氏名	MAWALIM, Candy Olivia		
学位の種類	博士（情報科学）		
学位記番号	博情第 465 号		
学位授与年月日	令和 4 年 3 月 24 日		
論文題目	Content and Privacy Protection Methods for Secure Speech Communication and Its Applications		
論文審査委員	主査　鵜木　祐史	北陸先端科学技術大学院大学	教授
	赤木　正人	同	教授
	党　建武	同	教授
	吉高　淳夫	同	准教授
	岡田　将吾	同	准教授
	伊藤　彰則	東北大学	教授

論文の内容の要旨

Various forms of speech are utilized throughout social media. Advanced speech technology, such as voice conversion techniques and speech synthesis, can synthesize or clone speech entirely as a human voice. Distributing users' speech publicly on a social network without privacy measures affects the security of speech technology and privacy protection. Without protection, speech samples on the internet could be used for theft of personally identifiable information, fraud, and/or authentication of the automatic speaker verification (ASV) system for criminal purposes. Therefore, there must be a solution to the emerging threat of unauthenticated speech signals, such as synthesizing, cloning, and speech conversion.

Speech information hiding (SIH) is one of the approaches for promoting secure speech communication, which is also the main part of this study. Information-hiding-based methods preserve the privacy and security of speech data by imperceptibly embedding particular information that needs to be hidden. SIH has at least three requirements: inaudibility (manipulation does not cause distortion perceivable by the human auditory system), blindness (accurate detection without the original signal), and robustness against common signal processing operations. Although each existing method has advantages, they have shortcomings and need improvement, especially in balancing the trade-off between inaudibility and robustness.

Another approach to improve the trade-off between inaudibility and robustness is considering the features used in speech codecs. Speech codecs are widely applied before speech is transmitted through a communication channel. Thus, using features in speech codecs for speech information hiding improves robustness. Line spectral frequencies (LSFs) are used as features in speech codecs with several speech watermarking methods. LSFs can be directly modified in accordance with a particular speech codec quantization method or manipulated accordingly to control speech formants for representing hidden information.

We investigate a parameter that affects the formation of auditory images, namely the McAdams coefficient, for the feature of SIH in this study. The modification of the McAdams coefficient is useful for adjusting frequency harmonics in audio signals. It has also been introduced for de-identifying or anonymizing speech signals. Since the McAdams coefficient is related to the adjustment of frequency harmonics (related to LSFs), we hypothesize that this coefficient is suitable for speech watermarking.

Another novelty presented in this study is that we propose a speech watermarking method based on a machine learning model. Studies on digital image watermarking based on machine learning models have shown impressive results. However, due to the higher complexity of speech than image data, machine learning models for speech watermarking have not been widely explored. We constructed a machine-learning-based blind detection model by using a binary classification task based on a random forest algorithm (hereafter, we refer to this model as a random forest classifier). The results indicate that our method satisfies the speech watermarking requirements with a 16-bps payload under normal conditions and numerous non-malicious signal processing operations.

Besides the conventional speech codecs, we also analyze a neural vocoder based on the neural source-filter (NSF) model for secure speech communication. We propose a method of improving the primary framework by modifying the state-of-the-art speaker individuality feature (namely, x-vector). Our proposed method is constructed based on x-vector singular value modification with a clustering model. We also propose enhance the proposed technique by modifying the fundamental frequency and speech duration to enhance the anonymization performance. To evaluate our method, we carried out objective and subjective tests. The overall objective test results show that our proposed method improves the anonymization performance in terms of the speaker verifiability, whereas the subjective evaluation results show improvement in terms of the speaker dissimilarity. The intelligibility and naturalness of the anonymized speech with speech prosody modification were slightly reduced (less than 5% of word error rate) compared to the results obtained by the baseline system in Voice Privacy Challenge 2020.

Keywords: speech information hiding, speaker anonymization, McAdams coefficient, x-vector, speech security and privacy

論文審査の結果の要旨

近年，情報通信技術の急速な発達やインフラの整備により，インターネット上でのマルチメディア情報の利用が盛んになっている．特に，サイバーフィジカル空間における音声コンテンツの利用は，スマートフォンの普及や AI スピーカの登場とともに，この数年で急激な伸びを示している．このような急激な需要拡大に対して，音声情報を安心・安全に利用するための技術革新や法整備は相当な遅れをとっており，音声コンテンツの違法コピー・違法配信といった社会問題だけでなく，音声改ざんや音声プライバシー侵害，音声なりすましといった問題も招いている．そのため，サイバーフィジカル空間において，ディジタル表現された音声メディア情報を安心・安全に利用するために，音声コンテンツと音声プライバシーの両方を保護する技術基盤を確立する必要がある．特に，音声コンテンツ保護では，情報秘

匿が難しい音声信号に対し，どのような音声情報表現（符号化）を利用することが適切であるかを検討し, 知覚不可能性, 頑健性, 情報秘匿性を備えた音声情報ハイディング技術基盤を整備する必要がある．また，音声プライバシー保護では，話者情報に係わる音響特徴を明らかにし，話者秘匿が可能な匿名化技術基盤を整備する必要がある．

本研究では，音声コンテンツ保護の実現のため，音声符号化で利用される線スペクトル周波数（LSF）を特徴とした音声情報秘匿法と，McAdams 係数を利用した音声情報秘匿法を確立した．前者では発話内容に係わる特徴に情報秘匿を行い，後者では話者情報に係わる特徴に情報秘匿を行った．次に，音声プライバシー保護を実現するために，音声の個人性を表出する統計量（X-Vector）を対象に特異値分解に基づく話者の匿名化技法を確立した．この方法は，国際会議 Interspeech2020 で企画された Voice Privacy Challenge 2020 に参加し，基準をクリアしたことから高い評価を受けた．最後に，McAdams 係数を利用した音声情報秘匿法を匿名化技法にも活用し，音声コンテンツと音声プライバシーの両方を保護する処理フレームワークを確立した．これは，本研究の最大の成果であり，話者情報を情報秘匿した安心・安全な音声コミュニケーションを提供可能とした．

以上，本論文は，サイバーフィジカル空間における音声コミュニケーションのセキュリティを高めるための一つの革新的技術を提供した．本技術は，音声情報秘匿・話者匿名として，音声改ざん防止や音声なりすまし防止といった重要課題に対して応用範囲が広く，学術的に貢献するところも大きい．よって博士（情報科学）の学位論文として十分価値あるものと認めた．