x-vector に基づく話者照合システムに対する一般人と物真似タレントによる声真似攻撃の分析

五味　真奈美†　　岩野　公司†
† 東京都市大学 メディア情報学部 情報システム学科

1. はじめに

音声による個人認証（話者照合）の実用化・高性能化に向けて, 我々は GMM-UBM 法に基づく話者照合システムを用いて, 一般人と物真似タレントの「声真似」による詐称の攻撃力と特徴の分析を行っている[1]．しかし，近年登場した深層学習に基づく話者照合システムを用いた分析は行われていない．本研究では，x-vector[2]に基づく話者照合システムを用いて, 一般人と物真似タレントの声真似攻撃の分析を行う．

2. x-vector に基づく話者照合システム

x-vector は, 多数話者の音声から学習された話者識別用の深層ニューラルネットワークの中間層から抽出される，入力音声の話者性を効率的に表現したベクトルである．話者照合システムに登録された申告話者の x-vector と入力音声の x-vector の類似度を PLDA によって算出し[3]，その値がしきい値以上であれば，申告話者本人として受理する．

3. 話者照合性能による声真似攻撃の分析

3.1 使用する音声データ

x-vector 抽出用の深層ニューラルネットワークは, 話者認識用データベース VoxCeleb1, 2[4]に含まれる 7,323 名の話者による 100 万以上の発声で学習した．分析に使用するデータには，先行研究[1]と同じ，6 名の男子学生（一般人）と物真似タレント 1 名（40 代男性，キャリア約 20 年）の 4 桁数字発声のデータを利用した．システム登録用の音声は 6 名の学生の地声の発声（各話者あたり 50 発声）とし, 詐称攻撃時の入力音声には，学生・タレントが声真似を行ったときの発声（各攻撃対象者に対して各話者あたり 10 発声ずつ）と，地声による発声（各話者あたり 10 発声ずつ）を利用した．

3.2 分析結果

図 1 に，照合システムのしきい値の変化に対する詐称者受理率の変化を, 一般人の物真似タレントの地声・声真似に対して示す．これをみると，声真似による詐称者受理率の上昇は，物真似タレントの方が圧倒的に大きいことがわかる．この傾向は先行研究[1]の分析でも同様にみられている．また，深層学習に基づく話者照合手法であっても，一般人・タレントのどちらの声真似攻撃でも，誤り率の上昇は確認されることから，声真似による詐称への対策が必要であることがわかる．

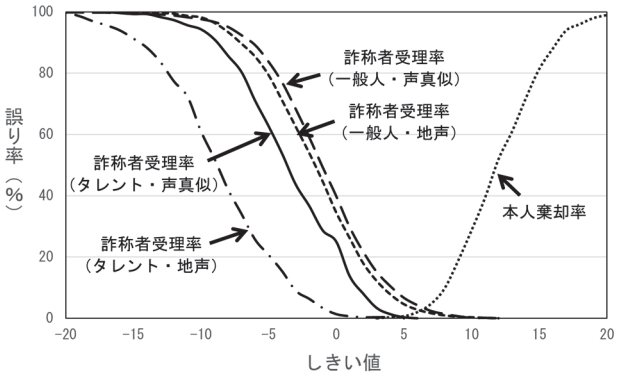

図 1　深層学習に基づく話者照合に対する地声と声真似攻撃の詐称者受理率の比較

4. まとめ

x-vector を用いた話者照合システムにおいて, 一般人の物真似タレントの声真似攻撃の分析を行い，タレントの詐称攻撃力の高さを確認した．今後は，声真似攻撃への対応手法の提案を目指す必要がある．

謝辞　本研究は JSPS 科研費 基盤研究(C) 19K12051 の助成を受けたものです．

参考文献

[1] 高木，岩野，“詐称者の物真似スキルを考慮した話者照合における声真似攻撃の分析，” 情報処理学会全国大会講演論文集，pp.455-456，2020.
[2] D. Snyder, et al., “X-vectors: Robust DNN embeddings for speaker recognition，” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)，pp. 5329-5333, 2018.
[3] P. Kenny, “Bayesian speaker verification with heavy tailed priors,” Proc. The Speaker and Language Recognition Workshop (Odyssey 2010), 2010.
[4] https://www.robots.ox.ac.uk/~vgg/data/voxceleb/