ISS-SP-003

話者照合システムに対する人間と声質変換の詐称攻撃力の比較

蒔田　博文†　　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

声による本人認証（話者照合）に対する攻撃として，登録話者の声を模倣した詐称音声の入力が考えられる．我々は，これまでに GMM-UBM 法に基づく話者照合システムを用いて，人間の声真似による攻撃について分析を行ってきた[1]．一方で近年，「声質変換」の研究が進み，詐称攻撃手段としての危険性が益々高まっている[2]．そこで本研究では，入手が容易な声質変換器を用いて，話者照合に対する攻撃力を調査し，人間の声真似による詐称との特徴の比較を行う．

2. 詐称音声の作成

声質変換器には，フリーで入手可能な Deep_VoiceChanger[3]と sprocket[4] を利用する．

音声データには，先行研究[1]で使用された 6 名の男子学生による 4 桁数字発声を使用する．各話者あたり 50 発声を登録音声とし，GMM-UBM 法に基づく話者照合システムの構築を行う．声質変換用モデルの学習には，各話者あたり 70 発声を使用する．ただし，変換モデルの学習には大量のデータが必要になることから，SoX[5]を利用した話速変換と，逆再生音声の作成を行い 22 倍に拡張した．照合を行う際には，各話者あたり（上記の 70 発声には含まれない）10 発声を変換元音声として，声質変換器を用いて，他の 5 人を模倣した詐称音声をそれぞれ作成し，入力する．

3. 声質変換による詐称攻撃の分析

3.1 話者照合性能による詐称攻撃力の分析

図 1 に，2 種類の声質変換器によって作成された詐称音声を照合システムに入力したときの，しきい値の変化に対する誤り率（詐称者受理率）の変化を示す．図中には，比較のため，先行研究[1]における人間（学生・物真似タレント）の声真似を詐称音声とした場合の詐称者受理率の曲線も併せて示す．声質変換器の詐称者受理曲線が人間の声真似による曲線を大幅に上回っていること，物真似タレントの声真似が，学生の声真似と Deep_VoiceChanger の中間の攻撃力を持っていることなどがわかる．

3.2 詐称音声の特徴の分析

図 2 に，声質変換器で作成された各詐称音声が，GMM-UBM 法における申告話者と対立話者（不特定話者）にどの程度近づいているかの分析結果を示す．この

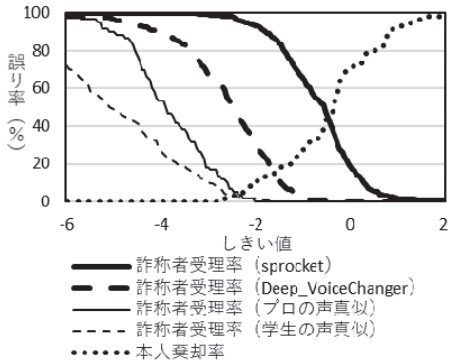

図 1　声質変換器による詐称攻撃を行った場合の話者照合性能

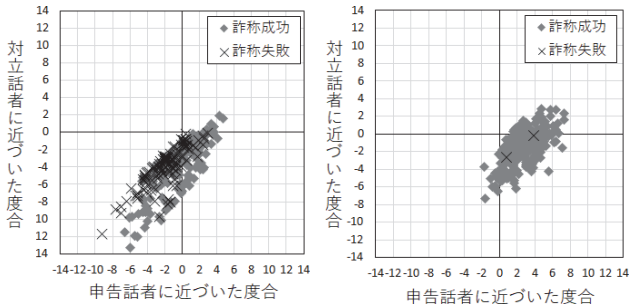

図 2　詐称音声の特徴分析（左：Deep_VoiceChanger，右：sprocket）

結果を見ると，Deep_VoiceChanger では申告話者に近づく効果よりも対立話者から遠ざかる効果によって，sprocket では申告話者の声に近づく効果によって，詐称に成功していることがわかる．前者は，先行研究[1]における学生の声真似の特徴と，後者は物真似タレントの声真似の特徴と近い傾向になっている．

4. まとめ

話者照合システムに対する，声質変換と人間の声真似の詐称攻撃について特徴比較を行った．

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] 高木ら，情報処理学会全国大会，pp.455-456，2020.
[2] Z. Wu, et al., IEEE Journal of Selected Topics in Signal Processing, 2017.
[3] https://github.com/pstuvwx/Deep_VoiceChanger
[4] 戸田ら，システム/制御/情報，vol.62, no.2, pp.69-75, 2018.
[5] http://sox.sourceforge.net/