TR-I-0183

発話変動にロバストなTDNNの検討

南 泰浩† 沢井 秀文
Yasuhiro MINAMI † Hidefumi SAWAI

1990.10.5.

概要

本報告では入力の変動に対する出力の変動を小さくする手法をTDNNに応用し、TDNNが発話様式の異なる発声中の音素に対してどのような効果があるかを調べた。この結果、入力の変動に対する出力の変動を小さくする手法は学習に非常に多くの時間を必要とし、パラメータの設定も非常に難しいことが確認された。

また、TDNNの入力層のウィンドウフレーム数を変化させたときの認識率を調べた。ウィンドウフレーム数を7フレームにし、出力層と第2隠れ層との間の重み係数を固定したとき、文節区切りを指定しない連続発声中の18子音に対して、従来型のTDNNに比べ6.5%の音素認識率の向上がみられた。

本報告は、学外実習生南泰浩(慶応大学)が行った実習の報告書である。

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

† Keio Univ.
† 慶応義塾大学

目次

1. はじめに

ＡＴＲでは従来より、ＴＤＮＮ（Ｔｉｍｅ-Ｄｅｌａｙ Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋｓ）を用いた音声認識を行ってきた(1)。ＴＤＮＮは特定話者の/ｂｄｇ/のタスクに対して９８．６％と高精度の音素認識率を達成することが確認された(1)。しかし、ＴＤＮＮを用いた連続音声認識システムを構築する過程で様々な問題が生じてきた。我々が構成したシステムは従来ＡＴＲにおいて構成した連続音声認識用システムのＨＭＭ（Ｈｉｄｄｅｎ Ｍａｒｋｏｖ Ｍｏｄｅｌ）－ＬＲの音声処理部（ＨＭＭ部）をＴＤＮＮの処理部に置きかえたものである(2)。このシステムを用いて連続音声認識を行った結果、ＨＭＭ－ＬＲに比べ文節認識率が低いことが確認された。この原因はＴＤＮＮの連続音声に対する音素認識率の低下であると考えられる。ＴＤＮＮが単語発声データにのみ特殊化され、他の発話様式に対応できないためであると考えられる。

そこで本報告では文献(3)で提案されている、汎化のアルゴリズムをTDNNに応用し、ロバストなTDNNの検討を行なった。この手法は入力の変動による出力の変動を少なくする手法である。しかし、この手法はパラメータの設定が非常に難しく、収束時間も多くの時間を必要とする。またTDNNでは重みに対してシフトレラントにするために重みに拘束をつけているため、今回示したアルゴリズムのようにさらに重みの拘束をつけるとうまく収束しないことが確認された。

また、３章では文献(4)で述べたようにＴＤＮＮの構造によるシフトトレラント性について調べ、ＴＤＮＮの改良の可能性を調べる。今回は入力のウィンドウのフレーム幅の変化について調べた。この結果入力のウィンドウの幅を７フレームにし、出力層と隠れ層第二の重みを固定した18子音用のＴＤＮＮは、句切りを指定しない発声において従来法のＴＤＮＮに比べ音素認識率で6.5％の向上を示した。

2. 出力の平滑化を行うＴＤＮＮ

基本的なアルゴリズムは入力の変動に対する出力の変動をできるだけ小さくするということである。入力に対する出力の変動を小さくするということは出力の関数をできるだけ滑らかにする方法である。この方法には２種類の手法が考えられる。一つは与えられたサンプル点の周りのみを滑らかにする手法である。この方法はサンプル点にノイズを加え、近傍の周りもサンプル点と同じ様な出力値をとるようにする。もう一つは出力を滑らかにするための評価関数を従来のエネルギー関数に加え、この値を従来のバックプロパゲーションの改良により小さくするように学習を行なう。しかし、この評価関数は陽に計算できないため、この評価関数よりいつも大きいか等しい関数を用いて、評価関数を極小化する手法を用いる。以下にこの２つの手法を示す。

[方法１]

この方法は学習パターンx_p (p＝１…Ｐ）に対してのみ入力に対する出力の変動を小さくする。入力に対する出力の変動を$\|\partial F(x_p)/\partial x_p\|^2$のように定義する。ここでＦはニューラルネットの出力ベクトルを示す。また$\|\ \|$は行列ノルムをしめし、Ｙは行列とし、各要素をy_{ij}とすると以下のように定義される。

$$\|Y\| = \sqrt{\sum_{i,j} y_{i,j}^2}$$

この場合は極小化する関数は

$$Q(W^1,\cdots,W^L) = (1/2)<|F(x_p) - op|^2> + (1/2)\,\varepsilon<\|\partial F(xp)/\partial xp\|^2>$$
$$= (1/2)<|F(x_p) - o_p|^2> + (1/2)\,\varepsilon<|(\partial F(x_p)/\partial x_p)\,d|^2>$$
$$\approx (1/2)<|F(x_p+d) - o_p|^2>$$

となる。＜・＞は確率変数・の期待値を表す。さらにｄは＜ｄ＞＝０、$<d\,d^T> = \varepsilon I$となるベクトルであり、またW^iはi-1層からi層への重みを表す。この評価関数は入力に雑音ｄを加えた出力値を従来の出力値と置き換えたものである。

[方法２]

入力空間の全てのｘに対して入力に対する出力の変動を小さくする。

$Q(W^1,\cdots,W^L) = (1/2) < | F(x_p) - op |^2 > + (1/2)\ \varepsilon < \| \partial F(xp) / \partial xp \|^2 >$

この場合は以上の評価関数をバックプロパゲーションにより極小化すればよいが、$\| \partial F(xp) / \partial xp \|^2$は通常陽に計算できない。そこで、この評価関数よりいつも大きいか等しい関数を用いて、評価関数を極小化する手法を用いる。

ここでx^LはＬ層への入力とし、$S(x^L)$はニューラルネットのＬ層への入力x^Lにシグモイド関数を通す関数とする。

ゆえに　$F(x) = S(x^L)$

$x^L = W^L S(x^{L-1})$

$x^1 = x$である。

$$\frac{\partial F(x^L)}{\partial x^1} = \frac{\partial S^L(x^L)}{\partial x^1} = \frac{\partial S(x^L)}{\partial x^L}\frac{\partial x^L}{\partial x^1} = \frac{\partial S(x^L)}{\partial x^L} W^L \frac{\partial S(x^{L-1})}{\partial x^1}$$

$$G^L(x^L) = \frac{\partial S(x^L)}{\partial x^L}$$

と置くと

$$\frac{\partial F(x)}{\partial x} = G^L(x^L)W^L \cdots G^1(x^1)W^1$$

となる。

ゆえに

$\| \partial F(x) / \partial x \|^2$

$= \| G(x^L)\ W^L \cdots G(x^1)\ W^1 \|^2$

通常　$A = (a_1, a_2, \cdots, a_L)$　$B^T = (b_1, b_2, \cdots, b_L)$とすると、各要素はベクトルである。

$\| A \|^2 = (a_1 \cdot a_1)^2 + (a_1 \cdot b_2)^2 + \cdots + (a_L \cdot b_L)^2$

$\| A \|^2 \| B \|^2 = | a_1 |^2 | b_1 |^2 + \cdots + | a_L | \ | b_L |$

ここで$a_1 \cdot b_1 = | a_1 | \ | b_1 | \cos(\Psi)$

ここでΨはa_1とb_1のなす角である。

これより

$| a_1 \cdot b_1 |^2 \leqq | a_1 |^2 | b_1 |^2$

となり

$\| AB \|^2 \leqq \| A \|^2 \| B \|^2$

が成り立つ。

よって

$\| \partial F(x) / \partial x \|^2$

$= \| G(x^L)\ W^L \cdots G(x^1)\ W^1 \|^2$

$\leqq \| G(x^L)\ W^L \|^2 \cdots \| G(x^1)\ W^1 \|^2$

さらにＧ（x^L）はx^Lにシグモイドの微分関数を通したものを対角要素にもつ行列であるから、任意のx^Lに対してある適当な値ｍによって以下のようにおさえられる。

$\| G(x^L)\ W^L \|^2 \leqq m$

よって以上に示した式は以下のようにおさえられる。

$\| \partial F(x) / \partial x \|^2$

$= \| G(x^L) W^L \cdots G(x^1) W^1 \|^2$
$\leq \| G(x^L) W^L \|^2 \cdots \| G(x^1) W^1 \|^2$
$\leq M \| W^L \|^2 \cdots \| W^1 \|^2$

以上より、評価関数を以下のように設定すればもともとの評価関数を上からおさえられる。
新たな評価関数は
$Q(WL,\cdots,W1) = (1/2) < | F(xp) - op |^2 > + (1/2) \delta \| W^L \|^2 \cdots \| W^1 \|^2$
となる。ここでδはϵMである

[学習アルゴリズム]
学習アルゴリズムは通常のバックプロパゲーションアルゴリズムにいくらかの改良を加えたようになる。これを以下に示す。

方法1

従来のバックプロパゲーションにノイズを加えた学習を行う。

方法2

$W^k(n+1)$（k-1層からk層へ接続する学習n回めの重み）に対しては、従来のバックプロパゲーション学習則に加え以下のような項を加える。

$$-\delta \| W^1(n) \|^2 \cdots \| W^{k-1}(n) \|^2 \cdot \| W^{k+1}(n) \|^2 W^k(n)$$

2.1 実験環境

まず、はじめに認識を行なう環境を以下に説明する。

2.1.1 認識対象音素

認識データとしては発声様式の異なった以下の4種類のデータから切り出した音素を用いる。認識に用いた音素の数は全ての手法で同じ数とした。

（1）５２４０単語の奇数番目のデータ（以下単語発話とする）（５．６８モーラ／ｓｅｃ）
（2）複合語を許さない文節発話（以下短い文節発話とする）（７．１４モーラ／ｓｅｃ）
（3）句切り指定を行わない発話（以下連続発話とする）（９．５６モーラ／ｓｅｃ）

2.1.2 学習音素

学習音素は重要５２４０単語の偶数番目のデータから切り出した音素を用いた。単語発話から切り出したデータだけで、どれだけ連続音声認識に対して有効であるかを調べるため学習データには文節発声や自由発声データを用いていない。また学習データ数は各子音２００個とした。

2.1.3 発声話者

発声はアナウンサ(MAU)1名の発声した音声で特定話者の認識を行った。

2.1.4 認識率の計算

ここで求めた認識率は各子音毎に認識率を求め、その値を各子音の個数に対応する重みを掛けて総和したものを全体の認識率とした。

2.1.5 ＴＤＮＮにおける分析

ＴＤＮＮで用いる1フレームの音声データは１２ＫＨｚでサンプリングし、２５６ポイントのハミング窓で、ＦＦＴを計算した後、１０ｍｓ毎に１６次のメル尺度のフィルタバンクを通し、１５フレーム内で平均0、±1に正規化したものである。

2.2 認識実験

実験には図1に示すような/b，d，g，m，n，N/の認識を行うネットワークを用いた。表1にノ

イズを加えた方法１の実験結果を示す。加えたノイズは各入力ユニットに対して［－δ、δ］の間の一様分布のノイズである。表２は方法２を行った結果である。また学習にはＤＣＰ２を用いた(5)。

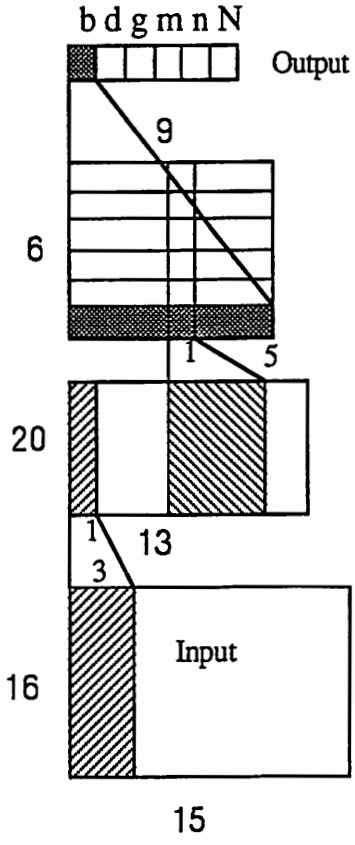

図1　出力の変動を小さくする実験で用いた
TDNN(/b,d,g,m,n,N/)

[検討]

方法１も方法２もどちらの場合もＴＤＮＮでは収束に非常に多くの時間を必要とした。しかも、どちらの場合も学習データに対しても認識率がある一定の値で抑えられてしまう。

これは方法１の場合には各学習毎にノイズを加えているため重みの変化の方向が絶えずゆらいでしまうためと考えられる。このような問題の改善方法として学習をある長時間に渡って平均するようにすればよい。しかし、このようにノイズの付加によって収束をしないというのは、TDNNによって構成される境界面とサンプルデータの間に十分に余裕がないからであるとも考えられる。

また方法２ではＴＤＮＮの拘束条件に加えさらに拘束条件を加えているため学習データに対しても認識率が上昇しないと考えられる。

以上のようなことから考えてこのような手法がうまく行かないというよりむしろTDNNがサンプルの近傍に境界面を構成するのではないかと考えられる。

３．ＴＤＮＮのウィンドウ幅による認識率

２．で示した結果からTDNNの入力層のウィンドウが３フレームであるために、３フレーム分の特徴だけを用いて境界の判別を行なっている可能性があるとも考えられる。そこで、ここでは入力ウィンドウのフレーム数の変化にともない認識率がどのように変わるかを調べた。

３．１　/b、d、g、m、n、N/の認識

入力のウィンドウの違いによる認識率の違いを調べるために次のような３つのネットワークによる認識率の差を調べる。これを図２(1)～(3)に示す。この図は入力のウィンドウを３フレーム、５フレーム、７フレームと増やしたものである。実験環境は２．とまったく同じである。実験としては以下の２つの実験を行った。（２）の手法は以前[4]の文献で音素認識率に対しての向上がみられた。

（１）通常のＴＤＮＮ（学習にはDynet[6]を使用）

（２）出力層と隠れ層第２の結合を1.0に固定し、しきい値も0.0に固定する。（学習にはDCP2[5]を使用）

図２の３種のTDNNに対して（１）と（２）の両方の実験を行なった。

結果を表３～表８に示す。この表は各発話様式に対する音素認識率である。横方向の数字は累積の認識率を示す。また、表９～表１４までは各TDNNの短い文節発話に対するシフトトレラント性の結果を示した。縦方向の数字はサンプルの中心が何msecずれているかを示し、横方向は累積の認識率を示す。

[検討]

表３～表８から、重みを固定したものはどの場合も連続発話に対して、固定しないものに比べよい認識率を示す。これは重みを固定することによって入力サンプルの複数の特徴に注目するようになるためと考えられる。

入力のウィンドウのフレーム数は３フレームより５フレーム、７フレームの方がよい認識率を示した。しかし、この認識率は音素毎にかなりのばらつきがあり、各音素毎に適したフレーム数が存在するようである。

この実験の中で特に、入力ウィンドウを７フレームにし、重みを固定したTDNNは、連続発話のデータに対して非常によい認識率を示した。

表９～表１４から、入力のウィンドウのフレーム数を変化させてもシフトトレラント性の劣化はみられないことがわかる。また、重みを固定したものは固定しないものに比べシフトトレラント性が向上していることがわかる。

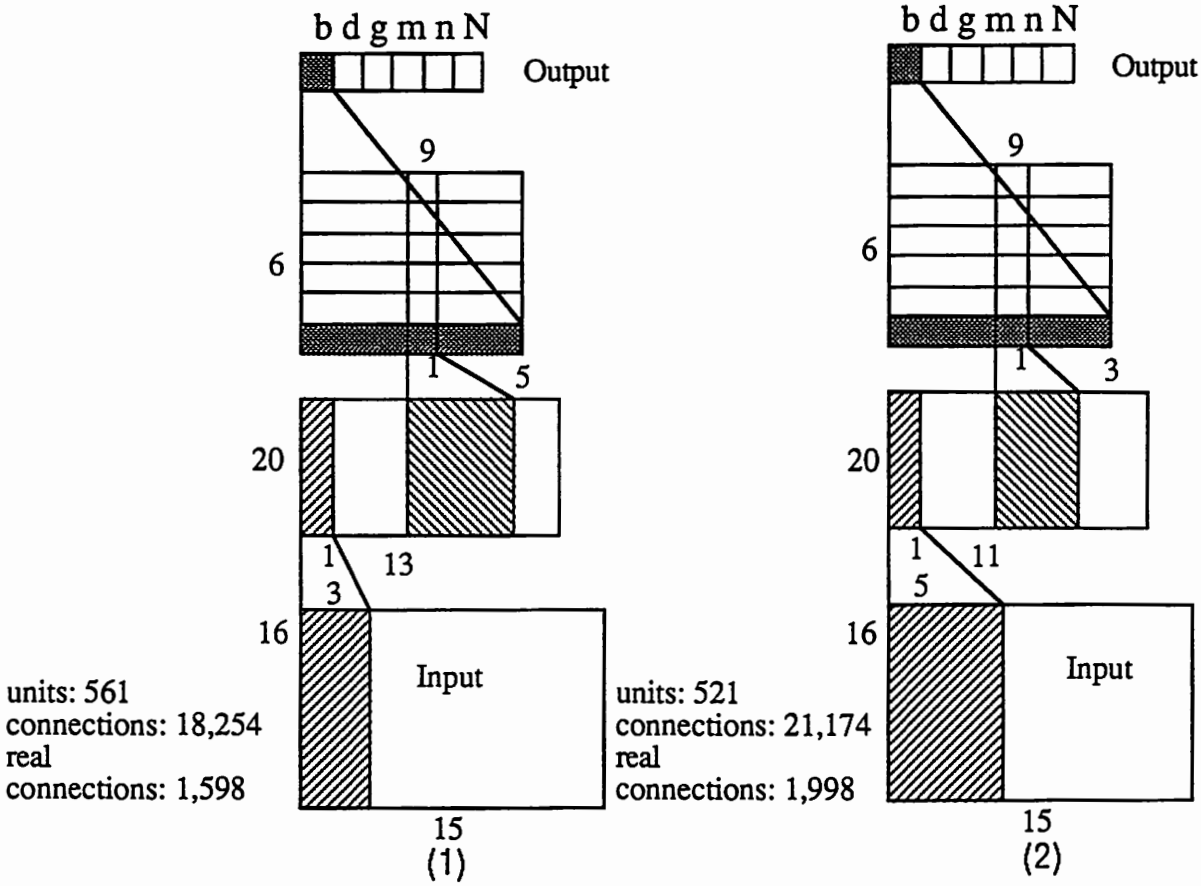

図2　実験で用いた入力ウィンドウの違う TDNN(/b,d,g,m,n,N/)

３．２　１８子音の認識率

１８子音では３．１の実験結果を踏まえて、図３、図４のようなネットワークを用いた。図３は従来型のTDNNである(7)。図４に関しては３．１と同様に出力層と隠れ層第二の間の重みを固定することも実験した。出力層と隠れ層第二の間の重みを固定する場合には学習にＤＣＰ２(5)を使い、他はDynet(6)を使った。

１８子音は/b,d,g,p,t,k,m,n,N,s,s h,h,z,c h,t s,r,w,y/であり音素区間は以下の条件を満たすもののみを用いた。

（1）/c h/は/c h i/のみから切り出す。

（2）/s h/は/s h i/のみから切り出す。

（3）/z/は/z a/, /z e/, /z u/, /z o/のみから切り出す。

（4）/h/は/h a/, /h i/, /h e/, /h o/のみから切り出す。

（5）その他はラベルに従いその音素の存在する区間を切り出す。以上のラベルの表記はＡＴＲデータベースの表記法(8)に従い、音素はこのラベルにより切り出したものである。その他の条件は２．で示したものと同じである。

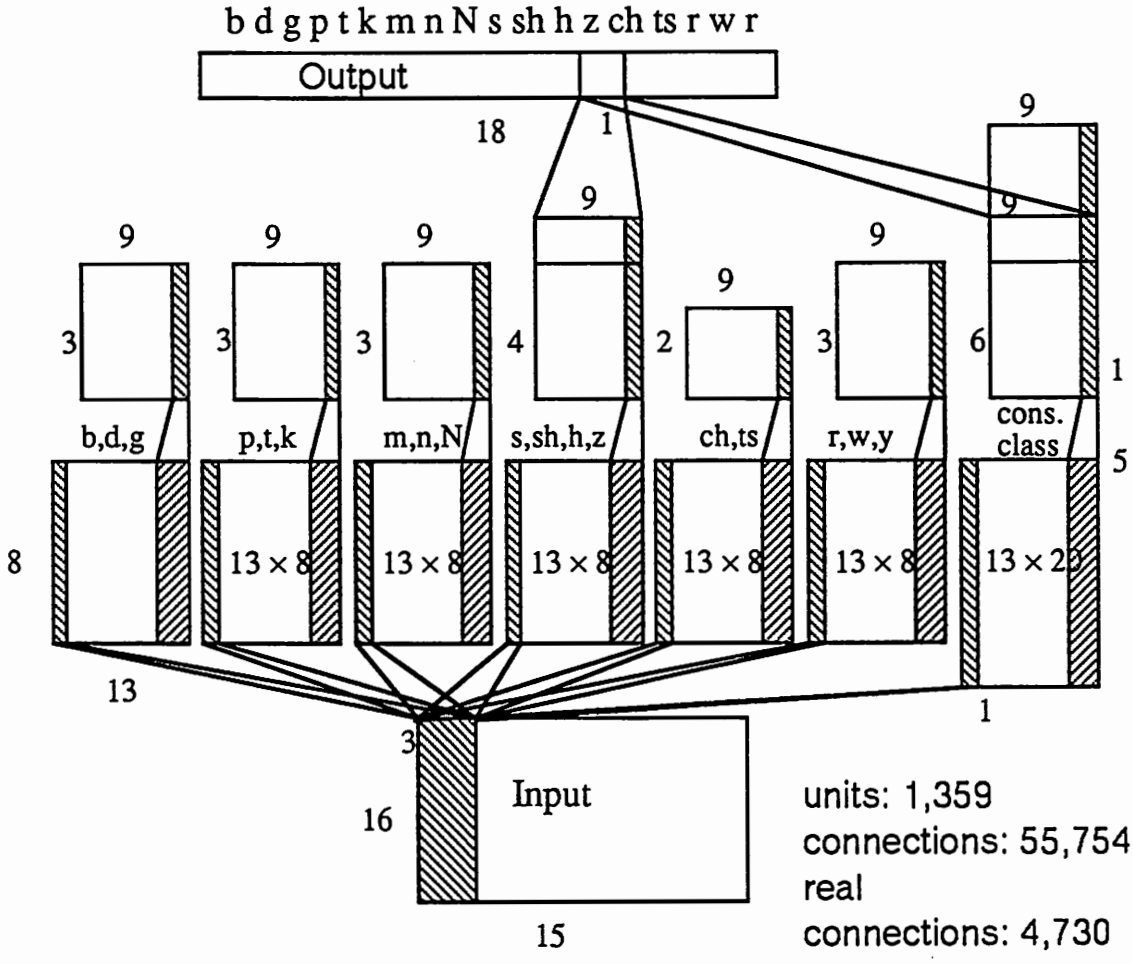

図3　従来型のTDNN(18 子音)

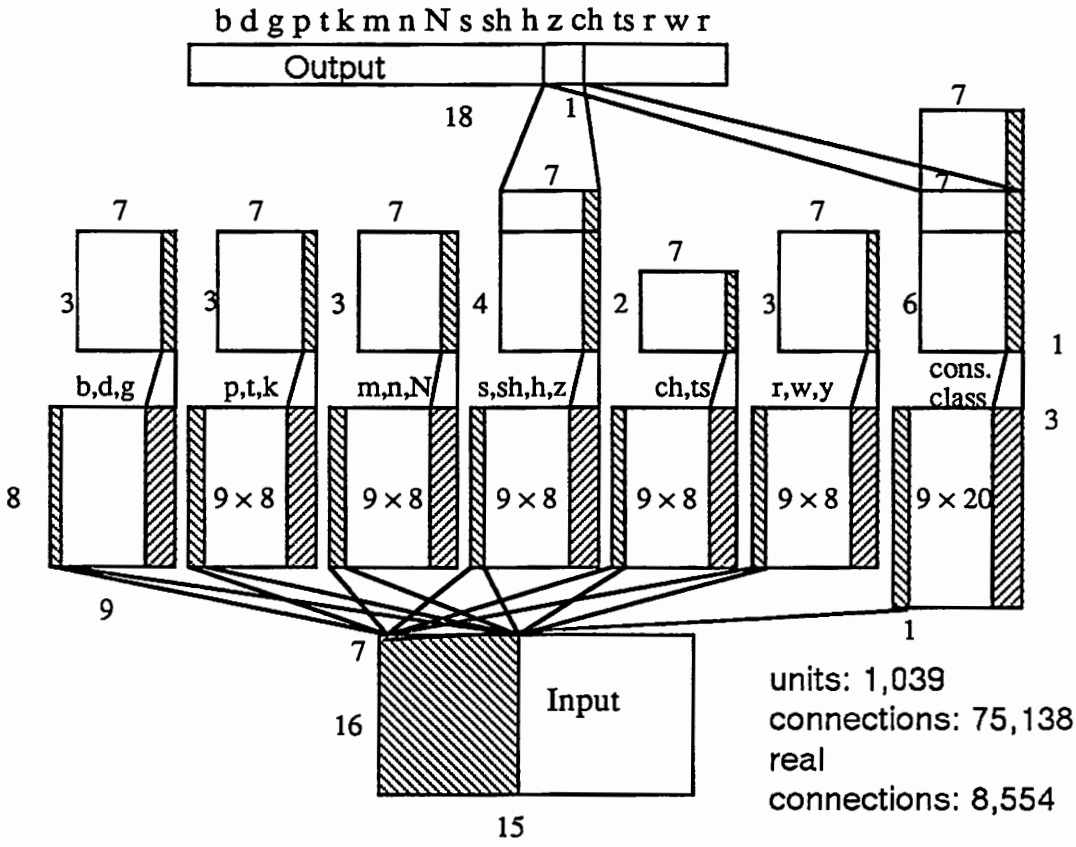

図4　入力ウィンドウが7フレームのTDNN(18 子音)

結果を表１５～表２０に示す。表１５～表１７までは各発話様式に対する音素認識率である。横方向の数字は累積の認識率を示す。また、表１８～表２０までは各TDNNの短い文節発話に対するシフトトレラント性の結果を示した。縦方向の数字はサンプルの中心が何msecずれているかを示し、横方向は累積の認識率を示す。

[検討]

表１５～表１７から、入力のウィンドウを７フレームにしたもの方が３フレームのものより若干認識率が高いことがわかる。さらに、重みを固定することにより認識率の向上がみられた。短い文節発話、連続発話に対して、従来のTDNNに比べそれぞれ、６.４％、６.５％の認識率の向上がみられた。また表１８～表２０より、入力ウィンドウを７フレームにしてもシフトトレラント性は悪くならないことがわかる。また重みを固定した場合は非常によいシフトトレラント性を示す。

まとめ

今回、発話の変動に強いTDNNを得るために各種の実験を行なった。　　入力の変動に対する出力の変化を小さくする項を評価関数に加えた方法は、TDNNに用いた場合収束に時間がかかり、パラメータの設定が難しいことが確認された。また、入力ウィンドウのフレーム数を変える手法は、出力層と隠れ層第二の間の重みを固定する手法と組合わせることにより発話変動に非常に強くなることが確認された。

謝辞

本研究の機会を与えて頂いたＡＴＲ自動翻訳電話研究所樽松明社長に深謝致します。また、熱心に御討論頂いた音声情報処理研究室室長の嵯峨山茂樹氏、杉山雅英氏、阿部匡伸氏をはじめとする音声情報処理研究室の皆様に感謝致します。

参考文献

(1) Ａ．ワイベル：”時間遅れ神経回路網（ＴＤＮＮ）による音韻認識”、信学会、信学技法 SP87-100(1987.12)

(2) 南泰浩、宮武正典、沢井秀文、鹿野清宏：”ＴＤＮＮ音韻スポッティングと拡張LRパーザを用いた文節音声認識”,音響講論集　３－１－１１(1989.10).

(3) 松岡清利：”誤差逆伝搬法の汎化問題に対する一手法”、電子情報通信学会論文誌 Vol. J 73 D-II No.6 pp. 897-904,1990.

(4) 南泰浩、沢井秀文：”ＴＤＮＮの構造の音韻認識率、シフトインバリアント性への影響”、ATR Tech. Report TR-I-0145(1990.2).

(5) 中村雅巳、鹿野清宏：”ニューラルネットにおけるバックプロパゲーション学習の効率化方法” ATR Tech. Report TR-I-0119(1989.10).

(6) P.Haffner,H.Sawai,A.Waibel and K.Shikano：” Fast Back-Propagation Learning Methods for Large Phonemic Neural Networks” 、音響講論集1-6-14(1989.3)または P.Haffner,A.Waibel,H.Sawai and K.Shikano:"Fast Back-Propagation Learning Methods for Neural Networks in Speech" ATR Tech. Report TR-I-0058(1988.11).

(7)A. Waibel, H. Sawai and K. Shikano,"Consonant Recognition by Modular Construction of Large Phonemic Time-Delay Neural Networks ",ICASSP'89,pp.112-115(1989).

(8) 武田一哉、匂坂芳典、片桐滋、桑原尚夫：”研究用日本語音声データベースの構築”、音響学会誌, 44,10,pp.747-754(昭63.10).

表1　入力にノイズを加えた場合の学習データに対する認識率(方法1)

ノイズの範囲	学習データに対する認識率
δ =0.02	93%程度
δ =0.01	93%程度

表2　入力に対する出力の変動を小さくする評価関数を用いた場合の各データに対する認識率(方法2)

手法	学習データ	単語発話	短い文節発話
通常のＴＤＮＮ	99.9%	95.6%	75.7%
δ =0.00001	93.0%	———	———
δ =0.000005	94.1%	91.5%	73.9%
δ =0.0000003	97.0%	93.7%	73.7%

表3　各発話様式に対する音韻認識率(/b,d,g,m,n,N/)
条件(入力3フレームウィンドウ)

発話様式	1位	2位まで	3位まで
単語発話	95.1%	98.6%	99.7%
短い文節発話	74.8%	88.9%	95.0%
連続発話	64.7%	83.2%	90.4%

表4　各発話様式に対する音韻認識率(/b,d,g,m,n,N/)
条件(入力3フレームウィンドウ、出力層と隠れ層第2の重み固定)

発話様式	1位	2位まで	3位まで
単語発話	95.2%	98.6%	99.2%
短い文節発話	77.6%	90.6%	95.5%
連続発話	69.8%	88.4%	94.7%

表5　各発話様式に対する音韻認識率(/b,d,g,m,n,N/)
条件(入力5フレームウィンドウ)

発話様式	1位	2位まで	3位まで
単語発話	95.9%	98.9%	99.5%
短い文節発話	78.2%	90.8%	94.8%
連続発話	69.3%	86.1%	91.4%

表6　各発話様式に対する音韻認識率(/b,d,g,m,n,N/)
条件(入力5フレームウィンドウ、出力層と隠れ層第2の重み固定)

発話様式	1位	2位まで	3位まで
単語発話	95.7%	99.1%	99.7%
短い文節発話	80.5%	93.4%	97.1%
連続発話	67.2%	86.4%	93.3%

表7　各発話様式に対する音韻認識率(/b,d,g,m,n,N/)
条件(入力7フレームウィンドウ)

発話様式	1位	2位まで	3位まで
単語発話	96.4%	99.3%	99.7%
短い文節発話	80.0%	91.3%	95.1%
連続発話	68.7%	85.5%	92.1%

表8　各発話様式に対する音韻認識率(/b,d,g,m,n,N/)
条件(入力7フレームウィンドウ、出力層と隠れ層第2の重み固定)

発話様式	1位	2位まで	3位まで
単語発話	95.7%	99.1%	99.7%
短い文節発話	79.1%	92.7%	97.4%
連続発話	74.5%	88.8%	95.6%

表9　短い文節発話に対する各TDNNのシフトトレラント性(/b,d,g,m,n,N/)
条件(入力3フレームウィンドウ)

シフト幅（ｍｓ）	1位	2位まで	3位まで
-20	60.3%	82.8%	92.6%
-10	72.5%	88.5%	94.7%
0	75.7%	89.3%	95.2%
10	74.1%	88.7%	94.4%
20	69.2%	86.1%	93.2%

表10　短い文節発話に対する各TDNNのシフトトレラント性(/b,d,g,m,n,N/)
条件(入力3フレームウィンドウ、出力層と隠れ層第2の重み固定)

シフト幅（ｍｓ）	1位	2位まで	3位まで
-20	68.1%	85.7%	92.9%
-10	73.3%	89.7%	94.4%
0	77.3%	90.7%	95.4%
10	76.8%	91.3%	95.5%
20	75.0%	89.2%	94.9%

表11　短い文節発話に対する各TDNNのシフトトレラント性(/b,d,g,m,n,N/)
条件(入力5フレームウィンドウ、出力層と隠れ層第2の重み固定)

シフト幅（ｍｓ）	1位	2位まで	3位まで
-20	62.4%	84.1%	90.5%
-10	73.5%	88.2%	93.5%
0	78.2%	90.8%	94.8%
10	77.8%	90.1%	94.9%
20	74.8%	89.1%	94.4%

表12　短い文節発話に対する各TDNNのシフトトレラント性(/b,d,g,m,n,N/)
条件(入力5フレームウィンドウ、出力層と隠れ層第2の重み固定)

シフト幅（ｍｓ）	1位	2位まで	3位まで
-20	70.2%	89.4%	95.2%
-10	77.6%	91.6%	96.6%
0	80.5%	93.4%	97.1%
10	79.8%	93.4%	97.4%
20	77.2%	91.8%	96.4%

表13　短い文節発話に対する各TDNNのシフトトレラント性(/b,d,g,m,n,N/)条件(入力7フレームウィンドウ)

シフト幅（ｍｓ）	1位	2位まで	3位まで
-20	66.4%	86.4%	93.9%
-10	77.6%	92.0%	96.3%
0	80.0%	91.3%	95.2%
10	78.9%	90.9%	95.1%
20	67.1%	86.7%	92.4%

表14　短い文節発話に対する各TDNNのシフトトレラント性(/b,d,g,m,n,N/)条件(入力7フレームウィンドウ、出力層と隠れ層第2の重み固定)

シフト幅（ｍｓ）	1位	2位まで	3位まで
-20	71.6%	90.9%	95.9%
-10	78.7%	93.6%	97.8%
0	79.1%	92.7%	97.4%
10	78.8%	92.4%	97.6%
20	73.3%	89.7%	96.8%

表15　各発話様式に対する音韻認識率(18子音)

条件(入力3フレームウィンド)

発話様式	1位	2位まで	3位まで
単語発話	96.2%	98.9%	99.6%
短い文節発話	76.2%	86.3%	91.5%
連続発話	56.6%	70.9%	78.5%

表16　各発話様式に対する音韻認識率(18子音)

条件(入力7フレームウィンドウ)

発話様式	1位	2位まで	3位まで
単語発話	96.1%	98.7%	99.4%
短い文節発話	79.5%	90.2%	93.8%
連続発話	57.9%	74.5%	81.6%

表17　各発話様式に対する音韻認識率(18子音)

条件(入力7フレームウィンドウ、出力層と隠れ層第2の重み固定)

発話様式	1位	2位まで	3位まで
単語発話	95.8%	98.6%	99.2%
短い文節発話	82.6%	92.6%	95.8%
連続発話	63.1%	78.3%	85.9%

表18　短い文節発話に対する各TDNNのシフトトレラント性(18子音)
条件(入力3フレームウィンドウ)

シフト幅（ｍｓ）	1位の認識率(%)	2までの累積認識率(%)	3までの累積認識率(%)	4までの累積認識率(%)	5までの累積認識率(%)
-20	55.9	72.4	81.4	87.2	91.2
-10	69.2	82.4	89.0	92.9	95.3
0	76.2	86.3	91.5	94.7	96.6
10	71.7	84.6	90.1	94.0	95.7
20	55.6	75.4	84.5	89.7	92.3

表19　短い文節発話に対する各TDNNのシフトトレラント性(18子音)
条件(入力7フレームウィンドウ)

シフト幅（ｍｓ）	1位の認識率(%)	2までの累積認識率(%)	3までの累積認識率(%)	4までの累積認識率(%)	5までの累積認識率(%)
-20	53.6	75.2	84.3	89.9	93.0
-10	72.9	86.8	92.3	95.3	96.8
0	79.5	90.2	93.8	95.7	97.1
10	73.9	87.5	92.1	94.3	95.6
20	54.4	75.1	83.1	87.8	90.1

表20　短い文節発話に対する各TDNNのシフトトレラント性(18子音)
条件(入力7フレームウィンドウ、出力層と隠れ層第2の重み固定)

シフト幅（ｍｓ）	1位の認識率(%)	2までの累積認識率(%)	3までの累積認識率(%)	4までの累積認識率(%)	5までの累積認識率(%)
-20	57.4	78.0	87.2	92.0	94.6
-10	75.3	88.8	93.5	96.1	97.2
0	82.6	92.6	95.8	97.1	98.1
10	78.5	90.2	95.0	97.2	97.8
20	60.5	82.3	89.6	93.1	94.5