声真似攻撃に対する話者照合システムの脆弱性の分析

Analysis of Vulnerability of a Speaker Verification System against Voice Mimicry Attacks

曽根 泰斗[*1]　岩野 公司[*1*2]
Taito Sone　Koji Iwano

[*1] 東京都市大学 環境情報学部　[*2] 東京都市大学 メディア情報学部
[*1] Faculty of Environmental and Information Studies, Tokyo City University　[*2] Faculty of Informatics, Tokyo City University

1. はじめに

近年，音声による個人認証（話者照合）への期待が高まっており，実用製品の開発なども進められるようになった．セキュリティーへの応用を考えると，話者照合システムの様々な攻撃に対する脆弱性を正しく把握し，その対策を考えることは極めて重要である[1].

本研究では，最も手軽な攻撃手段である「声真似（模倣）」に焦点をあて，「一般人（物真似の素人）が，自身の声質の近さとは無関係に設定された対象者に成りすますための模倣を行う」状況を想定し，そのときの物真似音声が話者照合性能に与える影響について調査・分析を行う．

2. 模倣音声データの収録

実験には，本学学生の男性 6 名のグループで収録した模倣音声を利用した．収録は約 2 日ごと，3 週間にわたり継続的に実施した（合計 9 日分）．各話者は 1 日の収録で，

① 本人の声として自然に行う発声
② グループ内の他者（5 名）を模倣しようと努力して行った発声
③ 本人として受理されようと努力して行った発声

を行う．②，③については，対象となる人物の初日の発声を聴取した上で発声を行う．しかし，素人が聴取のみで模倣を行うことには困難が予想されるため，模倣の練習支援を行いながら音声収録を行うことができるシステムを用意し，それによる収録も併せて行う．この収録システムには，前日までに収録されたデータを利用して構築された話者照合システムから出力される照合スコア（申告者らしさ）を，発声ごとに被験者に提示する機能がついており，被験者はこのスコアができるかぎり大きくなるように模倣を練習することができる．なお，各話者はそれぞれの収録セッションについて 10 個の 4 桁連続数字を発声している．

3. 話者照合性能への影響の調査

調査には，HMM でモデル化された申告者モデルと不特定話者モデル（UBM）を利用する話者照合を利用する．この手法では，各話者を 3 状態の HMM でモデル化するが，GMM-UBM 法[2]と同じ照合の原理を利用している．

学習には 3 週間のうちの前半 2 週間（6 日分）で収録された，①のセッションで発声された音声を利用する．評価には，後半 1 週間（3 日分）で収録されたデータを利用する．模倣発声を含まない評価（「模倣なし」）を行う場合には，③で発声された音声のみを評価に利用する．このとき，申告者以外の全ての話者の音声を詐称者データとして用いる．模倣発声を含む評価（「模倣あり」）行う場合には，③で発声された音声を申告者受理の評価に，②で発声された模倣音声を詐称者棄却の評価に使用する．

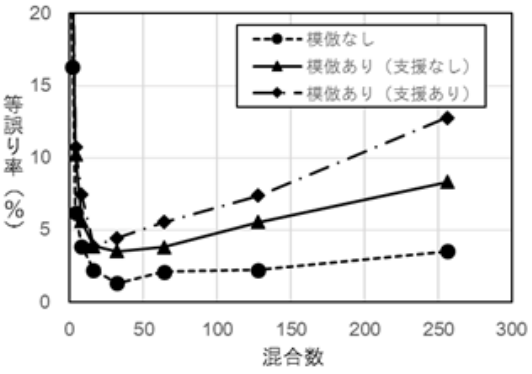

図 1　模倣が話者照合性能に与える影響の評価

図 1 に，HMM の混合数を変化させたときの，「模倣あり」「模倣なし」の等誤り率の変化を示す．混合数は申告者モデルと UBM で共通とした．結果を見ると，模倣支援の有無に関わらず，模倣行為による等誤り率の上昇が見られ，声質の近さを考慮せずに構成された話者グループの内の素人の模倣であっても，成りすましが成功している状況が確認される．また，模倣支援を用いた場合の方が，性能劣化が大きいことも確認された．

4. まとめ

本研究では，「声質の近さを考慮せずに構成された話者グループの内の素人の模倣」が話者照合システムに与える影響の評価を行った．その結果，このような攻撃であっても照合性能に悪影響を及ぼすことが確認された．また，照合スコアに基づく模倣支援を行いながら収録した模倣発声の方が，聴取のみで模倣を行った場合よりも照合性能の劣化に大きな影響を及ぼすこともわかった．今後の課題として，話者や試行ごとの詳細な分析や，i-vector に基づく話者照合システムに対する影響評価などがあげられる．

謝辞　本研究は JSPS 科研費 基盤研究（C）25330206 の助成を受けたものです．

参考文献

[1] N. Evans, et al., “Spoofing and countermeasures for automatic speaker verification,” Proc. INTERSPEECH, pp. 925-929, 2013.

[2] D. A. Reynolds, et al., “Speaker verification using adapted Gaussian Mixture Models,” Digital Signal Processing, vol. 10, pp. 19-41, 2000.