6D−4

不均一標本化されたスペクトログラムを入力とするニューラルネットによる数字音声認識

加藤誠巳　鵜飼敏之

(上智大学理工学部)

1. まえがき

ニューラルネットを用いて、音声認識を行おうとする研究が各方面で盛んに行われている。しかし通常のニューラルネットは音声パターンのような時間的に変化する構造を処理する能力に欠けているため、時間構造を考慮できるようなニューラルネットについて多くの報告がなされている。ここでは不均一標本化された特徴量[1]によって、数字音声を認識するニューラルネットについて検討を行った結果を報告する。

2. 不均一標本化入力データおよびネットワークの構造

音声認識の際、音声の特徴を効率よく表現するためには個々の音素の情報を有効に利用する必要がある。そのために周波数軸上での情報を的確に表す特徴量を選択し、さらに時間軸上での情報量の密度の変化によって特徴量の標本化間隔を変化させる必要がある。すなわち定常的な母音等では標本化間隔を疎に、音響的特徴が急激に変化するような音に関しては密に標本化を行う。これによって音声信号処理に於ける記憶容量や処理時間の低減が可能となってきた。ここではこの不均一標本化された特徴量を更に低減して入力層に数の上で制限のあるニューラルネットへの入力とし、ニューラルネットによる単語音声認識に応用した。

認識対象は離散数字(0〜9)音声とした(読みが二通りあるものがあるので計14種類)。まず入力データの作成方法について述べる。16チャネルのフィルタから10ms間隔でサンプリングした出力値から音声の始終端を検出する。その検出した音声区間に対してパワーがある閾値を越えてから不均一標本化の処理を行い、これを入力特徴量とする。この入力特徴量の作成には現在のところ富士通製音声入出力装置FACOM2360を使用している。

この16次元特徴量の時系列を用いて学習を行うが、これをニューラルネットの入力とする際、時間方向に長さが一定でないため、この特徴量からさらに変化の激しい上位16点を用いてニューラルネット入力特徴量とした。変化点抽出の処理としては定常的な波形にも関わらず波形のパワーの変化によって変化点となってしまう場合を避けるため、サンプル間の2階差分によって変化点を抽出した。これによってニューラルネットの入力ノード数は16次元×16サンプル=256ノードとなる。

図1にここで用いたネットワークの構造を示す。また図2に前述した処理をした入力波形の例を示す。図2 (a), (c)は不均一標本化された二つの特徴量波形/イチ/の例であり、(b), (d)は(a), (c)各波形から変化の激しい16点を抽出したニューラルネットへの入力特徴量の波形である。

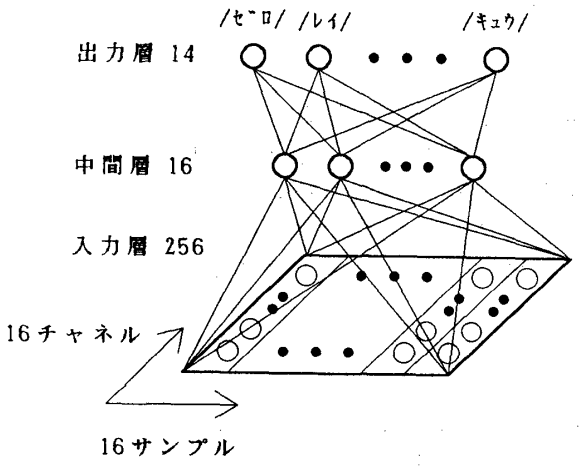
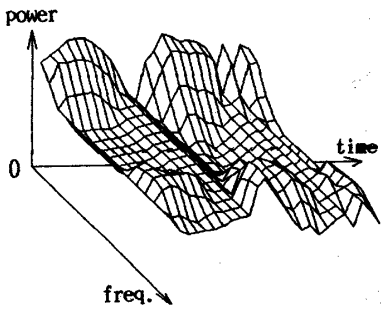
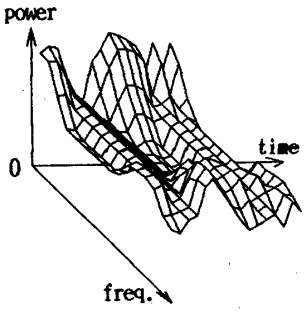


図1　ネットワークの構造

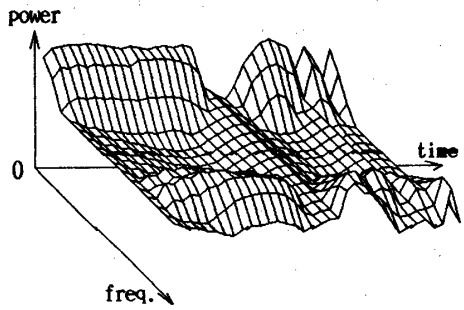
Spoken Number Recognition by Neural Nets using Non-uniform Sampling of Spectrum
Masami Kato, Toshiyuki Ukai
Sophia University



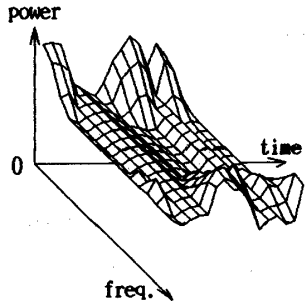
(a) 不均一標本化波形/イチ/の例



(b) ニューラルネット入力波形



(c) 不均一標本化波形/イチ/の例



(d) ニューラルネット入力波形

図2　不均一標本化波形

3. 認識結果

男性9名が各10回ずつ発声した離散10数字のうち、4回分を学習用に、6回分を認識用に用いた。学習は前述した様に特徴量の変化の激しい16点を抽出して行った。認識実験では、学習データを作成したのと同じ方法によって入力データを作成し、中間層の数を様々変えて認識実験を行った。この結果を表1に示す。この様に比較的少ないデータ数であるにもかかわらずある程度の認識率を得ることが出来た。

表1 認識結果

中間層数 (個)	認識率 (%)(closed)	認識率 (%)(open)
14	100.0 (504/504)	92.6 (700/756)
16	100.0 (504/504)	94.7 (716/756)
18	99.8 (503/504)	92.6 (700/756)
20	100.0 (504/504)	93.0 (703/756)

4. むすび

ニューラルネットを用いた単語音声認識で問題となることの一つとして、それぞれ単語によって時間長が極端に変わってしまうということがあげられる。そのため、ここでは不均一標本化されたスペクトログラムのうち変化の激しい16点を単純に選び学習および認識データとしたが、データ作成の際にＤＰマッチング等を利用する方法を検討中である。

終わりに有益な御討論戴いた本学マルチメディアラボの諸氏に謝意を表する。

参考文献

[1] 藤崎, 高村, 佐藤: "特徴量の不均一標本化と単語音声認識への応用", 音響学会音声研究会資料, S77-12(1977-6).