2—P—11

時間遅れ神経回路網による音節スポッティングの検討*

○沢井 秀文　アレックス・ワイベル　鹿野 清宏
(ATR自動翻訳電話研究所)

1. はじめに

日本語の単語や連続音声認識を行うためにCV音節をスポッティングする方法がある。

我々は先に時間遅れ神経回路網(Time-Delay Neural Network:TDNN)が非常に高い認識率(bdgタスクで98.6%)が得られることを示した[1]。本稿では、TDNNを音節スポッティングに適用した検討結果を示す。

従来、音節(Demisyllable)のスポッティングをニューラルネットを用いて行った実験として、"ド"、"レ"、"ミ"、"ファ"、"ソ"、"ラ"、"シ"の7音節の実験[2]があるが、カテゴリー数が限られている。日本語の約100音節のスポッティングに適用するためには、多くのカテゴリーを扱えるニューラルネットを如何に構成するかの問題点がある。

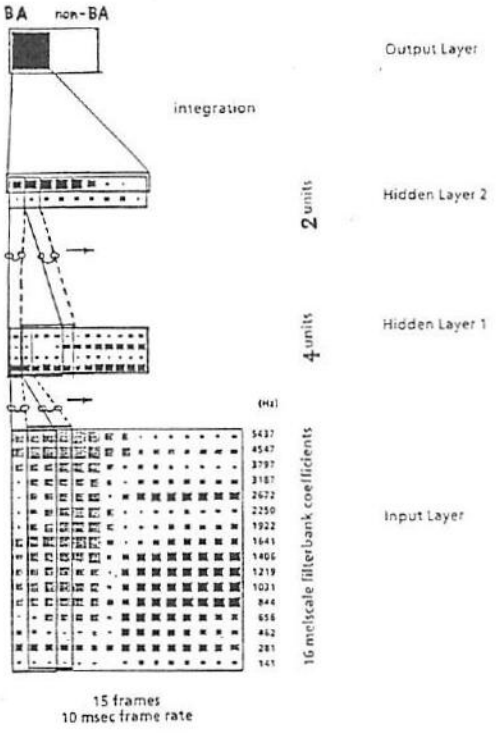

図1　音節スポッティング用TDNN

我々は、日本語音声中のCV音節をスポッティングするためのニューラルネットワークとして、ある音節とその音節以外とを識別するものを構成する。もし、このようなネットワークが学習でき、精度良く音節をスポッティングすることが可能となれば、音節数だけのネットワークを用意することにより、原理的に全ての音節をスポッティングすることが可能となる。しかし、ある音節以外の学習データを如何に選択し、ニューラルネットに学習させるかは一つの大きな問題である。

今回は、日本語の単音節の一例として"BA"を取り上げる。"BA"以外の音節としては約100音節存在するが、全てを使用することは学習上の効率から好ましくないので、"BA"とコンヒュージョンを起こし易いと考えられる"DA","GA","PA","TA","KA"の5音節を使用する。

2. 音節スポッティング用TDNNの学習

図1に、音節スポッティング用のTDNNを示す。学習用のサンプルとしては、重要語5240単語中の半数から"BA"を含む単語53語を抽出し、"BA"の部分15フレーム(10ms周期)を切り出した。尚、特徴パラメータは16次のFFTメルスペクトラムである。

"BA"以外の音節として、"DA","GA","PA","TA","KA"を含む単語の該当音節部分15フレームを同様に切り出した。学習サンプル数は全部で1014音節である。学習は、Back-Propagation学習則[3]に従って行った。

3. 音節スポッティング実験

図2に音節スポッティング実験の様子を示す。未知入力音声中を、図1のTDNNを3フレームずつシフトしながらスキャンした。教師データの与え方としては、"B"と"A"の境界と、図1のTDNNの中心とのずれが一定時間内(30msないし50ms)のときに"BA"とした。但し、境界の曖昧な部分は評価対象から除外した。未知入力音

*A Preliminary Study on Spotting Japanese CV-Syllables by Time-Delay Neural Networks
By Hidefumi Sawai, Alex Waibel and Kiyohiro Shikano
(ATR Interpreting Telephony Research Laboratories)

声の部分が"BA"であるか"non-BA"であるかの判定は、出力ユニットの値0≦o(BA),o(non-BA)≦1を用い、次の判定条件に従って決定した。

<判定条件>

1. o(BA) >o(non-BA) なら"BA"と判定
2. o(non-BA) >o(BA) なら"non-BA"と判定

表1と表2にスポッティングの実験結果を示す。評価用の単語としては、学習用の単語とは異なる"BA"を含む61単語を用いた。表3に一覧表を示す。"BA"以外の音節は138音節ある。3フレームずつシフトした時のサンプル数は、"BA"が156個、"non-BA"が1018個である。音節サンプルの同定率は"BA"が83.3%、"non-BA"が98.7%である。また音節単位では、"BA"は95.1%の確率で同定でき、"non-BA"は99.3%の確率で抑圧できた。音節単位での誤りは語頭の/a/(ashiba)、語中の /ba/ (keibatsu, shibaraku, jibaN)で生じた。

4. まとめ

日本語の音節スポッティングの検討をTDNNを用いて行った。ある音節とそれ以外の音節を識別できるように学習したニューラルネットを用いてある音節(例:"BA")を95%識別でき、他の音節を99.3%で抑圧できることを示した。これにより、他の音節検出用ニューラルネットを用意しておけば、任意の音節のスポッティングができる可能性を示した。

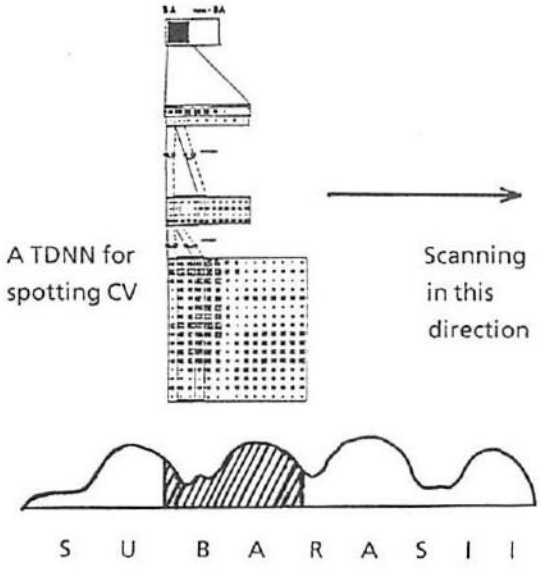

図 2. 単音節のスポッティング実験

表 1. CV スポッティング結果(音節単位)

音節名	BA	non-BA	合計
音節数	61	138	199
音節同定率	58/61 (95.1%)	137/138 (99.3%)	195/199 (98.0%)

表2. CV スポッティング結果(サンプル単位)

音節名	BA	non-BA	合計
サンプル数	156	1018	1174
音節サンプル同定率	130/156 (83.3%)	1005 /1018 (98.7%)	1135 /1174 (96.7%)

表 3. "BA"を含む単語リスト(評価用)

ashiba arubamu ichiba iwaba ubau oohaba oba obasaN kabaa kabaN gaNbaru kubaru keibatsu ketobasu geNba kouba koubaN kokubaN kobamu koNbaN saibaN sakubaN satsubatsu sabaku shibaraku shoubai shokuba jibaN juNbaN subarashii sokubaku soba tachiba tatoeba taba tabaneru tsubaki tsubame tebanasu toubaN nobasu habakaru bai baikai baishuu bakugeki bakudai bakuhatsu baketsu basho bachi batsu bameN baN baNbumi baNchi hyoubaN mabara meNbaa yakuba yabaN

[文献]

[1]Alex Waibel : "時間遅れ神経回路網による音韻認識 、音声研資料SP87-100 (1987.12).

[2]C.Kamm et al : "Training an Adaptive Network to Spot Demisyllables in Continuous Speech." ATR Work Shop on Neural Networks and Pararel Distributed Processing(July 1988).

[3]D.E.Rumelhart et al. PDP, MIT Press (1986).