

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閾値最適化法の検討

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、ストリーム重みと話者照合スコアの閾値を自動的に最適化する手法について述べる。提案手法では、スペクトル情報と基本周波数 (F_0) 情報をマルチストリーム HMM によって融合して用いる雑音に頑健な話者照合手法において、線形判別分析 (LDA) と Adaboost を組み合わせることで、ストリーム重みと閾値を同時に最適化する。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。実験では、詐称者受理誤り率 (FAR: False Acceptance Rate) と本人拒否誤り率 (FRR: False Rejection Rate) が等誤り率 (EER: Equal Error Rate) となるように照合スコアの閾値を最適化することを目標とした。実験の結果、SN 比 5 ~ 20dB において、FAR と FRR が EER となるように正しくストリーム重みと閾値を最適化できることを確認した。

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

A threshold optimization method based on Adaboost 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 and threshold optimization method for speaker verification using multi-stream HMMs. In our speaker verification framework using multi-stream HMMs integrating spectral and F_0 features, the proposed method simultaneously optimizes stream-weights and a decision threshold by combining the linear discriminant analysis (LDA) and the Adaboost techniques. Experiments were conducted using Japanese four-connected-digit utterances contaminated by white noise with various SNRs. In the experiments, a target ratio of false acceptance rate (FAR) and false rejection rate (FRR) was set at 1:1 so as to adjust them to approach an equal error rate (EER). Experimental results show that the proposed method effectively estimates the stream-weights and the thresholds in 5 - 20dB SNR conditions.

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

1. はじめに

実用的な話者照合システムを実現するために、雑音に対する頑健性の向上が重要な課題となっている。我々は先行研究において、音韻(スペクトル)情報と韻律(F_0)情報を組み合わせて用いる頑健な話者照合手法を提案している[1]。この手法では、

スペクトル情報と F_0 情報の融合にマルチストリーム HMM を利用している。ハフ変換に基づく雑音に頑健な抽出法[2]によって得られた F_0 特徴量とスペクトル特徴量を組み合わせることで雑音環境において話者照合性能が向上することを確認している。

マルチストリーム HMM を用いた話者照合では、照合の際

にストリーム重みと話者照スコアの閾値を設定する必要がある。実用性を考えると、これらのパラメータは照合の前に設定されることが望ましい [3-5]。我々はこれまで、線形判別分析 (LDA) と Adaboost [6] を用いてストリーム重みを事前に最適化する手法を提案している [3]。これは、ディベロップメントセットの雑音環境に応じて自動的にストリーム重みを最適化する手法で、白色雑音を重畳した 4 桁連続数字音声による話者照合実験において、様々な SN 比で最適なストリーム重みが得られることを確認した。しかしこのとき、照合スコアの閾値については等誤り率 (EER: Equal Error Rate) となるように手動で事後的に設定していた。

そこで本稿では、照合スコアの閾値を自動的に最適化する手法を提案する。閾値の最適値はストリーム重みの値によって変化するため、ストリーム重みと閾値は同時に推定される必要がある。提案手法では、文献 [3] のストリーム重み最適化の枠組みを用いて、ストリーム重みと同時に閾値を最適化する。また、閾値によって詐称者受誤り率 (FAR: False Acceptance Rate) と本人拒否誤り率 (FRR: False Rejection Rate) のバランスが定まるが、最適な FAR と FRR のバランスはシステムの用途によって異なる。例えば、金融・商取引の際の認証など、高いセキュリティが必要とされる場合には FAR を低く抑え、家庭用ロボットが話者を判断する際など、セキュリティよりもユーザビリティを優先する場合には逆に FRR が低くなるように設定する必要がある。提案手法は、システム構築者が意図した FAR と FRR のバランスとなるように閾値の最適化を行うことが可能である。

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、ストリーム重みと話者照スコアの閾値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

提案手法では、音韻情報と韻律情報をマルチストリーム HMM によって融合して話者照合に用いる。以下では、融合方法とマルチストリーム HMM を用いた話者照合手法について述べる。

2.1 音韻・韻律特徴量の融合

音韻特徴量としては 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。

韻律特徴量は $\log F_0$ と $\Delta \log F_0$ の計 2 次元で構成され、音韻特徴量と同じフレーム周期で抽出される。 F_0 はハフ変換を用いた雑音に頑健な方法 [2] で抽出する。

音韻特徴量と韻律特徴量をフレーム毎に結合し、計 27 次元の音韻・韻律融合特徴量を構成する。

2.2 音韻・韻律モデルの融合

本稿では、日本語 4 桁連続数字音声を用いて話者照合を行う。日本語連続数字発声では、CV 音節を単位として韻律 (F_0)

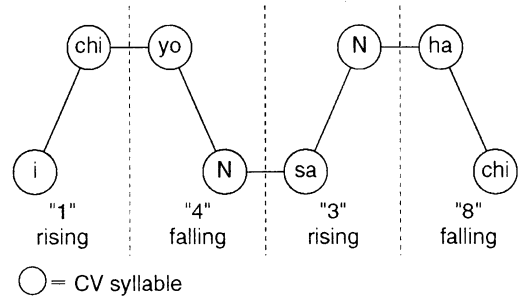


図 1 4 桁連続数字発話の F_0 パターン。

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、話者照合に用いる音響モデルを、CV 音節単位の 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 (/ichu/) 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。

音韻・韻律融合モデル (SP-HMM: Segmental-Prosodic HMM) を作るために、まず、音韻特徴量から音韻 HMM (S-HMM: Segmental HMM)、韻律特徴量から韻律 HMM (P-HMM: Prosodic HMM) を別々に学習する。このとき、音節単位の HMM の状態数は、音韻 HMM で音素数 \times 3、韻律 HMM で 1 とする。次に、学習した S-HMM と P-HMM のガウス混合分布を共有することによって SP-HMM をマルチストリーム HMM として構築する。

2.3 マルチストリーム HMM

マルチストリーム HMM は、音韻特徴量と韻律特徴量を 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て合わせることで、融合特徴量の出力確率を得る。 t フレーム目の音韻・韻律融合特徴量 O_{sp}^t が与えられたときの、状態 j における出力確率の対数 $b_j(O_{sp}^t)$ は以下の式で計算される。

$$b_j(O_{sp}^t) = \lambda_s b_j(O_s^t) + \lambda_p b_j(O_p^t). \quad (1)$$

ここで、 $b_j(O_s^t)$ 、 $b_j(O_p^t)$ は音韻特徴量 O_s と韻律特徴量 O_p から得られる出力確率の対数である。 λ_s 、 λ_p は音韻・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。

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		

2.4 話者照合スコア

特徴量系列 O が入力されたときの話者照合スコア $q(O)$ は次の式によって計算される。

$$q(O) = l(O|C) - l(O|G). \quad (2)$$

$l(O|C)$ と $l(O|G)$ はそれぞれ、申告話者モデル C から得られた対数尤度のフレーム平均、不特定話者モデル G から得られた対数尤度のフレーム平均を表している。

式 (1) を用いると、音韻・韻律融合特徴量 O_{sp} から得られる対数尤度は次のように書くことができる。

$$l(O_{sp}|C) = \lambda_s l(O_s|C) + \lambda_p l(O_p|C), \quad (3)$$

$$l(O_{sp}|G) = \lambda_s l(O_s|G) + \lambda_p l(O_p|G). \quad (4)$$

このとき、話者照合スコア $q(O_{sp})$ は、

$$q(O_{sp}) = \lambda_s q(O_s) + \lambda_p q(O_p) \quad (5)$$

と計算される。

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 $z = q(O_{sp}) - \theta$ 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。話者照合の流れを図 2 に示す。

3. 閾値とストリーム重みの自動推定法

本節では、まず LDA によってストリーム重みと閾値を同時に推定する方法について述べ、次に、この手法に Adaboost を適用してパラメータの推定値を最適化する方法について述べる。

3.1 LDA による推定法

2.4 節で述べたように、SP-HMM による話者照合は次の判別式 z を用いて行われる。

$$z = q(O_{sp}) - \theta \quad (6)$$

$$= \lambda_s q(O_s) + \lambda_p q(O_p) - \theta. \quad (7)$$

z は線形であるため、LDA によって得られる線形判別式の係数からストリーム重み λ_s, λ_p と閾値 θ を推定することができる。

まず、申告話者と詐称者のラベルの付いたディベロップメントセット中の各データについて音韻スコア $q(O_s)$ と韻律スコア $q(O_p)$ を計算し、 $q(O_s)$ と $q(O_p)$ からなる 2 次元空間上にプロットする。SN 比 20dB におけるディベロップメントセット

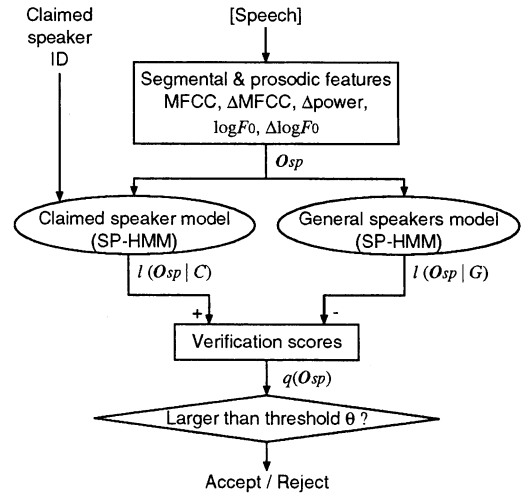


図 2 話者照合の流れ。

のスコア分布の一例を図 3 に示す。図では、横軸に音韻スコア $q(O_s)$ 、縦軸に韻律スコア $q(O_p)$ を取り、○が申告話者のスコア、×が詐称者のスコアを表している。この分布に対して LDA を適用し、申告話者と詐称者を分類する線形判別式 z を得る。図 3 の例では、LDA によって点線で示す直線が得られる。得られた判別式 $z = a_s q(O_s) + a_p q(O_p) - b$ は $a_s + a_p = 1$ を満たさないため、ストリーム重みの和が 1 となるようにこの式を変形する。最終的に、ストリーム重みと閾値の推定値は、

$$\hat{\lambda}_s = \frac{a_s}{a_s + a_p}, \quad \hat{\lambda}_p = \frac{a_p}{a_s + a_p}, \quad \hat{\theta} = \frac{b}{a_s + a_p} \quad (8)$$

となる。このように推定を行うことによって、全てのパラメータは LDA の基準に従って判別性能を最大化するように決定される。

LDA は判別する 2 つの分布の共分散が等しいことを仮定しているため、この推定方法では、ディベロップメントセット中の申告話者のスコア分布と詐称者のスコア分布の共分散が異なる場合に、得られる判別式の性能が劣化してしまう。また、この方法では詐称者受理誤り (FA: False Acceptance) と本人拒否誤り (FR: False Rejection) のバランスを考慮して閾値 θ を推定することができない。これらの問題に対処するため、LDA

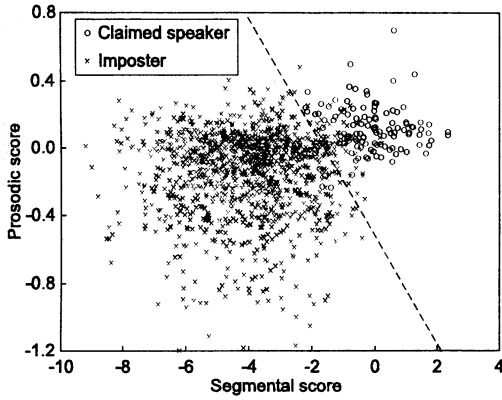


図3 SN比20dBにおけるディベロップメントセットのスコア分布

に Adaboost を適用して推定値の最適化を行う。

3.2 Adaboost を用いた最適化法

Adaboost [6] は、繰り返し学習した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手法である。ここでは、単純な識別器として、LDA で得られる線形判別式を用いる。こうすることによって、LDA のみによって得られる判別式よりも精度の高い判別式が得られる。さらに、FAR と FRR がシステム構築者の意図したバランス、 $FAR:FRR = \alpha:(1-\alpha)$ 、となるようにストリーム重みと閾値を推定することができる。

最適化アルゴリズムを以下に示す。 n はディベロップメントセット中のデータ数、 T は繰り返し回数である。 x_i ($i=1, \dots, n$) は $q(O_s)$ と $q(O_p)$ からなる 2 次元空間にプロットされたディベロップメントセット中のデータ、 w_i ($i=1, \dots, n$) は各データの重みを表す。

- 1 各データの重みを $w_i = 1/n$ で初期化する。
- 2 $t=1, \dots, T$ 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

$$z_t = \lambda_s^{(t)} q_s(O_s) + \lambda_p^{(t)} q_p(O_p) - \theta^{(t)} + \delta_{t-1} \quad (9)$$

を得る。 δ_{t-1} は FAR と FRR の比を $\alpha:(1-\alpha)$ とするためのオフセット量である。オフセット量は $t-1$ 回目の繰り返しで得られる。ただし、 $\delta_0 = 0$ とする。

iii) z_t を使って全学習データ $\{x_i\}$ に対して照合を行い、FAR, FRR, 重み付き識別誤差 $\epsilon_t, \epsilon_{FA}, \epsilon_{FR}$ を以下の式で求める。

$$FAR_t = \frac{\sum_i: x_i \text{ を FA}}{\sum_i: x_i \text{ が詐称者}} \quad (10)$$

$$FRR_t = \frac{\sum_i: x_i \text{ を FR}}{\sum_i: x_i \text{ が申告話者}} \quad (11)$$

$$\epsilon_t = \sum_{i: x_i \text{ を誤識別}} w_i \quad (12)$$

$$\epsilon_{FA} = \frac{\sum_i: x_i \text{ を FA } w_i}{\sum_i: x_i \text{ が詐称者 } w_i} \quad (13)$$

$$\epsilon_{FR} = \frac{\sum_i: x_i \text{ を FR } w_i}{\sum_i: x_i \text{ が申告話者 } w_i} \quad (14)$$

さらに、オフセット量 δ_t を次の式で計算する。

$$\delta_t = \beta \cdot \frac{1 - FAR_t / FRR_t}{1 + FAR_t / FRR_t} \quad (15)$$

- iv) 判別式 z_t の信頼度 c_t を以下のように計算する。

$$c_t = c_{\epsilon_t} \cdot c_{d_t} \quad (16)$$

$$c_{\epsilon_t} = \frac{1}{2} \log \frac{1 - \epsilon_t}{\epsilon_t} \quad (17)$$

$$c_{d_t} = \frac{1}{2} \log \frac{1 - d_t}{d_t} \quad (18)$$

$$d_t = [(1 - \alpha) \cdot FAR_t - \alpha \cdot FRR_t] \quad (19)$$

- v) 次の式によって学習データの重み w_i を更新する。

$$w_i = \begin{cases} w_i \cdot e^{-c_{cost_t}} & (i: x_i \text{ を正しく識別}) \\ w_i \cdot e^{c_{cost_t}} & (i: x_i \text{ を誤識別}) \end{cases} \quad (20)$$

ここで、 c_{cost_t} はコスト関数 $cost_t$ によって次のように計算される。

$$c_{cost_t} = \frac{1}{2} \log \frac{1 - cost_t}{cost_t} \quad (21)$$

$$cost_t = (1 - \alpha) \cdot \epsilon_{FA} + \alpha \cdot \epsilon_{FR} \quad (22)$$

- 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 の重み付き和、

$$z = \sum_{t=1}^T (c_t \cdot z_t) \quad (23)$$

とする。

- 4 ストリーム重みの和が 1 となるように z の係数を正規化する。

- 5 正規化された z の各係数をストリーム重みと閾値の推定値とする。

本来の Adaboost アルゴリズムでは、最終的な識別器 z は、各繰り返しで得られた識別器 z_t の重み付き多数決として $z = \sum_{t=1}^T c_t \cdot \text{sign}(z_t)$ の式によって得られる。しかし、この式は線形判別式ではなく、パラメータの推定に直接用いることができない。そのため、本研究では z を式 (23) のように z_t の重み付き和によって近似した。

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 カ月毎に 5 時期

	<Training>	<Test and development>	
Speaker ID	Session 1, 2, 3	Session 4, 5	
#01 • • • • #12	Used for speaker model	True speaker Imposters	<Group 1>
#13 • • • • #24	Used for general speakers model		<Group 2>
#25 • • • • #36		Used for weight and threshold optimization (Development set)	<Group 3>

図 4 話者 #01 を申告話者とし、グループ 2 を不特定話者モデルの学習セットとして用いた場合のデータの分割。

に渡って収録を行っている。男性話者 36 名が 1 時期に 50 回の 4 桁連続数字を発声しており、音声は 16kHz、16bit で標準化・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と閾値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ことにより、それぞれのデータセット間で話者の重複が生じないようにした。話者 #01 を申告話者とし、グループ 2 を不特定話者モデルの学習に用いる場合のデータの分割方法を図 4 に示す。話者 #01 を申告話者とする場合には、これ以外に、グループ 3 を不特定話者モデルの学習セット、グループ 2 をディベロップメントセットとして用いた実験も行う。学習セット、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には 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モデルと不特定話者モデルを学習する。このとき、音韻 HMM と韻律 HMM の混合数は 4 とした。これは予備的に行った 30dB の白色雑音環境での実験で最も良い性能を示した値である。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、3. 節で述べた方法でストリーム重みと照合スコアの閾値を最適化する。ここで得られたパラメータを用いて評価データからスコアを計算し、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う。

今回は、FAR と FRR を EER に調整することを目標とし、 $\alpha = 0.5$ とした。また、Adaboost の繰り返し回数 T とオフセット量 δ_t を決める係数 β は、それぞれ 200 と 0.005 に設定した。

4.2 実験結果

ストリーム重みと照合スコアの閾値を 3.1 節で述べた LDA

表 2 各パラメータ最適化法による照合結果の比較

Optimization method		SNR (dB)				
		30	20	15	10	5
LDA only	FAR (%)	1.22	4.82	8.15	9.55	10.9
	FRR (%)	0.51	2.68	11.4	27.6	50.9
Adaboost	FAR (%)	1.47	3.72	9.35	17.2	26.4
	FRR (%)	0.68	3.43	8.66	15.6	23.2
Manual optimization	EER (%)	0.73	3.39	9.21	16.3	24.3

のみを用いる方法で推定した場合、3.2 節で述べた LDA と Adaboost を組み合わせた方法で最適化した場合、手動で最適化した場合の各 SN 比における照合結果を表 2 に示す。30dB 以外の全ての雑音環境において、Adaboost によって最適化したストリーム重みと閾値を用いることによって、LDA のみで推定したパラメータを使うよりも FAR と FRR が近付いていることが分かる。30dB で FAR と FRR の差が広がった原因は、誤判別したデータの数が少なく、ディベロップメントセットと評価セット間のスコア分布のずれの影響を大きく受けてしまったためだと考えられる。

SN 比 20dB における、ディベロップメントセットとしてグループ 3、不特定話者モデルの学習セットとしてグループ 2 を用いた場合の、Adaboost の繰り返し回数を 1～200 回の間で変化させたときのコスト関数 $C = (1 - \alpha) \cdot FAR + \alpha \cdot FRR$ の遷移を図 5 の左図、FAR と FRR の遷移を右図に示す。LDA のみで推定した場合 ($t = 1$) から、繰り返し回数が増加するにつれてコスト関数が滑らかに減少しており、ストリーム重みが正しく最適化されていることが分かる。また、同時に FAR と FRR も EER に近付いており、提案手法によって、文献 [3] で述べられているストリーム重み最適化の効果を損なうことなく、同時に閾値も最適化できていることが確認できる。さらにグラフから、50 回程度の繰り返しで結果がほぼ収束し、その後繰り返し回数を増やしても性能が劣化しないことが分かる。

5. まとめ

マルチストリーム HMM を用いた話者照合において、LDA と Adaboost によって、ストリーム重みと照合スコアの閾値を同時に最適化する手法を提案した。白色雑音を重畳した 4 桁連続数字音声を用いた実験において、5～20dB の雑音環境で、提案手法によって FAR と FRR が EER に近づくように正しくパラメータを最適化できることを確認した。さらに、提案手法は原理的には FAR = FRR 以外のバランスにも調整することが可能である。

今後の課題としては、1) FAR と FRR の比 α の値を変化させた際のパラメータ最適化精度の評価、2) より多数のストリームを用いた話者照合における本手法の有効性の確認、3) ラベル付きのディベロップメントセットではなく、申告話者が詐称者か分かっていない評価セットのみでパラメータを最適化する手法の検討などが挙げられる。

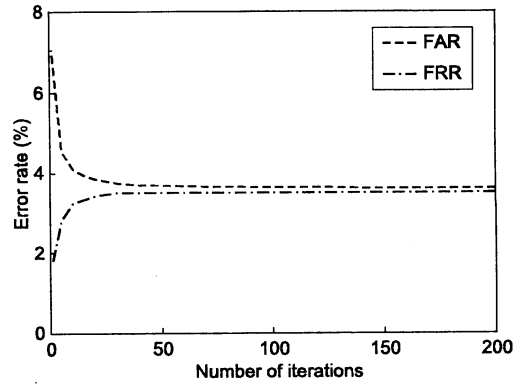
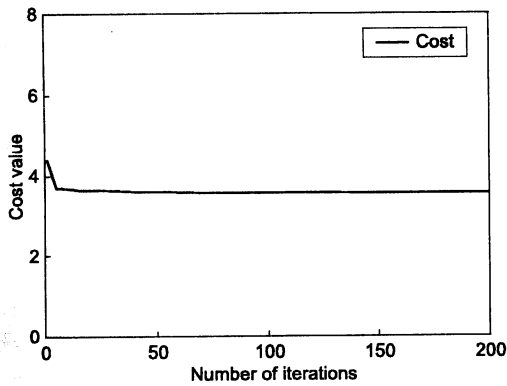


図5 Adaboost の繰り返し回数を変化させたときのコスト関数の遷移 (左) と FAR と FRR の遷移 (右).

6. 謝 辞

本研究の一部は文部科学省科学研究費補助金若手研究 (B) No.17700141 の支援を受けて実施された。

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] K. Iwano, T. Seki and S. Furui, “Noise robust speech recognition using F_0 contour information,” IEICE Transactions on Information and Systems, vol.E87-D, no.5, pp.1102-1109 (2004-5).
- [3] T. Asami, K. Iwano and S. Furui, “Stream-weight optimization by LDA and Adaboost for multi-stream speaker verification,” *Proc. Interspeech 2005*, Lisbon, Portugal, pp.2185-2188 (2005-9).
- [4] 松井知子, 西谷隆, 古井貞熙, “話者照合におけるモデルと閾値の更新法,” 信学論, vol.J81-D-II, no.2, pp.268-276 (1998-2).
- [5] J. Lindberg, J. Koolwaaij, H.-P. Hutter, D. Genoud, J.-B. Pierrot, M. Blomberg and F. Bimbot, “Techniques for a priori decision threshold estimation in speaker verification,” *Proc. RLA2C*, Avignon, France, pp.89-92 (1998-4).
- [6] 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).