

# 残響環境下における話者照合のための 逐次適応を用いたスコア統合重みの推定

中西 亮介<sup>†</sup> 塩田さやか<sup>†</sup> 貴家 仁志<sup>†</sup>

<sup>†</sup> 首都大学東京大学院システムデザイン研究科 〒191-0065 東京都日野市旭が丘 6-6

あらまし 本稿では、スコア統合による話者照合のための残響に頑健なスコア統合重みの推定法を提案する。実環境下において、話者照合システムの性能は背景雑音や残響による影響を受けて低下してしまうという問題がある。そのため背景雑音や残響に対する頑健性を改善するためのさまざまな手法が研究されている。頑健性を改善するために複数の識別器を用い、各識別器の識別結果を統合するスコア統合がある。本研究ではスコア統合で重要となる適切な統合重みを推定するために、残響環境下における複数の SNR モデルを用いた逐次適応を行う。提案法の性能を評価するために、QUT-NOISE-SRE protocol を使用して残響を重畳したデータを用いた話者照合実験を行い、提案法は残響に対して頑健であることを確認した。

キーワード 話者照合, GMM-UBM, スコア統合, 重み推定, 残響環境下

## Speaker verification using weight-adapted score fusion in noisy reverberant environments

Ryosuke NAKANISHI<sup>†</sup>, Sayaka SHIOTA<sup>†</sup>, and Hitoshi KIYA<sup>†</sup>

<sup>†</sup> Department of Information and Communication Systems Engineering, Tokyo Metropolitan University  
6-6 Asahigaoka, Hino-shi, Tokyo, 191-0065 Japan

**Abstract** This paper proposes a weight estimation method for automatic speaker verification (ASV) using score fusion in noisy reverberant environment. It is well known that background noise and reverberation affect the performances of ASV systems. Various techniques have been reported to improve robustness against these background noise and reverberation, and score fusion is one of the effective methods. However, the score fusion requires an adequate weight estimation method. The proposed method estimates an adequate weight in noisy reverberant environment by using a supervised adaptation with some SNR models. The QUT-NOISE-SRE protocol is used for simulating the reverberation of the clean speech in the experiments. The experimental results show that the proposed weight estimation method provides a high performance in noisy reverberant environment.

**Key words** speaker verification, GMM-UBM, score fusion, weight estimation, noisy reverberant environments

### 1. ま え が き

携帯電話や ATM のセキュリティとして生体認証の使用が普及しつつある。生体認証には指紋や虹彩、静脈、顔画像などを用いたものがある。その中でも声を用いた生体認証である話者照合はマイクのような簡単な装置で実現可能であることや、ユーザが使用するために事前練習を行う必要がないことから生体認証の中でもユーザが使用するに対する抵抗が少ない方法となっている。近年では、話者照合の手法として i-vector [1] や PLDA (Probabilistic linear discriminant analysis) [2-4] といったものが提案されており照合性能も大幅に向上してきているため更なる普及が期待されている。しかしながら、実際に話

者照合システムを使用する場合、システムに入力された音声は背景雑音および残響の影響を受ける可能性がある。学習モデル間の環境の不一致は話者照合システムの性能低下の要因となることが報告されているが、入力音声の環境をすべて学習することは難しい。そのため、背景雑音や残響に対する頑健性を向上するためのさまざまな手法が研究されている [5-7]。これらの手法は大きく分けると特徴量に関する手法 [8-13] とモデルに関する手法 [14-20] の二種類に分類することができる。特にモデルに関する手法は、バックエンドで処理を行うものが多い。その手法のひとつに話者モデルと入力音声の雑音や SNR の条件を一致させる手法がある。しかし、実際に用いる場合は入力音声の雑音や SNR が未知であるために、SNR の推定を行う必要

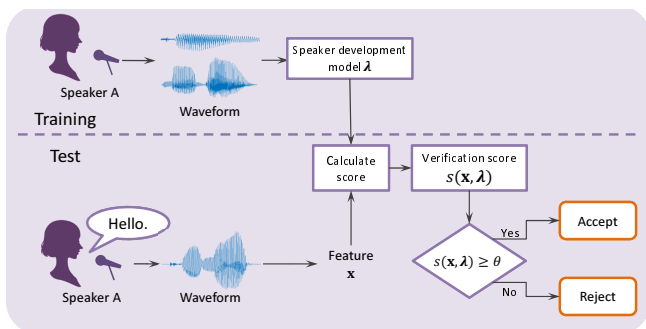


図 1: 話者照合システムの概要

がある。また、特定話者モデルの SNR が入力音声に対して一致するときに必ずしも最も性能が良い SNR となるとは限らないという問題もある。

一方で、モデルに関するアプローチの別の手法に、複数の識別器から得られた識別結果の統合がある。これは認識結果の統合および特徴量の統合、スコアの統合に分けることができる。スコア統合は複数の識別器から出力されたスコアを統合することで認識精度を改善する手法である。本稿ではあらかじめ SNR ごとに学習した話者モデルを用いて入力音声に対する照合スコアを計算し、各 SNR モデルのスコアを統合する手法および逐次適応を用いたスコア統合重みの推定アルゴリズムを提案する。また、雑音重畳には QUT-NOISE-SRE protocol [21] を使用し、残響を重畳した音声と残響を重畳していない音声のそれぞれで評価実験を行い、提案法の有効性を確認した。

## 2. 話者照合

### 2.1 GMM-UBM に基づく話者照合 [22]

話者照合システムの概要を図 1 に示す。図 1 に示すとおり、話者照合は登録部と照合部の二つに分けられており、登録部ではあらかじめ登録ユーザの声から特定話者モデルを学習しておく。照合部では、入力された音声から特徴量を抽出し、得られた特徴量  $\mathbf{x}$  を登録時に学習されたユーザモデル  $\lambda_U$  と詐称者モデル  $\lambda_I$  に入力し、それぞれのモデルから入力音声のユーザらしさ、詐称者らしさを尤度として  $p(\mathbf{x}|\lambda_U)$ ,  $p(\mathbf{x}|\lambda_I)$  で算出する。最後に  $p(\mathbf{x}|\lambda_U)$ ,  $p(\mathbf{x}|\lambda_I)$  から照合スコア  $s(\mathbf{x}, \lambda)$  を計算し、閾値  $\theta$  と  $s(\mathbf{x}, \lambda)$  を比較することで判定を行う。照合スコアの計算式は以下のとおりである。

$$s(\mathbf{x}, \lambda) = \log \frac{p(\mathbf{x}|\lambda_U)}{p(\mathbf{x}|\lambda_I)}. \quad (1)$$

### 2.2 複数モデルによるスコア統合

図 2 に複数の識別器を用いたスコア統合の例を示す [23]。スコア統合は複数の識別器から出力されたスコアを組み合わせることでシステムの精度を高める手法である。手順としてはまずはじめに、あらかじめ複数の識別器を学習しておき、テストデータが入力されたときには各識別器との照合スコアがそれぞれ計算される。次に識別器ごとに得られたスコアの統合を行い、最終的に統合されたスコアを用いて照合を行う。スコアの統合時に、各モデルから計算されたスコアをどのように統合するかを決める必要がある。あらかじめ教師ありデータを用いて重み

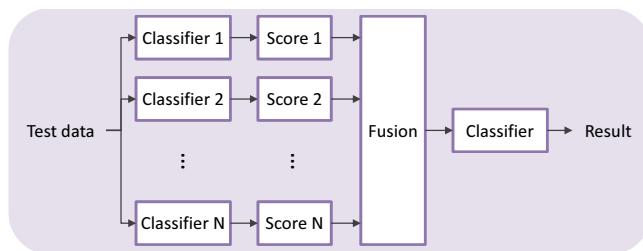


図 2: スコア統合の概要

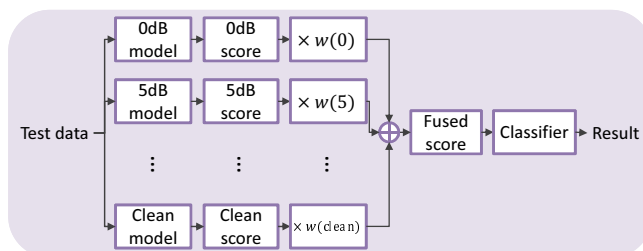


図 3: SNR ごとに学習した話者モデルを用いたスコア統合のフローチャート

の学習を行い逐次的に更新する手法を Boosting と呼び、その代表的なアルゴリズムに AdaBoost [24] がある。また重みの事前学習を行わず、各スコアに対する重みを同一とする手法を Bagging [25] と呼ぶ。Bagging は重みの推定を行わないために処理が早いですが、十分な精度が出ないという問題がある。一方 AdaBoost は重みの事前学習を行うため、モデルごとにより適切な重みを学習できるという利点がある。重みの推定法にはさまざまな手法が提案されているが、処理が複雑になってしまうという問題がある。

## 3. 提案法

### 3.1 複数の SNR モデルを用いたスコア統合

SNR ごとに学習した話者モデルを用いたスコア統合について説明する。図 3 に提案法によるスコア統合のフローを示す。ここで SNR ごとに学習した話者モデルは、図 2 の Classifier に相当する。本稿では逐次適応を用いたスコア統合重みの推定手法を提案する。教師あり評価データを用いて SNR ごとの特定話者モデルとのスコアを計算し、照合精度が高くなるように逐次的に更新を行う。

### 3.2 スコア統合重みの推定法

図 4 にスコア統合重みのアルゴリズムを示す。スコア統合重みの推定手順は (1) 前処理, (2) 重みの初期化, (3) 重みの推定の 3 つに分けられる。まず (1) 前処理として、SNR ごとに評価データと話者モデルとの重み学習用スコア  $s_m(i)$ ,  $i = 1, 2, \dots, N$  を計算する。ここで  $m \in M$  は重畳した SNR を示す。次に、(2) 各 SNR モデルに対するスコア統合重み  $w(m)$  の初期化を行う。本稿では  $m$  dB モデルに対する重み  $w(m)$  の初期値は、

$$w(m) = \frac{1}{M} \quad (2)$$

とする。(3) の重みの学習はさらに (i) 重みの更新量の計算, (ii) 重みの仮更新, (iii) 仮更新後の重みの正規化, (iv) 重みの

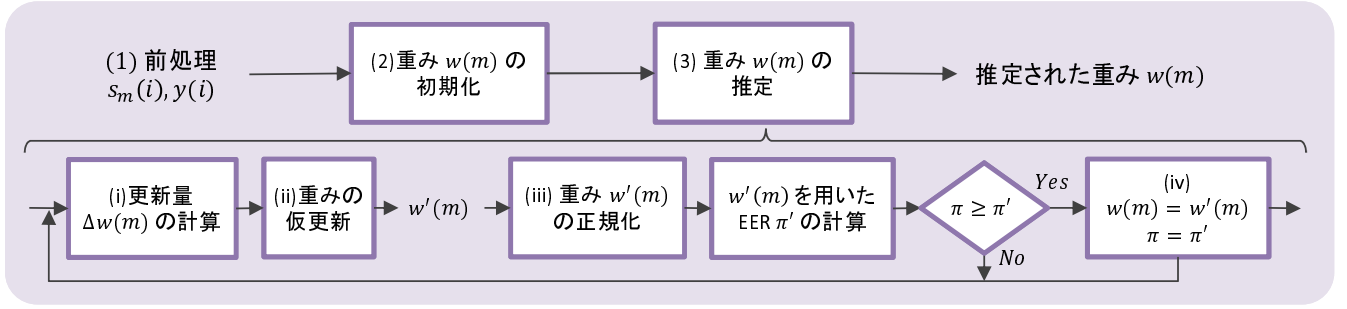


図 4: スコア統合重みの推定法のアルゴリズム

更新の 4 段階に分けられる。はじめに, (i) 各 SNR のスコアに対する重みの更新量を計算する。重みの更新量  $\Delta w(m)$  は,

$$\Delta w(m) = s_m(i) \cdot \rho \cdot y(i) \quad (3)$$

により計算される。ここで,  $\rho$  は重みの更新率,  $y(i)$  は教師データが登録話者か否かのラベルを表す。更新量  $\Delta w(m)$  を計算したあと, (ii) 重み  $w(m)$  を仮更新する。仮更新後の重み  $w(m)'$  は,

$$w(m)' = w(m) + \Delta w(m) \quad (4)$$

により計算される。SNR ごとに重みを仮更新したあと, (iii) 重み  $w(m)'$  の正規化を行う。正規化式を以下に示す。

$$w(m)' = \frac{w(m)'}{\sum_{k \in M} w(k)'}. \quad (5)$$

仮更新した重みを用いて (iv) 重みを更新するか判定を行う。正規化された重み  $\Delta w(m)$  を用いてスコアを統合し, 統合されたスコア  $s'_m(i)$  を用いて等価エラー率 (Equal Error Rate; EER) を計算する。EER は算出されたスコアから本人拒否率 (False rejection rate; FRR) と他人受入率 (False acceptance rate; FAR) を計算し, 全話者共通の閾値を設定して求められる値である。仮更新後の重みから算出した EER  $\pi'$  を更新前の重みから算出した EER  $\pi$  と比較し,  $\pi' \leq \pi$  であれば重みを  $w(m) = w(m)'$  とする。以上の 4 つの手順を評価データを用いて繰り返し, 各 SNR の最終的な重み  $w(m)$  を決定する。

## 4. 実験

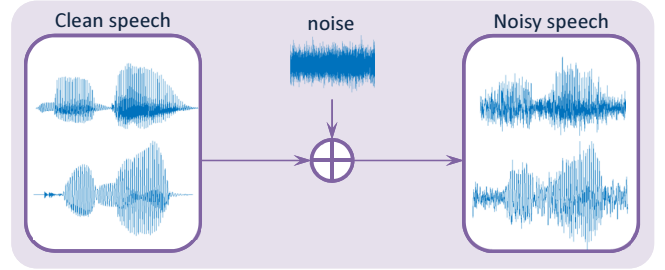
提案法の有効性について考察するために, 話者照合実験を行った。

### 4.1 データベース

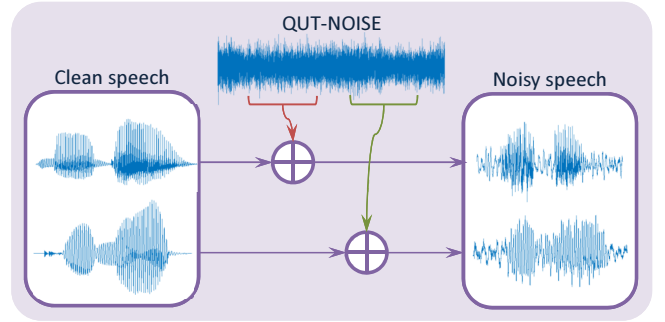
特定話者モデル用学習データおよびテストデータ, 評価データには VLD データベース [26] からポップフィルタを装着したヘッドセットマイク (SHURE SM10A-CN) で収録されたものを使用した。UBM は, SNR を 0dB から 30dB で 5dB きざみで重畳した雑音重畳音声と clean 音声をすべて混ぜてマルチコンディション学習を行った。学習用データへ重畳する雑音データベースは電子協騒音データベース [27] から走行自動車内 (1500cc クラス) 雑音を使用した。また, 雑音の重畳には Filtering and Adding Noise Tool (FaNT) [28] を使用した。

#### 4.1.1 QUT-NOISE database [29]

QUT-NOISE database は各 30 分以上の 20 のノイズセッション



(a) FaNT による雑音重畳



(b) QUT-NOISE-SRE protocol を用いた雑音重畳

図 5: 各雑音重畳手法のフロー

ンで構成されている雑音データベースである。CAFE, HOME, STREET, CAR, REVERB の 5 つのシナリオから成り, 収録は 1 シナリオにつき 2 箇所, 計 10 箇所で行われている。CAR シナリオを除いて, 各場所で録音日ごとにセッションが分けられている。また, CAR シナリオと REVERB シナリオは残響環境下で収録されている。本実験では特定話者モデル用学習データおよびテストデータ, 評価データへ重畳する雑音は QUT-NOISE database の CAR シナリオを使用した。雑音重畳には QUT-NOISE-SRE protocol を使用した。

QUT-NOISE-SRE protocol は clean 環境で収録された音声データベースに QUT-NOISE database を重畳するためのプロトコルである。図 5(a) に FaNT による雑音重畳, 図 5(b) に QUT-NOISE-SRE protocol を用いた雑音重畳のフローをそれぞれ示す。FaNT による雑音重畳手法では, clean 音声に雑音データを乗せるときにまったく同じセグメントの雑音データを重畳していた。一方, QUT-NOISE-SRE protocol は clean 音声に雑音データを乗せるときに QUT-NOISE database の選択したシナリオの雑音データから重畳するセグメントをランダムに選択する。また, CAR シナリオと REVERB シナリ

表 1: 実験条件

登録話者データベース	VLD データベース
学習データ (特定話者モデル)	70 文章 × 17 名 (計 1190 文章)
テストデータ	20 文章 × 17 名 (計 340 文章)
評価データ (重み学習用データ)	10 文章 × 17 名 (計 170 文章)
UBM 用データベース	JNAS (女性のみ)
UBM 学習データ	23657 文章 × 7 SNR (計 165599 文章)
GMM 混合数	1024
サンプリング周波数	16 kHz
フレーム長	25 msec
フレームシフト	10 msec
特徴量	MFCC 19 次 + $\Delta + \Delta\Delta$

表 2: 比較する手法

手法	特定話者モデル	スコア統合
MC	テストデータの SNR と同 SNR のモデル	—
BC	テストデータに対して EER の最も低いモデル	—
Uniform	各 SNR モデルでスコア計算後, 統合	統合重みは一定
Proposed method	各 SNR モデルでスコア計算後, 統合	評価データに対する EER が下がった場合に重みを更新

オの雑音を重畳する場合, clean 音声に残響を重畳したあとに雑音データを重畳する. このように残響を考慮することで QUT-NOISE-SRE protocol はより実環境に近い雑音環境を表すことができる.

#### 4.2 実験条件

実験条件を表 1 に示す. 雑音環境下における複数の異なる SNR のテストデータが入力された場合の性能を比較するために, CAR シナリオ雑音を QUT-NOISE-SRE protocol を用いて, SNR を 0 dB から 30 dB で 5 dB きざみで重畳した. 重畳した SNR を  $M = \{0, 5, 10, 15, 20, 30, \text{clean}\}$  とする. ただし clean は雑音を重畳していない音声を表す. また, 残響環境下での性能比較をするために, 残響を重畳した音声データと重畳しない音声データを作成し, それぞれの条件で同様の実験を行った. このとき残響を重畳した場合と重畳しない場合の提案法における重みの更新率  $\rho$  はそれぞれ 0.003, 0.005 とした.

表 2 に話者照合実験で比較する各条件を示す. 特定話者モデルとテストデータの収録条件を合わせたものをマッチドコンディション (Matched Condition; MC), テストデータに対してもっとも EER が低くなる SNR の特定話者モデルを 1 つ選択し用いるものをベストコンディション (Best Condition; BC) とする. Uniform は統合重みをすべての SNR において  $w(m) = 1/M = 1/7 \simeq 0.143$  としてスコアを統合する.

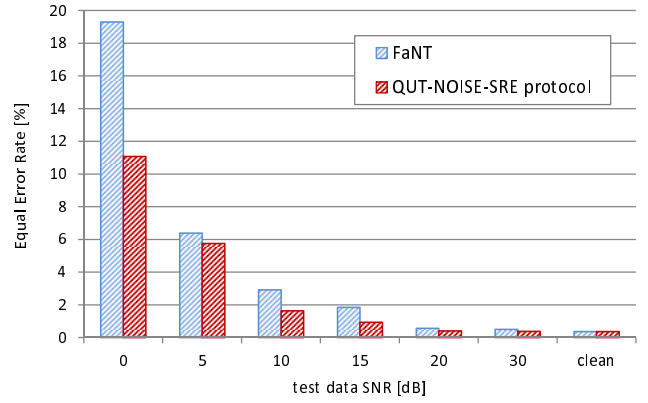


図 6: 2つの雑音重畳プロトコルによるテストデータの SNR ごとの EER

### 4.3 実験結果

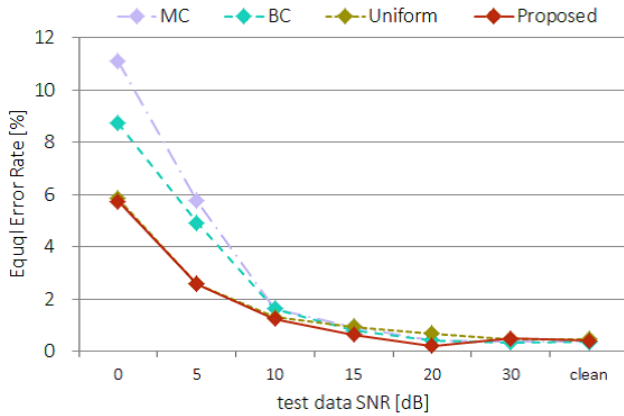
#### 4.3.1 雑音重畳プロトコルのちがいによる比較

はじめに FaNT による雑音重畳と残響を重畳していない場合の QUT-NOISE-SRE protocol による雑音重畳の 2つのプロトコルによる照合結果のちがいを比較するための話者照合実験を行った. どちらのプロトコルも特定話者モデルとテストデータの条件は合わせている. FaNT による雑音重畳と QUT-NOISE-SRE protocol による雑音重畳のテストデータの SNR ごとの EER を図 6 に示す. 図 6 より, 全 SNR のテストデータにおいて QUT-NOISE-SRE protocol による雑音重畳を行った場合の EER が FaNT による雑音重畳のものよりも低くなっている. これより, QUT-NOISE-SRE protocol は学習データに使用される雑音の種類が多いため, FaNT と比較してより頑健な話者モデルを学習できるという特徴があることがわかる.

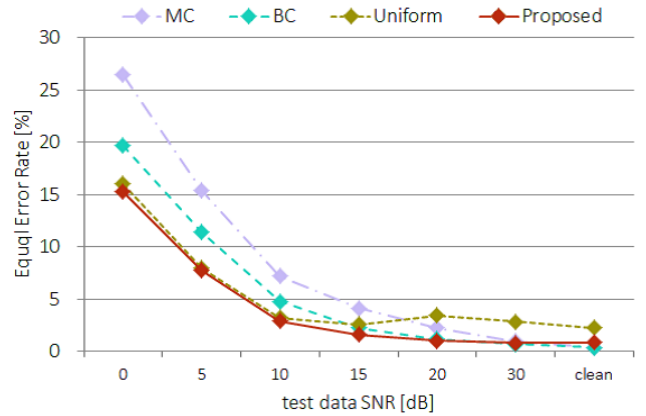
#### 4.3.2 残響を重畳しない場合

図 7 (a) に雑音だけを重畳し, 残響は重畳していない場合の MC および BC, Uniform, 提案法によるテストデータの SNR ごとの EER を示す. まず, MC と BC を比較すると, テストデータの SNR が 0, 5, 15, 30 dB のときに BC の EER は MC の EER よりも低くなっている. このことから話者モデルとテストデータの SNR が一致しているときの EER が最も低い EER であるとは限らないことが確認できる. 次に, BC と Uniform, 提案法による実験結果を比較する. テストデータの SNR が 20dB 以下のときに Uniform および提案法は BC よりも EER が低く, 30dB 以上のときも BC とほぼ等しくなっている. 以上より, 提案法を含むスコア統合手法は雑音に頑健であることがわかった.

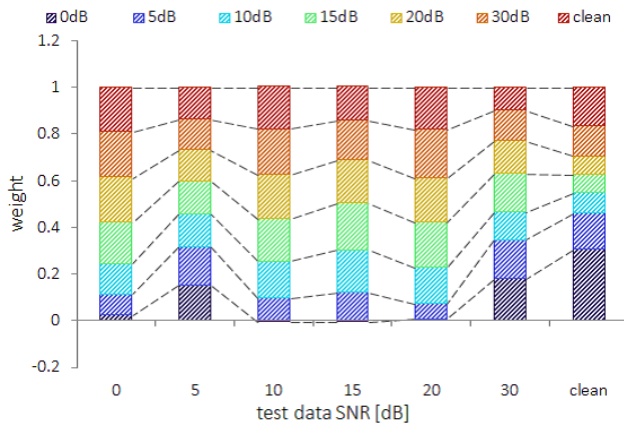
次に提案法により最終的に推定されたスコア統合重みを図 7 (b) に示す. 各項目は提案法により推定された, 各 SNR モデルから得られたスコアに対する重みをテストデータの SNR ごとに表している. 横軸はテストデータの SNR, 縦軸は最終的に推定された重みの値を表している. 図 7 (b) より, テストデータの SNR ごとに推定される重みにばらつきがあることがわかる. 特に 0dB のモデルの重みが大きく変化している.



(a) 提案法やその他手法によるテストデータの SNR ごとの EER

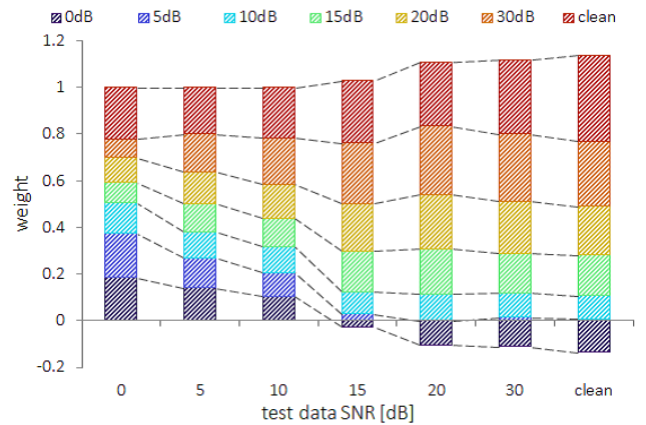


(a) 提案法やその他手法によるテストデータの SNR ごとの EER



(b) 提案法の最終的な重み

図 7: 残響を重畳しない場合の EER と推定された重み



(b) 提案法の最終的な重み

図 8: 残響を重畳した場合の EER と推定された重み

#### 4.3.3 残響を重畳した場合

図 8 (a) に雑音だけでなく残響も重畳した場合のテストデータの SNR ごとの各手法における EER を示す. 図 7 (a) と比較すると, 残響が重畳された場合は全体的に EER が高くなるのがわかる. BC と Uniform, 提案法を比較すると, 残響がない場合と同様にテストデータの SNR が 15dB 以下のときに Uniform および提案法の EER は BC よりも低くなっている. テストデータが低 SNR の場合は残響環境下においてもスコア統合が有効であることがわかる. しかし, MC と Uniform, 提案法を比較すると, テストデータの SNR が 20 dB 以上のときに Uniform の EER は MC よりも高くなってしまふ. 残響環境下において, Uniform はテストデータの SNR が高いときに照合性能が低下してしまうことがわかる. 一方, 提案法は clean を除く全 SNR のテストデータにおいて MC よりも EER が低くなっている. Uniform と比較しても, 提案法の方が高い性能を得られている. このことから, 提案法である重みの推定は残響環境下においても雑音に頑健であることがわかった.

提案法により最終的に推定された重みを図 8 (b) に示す. テストデータの SNR が低いときは推定後の重みの変動は小さく, 図 8 (a) から Uniform と提案法の EER の値は近いことがわかる. 一方, テストデータの SNR が高いときは低 SNR モデルから得たスコアの重みは小さく, 高 SNR モデルから得た

スコアの重みが大きくなるように重みが学習されている. このため提案法はテストデータが高 SNR の場合において Uniform で起こるような照合性能の低下がなく, BC よりも良い照合性能となったことが考えられる. 以上より, 提案法によるスコア統合重みの推定は残響環境下においても頑健であることがわかった.

## 5. むすび

本稿ではあらかじめ SNR ごとに別々に学習した話者モデルから得られた照合スコアを統合する手法と, そのスコア統合重みの逐次適応を用いた推定手法を提案した. 残響環境下における提案法の頑健性を調査するために, QUT-NOISE-SRE protocol を使用して clean 音声に残響を重畳した. 残響を重畳したものと重畳しないものを作成して, それぞれの環境下で GMM-UBM を用いた話者照合実験を行った. 実験結果より, 提案法は残響環境下においても照合性能の低下が少なく, 雑音に対して頑健であることがわかった.

今後の課題は, 実験に使用するデータを増やすことや, 評価関数を基とした重みの推定手法の検討が挙げられる. また, 本実験では特定のドメインから評価データを選択して実験を行ったが, 別ドメインの評価データを用いた場合に実験結果が変化することが考えられる. そのため別ドメインの評価データを用

いた場合にどれだけ結果が変わるか、ドメインの依存性について調査する。

## 文 献

- [1] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788–798, 2011.
- [2] S. Ioffe, "Probabilistic linear discriminant analysis," *Computer Vision–ECCV 2006*, pp. 531–542, 2006.
- [3] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, 2007.
- [4] P. Kenny, "Bayesian speaker verification with heavy-tailed priors." *Proc. Odyssey 2010 - The Speaker and Language Recognition Workshop*, 2010.
- [5] M. I. Mandasari, M. McLaren, and D. A. Van Leeuwen, "The effect of noise on modern automatic speaker recognition systems," *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 4249–4252, 2012.
- [6] R. Togneri and D. Pullella, "An overview of speaker identification: Accuracy and robustness issues," *Circuits and systems Magazine, IEEE*, vol. 11, no. 2, pp. 23–61, 2011.
- [7] J. Ming, T. J. Hazen, J. R. Glass, and D. A. Reynolds, "Robust speaker recognition in noisy conditions," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 5, pp. 1711–1723, 2007.
- [8] Y. Shao, S. Srinivasan, and D. Wang, "Incorporating auditory feature uncertainties in robust speaker identification," *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4, pp. IV–277, 2007.
- [9] S. O. Sadjadi, T. Hasan, and J. H. Hansen, "Mean hilbert envelope coefficients (mhec) for robust speaker recognition." *INTERSPEECH*, pp. 1696–1699, 2012.
- [10] Y. Shao and D. Wang, "Robust speaker identification using auditory features and computational auditory scene analysis," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 1589–1592, 2008.
- [11] Q. Li and Y. Huang, "Robust speaker identification using an auditory-based feature," *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 4514–4517, 2010.
- [12] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," *International Speech Communication Association (ISCA)*, 2001.
- [13] X. Zhao, Y. Wang, and D. Wang, "Robust speaker identification in noisy and reverberant conditions," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 4, pp. 836–845, 2014.
- [14] R. Saeidi, K.-A. Lee, T. Kinnunen, T. Hasan, B. Fauve, P.-M. Bousquet, E. Khoury, P. Sordo Martinez, J. M. K. Kua, C. You *et al.*, "I4U submission to NIST SRE 2012: A large-scale collaborative effort for noise-robust speaker verification," 2013.
- [15] S. O. Sadjadi, J. W. Pelecanos, and W. Zhu, "Nearest neighbor discriminant analysis for robust speaker recognition." *INTERSPEECH*, pp. 1860–1864, 2014.
- [16] S. Sarkar and K. S. Rao, "A novel boosting algorithm for improved i-vector based speaker verification in noisy environments," *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [17] D. A. van Leeuwen and R. Saeidi, "Knowing the non-target speakers: The effect of the i-vector population for plda training in speaker recognition," *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 6778–6782, 2013.
- [18] Y. Lei, L. Burget, L. Ferrer, M. Graciarena, and N. Scheffer, "Towards noise-robust speaker recognition using probabilistic linear discriminant analysis," *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 4253–4256, 2012.
- [19] T. Hasan, S. O. Sadjadi, G. Liu, N. Shokouhi, H. Boril, and J. H. Hansen, "Crss systems for 2012 nist speaker recognition evaluation," *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 6783–6787, 2013.
- [20] P. Rajan, T. Kinnunen, and V. Hautamäki, "Effect of multicondition training on i-vector plda configurations for speaker recognition." *Interspeech*, pp. 3694–3697, 2013.
- [21] D. B. Dean, A. Kanagasundaram, H. Ghaemmaghami, M. H. Rahman, and S. Sridharan, "The QUT-NOISE-SRE protocol for the evaluation of noisy speaker recognition," *Proceedings of the 16th Annual Conference of the International Speech Communication Association, Interspeech 2015*, pp. 3456–3460, 2015.
- [22] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1, pp. 19–41, 2000.
- [23] M. Ichino, N. Komatsu, W. Jian-Gang, and Y. W. Yun, "Speaker gender recognition using score level fusion by adaboost," *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on*, pp. 648–653, 2010.
- [24] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [25] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [26] S. Shiota, V. Fernando, J. Yamagishi, N. Ono, I. Echizen, and T. Matsui, "Voice liveness detection algorithms based on pop noise caused by human breath for automatic speaker verification," *Proc. Interspeech*, pp. 239–243, 2015.
- [27] 電子協騒音データベース, <http://research.nii.ac.jp/src/JEIDA-NOISE.html>.
- [28] H.-G. Hirsch, "F a NT-Filtering and Noise Adding Tool," 2005.
- [29] D. B. Dean, S. Sridharan, R. J. Vogt, and M. W. Mason, "The QUT-NOISE-TIMIT corpus for the evaluation of voice activity detection algorithms," *Proceedings of Interspeech 2010*, 2010.