論文 / 著書情報
Article / Book Information

論題(和文)	話者照合におけるマルチストリーム重みの教師なし推定法の検討
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年春季講演論文集, Vol. , No. 3-10-5, pp. 113-114
Citation(English)	, Vol. , No. 3-10-5, pp. 113-114
発行日 / Pub. date	2007, 3

話者照合におけるマルチストリーム重みの教師なし推定法の検討*

◎小島慎也, 岩野公司, 古井貞熙 (東工大)

1 はじめに

マルチストリーム HMM を用いた話者照合手法では，雑音重畳によって信頼性が低くなった特徴量を有するストリームの重みを相対的に小さくすることで，話者照合の耐雑音性を向上させることが可能である．我々の先行研究 [1] では，ケプストラム情報と基本周波数情報の 2 つのストリームを利用し，雑音に頑健な基本周波数情報のストリーム重みを大きくすることで話者照合の耐雑音性が向上することを確認している．

一方，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たマルチストリーム HMM を用い，雑音環境に応じて各次元の信頼度重みの推定を行うことで，音声認識の耐雑音性が向上することも確認している [2]．[2] では，線形判別分析 (LDA) によるストリーム重み推定手法を提案しており，この手法を用いて教師ありで重み推定を行っている．雑音の帯域性を効率的に抑制できることから，スペクトル特徴量の方が，ケプストラム特徴量を用いるよりも耐雑音性が高いことを確認している．

本研究では，この [2] と同様の手法で，スペクトル特徴量である SPEC [3] と，ケプストラム特徴量である MFCC を用いてマルチストリーム HMM を構築し，それぞれを用いた話者照合手法の耐雑音性について検討を行う．また，本研究では，教師なしの条件で LDA を利用したストリーム重みの推定を行う．

2 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 O_t に対する出力確率の対数 $b(O_t)$ は次のように計算される．

$$b(O_t) = \sum_{s=1}^{S} \lambda_s \cdot b(O_{st}) \tag{1}$$

ここで，$b(O_{st})$ はストリーム s の特徴ベクトル 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．

本研究では，スペクトル特徴量 SPEC [3]，ケプストラム特徴量 MFCC の各次元を個別のストリームとして扱い，各次元の重みを推定する．実際用いる特徴ベクトルには Δ 項と Δ 対数パワー項が含まれているが，Δ 項はまとめて 1 つのストリームとして扱い，重みは推定対象とせず，1.0 に固定した．

3 話者照合スコア

入力特徴量 O に対する話者照合スコア $q(O)$ は，申告話者の特定話者モデル M^c から得られる尤度 $p(O|M^c)$ を，不特定話者モデル M^g から得られる尤度 $p(O|M^g)$ で正規化することで得られる．

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

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 $p(O_s|M^c)$, $p(O_s|M^g)$ を用いて，この式を書き換えると，最終的に，

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

となる．$q(O_s)$ 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 $z = q(O) - \theta$ という線形関数となる．

4 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り，その係数をストリーム重みに見立てることが出来る．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，次元ごとのストリーム重みの信頼度を推定することができる [4]．

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のフレーム平均を求め，x_s 座標 (S' 次元空間上) にプロットする．ここで，Δ 項の重みは推定しないため S' は元の総ストリーム数 S より 1 つ小さい値となる．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．

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

このストリーム重みを，全てのモデルに対して共通に使用する．

教師なしでの推定を行う場合には，重み推定用のデータに対して，初期モデルによって申告者か詐称者かをラベル付けし，推定に利用する．

5 話者照合実験

5.1 音声データ

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, 4] 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．

*Unsupervised estimation of multi-stream weights for speaker verification, by KOJIMA Shinya, IWANO Koji and FURUI Sadaoki (Tokyo Institute of Technology)

Table 1　各雑音・SN 比条件における EER (%)

		ピンクノイズ				エレベータホール雑音				走行車内雑音				列車（在来線）雑音			
		20dB	15dB	10dB	5dB	20dB	15dB	10dB	5dB	20dB	15dB	10dB	5dB	20dB	15dB	10dB	5dB
MFCC	BASE	2.2	7.4	19.0	32.9	4.0	10.1	22.0	35.1	4.9	5.1	5.7	7.7	9.0	16.6	27.8	39.0
	LDA	2.2	7.0	19.2	34.6	3.2	9.1	22.0	35.5	3.5	3.6	3.8	5.3	6.2	14.9	27.9	40.0
SPEC	BASE	2.2	6.5	17.4	34.2	4.3	10.9	22.3	36.5	5.3	5.6	5.8	7.6	8.6	15.7	27.0	39.3
	LDA	1.9	6.1	17.9	35.2	3.6	10.0	22.4	36.8	3.4	3.4	3.6	4.7	5.5	12.9	25.4	40.0

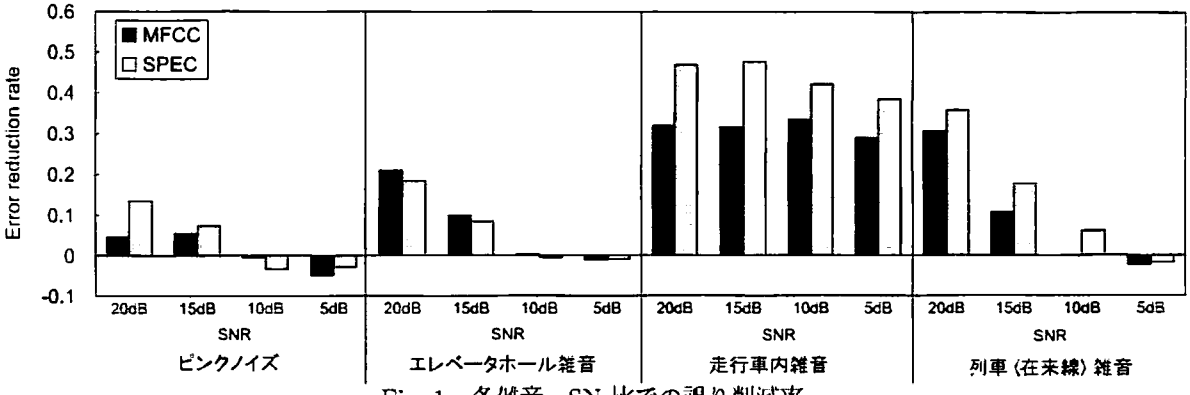

Fig. 1　各雑音・SN 比での誤り削減率

データは 12 名ずつ 3 グループに分け，不特定話者モデルの学習に用いる話者グループと，評価用の話者グループを 6 通りの組み合わせで選択し，それぞれについての照合実験の結果を平均して，全体の評価に用いる．詐称者としては，評価用のグループ中の，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．

学習データには SN 比 30dB のピンクノイズを付加させ，評価データには，ピンクノイズ，電子協騒音データベース [5] のエレベータホール，走行車内，列車（在来線）雑音をそれぞれ SN 比 5, 10, 15, 20 dB で付加させたものを用いた．特徴量として，MFCC は 25 次元ベクトル（12 MFCC，12 ΔMFCC，Δ 対数パワー）を，SPEC は 27 次元ベクトル（13 SPEC，13 ΔSPEC，Δ 対数パワー）を用いた．両者は次元数は異なるが，情報量としては等価である．

5.2　実験方法

まず，申告話者モデルと不特定話者モデルの学習を行う．その際には，通常の HMM の学習と同様に，特徴量の全次元を 1 つのベクトル（ストリーム）として扱う．モデルは音節単位である．出来上がったモデルについて，Δ 項以外の各次元を個別のストリーム，Δ 項を 1 つのストリームとして分割し，全てのストリーム重みを 1.0 として初期モデルを作成する．評価用の話者グループの全ての話者のデータについて，初期モデルを用いた照合実験を行い，その判定結果をもとに LDA により，ストリーム重みの推定を行う．本実験では，この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．最後に，推定された重みを用いて，再度照合を行う．

5.3　実験結果

各雑音・SN 比条件において，MFCC, SPEC それぞれについて初期モデル (BASE) の EER と 教師なし重み推定後のモデル (LDA) の EER を Table 1 に示す．また，重み推定による初期モデルからの EER の削減率を Fig. 1 に示す．Fig. 1 を見ると MFCC, SPEC どちらの特徴量を用いても，SN 比が 15, 20 dB では全ての雑音において初期モデルと比べて誤り率が削減されていることが分かる．特に，走行車内雑音で大きな改善が得られた．ストリーム重み推定の効果は，初期モデルによる照合誤りが大きくなると得られにくくなり，Table 1 と併せて見ると等誤り率がおおよそ 20% 以上になると改善効果が得られなくなる．また，MFCC と SPEC を比較すると，多くの雑音条件で SPEC の方が重みを付与することによる改善が大きく，音声認識と同様の傾向 [2] が確認された．

6　まとめ

マルチストリーム HMM を用いた話者照合において，ストリーム重みを教師なし推定する方法を提案し，4 桁連続数字音声を用いた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，様々な雑音条件の SN 比 15, 20 dB において，初期モデルからの誤りが削減され，耐雑音性の向上が確認された．

今後の課題として，Δ 項をストリーム重み推定の対象に含めた場合の話者照合性能の確認や，MLLR などの雑音適応手法との性能比較，MLLR との融合の検討などが上げられる．

参考文献

[1] 浅見太一，岩野公司，古井貞熙，“ハフ変換による基本周波数情報を用いた雑音に頑健な話者照合，” 音講論，vol.1, pp.177-178 (2004-3).

[2] 小島要，岩野公司，古井貞熙，“マルチバンド音声認識における尤度重み推定法の検討，” 音講論，pp.65-66 (2005-9).

[3] 西村義隆，篠崎隆宏，岩野公司，古井貞熙，“周波数帯域ごとの重みつき尤度を用いた音声認識の検討，” 音講論，vol.1, pp.117-118 (2004-3).

[4] 浅見太一，岩野公司，古井貞熙，“マルチストリーム話者照合におけるブースティングを用いた閾値の最適化，” 音講論，pp.127-128 (2005-9).

[5] http://www.sunrisemusic.co.jp/dataBase/fl/noisedata01_fl.html