TR-I-0222

VQ Neural Network による教師なし話者適応
Unsupervised Speaker Adaptation Using VQ Neural Network

山本 雅章、福沢圭二、杉山雅英
Masaaki YAMAMOTO , Keiji FUKUZAWA , Masahide SUGIYAMA

1991.07

概要
教師あり話者適応化は、教師信号に基づいて適応化の写像を構成するため強力であるが、あらかじめ決められた発声を行なわなければならないと言う点で柔軟性に乏しい。これに対して、教師なし話者適応化は、任意の発声を基に適応を行なえるため柔軟性に富みその用途が広い。本研究ではVQとニューラルネットが共に歪み最小化の基準で動作している点に着目し、VQとニューラルネットを組み合わせた教師なし適応化方式の有効性を評価した。5母音による評価実験の結果、72.0%であった適応化前の認識率が、適応化によりの認識率94.4%となり、本手法の有効性が確認された。

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

目次

図目次

1 はじめに

音声認識システムは話者の違い、発話様式の変動、発話環境の変動といった未学習データに対する識別性能の劣化という問題を持つ。この内、話者の違いについては話者適応化により改善が行なわれている。教師付き話者適応化は、教師信号に基づいて写像を構成するため強力であるが、あらかじめ決められた発声を行なわなければならないと言う点で柔軟性に乏しい。これに対して、教師なし話者適応化は、任意の発声を基に適応を行なえるため柔軟性に富みその用途が広い。既に、ニューラルネットによる教師付き話者適応化の方法が提案され [1],[2]、その有効性が確認されているが、本研究では VQ とニューラルネットが共に歪み最小化の基準で動作している点に着目し、VQ とニューラルネットを組み合わせた教師なし話者適応化方式 [3] の有効性を評価した。

2 VQ Neural Network による教師なし話者適応

未知話者の入力ベクトルを標準話者の対応するカテゴリ内のコードベクトルへ写像できれば、話者適応を行なうベクトル量子化が実現できる。この写像を NN(Neural Network) を用いて行なう。教師信号無しに学習を行なう為、自動的に教師信号を生成する。自動的に教師信号を生成するアルゴリズムとして、VQ(Nearest-neighbor) の基準によるアルゴリズム、階層的クラスタリングに基づいたアルゴリズム、そしてこの 2 つを組み合わせたアルゴリズム、の 3 つのアルゴリズムについて検討をおこなった。

2.1 VQ(Nearest-neighbor) の基準によるアルゴリズム

Nearest-neighbor の基準によるアルゴリズムは以下に示すように、未知話者の入力ベクトルに対する教師信号を NN により写像されたベクトルに最も近い標準話者のコードブックを選んで与える (図 1)。

Nearest-neighbor based algorithm

1. N : final iteration number
   $g^{(1)}$: initial NN mapping function
   (g : NN mapping function)
2. $n = 1$
3. $v'_j = g^{(n)}(v_j)$
   (v_j : Unknown Speaker Vector)
4. $i = \underset{i}{\operatorname{argmin}}\, d(u_i, v'_j)$
   (u_i : Standard Speaker VQ Code)
5. $u_i = g^{(n)}(v_j)$: NN Training
6. $n = n + 1$, if $n < N$ goto 3, else stop

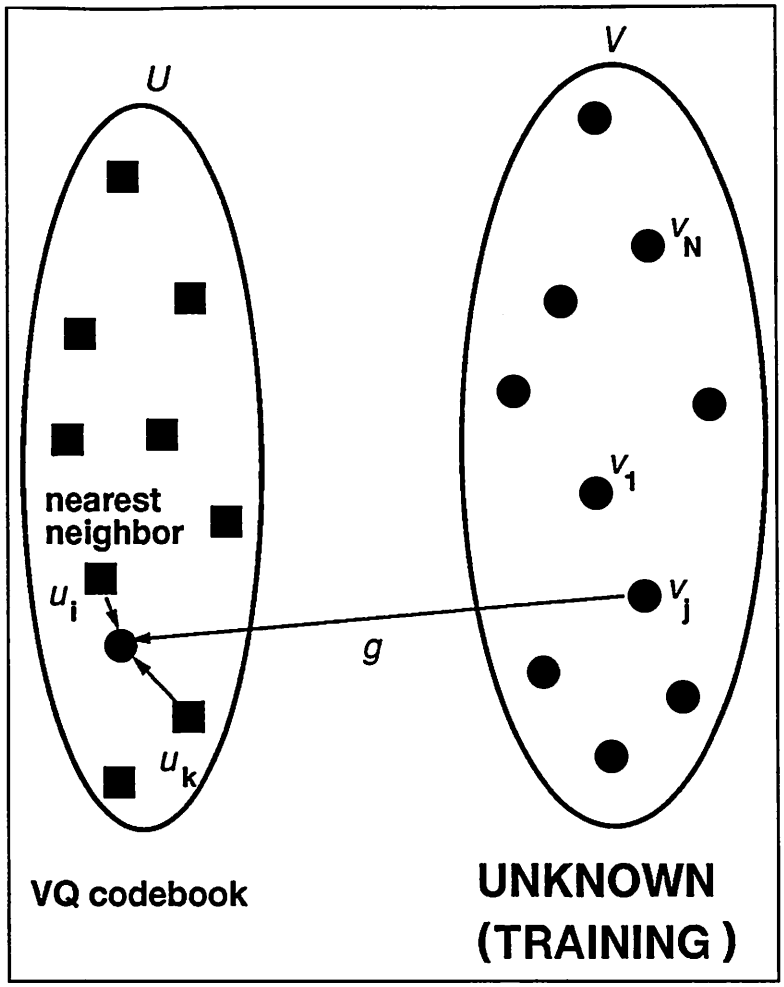

図 1: Nearest-neighbor based Algorithm

このアルゴリズムは、最初に NN による写像 ($g^{(1)}$) が好ましくない対応関係をつくり出した場合、その方向に学習が進んでゆくという問題点を持つ。従って、このアルゴリズムにおいては、NN の weighting parameter の初期値をいかに与えるかが重要となる。本研究では weighting parameter の初期値として以下の 3 通りの与え方を検討した。

- random weighting
- identity mapping [2]
- no mapping (set u_j equal to $\underset{i}{\operatorname{argmin}}\, d(u_i, v_j)$)

2.2 階層的クラスタリングに基づいたアルゴリズム

階層的クラスタリングに基づいたアルゴリズムを図 2に示す。NN の写像 g はクラスタリング手法 [4] により段階的に作られてゆく。このアルゴリズムにより大局的な特徴写像から精密な特徴写像の学習が行なわれる。以下にこのアルゴリズムの手順を示す。

Hierarchical-clustering based algorithm

1. n_s: initial cluster number, n_e: final cluster number
2. $n = n_s$
3. $V = V_1 \oplus \ldots \oplus V_n$, $V_i = \wp(\tilde{v}_i)$
   ($\wp$: division of V using $\tilde{v}_i$, $\oplus$: partition)
4. $\tilde{u}_i = C(U_i), U_i = \wp(g(\tilde{v}_i))$
   (C : centroid of a set)
   $U = \wp(g(\tilde{v}_1)) \oplus \ldots \oplus \wp(g(\tilde{v}_n))$
5. $g(v) = \{v - \sum_i w_i^* \tilde{v}_i\} + \{\sum_i w_i^* \tilde{u}_i\}$
   $w_i^* = \frac{d(v, \tilde{v}_i)^{-p}}{\sum_j d(v, \tilde{v}_j)^{-p}}$
6. if $n = n + 1$, $n < n_e$ goto 3, else stop

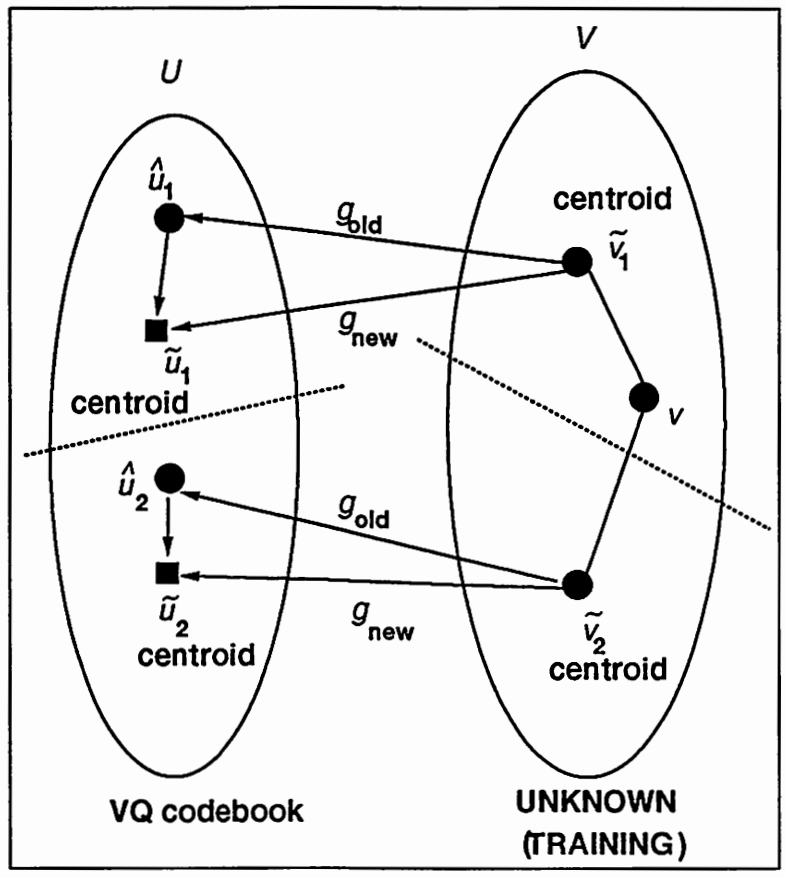

図 2: Hierarchical-clustering based algorithm ($n = 2$ case)

2.3 組合せアルゴリズム

前記2つのアルゴリズムを組み合わせることが可能である。始めに階層的クラスタリングに基き、カテゴリ間の重心移動を行なうようにNNの weight parameter を初期化する。次に Nearest-neighbor の基準によるアルゴリズムにより話者間の写像をより正確に行なえるようにNNの学習を重ねてゆく。

2.4 Neural Netowrk 構造

話者間の写像には、4層 feed-forward 型 NN を用いた。NN の構造を図3に示す。ネットワークの結合数は4216、入力層、隠れ層1、隠れ層2、出力層のユニット数は各々、16、50、50、16。

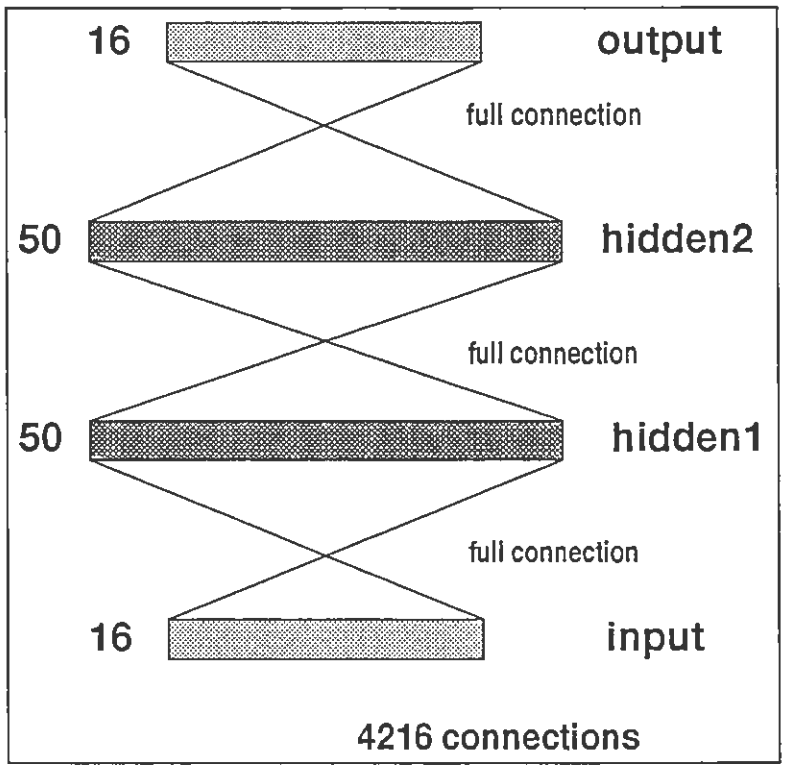

図 3: Speeker adaptation NN structure

3 評価実験

5母音を用いた実験により、教師なし話者適応アルゴリズムの評価を行なった。組合せアルゴリズムによる実験では、階層クラスタ数(n_e)を1として未知話者サンプルの重心移動を行ない、VQにより対応づけされたコードベクトルを教師信号としてNearest-neighborの基準に基づくNNの写像学習を開始した。また比較の為、未知話者と標準話者の同一単語発声データを用いた教師あり話者適応[1]による評価も行なった。

3.1 実験条件

実験条件を表1に示す。

表 1: 実験条件

話者	標準話者 - 男性1名(MAU)、未知話者 - 男性1名(MHT)	
認識タスク	5母音 /a,i,u,e,o/	
特徴パラメータ	16次 LPC ケプストラム	
分析条件	LPC分析次数	14次
	標本化周波数	12 kHz
	窓長	256 点 (21.3ms)
	窓関数	Hamming
サンプルデータ	コードブック作成	500 samples [100 samples × 5 vowels] (標準話者 5240 単語偶数番から、母音中心部を抽出)
	適応学習サンプル	500 samples [100 samples × 5 母音] (未知話者 5240 単語偶数番から、母音中心部を抽出.)
	テストサンプル	500 samples [100 samples × 5 母音] (未知話者 5240 単語奇数番から、母音中心部を抽出)
VQ コードブックサイズ	40 [8 samples × 5 母音] (母音カテゴリごとに VQ コードをつくり出す)	
NN の学習繰り返し回数	$N = 10$	
階層クラスタ数	$n_e = 1$	
教師あり学習データ	25 単語、5 単語 (両話者の 5240 単語偶数番から、音素バランスを考慮して選択)	

認識率は以下の式に従って、算出する。

$$R = \frac{N_s}{N_a} \times 100$$

R : 認識率 (%)
N_a : テストサンプル数
N_s : v_j と u_i が同じ音素カテゴリに含まれるサンプル数
(v_j : Unknown Speaker Vector , u_i : Standard Speaker VQ Coed)
音素カテゴリの対応 : $i = \underset{i}{\operatorname{argmin}}\, d(u_i, g^{(N)}(v_j))$ ($g^{(N)}$: NN mapping function)

3.2 実験結果

実験結果を表2に示す。また各教師なしアルゴリズムによる学習過程での歪みとカテゴリ間の対応の変化をAppendix Aに、認識時のカテゴリ間の対応をB適応前と適応後のスペクトルの変化をAppendix Cに示す。

実験の結果、組合せアルゴリズムによる認識率が94.4%となり、適応化前と比べて約20%の認識率の向上が得られた。また25単語、5単語による教師あり話者適応とほぼ同じ性能が得られることが明らかとなった。

表 2: Phoneme Recognition Accuracy

methods		recognition rate (%)
適応化前		72.0
教師なし適応	nearest-neighbor	
	• random weighting	19.8
	• identity mapping	76.8
	• no mapping	75.2
	階層的クラスタリング	87.0
	組合せアルゴリズム	94.4
教師あり適応	25 words	95.8
	5 words	94.2

4 まとめ

VQ とニューラルネットを組み合わせた教師なし適応化方式の有効性を評価した。5 母音による評価実験の結果、VQ(Nearest-neighbor) の基準によるアルゴリズムと階層的クラスタリングに基づいたアルゴリズムを組み合わせたアルゴリズムにより、72.0% であった適応化前の認識率が、94.4% となり、本手法の有効性が確認された。

今後の検討課題として以下の項目が挙げられる。

- 男女話者間での評価
- 子音、全音素での評価

謝辞

本報告は、1991 年 2 月 25 日から 4 月 5 日に ATR にアルバイト学生として来た京都工芸繊維大学の山本君の協力を得てなされたものである。報告者の不十分な指導にもかかわらず熱心に本研究を進めてくれた山本君に感謝します。また研究の機会を与えて頂いた樽松社長、貴重な御助言、御検討頂いた嵯峨山研究室長、VQ のソフトウェアを提供して頂いた服部氏、NN のソフトウェアを提供して頂いた沢井氏（現 リコー中央研究所情報エレクトロニクス研究センター）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] 磯, 他, ニューラルネットワークによる話者適応, 音学講論, 1-6-16(1989-3).

[2] 福沢, 他, ニューラルネットワークによる恒等写像を用いた話者適応, 音学講論, 1-8-16(1990-9).

[3] 杉山, 他, ニューラルネットによる集合間写像の教師なし話者適応, 音学講論, 2-P-10(1990-9).

[4] 白木, 誉田, 時空間パタン符号化における話者適応化, 音学講論, 3-6-9(1987-10).

Appendix

A 学習過程での歪みとカテゴリ間の対応の変化

```
[ Random_Weighting ]
Training Iteration N = 0  ; distance = 0.952226
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0  100    0    0
/i/ |   0    0  100    0    0
/u/ |   0   15   85    0    0
/e/ |   0    6   94    0    0
/o/ |   0    0  100    0    0
Recognition Rate = 17.0 %

Training Iteration N = 1  ; distance = 0.013367
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0  100    0    0
/i/ |   0    0  100    0    0
/u/ |   0    0  100    0    0
/e/ |   0    0  100    0    0
/o/ |   0    0  100    0    0
Recognition Rate = 20.0 %

Training Iteration N = 2  ; distance = 0.013367
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0  100    0    0
/i/ |   0    0  100    0    0
/u/ |   0    0  100    0    0
/e/ |   0    0  100    0    0
/o/ |   0    0  100    0    0
Recognition Rate = 20.0 %

Training Iteration N = 3  ; distance = 0.013367
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0  100    0    0
/i/ |   0    0  100    0    0
/u/ |   0    0  100    0    0
/e/ |   0    0  100    0    0
/o/ |   0    0  100    0    0
Recognition Rate = 20.0 %
```

[Identity_Mapping]
Training Iteration N = 0 ; distance = 0.769996

	/a/	/i/	/u/	/e/	/o/
/a/	5	0	40	0	55
/i/	0	97	1	2	0
/u/	0	0	86	3	11
/e/	0	19	5	76	0
/o/	0	0	13	0	87

Recognition Rate = 70.2 %

Training Iteration N = 1 ; distance = 0.159913

	/a/	/i/	/u/	/e/	/o/
/a/	0	0	37	0	63
/i/	0	98	0	2	0
/u/	0	0	91	3	6
/e/	0	5	0	95	0
/o/	0	0	12	0	88

Recognition Rate = 74.4 %

Training Iteration N = 2 ; distance = 0.099544

	/a/	/i/	/u/	/e/	/o/
/a/	0	0	36	0	64
/i/	0	98	0	2	0
/u/	0	0	97	2	1
/e/	0	1	0	99	0
/o/	0	0	12	0	88

Recognition Rate = 76.4 %

Training Iteration N = 3 ; distance = 0.076546

	/a/	/i/	/u/	/e/	/o/
/a/	0	0	36	0	64
/i/	0	98	0	2	0
/u/	0	0	99	1	0
/e/	0	0	0	100	0
/o/	0	0	11	0	89

Recognition Rate = 77.2 %

Training Iteration N = 4 ; distance = 0.063849

	/a/	/i/	/u/	/e/	/o/
/a/	0	0	36	0	64
/i/	0	98	0	2	0
/u/	0	0	99	1	0
/e/	0	0	0	100	0
/o/	0	0	11	0	89

Recognition Rate = 77.2 %

```
Training Iteration N = 5  ; distance = 0.054756
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   35    0   65
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   11    0   89
Recognition Rate = 77.2 %


Training Iteration N = 6  ; distance = 0.048647
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   34    0   66
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   11    0   89
Recognition Rate = 77.2 %

Training Iteration N = 7  ; distance = 0.044959
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   33    0   67
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   11    0   89
Recognition Rate = 77.2 %

Training Iteration N = 8  ; distance = 0.041328
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   33    0   67
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   11    0   89
Recognition Rate = 77.2 %

Training Iteration N = 9  ; distance = 0.038258
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   33    0   67
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   10    0   90
Recognition Rate = 77.4 %
```

```
Training Iteration N = 10  ; distance = 0.034798
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0   33    0   67
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   10    0   90
Recognition Rate = 77.4 %
```

```
[ No_Mapping ]
Training Iteration N = 0  ; distance = 0.938744
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   4    0   41    0   55
/i/ |   0   98    1    1    0
/u/ |   0    0   87    3   10
/e/ |   0   17    5   78    0
/o/ |   0    0    7    0   93
Recognition Rate = 72.0 %

Training Iteration N = 1  ; distance = 0.168154
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   34    0   66
/i/ |   0   98    0    2    0
/u/ |   0    0   93    3    4
/e/ |   0    5    0   95    0
/o/ |   0    0   11    0   89
Recognition Rate = 75.0 %

Training Iteration N = 2  ; distance = 0.100528
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   31    0   69
/i/ |   0   98    0    2    0
/u/ |   0    0   98    2    0
/e/ |   0    3    0   97    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.0 %

Training Iteration N = 3  ; distance = 0.067335
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   98    2    0
/e/ |   0    1    0   99    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.4 %

Training Iteration N = 4  ; distance = 0.054795
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.8 %
```

```
Training Iteration N = 5  ; distance = 0.046952
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.8 %

Training Iteration N = 6  ; distance = 0.042653
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.8 %

Training Iteration N = 7  ; distance = 0.039651
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.8 %

Training Iteration N = 8  ; distance = 0.036370
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.8 %

Training Iteration N = 9  ; distance = 0.033643
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |   0    0   30    0   70
/i/ |   0   98    0    2    0
/u/ |   0    0   99    1    0
/e/ |   0    0    0  100    0
/o/ |   0    0   13    0   87
Recognition Rate = 76.8 %

Training Iteration N = 10  ; distance = 0.030689
```

	/a/	/i/	/u/	/e/	/o/
/a/	0	0	30	0	70
/i/	0	98	0	2	0
/u/	0	0	99	1	0
/e/	0	0	0	100	0
/o/	0	0	13	0	87

Recognition Rate = 76.8 %

```
[ Combination ( Centroid_move ) ]
Training Iteration N = 0  ; distance  = 0.556155
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |  72    0   13    3   12
/i/ |   0   97    0    3    0
/u/ |   0    0   75   13   12
/e/ |   0    7    0   93    0
/o/ |   1    0    1    0   98
Recognition Rate = 87.0 %

Training Iteration N = 1  ; distance  = 0.151462
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |  90    0    5    0    5
/i/ |   0   98    0    2    0
/u/ |   0    0   84   11    5
/e/ |   0    1    0   99    0
/o/ |   0    0    3    0   97
Recognition Rate = 93.6 %

Training Iteration N = 2  ; distance  = 0.086092
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |  99    0    1    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   84   10    6
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.6 %

Training Iteration N = 3  ; distance  = 0.062460
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   83    9    8
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.6 %

Training Iteration N = 4  ; distance  = 0.052994
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   84    7    9
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.8 %
```

```
Training Iteration N = 5  ; distance  = 0.047608
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   83    7   10
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.6 %

Training Iteration N = 6  ; distance  = 0.042082
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   83    7   10
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.6 %

Training Iteration N = 7  ; distance  = 0.036651
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   82    7   11
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.4 %

Training Iteration N = 8  ; distance  = 0.032384
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   82    7   11
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.4 %

Training Iteration N = 9  ; distance  = 0.029073
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   82    7   11
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 95.4 %

Training Iteration N = 10  ; distance  = 0.026865
```

	/a/	/i/	/u/	/e/	/o/
/a/	100	0	0	0	0
/i/	0	98	0	2	0
/u/	0	0	82	7	11
/e/	0	0	0	100	0
/o/	0	0	3	0	97

Recognition Rate = 95.4 %

```
[ 25words_training ]
Training Iteration N = 0  ; distance = 0.249160
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   90    7    3
/e/ |   0    0    1   99    0
/o/ |   0    0    3    0   97
Recognition Rate = 96.8 %

Training Iteration N = 1  ; distance = 0.105524
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   93    7    0
/e/ |   0    0    0  100    0
/o/ |   0    0    9    0   91
Recognition Rate = 96.4 %

Training Iteration N = 2  ; distance = 0.075794
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   92    8    0
/e/ |   0    0    0  100    0
/o/ |   0    0   14    0   86
Recognition Rate = 95.2 %

Training Iteration N = 3  ; distance = 0.061342
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0   15    0   85
Recognition Rate = 94.8 %

Training Iteration N = 4  ; distance = 0.053138
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0   14    0   86
Recognition Rate = 95.0 %
```

```
Training Iteration N = 5  ; distance = 0.044148
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0   10    0   90
Recognition Rate = 95.8 %

Training Iteration N = 6  ; distance = 0.036441
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0    8    0   92
Recognition Rate = 96.2 %

Training Iteration N = 7  ; distance = 0.029809
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0    8    0   92
Recognition Rate = 96.2 %

Training Iteration N = 8  ; distance = 0.026780
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0    8    0   92
Recognition Rate = 96.2 %

Training Iteration N = 9  ; distance = 0.024939
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   91    9    0
/e/ |   0    0    0  100    0
/o/ |   0    0    8    0   92
Recognition Rate = 96.2 %

Training Iteration N = 10  ; distance = 0.023767
```

	/a/	/i/	/u/	/e/	/o/
/a/	100	0	0	0	0
/i/	0	98	0	2	0
/u/	0	0	91	9	0
/e/	0	0	0	100	0
/o/	0	0	8	0	92

Recognition Rate = 96.2 %

```
[ 5words_training ]
Training Iteration N = 0  ; distance = 0.355547
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |  74    0    0    0   26
/i/ |   0   92    0    8    0
/u/ |   0    0   80    8   12
/e/ |   0    2    1   97    0
/o/ |   0    0   15    0   85
Recognition Rate = 85.6 %

Training Iteration N = 1  ; distance = 0.137433
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ |  91    0    0    0    9
/i/ |   0   96    0    4    0
/u/ |   0    0   94    6    0
/e/ |   0    0    0  100    0
/o/ |   0    0   11    0   89
Recognition Rate = 94.0 %

Training Iteration N = 2  ; distance = 0.093629
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   95    5    0
/e/ |   0    0    0  100    0
/o/ |   0    0   14    0   86
Recognition Rate = 95.8 %

Training Iteration N = 3  ; distance = 0.072331
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   95    5    0
/e/ |   0    0    0  100    0
/o/ |   0    0   16    0   84
Recognition Rate = 95.4 %

Training Iteration N = 4  ; distance = 0.059688
       /a/  /i/  /u/  /e/  /o/
-------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   95    5    0
/e/ |   0    0    0  100    0
/o/ |   0    0   16    0   84
Recognition Rate = 95.4 %
```

```
Training Iteration N = 5  ; distance = 0.051651
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   94    6    0
/e/ |   0    0    0  100    0
/o/ |   0    0   16    0   84
Recognition Rate = 95.2 %

Training Iteration N = 6  ; distance = 0.045798
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   94    6    0
/e/ |   0    0    0  100    0
/o/ |   0    0   16    0   84
Recognition Rate = 95.2 %

Training Iteration N = 7  ; distance = 0.042359
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   94    6    0
/e/ |   0    0    0  100    0
/o/ |   0    0   16    0   84
Recognition Rate = 95.2 %

Training Iteration N = 8  ; distance = 0.039780
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   94    6    0
/e/ |   0    0    0  100    0
/o/ |   0    0   16    0   84
Recognition Rate = 95.2 %

Training Iteration N = 9  ; distance = 0.036664
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   94    6    0
/e/ |   0    0    0  100    0
/o/ |   0    0   17    0   83
Recognition Rate = 95.0 %

Training Iteration N = 10  ; distance = 0.034170
```

	/a/	/i/	/u/	/e/	/o/
/a/	100	0	0	0	0
/i/	0	98	0	2	0
/u/	0	0	94	6	0
/e/	0	0	0	100	0
/o/	0	0	17	0	83

Recognition Rate = 95.0 %

B 認識時のカテゴリ間の対応

```
********** Recognition **************

TEST DATA: /a,i,u,e,o/ ( from 5240 odd )

[ Without Adaptation ]
distance = 0.917500
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |  10    0   33    0   57
/i/ |   0   99    1    0    0
/u/ |   0    0   86    4   10
/e/ |   0   16    8   76    0
/o/ |   0    0   11    0   89
Recognition Rate = 72.0 %

[ Random weighting ]
distance = 0.013880
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0  100    0    0
/i/ |   0    0  100    0    0
/u/ |   0    1   99    0    0
/e/ |   0    0  100    0    0
/o/ |   0    0  100    0    0
Recognition Rate = 19.8 %

[ Identity mapping ]
distance = 0.045439
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0   31    0   69
/i/ |   0   98    0    2    0
/u/ |   0    0   96    4    0
/e/ |   0    0    1   99    0
/o/ |   0    0    9    0   91
Recognition Rate = 76.8 %

[ No mapping ]
distance = 0.043910
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ |   0    0   29    0   71
/i/ |   0   98    0    2    0
/u/ |   0    1   95    4    0
/e/ |   0    0    1   99    0
/o/ |   0    0   16    0   84
Recognition Rate = 75.2 %

[ 25 words Training ]
distance = 0.032982
```

```
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   98    0    2    0
/u/ |   0    0   86   13    1
/e/ |   0    0    0  100    0
/o/ |   0    0    5    0   95
Recognition Rate = 95.8 %

[ 5words Training ]
distance = 0.041672
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   99    0    1    0
/u/ |   0    0   92    8    0
/e/ |   0    0    1   99    0
/o/ |   0    0   19    0   81
Recognition Rate = 94.2 %

[ Combination ( Centroid move ) ]
distance = 0.042752
       /a/  /i/  /u/  /e/  /o/
------------------------------
/a/ | 100    0    0    0    0
/i/ |   0   97    0    3    0
/u/ |   0    0   78    8   14
/e/ |   0    0    0  100    0
/o/ |   0    0    3    0   97
Recognition Rate = 94.4 %
```

C 適応前と適応後のスペクトルの変化

適応前と適応後のスペクトルの変化の例を図4と図5に示す。図中、適応前の未知話者のスペクトルを20dB下げて表示している。また、対応づけされたVQ Codeのスペクトルを破線で表示し、残りのスペクトルが適応後の未知話者のスペクトルを表している。

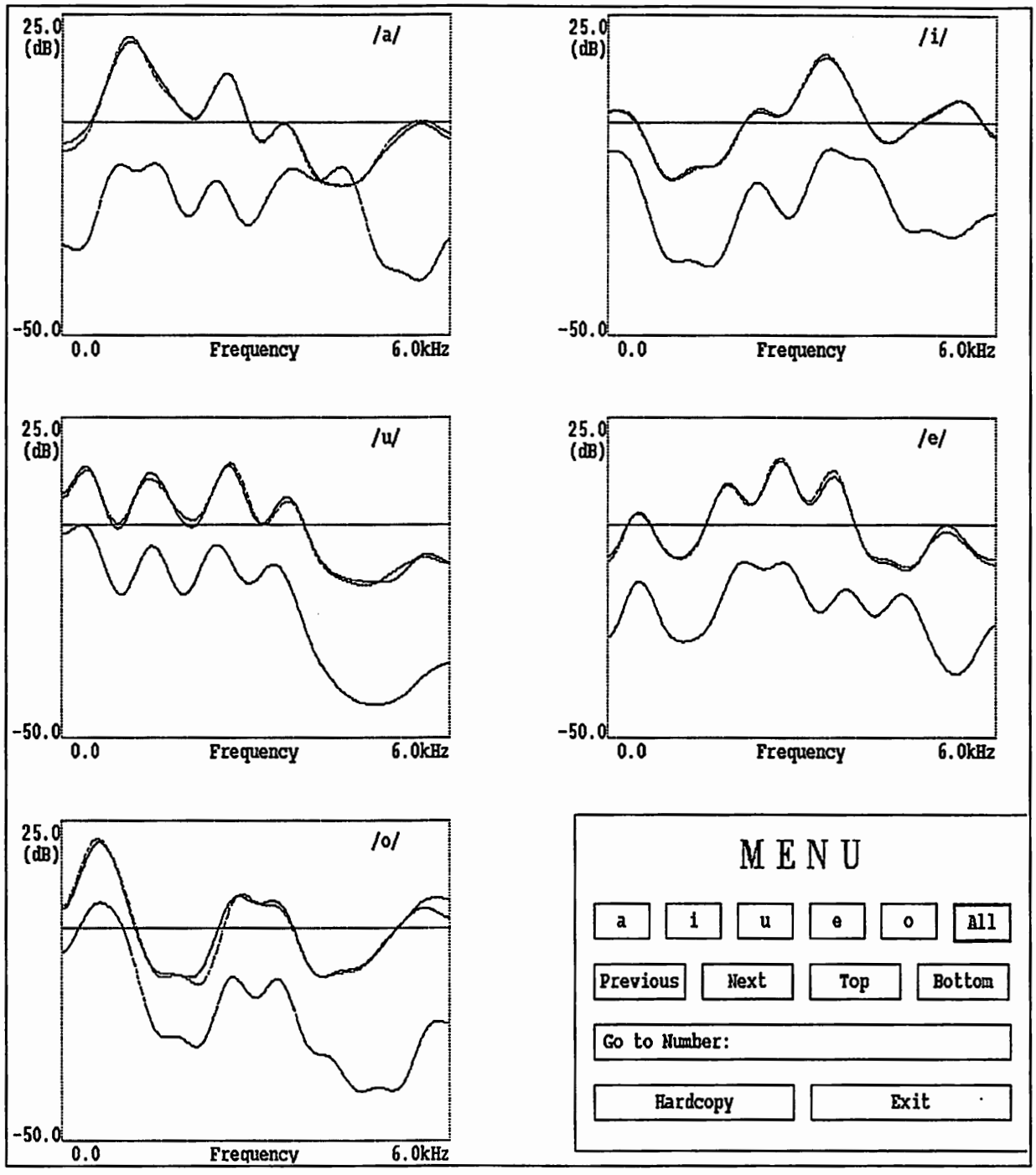

図 4: Spectrum 1

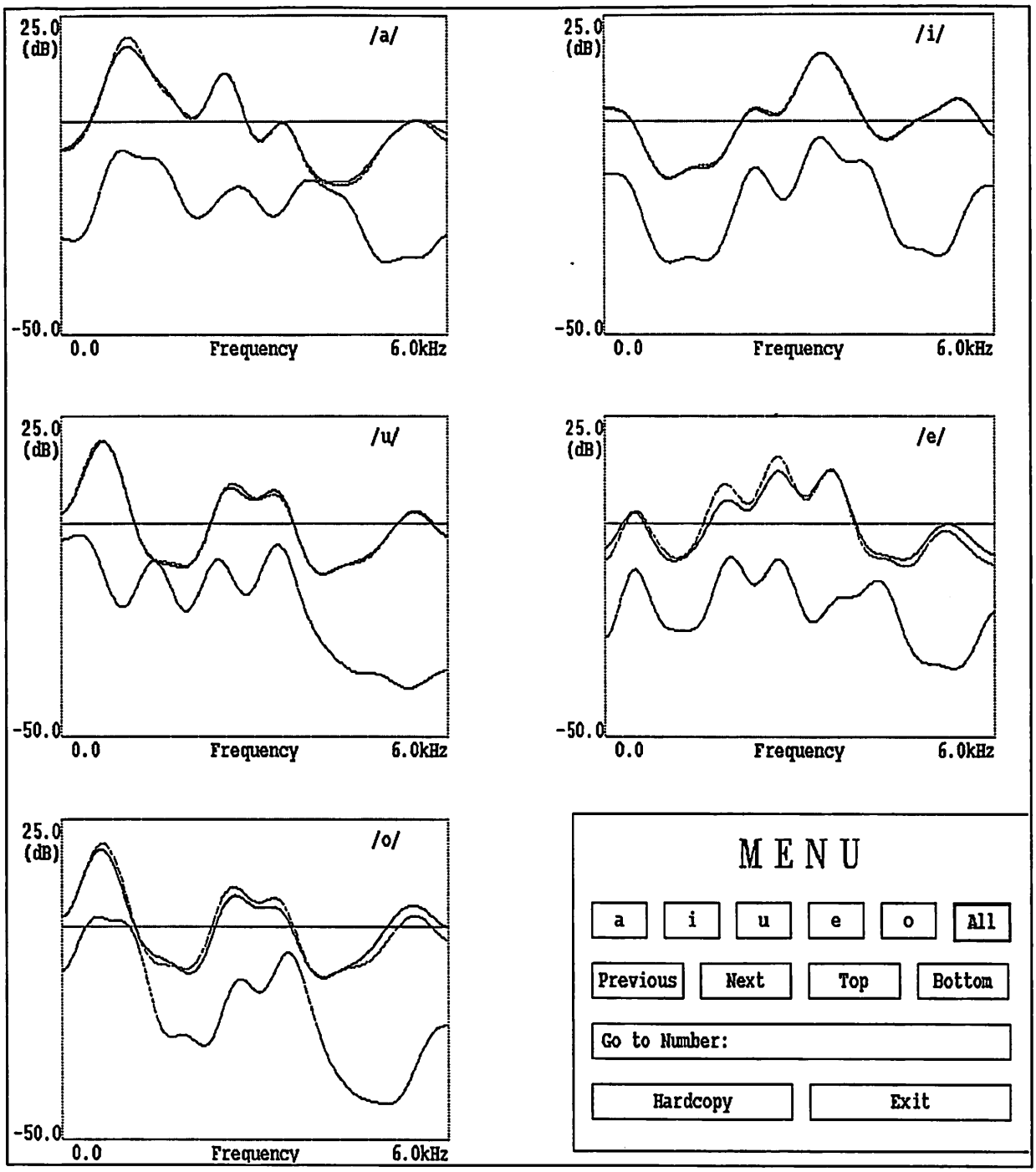

図 5: Spectrum 2

D 実験に用いたソフトウェア

実習生の山本君の作成したプログラム '91.04.05

1. 母音中心リストの作成
   source : /pooh/Adapt/Work/MKLIST/mklist_mk.c
   exec : /pooh/Adapt/Work/MKLIST/mklist_mk
   cshell : /pooh/Adapt/Work/MKLIST/mklist.sh
2. COR file 作成
   source : /pooh/Adapt/Work/WCR/WaveCorrRmix.c
   exec : /pooh/Adapt/Work/WCR/WaveCorrRmix
   cshell : /pooh/Adapt/Work/WCR/WaveCorrRmix.sh
3. Cep file & Cep Codebook の作成
   source : /pooh/Adapt/Adapt/VQ/src/make_cdbook1.c
   exec : /pooh/Adapt/Work/Adapt/VQ/make_cdbook1
   cshell : /pooh/Adapt/Work/Adapt/VQ/exe
   /pooh/Adapt/VQ/src: make_cdbook1.c , make_cdbook1.make
   /pooh/Adapt/VQ: make_cdbook1 , exe
4. Binary to ASCII for NN ASCII file
   /pooh/Adapt/Work/TOASCII: optin [-t : mapping sample]
5. NN Training Word sample
   /pooh/Adapt/VQ:
6. Word COR data
   /pooh/Adapt/Work/Work: (1-25 , 26-30 (5words))
7. VQ NN /pooh/Adaptation Training
   /pooh/Adapt/Net/AU_HT_word25:
   /pooh/Adapt/Net/AU_HT_word5:
   /pooh/Adapt/Net/Codebook:
   /pooh/Adapt/Net/Logfile:
   /pooh/Adapt/Net/Logfile/Id:
   /pooh/Adapt/Net/Logfile/Randam:
   /pooh/Adapt/Net/MAU_idmap.aiueo:
   /pooh/Adapt/Net/Net:
   /pooh/Adapt/Net/Net/Netdir:
   /pooh/Adapt/Net/Net/Netdir3552:
   /pooh/Adapt/Net/Net1:
   /pooh/Adapt/Net/Net1/Netdir:
   /pooh/Adapt/Net/Net1/Netdir0066:
   /pooh/Adapt/Net/Net2:
   /pooh/Adapt/Net/Net2/Netdir:
   /pooh/Adapt/Net/Net2/Netdir0026:

/pooh/Adapt/Net/Netdir:

/pooh/Adapt/Net/Netdir25:

/pooh/Adapt/Net/Netdir5:

/pooh/Adapt/Net/Netdir_id_odd:

/pooh/Adapt/Net/Netdir_r1_odd:

/pooh/Adapt/Net/Netdir_r2_odd:

/pooh/Adapt/Net/Netdir_word25_odd:

/pooh/Adapt/Net/Netdir_word5_odd:

8. X Window Specrtum display

/pooh/Spectrum: spec.c , exe (makefile) spec [mapped-file] [next-sample; cep-code]

new version: spec_new.c , exe (makefile) spec_new [mapped-file] [next-sample; cep-code]