T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	話者照合における雑音に非依存なマルチストリーム重みの検討
Title(English)	
著者(和文)	小島慎也, 岩野公司, 古井貞熙
Authors(English)	Shinya Kojima, Koji Iwano, SADAOKI FURUI
出典(和文)	日本音響学会 2007年秋季講演論文集, Vol. , No. 2-3-20, pp. 101-102
Citation(English)	, Vol. , No. 2-3-20, pp. 101-102
発行日 / Pub. date	2007, 9

Powered by T2R2 (Tokyo Institute Research Repository)

話者照合における雑音に非依存なマルチストリーム重みの検討*

◎小島慎也, 岩野公司, 古井貞熙 (東工大)

1 はじめに

我々は，スペクトル特徴量やケプストラム特徴量の各次元を個別のストリームとして扱ったマルチストリーム HMM による，雑音に頑健な話者照合手法の提案を行っている [1]．従来研究 [1] では，評価データそのものを重み推定用のデータとして利用する教師なしの条件で，各ストリーム（次元）の信頼度重みを，線形判別分析 (LDA) に基づく手法 [2, 3] で推定し，話者照合実験により得られた重みの有効性を確認している．本研究では，評価データと異なる雑音条件のデータを重み推定用データとして利用し，LDA に基づく手法で推定した「評価データの雑音条件に依存しない重み」が話者照合の耐雑音性の向上に有効であるかどうかを検討する．

また，従来までの研究では，特徴ベクトルの静的特徴量の次元のみを重み推定対象のストリームとしていたが，本研究では，動的特徴量を含む全ての次元を個別のストリームとして扱い，重みの推定対象とする．そこで，この両者の性能差についても比較を行う．

2 話者照合手法

2.1 マルチストリーム HMM

マルチストリーム HMM では，t フレーム目の入力特徴ベクトル $\boldsymbol{O}_t$ に対する出力確率の対数 $b(\boldsymbol{O}_t)$ は次のように計算される．

$$b(\boldsymbol{O}_t) = \sum_{s=1}^{S} \lambda_s \cdot b(\boldsymbol{O}_{st}) \tag{1}$$

ここで，$b(\boldsymbol{O}_{st})$ はストリーム s の特徴ベクトル $\boldsymbol{O}_{st}$ の出力確率の対数であり，S は総ストリーム数，λ_s はストリーム s の重みである．

2.2 話者照合スコア

話者照合スコア $q(\boldsymbol{O})$ は，申告話者の特定話者モデル M^c から得られるフレーム平均の尤度 $p(\boldsymbol{O}|M^c)$ を，不特定話者モデル M^g から得られるフレーム平均の尤度 $p(\boldsymbol{O}|M^g)$ で正規化することで得られる．

$$q(\boldsymbol{O}) = \log p(\boldsymbol{O}|M^c) - \log p(\boldsymbol{O}|M^g) \tag{2}$$

申告話者・不特定話者モデルにはマルチストリーム HMM が利用される．そこで，それぞれのモデルのストリーム s から得られる尤度 $p(\boldsymbol{O}_s|M^c)$, $p(\boldsymbol{O}_s|M^g)$ を用いて，この式を書き換えると，最終的に，

$$q(\boldsymbol{O}) = \sum_{s=1}^{S} \lambda_s \cdot q(\boldsymbol{O}_s) \tag{3}$$

となる．$q(\boldsymbol{O}_s)$ はストリームごとに計算される照合スコアとなる．照合スコアが，閾値 θ を超えた場合に，申告者本人であると判断する．したがって，判別式は $z = q(\boldsymbol{O}) - \theta$ という線形関数となる．

2.3 LDA によるストリーム重みの推定

各ストリームから得られる照合スコアで構成される多次元空間上に，申告話者の特徴量が正しく入力されたときと，詐称者の特徴量が入力されたときのデータをプロットすることで分布を作成し，この2つの分布を識別する関数を LDA を用いて求めると，得られる関数は，照合に用いる線形関数と同じ和の形となり，その係数をストリーム重みと見なすことが出来る．このようにすることで，重み推定用のデータの雑音条件に応じて，申告話者と詐称者の分布の識別性能が最大になるように，各ストリームの重みを推定することができる [1, 2]．

具体的には，各入力データについて，ストリーム s から得られる照合スコアを，x_s 座標 (S 次元空間上) にプロットする．申告話者と詐称者の2つの分布について LDA を適用して得られた判別関数は，

$$a_0 + \sum_{s=1}^{S} a_s x_s = 0 \tag{4}$$

となるが，係数 a_s に負の値が算出されることがあるため，その場合にはそのストリームの信頼度が著しく低いと見なし，そのストリームの重みを 0 とする．そこで，最終的なストリーム重み λ_s は次のように計算される．

$$\lambda_s = S \cdot \frac{a'_s}{\sum_{i=1}^{S} a'_i} \quad , \quad a'_i = \begin{cases} a_i & (a_i \geqq 0) \\ 0 & (\text{otherwise}) \end{cases} \tag{5}$$

3 話者照合実験

まず，従来までの静的特徴量の各次元に対して重み推定を行った場合と，動的特徴量を含む全次元について重み推定を行った場合について，雑音環境下における話者照合性能の比較を行う．次に，評価データとは異なる様々な雑音条件のデータを用いてストリーム重みを推定することで，雑音に依存しない重みの推定が可能であるかについて検討を行う．

3.1 実験条件

音声データには，1ヵ月毎に収録を行った，計 5 時期分のデータ [1, 2] を使用した．使用した話者は男性話者 36 名分で，各話者は 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・量子化されている．

データは 12 名ずつ 3 グループに分け，それぞれを評価用，不特定話者モデルの学習用，(教師あり条件での) ストリーム重み推定用に割り当てる．詐称者としては，評価用のグループの申告話者以外の全話者を用いる．申告話者・不特定話者のマルチストリーム HMM の学習には 1 ～ 3 時期目の音声データを用い，ストリーム重みの推定と評価には，4，5 時期目のデータを用いる．

特徴量としては，ケプストラム特徴である MFCC の 25 次元ベクトル（12 MFCC，12ΔMFCC，Δ 対数パワー），スペクトル特徴量 SPEC [4] の 27 次元ベクトル（13 SPEC，13ΔSPEC，Δ 対数パワー）を用いた．両者は次元数は異なるが，これはスペクトル特徴量では，全帯域のエネルギーの総和が一定値になるように正規化されているためであり，情報量としては等価である．

*Estimation of noise-independent multi-stream weights for speaker verification. by KOJIMA Shinya, IWANO Koji and FURUI Sadaoki (Tokyo Institute of Technology)

Table 1　雑音種ごとの平均誤り削減率

手法		ピンクノイズ	エレベータホール雑音	駅雑音
教師あり	MFCC-S	2.0 %	18.6 %	20.5 %
	MFCC-SD	6.0 %	21.5 %	23.8 %
	SPEC-S	17.8 %	17.0 %	22.6 %
	SPEC-SD	20.4 %	24.7 %	30.7 %
教師なし	MFCC-S	0.6 %	7.3 %	2.6 %
	MFCC-SD	2.4 %	8.5 %	4.4 %
	SPEC-S	3.0 %	6.2 %	11.7 %
	SPEC-SD	2.7 %	9.6 %	12.6 %

申告話者・不特定話者モデルは数字単位のマルチストリーム HMM であり，照合スコアを得るときには，それぞれのモデルを使用して桁数制限のある連続数字認識を行う．初期モデルは，1) まず，通常の HMM を学習し，2) 出来上がったモデルの特徴量ベクトルをストリームに分割し，3) 全てのストリーム重みを 1.0 とすることで作成する．推定により得られたストリーム重みは，全てのモデルに対して共通に使用する．

3.2　動的特徴量に対する重み付与の効果

3.2.1　実験方法

動的特徴量への推定重みの付与の効果を確認する実験を，教師あり・なしの両方の条件で行う．教師なしの実験では，ストリーム重み推定用の話者グループは使用せず，評価用のグループの全話者のデータについて，初期モデルによる照合実験を行って，申告者か詐称者かをラベル付けし，その判定結果をもとにして重み推定を行う．初期モデルを用いた照合の際の閾値は，等誤り率（EER）が最小となるように定めた．教師あり・なしのどちらの実験についても，6 通りのグループの組み合わせで照合実験を行い，全ての結果を平均して全体の評価に用いた．

学習用データには SN 比 30 dB でピンクノイズを，評価用データには，ピンクノイズ，電子協騒音データベース [5] 中の，エレベータホール，駅雑音を SN 比 5, 10, 15, 20 dB で付加した．教師あり条件では，重み推定用データに，評価データと同じ条件で雑音を重畳した．

3.2.2　実験結果

MFCC, SPEC それぞれについて，従来までの静的特徴量のストリームのみを重みの推定対象にした場合（動的特徴の Δ 項は重みの推定対象とせず，まとめて 1 つのストリームとして扱って重みを 1.0 に固定）(MFCC-S, SPEC-S) [1] と，静的特徴量のみでなく，動的特徴量の各次元も重み推定対象とした場合 (MFCC-SD, SPEC-SD) について比較を行う．各手法の初期モデルからの EER の平均誤り削減率を，雑音種ごとに各 SN 比から得られた結果を平均して Table 1 に示す．この表から，教師あり・なしのどちらの場合でも，動的特徴量のストリームに推定された重みを付与することで，誤り率が削減されていることが分かる．また，スペクトル特徴量である SPEC の方が，MFCC よりも改善が大きい傾向があることが分かる．

3.3　雑音環境に非依存なストリーム重みの推定

3.3.1　実験方法

評価データに重畳する，ピンクノイズ，エレベータホール雑音，駅雑音の他に，電子協騒音データベースの走行車内，列車（在来線），計算機室雑音も用意する．重み推定用データに評価データと異なる 5 種類の雑音を，SN 比 5, 10, 15, 20 dB で重畳し，これら全ての雑音条件での重み推定用データを用いて教師あり重み推定を行って評価データの雑音条件に依存しない重みの推定を行う．また，比較のために，評価データと同じ雑音，SN 比条件で重み推定用データに雑音を重畳して重み推定を行った場合（雑音条件が既知とした場合）についての照合実験も行った．

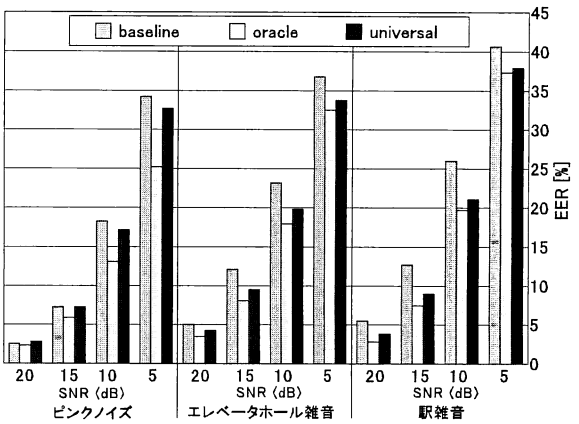

Fig. 1　雑音に非依存なストリーム重みの推定による SPEC での各雑音・SN 比における EER

本実験では，3.2 節の実験結果を受け，静的・動的特徴量の全ての次元に対して重み推定を行っている．

3.3.2　実験結果

特徴量として SPEC を用いたときの，各雑音・SN 比条件における，初期モデル (baseline)，雑音環境を既知として重み推定を行った場合 (oracle)，雑音環境に非依存な重み推定をした場合 (universal) の EER を Fig. 1 に示す．全ての雑音実験において，雑音環境を既知とした場合には及ばないが，初期モデルからの誤り率の削減が得られ，評価データの雑音環境に非依存な重みが推定されていることが分かる．なお，MFCC においても同様の傾向が得られている．

4　まとめ

マルチストリーム HMM を用いた話者照合において，評価データの雑音環境に依存しないストリーム重みを推定する方法について検討を行い，様々な雑音条件で，推定された重みによる耐雑音性の向上を確認した．また，静的特徴量のストリームのみでなく，動的特徴量のストリームについても重み推定対象として，推定重みを付与する方が，高い耐雑音性が得られることを確認した．

今後の課題として，より多くの音声データを用いた場合の本手法の有効性の確認や，他のストリーム重みの推定法の導入などが挙げられる．

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．

参考文献

[1] 小島他，音講論, pp.113-114 (2007-3).
[2] 浅見他，音講論, pp.101-102 (2004-9).
[3] 小島他，音講論, pp.65-66 (2005-9).
[4] 西村他，音講論, pp.117-118 (2004-3).
[5] S. Itahashi, *ICSLP*, vol.2, pp.1081-1084 (1990-11).