社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2004－SLP－54（30）
2004／12／21

マルチストリーム話者照合における ブースティングに基づく重み最適化法の検討

浅見　太一†　　岩野　公司†　　古井　貞熙†

† 東京工業大学大学院 情報理工学研究科 計算工学専攻
〒 152–8552 東京都目黒区大岡山 2–12–1
E-mail: †{taichi,iwano,furui}@furui.cs.titech.ac.jp

あらまし　本稿では，マルチストリーム HMM を用いた話者照合において，ストリーム重みを自動的に最適化する手法について述べる．我々はこれまで，スペクトル情報と基本周波数情報をマルチストリーム HMM によって融合した，雑音に頑健な話者照合手法を提案している．ここではまず，その手法においてマルチストリーム HMM を利用することの有効性を示す．次に，線形判別分析 (LDA) と Adaboost を組み合わせたストリーム重みの適応化手法を提案する．提案手法の有効性を確認するため，様々な SN 比の白色雑音を重畳した日本語 4 桁連続数字音声による話者照合実験を行った．実験の結果，全ての SN 比において，LDA によって推定されたストリーム重みを用いることで，スペクトル情報のみによって照合を行う場合よりも照合性能が改善した．また，SN 比 10 ～ 30dB という幅広い雑音環境において Adaboost による重み最適化の有効性が確認された．

キーワード　話者照合，耐雑音，マルチストリーム HMM，ストリーム重み推定，Adaboost

A stream-weight optimization method based on boosting for multi-stream speaker verification

Taichi ASAMI†, Koji IWANO†, and Sadaoki FURUI†

† Department of Computer Science, Tokyo Institute of Technology
2–12–1 Ookayama, Meguro-ku, Tokyo, 152–8552 Japan
E-mail: †{taichi,iwano,furui}@furui.cs.titech.ac.jp

Abstract　This paper proposes an automatic stream-weight optimization method for speaker verification using multi-stream HMMs. We have already proposed a noise-robust speaker verification method using multi-stream HMMs for integrating spectral and prosodic information. The paper first shows the effectiveness obtained by using multi-stream technique in our speaker verification framework. Next, a stream-weight adaptation method using both linear discriminant analysis (LDA) and Adaboost is proposed. Experiments were conducted using four-connected-digit utterances of Japanese contaminated by white noise with various SNRs. Experimental results show that 1) the verification performance was improved in all SNR conditions by using stream weights estimated by LDA and 2) the performance is further improved by using Adaboost in 10 - 30dB SNR conditions.

Key words　speaker verification, noise robustness, multi-stream HMM, stream-weight estimation, Adaboost

1. はじめに

話者照合はユビキタス社会での個人認証技術として注目されており，高精度な話者照合システムの需要が高まっている．実用的な話者照合システムを実現するためには，特に雑音に対する頑健性の向上が大きな課題となっている．

話者照合の耐雑音性を高める手法の一つとして，マルチストリーム HMM によって，耐雑音性に優れた情報をスペクトル情報と組み合わせて利用する方法が有効である．我々は，ハフ変換によって雑音に頑健に抽出した基本周波数情報を韻律情報としてスペクトル (音韻) 情報と融合させることによって，話者照合の耐雑音性が向上することを確認した [1]．しかしこれまでは，マルチストリーム HMM のストリーム重みは事後的に手動で最適値に設定していた．実用性を考えると，ストリーム

重みは，システムが使用される環境に合わせて最適な値に自動的に設定される必要がある．そこで本研究では，ディベロップメントセットを用いて素早くストリーム重みを最適化する方法として，線形判別分析 (LDA) と Adaboost [2] を組み合わせたストリーム重みの適応化手法を提案する．Adaboost は複数の識別器を組み合わせることによって精度の高い識別器を構成する手法で，音声認識にこの手法を用いることで認識性能が向上したという報告がなされている [3-6]．特に文献 [4] では複数の特徴量を利用するマルチモーダル音声認識に Adaboost を適用したときの性能向上について報告している．

以下では，まず，音韻情報と韻律情報をマルチストリーム HMM によって融合させて用いる話者照合について説明し，その有効性を検証する．次に，LDA と Adaboost を組み合わせることによって，ストリーム重みを自動的に最適化する手法について述べる．最後に提案手法の有効性を確認する 4 桁連続数字音声を用いた話者照合実験について述べる．

2. マルチストリーム HMM による話者照合

音韻情報と韻律情報を融合したマルチストリーム HMM による話者照合は，次のように行われる [1]．

2.1 音韻・韻律特徴量の融合

音韻特徴量は MFCC 12 次元，Δ MFCC 12 次元，Δ power の計 25 次元を用いる．特徴量はフレーム長 25ms，フレーム周期 10ms で抽出し，入力音声ごとに CMS を行っている．韻律特徴量も音韻特徴量と同じフレーム周期で抽出される．特徴量として $\log F_0$ と $\Delta \log F_0$ の計 2 次元を用いる．F_0 はハフ変換を用いた雑音に頑健な方法 [7] で抽出する．

音韻特徴量と韻律特徴量を結合した計 27 次元を融合特徴量として用いる．

2.2 音韻・韻律モデルの融合

話者照合実験には日本語 4 桁連続数字音声を利用する．日本語連続数字発声では，CV 音節を単位として韻律 (F_0) のパターンを表現するのが容易である．特に 4 桁連続数字音声の F_0 遷移パターンは，図 1 のように，CV 音節を単位として「上昇」「下降」といった韻律ラベルを付与することで表現することができる．そこで，音声認識に用いる音響モデルは，韻律情報付きの CV 音節 HMM として作成する．

全ての数字は 2 つの CV 音節 (2 モーラ) で構成される (「2」は /ni:/，「5」は /go:/ と最終母音が長音化した形で扱う)．ここでは数字内部の音韻環境のみを考慮し，音韻・韻律融合モデルは左右どちらかのコンテキストにのみ依存する音節モデルとする．そこで，融合モデルを，左コンテキスト (LC) 依存の音節 (SYL)「LC-SYL, PM」と，右コンテキスト依存 (RC) 依存の音節 (SYL)「SYL+RC, PM」と表現する．ここで「PM」は F_0 パターンの遷移を示し，上昇 (U)・下降 (D) のいずれかとなる．例えば，「上昇型数字 1 (/ichi/) の第一音節 /i/」は「i+chi, U」と表記される．表 1 に融合モデルの一覧を示す．sil は連続数字の最初と最後に入る無音区間を表現し，sp は数字間に入る短い無音区間を吸収するモデルである．

実際に融合モデルを作るためには，まず，音韻特徴量から音韻 HMM (Segmental HMM，以後 S-HMM と呼ぶ)，韻律特徴量から韻律 HMM (Prosodic HMM，以後 P-HMM と呼ぶ) をそれぞれ別に学習する．そして，両者の混合分布を共有することによって音韻・韻律を融合したマルチストリーム HMM (Segmental-Prosodic HMM，以後 SP-HMM と呼ぶ) を構築する．

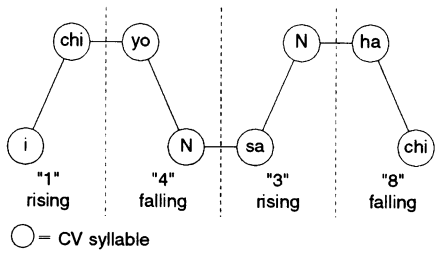


図 1　4 桁連続数字発話の F_0 パターン．

2.3 マルチストリーム HMM

マルチストリーム HMM は，音韻と韻律特徴量を 2 つのストリームに分け，それぞれから得られる出力確率を重み付けし，合わせることで，融合特徴量の出力確率を得る．音韻特徴量 $\boldsymbol{O}_s$ と韻律特徴量 $\boldsymbol{O}_p$ で構成される融合特徴量 $\boldsymbol{O}_{sp}$ が与えられたときの状態 j における出力確率の対数 $b_j(\boldsymbol{O}_{sp})$ は以下の式でフレーム毎に算出される．

$$b_j(\boldsymbol{O}_{sp}) = \lambda_s b_j(\boldsymbol{O}_s) + \lambda_p b_j(\boldsymbol{O}_p) \tag{1}$$

ここで $b_j(\boldsymbol{O}_s)$，$b_j(\boldsymbol{O}_p)$ は状態 j における $\boldsymbol{O}_s$，$\boldsymbol{O}_p$ の出力確率の対数である．λ_s，λ_p はそれぞれ音韻・韻律ストリーム重みであり，$\lambda_s + \lambda_p = 1$ $(0 \leq \lambda_s, \lambda_p \leq 1)$ とする．

2.4 話者照合スコア

特徴量 $\boldsymbol{O}$ が入力されたとき，申告話者 S^c である確率 $p(S^c|\boldsymbol{O})$ は以下のように定義される．

$$p(S^c|\boldsymbol{O}) = \frac{p(\boldsymbol{O}|S^c)p(S^c)}{p(\boldsymbol{O})} \tag{2}$$

ここで，特徴量の生起確率 $p(\boldsymbol{O})$ を，不特定話者モデルからの特徴量の出現確率 $p(\boldsymbol{O}|S^g)$ を用いて表すと，

$$p(S^c|\boldsymbol{O}) = \frac{p(\boldsymbol{O}|S^c)p(S^c)}{p(\boldsymbol{O}|S^g)p(S^g)} \tag{3}$$

となる．各話者について，申告話者の出現確率 $p(S^c)$ は共通であると仮定し，さらに不特定話者モデルの生起確率は定数となるため，

$$p(S^c|\boldsymbol{O}) \propto \frac{p(\boldsymbol{O}|S^c)}{p(\boldsymbol{O}|S^g)} \tag{4}$$

となる．これは，特定話者モデルから得られた尤度を不特定話者モデルから得られた尤度で正規化することを意味している．式 (4) の右辺の分子，分母は，

$$p(\boldsymbol{O}|S^c) = \Sigma_w p(\boldsymbol{O}|S^c, w)p(w) \approx \max_w p(\boldsymbol{O}|S^c, w) \tag{5}$$

$$p(\boldsymbol{O}|S^g) = \Sigma_w p(\boldsymbol{O}|S^g, w)p(w) \approx \max_w p(\boldsymbol{O}|S^g, w) \tag{6}$$

のように計算される．w は 4 桁連続数字列である．これは，通

表 1 融合モデル (SP-HMM) の一覧．融合モデルは「LC-SYL,PM」「SYL+RC,PM」と表記され，「LC-SYL,PM」は左コンテキスト依存の音節モデル，「SYL+RC,PM」は右コンテキスト依存の音節モデルとなる．「PM」は F_0 パターンの遷移を示し，上昇 (「U」)・下降 (「D」) で表現される．

digit	model		digit	model		digit	model	
0	ze+ro,U	ze+ro,D	4	yo+N,U	yo+N,D	8	ha+chi,U	ha+chi,D
/zero/	ze-ro,U	ze-ro,D	/yoN/	yo-N,U	yo-N,D	/hachi/	ha-chi,U	ha-chi,D
1	i+chi,U	i+chi,D	5	go+o,U	go+o,D	9	kyu+u,U	kyu+u,D
/ichi/	i-chi,U	i-chi,D	/go:/	go-o,U	go-o,D	/kyu:/	kyu-u,U	kyu-u,D
2	ni+i,U	ni+i,D	6	ro+ku,U	ro+ku,D		sil	sp
/ni:/	ni-i,U	ni-i,D	/roku/	ro-ku,U	ro-ku,D			
3	sa+N,U	sa+N,D	7	na+na,U	na+na,D			
/saN/	sa-N,U	sa-N,D	/nana/	na-na,U	na-na,D			

常の音声認識に用いられるのと同様の尤度計算を特定話者モデルと不特定話者モデルに対して行い，得られた 2 つの尤度を照合に用いることを意味している．

そこで，話者照合スコア $q(\boldsymbol{O})$ を対数を用いて，

$$q(\boldsymbol{O}) = \log p(\boldsymbol{O}|S^c) - \log p(\boldsymbol{O}|S^g) \tag{7}$$

と定義する．このスコアが閾値 θ を越えたときに，申告話者本人であると判断する．したがって，判別式は，

$$z = q(\boldsymbol{O}) - \theta \tag{8}$$

となり，z が正であれば本人として受理，0 以下であれば詐称者として棄却する．

2.5 マルチストリーム HMM の有効性の検証

まず，音韻情報と韻律情報の融合にマルチストリーム HMM を用いることの有効性を検証する．そこで，

(a) 申告話者・不特定話者モデルの SP-HMM から得られる照合スコア $q_{sp}(\boldsymbol{O}_{sp})$ を閾値と比較して照合を行った場合，

(b) S-HMM から $q_s(\boldsymbol{O}_s)$，P-HMM から $q_p(\boldsymbol{O}_p)$ を独立に計算し，後でそれらに重みを付けて加算した $\omega_s q_s(\boldsymbol{O}_s)+\omega_p q_p(\boldsymbol{O}_p)$ を閾値と比較して照合を行った場合，

の 2 通りについて EER (Equal Error Rate) を比較した．照合スコア q_m の添字 m はスコアを計算したモデル (S-HMM, P-HMM, SP-HMM) を表す．この実験では $m = s, p, sp$ のいずれかとなる．2 通りの話者照合の処理の流れをそれぞれ図 2 の (a), (b) に示す．この実験では，スコアの融合重み ω_s，ω_p とストリーム重み λ_s，λ_p は 0.0 から 1.0 まで 0.1 刻みで変化させて実験を行い，事後的に最適値を選択した．実験に用いたデータは後に述べる話者照合実験に用いたものと同じである．

実験結果を表 2 に示す．全ての SN 比において，S-HMM と P-HMM から別々に得られたスコアを後で融合する方法よりも，SP-HMM によって特徴量の段階で融合を行う方法の方が EER が低くなっており，音韻情報と韻律情報を用いた話者照合において，マルチストリーム HMM を利用することの有効性が確認された．我々の先行研究 [7] では，音声認識実験において，音韻・韻律を組み合わせたマルチストリーム HMM を用いることによって，音韻 HMM を単独で用いるよりも数字のアライメント精度が向上し，その結果，認識性能が改善されることを確認している．話者照合においても同様に，マルチストリーム HMM を用いることによるアライメント精度の向上が性能の改善に有効であったと考えられる．

表 2 各 SN 比における，S-HMM・P-HMM を別々に用いた場合と SP-HMM を用いた場合の EER (%) の比較．

SNR (dB)	S-HMM & P-HMM	SP-HMM
30	0.67	0.65
20	3.46	3.36
15	9.62	9.02
10	17.91	16.28
5	25.62	24.32

3. ストリーム重みの自動推定法

マルチストリーム HMM の利用は話者照合に有効であるが，これまではストリーム重みを事後的に設定していた．実用性を考えると，重みは事前に自動的に決められる必要がある．そこで，LDA と Adaboost を組み合わせて用いることによってストリーム重みを自動的に最適化する手法を提案する．本節では，まず，LDA によってストリーム重みを推定する方法について述べ，次に，この手法に Adaboost を適用して重みの推定値を最適化する方法について述べる．

3.1 LDA によるストリーム重み推定法

2.4 節で述べたように，SP-HMM による照合は判別式，

$$z = q_{sp}(\boldsymbol{O}_{sp}) - \theta \tag{9}$$

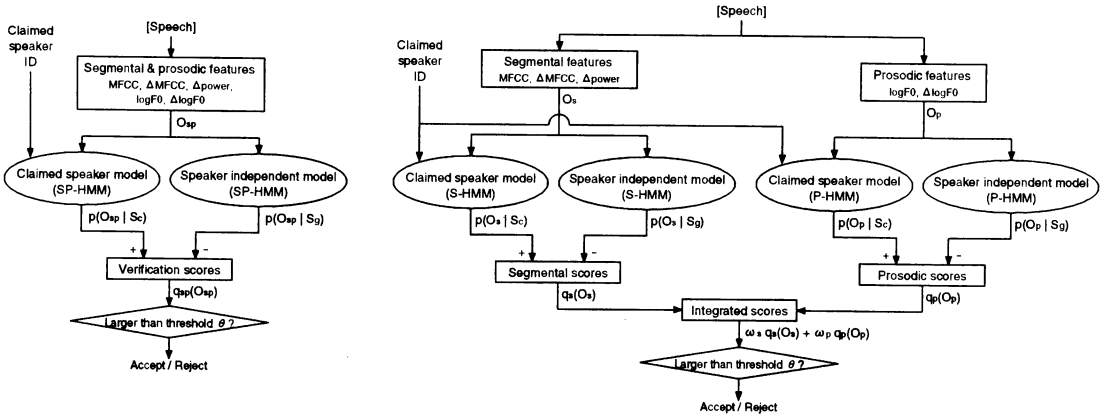
によって行われるが，本研究では重み推定の計算時間削減のため，z を，

$$z = \lambda_s q_{sp}(\boldsymbol{O}_s) + \lambda_p q_{sp}(\boldsymbol{O}_p) - \theta \tag{10}$$

$$\approx \lambda_s q_s(\boldsymbol{O}_s) + \lambda_p q_p(\boldsymbol{O}_p) - \theta \tag{11}$$

と近似する．これは，SP-HMM から計算されるスコア q_{sp} を，S-HMM から得られるスコア q_s と P-HMM から得られるスコア q_p によって近似することを意味する．

この式において，LDA によって z を求めることで λ_s と λ_p を推定する．まず，ディベロップメントセットを用いて $q_s(\boldsymbol{O}_s)$ と $q_p(\boldsymbol{O}_p)$ を個別に計算し，LDA を行う．得られる判別式 $z = \lambda_s q_s(\boldsymbol{O}_s) + \lambda_p q_p(\boldsymbol{O}_p) - \theta$ は $\lambda_s + \lambda_p = 1$ を満たしていないため，次に $q_s(\boldsymbol{O}_s)$ と $q_p(\boldsymbol{O}_p)$ の係数の和が 1 となるように変形を行う．得られた z を



(a) SP-HMM　　(b) S-HMM & P-HMM

図 2　話者照合の処理の流れ．(a) は SP-HMM によって融合特徴量 $\boldsymbol{O}_{sp}$ から照合スコア $q_{sp}(\boldsymbol{O}_{sp})$ を計算して照合を行う場合，(b) は 音韻特徴量 $\boldsymbol{O}_s$ と S-HMM から $q_s(\boldsymbol{O}_s)$，韻律特徴量 $\boldsymbol{O}_p$ と P-HMM から $q_p(\boldsymbol{O}_p)$ を計算し，後で両者を融合する場合を示している．

$$z = (\lambda_s+\lambda_p)\left(\frac{\lambda_s}{\lambda_s+\lambda_p}q_s(\boldsymbol{O}_s)+\frac{\lambda_p}{\lambda_s+\lambda_p}q_p(\boldsymbol{O}_p)-\frac{\theta}{\lambda_s+\lambda_p}\right) \tag{12}$$

と変形すると，z の正負は

$$\frac{\lambda_s}{\lambda_s+\lambda_p}q_s(\boldsymbol{O}_s)+\frac{\lambda_p}{\lambda_s+\lambda_p}q_p(\boldsymbol{O}_p)-\frac{\theta}{\lambda_s+\lambda_p} \tag{13}$$

の正負と等しくなる．そこで，

$$\frac{\lambda_s}{\lambda_s+\lambda_p},\ \frac{\lambda_p}{\lambda_s+\lambda_p},\ \frac{\theta}{\lambda_s+\lambda_p} \tag{14}$$

の各値をストリーム重みと照合の閾値の推定値とする．照合の閾値については，システムの用途によって最適な値が変化するため，今回はストリーム重みのみをこの方法で推定した．

3.2 Adaboost を用いたストリーム重み最適化法

Adaboost [2] は，単純な識別器を複数組み合わせることによって精度の高い識別器を構成する Boosting 法の中でも顕著な性能を示す手法である．以下，LDA で得られる線形判別式を識別器として用いた Adaboost のアルゴリズムと，Adaboost を用いたストリーム重み最適化法について説明する．

3.2.1 Adaboost アルゴリズム

Adaboost では，毎回の繰り返しごとにデータに付けられた重みにしたがって学習セットのリサンプリングを行う．リサンプリングした学習セットを使って識別器を学習し，得られた識別器の精度によって識別器に重みを与え，各データの重みを変更する．変更された重みを用いて，学習セットのリサンプリングから繰り返す．最終的な識別器は，それまでに得られた識別器の重み付き多数決となる．

学習セットのデータ数 n，繰り返し回数 T のときの Adaboost のアルゴリズムは以下のようになる．学習セットを $\{x_i\}$ $(i=1,\cdots,n)$，各データの重みを $\{w_i\}$ $(i=1,\cdots,n)$ とする．

（1） 各データの重みを $w_i := 1/n$ で初期化する．

（2） $t=1,\cdots,T$ で以下を実行する．

i) $\{w_i\}$ を確率分布として，$\{x_i\}$ から重複を許して n 個，重み付きリサンプリングしたものを $\{x'_i\}$ とする．

ii) $\{x'_i\}$ に対して LDA を行い，線形判別式

$$z_t = \lambda_s^{(t)}q_s(\boldsymbol{O}_s)+\lambda_p^{(t)}q_p(\boldsymbol{O}_p)-\theta^{(t)}$$

を得る．

iii) z_t を使って学習セットの全データ $\{x_i\}$ に対して照合を行い，次の重み付き識別誤差 ϵ_t を計算する．

$$\epsilon_t = \sum_{i:x_i\text{を誤識別}} w_i$$

ただし，$0<\epsilon_t\leqq 1/2$ とする．$\epsilon_t > 1/2$ のときは z_t の判断を逆にし，$\epsilon_t := 1-\epsilon_t$ とする．$\epsilon_t = 0$ ならば $w_i := 1/n$ として i) からやり直す．

iv) z_t の重みを $c_t = \frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}$ とする．

v) 次の式によって w_i を更新する．

$$w_i := \begin{cases} w_i \times e^{-c_t} & (i\text{: } x_i \text{ を正しく識別}) \\ w_i \times e^{c_t} & (i\text{: } x_i \text{ を誤識別}) \end{cases}$$

vi) $\sum_{i=1}^{n} w_i = 1$ となるように w_i を正規化する．

（3） 最終的な識別器を z_t の重み付き多数決．

$$z = \sum_{t=1}^{T}\{c_t \times sign(z_t)\}$$

とし，z の正負によって識別を行う．

3.3 Adaboost を用いたストリーム重み推定法

前節で述べた Adaboost アルゴリズムの出力は，線形判別式の形をしていないので，ストリーム重みの推定に直接用いることができない．そこで今回は Adaboost の結果を，

$$z = \sum_{t=1}^{T} \left(c_t z_t \right) \tag{15}$$

と近似した．これにより，

$$\lambda_s^{(boost)} = \sum_{t=1}^{T} \left(c_t \lambda_s^{(t)} \right) \tag{16}$$

$$\lambda_p^{(boost)} = \sum_{t=1}^{T} \left(c_t \lambda_p^{(t)} \right) \tag{17}$$

$$\theta^{(boost)} = \sum_{t=1}^{T} \left(c_t \theta^{(t)} \right) \tag{18}$$

とすると，z は

$$z = \lambda_s^{(boost)} q_s(\boldsymbol{O}_s) + \lambda_p^{(boost)} q_p(\boldsymbol{O}_p) - \theta^{(boost)} \tag{19}$$

と線形判別式の形に表すことができる．

この式を 3.1 節と同様に $q_s(\boldsymbol{O}_s)$ と $q_p(\boldsymbol{O}_p)$ の係数の和が 1 となるように正規化した，

$$\frac{\lambda_s^{(boost)}}{\lambda_s^{(boost)} + \lambda_p^{(boost)}}, \quad \frac{\lambda_p^{(boost)}}{\lambda_s^{(boost)} + \lambda_p^{(boost)}} \tag{20}$$

の各値を音韻・韻律ストリーム重みの推定値とする．

4. 話者照合実験

4.1 実験条件

4.1.1 音声データ

音声データは時期差による変化を考慮し，1 ヵ月毎に 5 時期に渡って収録を行っている．男性話者 36 名が 1 時期に 50 個の 4 桁連続数字を発声しており，音声は 16kHz，16bit で標本化・量子化した．

1 ～ 3 時期目のデータをマルチストリーム HMM の学習セットとし，4，5 時期目のデータをストリーム重みの推定に使うディベロップメントセットと評価セットとして用いる．データは 12 名ずつ 3 グループに分け，各グループを不特定話者モデルの学習セット，ディベロップメントセット，評価セットとして用いる．図 3 に不特定話者モデルの学習に第 2 グループを用いて話者 M01 に対して照合を行うときのデータの使い方を示す．話者 M01 に対しての照合は，不特定話者モデルの学習に第 3 グループを使い，ディベロップメントセットに第 2 グループを使う場合についても実験する．学習セットとディベロップメントセット，評価セットの 3 グループの組み合わせの計 6 通りについて実験を行い，得られた結果の平均によって評価を行う．

学習セットには SN 比 30dB の白色雑音を付加させ，ディベロップメントセットと評価セットは SN 比 5，10，15，20，30dB の白色雑音を付加させたものを用いる．

4.1.2 実験方法

照合を行う際は，まず学習セットを用いて各話者の個人モデルと不特定話者モデルを学習する．このとき不特定話者モデルの学習セットに申告話者が含まれないようにする．次に評価セットと同じ SN 比の雑音が重畳したディベロップメントセットを使ってストリーム重み推定を行う．そして，推定されたストリーム重みを用いてマルチストリーム HMM により照合を行う．このとき，評価セット中の申告話者以外の全話者を詐称者として実験を行い，本人棄却誤り率 (False Rejection Rate) と詐称者受理誤り率 (False Acceptance Rate) を調べる．

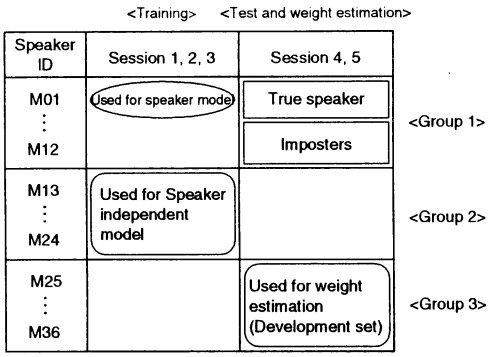


図 3 不特定話者モデルの学習に第 2 グループを用いて話者 M01 に対して照合を行うときのデータの使い方．

ストリーム重みを推定する際，Adaboost の繰り返し回数を 1 ～ 5 まで変化させ，それぞれの場合について照合実験を行った．

HMM の混合数は，申告話者モデル・不特定話者モデルの音韻・韻律ストリームともに 4 とした．これは SN 比 30dB において最も高い性能を示した混合数である．

4.2 実験結果

各 SN 比において，Adaboost の繰り返し回数を変化させ，それぞれの場合で推定されたストリーム重みを用いて照合を行ったときの EER を表 3 に示す．各 SN 比において最も低くなった EER を太字で表記した．表中の Baseline は，マルチストリーム HMM を用いず，音韻情報のみによって照合を行ったときの結果である．また，Adaboost i=1 は LDA のみで推定したストリーム重みを用いて照合を行った場合の結果と等しい．2.5 節の，重みを事後的に設定した実験の結果を表の最右欄に示した．この値よりも本実験の結果の方が EER が低くなっているところがあるが，これは 2.5 節では重みを 0.1 刻みで変化させたのに対して，本実験では重みを 0.01 単位で最適化したためである．

全ての SN 比において Adaboost を用いることでベースラインから EER が減少していることが分かる．繰り返し回数を増やしていくと，EER は増加するが，このときにも極端な性能劣化はなく，ベースラインよりも良い性能を保っている．この結果から，LDA によるストリーム重み推定法と Adaboost を用いた重み最適化法によって，妥当なストリーム重みが得られることが分かる．また，SN 比 10 ～ 30dB という広い範囲の雑音条件において，i=2 のときに最も性能が良くなっている．これにより，LDA によるストリーム重み推定法に Adaboost を適用することによる重み最適化の効果が確認された．SN 比 20dB において，Adaboost の繰り返し回数を変化させたときの Detection Error Tradeoff (DET) カーブの推移を図 4 に示す．Adaboost の繰り返し回数を増やすとカーブが左下方向に移動し，i=2 のときに最も性能が高くなっていることが分かる．

表 3 各 SN 比において，Adaboost の繰り返し回数を変化させたときの EER (%) の比較.

SNR (dB)	Baseline (S-HMM only)	Adaboost i=1 (LDA)	Adaboost i=2	Adaboost i=3	Adaboost i=4	Adaboost i=5	Manually optimized
30	0.88	0.74	**0.67**	0.75	0.77	0.77	0.65
20	4.91	3.54	**3.38**	3.41	3.49	3.48	3.36
15	14.67	9.04	**8.89**	8.90	9.06	8.98	9.02
10	27.10	16.29	**16.27**	16.31	16.73	16.63	16.28
5	37.48	**23.79**	23.89	23.94	24.04	24.01	24.32

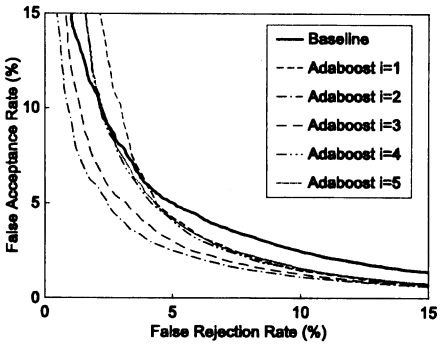


図 4 SN 比 20dB において Adaboost の繰り返し回数を変化させたときの DET カーブの推移.

5. ま と め

本稿では，まず，音韻と韻律を融合した話者照合にマルチストリーム HMM を用いることの有効性を示した．SP-HMM によって照合を行った場合と，S-HMM と P-HMM を独立に使って照合を行った場合の 2 通りについて実験を行った．その結果，全ての雑音環境において SP-HMM を用いた場合の方が EER が低くなり，話者照合にマルチストリーム HMM を用いる効果が確認された．次に，そのストリーム HMM を自動的に推定する手法を提案した．提案手法では，LDA によって自動的に重みを推定する手法に Adaboost を適用することによって重みの推定値を最適化した．4 桁連続数字音声を用いた話者照合実験において本手法の有効性を確認した．提案手法を用いることによって，実験を行った全ての SN 比において妥当なストリーム重みが得られることを確認した．また，SN 比 10 ～ 30dB において，LDA のみで推定した重みを用いるよりも，Adaboost を適用して得られる重みを用いた方が照合性能が高くなり，Adaboost による重み最適化の効果が確認された．

今後の課題としては，1) 照合スコアの閾値を推定する方法の検討，2) 3.1 節で行った近似を用いず，SP-HMM で計算したスコア $q_{sp}(\boldsymbol{O}_{sp})$ によってストリーム重みを推定する方法の検討，3) 3.3 節で行った Adaboost の出力結果の近似をしない重み推定法の検討，4) 白色雑音以外の雑音環境での効果の確認などが挙げられる．

文 献

[1] 浅見太一，岩野公司，古井貞熙，"雑音に頑健な話者照合のための基本周波数情報の利用，" 信学技報，vol.104, no.87, pp.1-6 (2004-5).

[2] Y. Freund and R.E. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Science, 55(1), pp.119-139 (1997).

[3] S.W. Foo and L. Dong, "A boosted multi-HMM classifier for recognition of visual speech elements," *Proc. ICASSP 2003*, vol.2, pp.285-288, Hong Kong (2003-4).

[4] P. Yin, I. Essa and J.M. Rehg, "Boosted audio-visual HMM for speech reading," *Proc. AMFG 2003*, pp.68-73 (2003-10)

[5] C. Dimitrakakis and S. Bengio, "Boosting HMMs with an application to speech recognition," *Proc. ICASSP 2004*, vol.5, pp.621-624, Montreal, Quebec, Canada (2004-5).

[6] C.Meyer, "Utterance-level boosting of HMM speech recognition," *Proc. ICASSP 2002*, vol.1, pp.109-112, Orlando, Florida (2002-5).

[7] Koji Iwano, Takahiro Seki, and Sadaoki Furui, "Noise robust speech recognition using F0 contour information," IEICE Transactions on Information and Systems, vol.E87-D, no.5, pp.1102-1109 (2004-5).