2ZA-05

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

髙木　啓行†　　岩野　公司†
東京都市大学†

1. はじめに

近年，高いセキュリティを確保するため，様々な情報システムに生体認証技術の導入が進んでいる．「音声による個人認証（話者照合）」は，その手軽さなどから実用への期待が高まっており，様々な研究が進められている[1]．話者照合の実用化を進めるためには，様々な「詐称（成りすまし）」による攻撃に対する脆弱性を十分に把握し，その対策を講じる必要がある．

我々は，最も簡単な詐称攻撃である「声真似（模倣）」に着目し，その攻撃力と特徴について分析を進めている．先行研究[2]では，声真似に特別なスキルを有さない「一般人」と，声質を他人に似せることを日ごろから訓練し，高いスキルを有している「物真似タレント」が行う声真似攻撃について分析を行っている．その結果，「一般人の声真似は，必ずしも攻撃対象話者の声質には近づいていないが，低確率で詐称に成功する」「物真似タレントの声真似は，効率的に攻撃対象者の声質に近づいており，一定の確率で詐称に成功する」傾向が明らかになった．本研究では，この物真似のスキルによる攻撃力の違いをより詳細に分析するため，発話ごとの分析を行って現象の解明を目指す．

2. 実験条件

2.1 使用する音声データ

音声データには先行研究[2]と同じ，6 名の男子学生（一般人）とプロの物真似タレント 1 名（40 代男性，キャリア約 20 年）の 4 桁連続数字発声を利用する．分析用に一般人 6 名を登録話者とする話者照合システムを構築するが，その際には，約 2 日ごとに収録した 5 日分の地声による発声（話者 1 人あたり，1 日 10 回ずつ発声）を学習データとして利用する．この 5 日とは別の日に行われた，本人の地声（10 発声）と他の 5 人の声真似を行ったときの発声（それぞれ 10 発声ずつ）を，照合性能の評価と分析に用いる．また，物真似タレントに男子学生 6 名の声真似による連続数字発声（各 10 回）を依頼し，収録したデータも評価と分析に用いる．

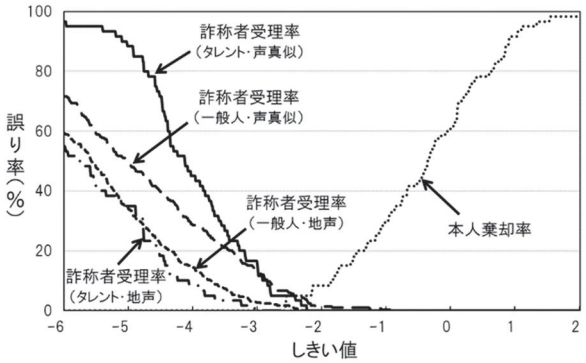

図 1　一般人と物真似タレントによる声真似攻撃時の話者照合性能

2.2 分析用話者照合システム

話者照合には，GMM-UBM 法[3]と同じ原理に基づく手法を利用する．まず，学習データを用いて，各登録話者の話者モデル（申告話者モデル）と，不特定話者モデル（UBM）を 3 状態の隠れマルコフモデル（HMM）で構築する．このとき，不特定話者モデルは 6 名 5 日分の計 300 発声で学習する．音響特徴量には 12 次元 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 次元のベクトルを利用する．照合の際には，入力音声の特徴量系列 X を申告話者モデル（C）と不特定話者モデル（UBM: U）に入力し，それぞれからフレームあたりの平均尤度 $P(X|C)$，$P(X|U)$ を算出する．式(1)で定義される照合スコア $S(X)$ が設定したしきい値よりも大きければ入力音声は申告話者のものとして受理され，小さければ詐称者とみなされ棄却する．

$$S(X) = \log P(X|C) - \log P(X|U) \qquad (1)$$

図 1 に一般人と物真似タレントの地声と声真似発声をそれぞれ詐称者発声として用いたときの，しきい値に対する詐称者受理率を，本人棄却率と共に示す．なお，話者モデルの混合数は照合

Analysis of voice mimicry attack on speaker verification considering impersonators' skills
Haruki Takagi†, Koji Iwano†, †Tokyo City University

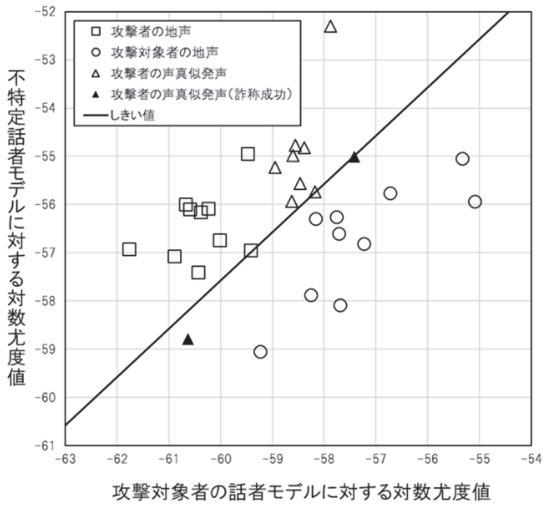

図 2　一般人による声真似発声の尤度分布

性能が最大となるように最適化を行い，64 と設定した．一般人・物真似タレントともに，地声を詐称発声として用いているときには詐称者受理率は低いが，声真似を行うことによって詐称者受理率の上昇が見られ，特にその上昇度合は物真似タレントの方が大きいことが確認できる．

3. 尤度に基づく発声ごとの分析

声真似によって詐称が成功する現象を，より詳細に分析するため，「攻撃者の地声」，「攻撃対象者の地声」，「攻撃者の声真似発声」の三者について，発声ごとに話者照合時の尤度 $P(X|C)$，$P(X|U)$ を求め，尤度空間でどのように発声が分布しているかを調査した．

図 2 は，一般人のある攻撃者がある攻撃対象者の声真似をしたときの尤度分布の例を示す．横軸が $\log P(X|C)$，縦軸が $\log P(X|U)$ の値であり，四角（□）が攻撃者の地声，丸（○）が攻撃対象者の地声，三角（△）が攻撃者の声真似発声を表している．直線は等誤り率を与えるしきい値を表しており，この線の右下が本人として受理される領域となる．したがって，三角のうち黒塗りになっているもの（▲）は，声真似によって詐称が成功した発声である．この攻撃者の多くの声真似発声（△）は地声（□）から右上の方向に移動している．これは，攻撃対象者と不特定話者モデルで表現されている別の話者（対立話者）の双方に近づいていることを意味している．このうち，攻撃対象者に近づく度合いが大きい 1 例で詐称が成功している．一方，声真似によって下方向に大きく移動している発声も 1 例見られ，詐称に成功している．これは，攻撃対象者には近づいていないが，それ以上に不特定話者モデルで表現される対立話者から大きく遠ざかっている（声真似の質は低いが，声の特徴量の変化は大きく，その変化によって偶然，対象者の本人受理領域に到達した）ことを意味している．このように，詐称の成功要因としては，「(a) 攻撃対象者に近づく効果によるもの」と「(b) 攻撃対象者に近づくよりも，不特定話者モデルで表現される対立話者から遠ざかる効果が大きいもの」の 2 種類が考えられる．

表 1　物真似スキル別の詐称成功要因の分析

物真似スキル	攻撃回数	詐称成功数	
		(a) 攻撃対象者に近づく	(b) 対立話者から遠ざかる
低（一般人）	300	12	3
高（タレント）	60	3	0

表 1 に，等誤り率を与えるしきい値のもとで，詐称に成功した声真似発声のうち，要因が (a), (b) に分類されるものの数を，一般人・物真似タレントのそれぞれについて示す．物真似スキルの低い一般人では，(b) の要因で詐称に成功するケースが見られるのに対し，スキルの高い物真似タレントには見られないことがわかる．

以上より，声真似攻撃による詐称が成功する要因は大きく 2 つに分類され，特に物真似のスキルが低い人の声真似では，攻撃対象者の声質には近づかないが，その変化の方向が偶然，本人受理領域に近づくことで詐称に成功してしまう現象がみられることがわかった．

4. まとめ

本研究では，攻撃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を，発声ごとの尤度分布に基づいて行った．その結果，特に物真似のスキルが低い一般人の声真似攻撃の詐称が成功してしまう要因が大きく 2 種類に分類できることがわかった．今後は，それぞれの現象の違いを考慮した声真似攻撃への対応手法の提案を検討する必要がある．

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] 越仲，篠田，“話者認識の国際動向，”日本音響学会誌，vol.69, no.7, pp.342-348, 2013.

[2] 岩野ら，“プロの物真似タレントの声真似が話者照合に与える影響と音響特徴の分析，” 電子情報通信学会技術研究報告，vol. 117, no. 189, pp. 55-60, 2017.

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