TR-IT-0233

3 次元ビタビアルゴリズムを用いた話者正規化音響モデルの作成
Speaker normalized acoustic modeling based on 3-D Viterbi search algorithm

杉村 耕司 Sugimura Kouji
深田 俊明 Fukada Toshiaki

1997.9.12

高性能な音声認識システムを実現する上で、認識性能の高い音響モデルの構築は必要不可欠な技術である。そのための一つの方法として、周波数ワーピングを用いた話者正規化音響モデルの作成が行なわれている。しかし、従来の方法では一つの発話内においては、ワーピング係数を固定とする制限が付けられていた。そこで今回、ワーピング係数を発話内においても可変とし、より精密な音響モデルの作成を試みた。その結果、ワーピング係数を発話内で固定として作成した音響モデルと比較した場合、学習データに対して常に高い尤度を示す、音響モデルが作成できた。

目次

第 1 章

はじめに

高性能な音声認識システムを実現する上で、認識性能の高い音響モデルの構築は必要不可欠な技術である。しかし、発話は話者毎に異なった音響的な広がりを持つために、大量の不特定話者音声を用いて1つの音響モデルを学習する場合、音素モデルの平均がぼやけたり、分散が大きくなり、異なった音素間で分布の重なりが生じる。そのため、いくら混合数や状態数、コンテキストを増やしても、認識性能が頭打ちとなる可能性が考えられる。

このため、近年、周波数ワーピングによって話者正規化を行ない、頑健 (robust) で精密 (precise) な音響モデルを作成する研究が行なわれている [1] [2] [3] 。この方法は、認識時に入力音声の特徴パラメータを幾通りかの周波数ワーピングを行ない、尤度最大となるワーピング係数から得られる結果を認識結果とするものである。しかしこれまで提案されてきた手法は、一つの発話では発話全体に対して、1種類のワーピング係数しか許していないという制限がある。

そこで本研究の目的は、周波数ワーピングに基づく話者正規化音声認識におけるワーピングの方法を、 1）発話内において固定とした固定周波数ワーピング と、 2）1つの発話内においても、3次元ビタビアルゴリズムを用いて可変とした可変周波数ワーピング の2通りについて、それぞれの場合の比較を行なうことである。

以下、第2章では今回特徴パラメータとして用いたメルケプストラム分析法の説明、固定周波数ワーピング、可変周波数ワーピングにおける学習の手順を示し、第3章では、これら2種類のワーピング方法により音響モデルを作成し、その比較を行なう。第4章をまとめとする。また、完全な比較は出来ていないが、認識実験の結果の一部を付録に示す。

第 2 章

アルゴリズム

2.1 Pre-processing

2.1.1 特徴パラメータ

本報告では、特徴パラメータとしてメルケプストラム分析法 [5] により得られるものを用いる。メルケプストラムの定義はいくつかあるが、ここでは、ある適当な、位相特性を持つ因果的なオールパス関数を

$$\tilde{z}^{-1} = \Psi(z) \tag{2.1}$$

ただし、

$$\Psi(e^{j\omega}) = \exp(-j\tilde{\omega}) \tag{2.2}$$

として、

$$\tilde{c}(m) = \frac{1}{2\pi j} \oint_c \log \tilde{X}(\tilde{z}) \tilde{z}^{m-1} d\tilde{z} \tag{2.3}$$

$$\log \tilde{X}(\tilde{z}) = \sum_{m=-\infty}^{\infty} \tilde{c}(m) \tilde{z}^{-m} \tag{2.4}$$

でメルケプストラム $\tilde{c}(m)$ を定義する。ここで、$X(z)$ は、安定な実系列 $x(n)$ の z 変換、C は単位円を含む $\log X(z)$ の収束領域内で、原点を左回りに一周する閉路とする。式 (2.3) 式 (2.4)、は、単位円周上で、

$$\tilde{c}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \tilde{X}(e^{j\tilde{\omega}}) e^{j\tilde{\omega}m} d\tilde{\omega} \tag{2.5}$$

$$\log \tilde{X}(e^{j\tilde{\omega}}) = \sum_{m=-\infty}^{\infty} \tilde{c}(m) e^{-j\tilde{\omega}m} \tag{2.6}$$

但し、

$$\tilde{X}(e^{j\tilde{\omega}}) = \tilde{X}(e^{j\beta(\omega)}) = X(e^{j\omega}) \tag{2.7}$$

と書くことができ、$\tilde{c}(m)$ は対数スペルトル $\log X(e^{j\omega})$ を非直線周波数軸 $\tilde{\omega} = \beta(\omega)$ に周波数変換した時のフーリエ係数となる。ここで、オールパス関数を

$$\tilde{z}^{-1} = \Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}, \qquad |\alpha| < 1 \tag{2.8}$$

とすれば、$\tilde{z}^{-1} = e^{-j\tilde{\omega}}$ の位相特性は、

$$\tilde{\omega} = \beta(\omega) = \tan^{-1} \frac{(1-\alpha^2)\sin\omega}{(1+\alpha^2)\cos\omega - 2\alpha} \tag{2.9}$$

で、与えられる。式 (2.9) は、標本化周波数が $16kHz$ の場合、α を 0.42 と選べば、人間の音の高さに対する聴覚特性を表すメル尺度によく近似し、また、同様に $\alpha = 0.46$ と選べば、MFCC 分析におけるメル尺度によく近似するため、ここでは式 (2.8) で表される一次のオールパス関数を用いることにする。このような非直線周波数軸 $\tilde{\omega}$ 上でのケプストラム $\tilde{c}(m)$ を、メルなどの心理尺度を含む周波数目盛上でのパラメータという意味で、メルケプストラムと呼んでいる。

表 2.1 に、幾つかの特徴パラメータを用いて、3 状態各 5 混合の環境非依存音素 HMM を作成した場合の認識性能の違いを示す (TIMIT データベース使用)。表から、メルケプストラム分析法が音声認識において有効な分析法の一種であると考えられる。

表 2.1: 認識性能 (音素認識 %)

特徴パラメータ	w/o LM	with LM
LPC-cepstrum	43.98	52.15
MFCC	48.18	55.46
メルケプストラム	49.64	56.38

2.1.2 周波数ワーピング

周波数ワーピングは式 (2.8) における α を変えることにより実現する。図 2.1に α を変化させた場合の周波数ワーピングの例を示す。

2.2 話者正規化モデルの作成

2.2.1 3 次元ビタビアルゴリズム

従来法の周波数ワーピングによる話者正規化では、各ワーピング係数に対し、個別に状態とフレームという 2 次元の格子構造においてビタビ探索を行ない、話者毎に最大の尤度を与えるワーピング係数を決定してきた (図 2.2参照)。

本報告で用いる 3 次元ビタビアルゴリズムは、従来法の、状態とフレームという 2 次元に、ワーピング係数を次元に加えた 3 次元格子構造においてビタビ探索を行なう (図 2.3参照)。

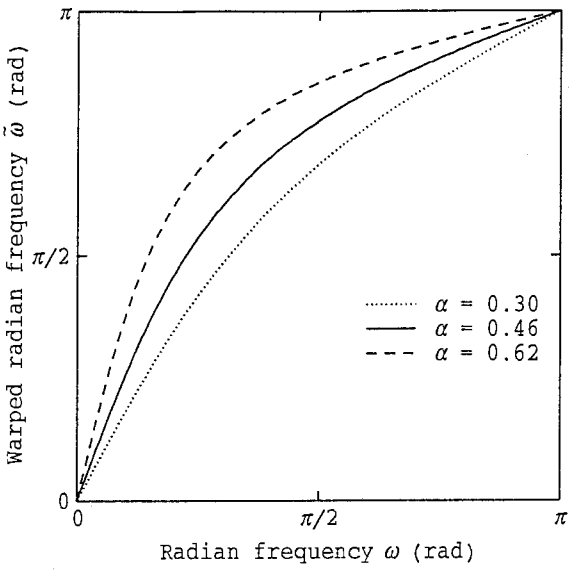


図 2.1: 周波数ワーピング

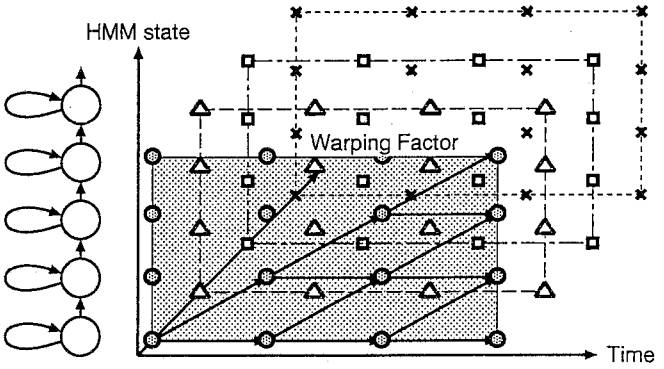


図 2.2: 固定周波数ワーピング

2.2.2 学習アルゴリズム

固定周波数ワーピングにおけるモデルの学習 (SNC-HMM)

1 つの発話内においてはワーピング係数を固定とした固定周波数ワーピングにおける、学習の手順を以下に示す。

1. 全ての話者に対して、一定のワーピング係数の初期値 $\alpha = 0.46$ を与え、初期モデルを作成する。

2. 9 種類にワーピングした学習データに対し、2 次元ビタビ探索を行ない、話者毎又は発話毎に最大の尤度を与えるワーピング係数を決定する。

3. ステップ 2 で得られたワーピング係数から求めた特徴パラメータを用いて、HMM の再学習を行なう。

4. ステップ 2 にもどる。

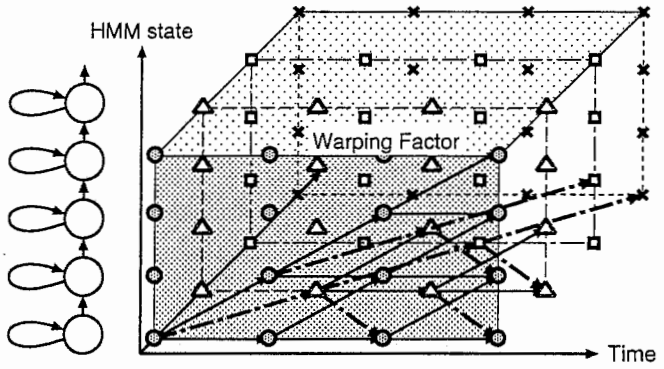


図 2.3: 可変周波数ワーピング

可変周波数ワーピングにおけるモデルの学習 (SNV-HMM)

可変周波数ワーピングによる学習の手順をしめす。

1. 全ての話者に対して、一定のワーピング係数の初期値 $\alpha = 0.46$ を与え、初期モデルを作成する。

2. 現在の HMM において、3 次元ビタビ探索を行ない、学習データに対してビタビアライメントをとり、最大の尤度を与えるワーピング係数列を決定する。

3. 最適なワーピング係数列により得られる特徴ベクトル系列 (図 2.4参照) によって、HMM の再学習を行なう。

4. ステップ 2 にもどる。

今回は、ワーピング係数が急峻に変化するのを防ぐため、3 次元ビタビ探索において制限を加えた (2.2.3参照)。

2.2.3 認識アルゴリズム

認識手順においては、尤度最大となる最適経路を求めることにより、ワーピング係数の推移と発話内容を表す音素（単語）列を得る。

図 2.5 に認識アルゴリズムを示す。ここで、$\mathbf{S}$、Q、D、N は、初期状態集合、状態数、ワーピング係数の数、フレーム数を。π 、 α、 $a(q',q)$ 、 $a(d',d)$、 b、 $\mathbf{x}$ は、初期状態確率、最適状態系列における尤度、状態 q' から q への遷移確率、ワーピング係数 d' から d への遷移確率、出力確率、特徴ベクトルを表す。

今回、可変周波数ワーピングにおいてワーピング係数 d' から d への遷移確率 $a(d',d)$ は、

$$a(d',d) = \begin{cases} 1.0, & |d'-d| \leq w \\ 0.0, & |d'-d| > w. \end{cases} \tag{2.10}$$

として、今回の実験では $w = 1$ とした。さらに、音素モデル内での状態遷移においては w を 0 とした。

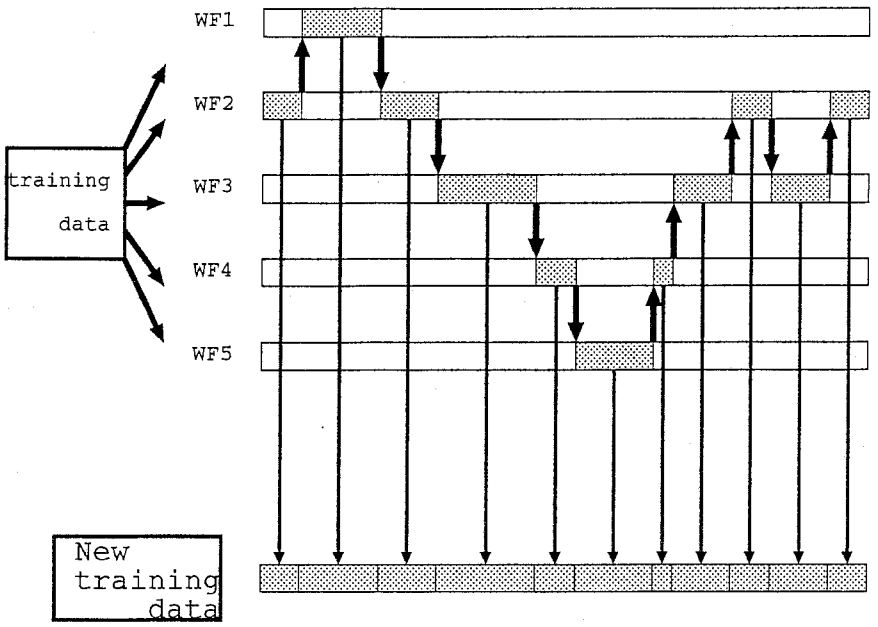


図 2.4: 3D ビタビ探索による特徴パラメータの生成

ここで、全ての状態遷移において w を 0 とすると、可変周波数ワーピングは固定周波数ワーピングと等しくなる。

Initialization:

for $q = 1$ **to** Q

  for $d = 1$ **to** D

    if $(q, d) \in \mathbf{S}$ **then**

      $\alpha(q, d, 0) = \log \pi(q, d)$, where $\sum_{(q,d)\in \mathbf{S}} \pi(q, d) = 1$

    else

      $\alpha(q, d, 0) = -\infty$

Recognition:

for $n = 1$ **to** N

  for $q = 1$ **to** Q

    for $d = 1$ **to** D

      $\alpha(q, d, n) = \max_{q', d'} \{\alpha(q', d', n-1) + \log a(q', q) + \log a(d', d)\} + \log b(q, \mathbf{x}(d, n))$

図 2.5: 3 次元ビタビ探索による認識アルゴリズム

第 3 章

実験

3.1 音響モデルの作成

TIMIT のデータベースの学習データ 462 人、各 8 文章の 3696 文章を用いて、音響モデルを作成する。

3.1.1 実験条件

特徴パラメータは、フレーム長 25.6$msec$、フレーム周期 10$msec$ により計算した、メルケプストラム ($c(0) \sim c(12)$) と、その一次回帰係数の合計 2 6 次元を使用した。ワーピング係数は、$\alpha = 0.30$ から 0.62 の 9 種類とした。学習条件を表 3.1に示す。

表 3.1: 学習条件

トポロジー	ML-SSS により作成
状態数	1000 状態
混合数	5 混合
話者数	男性 326 人 、女性 136 人、各 8 発話 (TIMIT データベース)
学習方法	ラベル学習 (F-B 学習) 連結学習 (Viterbi 学習)
学習の繰り返し	ラベル学習 20 回、連結学習 10 回

ワーピング係数が 0.46 の時の特徴ベクトルを用いて、尤度最大化基準逐次分割アルゴリズム (ML-SSS) により、 1000 状態、各 5 混合の HMnet を作成し、これを固定周波数ワーピング、可変周波数ワーピングによる音響モデル学習時のトポロジーとした。

3.1.2 固定周波数ワーピング

音響モデルの学習

固定周波数ワーピングにおけるモデル学習時に、学習データに対しビタビ探索を行なった際の総尤度の推移を図 3.1に示す。

また、学習データにおいて尤度最大となるワーピング係数を話者毎、発話毎に選択した時の分布の推移を図 3.2から図 3.5に示す。さらに、それぞれの繰り返しにおいて選択されたワーピング係数の平均値と分散を 図 3.6、図 3.7に示す。

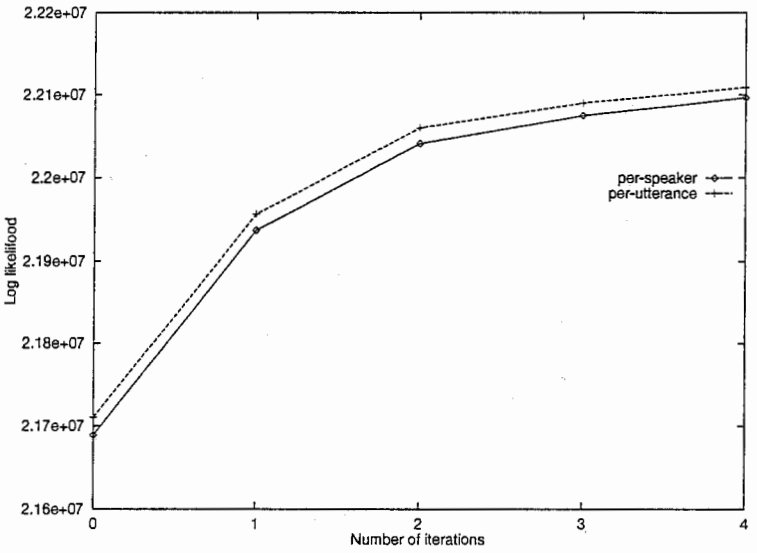
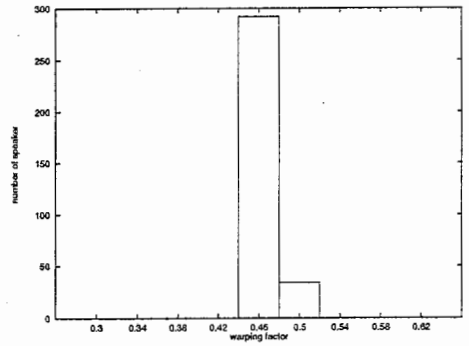
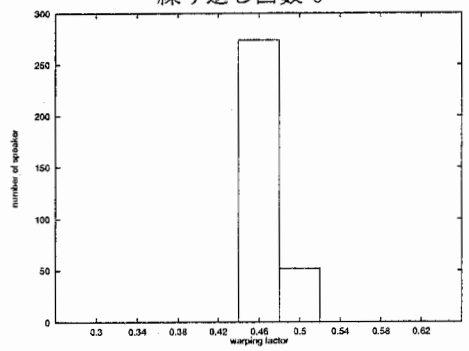


図 3.1: 固定周波数ワーピングにおけるモデル学習時の総尤度の変化

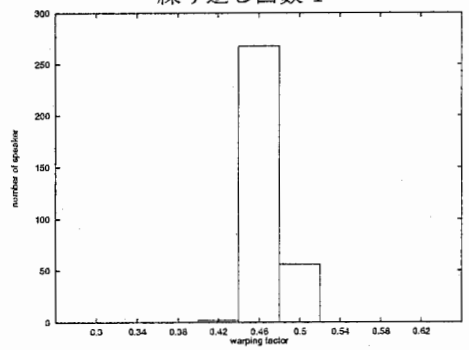
ここで、0.46 より小さいワーピング係数は低域周波数での圧縮、高域周波数での伸長を、0.46 より大きいワーピング係数は、低域周波数での伸長、高域周波数での圧縮を意味している。そのため、図 3.6から、男性話者には 0.46 より大きなワーピング係数選ばれ、女性には 0.46 より小さなワーピング係数が与えられている。これは、男声のフォルマント周波数が、女声よりも低いためとみられ、文献 [1] などの結果にも同じ傾向が見られる。またこの傾向は、学習を繰り返すたびに大きくなっていく。これは、学習の繰り返しによって、精密なモデルが得られているためと考えられる。次に分散の推移を見てみると学習の繰り返しとともに、男性、女性毎にワーピング係数の分布が広がっていく。これから、学習を繰り返すことによって、各話者の個人差に適応したワーピング係数が選ばれていると考えられる。



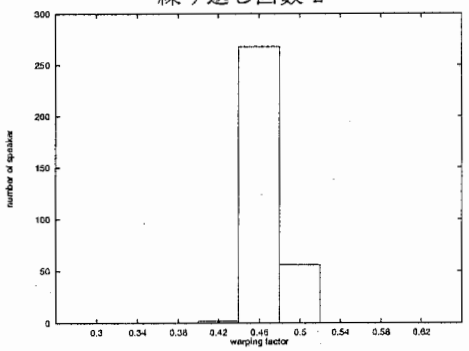
繰り返し回数 0



繰り返し回数 1

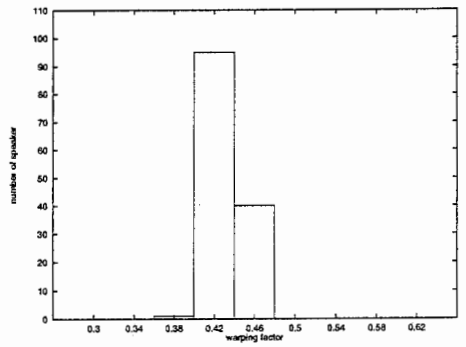


繰り返し回数 2

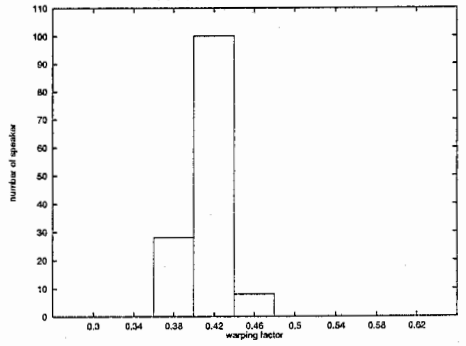


繰り返し回数 3

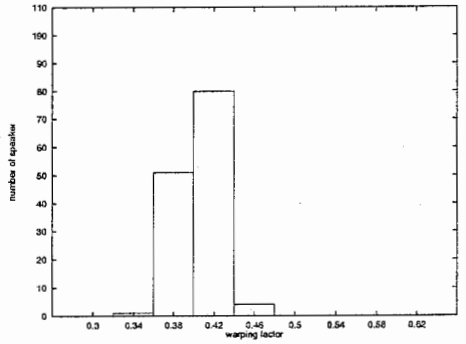
図 3.2: 男性話者のワーピング係数の分布の変化 (ワーピング係数を話者毎に選択)



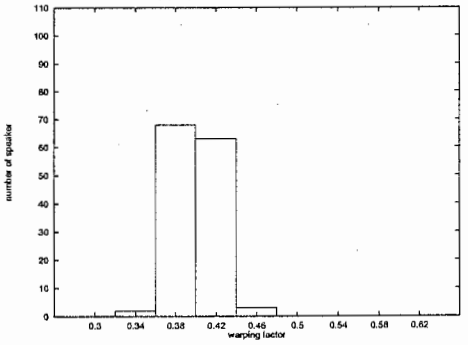
繰り返し回数 0



繰り返し回数 1

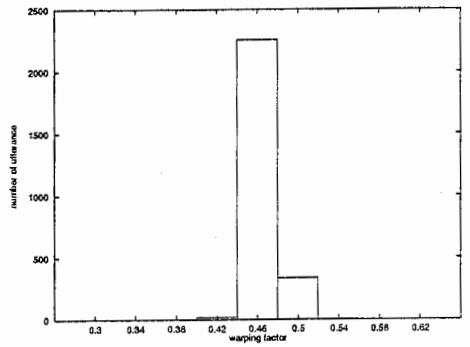


繰り返し回数 2

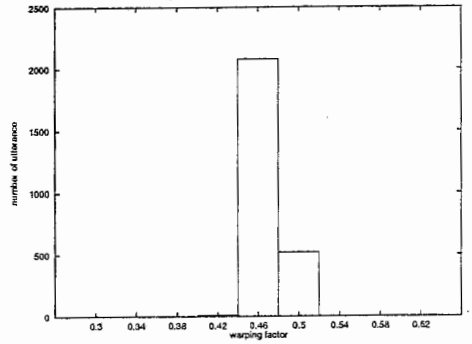


繰り返し回数 3

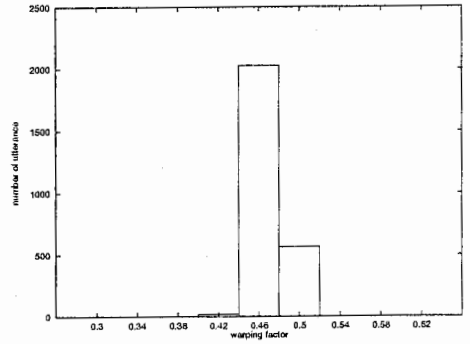
図 3.3: 女性話者のワーピング係数の分布の変化 (ワーピング係数を話者毎に選択)



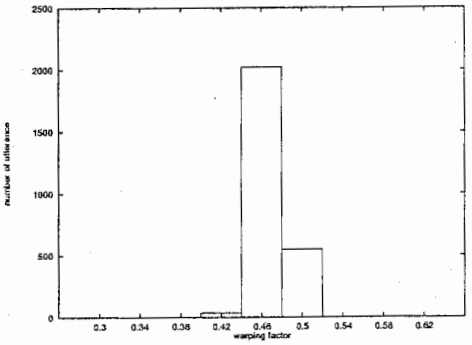
繰り返し回数 0



繰り返し回数 1

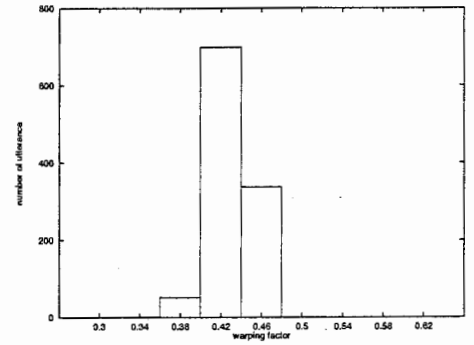


繰り返し回数 2

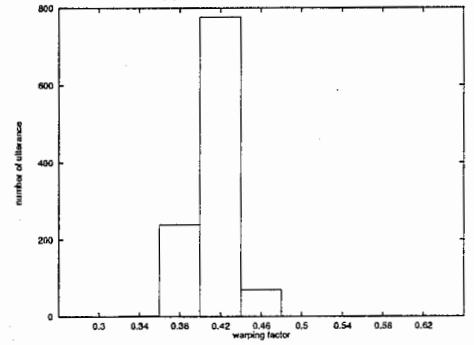


繰り返し回数 3

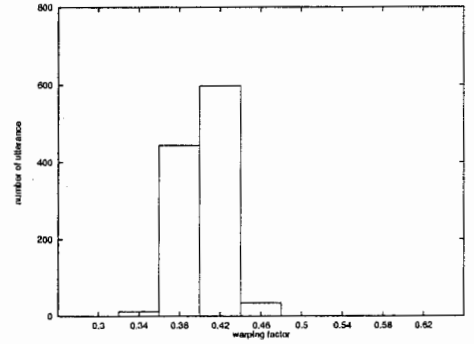
図 3.4: 男性話者のワーピング係数の分布の変化 (ワーピング係数を発話毎に選択)



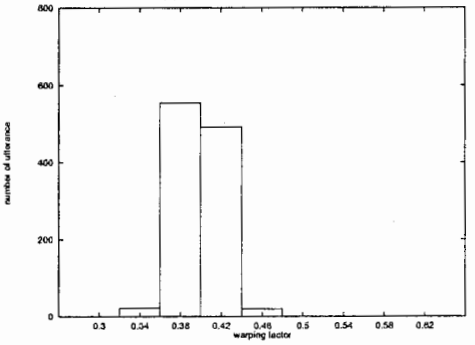
繰り返し回数 0



繰り返し回数 1



繰り返し回数 2



繰り返し回数 3

図 3.5: 女性話者のワーピング係数の分布の変化 (ワーピング係数を発話毎に選択)

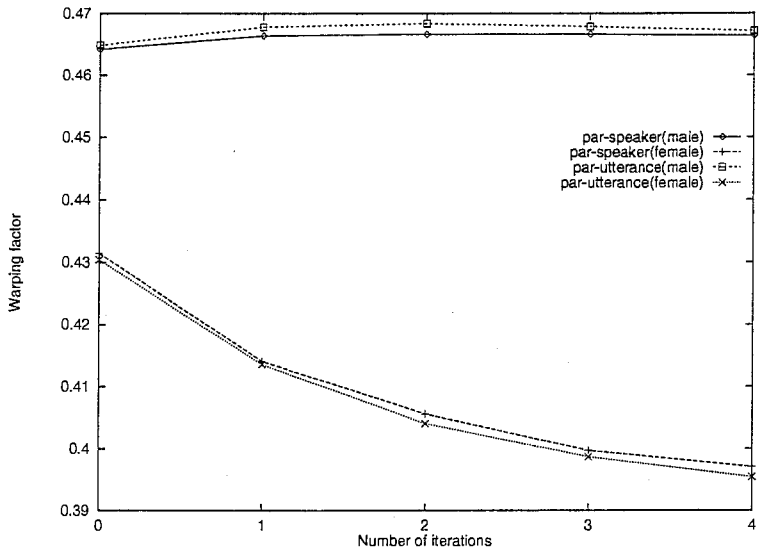


図 3.6: 固定周波数ワーピングにおけるワーピング係数の平均値

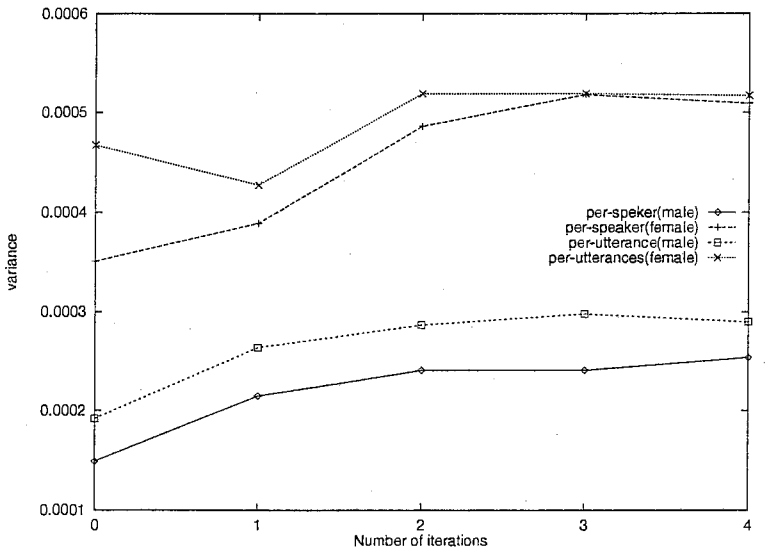


図 3.7: 固定周波数ワーピングにおけるワーピング係数の分散

3.1.3 可変周波数ワーピング

次に、本報告の提案法であるワーピング係数を1つの発話中においても、可変として実験を行なった。

音響モデルの学習

可変周波数ワーピングにおけるモデル学習時に、学習データに対し3次元ビタビ探索を行なった際の総尤度の推移を図3.8に示す。図中において、"constant" は、固定周波数ワーピングにおいて話者毎にワーピング係数を求めた際の総尤度を表している。この図から、可変周波数ワーピ

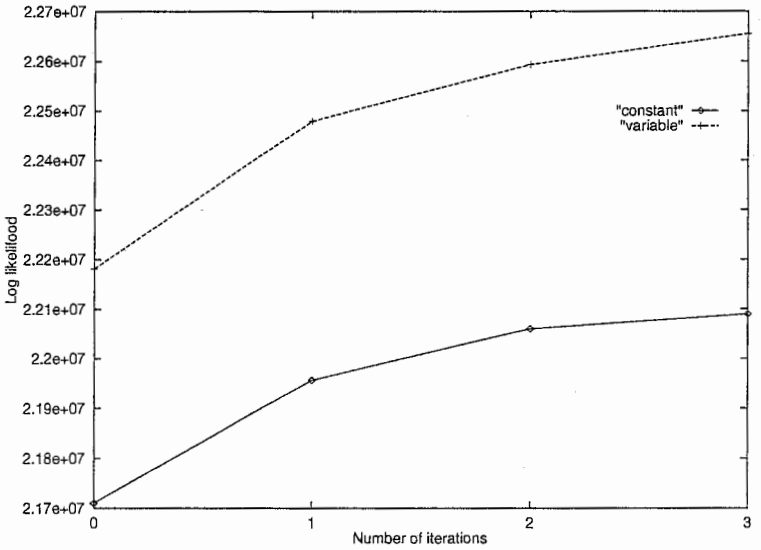
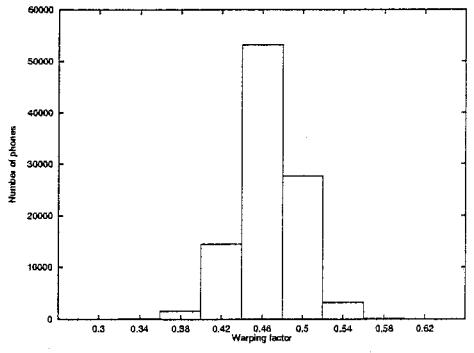


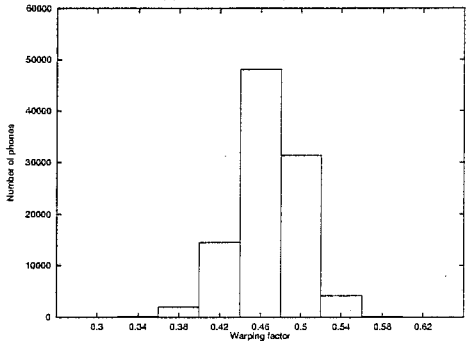
図 3.8: 固定周波数ワーピングモデル学習時の総尤度の変化

ングによるモデルの学習は固定周波数ワーピングに対し、常に高い尤度を示している。このことから、より精密なモデルが作成されていると考えられる。

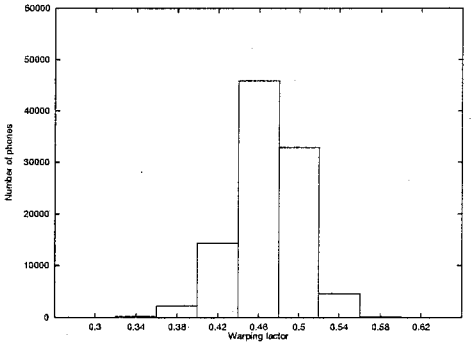
可変周波数における男性のワーピング係数の分布の変化を図3.9に、女性の場合を図3.10に示す。可変周波数ワーピングにおいても、選択されるワーピング係数の推移は男性、女性共に固定周波数ワーピングの場合と同じように推移している。しかしその分布の広がりは大きくなっている。



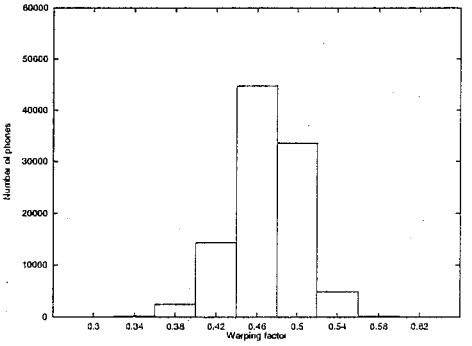
繰り返し回数 0



繰り返し回数 1

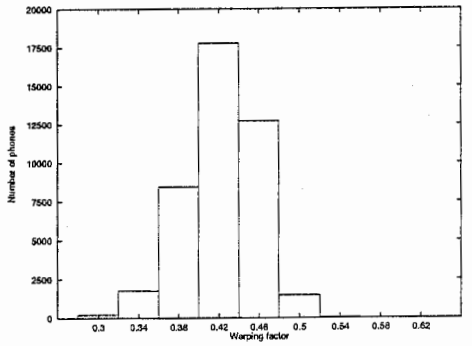


繰り返し回数 2

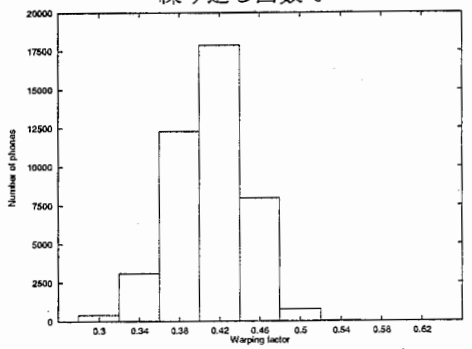


繰り返し回数 3

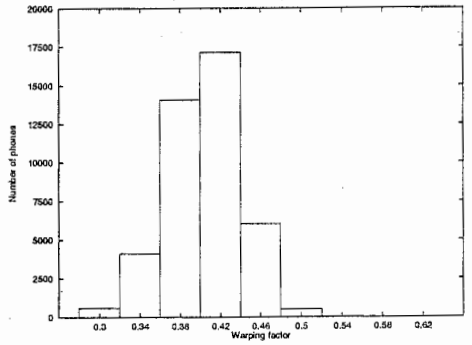
図 3.9: 男性話者のワーピング係数の分布の変化 (可変周波数ワーピング)



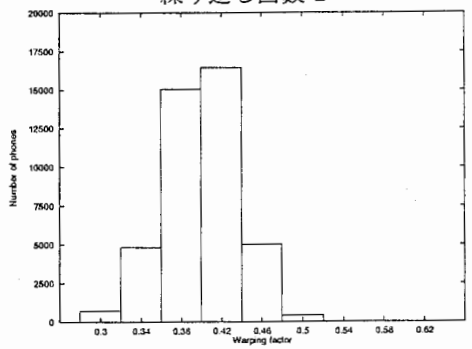
繰り返し回数 0



繰り返し回数 1



繰り返し回数 2



繰り返し回数 3

図 3.10: 女性話者のワーピング係数の分布の変化 (可変周波数ワーピング)

図 3.9, 図 3.10から求めた、ワーピング係数の平均値と分散を図 3.11、図 3.12に示す。図 3.11、図 3.12において、"constant" は、固定周波数ワーピングにおいて、話者毎にワーピング係数を求めた際の結果を表している (図 3.6および、図 3.7参照)。

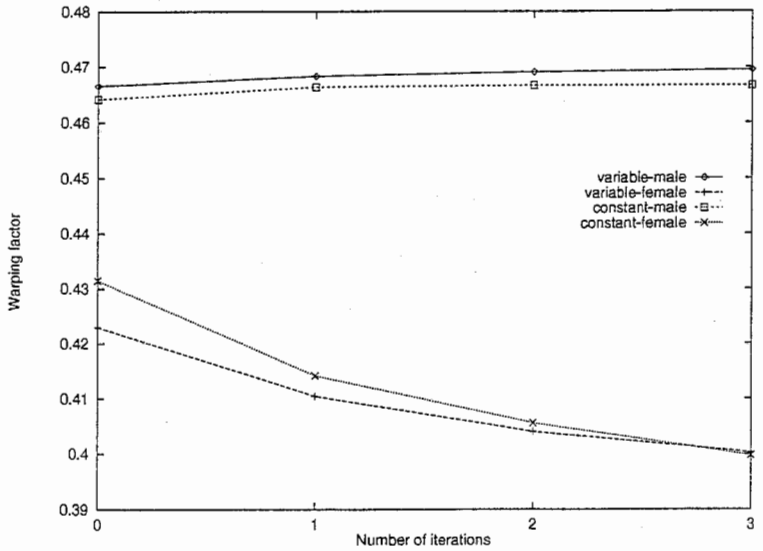


図 3.11: 可変周波数ワーピングにおけるワーピング係数の平均値

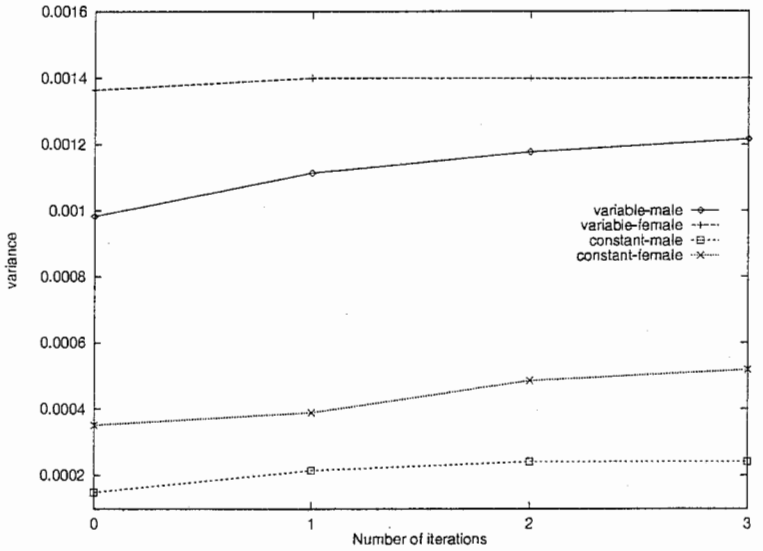


図 3.12: 可変周波数ワーピングにおけるワーピング係数の分散

ワーピング係数を発話中にも可変とした場合においても、ワーピング係数の平均値は固定周波数ワーピングの場合と同じよう男性話者には、0.46 より大きな、女声話者には 0.46 より小さなワーピング係数が与えられ、学習を繰り返す毎にその傾向は大きくなっている。しかし図 3.12に見られるように、その分散は男性、女声共に大きくなっている。これは、発話中においても、各音素に対して異なったワーピング係数が選ばれていることを示している。

第 4 章

まとめ

本稿は、周波数ワーピングを用いた話者正規化音声認識において、ワーピング係数を発話内においても可変とすることで、より精密な話者正規化音響モデルの作成を目的とし、また従来法であるワーピング係数を発話内において固定とした場合との比較を行なった。

その結果、音響モデルの学習時に、学習データに対しビタビ探索を行なった際の総尤度は、繰り返しの回数に関わらず、常に提案法の方が高かった。このことからより精密な音響モデルが作成出来ている考えられる。

今後、実際に認識実験を行ない、不特定話者、性別依存、それに固定周波数ワーピングといった、他の音響モデル用いた場合との認識性能の比較をし、その有効性が示されることを期待する。

更には、今回の実験においてはワーピング係数間の遷移に対し、制限を与えたがこの制限を変化させた場合の影響を調べるのも、今後の課題であろう。

謝辞

本研究を進めるにあたり、匂坂芳典室長をはじめとして、ATR音声翻訳通信研究所第一研究室の皆様方には、温かい御支援、御指導をいただきました。本当に有難う御座いました。また、同じ実習生の目から、的確なアドバイスをくれた、早稲田大学の杉田洋介くん有難う。最後になりましたが、実務訓練の機会を与えて下さった、ATR音声翻訳研究所の山本誠一社長、奈良先端科学技術大学院大学の鹿野清宏教授、ならびに、中村哲助教授に深く感謝致します。

参考文献

[1] L. Lee and R. C. Rose: "Speaker normalization using efficient frequency warping procedures," *Proc. ICASSP-96*, pp. 353–356 (1996).

[2] P. Zhan and M. Westphal: "Speaker normalization based on frequency warping," *Proc. ICASSP-97*, pp. 1039–1042 (1997).

[3] S. Wegmann, D. McAllaster, J. Orloff and B. Peskin: "Speaker normalization on conversational telephone speech," *Proc. ICASSP-96*, pp. 339–341 (1996).

[4] 清水 徹, 山本 博史, 政瀧 浩和, 松永 昭一, 匂坂 芳典: "大語い連続音声認識のための単語仮説数削減," 信学論 (D-II), **J79-D-II**, 12 pp. 2117–2124 (1996-12).

[5] 徳田 恵一, 小林 隆夫, 深田 俊明, 斉藤 博徳, 今井 聖: "メルケプストラムをパラメータとする音声のスペクトル推定", 信学論 (A), **J74-A**, 8 pp. 1240–1248 (1991-8).

付録・認識実験

認識にはTIMITデータベースの評価データ (男性112人、女性56人 各8発話) を用いた。
ソフトウエアは、ATRSPREC r04r02 (およびその修正版) をもとに行なった。

A　不特定話者モデル

ML-SSSにおいて作成したトポロジーに対しワーピング係数$\alpha = 0.46$による特徴パラメータにより学習を行なった音響モデル（固定、可変周波数ワーピングにおける音響モデル学習時の初期モデルに等価、GI-HMM）を用いて行なった認識実験の結果を表B.1に示す。この時、ビームサーチのビーム幅は、30と40の2通りで行なった。

また、認識実験においては、表A.1の61音素における音素を39音素における音素と等価とすることで39音素とした場合の認識結果も併せて示す。

表から、ビーム幅30と40の場合を比較すると、ビーム幅を40の方が、認識率が低下している。これから、不特定話者モデルでの認識においてはビーム幅を30以上にする必要はないと考えられる。

表 A.1: 61音素から39音素に変換する条件

61音素	sil epi dx pcl tcl kcl bcl dcl	en nx	axh	hv	em	eng	ux	el	axr	zh	ao	ih
39音素	pau	n	ax	hh	m	ng	uw	l	er	sh	aa	ix

B　話者クラスタモデル

不特定話者モデルと同様に、ML-SSSによって作成したトポロジーに対し性別毎にラベル学習、連結学習行ない、性別依存モデル (GD-HMM) を作成する。これを用いて行なった認識結果を表B.1に示す。

次に、性別依存モデルを用いた場合に、実際の入力音声の性別と、認識時に選ばれたモデルの性別との関係をB.2に示す。

これらの結果から、性別依存モデルにおいては、正しい音響モデルがかなりの確率で選択されているにもか変わらず、不特定話者モデルを用いた認識性能に較べ、大きな改善がみられない。

表 B.1: 不特定話者モデルと性別依存モデルを用いた認識結果

	認識率 Acc(%)	
評価セット	61音素	39音素
不特定話者モデル (ビーム幅 30)	58.69	65.87
不特定話者モデル (ビーム幅 40)	58.05	65.25
性別依存モデル	58.75	65.98

表 B.2: 性別依存モデルの選択結果

	男性モデル	女性モデル	モデル選択正解率 (%)
男性発話	864	32	96.4
女性発話	417	31	93.1

この原因として、男性、女性という、2種類のクラスでは、クラスタリングが不十分である、またクラスを2つに分割する事により、それぞれの与えられる学習データが半減してしまうために十分な学習ができない、といったことが考えられる。

C 固定周波数ワーピング

C.1 GI-HMM を用いた認識結果

初めに、不特定話者モデル (GI-HMM) に対し固定周波数ワーピングによる認識実験を行なった。

まず初めに予備実験として、テストデータに対しワーピング係数を 0.30 0.34 … 0.58 0.62 の9通りに変化させ、得られる特徴ベクトルをそれぞれ、Aで用いた GI-HMM により認識を行なった。結果を図 C.1に示す。Aで用いた GI-HMM はワーピング係数 0.46 の学習データによって作成しているため、ワーピング係数が 0.46 の時、認識性能最大となり、ワーピング係数が 0.46 より離れるにしたがって、認識性能はしだいに低下する。

表 C.1: 各周波数ワーピング係数における認識結果

ワーピング係数	0.30	0.34	0.38	0.42	0.46	0.50	0.54	0.58	0.62
認識率 Acc (%)	27.98	35.66	45.97	54.99	58.69	55.61	47.76	36.87	25.68

つぎに、1つの発話に対し各ワーピング係数毎にビームサーチを行ない、得られる9種類の結果から認識率最大化基準および尤度最大化基準のもとで、ワーピング係数を発話毎、話者毎に決定した場合の認識結果を、表 C.2に示す。この時の認識率最大化基準における認識性能は、GI-HMM を用いて固定周波数ワーピングを行なった場合の認識性能の上限値を示している。

表 C.2: 認識率最大化基準の認識結果

	認識率 Acc (%)	
評価セット	話者毎に α を決定	発話毎に α を決定
6 1 音素	60.15	63.27
3 9 音素	67.05	69.34

表 C.3: 尤度最大化基準の認識結果

	認識率 Acc (%)	
評価セット	話者毎に α を決定	発話毎に α を決定
6 1 音素	59.08	59.10
3 9 音素	65.87	66.14

認識率最大化に基づく認識結果では、ワーピング係数を話者毎、発話毎に選択しても認識結果に大きな違いは見られず、共にベースライン (GI-HMM) の認識性能を若干上回る。

次に、9 種類の特徴ベクトルに一括してビームサーチを行なった認識結果を表 C.5に示す。ワーピング係数毎に認識を行なった時に比べ、約 2 % 程度の認識率の低下が見られる。

表 C.4: GI-HMM において各ワーピング係数に対し一括してビームサーチを行なった認識結果

評価セット	認識率 Acc(%)
6 1 音素	57.09
3 9 音素	64.33

C.2 正規化モデルを用いた認識実験

次に、固定周波数ワーピングによる話者正規化の学習を 1 回行なった音響モデルによる認識実験の結果を表 C.5に示す。

このときのビーム幅を変化させた場合の認識率への影響に対し予備的な実験を行なった。

評価データの一部 (dr1 および dr2、37 人) に対し、ビーム幅を変化させて認識実験を行なった。その結果を図 C.1 に示す。これより、固定周波数ワーピングにおける認識実験においては、ビーム幅を 40 にしたとき、認識率が最大となっている。

表 C.5: 固定周波数ワーピングにおける話者正規化モデルによる認識結果

評価セット	認識率 Acc(%)	
	ビーム幅 30	ビーム幅 40
6 1 音素	56.80	57.86
3 9 音素	64.13	65.23

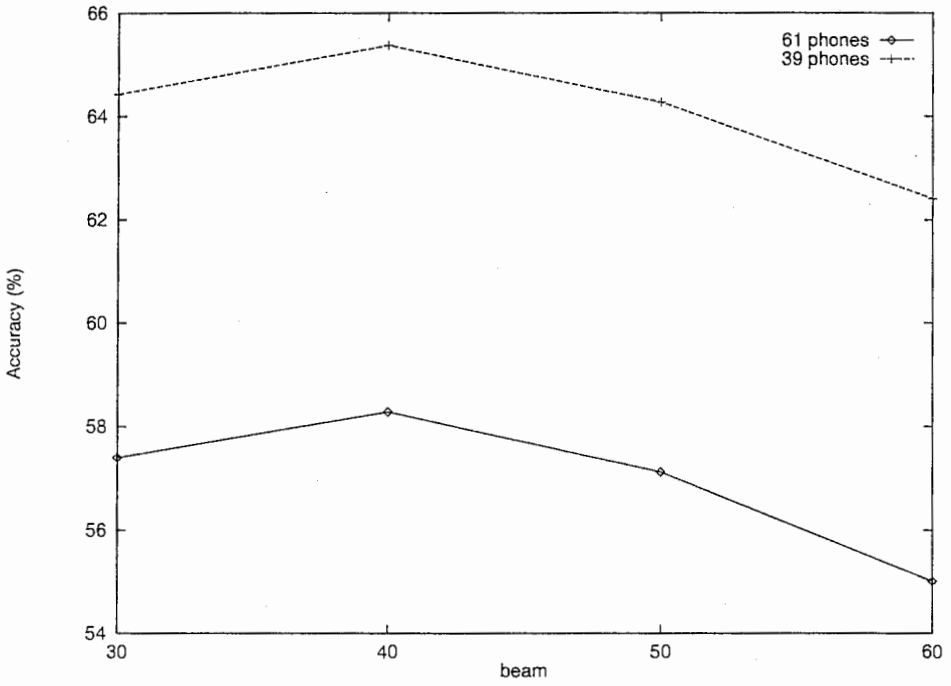


図 C.1: ビーム幅の変化が認識に与える影響

D 可変周波数ワーピング

D.1 GI-HMM を用いた認識結果

不特定話者モデルに対して、3 次元ビタビによる認識を行なった結果を表 D.1に示す。

表 D.1: GI-HMM に対し、可変周波数ワーピングを行なった認識結果

	認識率 Acc(%)
6 1音素	58.23
3 9音素	65.39

以上の現在認識率が出ている、不特定話者モデル (GI-HMM)、性別依存モデル (GD-HMM)、不特定話者モデルに対する固定周波数ワーピングによる認識実験で、それぞれのワーピング係数毎にビームサーチを行なった結果 (GI-CFWA)、全てのワーピング係数を一括してビームサーチした結果 (GI-CFWB)(固定周波数ワーピングではワーピング係数は話者毎に決定の場合を示した)、および可変の周波数ワーピングによる認識 (GI-VFW)、それに固定周波数ワーピングにより話者正規化を行なった固定周波数話者正規化モデル (SNC-HMM1)、による認識結果を図 D.1に示す。

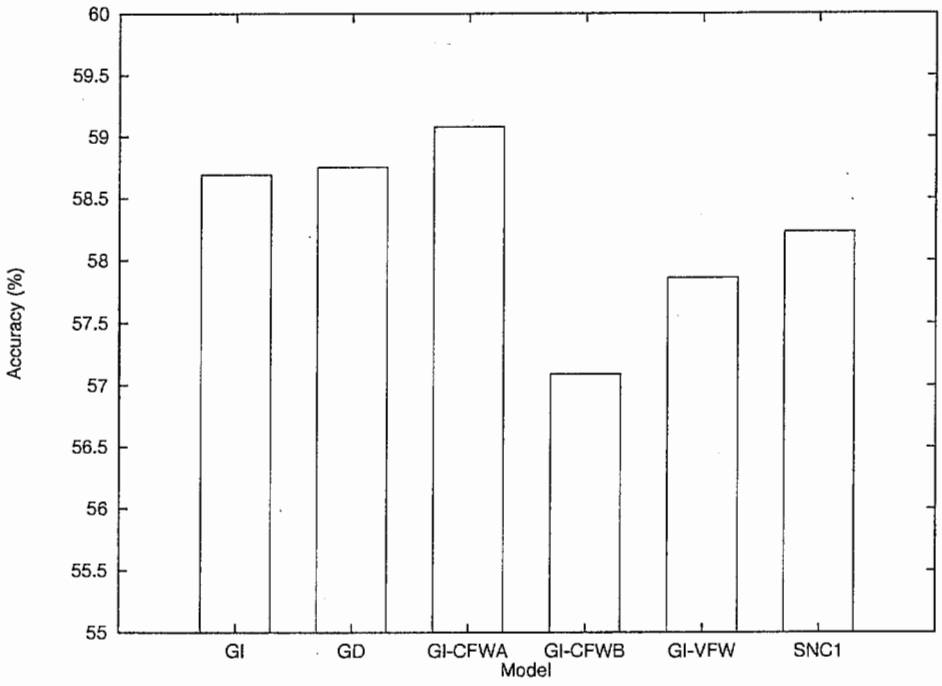


図 D.1: 各音響モデルによる認識結果