期5msで分析した対数ワースペクトル(モデル1と2用)と、聴覚フィルタ出力後、フレーム長・周期共に5msで求めた対数ワースペクトル(モデル3用)を用いた。聴覚フィルタ²⁾は1/2Bark毎に1Bark帯域のBPFを34個並べたもので、各フィルタは基底膜の共振特性を基に設計した。

[3.2 学習データ] LPC分析から得られた種をスペクトラムに照らし合わせ、誤抽出データを修正して修正した6単語のホルマントデータを作成する。次に入力信号のワードが閾値以下の無音区間はホルマントなしと判断し、閾値以上の区間についてのみ各モデル入力に対応するホルマント学習データを作成する。6単語から得られたモデルの学習データ数は各々1459, 29504, 785個である。

4. 実験結果と考察 学習・未学習データの例、および各モデルの出力を Fig.4に示す。モデル1の出力には不連続部が見られ、フレーム間情報の必要性が伺える。モデル2は対象とする周波数・時間の小領域が小さすぎるのではないかとと思われる。モデル3にはピッチ周期の影響が見られる。このため聴覚フィルタ出力と Fig.4 (a0), (b0)とを比較検討した結果、ホルマントに対するピッチの影響の割合が相対的に聴覚フィルタ出力の方が大きいためであると思われる。何れのモデルも、部分情報の学習にも拘らず学習した音韻のホルマントを抽出できていると判断でき、神経回路網モデルの有効性が確認された。

5. 今後の課題 (1)各種ホルマント構造の学習、(2)認識部との結合、が必要である。認識部も神経回路網モデルで構成すれば多層化した一つのモデルで音声認識実現の可能性がある。また、学習用ホルマントデータ作成の労力軽減のため、教師なし学習モデルの検討も重要な課題である。

1) Rumelhart, D.E. et al., "Learning Representation by Back-Propagation Errors." Nature, Vol. 323, pp533-536, Oct. 9 1986

2) 駒木根他, "蝸牛の周波数分析機能を模擬するフィルタバンクの構成法" 信学技報 SP87-45, 昭和62年7月

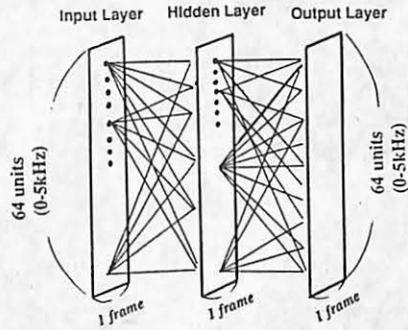


Fig. 1 Model 1 reads spectrum frame by frame and has no time information. (unit number: 64-64-64, connection number: 8192)

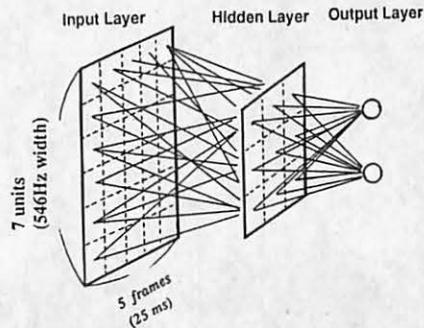


Fig. 2 Model 2 sweeps on spectrogram and reads partial spectrogram. (unit number: 35-9-2 connection number: 333)

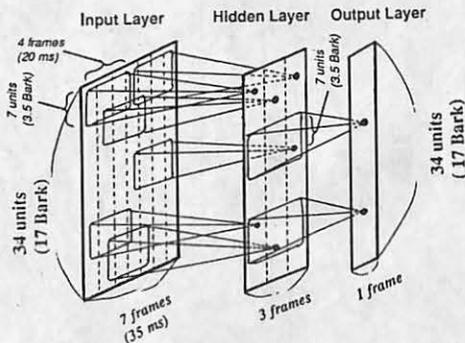


Fig. 3 Model 3 reads auditory spectrogram. A unit in upper layer is connected to units just under itself in lower layer. (unit number: 238-102-34, connection number: 3570)

* "Pseudo-Formant Extraction by Neural Net," by Hideyuki TAKAGI and Eiichi TUBOKA (Matsuchita Electric Industrial Co., LTD)

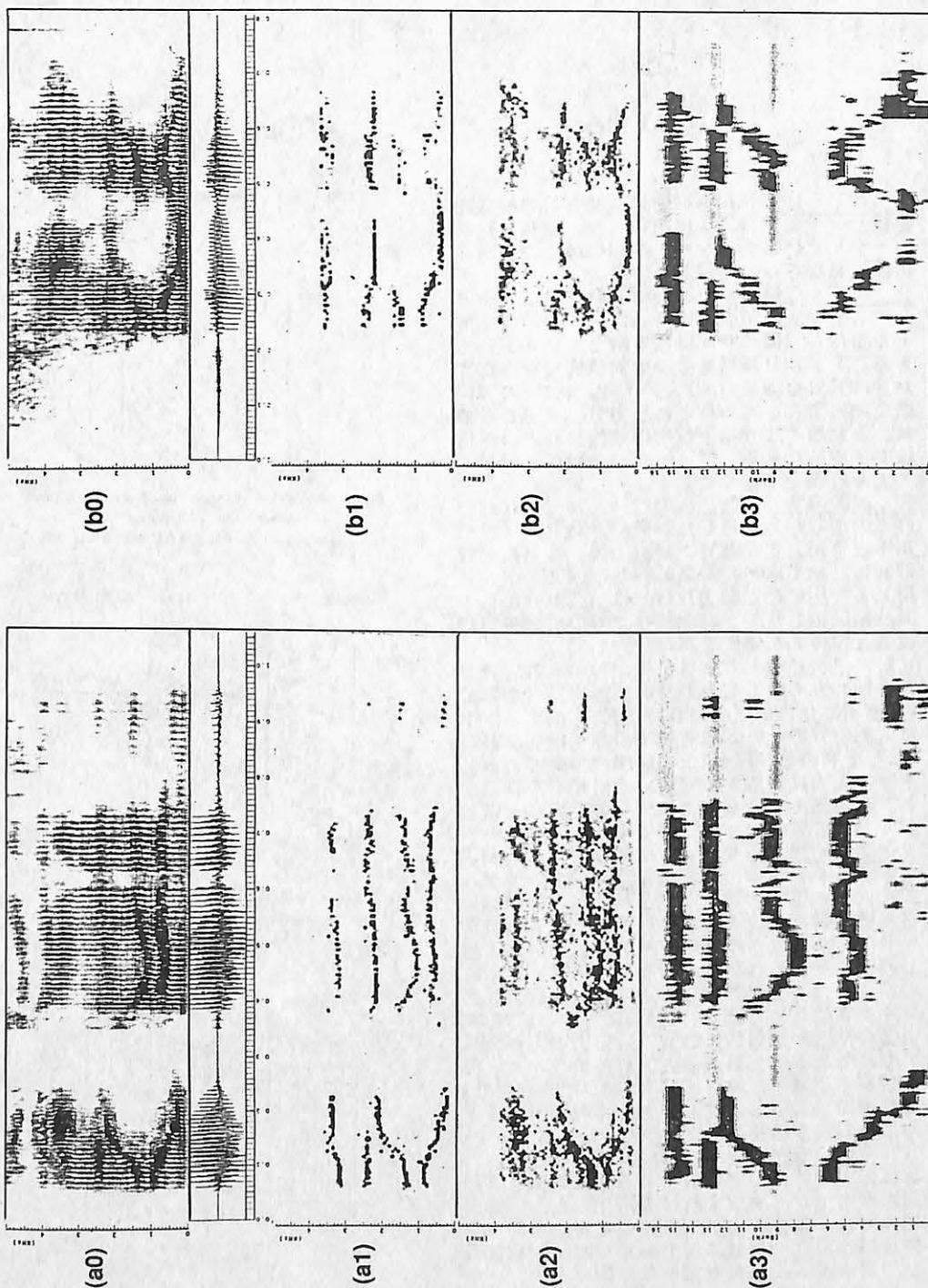


Fig.4 (a0) and (b0) are spectrograms of learning data and unlearning data respectively. (a1), (a2), and (a3) are outputs of Model 1, 2, and 3, when (a0) is input to each model. (b1), (b2), and (b3) are outputs of Model 1, 2, and 3, when (b0) is input to each model.