

頸椎損傷患者の口の形の認識について

2B-2

井上 晋¹・井上 倫夫¹・小林 康浩¹・加納 尚之²・井上 公明³¹（鳥取大学工学部）・²（米子高専）・³（鳥取赤十字病院）

1. はじめに

筋萎縮性側索硬化症、頸椎損傷、脳内出血等の患者は、自らの意志を周囲の健常者に伝えることが困難である。このような人の意志伝達を補助できる装置（Communication Aid:CA）があれば、意志疎通の道が開ける。

本報告は、CAへの患者からの合図を、口唇の形状変化から判別する方法について述べる。具体的には、母音発声時（実際には発声不能）の口唇の形状変化をビデオカメラで捉え、3層ニューラルネットワークを用いて識別を行った。

2. 母音（口形）の識別

現在のCAは、スイッチ1つで入力できるように、縦横走査法を採用している。

縦横走査法は、画面上に文字盤を提示しておき、カーソルを『あ・か・さ・・・』と横方向に移動させる。使用者は、自分の入力したい文字のある列にカーソルが来た時に1回目のスイッチ操作を行うと、カーソルはその列を『あ・い・う・・・』のように縦方向に1文字ずつ移動するので、欲しい文字に来たときに2回目のスイッチ操作を行う。この2回のスイッチ操作で文字盤上の1文字を選択する方法である。

口唇の形状変化より母音が識別できるとすると、使用者は文字盤を見ながら、『あ、い、う、え、お』の何れかの口の動きを示すことで文字が選択できる。『す』を選択するには、カーソルが『さ』行に来たときに、『う』と言う動作をする。これにより、カーソルの縦方向の走査

が無くなるため、1回の動作で1文字選択できる。

3. ニューラルネットワークの構造

口の動きを時間変化を追って捉え、かつ高速に処理するために、ビデオ画像より口の部分を32×32ドットで切り出し、1秒間に15枚の画像フレームを処理した。

ネットワークの構造を図1に示す。約1秒間の動きを捉えるため、入力層は16グループに分かれ、1グループは32×32ドットからなるフレームに対応する1024個の入力ノードからなっている。中間層は、入力層の1グループに6ノードとし合計96ノード。出力層は6ノードで『あ、い、う、え、お、開口』を出力する。

中間層のウェイトの初期値は、乱数を用いるのではなく、学習用パターンを用いて設定した。具体的には、各グループの中間層のウェイトは、第1ノードには、『あ』に強く反応するように、『あ』のパターンと、その他の『い、う、え、お』のパターンとの差を。第2ノードには、『い』のパターンと、その他の『あ、う、え、お』のパターンとの差を、といったようにそれぞれ設定した。第6ノードには、口の開きに反応するように、『あ、い、う、え、お』の平均のパターンと、閉じたパターンとの差を設定した。こうすることで、中間層で口の形に、出力層で口の動きに反応するネットワークが作成できた。

How to Discriminate Shapes of Spinal Cord Injured Patient's Mouth

Susumu Inoue¹, Michio Inoue¹, Yasuhiro Kobayashi¹, Naoyuki Kanoh², Kimiaki Inoue³

1:Tottori University, 2:Yonago national College of Technology, 3:Tottori Red Cross hospital

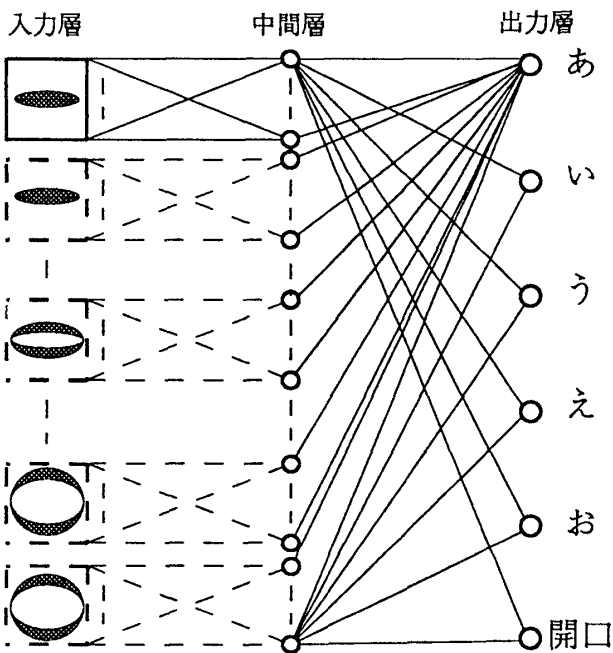


図1 ネットワーク構造

フレーム番号	教師信号
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	
---	ON
---	OFF
---	ON
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	OFF
---	OFF

図2 学習パターン

4. 学習と判定

学習は、出力層のウェイトとオフセット値のみ行った。基本となる16枚の『あ、い、う、え、お』の連続画像フレームを取り込み、これをONのパターンとし、このパターンを用いて、OFFのパターンを作成した。図2に作成した学習パターンの『あ』についてのパターンを示す。『い、う、え、お』についても同様にパターンを作成した。学習パターンは、計21種類である。

出力層の6番目のノードは、口唇の開閉に反応するノードである。判定の際には、この6番目のノードと、その他の『あ、い、う、え、お』に反応するノードが、同時にONになった場所をそれぞれ『あ、い、う、え、お』と判定する。

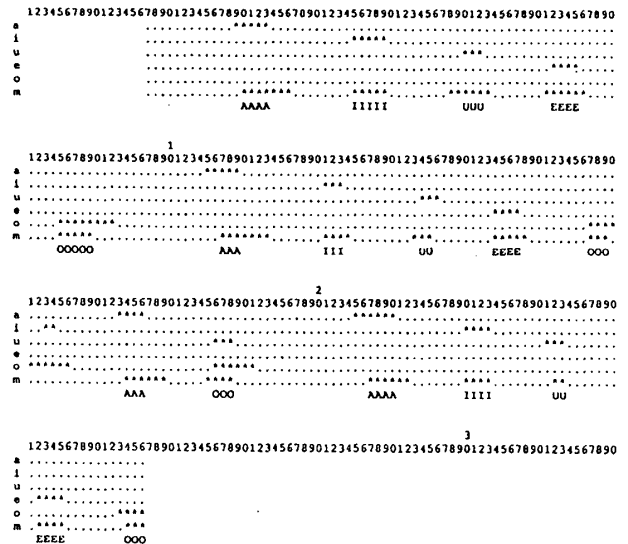
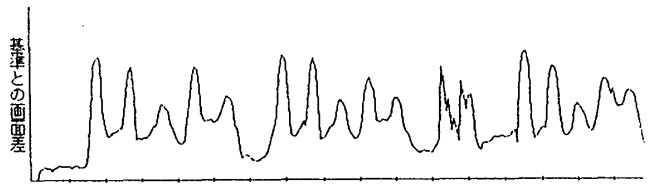


図3 判定結果

はっきりと『あ、い、う、え、お』と言った連続画像を400回学習させた学習効果を図3に示す。上のグラフは、先頭の画面との画面差のグラフで、口の動きを示している。例えば、先頭画面から約25枚目で最初の口の動きがあったことが解る。下のグラフは6個の出力層の出力と、『あ、い、う、え、お』の判定結果を示している。この結果より、ほぼ正しく判定できていることが解る。170~190枚目は、『あいうえお』と早口で言った部分で、『あ』と『お』しか判定できなかった。

6. 終わりに

3層ニューラルネットワークを用いることにより、口の形のみで、母音の識別が可能であることが解った。現在は実際にCAの入力装置として使用できるように検討中である。

[参考文献]

[1] 井上, 江口他: “頸椎損傷患者の口の開閉の検出について” 第48回情報処理学会全国大会論文集 5T-8, 1993