ニューラルネットワークによる口形からの母音認識

5W-9

田村進一，光本浩士，　河合秀夫，　岡崎耕三，副井裕，　黒須顕二
（大阪大学基礎工学部）（大阪電通大）（鳥取大学工学部）（九州工大）

I.　序

従来より我々は，口形画像の画像的特徴と音響的特徴とから音声を認識する研究を行ってきた[1]．本稿では，口形のみであるが，その二次元輪郭形状を直接ニューラルネットワークへ入力し，エラーバックプロパゲーションにより口形母音を学習・認識する実験を行ったので報告する．

II.　実験システム

実験準備の都合上，不十分ではあるが5×5 入力面を用いた．口形は，母音「a」「i」「u」「e」「o」を発生したときのそれを自動[2]または手でトレースした．そして，その輪郭が5×5 の各画素を横切る長さに比例した量を入力とした．

図１には，「a」から「o」までの入力口形を示す．斜線部の大きさが，その画素値を表している．たとえば，「o」の入力画像値を数値的に表すと，図２のようになる．但し，図２では図１の値を{0,1}から{0.1,0.9}に変換してある．

ニューラルネットワークの構成ユニット数は，入力層25個（5×5），中間層10個，出力層5個（「a」から「o」）である．

学習サンプルは，図２等に，[-0.2,0.2]なる一様分布雑音を各画素に加えたものを用いた．学習は，「a」から「o」までを１サイクルとして，140回（サイクル）行った．

認識実験は，学習サンプルと同じく，[-0.2,0.2]なる一様分布雑音が各画素に乗ったものを用いた．誤り率は，事実上０であった．また，[-0.3,0.3]なる付加雑音入力パターンに対しては10%程度の誤認識を生じた．

III.　むすび

実験準備の都合上，荒い画素の画像しか取り扱えなかった．また，自然な口形のvariationではなく，ランダム雑音で代用した．このような問題はあるが，口形の認識を，ニューラルネットワークで行えることを確かめた．

今後は，これらの問題の対策，特に後者の自然な口形の入力（多数の実画像）による識別・認識を行う必要がある．本研究は重点領域の援助を受けた．

図１　口形画像（雑音なし）

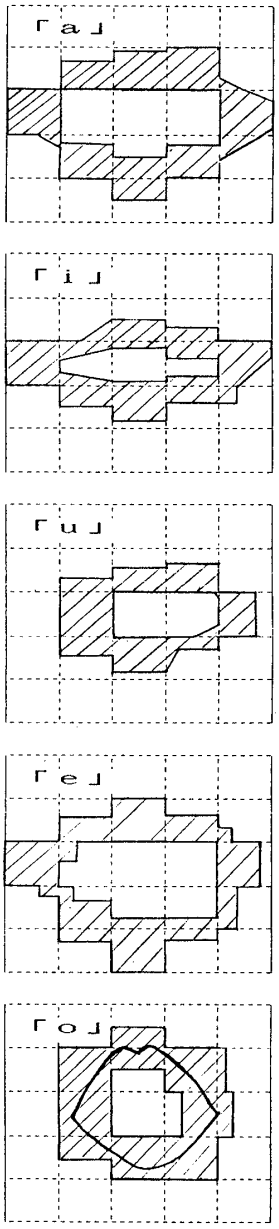

図２　「o」の画素値

0.1	0.1	0.5	0.1	0.1
0.1	0.9	0.5	0.9	0.3
0.1	0.9	0.1	0.3	0.2
0.1	0.5	0.9	0.9	0.1
0.1	0.1	0.1	0.1	0.1

参考文献
[1]田村他，口形の画像情報と音声による単語認識，第30回自動制御連合講演会3062,pp 511-512,昭62.10.
[2]松岡他，画像処理による読唇の試み－母音口形の識別およびそれに基づく単語認識－計測自動制御学会論文集，22巻2号，pp.191-198（1986）

Vowel recognition from mouth shape by using neural netwok
S.Tamura,H.Mitsumoto,H.Kawai,K.Okazaki,Y.Fukui and K.Kurosu