TR-I-0103

時間遅れ神経回路網(TDNN)による音韻スポッティングのための効果的学習法
Effective Training Methods for Spotting Japanese Phonemes Using Time-Delay Neural Networks

宮武 正典　　沢井 秀文　　鹿野 清宏
M. MIYATAKE, H. SAWAI and K. SHIKANO

Aug. 1989

概要

ニューラル･ネットワークを用いた連続音声認識を実現するため、時間遅れ神経回路網(TDNN)による音韻スポッティングを試みた。まず、音韻認識用に学習された音韻統合TDNNを評価用単語音声2,620語に適用したところ、TDNNの持つ時間方向に対するシフト･インバリアントな性質が確認された。さらに予め定めた基準により音韻スポッティング結果を集計したところ、92.5%の音韻が正しく抽出されることが判明し、TDNNの音韻スポッティング能力の高さが確認された。また、スポッティング誤りの傾向が明らかになった。この結果を基に、顕著な挿入や脱落の誤りを除去するために、学習データの抽出位置を考慮して、音韻スポッティングに適した学習方法を提案する。この方法を用いてTDNNを再学習させたところ、全音韻の98.0%が正しく抽出されるとともに、挿入誤りの75%以上が除去され、極めて精度の高い音韻スポッティングを実現した。この結果、ニューラル･ネットワークを用いた連続音声認識の実現の可能性が高まった。

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

内容

1. はじめに

近年ニューラル・ネットワークの音声認識への応用がさかんに研究されている。ネットワークのモジュール化(1)(2)(3)や学習の高速化(4)などの手法が種々提案され、従来困難だった大規模なネットワークの学習が容易になるとともに、認識実験においても良好な結果が得られている。例えば、時間遅れニューラル・ネットワーク(Time-Delay Neural Networks, 以下TDNNと略す)を特定話者の評価用単語音声中の音韻認識に適用し、/b//d//g/の3音韻に対して98.6%、18子音に対して96.7%、母音を含めた23音韻に対しても94.7%と高い音韻認識率を得ている(2)(3)(4)(5)。

我々は現在、連続音声認識に適したニューラル・ネットワークを確立することを目的に研究を行っている。ニューラル・ネットワークを連続音声に適用する方法としては、大きく分けてセグメンテーション手法との組み合わせによる音韻認識(6)と音韻スポッティングの2つが考えられる。我々は、TDNNの高い音韻認識性能と時間方向に対するシフト・インバリアントな性質を活かした音韻のスポッティングのためのアーキテクチャを提案し、予備的な実験で実現の見通しを得ている(7)。本報告では、日本語の全音韻を統合したTDNNを用い、音韻スポッティングに適した学習方法を提案する。さらにこれを単語音声に適用し、精度の高い音韻スポッティングを実現したので、その結果を報告する。

従来のTDNNの学習では、視察(8)に基づく特定位置(以下、音韻代表点という)の音声データ(以下、学習用音韻データという)のみを用いたが、これを音韻スポッティングに適用する場合、学習用音韻データを追加し再学習する必要があると考えられる。その一方、TDNNのシフト・インバリアントな性質により、音韻代表点の周辺においては同様のスポッティング結果が期待できる。そこで本報告ではまず、学習用音韻データのみで学習されたTDNNを単語音声に適用し、未学習区間に対するシフト・インバリアント性や音韻カテゴリ間の側抑制能力を調べる。次にこの結果に基づき、音声データを追加して再学習させ、その効果を確認し、連続音声認識への適応の可能性を示す。

2. 音韻認識用TDNNによる音韻スポッティング

ニューラル・ネットワークを連続音声認識に適用する手法の一つに、音韻スポッティングがある。これは音韻のセグメンテーションを必要としないため、従来の音韻認識用TDNNを拡張する方向でそのまま適用できる利点がある。

適用の方法として、それぞれの音韻をスポッティングするTDNNを音韻グループ別に複数用いる方法が考えられる(7)。この場合TDNNは、該当音韻とそれ以外との識別を行えばよく、全音韻を一度に識別する場合と比較して、各音韻グループ用のネットワークは小規模で実現が可能である。これに対し、全音韻を同時に識別するTDNNを用いた場合、ネットワーク規模は大きくなるが、全音韻を一度に学習するので、有効な側抑制能力が期待できる。そこで本報告では、日本語の全23音韻(18子音 /b/ /d/ /g/ /p/ /t/ /k/ /m/ /n/ /N/ /s/ /sh/ /h/ /z/ /ch/ /ts/ /r/ /w/ /y/ +5母音 /a/ /i/ /u/ /e/ /o/)を識別する音韻統合TDNNを用いることにした。まず最初の試みとして、音韻認識用に学習した音韻統合TDNN(以下、音韻認識用TDNNという)を単語音声に適用した結果について述べる。

2.1 音韻認識用TDNN

音韻認識用TDNNを図1に示す。これは文献(3)によるもので、各音韻グループに対応するネットワーク・モジュールを基に、子音グループ判別用ネットワークを加え、全音韻判別用にスケールアップしたものである。入力層は15×16ユニットからなり、横方向が時間、縦方向が周波数を表す。音声データが入力され

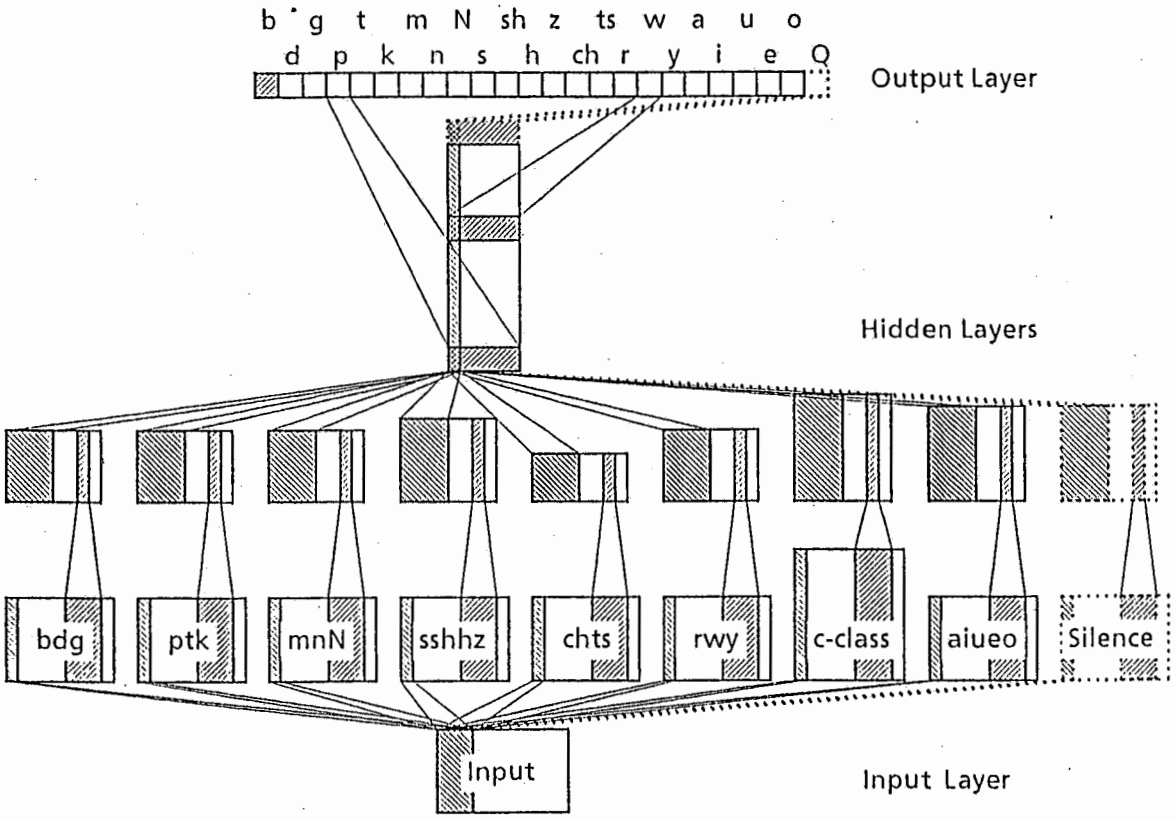

図1 音韻認識用の音韻統合TDNN
(破線は無音カテゴリの付加を表す。3章参照)

ると、全音韻のスポッティング結果が同時に出力される。なお、従来では、各モジュール毎に個別に学習が行われたのち一つのネットワークに統合されたが、現在では文献(4)により、ネットワーク全体でのランダムな初期値からの学習が可能である。

2.2 学習用音韻データ

学習用音韻データは、男性話者1名の発声した日本語の重要語5,240単語の偶数番目の単語中から視察により切り出し、23のカテゴリに分類した4,600サンプルである。1カテゴリ当たり200サンプルであるが、200に満たないカテゴリでは同じサンプルが重複して使われている。特徴パラメータは16次×15フレーム(150msec)のFFTメル・スペクトラムで、全体を平均値が0、絶対値の最大値が1になるように正規化した。

学習用音韻データの切り出し位置を /g/ を例にして図2に示す。横方向は時間、縦方向はFFTの次数を示している。音声データに予め与えられた音韻区間の終端を基準に、前100msec～後50msecの区間(音韻代表点＝図中の斜線部分)を切り出し、それぞれの学習用音韻データとした。

これらの学習用音韻データを用いて、音韻認識用TDNNの学習をバック・プロパゲーション法(9)により行い、連続音声に適用する。

2.3 連続音声への適用方法

図3に連続音声への適用方法を示す。図中、下側は音韻認識用TDNNの入力層と入力音声を、上側は出力層と出力結果を表している。隠れ層は表示を省略している。入力層は16次×15フレームの構成で、150msec分の音声の各入力に対して出

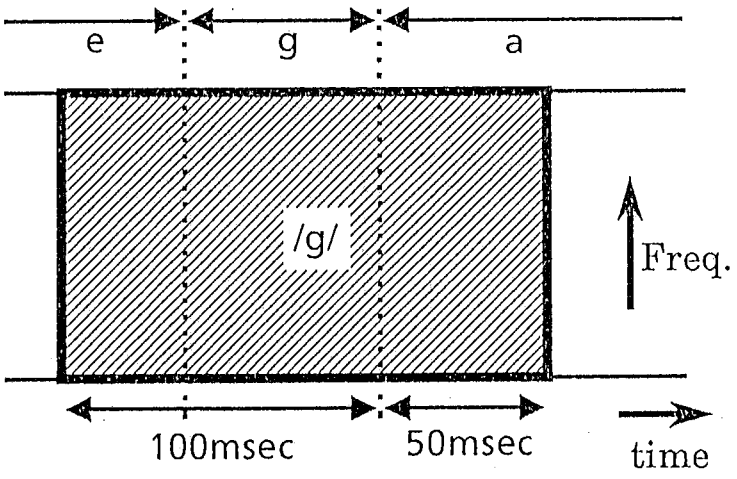

図2 学習用音韻データの切り出し位置(例; /g/)

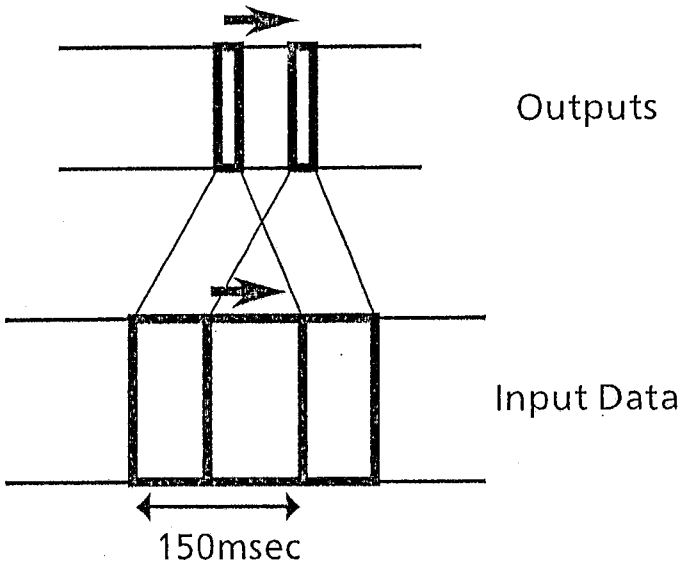

図3　連続音声への適用方法

力層から出力結果が得られる。このような操作を入力音声に対して1フレームずつシフトして行い、得られた出力層からの出力結果の時系列をスポッティング結果とした。

以上で述べた方法を、連続音声への適用を目指して単語音声に適用した結果を次に述べる。

2.4　音韻スポッティング実験

学習用音韻データで学習された音韻認識用TDNNを、同一話者の評価用単語音声(重要語5,240単語の奇数番目)に適用し、音韻スポッティング実験を行った。その結果の一例を図4に示す。図中、横方向が時間、縦方向が音韻カテゴリを表し、

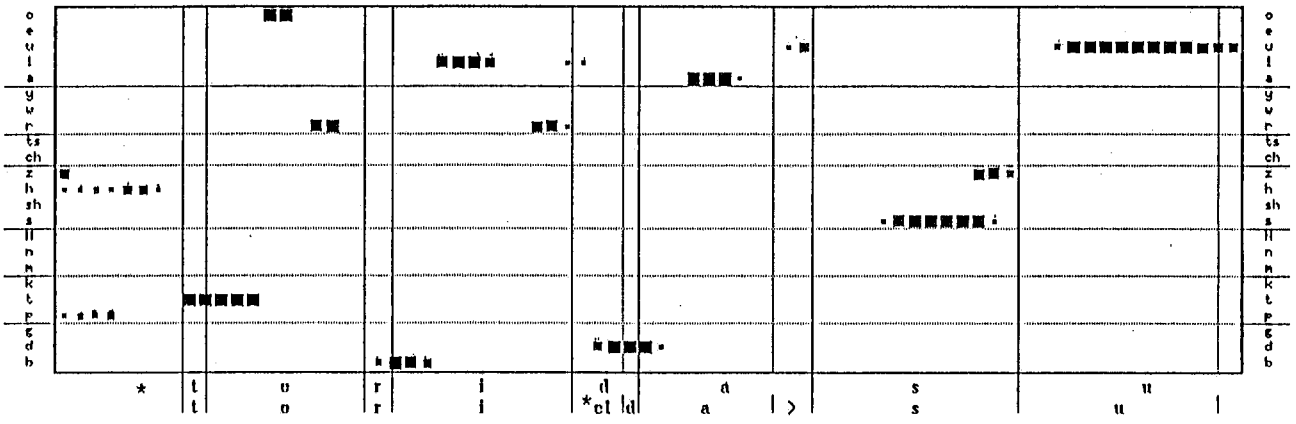

図4　音韻認識用TDNNを用いた音韻スポッティング結果の一例 /toridasu/
(ラベルの*は無音を表す)

黒い四角の大きさで出力層の発火の度合い(0~1)を示している。参考のため、予め視察で与えられたラベル情報を付記した。この図にも見られるように、いくつかの音韻に挿入誤り·脱落誤りが生じたが、ほぼ正しい位置で音韻が検出され、その他の音韻は十分小さい値に抑制されていることがわかった。また、発火はほとんどの場合30msec程度は継続することもわかり、TDNNのシフト·インバリアント性が有効に作用していることが確認された。

このスポッティング結果を表1の基準で集計したものを表2に示す。ここでは単語の前後を除いた音声区間のみを対象にした。この表にも示されているように、全音韻の92.5%が正しく検出できることが明らかになった。その一方で、音韻総数とほぼ同数の挿入誤りがあることも判明した。

さらに誤りを詳細に検討した結果、次のような傾向があることがわかった。

① 語頭·語尾や無声破裂音の閉鎖部など、無音部分の周辺での /h/ と /z/ の挿入誤りが多い(合計5,124個)。これは無音部分が学習されていないために、無音に類似した子音が誤って挿入されたものと考えられる。

② 語頭や過渡部分における /g/ /k/ /r/ の挿入誤りが多い。これらは、過渡部分や継続時間の長い音韻の始端に近い部分で多く見られた。特に /g/ と /r/ は、母音区間の前半部での挿入誤りが多いことがわかった。

③ /s/ と /ts/ 、 /sh/ と /ch/ の置換が多い。

これらの結果から、かなりの音韻が正しく検出できるものの、挿入誤りや脱落誤りを減らし音韻スポッティング性能の向上をはかるためには、無音部分や過渡部分の未学習データを追加して再学習を行う必要があることがわかった。

表1　スポッティング結果の集計基準

出力が0.5以上のひとつ以上連続する区間を発火区間とし、 ·正解　ラベルに基づく音韻区間±30msecに該当音韻の発火区間が存在する ·脱落　上記音韻区間に該当音韻の発火区間が存在しない ·挿入　どの音韻区間にも該当しない発火区間が存在する ただし、正解とされた発火区間が、ラベルに基づく音韻区間±30msecを越える場合は、越える部分を挿入とする

表2　音韻認識用TDNNによる音韻スポッティング結果
(評価用単語2620語、カッコ内は%)

音韻総数	正解	脱落	挿入
13974	12928 (92.5)	1046 (7.5)	13540

3. 再学習による音韻スポッティングの改良

以上の結果を基に、未学習データを追加してTDNNの再学習を行った。まずはじめに、無音部分の音声データを追加して、/h/ や /z/ の挿入誤りの除去を試みた。無音区間を明確にし、音声区間の判定を容易にするため、音韻認識用TDNNに無音判別TDNNを付加し、無音を加えた24カテゴリとした。予備実験として無音判別TDNNを構築し、その有効性の確認を行ったので、その結果を以下に述べる。

3.1 無音判別TDNN

無音判別TDNNは、文献(3)を参考にして、図5に示す構造のものを用いた。学習には、学習用音韻データの一部200個と、単語の前後の無音部から切り出した学習用無音データ200個の計400個を用いた。その結果、表3に示すように、平均で

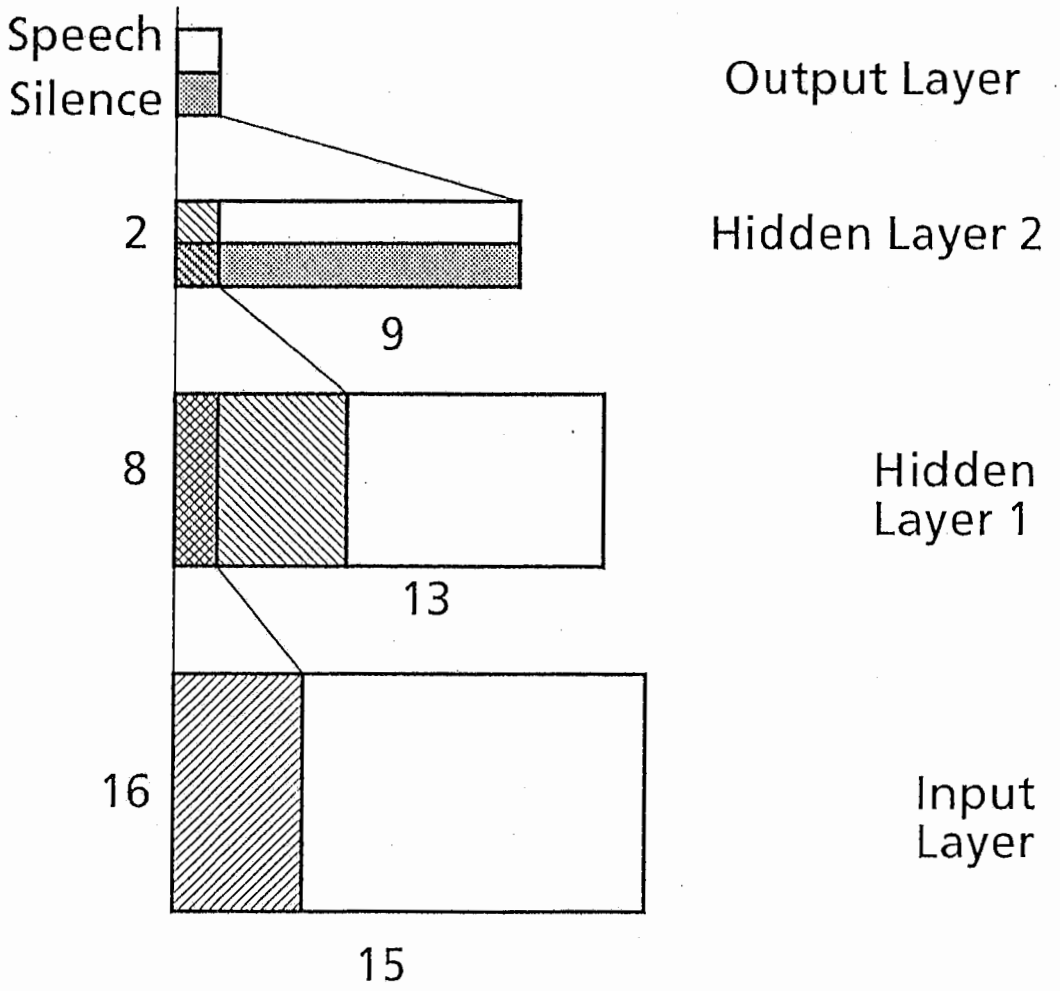

図5 無音判別TDNNの構造

表3　無音判別結果

入力＼出力	音声	無音	認識率
音声(学習用データ)	199	1	99.5%
無音(学習用データ)	0	200	100%
音声(評価用データ)	198	2	99.0%
無音(評価用データ)	0	200	100%

学習用データに対し99.8%、評価用に別に抽出したデータに対し99.5%という高い認識率が得られ、無音判別TDNNの有効性が確認された。これを無音判別モジュールとして音韻認識用TDNNに付加･統合した(以下、無音付加TDNNという)。そして上述の学習用無音データ200個を学習用全音韻データ4,600個に追加した計4,800個のサンプルで再学習を行い、前述の評価用単語音声に適用して、音韻スポッティングの結果を調べた。

3.2　無音付加TDNNによる音韻スポッティング

図6に、無音付加TDNNによる音韻スポッティングの結果の一例を示す。この図にも示されるように、無音を学習する前には語頭･語尾や無声破裂音の閉鎖部など無音部分の周辺において/h/や/z/の挿入誤りが見られたが、無音判別TDNNの付加により、これらの挿入誤りがかなり改善されることが明らかになった。

これらのスポッティング結果を用い、表1の基準で集計したものを表4に示す。この表にも示されるように、95.8%の音韻が正しく抽出でき、なおかつ挿入誤りは音韻総数の62.2%と、音韻認識用TDNNを用いた場合と比較して3分の2以

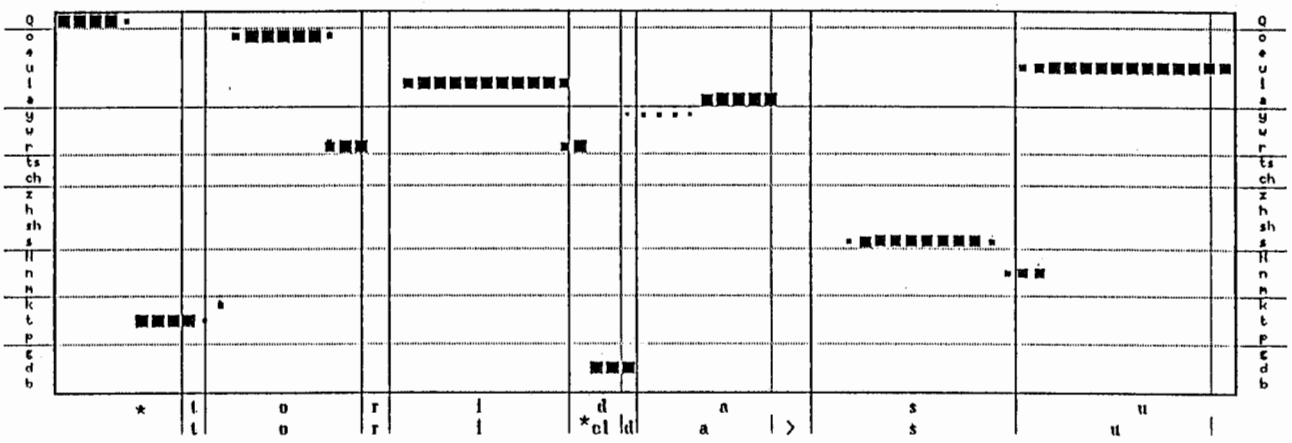

図6　無音付加TDNNによる音韻スポッティング結果の一例 /toridasu/
(ラベルの*、カテゴリのQは無音を表す)

表4　無音付加TDNNによる音韻スポッティング結果
(評価用単語2620語)　(カッコ内は%)

音韻総数	正解	脱落	挿入
13974	13387 (95.8)	587 (4.2)	8689

下になった。誤りの減少傾向を個々に調べると、/h/ と /z/ の挿入誤りの80%以上が除去されていることがわかり、無音付加TDNNの有効性が確認された。

次に、未学習区間の存在による挿入誤り·脱落誤りを除去するために、各音韻区間内の種々の位置から抽出したデータを用いて再学習を行ったので、その方法を以下に述べる。

3.3　複数音韻代表データによる再学習

未学習区間をなくすため、学習用単語2,620語中の各音韻区間に対しそれぞれ複数個のデータ(以下複数音韻代表データという)を抽出し再学習に用いることにした。音韻認識用TDNNによるスポッティング実験において、時間方向のシフト·インバリアント性はおよそ30msecの間有効であることが明らかになっているので、複数音韻代表データは各音韻区間内の20msec毎の各点を中心にした150msecの区間を抽出することとした。この様子を /s/ を例にして図7に示す。この例では矢印で示される5つの区間が学習用データとなる。ただし、目視により決定された音韻境界の誤差や、音韻境界付近を2つのカテゴリに分離する困難さを避けるため、TDNNの時間方向のシフト·インバリアント性に期待して、音韻境界を中心とする20msecの区間のデータは、他にデータがない場合以外は用いなかった。また、無音カテゴリのデータは、単語の始端、終端から20msec離れた部分を用いた。

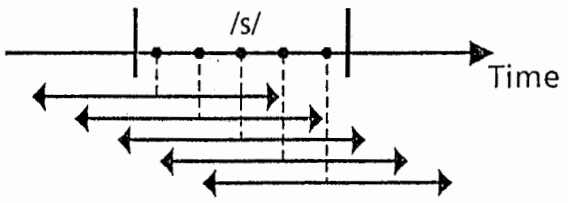

図7　学習用データの抽出方法(音韻区間/s/)

カテゴリ当たり最大400個と1,000個の2通りの上限を設けて再学習を行った無音付加TDNNを用いて音韻スポッティングを行ったので、その結果を以下に示す。

3.4 再学習された無音付加TDNNによる音韻スポッティング

上述の複数音韻代表データで再学習された無音付加TDNNを前記評価用単語音声2,620語に適用した結果の例を図8に示す。この図からもわかるように、それぞれの音韻区間全体に渡って正しく音韻がスポッティングされ、それ以外の音韻は十分抑制されていることが明らかになった。この結果を表1の基準で集計したものを表5に示す。この表にも示されるように、全音韻の96.7~98.0%が正しくスポッティングされることがわかるとともに、個々の音韻について再学習前の無音付加TDNNと比較したところ、/r/ の挿入誤りの90%以上、/g/ /k/ の挿入誤りに関しても60~80%が除去され、TDNNの音韻スポッティング能力の高さが確認された。

また、学習用データ数の増加が脱落誤りの除去に有効であることもわかった。母音や撥音 /N/ の学習用データはそれぞれ5,000~8,000サンプルあり、400サンプルでは十分ではなく、サンプル数を増やすことで判別能力が向上したものと考えられる。

その一方で、挿入誤り・脱落誤りがまだ残っている。しかも、表5の /g/ /k/ /r/ などのように、個々の音韻の挿入誤りと脱落誤りの増減が相反し、一方が減少すると他方が増加する状況が見られる。すでに正解率は98%に達しているが、今後さらに発声変動や話者による変動を吸収するためには、言語処理等の後処理に及ぼす影響を十分に考慮した柔軟な判別を可能とする学習方法の検討が必要である。

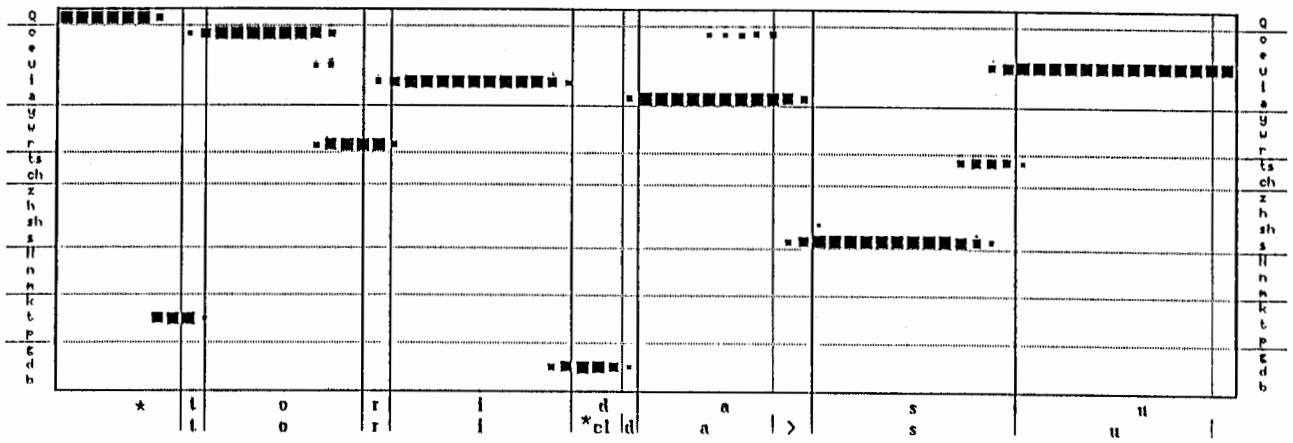

図8 複数音韻代表データで学習された無音付加TDNNによる音韻スポッティング結果の一例 /toridasu/ (ラベルの*、カテゴリのQは無音を表す)

表5　複数音韻代表データで再学習したTDNNによる音韻スポッティング結果
(評価用単語2620語、カッコ内は%)

音韻	総数	400個/音韻			1000個/音韻		
		正解	脱落	挿入	正解	脱落	挿入
/b/	231	228	3	268	225	6	104
/d/	180	175	5	106	171	9	71
/g/	265	230	35	198	210	55	57
/p/	28	25	3	203	26	2	104
/t/	461	452	9	178	459	2	235
/k/	1300	1218	82	116	1283	17	245
/m/	485	482	3	323	479	6	213
/n/	273	258	15	84	263	10	63
/N/	488	487	1	161	488	0	163
/s/	572	570	2	175	572	0	100
/sh/	387	385	2	52	386	1	81
/h/	313	312	1	215	310	3	159
/z/	315	310	5	170	307	8	87
/ch/	141	140	1	57	141	0	163
/ts/	220	219	1	205	218	2	235
/r/	760	709	51	62	730	30	97
/w/	81	80	1	74	79	2	13
/y/	573	531	42	124	561	12	171
/a/	1772	1770	2	108	1771	1	85
/i/	1333	1282	51	155	1302	31	200
/u/	1615	1496	119	206	1543	72	200
/e/	829	822	7	222	827	2	254
/o/	1352	1337	15	97	1348	4	136
計	13974	13518 (96.7)	456 (3.3)	3559	13699 (98.0)	275 (2.0)	3236

最後に、最大1,000個の複数音韻代表データで学習された無音付加TDNNのスポッティング能力や一般化能力をさらに詳細に調べたので、その結果を述べる。

3.5　音韻スポッティング能力の評価

最大1,000個の複数音韻代表データで学習された無音付加TDNNを用い、学習用の複数音韻代表データに対する音韻認識、およびこれと同じ基準で前述の評価用単語2,620語から抽出したデータに対する音韻認識の実験結果を表6に示す。ただし、評価用データにおける誤りにはリジェクト(閾値=0.1)を含む。これらの表にも示されるように、学習用データに対し98.4%、評価用データに対し95.9%と、従来のTDNNによる音韻認識性能と同等以上の認識率を示している。

次に、3.4節のスポッティング結果を用い、音韻境界を含む全てのフレームに対して音韻認識率を調べた結果を表7に示す。また、学習用単語2,620語に対して同様のスポッティング実験を行い、全てのフレームに対して音韻認識率を調べた結果を表8に示す。また、表1の基準に従って集計した結果を参考のために表9に示す。これらの実験では、閾値(0.1)未満のものはリジェクトとし、学習には用い

表6　複数音韻代表データで再学習したTDNNを用いた音韻認識率
(音韻境界を除く)

音韻	学習用データに対し			評価用データに対し		
	総数	誤り	認識率(%)	総数	誤り	認識率(%)
/b/	472	2	99.6	507	13	97.4
/d/	448	0	100.0	400	15	96.3
/g/	609	66	89.2	585	128	78.1
/p/	48	0	100.0	28	2	92.9
/t/	454	0	100.0	461	7	98.5
/k/	1000	2	99.8	1696	38	97.8
/m/	1000	6	99.4	1045	53	94.9
/n/	460	2	99.6	459	39	91.5
/N/	1000	22	97.8	4130	198	95.2
/s/	1000	32	96.8	3348	240	92.8
/sh/	1000	7	99.3	2785	83	97.0
/h/	846	9	98.9	815	50	93.9
/z/	1000	10	99.0	1117	36	96.8
/ch/	565	4	99.3	533	25	95.3
/ts/	1000	4	99.6	842	22	97.4
/r/	838	7	99.2	812	28	96.6
/w/	124	0	100.0	117	6	94.9
/y/	1000	30	97.0	1147	116	89.9
/a/	1000	5	99.5	8432	177	97.9
/i/	1000	26	97.4	7049	383	94.6
/u/	1000	24	97.6	8417	530	93.7
/e/	1000	8	99.2	5585	145	97.4
/o/	1000	7	99.3	10036	144	98.6
/Q/	1000	29	97.1	5240	235	95.5
計	**18864**	**302**	**98.4**	**65586**	**2713**	**95.9**

表7　複数音韻代表データで再学習したTDNNを用いた音韻認識率
(評価用単語中の境界を含む全152,421フレームに対して)
(カッコ内は、無音/Q/を除いた場合)

1位以内(%)	2位以内	3位以内	4位以内	5位以内
82.7 **(84.6)**	**91.3** **(93.0)**	**92.2** **(93.8)**	**92.2** **(93.8)**	**92.2** **(93.8)**

表8　複数音韻代表データで再学習したTDNNを用いた音韻認識率
(学習用単語中の境界を含む全152,350フレームに対して)
(カッコ内は、無音/Q/を除いた場合)

1位以内(%)	2位以内	3位以内	4位以内	5位以内
83.2 **(85.1)**	**91.8** **(93.4)**	**92.5** **(94.1)**	**92.6** **(94.2)**	**92.6** **(94.2)**

られていないイベント層の/cl/は、無音/Q/に含めた。表7、表8に示されるように、学習用単語に対する音韻認識率と評価用単語に対する音韻認識率とにほとんど差はなく、いずれも認識率は第2位以内でほぼ飽和していることがわかった。

表9 複数音韻代表データで再学習したTDNNによる音韻スポッティング結果
(学習用単語2620語) (カッコ内は%)

音韻総数	正解	脱落	挿入
14020	13839 (98.7)	181 (1.3)	2864

表10 複数音韻代表データで再学習したTDNNによる音韻スポッティング結果
(音韻バランス単語216語) (カッコ内は%)

音韻総数	正解	脱落	挿入
1428	1306 (91.5)	122 (8.5)	556

表6に示されるように、音韻境界を除いた音韻認識率は評価用単語に対して95.9%であるから、その差が主に音韻境界部分に依存していると考えられる。第2位以下の認識率をいかに向上させるかが今後の課題である。

さらに、同じTDNNを同一話者の発声した音韻バランス単語216語に対して用い、音韻スポッティングを行い、表1の基準で集計したものを表10に示す。この表にも示されるように、正解音韻の抽出率は、重要語の評価用単語における98.0%と比較して91.5%とかなり悪い。音韻バランス単語と重要語に共通な単語190語に対するスポッティング結果だけを比較したところ、正しく抽出できた音韻は、音韻バランス単語において91.7%、重要語において97.1%となり、音韻環境の違いに依存した差ではない。認識率低下の原因として、録音やA/D変換の条件の違い、経時変化による発声の違い等の微妙な差がTDNNの音韻認識性能にかなり影響を及ぼしているのではないかと考えられるが、実際の音声を聞いてもそれほどの違いは感じられず、さらに原因の究明をする必要がある。

4. むすび

TDNNを連続音声の音韻スポッティングに適用するための方法について述べ、最初の試みとして大語彙単語音声に適用した。その結果、音韻認識用に学習されたTDNNでも92.5%の音韻が正しく抽出され、TDNNの音韻スポッティング性能の高さが判明した。また、TDNNの持つ時間方向のシフト・インバリアントな性質が確認された。その一方で、スポッティング誤りの傾向が明らかになった。この結果を基に、顕著な挿入や脱落の誤りを除去するために、TDNNに無音カテゴリを付加するとともに、学習データの抽出位置を考慮した効果的な学習方法を提案した。この方法を用いて無音付加TDNNを再学習し、単語音声に適用し

た結果、全音韻の98.0%が正しく抽出され、なおかつ、75%以上の挿入誤りが除去された。これは極めて高精度の音韻スポッティングであり、TDNNの音韻識別能力の高さが改めて実証された。また、TDNNを用いた音韻スポッティングにおいて今後解決すべき問題点も明らかになった。今後は連続音声への適応をはかるとともに、言語処理との結合をはかっていく。そのために、発声速度の違いや話者間/話者内変動に対応できる柔軟な判別能力を持ったネットワークを実現するための学習方法や構造の確立を目指す。

謝辞

日頃ご指導いただく樽松社長、たくさんの助言をいただいたA. Waibel氏をはじめ、活発なご討論をいただいた音声情報処理研究室の皆様に感謝します。

文献

(1) 松岡、浜田、中津「グループ分割型ニューラルネットによる単音節音声認識」信学技報SP88-15(1988)

(2) A. Waibel, H. Sawai and K. Shikano, "Phoneme Recognition by Modular Construction of Time-Delay Neural Networks", 昭63秋季音響学会講論集2-P-12 または 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)

(3) H. Sawai, A. Waibel, M. Miyatake and K. Shikano, "Phoneme Recognition by Scaling up Modular Time-Delay Neural Networks", 信学技報SP88-105(1988)

(4) P. Haffner, H. Sawai, A. Waibel and K. Shikano, "Fast Back-Propagation Learning Methods for Neural Networks in Speech Recognition", 信学春季全大SA-1-1(1989) または P. Haffner, H. Sawai, A. Waibel and K. Shikano, "Fast Back-Propagation Learning Methods for Large Phonemic Neural Networks", 平元春音響講論集1-6-14

(5) A. Waibel, "Phoneme Recognition Using Time-Delay Neural Networks", 信学技報SP87-100(1987) または A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks", IEEE Trans. on ASSP, Vol. 37, No. 3, pp.328-339 (1989)

(6) 小森、畑崎、田中、川端、鹿野「スペクトログラム・リーディング知識とニューラル・ネットワークを用いた音韻認識エキスパートシステム」、信学技報SP89-33(1989)

(7) 沢井、宮武、ワイベル、鹿野「連続音声認識のための時間遅れ神経回路網を用いた音韻/音節スポッティング」、信学論(D-II), J72-D-II, 8, pp.1151-1158(1989)

(8) 武田、匂坂、片桐、桑原「研究用日本語音声データベースの構築」、音響誌、Vol. 44, No. 10, pp.747-754(1988)

(9) D. E. Rumelhart, J. L. McClelland and the PDP Research Group, "Parallel Distributed Processing", MIT Press(1986)