

継続利用する話者照合システムにおける性能の経時変化に関する検討

榎本 耕二[†] 岩野 公司[†]

東京都市大学[†]

1. はじめに

近年、情報システムのセキュリティ向上のため、音声による認証（話者照合）の応用が期待されている。しかし、音声は時期で特徴が変動してしまい、それによって照合性能が劣化してしまうという問題がある。

このような性能劣化の分析を行った研究 [1] では、約 10 日間に渡って 2~3 日おきに収録されたデータを登録に利用することで、そこから 3 か月以上経過したデータを照合したときでも性能劣化が見られないことを確認している。この研究では、発声内容は 2 単語のみとなっており、多次元正規分布で構築された申告話者本人のモデルに対する未知音声の尤度を照合スコアとしている。

本研究では、より実践的な状況を想定して、話者照合性能の経時変化の分析を行う。具体的には、「約 2 ヶ月間に渡り、テキスト指定型話者照合システム [2] をほぼ毎日継続的に利用する」データを利用する。各ユーザはランダムに提示される 4 桁連続数字の発声を行っており、照合手法には現在一般的に用いられている GMM-UBM 法 [3] を利用する。この手法では、未知音声に対し、混合正規分布 (GMM) で構築される申告話者モデルから得られる尤度を、不特定話者モデル (UBM) の尤度を用いて正規化することで照合スコアを算出する。

このような環境において、登録が最後に行われた日から時間が空いて照合が行われる場合についての性能調査を行う。同時に、ユーザの負担の軽減を考慮して一日あたりの登録音声の量をどの程度まで抑えることができるかについて検討を行う。

2. 実験条件

2.1 使用データ

科研費若手研究 (B) 「公開鍵基盤 (PKI) と

An analysis of speaker verification performance degradation assuming long-term and continued system use

[†] Koji Enomoto, Koji Iwano, Tokyo City University

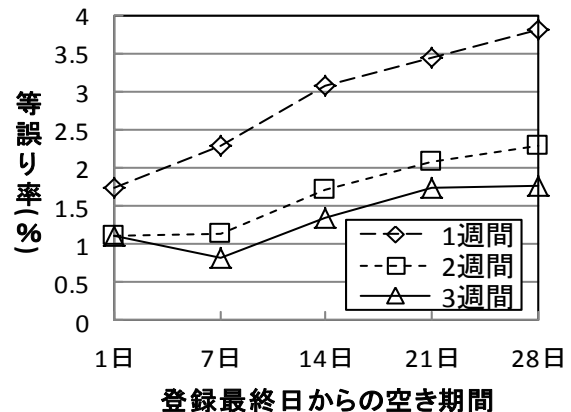


図1 登録データ蓄積期間を変化させたときの話者照合性能の比較

話者照合技術の融合に関する研究 (No. 17700141) [4] によって、約 2 ヶ月間に渡ってほぼ毎日収録された大学生・大学院生の男性 20 名の音声を使用する。各話者について一日あたり 10 個の 4 桁連続数字発声が収録されている。

2.2 音響特徴量とモデル学習

音響特徴量には、12 次元の MFCC とその 1 次微分成分、対数パワーの 1 次微分成分の計 25 次元のベクトルを使用する。

本実験では、週を単位として登録データの蓄積期間を定める。例えば、1 週間を蓄積期間とした場合には、その週の本人の発声を用いて申告話者モデルを GMM で学習し、全話者の発声を用いて不特定話者モデル (UBM) の学習を行う。なお、GMM の混合数は予備実験の結果により良好となった 256 に固定した。

3. 登録データ蓄積期間の違いによる性能変化の分析

登録データ蓄積期間が 1, 2, 3 週間のときの、登録最終日から照合日までの空き期間の長さに対する照合性能の遷移を図 1 に示す。空き期間は、1, 7, 14, 21, 28 日と変化させており、照合性能は等誤り率で示す。この実験では、10 発声全てを学習データとして利用している。

図より、2, 3 週間のデータを使用した場合に比べ、1 週間分のみを登録データに使用した場合に、

空き期間が長くなるほど大きな性能劣化が見られた。28日経過後については、1%以上誤り率が大きいことが確認され、従来研究[1]と同様の知見を得た。

4. 登録音声の量の違いによる性能変化の分析

3章の結果から、時期差による性能劣化を防ぐためには少なくとも2週間以上必要であることが示された。そこで、本章では、データ蓄積期間を2, 3週間としたときに、どの程度一日あたりの発声数を減らすことが可能かについて検討する。なお、1週目については10発声全てを登録に用いた。

4.1 データ蓄積期間2週間の実験

2週目の登録発声数を7, 5, 3個と減らした場合の照合性能の経時変化を図2に示す。図中の「10→5」という表記は、1週目は一日あたり10発声を、2週目は一日あたり5発声を登録に使用したことを意味している。

これを見ると、2週とも10発声を用いた場合に比べ、2週目の登録データ数を7発声にしたものはほとんど差が見られない。一方、2週目のデータ半分(5発声)以下にすると明らかな性能劣化が確認されることがわかった。

4.2 データ蓄積期間3週間の実験

4.1節の結果を受け、「1週目は10発声、2週目は7発声」を登録した上で、3週目の登録音声の数を7, 5, 3回と減らしたときの照合性能の経時変化を調べる。結果を図3に示す。参考のため、3週間に渡って10発声を使い続けた場合(10→10→10)と、3週目に登録を行わなかった場合(10→7)の結果も併せて掲載する。

この結果をみると、3, 5回に減らしたものは3週目に登録を行わなかった場合(10→7)とほぼ同性能となり、少量のデータの登録では意義が少ないことがわかる。一方、7発声を登録に使用した場合(10→7→7)は、最大限登録を行った場合(10→10→10)に比べ性能劣化がほとんど見られず、有効であることがわかる。以上より、今回の実験環境においてユーザの負担の軽減を目的として発声回数を段階的に減らす場合には、2週目以降の発声回数を半分以下に減らすことは危険で、7割程度まで減らすことが可能であることがわかった。

5. まとめ

本研究では、GMM-UBM法に基づくテキスト指定型話者照合システムの継続利用を想定した場合の、照合性能の経時変化について分析を行った。実験の結果、登録データの蓄積期間を2週

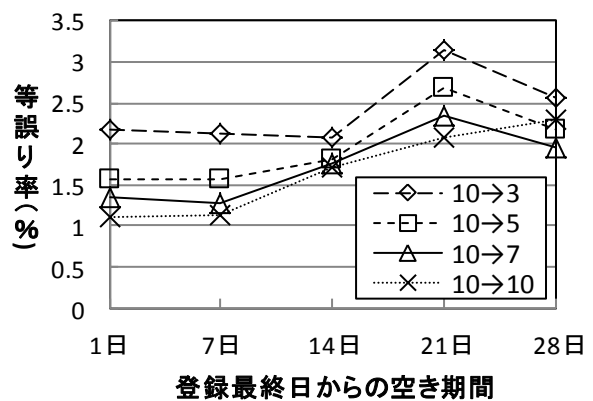


図2 登録データ蓄積期間を2週間とした際の発声数の違いによる話者照合性能の比較

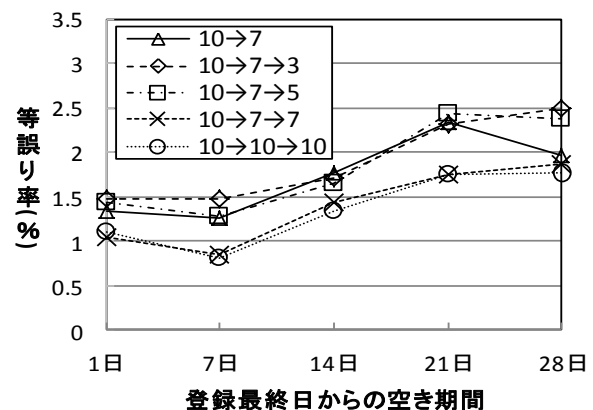


図3 登録データ蓄積期間を3週間とした際の発声数の違いによる話者照合性能の比較

間以上確保しなければ、経時による性能劣化が大きいことが確認され、文献[1]の知見と同様となった。また、ユーザの負担を減らしつつ、十分な話者照合性能を得るためには、2週目以降の登録データの発声回数を10から7まで減らす制御が有効であることがわかった。

今後は、2週目以降の話者モデルの更新に各種の適応化手法を用いた場合の分析や、別の話者照合手法での分析などを行う必要がある。

参考文献

- [1] 古井貞熙, “音声の個人性パラメータの時期的変動と話者認識,” 電子通信学会論文誌, vol.57-A, no.12, pp.880-887, 1974.
- [2] 松井知子, 古井貞熙, “テキスト指定型話者認識,” 電子情報通信学会論文誌, vol. J79-D-II, no. 5, pp. 647-656, 1996.
- [3] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, “Speaker Verification Using Adapted Gaussian Mixture Models,” Digital Signal Processing, vol. 10, pp. 19-41, 2000.
- [4] <http://kaken.nii.ac.jp/ja/p/17700141>