

TR-I-0219

TDNN を用いた不特定話者の音素認識 (2)

Speaker Independent Phoneme Recognition
Using Time-Delay Neural Networks

中村 悟†	沢井 秀文‡	杉山 雅英§
Satoru Nakamura	Hidefumi Sawai	Masahide Sugiyama
†慶応大学理工学部	‡Ricoh 中央研究所	§ATR 自動翻訳電話研究所

内容梗概

ニューラルネットワークを用い、多数話者学習による不特定話者の音素認識について報告する。予備実験として、有声破裂音 /b,d,g/ の音素認識実験を行なった。その結果、Modular-TDNN の認識性能が最も優れ、構造上の有意性が確認されている。学習用話者を 6 名、及び、12 名とした時の評価用話者 8 名の平均認識率は、それぞれ、91.3%、93.6% であった。

本報告では、更に、認識音素を 18 子音へと拡張し不特定話者音素認識を行なった。その結果、学習用話者 6 名、評価用話者 8 名とした時の認識率は、6 子音 /b,d,g,m,n,N/ において、86.8%、5 母音 /a,i,u,e,o/ において、93.7%、18 子音において、80.4% であった。これにより、Modular-TDNN により構成されるネットワークの有効性が示された。

© ATR Interpreting Telephony Research Laboratories

© ATR 自動翻訳電話研究所

目次

1	はじめに	1
2	ネットワーク	1
2.1	未知話者の /b, d, g/ 音素認識におけるネットワークの性能比較	1
2.2	18子音認識への拡張	3
3	実験	6
3.1	音声資料	6
3.2	ネットワークの学習	6
4	実験結果	6
4.1	6子音 /b, d, g, m, n, N/ 認識実験	7
4.2	5母音 /a, i, u, e, o/ 認識実験	7
4.3	18子音認識実験	8
5	考察	12
6	むすび	13
A	各話者の認識率	15
B	コンフュージョンマトリックス	18

図目次

1	Modular-TDNN の構造	3
2	不特定話者 6 子音 /b, d, g, m, n, N/ 認識用ネットワークの構造	4
3	不特定話者 18 子音認識用ネットワークの構造	5

表目次

1	未知話者 (8 名) の /b, d, g/ 音素認識率 — Architecture による比較	2
2	未知話者の /b, d, g/ 音素認識率 — 学習方法による比較	2
3	未知話者 (8 名) の /b, d, g/ 音素認識率 — キャパシティによる比較	2
4	学習及び評価に用いた話者	6
5	6 子音 /b, d, g, m, n, N/ の認識率 — closed speaker の場合	7
6	6 子音 /b, d, g, m, n, N/ の認識率 — open speaker の場合	7
7	5 母音 /a, i, u, e, o/ の認識率 — closed speaker の場合	7
8	5 母音 /a, i, u, e, o/ の認識率 — open speaker の場合	8
9	全音素の認識率	9
10	18 子音の認識率 — closed speaker の場合	10
11	18 子音の認識率 — open speaker の場合	11
12	不特定話者の認識実験	12
13	不特定話者の認識実験	13
14	Single-TDNN を用いた特定話者の音素認識 (各音素と音素グループの認識率)	15
15	Modular-TDNN を用いた Multi-Speaker の音素認識 (各音素と音素グループの認識率)	16
16	Modular TDNN を用いた未知話者の音素認識 (各音素と音素グループの認識率)	17

1 はじめに

近年、音声認識の分野において、ニューラルネットワークの応用が活発に行なわれていきている。計算機の進歩とともに、ニューラルネットワークの高速化アルゴリズムにより、大規模ネットワークの学習が容易になった。時間遅れ神経回路網 (Time-Delay Neural Network, 以下 TDNN と略す) を用いた特定話者音素認識において高い性能が報告されて以来 [1] [2]、Multi-speaker 音素認識へと拡張されてきた [3]。本報告では、TDNN を基本構造とするニューラルネットワークを不特定話者の音素認識に適用し、その能力を明らかにする。ニューラルネットワークを用いた不特定話者音素認識の方法には、大分して話者適応によるものと、多数話者学習による 2 つの方法が考えられている [4]。今回、特定話者や Multi-speaker 音素認識実験の知見をもとに、後者の多数話者学習による不特定話者の音素認識を試みる。多数話者学習の準備実験として、有声破裂音 /b, d, g/ 音素認識実験が行なわれ [5]、不特定話者の音素認識において高い認識性能が示された。しかし、認識音素数を 18 子音、更に 5 母音と、無音を加えた日本語 24 音素へと拡張をする場合、ネットワークの規模、学習サンプルが大きくなりネットワークの学習に時間がかかるという欠点がある。この問題に対して、部分的学習及びサンプルの分割学習を行なうことにより、不特定話者 18 子音音素認識へと拡張した。その結果、学習用話者 6 名、評価用話者 8 名とした時の認識率は、6 子音 /b, d, g, m, n, N/ において、86.8%、5 母音 /a, i, u, e, o/ において、93.7%、18 子音において、80.4% と、高い認識率が得られた。これにより、Modular-TDNN により構成されるネットワークの有効性が示された。

2 ネットワーク

2.1 未知話者の /b, d, g/ 音素認識におけるネットワークの性能比較

不特定話者音素認識の予備実験として、特定話者、及び Multi-speaker の音素認識用のネットワークを不特定話者音素 /b, d, g/ の認識に適用し、ネットワークアーキテクチャー、学習法、キャパシティーについて検討を行なった¹。

アーキテクチャーによる性能比較を行なったネットワークは、Single TDNN、SID network、Meta-Pi network、Modular TDNN、MSID network の 5 つであり、この実験結果を表 1 に示す。この結果、Modular-TDNN による認識率が最も良く、6 話者学習時及び 1 2 話者学習時での認識率は、それぞれ、91.3%、93.6% であった。ここで、MSID ネットワークの構造は話者識別用の Speaker I.D. network を Modular-TDNN に付加した構造になっているが、このサブネットを付加することによる認識率向上における有意性はみられないことが判る。Multi-Speaker の音素認識において優れた性能が報告されている Meta-Pi network の不特定話者に対する音素認識率は、予想された通り低いものとなった。

又、認識率の最も良かったネットワークである Modular-TDNN において、学習方法による認識実験の比較を行なった。2 段階学習 (2 stage learning)、つまり、Modular TDNN の構成単位である各話者毎に学習された Single TDNN の重みを初期値に用いて学習を行なった場合と、ランダム値から学習を行なった場合の比較実験を行なった。その結果を表 2 に示す。若干ではあるが、2 段階学習による認識率は、open speaker と closed speaker の両方において良く、2 段階学習の有効性が示された。

更に、ネットワークのキャパシティーによる実験は、Single TDNN を用いて、第 1 隠れ層のユニット数を増やすことにより行なった。その結果を表 3 に示す。表から、第一隠れ層のユニット数を増やすにつれ、ある程度の認識率を得ることが出来ることがわかる。しかし、Modular-TDNN と同程度のキャパシティーをもつ Single-TDNN の認識率は、Modular-TDNN の認識率を下回るものである。これは、Modular-TDNN がキャパシティーを有効に利用しているためと考えられる。

¹ 文献 [5] における評価話者数は 4 名であったが、認識率の信頼性を上げるため、評価話者数を 4 名増やし、8 名として再評価した。

表 1: 未知話者 (8 名) の /b, d, g/ 音素認識率 — Architecture による比較

Network Architecture	recognition rate (%)			
	6 話者学習時		12 話者学習時	
	closed	open	closed	open
Single TDNN	93.2	82.0	—	—
SID	96.4	79.2	95.4	82.4
Meta-Pi	96.9	79.3	95.4	82.4
Modular TDNN	97.7	91.3	97.1	93.6
MSID	97.2	89.4	97.3	93.4

表 2: 未知話者の /b, d, g/ 音素認識率 — 学習方法による比較

学習方法	6 話者学習時の認識率 (%)	
	closed	open
2 段階学習	97.7	91.3
ランダム値 からの学習	97.3	90.9

表 3: 未知話者 (8 名) の /b, d, g/ 音素認識率 — キャパシティによる比較

隠れ層 1 の ユニット数	コネクション 数	認識率 (%)	
		closed(6 話者平均)	open(12 話者平均)
8	6233	93.2	82.0
16	12409	94.6	85.6
24	18585	95.4	87.9
32	24761	95.9	90.0
40	30937	97.4	89.1
48	37113	97.4	88.9
56	43289	97.6	90.8

以上の実験結果から、以下のことが示された。

- Modular-TDNN のようにモジュール化を行なうことにより、限られたキャパシティーを有効に利用し、より良い認識率を得ることが出来る。
- ネットワークの各モジュールの初期値として、モジュール毎の学習済みの重みを用いた方が、ランダム値からの学習よりも、安定かつ迅速に収束する。
- 特定話者及び Multi-speaker に対して音素認識を行なう際には、Meta-Pi ネットワークのように話者情報を有効に利用できるが、不特定話者音素認識の場合はその限りではない。

2.2 18 子音認識への拡張

これらの実験結果をふまえ、認識する音素を $/b, d, g/$ から 18 子音及び 23 音素へと拡張することを考える。図 1 に Modular-TDNN の構造を示す。この Modular-TDNN 構造は、3 音素 $/b, d, g/$ における未知話者音素認識において報告されているように、2 段階学習と、モジュール構成により学習を安定して極小解に収束させ、かつ、限られたキャパシティーを有効に利用するためのものである。この Modular-TDNN を基本構造として、 $/b, d, g, m, n, N/$ の 6 子音認識用、18 子音用に拡張したネットワークの構造を図 2、図 3 に示す。ここで、各音素カテゴリに対応する Modular-TDNN に加え、音素カテゴリを識別するネットとして consonant class net がある。

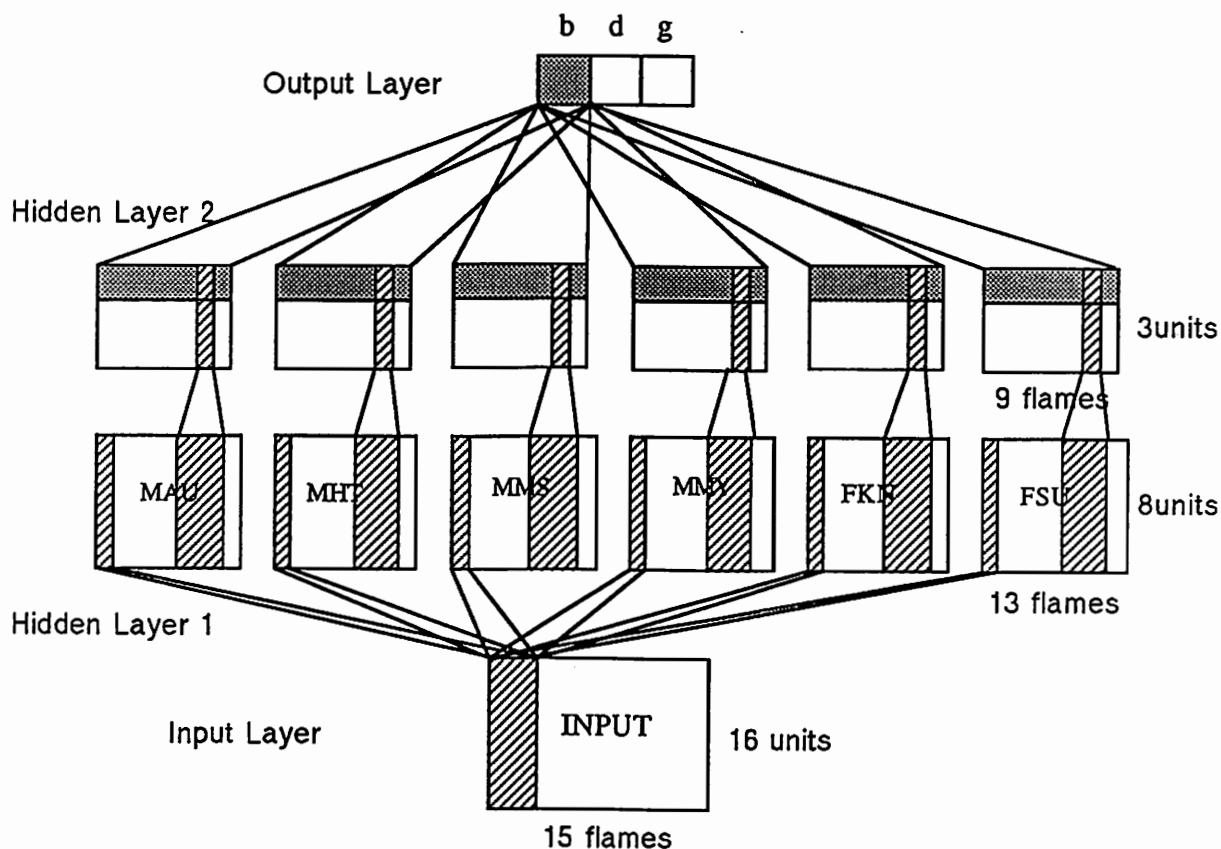


図 1: Modular-TDNN の構造

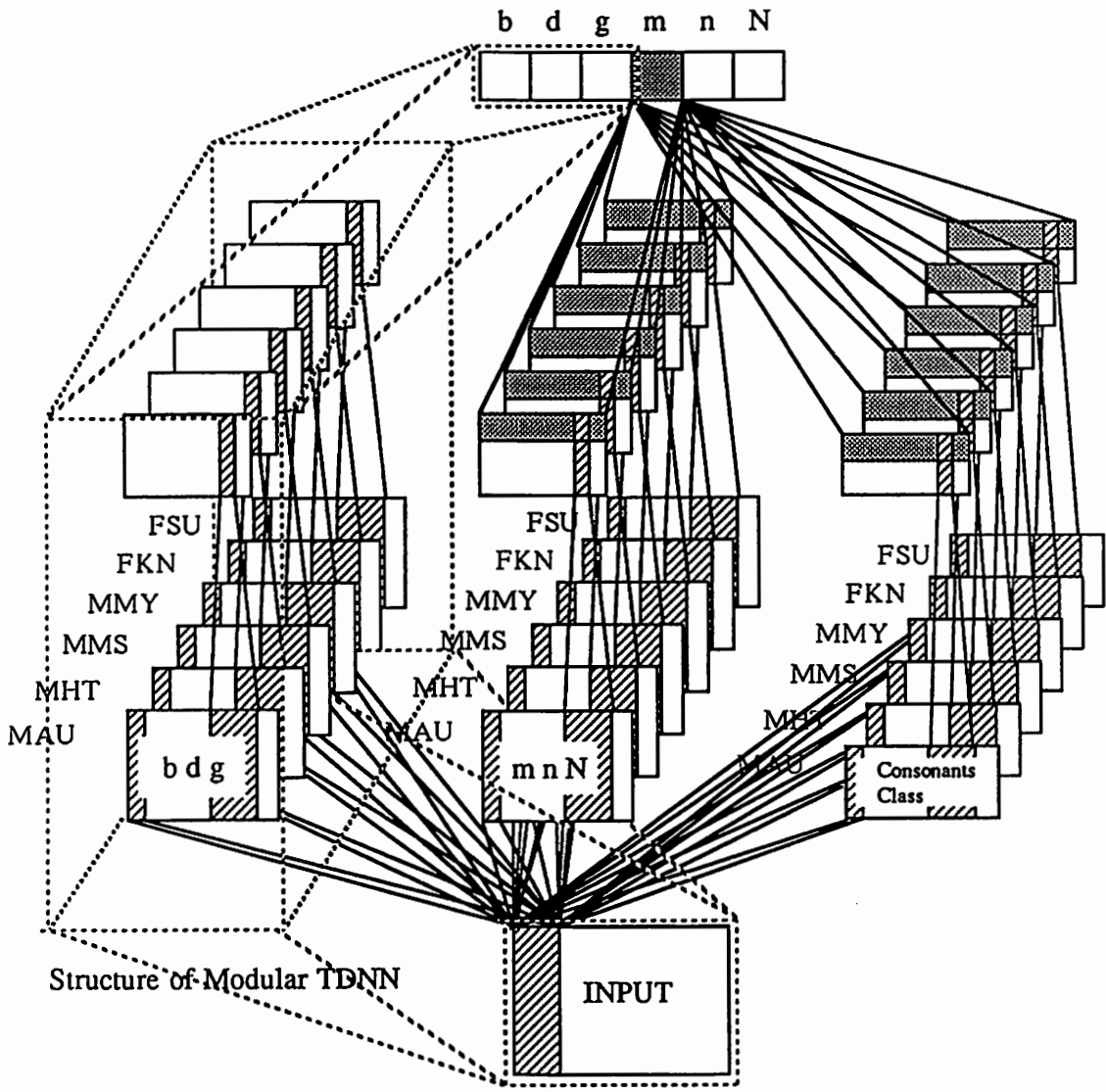


図 2: 不特定話者 6 子音 /b, d, g, m, n, N/ 認識用ネットワークの構造

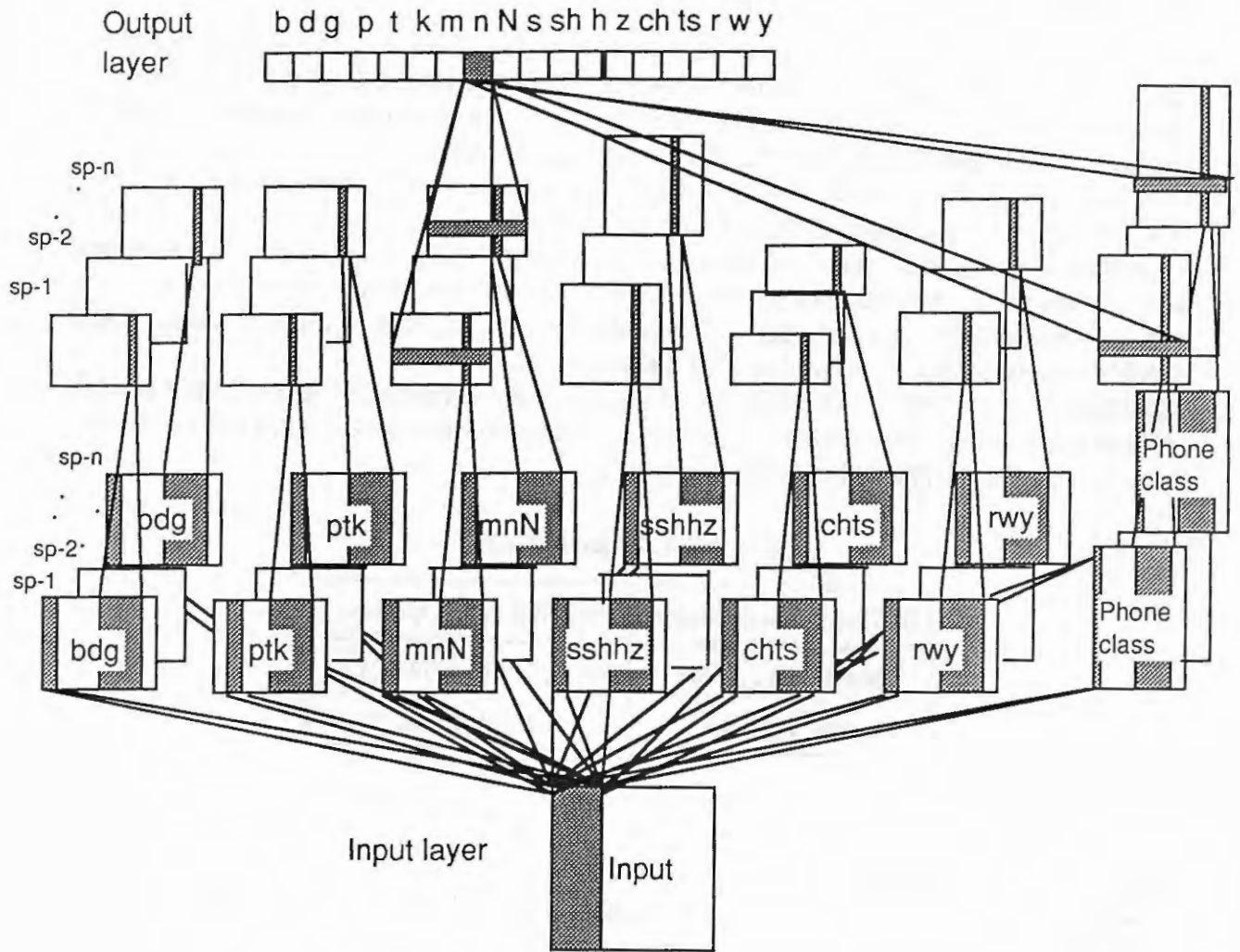


図 3: 不特定話者 18 子音認識用ネットワークの構造

3 実験

3.1 音声資料

ネットワークの学習用音素データは、男性4名、女性2名、計6名²のアナウンサーの発声した日本語重要単語 5240 語の偶数番目の単語中より視察ラベルにより切り出されたものである。各音素のデータ数は、子音で最大 500 サンプル、母音で最大 600 サンプルとした。1 カテゴリ当たりのサンプル数が最大値に満たない場合は、サンプル数の偏りをなくすため、同じサンプルを重複して用いた。

ネットワークへの入力データは、各サンプルにおいて、16 次×15 フレーム (150ms) の FFT メル・スペクトラム時系列であり、平均値 0、最大値 +1、最小値 -1 に正規化した。

各音素の切り出し位置は、音素により異なるものであった。音素 /b, d, g, p, t, k/ の場合、各子音の終端部が入力 150ms のウィンドの中心、つまり、75ms にくるように、他の子音 /m, n, N, s, sh, h, z, ch, ts, r, w, y/ の場合、子音終端部がウィンドの中心位置より 40ms 遅れるように、又、母音 /a, i, u, e, o/ の場合、始端部分がウィンドの中心位置より、40ms 前にくるように切り出した。

認識実験用データは、学習に用いた 6 話者 (closed speaker) に加え、学習には用いなかった男性 4 名、女性 4 名の計 8 名 (open speaker) の発声した、重要単語 5240 単語中の奇数番目から切り出されたサンプルを用いた。ここで、学習及び評価に用いた話者名を表 4 に示す。

表 4: 学習及び評価に用いた話者

学習話者 (closed speaker)	評価話者 (open speaker)
MAU, MHT, MMS	MNM, MTK, MTM, MTT
MMY, FKN, FSU	FAF, FKM, FYM, FYN

3.2 ネットワークの学習

学習則として、Back-Propagation アルゴリズムを用いた。更に、大規模ネットワーク及び膨大な学習データの学習を高速に処理するために高速化アルゴリズム [7] を用いた。各実験に用いたネットワークは Single-TDNN を構成単位とするモジュール構造になっている。音素グループの各モジュールは予め学習され、ネットワーク全体を構成するとき、ネットワークの初期値として使われる。

学習法をまとめると次のような 2 段階学習となる。

1. 各話者、各音素カテゴリ毎に Single-TDNN を用いて学習
2. 全ネットワークの学習

未知話者の音素認識を次の 3 つの音素グループで行なった。

- 6 子音 /b, d, g, m, n, N/
- 5 母音 /a, i, u, e, o/
- 18 子音 /b, d, g, p, t, k, m, n, N, s, sh, h, z, ch, ts, r, w, y/

4 実験結果

各 3 つの実験における評価は、closed speaker と open speaker に対して行なった。学習は 500 epoch 行ない、学習サンプルに対する認識率がほぼ収束した範囲で 10 epoch おきに各話者毎に認識実験を行なった。各 epoch 毎に各話者の平均認識率を求め、最良の平均認識率を示した epoch における結果を示した。

以下各実験結果を示す。

²学習話者は文献 [3]、[5] における実験結果と比較するため同様の学習話者を用いた。

4.1 6子音 /b, d, g, m, n, N/ 認識実験

不特定話者の6子音 /b, d, g, m, n, N/ 認識実験の結果を表5と表6に示す。

表5: 6子音 /b, d, g, m, n, N/ の認識率 — closed speaker の場合

Phoneme	speaker						Recognition Rate
	MAU	MHT	MMS	MMY	FKN	FSU	
b	96.90	98.60	96.40	91.80	97.10	96.10	96.15
d	97.80	99.40	94.30	95.30	95.90	97.60	96.72
g	92.90	97.60	96.80	97.60	94.90	96.30	96.02
m	98.50	97.90	96.90	98.70	97.50	97.90	97.90
n	92.50	96.60	91.40	91.20	96.60	90.60	93.15
N	98.80	99.60	99.00	99.50	97.70	99.50	99.02
Average	96.70	98.30	96.30	96.50	96.90	96.70	96.90

表6: 6子音 /b, d, g, m, n, N/ の認識率 — open speaker の場合

Phoneme & Phoneme Group	speaker								Recognition Rate(%)
	MNM	MTK	MTM	MTT	FAF	FKM	FYM	FYN	
b	84.7	98.6	86.5	95.7	92.8	87.9	85.5	94.0	90.7
d	96.0	92.9	96.5	66.5	83.6	92.4	86.5	84.9	87.4
g	91.7	88.9	91.3	94.5	72.8	66.4	88.1	78.5	84.0
m	91.3	94.2	78.2	97.7	89.4	64.5	72.2	96.1	85.5
n	61.5	59.2	82.6	66.0	78.1	83.0	86.5	58.4	71.9
N	95.0	97.7	94.8	98.0	96.4	97.3	99.3	96.8	96.9
Average	87.7	89.6	87.3	89.4	87.0	80.4	85.9	87.1	86.8

4.2 5母音 /a, i, u, e, o/ 認識実験

表7: 5母音 /a, i, u, e, o/ の認識率 — closed speaker の場合

Phoneme & Phoneme	speaker						Recognition Rate(%)
	MAU	MHT	MMS	MMY	FKN	FSU	
a	100.00	99.30	98.80	99.70	98.30	99.50	99.27
i	99.00	99.80	99.50	99.50	99.30	99.30	99.40
u	97.50	97.20	96.80	99.20	99.00	97.70	97.90
e	98.20	99.00	98.20	99.20	98.80	99.20	98.77
o	99.80	100.00	98.00	98.20	99.50	99.70	99.20
Average	98.90	99.10	98.30	99.10	99.00	99.10	98.92

表 8: 5 母音 /a, i, u, e, o/ の認識率 — open speaker の場合

Phoneme & Phoneme Group	speaker								Recognition Rate (%)
	MNM	MTK	MTM	MTT	FAF	FKM	FYM	FYN	
a	96.20	99.20	75.00	93.80	99.20	99.20	92.70	94.00	93.66
i	99.80	97.20	99.50	95.00	93.50	90.00	99.30	99.30	96.70
u	95.70	89.70	96.20	95.80	96.50	66.70	96.20	98.00	91.85
e	90.30	99.30	96.30	93.30	86.80	88.80	94.70	90.20	92.46
o	98.00	94.80	98.50	91.50	99.70	80.50	94.70	95.70	94.18
Average	96.00	96.00	93.10	93.90	95.10	85.00	95.50	95.40	93.75

4.3 18 子音認識実験

18 子音の学習に際して、他のネットワークと同様に 2 段階学習を行なったが、ネットワークの規模と学習サンプルが大きくなったため、ネットワークの部分的な学習、及び、学習サンプルの分割により学習を行なった。ネットワークの部分的な学習は、第 2 隠れ層と出力層間のみを学習することにより行なわれた。つまり、各サンプルに対し、学習済みの各モジュールより得られる第 2 隠れ層の全ユニット (1296 ユニット) の発火パターンを新たな学習サンプルとして、2 層ネットワークの学習を行なった。又、18 子音学習時の学習サンプル数は 5400 サンプルであり、10 個のサンプルに分割して学習を行なった。各サブ学習サンプルに対し 100 epoch の学習を行ない、これを 10 回繰り返した。しかしながら、第 2 隠れ層から出力層への 2 層のネットワークでは、学習データとして用いる各モジュールの発火パターンが十分線形分離可能でないために学習データの認識率は 80% 程度に収束し学習が進まなかった。そこで、学習方法として、未知話者の 3 音素 /b, d, g/ の認識実験の結果をふまえ、各音韻グループ毎に Modular TDNN で学習を行ない、最終的に先に述べた様に、発火パターンにより第 2 隠れ層と出力層間の学習を行なった。改めて、学習の手順を示すと以下のようになる。

1. 各話者、各音素カテゴリ毎の Single-TDNN を用いた学習
2. 各音素カテゴリ毎の Modular-TDNN を用いた学習
3. 全ネットワークの学習 (部分的学習、サンプルの分割学習)

その結果、学習データに対して 92% 程度まで学習が進んだ。2 段階目の学習の過程で、得られた各 Modular-TDNN の認識率を特定話者の 6 名平均認識率と共に表 9 に示す。又、18 子音認識用ネットワークの closed speaker 及び open speaker に対する認識率をそれぞれ表 10、表 11 に示す。

表 9: 全音素の認識率

Phoneme	特定話者		Multi-speaker		未知話者	
	Recognition Rate	Average	Recogniton Rate	Average	Recognition Rate	Average
b	98.12	98.27	97.10	97.65	89.06	90.56
d	98.48		96.78		94.28	
g	98.25		98.67		89.27	
p	86.50	98.17	78.58	98.35	66.66	92.38
t	98.20		98.52		90.50	
k	98.47		98.80		94.80	
m	98.22	98.22	97.83	97.15	86.56	87.24
n	96.42		92.55		71.03	
N	99.25		99.22		97.97	
s	99.68	99.62	99.62	99.43	96.38	95.90
sh	99.80		99.52		95.09	
h	99.38		99.20		94.55	
z	99.62		99.33		96.99	
ch	100.00	100.00	100.00	99.83	99.83	99.80
ts	100.00		99.75		99.79	
r	99.63	99.10	99.33	98.22	98.18	97.05
w	99.13		98.05		96.71	
y	97.45		94.98		93.97	
a	99.47	99.23	99.27	98.92	93.66	93.75
i	99.47		99.40		96.70	
u	98.62		97.90		91.85	
e	99.38		98.77		92.46	
o	99.23		99.20		94.18	

表 10: 18 子音の認識率 — closed speaker の場合

Phoneme	speaker						Recognition Rate (%)
	MAU	MIT	MMS	MMY	FKN	FSU	
b	95.20	98.60	95.10	84.10	84.50	82.40	89.98
d	95.00	98.80	91.50	95.30	91.20	90.00	93.63
g	87.70	91.70	79.80	81.80	90.10	69.10	83.37
p	100.00	100.00	78.60	80.00	86.70	93.30	89.77
t	97.50	97.30	92.50	92.00	89.30	95.90	94.08
k	94.80	93.00	94.00	93.40	91.80	96.20	93.87
m	94.20	97.30	95.00	93.90	92.50	94.00	94.48
n	89.40	91.70	81.60	80.50	87.20	83.40	85.63
N	95.50	93.50	96.00	99.80	93.40	98.30	96.08
s	98.20	99.00	96.40	93.10	88.40	98.80	95.65
sh	96.00	97.80	96.30	93.50	92.10	100.00	95.95
h	84.20	90.20	93.70	92.10	94.20	85.70	90.02
z	98.70	98.90	90.70	98.10	95.00	76.00	92.90
ch	74.60	97.10	85.70	84.80	85.50	100.00	87.95
ts	70.10	88.70	92.10	78.00	97.80	78.50	84.20
r	81.00	87.20	83.60	92.60	60.60	81.00	81.00
w	94.90	93.80	91.10	90.00	84.80	98.70	92.22
y	93.10	96.60	97.80	93.40	47.80	41.70	78.40
Average	91.80	94.50	91.70	91.50	86.30	88.20	90.67

表 11: 18 子音の認識率 — open speaker の場合

Phoneme & Phoneme Group	speaker								Recognition Rate (%)
	MNM	MTK	MTM	MTT	FAF	FKM	FYM	FYN	
b	71.80	85.50	77.30	81.20	67.10	54.60	67.60	83.70	73.60
d	94.90	95.30	96.50	87.60	93.60	62.90	95.30	97.10	90.40
g	71.70	56.50	65.60	74.30	43.70	44.30	68.40	53.40	59.74
p	53.30	73.30	53.30	66.70	80.00	66.70	66.70	73.30	66.66
t	69.80	80.00	63.20	77.30	74.50	83.00	75.00	94.80	77.20
k	91.00	88.20	91.00	95.80	78.00	67.80	97.40	92.80	87.75
m	87.80	72.40	56.20	93.40	51.70	47.50	89.80	81.70	72.56
n	69.40	54.30	65.70	60.80	78.90	80.40	94.70	84.40	73.58
N	85.40	98.90	94.30	96.60	93.30	87.90	99.80	94.30	93.81
s	87.40	95.80	87.80	86.60	86.00	89.80	97.80	81.00	89.02
sh	71.30	89.10	85.50	98.90	98.90	70.50	89.70	97.70	87.70
h	60.00	79.50	69.30	70.50	55.30	84.20	96.90	92.90	76.08
z	94.80	85.60	93.60	94.00	91.00	55.30	95.90	85.20	86.92
ch	95.40	94.20	89.90	97.20	95.70	95.70	68.60	62.70	87.43
ts	93.50	93.40	93.10	97.30	95.80	86.20	35.10	87.50	85.24
r	76.00	88.80	81.80	88.20	24.80	64.80	91.20	75.40	73.88
w	68.40	94.90	79.50	77.30	79.50	56.80	96.30	84.00	79.59
y	93.30	96.10	88.80	90.90	25.30	26.10	99.40	66.10	73.25
Average	81.20	84.50	79.60	86.60	69.50	69.50	88.20	84.40	80.44

5 考察

各実験における open speaker と closed speaker の音素認識率をまとめるて表 12に示す。結果より、不特定話者の音素認識においても、特定話者及び Multi-Speaker 同様に高認識率が得られた。

表 12: 不特定話者の認識実験

Phoneme	recognition rate (%)	
	closed speaker	open speaker
6 子音 /b, d, g, m, n, N/	96.9	86.8
5 母音 /a, i, u, e, o/	98.9	93.8
18 子音	90.7	80.4

表 13に各実験の音素毎の認識率を示したが、音素数の増加に伴い、コンフュージョンの程度が大きくなっていることがわかる。音素数を増やすことにより特に認識率が低下する音素は、/b/, /g/, /m/, /r/, /w/, /y/であった。話者毎に、その様子を見ると (Appendix A, B参照)、コンフュージョンの起こり方は同様であってもその程度が異なっている。18 子音認識の実験の際、音素カテゴリ毎に各話者に対して学習済みのネットワークを、Modular TDNN として構成してチューニングを行ない、更に、ネットワーク全体を構成し部分的学習により最終的な学習を行なった。これに対し、音素カテゴリ識別用のモジュールに関しては、各話者用のネットワークの規模が大きかったため、Modular TDNN としてチューニングは行なわれなかった。そのためもあって、カテゴリ間のコンフュージョンが大きいことがわかる。逆に音素カテゴリ認識用のモジュールについても、Modular TDNN 構造をとることにより高認識率を実現出来ると考えられる。

当然のことながら、各音素認識率のばらつきは、closed speaker より open speaker のばらつき方が大きく、特に女性話者のばらつきが大きいことがわかる。これは、学習話者が男性 4 名、女性 2 名の計 6 名とサンプルに偏りがあつたためと考えられる。

本研究では、不特定話者の音素認識を行なう方法として、多数話者学習を試みた。18 子音認識として 6 話者を学習に用いたが、ネットワークの規模・学習サンプルにおいて実験条件が限られ十分とはいえなかった。今回は、実験話者を無作為に選んだが、更に多くの学習話者を用いる際には、単純に学習話者数分のモジュールを用意するわけにはいかない。その場合、予め学習話者を何らかの方法で分類し、各話者カテゴリの代表話者、或いはそのカテゴリの特徴を反映した学習サンプルによりネットワークの学習を行なうべきであろう。それにより、多数話者学習を実現するとともにネットワークの規模や学習サンプル数を抑えることが出来るであろう。

表 13: 不特定話者の認識実験

Phoneme & Phoneme Group	recognition rate (%)						
	speaker dependent	closed speaker			open speaker		
		Modular- TDNN	/b,d,g,m,n,N/ Reco. network	18cons. Reco.network	Modular- TDNN	/b,d,g,m,n,N/ Reco. network	18cons. Reco.network
b	98.1	97.1	96.2	90.0	89.1	90.7	73.6
d	98.5	96.8	96.7	93.6	94.3	87.4	90.4
g	98.3	98.7	96.0	83.4	89.3	84.0	59.7
/b,d,g/	98.3	97.7			90.6		
p	86.5	78.6		89.8	66.7		66.7
t	98.2	98.5		94.1	90.5		77.2
k	98.5	98.8		93.9	94.8		87.8
/p,t,k/	98.2	98.4			92.4		
m	98.2	97.8	97.9	94.5	86.6	85.5	72.6
n	96.4	92.6	93.2	85.6	71.0	71.9	73.6
N	99.3	99.2	99.0	96.1	98.0	96.9	93.8
/m,n,N/	98.2	97.2			87.2		
/b,d,g,m,n,N/			96.9			86.8	
s	99.7	99.6		95.7	96.4		89.0
sh	99.8	99.5		96.0	95.1		87.7
h	99.4	99.2		90.0	94.6		76.1
z	99.6	99.3		92.9	97.0		86.9
/s,sh,h,z/	99.6	99.4			95.9		
ch	100.0	100.0		88.0	99.8		87.4
ts	100.0	99.8		84.2	99.8		85.2
/ch,ts/	100.0	99.8			99.8		
r	99.6	99.3		81.0	98.2		73.9
w	99.1	98.1		92.2	96.7		79.6
y	97.4	95.0		78.4	94.0		73.3
/r,w,y/	99.1	98.2			97.1		
18 consonants				90.7			80.4
a	99.5	99.3			93.7		
i	99.5	99.4			96.7		
u	98.6	97.9			91.9		
e	99.4	98.8			92.5		
o	99.2	99.2			94.2		
/a,i,u,e,o/	99.2	98.9			93.8		

6 むすび

ニューラルネットワークを用いた多数話者学習による不特定話者の音素認識を行なった。有声破裂音 /b,d,g/ の音素認識率が6話者学習時に 91.3% で、あったのに対し、認識音素を6子音 /b,d,g,m,n,N/、5母音 /a,i,u,e,o/、18子音へと拡張した結果は、以下のようになった。

- 6子音 /b,d,g,m,n,N/ 86.8%
- 5母音 /a,i,u,e,o/ 93.7%
- 18子音 80.4%

大規模なネットワーク、膨大な学習サンプルを必要とする欠点はあるものの不特定話者においても高い認識率が示された。

今後の課題

- 23 音素認識の実現
- 文節発声に対する音素認識実験

謝辞

研究の機会を与えて下さったATR自動翻訳電話研究所の樽松社長に深く感謝致します。また、御指導・助言を頂いた小森氏、鷹見氏、福沢氏をはじめとする音声情報処理研究室及び、データ処理研究室の皆様へ感謝致します。

参考文献

- [1] A. Waibel, T. Hanazawa, G. Hinton, K. Sikano and K. J. Lang, "Phoneme Recognition Using Time-Delay Neural Networks", IEEE Trans. Acoustics, Speech and Signal Processing (March 1989).
- [2] A. Waibel, H. Sawai, K. Sikano, "Consonant Recognition by Modular Construction of Large Phoneme Time-Delay Neural Networks", Proceedings of the 1989 IEEE International Conference on Acoustics, Speech and Signal Processing, May 1989, pp.112-115, (May 1989).
- [3] Hampshire, J., Waibel, A. The Meta-pi Network: Connectionist Rapid Adaptation for High-Performance Multi-Speaker Phoneme Recognition, Proceedings of the 1990 IEEE International Conference on Acoustics, Speech and Signal Processing, S3.9, pp.164-168 (1990).
- [4] 沢井, 中村, 福沢, 杉山, ニューラルネットワークによる不特定話者音声認識へのアプローチ法について, 音響学会講演論文集, 1-5-17, pp.37-38 (1991-03).
- [5] 中村, 沢井, "不特定話者の音素認識", 音声研究会, SP90-61 (Dec. 1990).
- [6] Rumelhart, D. E., McClelland, J. L., et al., "Parallel Distributed Processing", Vol. 1. Cambridge, MA.: MIT Press (1987).
- [7] P. Haffner, A. Waibel, K. Shikano, Fast Back-Propagation Learning Methods for Neural Networks in Speech, Rec. Fall Meet. Acoust. Soc. Jpn. 2-P-1, pp.203-204 (Oct. 1988).

A 各話者の認識率

実験の過程において得られた、各話者毎の各音素、音素カテゴリの認識率を示す。表 14に Single-TDNN による特定話者音素認識実験、表 15に Modular-TDNN による Multi-Speaker 音素認識実験、表 16に先の Modular-TDNN による不特定話者音素認識実験の結果を示す。

表 14: Single-TDNN を用いた特定話者の音素認識 (各音素と音素グループの認識率)

Phoneme &	Speaker						Recognition
Phoneme Group	MAU	MHT	MMS	MMY	FKN	FSU	Rate
b	99.10	99.00	97.80	94.70	98.60	99.50	98.12
d	97.80	99.40	95.50	100.00	99.40	98.80	98.48
g	98.80	99.60	97.60	98.00	98.80	96.70	98.25
/b,d,g/	98.60	99.40	97.10	97.40	98.90	98.20	98.27
p	93.30	86.70	85.70	80.00	80.00	93.30	86.50
t	98.90	97.70	97.00	98.20	99.50	97.90	98.20
k	99.60	98.40	99.00	98.20	96.60	99.00	98.47
/p,t,k/	99.20	97.90	97.90	97.90	97.70	98.40	98.17
m	98.30	99.00	98.50	97.30	97.70	98.50	98.22
n	96.60	98.10	94.40	95.80	95.50	98.10	96.42
N	99.60	99.80	98.10	99.80	98.40	99.80	99.25
/m,n,N/	98.50	99.10	97.50	97.80	97.50	98.90	98.22
s	99.80	100.00	98.70	100.00	99.60	100.00	99.68
sh	100.00	100.00	99.40	100.00	99.40	100.00	99.80
h	100.00	98.80	100.00	99.20	100.00	98.30	99.38
z	100.00	100.00	99.10	100.00	99.50	99.10	99.62
/s,sh,h,z/	99.90	99.70	99.20	99.80	99.70	99.40	99.62
ch	100.00	100.00	100.00	100.00	100.00	100.00	100.00
ts	100.00	100.00	100.00	100.00	100.00	100.00	100.00
/ch,ts/	100.00	100.00	100.00	100.00	100.00	100.00	100.00
r	100.00	99.80	98.60	99.80	100.00	99.60	99.63
w	100.00	100.00	97.50	98.60	98.70	100.00	99.13
y	100.00	99.40	97.20	92.20	98.30	97.60	97.45
/r,w,y/	100.00	99.70	98.20	98.00	99.50	99.20	99.10
a	99.80	99.70	99.00	99.80	98.80	99.70	99.47
i	99.30	99.80	99.00	99.50	99.70	99.50	99.47
u	97.30	99.30	97.20	99.20	99.50	99.20	98.62
e	99.50	99.30	98.70	99.80	99.50	99.50	99.38
o	98.80	100.00	97.80	100.00	99.00	99.80	99.23
/a,i,u,e,o/	99.00	99.60	98.30	99.70	99.30	99.50	99.23

表 15: Modular-TDNN を用いた Multi-Speaker の音素認識 (各音素と音素グループの認識率)

Phoneme & Phoneme Group	Speaker						Recognition
	MAU	MHT	MMS	MMY	FKN	FSU	Rate
b	97.80	98.60	98.20	95.70	95.70	96.60	97.10
d	98.90	99.40	96.60	94.10	94.10	97.60	96.78
g	97.60	99.20	98.40	99.60	99.60	97.60	98.67
/b,d,g/	98.00	99.10	97.90	96.80	96.80	97.30	97.65
p	100.00	86.70	71.40	66.70	66.70	80.00	78.58
t	99.30	99.30	99.10	96.60	98.20	98.60	98.52
k	98.80	98.80	99.40	99.60	97.20	99.00	98.80
/p,t,k/	99.10	98.80	98.80	97.70	97.20	98.50	98.35
m	97.10	98.10	99.00	96.00	98.30	98.50	97.83
n	93.20	95.80	90.30	90.00	94.30	91.70	92.55
N	99.00	99.80	99.20	99.80	97.50	100.00	99.22
/m,n,N/	97.00	98.20	97.10	96.00	97.10	97.50	97.15
s	99.80	100.00	98.90	99.80	99.40	99.80	99.62
sh	100.00	100.00	98.80	99.40	98.90	100.00	99.52
h	100.00	98.80	100.00	99.60	99.60	97.20	99.20
z	100.00	100.00	98.70	100.00	99.10	98.20	99.33
/s,sh,h,z/	99.90	99.70	99.10	99.70	99.30	98.90	99.43
ch	100.00	100.00	100.00	100.00	100.00	100.00	100.00
ts	99.40	100.00	99.10	100.00	100.00	100.00	99.75
/ch,ts/	99.60	100.00	99.40	100.00	100.00	100.00	99.83
r	99.20	99.40	98.60	100.00	99.00	99.80	99.33
w	100.00	98.80	94.90	97.10	97.50	100.00	98.05
y	99.40	100.00	99.40	98.80	95.50	76.80	94.98
/r,w,y/	99.30	99.50	98.40	99.50	98.00	94.60	98.22
a	100.00	99.30	98.80	99.70	98.30	99.50	99.27
i	99.00	99.80	99.50	99.50	99.30	99.30	99.40
u	97.50	97.20	96.80	99.20	99.00	97.70	97.90
e	98.20	99.00	98.20	99.20	98.80	99.20	98.77
o	99.80	100.00	98.00	98.20	99.50	99.70	99.20
/a,i,u,e,o/	98.90	99.10	98.30	99.10	99.00	99.10	98.92

表 16: Modular TDNN を用いた未知話者の音素認識 (各音素と音素グループの認識率)

Phoneme & Phoneme Group	Speaker								Recognition Rate
	MNM	MTK	MTM	MTT	FAF	FKM	FYM	FYN	
b	81.90	92.30	91.80	93.70	90.30	81.20	84.10	97.20	89.06
d	96.60	97.60	99.40	79.40	95.90	98.20	90.60	96.50	94.28
g	99.20	84.20	96.00	99.20	77.60	71.90	91.70	94.40	89.27
/b,d,g/	92.70	90.50	95.60	92.10	86.70	82.10	88.90	95.90	90.56
p	73.30	80.00	66.70	73.30	86.70	53.30	40.00	60.00	66.66
t	88.90	82.50	87.50	89.30	93.60	89.50	94.50	98.20	90.50
k	95.20	95.60	98.00	98.80	90.40	84.40	98.60	97.40	94.80
/p,t,k/	91.90	89.30	92.70	94.00	91.80	86.30	95.80	97.20	92.38
m	98.30	96.50	86.10	99.80	85.50	61.20	70.50	94.60	86.56
n	61.50	54.70	72.10	50.60	78.50	89.80	85.00	76.00	71.03
N	95.40	99.30	98.60	97.10	97.50	99.30	99.80	96.80	97.97
/m,n,N/	89.10	88.20	87.60	87.50	88.40	81.20	84.60	91.30	87.24
s	91.00	99.80	91.20	98.40	98.00	99.20	94.20	99.20	96.38
sh	83.20	98.90	98.80	100.00	100.00	93.40	86.40	100.00	95.09
h	94.50	99.60	95.30	92.60	80.20	97.40	98.80	98.00	94.55
z	100.00	99.50	100.00	99.10	99.10	91.20	100.00	87.00	96.99
/s,sh,h,z/	92.30	99.60	94.90	97.50	94.60	96.40	95.10	96.80	95.90
ch	100.00	100.00	100.00	98.60	100.00	100.00	100.00	100.00	99.83
ts	99.40	100.00	100.00	98.90	100.00	100.00	100.00	100.00	99.79
/ch,ts/	99.60	100.00	100.00	98.80	100.00	100.00	100.00	100.00	99.80
r	99.60	100.00	97.00	99.80	96.80	95.20	97.60	99.40	98.18
w	96.20	93.70	100.00	98.70	100.00	85.10	100.00	100.00	96.71
y	99.40	98.30	97.80	94.50	83.10	87.30	100.00	91.40	93.97
/r,w,y/	99.20	98.90	97.50	98.50	93.90	92.40	98.40	97.60	97.05
a	96.20	99.20	75.00	93.80	99.20	99.20	92.70	94.00	93.66
i	99.80	97.20	99.50	95.00	93.50	90.00	99.30	99.30	96.70
u	95.70	89.70	96.20	95.80	96.50	66.70	96.20	98.00	91.85
e	90.30	99.30	96.30	93.30	86.80	88.80	94.70	90.20	92.46
o	98.00	94.80	98.50	91.50	99.70	80.50	94.70	95.70	94.18
/a,i,u,e,o/	96.00	96.00	93.10	93.90	95.10	85.00	95.50	95.40	93.75

B コンフュージョンマトリックス

18 子音認識実験における、各話者のコンフュージョンマトリックスを以下に示す。

closed Speaker

MAU

216	2	2	3			1	1			1			1			::	11/	227 = 95.2
1	170	5	1	2												::	9/	179 = 95.0
3	1	221	2			10	2	2					2	1	8	::	31/	252 = 87.7
			15													::	0/	15 = 100.0
			5	429	5					1						::	11/	440 = 97.5
			5	4	474	3	2	2		1	2	3	4			::	26/	500 = 94.8
		3				453	12			2			11			::	28/	481 = 94.2
1		5	1			14	237			3			4			::	28/	265 = 89.4
1		6				11	2	466							2	::	22/	488 = 95.5
				2				491	3				4			::	9/	500 = 98.2
				1				170	4		1		1			::	7/	177 = 96.0
			1	2		9				223		8	9	1	12	::	42/	265 = 84.2
										224	2	1				::	3/	227 = 98.7
								18			53					::	18/	71 = 74.6
								53				124				::	53/	177 = 70.1
2	1	30	1	1	14	9	6	5		7	15		405	4		::	95/	500 = 81.0
		3								1				74		::	4/	78 = 94.9
		1	2			1				1	1		6	162		::	12/	174 = 93.1

Total Error:: 409/ 5016 = 91.8

test.MHT.best

205	2	1												::	3/	208 = 98.6
1	168		1											::	2/	170 = 98.8
	233		2	8	2			3				6	::	21/	254 = 91.7	
		15												::	0/	15 = 100.0
	1	7	428	4										::	12/	440 = 97.3
1		4	3	465	11		8			4	4			::	35/	500 = 93.0
	2	3		469	4	1				3				::	13/	482 = 97.3
	2	1		5	243					14				::	22/	265 = 91.7
1	5			1	3	417		2		11	6		::	29/	446 = 93.5	
					1	495	1			3				::	5/	500 = 99.0
						174	1	3						::	4/	178 = 97.8
			4	2	1	14	230	1		3				::	25/	255 = 90.2
			1		1			175						::	2/	177 = 98.9
							2		67					::	2/	69 = 97.1
					21					165				::	21/	186 = 88.7
4	1	10		1	17	17	2	5	6	436	1			::	64/	500 = 87.2
		3	1								75	1		::	5/	80 = 93.8
		1		1						4		171		::	6/	177 = 96.6

Total Error:: 271/ 4902 = 94.5

test.MMS.best

212		3	4							4		::	11/	223 = 95.1		
6	161	3	4					2				::	15/	176 = 91.5		
1		202			2	14		5	1	1	2	25	::	51/ 253 = 79.8		
			11	3									::	3/ 14 = 78.6		
1	2	1	20	407	7				1			1	::	33/ 440 = 92.5		
1		2	4	3	470			6	1	4	3	2	::	30/ 500 = 94.0		
		4	2		1	458	7	1		1	8		::	24/ 482 = 95.0		
1		10				25	218			13			::	49/ 267 = 81.6		
		3			2	10	1	458	3				::	19/ 477 = 96.0		
	1	1		1				3	431	1		7	1	1	::	16/ 447 = 96.4
1								156	2	1	2		::	6/ 162 = 96.3		
2					3			236		3	2	1	3	2	::	16/ 252 = 93.7
1			2		10	1	2	2		204		1	2		::	21/ 225 = 90.7
					1				2	5	48		::	8/ 56 = 85.7		
	1		1	1				4		2	105		::	9/ 114 = 92.1		
6		12	9	1	8	15	13	6	3	3		418	1	5	::	82/ 500 = 83.6
		4					2						72	1	::	7/ 79 = 91.1
		2										2	174		::	4/ 178 = 97.8

Total Error:: 404/ 4845 = 91.7

test.MMY.best

174	13	9	4						6							1	::	33/	207 = 84.1	
2	162	1	2	2											1		::	8/	170 = 95.3	
		1	202	2		8	18	3	9						1	3	::	45/	247 = 81.8	
				12	1	2											::	3/	15 = 80.0	
			1	5	405	17							4		7	1	::	35/	440 = 92.0	
			1	1	1	467	5			1			11		4	8	1	::	33/	500 = 93.4
			4	1			450	15							6	3	::	29/	479 = 93.9	
1		2					23	210					1		24		::	51/	261 = 80.5	
							1		404								::	1/	405 = 99.8	
										457	1		1		32		::	34/	491 = 93.1	
											157	5		6			::	11/	168 = 93.5	
						3	3	1		1		233		1	4	7	::	20/	253 = 92.1	
1				1											208		::	4/	212 = 98.1	
											10			56			::	10/	66 = 84.8	
											34	2		1	131		::	37/	168 = 78.0	
	5		2	13			7	4	1				4		463	1	::	37/	500 = 92.6	
	1						2	3					1			63	::	7/	70 = 90.0	
					1										10	155	::	11/	166 = 93.4	

Total Error:: 409/ 4818 = 91.5

:::::::::::::
 test.MNM.best
 :::::::::::::::

155	33	14	3	2	1	6													1	::	61/	216 = 71.8						
	2	168	1	2	3	1															::	9/	177 = 94.9					
12	2	182	1	1	36	15	1	1			2									1	::	72/	254 = 71.7					
	1			8	3	1														1	1	::	7/	15 = 53.3				
		49	2	3	307	78															1	::	133/	440 = 69.8				
				3	29	455	2	3	1												1	6	::	45/	500 = 91.0			
8		22	5			423	22	1													1	::	59/	482 = 87.8				
19		10	1	3		43	184															5	::	81/	265 = 69.4			
	1		5			6	2	393														48	1	4	::	67/	460 = 85.4	
								436														10	53		::	63/	499 = 87.4	
					4					119	43											1			::	48/	167 = 71.3	
			2		5	8	3			9	153	8	6	55	2	4									::	102/	255 = 60.0	
	1									1	1		202	7	1										::	11/	213 = 94.8	
					2						1												62		::	3/	65 = 95.4	
										9													2	157		::	11/	168 = 93.5
15		35	3	1	5	11	8	3				14	23										380	2	::	120/	500 = 76.0	
			14	6		1	3						1											54		::	25/	79 = 68.4
			5	4				1						1										1	166	::	12/	178 = 93.3

Total Error:: 929/ 4933 = 81.2

:::::::::::::
 test.MTK.best
 ::::::::::::::

177	14	2	9			1	1			1				2	::	30/	207 = 85.5			
1	162		1	6											::	8/	170 = 95.3			
42	23	143	8			2	21			5				1	7	1	::	110/	253 = 56.5	
			11			2								1			::	4/	15 = 73.3	
	4		18	352	50									3	4		::	88/	440 = 80.0	
1			31	7	441					1				4	7		::	59/	500 = 88.2	
52		11	37			349	8	3						3	11		::	133/	482 = 72.4	
20		8	18			72	144	1									::	121/	265 = 54.3	
		1					3	430								1	::	5/	435 = 98.9	
				1				479						16	4		::	21/	500 = 95.8	
					1				163	6		12		1			::	20/	183 = 89.1	
		4	4			2				205					43		::	53/	258 = 79.5	
8	2	6	1			3	3	2			2	185		2	2		::	31/	216 = 85.6	
											3	65	1				::	4/	69 = 94.2	
								12						170			::	12/	182 = 93.4	
9		11	1			5	19	3			2	3		444	2	1	::	56/	500 = 88.8	
		3			1											75		::	4/	79 = 94.9
		3												4		171	::	7/	178 = 96.1	

Total Error:: 766/ 4932 = 84.5

:::::::::::::
 test.MTM.best
 ::::::::::::::

160	28		7						3					2				4	2	1	::	47/	207 = 77.3
	1	164							4										1		::	6/	170 = 96.5
9	16	166	15	3	5	16	3	7						5	1			1	2	4	::	87/	253 = 65.6
	1	1		8	3	2															::	7/	15 = 53.3
		104		4	278	54															::	162/	440 = 63.2
			2	6	3	455								31		2				1	::	45/	500 = 91.0
3	1	3	89		1	271	85	1						4			19	5			::	211/	482 = 56.2
	1		3	46			20	174	3					1			16		1		::	91/	265 = 65.7
	2		3				6	2	415					3	4		2		3		::	25/	440 = 94.3
							1		439	5	2	7					46				::	61/	500 = 87.8
					1				148	19		5									::	25/	173 = 85.5
1		2	1	28	1		1		2	178	8	11	23						1		::	79/	257 = 69.3
		2	2				6	3			1	204									::	14/	218 = 93.6
										6	1		62								::	7/	69 = 89.9
									5	1	2	3	1	163							::	12/	175 = 93.1
4	1	11	19		11	13	6	3			11					409	8	4		::	91/	500 = 81.8	
			3	4		1		3			4							58			::	15/	73 = 79.5
			1	16													3		158		::	20/	178 = 88.8

Total Error:: 1005/ 4915 = 79.6

:::::::::::::
 test.MTT.best
 ::::::::::::::

168	3	7				17	1			3	4		2	2	::	39/	207 = 81.2	
12	149	3	2	1		1				2					::	21/	170 = 87.6	
3	8	188	5			1	40		4		2		1	1	::	65/	253 = 74.3	
			10	2	3										::	5/	15 = 66.7	
	4		39	340	52					3	1		1		::	100/	440 = 77.3	
				3	479	11					1	2	3	1	::	21/	500 = 95.8	
		3			5	450	3			1			18	2	::	32/	482 = 93.4	
2		12			4	65	161			2	8		11		::	104/	265 = 60.8	
		3				8	395						3		::	14/	409 = 96.6	
								433	6				61		::	67/	500 = 86.6	
									180	1	1				::	2/	182 = 98.9	
1				22	5	1		7	1	182	4	8	15	12	::	76/	258 = 70.5	
1	1	7		1				1		2	203				::	13/	216 = 94.0	
									2			69			::	2/	71 = 97.2	
								5				181			::	5/	186 = 97.3	
5		17			7	9	2			8	7		441	4	::	59/	500 = 88.2	
		7				10								58	::	17/	75 = 77.3	
					1	1				1			11	1	150	::	15/	165 = 90.9

Total Error:: 657/ 4894 = 86.6

:::::::::::::
 test.FKM.best
 ::::::::::::::

113	22	3	22		29	2				4		5		7	::	94/	207 = 54.6		
1	107		12	26	12					7		4	1		::	63/	170 = 62.9		
10	18	112	21	4	41	6	3	14		1	2	3	10	2	3	3	::	141/	253 = 44.3
			10	3	1								1		::	5/	15 = 66.7		
	1		42	365	12					9		3	5		3	::	75/	440 = 83.0	
			64	46	339	1		1		9		31	9		::	161/	500 = 67.8		
16		9	18	1	22	229	152	18	1	2	6		2		1	5	::	253/	482 = 47.5
1		12	2		9	19	213	2			4	1	1		1		::	52/	265 = 80.4
	2		9		9	1	1	356					1		26		::	49/	405 = 87.9
					1			449		3			47			::	51/	500 = 89.8	
								13	129	1		34	6			::	54/	183 = 70.5	
				15	1	1		6	1	223	1	2	13		2	::	42/	265 = 84.2	
2		2		38		3		19		12	120	10	11			::	97/	217 = 55.3	
									1		1	67	1			::	3/	70 = 95.7	
								22		4		162				::	26/	188 = 86.2	
3	8	5	2	1	72	18	12	23	2	1	11	12		324		6	::	176/	500 = 64.8
5		3	4		3	2	2				8			1	42	4	::	32/	74 = 56.8
	6	17	22		9	25	9	13	2	1	5	13				43	::	122/	165 = 26.1

Total Error:: 1496/ 4899 = 69.5

:::::::::::
test.FYM.best
:::::::::::

140	50	2	10	2	2	1									67/	207 =	67.6		
3	162	1		2	1					1					8/	170 =	95.3		
11	27	173	2	1	3	28	1	3				1			80/	253 =	68.4		
			10	4	1										5/	15 =	66.7		
1	29		66	330	10	2							2		110/	440 =	75.0		
	1		6	4	487	1				1					13/	500 =	97.4		
4		8	10			433	20				7				49/	482 =	89.8		
1		7	3			3	252								14/	266 =	94.7		
		1					438								1/	439 =	99.8		
							489			2				9		11/	500 =	97.8	
							1	165	12	5	1					19/	184 =	89.7	
								251	1				7			8/	259 =	96.9	
												2				9/	217 =	95.9	
										3	208	2	2			22/	70 =	68.6	
										9	13	48				120/	185 =	35.1	
										6	114			65		44/	500 =	91.2	
	3	3			7	9	10	2			10			456		3/	81 =	96.3	
		1						1					1		78		1/	178 =	99.4
		1												177					

Total Error:: 584/ 4946 = 88.2

:::::::::::
 test.FYN.best
 :::::::::::

180	11	3		1	2	16	2									35/	215 = 83.7
4	167		1													5/	172 = 97.1
9	6	134	1	1	15	46	16	14	1	8					117/	251 = 53.4	
			11	3					1						4/	15 = 73.3	
			15	417	3	1								4	23/	440 = 94.8	
1			15	12	464				5			1	2		36/	500 = 92.8	
3		12	4	1		394	66	1		1					88/	482 = 81.7	
		4	3			33	221			1					41/	262 = 84.4	
		3				8	9	414		5					25/	439 = 94.3	
									405	1	2				92	95/	500 = 81.0
									170						4	4/	174 = 97.7
		1		8		2	1		235		1	5			18/	253 = 92.9	
							18	5	1	184	1	7			32/	216 = 85.2	
								25				42			25/	67 = 62.7	
					1				20				154	1	22/	176 = 87.5	
3	1	9	1		13	46	28	6		9	2		377	1	4	123/	500 = 75.4
3						8	1							63		12/	75 = 84.0
1		5	7			28	6	11						1	115	59/	174 = 66.1

Total Error:: 764/ 4911 = 84.4