7ZA-05

走行中の自動車内環境での音声による個人認証 *

田本 篤喜 †, 伊藤 克亘 †,

1 まえがき

現在, 主に個人の認証では, パスワードやカードなどの, ユーザーの所有物に基づく認証が行われてきた. しかし人間の生体特徴を個人の認証に用いることで, 盗難・紛失・忘却などの危険は減らすことができる. そのうち, 声による人物特定技術 (話者認識) は, 既に実用化されている例もある.

上記の利点に着目し本研究では, 運転中におけるスマートフォンのロック解除を想定し, 走行中の自動車内環境での音声による個人認証精度の向上に取り組む.

2 話者照合

iPhone の Siri の, 声による持ち主登録機能はこのテキスト依存型の話者照合であり, 「Hey,siri」という発話を用いて照合を行う [1]. 主な手法として, DTW[2] を用いて時間的な情報を踏まえて類似度を計算することで高い精度が得られる. DTW は少ないデータからテンプレートを作成するという利点がある. このデータ量の少なさは精度低下を招くことがあるが, それを補う手法として, GMM Posteriorgram がある [3].

話者照合はスマートフォンのようなウェアラブル端末に実装される場合, 多くの場面での使用が想定される. 本研究ではその一つとして, 運転中の話者照合に取り組み, 雑音, 発声時期の違いによる話者内特徴変動, の二つの問題に対して頑健な話者照合を目指す.

雑音 - 走行中の自動車内は, 静かな室内と違い, 走行音や風切り音, エンジン音などの雑音がある.

発声時期の違いによる話者内特徴変動 - 話者照合システムは一度の登録から長期的な使用が見込まれる. 一方で, 人間の発話は, 同じ人が同じ言葉を繰り返し発声しても, ある程度時期をおくと, 発声が変わるため, 認識系の性能に影響を及ぼす.

3 処理の流れ

DTW を用いたテキスト依存型の話者照合は, 3 発話程度の正解テンプレートを登録しておき, 入力されたときには, テンプレートと入力音声の特徴を比較し, 予め決めたしきい値に基づいて照合を行う. 本研究では, 登録音声発話全ての特徴量で学習した GMM を用いてテンプレートを構成することで, 正解話者の話者内変動を考慮したテンプレート表現ができ, また GMM で多くの発話の情報を考慮できるという利点に着目し, GMM Posteriorgram をテンプレート表現として導入する. 一般的な話者照合の流れに加えて, 音声が入力されたときの車内雑音抑圧処理を加える.

特徴量としては, 人間が真似することが困難な, 先天的個人性が効果的であるとされている. 実際に話者照合の研究で MFCC が先天的個人性情報として用いられている [4]. 本研究では, フレーム長 25ms, フレームシフト 10ms で抽出を行い, 13 次の MFCC と ΔMFCC, ΔΔMFCC を特徴量として使用する.

3.1 GMM Posteriorgram

発話テンプレートを表現する特徴ベクトルとして, GMM Posteriorgram を導入する [3]. GMM posteriorgram とは, 異なる GMM コンポーネントによって生成される, 出力確率を要素に持つ特徴ベクトルの確率表現である. GMM posteriorgram の生成には, 以下の処理を施す. p 個のフレームを含む, 長さ $\mathbf{S}$ の音声データを考える. $\mathbf{S}$ は,

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_p] \tag{1}$$

と表せる. i 番目のフレームに対応する GMM posterior vector は,

$$\mathbf{x}_i = [P(\lambda_1|\mathbf{s}_i), P(\lambda_2|\mathbf{s}), ..., P(\lambda_N|\mathbf{s}_i)]^T \tag{2}$$

で与えられる. ここで, λ_j は j 番目のガウシアンコンポーネントを表し, N はガウシアンコンポーネントの総数を表す. GMM posteriorgram は, 1 発話中の各フレームの posterior vector を連結する,

$$X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p] \tag{3}$$

3.2 照合

照合では, 登録音声 3 発話の GMM Posteriorgram を用いて決定したしきい値と入力音声・登録音声間の距離を比較することで照合を行う. 事前に決定するしきい値を式 (4) で示す [5].

$$T = \frac{\mu_1\sigma_2\omega_2 + \mu_2\sigma_1\omega_1}{\sigma_1\omega_1 + \sigma_2\omega_2} \tag{4}$$

μ_1,σ_1 は正解話者音声内の距離の平均と標準偏差, μ_2,σ_2 は正解話者・詐称者音声間の距離の平均と標準偏差, ω_1,ω_2 は足して 1 になるような重みであり, しきい値を調整する際に変更する値である. 入力音声と 3 つのテンプレート音声の平均距離が式 (4) のしきい値よりも小さいときに正解話者, 大きいときに詐称話者と決定する.

3.3 雑音対策

車内雑音が低周波成分に集中するという特徴から, 100Hz 以下の低周波成分を除去する [6]. 100Hz という数字は, スペクトル領域での個人の特徴をつぶさない程度で決定した値である.

定常雑音抑圧の手法として SS 法を施す [6]. SS 法とは, 雑音が重畳した音声のスペクトルから雑音のスペクトルの推定値を周波数領域で減算する, スペクトル減算に基づく雑音抑圧法である.

4 評価

一般に, 話者照合の研究の評価には等誤り率 (EER) が用いられる. 等誤り率とは, ふたつの誤り率が等しく

*: Voice authentication by text dependent single utterance for in-car environment *Atsuki Tamoto (Hosei Univ.) et al.*

†法政大学 情報科学部

なるしきい値での誤り率である．本研究で実装した話者照合における，トレードオフのグラフを図 1 に示す．

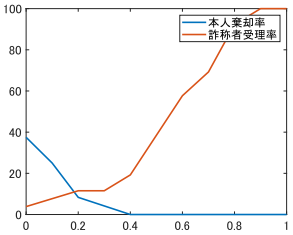

図 1. 誤り率と EER

図 1 の二つの誤り率の交点が EER である．本研究では，雑音抑圧処理の評価，長期的特徴変動の影響の評価ともに，この EER となるしきい値 ($\omega_1 = 0.15$) で実験を行う．

雑音対策に関する評価としては，登録・照合ともに同じ時期に発話したデータを用いて精度評価を行う．また，長期的特徴変動に関する実験としては，テンプレート更新の有無での二つの実験を行う．

使用するデータに関して，先行研究 [7] では，擬似雑音環境下における話者照合システムの評価に用いる音声として，目的の音声信号と雑音レベルの比 (SNR) が指定した値になるような雑音を付与し，いくつかの雑音レベルで実験を行っている．本研究においても，雑音を付与することで，多くのデータを集める．

時期の違いによる特徴変動を考慮したデータとして，1 週間に 1 日，朝昼夜の録音を 1 年間行った音声コーパスである，AWA 長期間収録音声コーパス (AWA-LTR) を用いて評価を行う．

雑音抑圧処理の結果として，SNR ごとの本人棄却率 (FRR) と詐称者受理率 (FAR) を表 1 に示す．

表 1. 雑音抑圧処理の結果 (%)

SNR		抑圧あり	抑圧なし
15dB	FRR	45.8	100
	FAR	3.8	0
10dB	FRR	50.0	100
	FAR	0	0
ノイズなし	FRR	−	12.5
	FAR	−	11.5

15dB,10dB ともに，抑圧ありと抑圧なしの FRR を比較すると，どちらも抑圧ありの方が低い FAR を保ちつつ FRR も改善される結果となり，雑音抑圧処理の有効性が示される結果となった．一方で，表 1 より，雑音抑圧処理によってノイズなしの精度より FRR が 30 %以上低下しているのがわかる．これはスペクトルサブトラクションによる雑音成分が影響したと考えられる．ここで，雑音抑圧処理を施した，受理された音声をテンプレートに加え，15dB でもう一度実験を行った結果，FRR0 %，FAR3.8 %が得られ，テンプレート追加処理による精度の向上が見られた．

テンプレートの更新処理をせずに，登録直近の 2ヶ月で発話された音声で実験した結果 FRR は表 2 より，8.3 %となった，登録から 10ヶ月の期間を空けて発話された音声で実験した結果 FRR は 4.2 %という結果が得られ，本研究で実装した手法では時期差による精度への影響がなかった．MFCC を直接テンプレートとして用いた本人棄却率と，特徴量 MFCC の GMM Posteriorgram(GP) を用いた本人棄却率を表 2 に示す．なお，どちらも EER となるしきい値 (MFCC,GP, それぞれ $\omega_1 = 0.15, 0.85$) である．

表 2. 特徴変動に関する結果の特徴量ごとの比較 (%)

FRR	DTW-MFCC	DTW-GP
直近 2ヶ月	54.2	8.3
10ヶ月後	79.2	4.2

表 2 より，テンプレート表現として話者内変動を考慮できる GMM Posteriorgram を導入したことにより，発声時期の違いによる精度への影響はなかったと言える．

5 あとがき

本研究では，走行中の自動車内での話者照合として，走行車内での問題 (雑音)，話者照合の問題 (特徴変動) の二つを問題視し，評価を行った．雑音抑圧処理の結果，実際に雑音抑圧処理を施すことで SNR10dB,15dB 平均で FRR47.9 %の向上が確認できた．また，雑音抑圧処理によって生じる精度への影響は，雑音抑圧処理を施した，受理された音声をテンプレートに加えることでさらなる向上が見られた．長期的特徴変動に関する実験の結果は，登録から 10ヶ月空いた発話では精度への影響が出ない結果となった．原因として，実際に MFCC のみによるテンプレートと比較し，GMM Posteriorgram が長期的変動にも有効であることが確認できた．本研究では，雑音と発声時期の違いによる特徴変動のみを対象とし，評価にも二つの問題のみを考慮した評価データを使用し実験をおこなった．今後はさらに，ロンバード効果も車内で行う話者照合の問題点として扱う必要がある．

参考文献

[1] Siri Team, `https://machinelearning.apple.com/2018/04/16/personalized-hey-siri.html`, “ Personalized Hey Siri ”

[2] M. Pandit et al.,“ Feature Selection for a DTW-based Speaker Verification system ”, IEEE International Conference on Acoustics, Speech and Signal Processing, May 1998.

[3] S. Jelil et al.,“Speaker Verification Using Gaussian Posteriorgrams on Fixed Phrase Short Utterances ”INTERSPEECH-2015, pp. 1042-1046, Oct.2015.

[4] 王 他, “ 話者認識におけるロバストネス ”, 音響誌, vol. 69, no. 7, pp. 357-364, 2013.

[5] S. Paul et al., “ Comparative Analysis of Two Different System’s Framework for Text Dependent Speaker Verification ”ICCPCT-2015, march.2015.

[6] 武田一哉,“自動車の中での音声認識”, IPSJ Magazine, vol.45, no.10, Oct.2004.

[7] 鎌田 他, “ 雑音環境下における話者照合 ”, 信学技報 TECHNICAL REPORT IEICE, pp.55-60, March. 2007.