2005－SLP－55 (5)
2005／2／4

加重等分解度特徴量を用いたテキスト独立型話者識別

茂原 康博† Xugang Lu† 党 建武†
†北陸先端科学技術大学院大学 情報科学研究科
〒 923-1292 石川県能美郡辰口町旭台 1-1
E-mail: †{yosiyasu,xugang,jdang}@jaist.ac.jp

あらまし 本研究では，話者の生理学的特徴を捉える音響特徴に着目して加重等分解度特徴量を提案し、さらに話者識別システムに取り込み，話者識別を行った．音声の個人性については，口・鼻腔の音響結合度合いによりスペクトル上の 300 Hz と 3000 Hz 付近において極零対が生じ，または，梨状窩の個人差により 4000 Hz から 6000 Hz までの範囲において特徴的スペクトルが形成されているとの音声生成の研究報告があった．話者の生理学的特徴を取り入れるため上記の周波数領域を局部細分化して話者認識を行ったところ，高い識別率を得るために，メルフィルタのバンドの最適な細分割数は，高周波数領域において高くなる傾向になることがわかった．この結果は，話者個人特徴の詳細を捉えるため，全周波数領域でメルスケールより線形スケールの方が有効であろうということを示唆する．この知見をもとに，本研究では，線形スケールを用い上記の生理学的特徴に関わる周波数領域に大きな重み係数をつけ DCT を施して音響特徴量（加重等分解度特徴量）を抽出した．提案した音響特徴量を MFCC と組み合わせたハイブリッド GMM モデルでは，従来の GMM モデルより話者識別率が顕著に改善された．

Text Independent Speaker Identification Using Weighted Linear Scale Spectral Feature

Yasuhiro YOSHIHARA† Xugang LU† and Jianwu DANG†
† School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Nomi, Ishikawa, 923-1292 Japan
E-mail: †{yosiyasu, xugang, jdang}@jaist.ac.jp

Abstract. In this research, we proposed a Weighted Linear Scale Spectral Feature to emphasize speakers' physiological individuals, and applied the proposed feature into a speaker identification system by combining it with the traditional Mel Frequency Cepstrum Coefficient (MFCC). Studies of speech production reported that the coupling degree of the nasal and oral cavities can induce pole-zero pairs around 300 Hz and 3000 Hz, and the piriform fossa, a side branch of the vocal tract, shapes the spectra in the region from 4000 Hz to 6000Hz. A local subdivision method of the concerned frequency region has been proposed to utilize the physiological information in speaker recognition. The results indicated that to reach the optimal performance more subdivisions is required for higher frequency region than for lower frequency region. This suggested that for extracting the individual details the analysis with a linear scale in frequency domain may be more efficient than that of a log scale. Based on this finding, this study adopted linear scale sub-bands in frequency domain, weighted the frequency regions that are concerned with the physiological events, and then carried out the DCT on it to get the target feature, named Weighted Linear Scale Spectral Feature. As a result, the performance of speaker identification was greatly improved when combining the proposed feature with the conventional MFCC.

1. はじめに

近年の高度情報化社会の到来により，音声を用いた個人認識技術（話者認識）の高精度化への要望がますます高まっている．

話者認識分野において，音声の個人性を捉える物理特徴量として，声道の大局的な形状を捉えている音響特徴量であるMFCCが主流として用いられてきた．多くの先行研究において，その有効性が示され，話者照合同様，話者識別においても最も有効な話者特性を表す音響特徴量の１つであるといえる[4][5].

一方，近年の声道形状の微細構造と音響特性との関連性に関する研究により，発話時の口腔と鼻腔との結合度合によりスペクトル上において，300 Hz と 3000 Hz 付近に極零対を起こすことが明らかになった[1]. 人によって発話時口腔と鼻腔の結合度合が異なるので，その形状の差異が話者の個人性に深く関わっていると考えられる．また，梨状窩が声道の分岐として4000 Hz から 6000 Hz において大きな極零対をおこす[13]. 梨状窩の形状は発話時に相対的不変なので，その形状的な差異は音声の個人性に深く関っていると考えられる[1][2]. 音声認識に良く用いられる MFCC ではメルスケールフィルタを使っているので，上記のような個人性が存在する周波数領域において周波数分解度が粗く，十分に話者特性の詳細を捉えきれていない恐れがある．

そこで筆者らは，これらの音声生成側知見に基づき，個人性情報が存在するそれぞれの周波数領域を局部細分化したケプストラム抽出法を提案し，従来のテキスト独立型話者識別システムに組み込むことにより(図１参照)，システムの識別精度を向上させることを試みた[3]. MFCC と 3000 Hz 付近を局部細分化したケプストラムとを組み合わせた識別実験を行い，MFCC+基本周波数，MFCC+予測残差ケプストラムを組み合わせた場合の識別結果との比較をとおして,生理学的特徴を重視した音響特徴量の話者識別に対する有効性を確認した．本研究では，さらに 300 Hz，または 5000 Hz 付近を局部細分化した識別実験を行い，話者識別精度に対する有効性について考察する．また，生理学的特徴量は，時期差変動に対して頑健性を持つのかを調査した．さらに，これらの識別実験の結果をもとに得られた知見に基づき，新たな話者特徴量として加重等分解度特徴量を提案し，話者識別実験を行った．

以下，第２章では，本研究で用いる特徴量について，第３章では，本研究で用いている認識部であるマルチストリーム GMM モデルに基づく話者識別法について述べる．第４章では，局部細分化ケプストラムを用いたときの識別結果について述べ，第５章では加重等分解度特徴量とその識別結果について，最後に第６章ではまとめと今後の課題を述べる．

2. 音響パラメータの基本設定

本章では，本研究で用いる音響パラメータの基本設定について説明する．

2.1 MFCC

MFCC は，人間の聴覚特性を考慮した音響特徴量であり，低周波数領域の周波数分解能を細かく，高周波数領域を粗くしてスペクトル包絡に関する情報を抽出している手法である．ある意味で，MFCC は音声の全局的な特徴を重視し捉えているといえる．本研究では，MFCC を音響特徴量のひとつとして用いる．次元数として１次元から１４次元までを用いる．また，０次元の一階差分をとり，⊿パワーとして特徴量ベクトルに付加し，全体で１５次元にした．

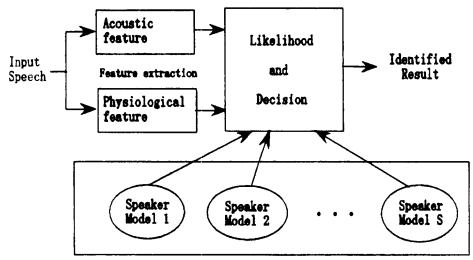


図 1. 生理学的特徴量を考慮した話者識別システム

2.2 基本周波数

基本周波数は話者特徴のひとつとして注目されている[9][10]. 本研究で用いた基本周波数抽出手法は，入力した音声波を線形予測フィルタに通し，線形予測残差信号を算出し，正規化自己相関法によって算出した最大ピーク位置を基本周期として抽出する手法を用いる．有声音・無声音判定には，最大ピークエネルギーに閾値 θ_r を設け，θ_r を超えた場合有声音と判定し，それ以下ならば無声音として判定する．その後，有声音全体にわたって，基本周波数に平滑化処理を施した．

基本周波数は，無声音区間には存在しないため，有声音部と判定された場合に，式（2）のように MFCC に基本周波数成分を付加したものを１個目の話者モデルの特徴量ベクトルとして，２個目の話者モデルの特徴量として，全区間（音声区間）から抽出した MFCC を用いての識別実験を行う．ここで，対数基本周波数を用いた．

$$X^{All} = \{\Delta p, c_1, c_2, \cdots, c_N\} \quad (1)$$

$$X^{Voiced} = \{\Delta p, c_1, c_2, \cdots, c_N, \log F_0\} \quad (2)$$

2.3 線形予測残差ケプストラム係数

線形予測残差ケプストラム係数は話者特徴量のひとつとして用いた研究もあった[7][8][9]. 予測残差信号には,主に音源特性に関する情報を捉えており，もしくは高次の声道形状に関する情報を捉えている可能性もある．線形予測残差ケプストラムの抽出法としては，線形予測残差信号にフーリエ変換を施し，対数パワースペクトルを求めた後，逆フーリエ変換を施し，ケプストラム係数に変換する．線形予測残差信号のスペクトルに調波構造が現れるのは，有声音部のみであるため，音声分析区間を有声音部のみに限定し，無声音部は用いない．なお，次元数は１から１４次元のみを用い，０次元部は用いないこととする．

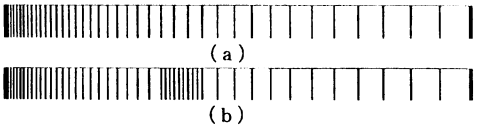


図 4. (a)等メルスケールフィルタバンクを
線形周波数上で表した例
(b)3000Hz付近の各帯域を3等分割した例
(フィルタバンク数40)

2.4 局部細分化ケプストラム法

この手法は我々が先行研究で提案した方法である[3]. 従来の対数関数を用いた MFCC は, 式（3）に基づいた, メルスケール上で等しいフィルタバンクに分割する. それの線形周波数領域での表現を図4に示す. N個のメルスケール帯域に対して平均対数パワースペクトル $m_i(i=1,2,\cdots,N)$ を算出し, 最後に式（4）のような離散コサイン変換を施し, C_i を算出する.

$$MF_f = 2595 \log_{10}(1 + \frac{f}{700}) \quad (3)$$

$$C_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j \cos(\frac{\pi i}{N}(j - 0.5)) \quad (4)$$

目標周波数領域の分析帯域を局部細分化するため, 各メルスケール分析帯域においてD個のサブバンドに再分割する. ここで,着目している領域には, P 個のメルスケール分析帯域が存在し, それぞれの低遮断周波数は $LF_t(t=s,s+1,\cdots,s+P-1)$ とする. s はその領域の先頭帯域とすると, 新しいサブバントを式(5)により求めることができる.

$$NewBand_k = LF_t + k\frac{LF_{t+1} - LF_t}{D} \quad (5)$$

$$(k = 0,1,2,\cdots,D; t = s,s+1,\cdots s+P-1)$$

ここで, 各元のメルフィルタバンクをD等分割することになる. 図4（b）には, 3000 Hz付近の3つの分析帯域幅をそれぞれ3分割した例を示す. 以上の過程で細分化した分析帯域幅を用いて, ケプストラム係数を抽出する. 次元数は, 1次元から14次元までを用い, それに⊿パワーの1次元を付加し, 全体で15次元を特徴量として用いる.

3. マルチストリーム GMM 話者識別法

本研究では, 異なる複数個の特徴量を用いて話者の識別率の向上を図る. 従来の Single GMMモデルに基づく話者識別法を用いた場合において, 異なる複数の特徴量ベクトルをそのまま連結させ次元数を増やすことにより, 話者識別を行うことが可能である. しかし, この場合は特徴量間のレベルの重みを決定するのは困難である. そこで, 本研究では, GMMを複数個に拡張した話者モデル（図5参照）を認識部として用いることにする. いま, p 個の特徴量である. $X=\{x_k \mid k=1,2,\cdots,P\}$ の話者モデル化を考えた場合, ある話者モデルは, 個々の特徴量ごとに GMM モデルを生成し, p 個の完全に独立した話者モデルを構築する. ある話者モデルは, 次式により完全に表せる. ここで, p,μ,Σ はそれぞれガウス分布係数, 平均ベクトル, 共分散行列を表し, M は混合数である.

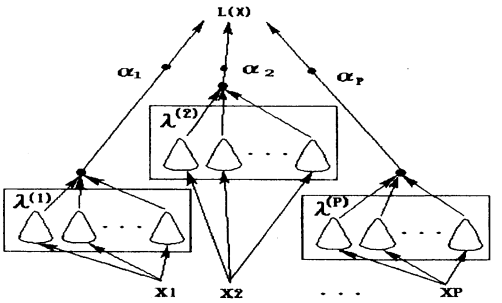


図 5. マルチストリームGMM

表 1. 分析条件

分析区間	32ms	C300の周波数範囲	
分析周期	10ms	TIMIT	200~400 Hz
高域強調	0.97	NTT-VR	200~400 Hz
窓関数	ハミング窓	C3000の周波数範囲	
MFCCパラメータ	14次元	TIMIT	2700~3300 Hz
フィルタバンク数	40	NTT-VR	2300~3700 Hz
予測残差パラメータ	14次元	C5000の周波数範囲	
FFTポイント数	1024	TIMIT	4000~6000 Hz
混合数	32	NTT-VR	4000~6000 Hz

$$\lambda^{(1)} = \left\{p_i^{(1)}, \mu_i^{(1)}, \Sigma_i^{(1)} \mid i = 1,2,\cdots M\right\}$$
$$\lambda^{(2)} = \left\{p_i^{(2)}, \mu_i^{(2)}, \Sigma_i^{(2)} \mid i = 1,2,\cdots M\right\}$$
$$\vdots$$
$$\lambda^{(P)} = \left\{p_i^{(P)}, \mu_i^{(P)}, \Sigma_i^{(P)} \mid i = 1,2,\cdots M\right\} \quad (6)$$

話者判定の際に用いる最終出現確率は, 式 (7) のように各々のモデルの出現確率を係数 α_k で重み付けした線形結合和で表す.

$$L(X) = \sum_{k=1}^{P} \alpha_k \overline{P}(x_k \mid \lambda^{(k)}) \quad (7)$$

$$\sum_{k=1}^{P} \alpha_k = 1 \quad (8)$$

但し, $\overline{P}(x_k \mid \lambda^{(k)})$ は全音声区間における平均対数尤度である. また,すべての k に対し $0.0 \le \alpha_k \le 1.0$ が成り立つ. 話者判定は, 式 (7) の最終出現確率が最も大きい話者を入力音声の話者とする.

4. 局部細分化ケプストラムを用いた話者識別実験

4.1 実験・分析条件

本研究では, 2種類の音声データベース（TIMIT, MTT-VR）を用いて識別実験を行った. 以下にそれぞれのデータベースと分析条件について説明する.

4.1.1 NTT-VR データベース

時期差のある音声データベース（サンプリングレート：16kHz、量子化精度：16bit）として NTT-VR を用いた．話者数は 35 人（男 22 人，女 13 人）である．また、収録時期は 5 時期分(90'8, 90'9, 90'12, 91'3, 91'6)である．予備実験に用いる音声データとしては，テストデータとして 4 時期につき 1 文章(約 4 秒)ずつ用いる．学習データは，テストデータでは用いない残りの 1 時期のみの 10 文章（約 40 秒）のデータを用いる。また，学習において，時期差の異なる音声を順次繰り返して識別実験（5 fold cross validation 法）を行った．全体の総テスト数は，1(回・テスト)×4(時期・テスト)×5（時期・学習）×35(人)=700 回である．本実験に用いる音声データは，予備実験で用いなかったテストデータを用いる．また、テストデータ数を各 4 時期につき 4 文章に増やした．全体の総テスト数は，4(回・テスト)×4(時期・テスト)×5(時期・学習)×35(人)=2800(回)である．

分析条件は，表 1 に記すように設定した．また，マルチストリーム GMM における重み係数は，事後的に最高識別率になるように設定した．

4.1.2 TIMIT データベース

時期差のない音声データベースとしては，TIMIT を用い（16ｋＨｚ、16bit），話者数は 630 人（男 438 人，女 192 人）である．テストデータ長は約 1.5 秒，学習データ長は約 24 秒を用いた．分析条件は，NTT-VR データベースを用いたときと同様である．

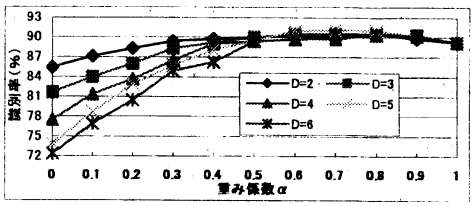
4.2 話者識別実験(１)

4.2.1 細分割値 D および重み係数の決定

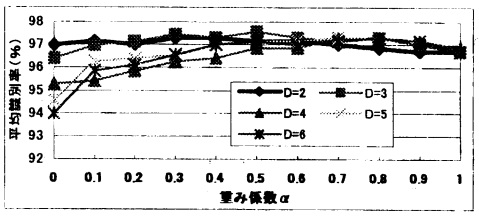
予備実験にて，重み係数および細分割値 D の決定を最高識別率がなるように事後的に行った．300 Hz 付近を局部細分化したケプストラムと MFCC を組み合わせたときの認識結果を図 6 に示す．3000 Hz 付近および 5000 Hz 付近を局部細分化したときの識別結果をそれぞれ図 7 と図 8 に示す．そこで，(a)は音声データベース TIMIT で，(b)は NTT-VR である．なお，縦軸が平均識別率を表し，横軸が重み係数を示す．細分割値は 2,3,4,5,6 と変化させ，重み係数を 0.0 から 1.0 まで 0.1 ステップずつ変化させることによって，特徴量組み合わせの重み係数を求めた．

4.2.2 識別結果

前述のように求めた値を用いて，話者識別精度に対する有効性，または生理学的特徴の時期差による話者内変動に対する頑健性について調査を行った．TIMIT，NTT-VR を用いたときのそれぞれの識別結果を表 2 に記す．また，MFCC をベースに求めた誤り削減率を図 9 に示す．なお，特徴量組み合わせは以下に記す順で行った．なお，Re は予測残差ケプストラムを指し，C300，C3000，C5000 はそれぞれ 300 Hz,300 0Hz, 5000 Hz 付近を局部細分化したケプストラムを指す．識別実験の結果，(IV)鼻・口腔の音響結合を考慮した場合，TIMIT で 29.6%, NTT-VR で 16.3%の誤り削減率が得られた．また，（V）梨状窩を考慮した場合，TIMIT で 45.3 %, NTT-VR で 20.9 %の誤り削減率が得られた．

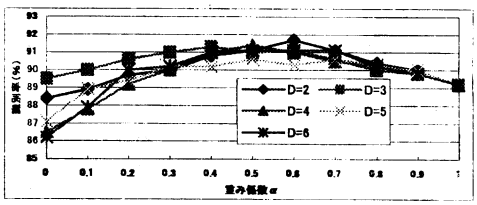


(a) TIMIT

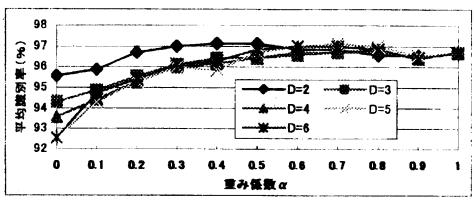


(b) NTT-VR

図 6. 300 Hz 付近を局部細分化した識別結果

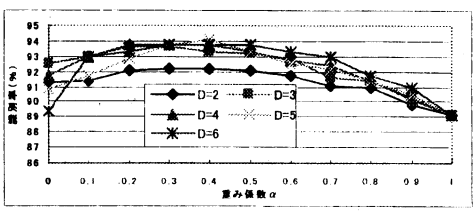


(a) TIMIT

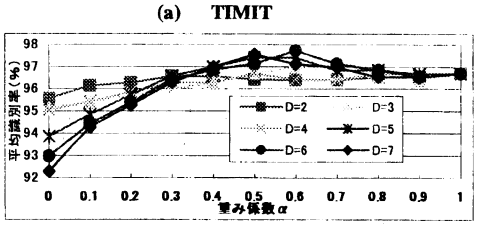


(b) NTT-VR

図 7. 3000Hz 付近を局部細分化した識別結果



(a) TIMIT



(b) NTT-VR

図 8. 5000Hz 付近を局部細分化した識別結果

これらの識別結果は，従来法である(Ⅰ)MFCC のみを用いたとき，または(Ⅱ)MFCC+F0，(Ⅲ)MFCC+Re の識別結果を上回っている．(Ⅷ)すべての生理学的特徴量を組み合した場合では，TIMIT で 65.7 %，NTT-VR で 44.1 %の誤り削減率が得られた．

4.3 考察

音声生成側の知見に基づき話者の音響特徴量を抽出し，従来の話者識別システムに組み込むことにより識別精度を向上させたので，生理学的特徴量の有効性が示された．時期差のある音声データベースにおいても有効性が示され，生理学的特徴量は時期差変動による話者内変動に頑健性を持つことを示した．

5. 加重等分解度特徴量による話者識別実験

5.1 加重等分解度特徴量の概要

前章の局部細分化による認識実験では，高周波数領域では必要な分割数が多くなる傾向から，メルスケールである MFCC より，線形スケールの方が高周波数領域に存在する個人性の詳細を捉えられることがわかった．その結果は，話者の個人特徴を捉えるため，全周波数領域でメルスケールより線形スケールの方が最も適切であろうと示唆している．それゆえ，我々は線形的に等分割したフィルタを用い，かつ生理学的特徴に密接に関連する周波数領域に重みを付けた特徴量を提案し，話者識別に対する有効性について調査を行う．便宜上，この特徴量を加重等分解度特徴量と称する．加重等分解度特徴量を抽出する流れ図を図 10 に示す．この抽出過程では，音声波を周波数変換し，細かい等分解度フィルタ処理を施した後に対数をとり，生理学的特徴を考慮した重み関数を乗算し，最後に離散コサイン変換を施し特徴量ベクトルを求める．重み関数は，以下の指数関数で構築した．

$$w(\omega,\varepsilon_i)=m_i\exp\left(-\frac{(w-CFre_i)^2}{\sigma_i^{\,2}/8}\right)+\beta \qquad (9)$$

$$\hat{w}(\omega,\varepsilon_{300},\varepsilon_{3000},\varepsilon_{5000}) =\max\{w(\omega,\varepsilon_{300}),w(\omega,\varepsilon_{3000}),w(\omega,\varepsilon_{5000})\} \qquad (10)$$

但し，ω は周波数成分を指し，$\varepsilon_i=\{m_i,\beta,CFre_i,\sigma_i\}$ であり，m_i，$CFre_i$，σ_i はそれぞれ強調したい周波数領域における重み，中心周波数，帯域幅を指す．また，フロアリング値 β を設定し，ある程度のパワーを残すことで，パワースペクトルが零になる領域をなくす．提案法は，複数の生理学的特徴を一括に精度良く抽出可能なので，複数の GMM モデルを用いなくて済むという点でメリットがある．こうして求めた生理学的特徴を捉えた特徴量と大局的な声道形状に関する情報を捉えた MFCC を組み合わせて話者識別実験を行った．

5.2 話者識別実験（2）

5.2.1 提案法における次元数の調査

提案法のフィルタは細かく分割されているため，高い次数まで話者の個人性を内包している可能性がある．そこで，NTT-VR を用いて有効次元数の調査の予備実

表 2. 特徴量組み合わせ識別結果

特徴量	TIMIT	NTT-VR
(Ⅰ)MFCC	89.2	95.7
(Ⅱ)MFCC + F0	90.2	96
(Ⅲ)MFCC + Re	92.2	96.2
(Ⅳ)MFCC + C300 + C3000	92.4	96.4
(Ⅴ)MFCC + C5000	94.1	96.6
(Ⅵ)MFCC + C300 + C3000 + C5000	95.4	97.3
(Ⅶ)MFCC + C300 + C3000 + C5000 + Re	96.3	97.6

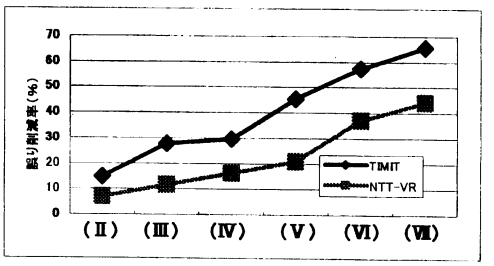


図 9. 特徴量組み合わせに対する誤り削減率
（組み合わせ番号は表 2 を参照）

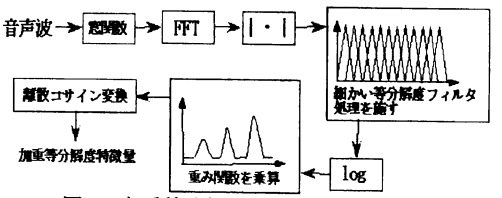


図 10. 加重等分解度特徴量の抽出過程

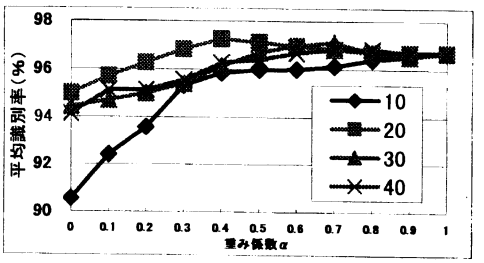


図 11. 次元数の話者識別精度に与える影響

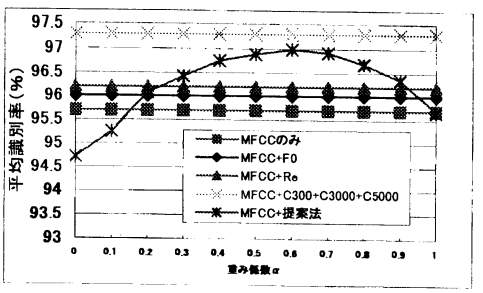


図 12. 提案法と他の手法との識別結果の比較

験を行った．次元数を 10 次元から 40 次元まで 10 次元ずつ変化させたときの識別結果を図 11 に示す．なお今回，線形等分割フィルタにおける帯域幅，重み係数は，局部細分化ケプストラムの実験結果に基づいて経験的に設定した．その結果，帯域幅は 50Hz，重み係数はそれぞれ α_{300}=0.6, α_{3000}=0.4, α_{5000}=0.8, β=0.2 にした．横軸は重み係数を指し，縦軸は平均識別率を指す．識別実験の結果，最高識別率は，次元数 20，重み係数 α=0.4 のときの 97.3%であった．

5.2.2 実験結果

前項で求めた重み係数，次元数をもとに話者識別実験を行った．比較の為に，(Ⅰ)MFCC，(Ⅱ)MFCC+F0，(Ⅲ) MFCC+Re，(Ⅵ) MFCC＋C300＋C3000+C5000 のそれぞれの最高識別率を併記した．縦軸が重み係数を指し，横軸が平均識別率を指す．識別実験の結果，提案法（＋MFCC）の識別結果は重み係数 0.6 のとき 97.0%となり，従来法である(Ⅰ)，(Ⅱ)，(Ⅲ)の識別結果をすべて上回ったが，(Ⅳ)の識別結果 97.3%には 0.3%及ばなかった．

5.3 考察

生理学的特徴を考慮した特徴量として，加重等分解度特徴量を提案し，テキスト独立型話者識別実験を行ったところ，97.0%の識別結果が得られ，従来法である(Ⅰ)，(Ⅱ)，(Ⅲ)の識別結果をすべて上回った．加重等分解度特徴量を用いた場合，局部細分化ケプストラム(Ⅳ)に比べて識別結果は若干劣っている．これは，提案法における重み係数を経験的に設定したため最適ではない可能性があると思われる．しかしながら，話者モデルにおいて，局部細分化ケプストラムを用いた４話者モデルから２話者モデルに削減でき，計算・時間コストも大幅に削減できた．また，最適な重みを決定できれば，さらなる識別精度の向上が期待できる．

6 まとめと今後の課題

本稿では，まず音声生成側の知見に基づき局部細分化ケプストラム法を用いたテキスト独立型話者識別実験を通して，生理学的特徴の話者識別精度に対する有効性を示した．さらに，これらの識別実験から得られた知見に基づき，新たな話者特徴量として加重等分解度特徴量を提案し，その話者識別に対する有効性を調査した．その結果，良い識別結果が得られ，提案法の有効性が示せた．

今後の課題として，提案法において経験的に定めていた重み係数の決定の自動化を考える必要がある．また，認識部において，事後的に重み係数を設定していたので，自動最適化する必要がある．また，実環境下を考慮して，雑音データベースを用いた話者識別実験を行い，生理学的特徴が雑音に対して頑健性を持つかを調査する予定である．

参考文献

[1] J. Dang, K. Honda, "An Improved vocal tract model of vowel production implementing piriform fossa resonance and transvelar nasal coupling," Proc. ICSLP96, pp.965-968(1996).

[2] 北村達也，本多清志，“母音発声時の声道形状における不変部位とその音響特性，” 音講論，Vol.2, pp.337-338 (2003).

[3] 葭原康博，Xugang Lu，党 建武，“生理学的特徴量のテキスト独立型話者識別に対する有効性についての検討，” 信学技報，Vol.104,No.221,pp.1-6(2004).

[4] D. A. Reynolds, "Experimental evaluation of features for robust speaker identification," IEEE Trans, Speech and Audio Processing, Vol.2, pp639-643(1994).

[5] J. P. Openshaw, Z. P. Sun, J. S. Mason, "A comparison of composite features under degraded speech in speaker recognition," ICASSP93, pp.371-374(1993).

[6] Li. Liu, J. He, and G. Palm, "Single modeling for speaker identification, " Proc. ICASSP96, Vol1.1, pp.5-8(1996).

[7] P. Thevenaz, and H. Hugli, "Usefulness of the LPC residue in text-independent speaker verification," Speech Communication, Vol.17, pp.145-157(1995).

[8] J. He, L. Liu, and G. Palm, "On the use of features from prediction residual signals in speaker identification," Proc. EUROSPEECH95 , Vol.1, pp.313-316(1995).

[9] K. P. Markov, and S. Nakagawa, "Integrating pitch and LPC-residual information with LPC-cepstrum for text-independent speaker recognition," The Journal of the Acoustical Society of Japan(E)," 20, 4, pp.281-291(1999).

[10] 松井知子，古井貞熙，“音源・声道特徴を用いたテキスト独立型話者認識，” 電子情報通信学会論文誌，Vol. J75-A, No. 4, pp. 703-709(1992).

[11] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," Speech Communication, vol.17, pp.91-108(1995).

[12] S. Furui , "Cepstrum analysis technique for automatic speaker verification," IEEE trans. Speech, Signal Processing, vol.ASSP-29, pp.1-12(1981).

[13] F. Soong and A. Rosenberg, "On the use of instantaneous and transitional spectral information in speaker recognition," IEEE Trans. Speech.Signal Processing, vol.36, pp.871-879(1988).

[14] Dang, J., and Honda, K., "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. 101, 456-465(1997).