様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ６ 月　 ２ 日現在

機関番号：６２６１５
研究種目：基盤研究(B)（特設分野研究）
研究期間：2018～2020
課題番号：１８ＫＴ００５１

研究課題名（和文）生体検知とASVspoofチャレンジによる安全・安心な音声情報処理システムの実現

研究課題名（英文）Safe and secure speech information processing based on liveness detection and ASVspoof challenge

研究代表者
山岸　順一（Yamagishi, Junichi）

国立情報学研究所・コンテンツ科学研究系・教授

研究者番号：７０７０９３５２

交付決定額（研究期間全体）：（直接経費）　　14,200,000円

研究成果の概要（和文）：音声情報処理が社会に普及するにつれ、話者照合や音声認識に対する攻撃が起き始めた。 本研究の目的は、音声の生体検知技術を高度化し、問題の一解決法を提示することである。 生体検知とは「他人により許可なく 取得され、加工され、外部デバイスにより再生された音声」と「生体からその場で発声した生の音声」を区別する機械学習技術である。そこで最先端の音声合成や声質変換技術により合成された音声を大量に含むDBを構築、コンペティションを開催、分野全体で生体検知技術を高度化することを行った。聴覚上は差がわからない様な加工音声に対しても、生体検知を行える様になり、安全安心な音声情報処理を実現する一解決法が得られた。

研究成果の学術的意義や社会的意義
音声情報処理は多くのスマートデバイスで利用されており、社会を支える基盤技術である。音声の生体検知は音声インターフェースの手軽さとトラストの両方を同時に実現する技術であり、社会的意義は高い。実際、本研究を通して構築し、一般公開したDBは、世界のアカデミック組織のみならず、多くの企業にも利用されている。学術的意義も高く、多くの国際会議論文が本DBを利用している。現在AI技術により生成されたメディアが悪用される事が危惧され、deepfakeと呼ばれることもある。本研究は、音声を対象に研究を行ったが、その成果は映像や文字等にも応用可能であると考えられ、今後さらに発展させることが可能であると期待される。

研究成果の概要（英文）：As speech processing became more widespread in society, attacks on speaker verification and speech recognition began to occur. The purpose of this research is to improve the liveness detection technology of speech and to present a solution to the problem. Liveness detection is a machine learning technology that distinguishes between "voice obtained by another person without permission, processed, and reproduced by an external device" and "live voice uttered on the spot from the living body". Therefore, we built a DB containing a large amount of audio files synthesized by the latest speech synthesis and voice conversion technology, held a competition, and advanced the liveness detection technology in the field. It has become possible to detect artificial voices for which there is no perceived audible difference, and a solution has been obtained that realizes safe and secure voice information processing.

研究分野：音声情報処理

キーワード：音声情報処理　話者照合　生体検知　生体認証　音声インタフェース

１．研究開始当初の背景

研究開始当時、音声インターフェースが社会へ爆発的に普及していた。また同時に要素技術である話者照合や音声認識システムの脆弱性を悪用した攻撃も起き始めていた。話者照合による個人認証に対しては、所望の人の声をテキストから合成する、もしくは、他人から所望の人の声へ変換する技術により、他人になりすませることが以前から報告されていたが、それだけでなく、音声認識システムに対する攻撃も新たに報告され始めていた。そこで我々は、この問題を解決するためには、提示された音声が、生体つまり人間がその場で発声した生の音声かどうかを判断する「生体検知技術」を研究し、発展させる必要があると考え、生体検知技術を高度化・高精度化させる本研究を開始するに至った。更に個々の研究室単位での研究だけでは限界がある事から、本技術に関する競争型研究を実施することで分野全体の研究速度を加速させ、音声インターフェースの手軽さとトラストの両方を同時に実現する研究を加速させる必要があるとも考えた。

２．研究の目的

本研究課題の核心を成す科学的・技術的問いは、「他人により許可なく取得され、場合によっては加工され、何らかの外部デバイスにより再生された音声」と「生体からその場で発声した生の音声」を自動でかつ精密に区別できるか、ということである。もしこの問いに対する回答が Yes であれば、音声認識や話者照合の前段で、非生体による音声信号を自動的に棄却することができ、トラストな信号に対してのみ、音声認識や話者照合を実行できる様になる。

どの様に特徴量を抽出し、モデルを構築すれば、この様な自動検出が可能であろうか？もし、攻撃者により音声が加工されている場合、何らかの音響特性が変わっている可能性が高い。その通常とは異なる音響特性を検出できるかどうかを見極める必要がある。もし、他人により音声が取得され、外部デバイスにより再生された場合、音声が意図的に加工されていない場合でも、攻撃者が利用する収録装置と再生装置により音響特性が変わり、さらに、攻撃者が音声を取得した際の背景音が実際の環境音に重畳され、異なる音響特性になり得る。現実には、これらの複数の要素が組み合わされ、システム側はその様な微少な差異を頑健に検出する必要がある。

また、我々はこれまで話者照合システムの生体検知精度を競う ASVSpoof challenge を開催し、分野における牽引を鋭意行ってきた。しかし、これまで構築したデータベースでは、プレイバック攻撃、音声合成攻撃、声質変換攻撃などの主要な攻撃法を個別に評価していた。現実には、どの攻撃が使われるかは未知であり、多数の攻撃を網羅的にカバーした新たなデータベースを構築し、攻撃手法に依存しない生体検知技術の確立、及び、その統一的な評価法を実現する必要もある。

更に、前述した攻撃法以外の攻撃法も想定し、未知の脆弱性に対する検証も先回りして行うことも必要である。例えば、他のバイオメトリクスにおける問題が、音声インターフェースに対しても起きるのかどうか、どの様な対応が必要かを事前に検証する事も目的の１つである。

３．研究の方法

上記研究目的を達成するため、以下の３つの課題に取り組む。

【その１：音声生体検知技術の更なる向上】
これまでの生体検知研究は収録音声のプレイバック攻撃、音声合成攻撃、声質変換攻撃などの攻撃法を個別に評価していた。そこで、これら全てを網羅的にカバーした新たなデータベースを構築し、全て攻撃に頑健な生体検知技術の確立、及び、その統一的な評価法を実現する。さらに、2019 年に ASVspoof challenge 2019 を開催することで、音声の生体検知分野を牽引する。
【その２：音声生体検知技術の対象拡大】
生体検知の対象を話者照合以外にも拡大し評価を行う。その際、one-class learning など異常値検出法を新たな生体検知技術も検討する。
【その３：音声インターフェースに対して近い将来に想定される攻撃に対する検証と対策】
音声インターフェースを利用するデバイスの種類は毎年数倍に増えており、新たな攻撃法が常に考え出されると予想される。それ故、未知の脆弱性に対する検証も先回りして行う。

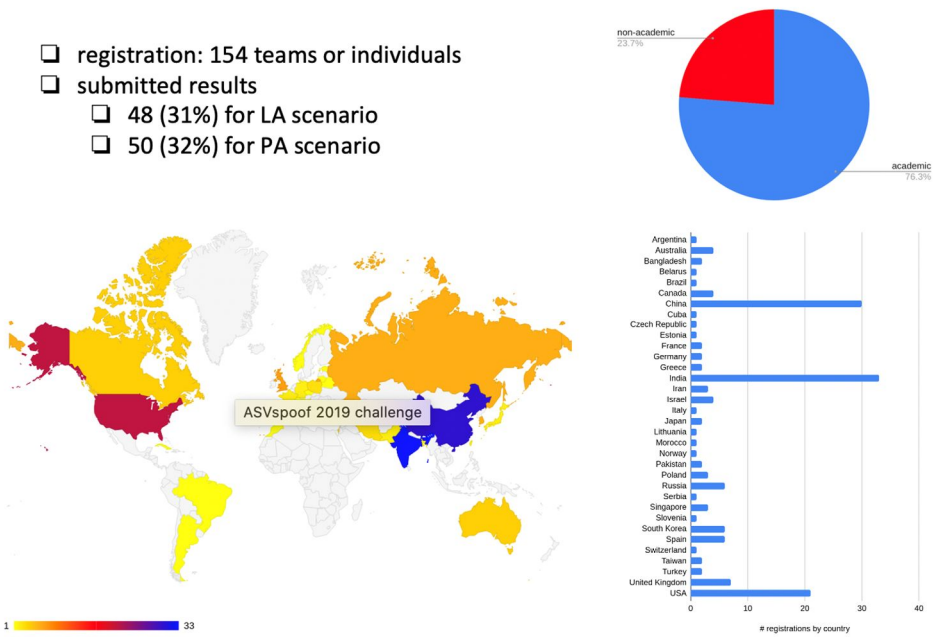
４．研究成果

本プロジェクトでは、上記課題に対して以下の成果を挙げた。
【その１：音声生体検知技術の更なる向上】
近年の深層学習を利用した音声合成や声質変換は、エンターテインメント等にて新たな価値をもたらすと考えられるものの、悪用された場合には話者認識システム等においてセキュリティー上の問題も起こす。話者認識の安全性と頑健性の向上のため、仏 Eurecom 研究所および東フ

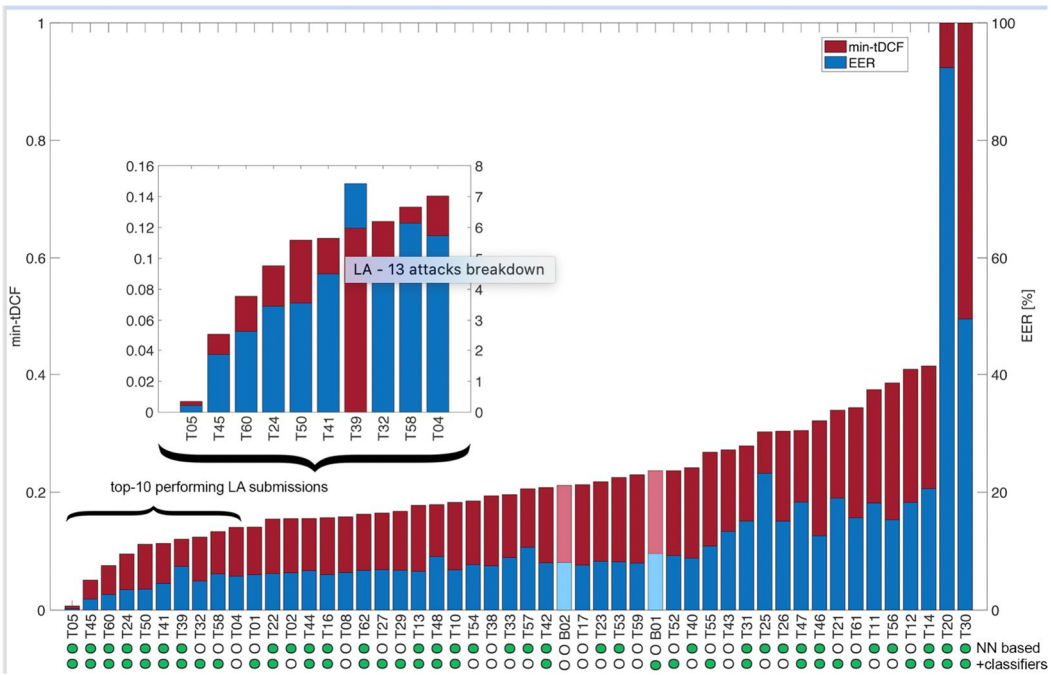
ィンランド大と協力し、話者照合に対するなりすまし攻撃を自動的に防御する生体検知技術を共通コーパスで比較する ASVspoof challenge 2019 を世界規模で開催した［1,2］。

音声合成や声質変換によるなりすまし攻撃を想定した「Logical access タスク」と、音声の単純な再生によるリプレイ攻撃を想定した」Physical access タスク」の 2 種類を想定し、それぞれ大規模データベースを構築した[3]。Logical access タスクにおいては、19 種類の異なるアルゴリズムによる合成音声や変換音声を、Google，iFlytek，NTT 等の複数企業の協力のもと用意した。Physical access タスクにおいては、様々な条件によるリプレイをシミュレーションにより大量に生成した。

大規模データベースはチャレンジ参加希望の 154 組織に配布され、そのうち 50 組織が実際に生体検知モデルを構築し、未知の攻撃手法が大量に含まれる評価データの真贋判定を行った。国別チャレンジ参加組織数およびアカデミック・企業の比率を下図に示す。データベースは 2019 年秋に無償公開され、更に多くの機関に利用されている。



統一指標についても検討を行なった。チャレンジ参加者のライブネス検知モデルの精度評価およびランキングは、等価誤り率(EER)に加え、個人認証の性能と生体検知の性能を統合した tandem DCF と呼ぶ新たな指標を提案し、評価を行なった[4]。下図は Logical access タスクの結果である。様々な分析の結果、人間には聴覚上差がわからない様な詐称音声でも識別可能である事が確認されている。



また、音声の生体検知に関する学術発表の場を研究コミュニティーに提供するため、国際会議 Interspeech 2019 および ASRU 2019 の両方において、スペシャルセッションを開催した。それ

に加え、国際ジャーナル誌 Computer Speech Language における特集号も企画した。その他、データベース構築に協力した Google およびチャレンジ参加企業 ID R&D Inc のプレスリリースがあったことから、米国における新聞報道も多数あり、社会的に大きな反響も得た。

更に、話者認識システムと音声の生体検知システムとを同時に学習する枠組みについても新たに検討した。生体認証および生体検知の指標(EER や DCF)は通常微分不可能である事から、深層モデルの学習には直接利用されず、単純な end-to-end 学習は不可能である。そこで、教師あり学習ではなく、強化学習を新たに導入することで、話者認識システムと生体検知システムの両方を同時に高精度化し、安全性と頑健性を高める研究も行なった[5]。これは東フィンランド大との共同研究成果である。

【その２：音声生体検知技術の対象拡大】
話者照合以外の生体検知手法として、手のひら画像に対象を拡大し、評価を行った。音声を含む生体偽造物は、生体模倣を目的とした偽造物と、生体認証システムの誤動作を目的とした（生体模倣を目的としない）偽造物の両方が存在する。そこで、両者の攻撃を想定したデータベースを構築するとともに、One-Class-Learning を用いた異常値検知により、これらの偽造物を高精度で検知する手法を提案した[6]。

【その３：音声インターフェースに対して近い将来に想定される攻撃に対する検証と対策】
音声インターフェースは人間の声だけでなく，環境に存在する様々な音響を取得している。ここでは，スマートフォンのキー入力に伴い発生するタップ音に着目し，「正規ユーザがスマートフォンにキー入力を行う際のタップ音を，音声インターフェース等のマイクで盗聴する」という攻撃シナリオの下で、タップ音から入力内容が推定可能であるという脅威を検証した[7]。

参考文献
[1] Massimiliano Todisco, Xin Wang, Ville Vestman, Md Sahidullah, Hector Delgado, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Tomi Kinnunen, Kong Aik Lee, “ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection,”Proc. Interspeech 2019, Sept. 2019
[2] Andreas Nautsch, Xin Wang, Nicholas Evans, Tomi Kinnunen, Ville Vestman, Massimiliano Todisco, Héctor Delgado, Md Sahidullah, Junichi Yamagishi, Kong Aik Lee, "ASVspoof 2019: Spoofing Countermeasures for the Detection of Synthesized, Converted and Replayed Speech," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 2, pp. 252-265, April 2021
[3] Xin Wang, Junichi Yamagishi, Massimiliano Todisco, Héctor Delgado, Andreas Nautsch, Nicholas Evans, Md Sahidullah, Ville Vestman, Tomi Kinnunen, Kong Aik Lee, Lauri Juvela, Paavo Alku, Yu-Huai Peng, Hsin-Te Hwang, Yu Tsao, Hsin-Min Wang, Sébastien Le Maguer, Markus Becker, Fergus Henderson, Rob Clark, Yu Zhang, Quan Wang, Ye Jia, Kai Onuma, Koji Mushika, Takashi Kaneda, Yuan Jiang, Li-Juan Liu, Yi-Chiao Wu, Wen-Chin Huang, Tomoki Toda, Kou Tanaka, Hirokazu Kameoka, Ingmar Steiner, Driss Matrouf, Jean-François Bonastre, Avashna Govender, Srikanth Ronanki, Jing-Xuan Zhang, Zhen-Hua Ling, “ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech,” Computer Speech & Language, Volume 64, 2020
[4] Tomi Kinnunen, Hector Delgado, Nicholas Evans, Kong Aik Lee, Ville Vestman, Andreas Nautsch, Massimiliano Todisco, Xin Wang, Md Sahidullah, Junichi Yamagishi, Douglas A. Reynolds, "Tandem Assessment of Spoofing Countermeasures and Automatic Speaker Verification: Fundamentals," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2195-2210, 2020
[5] Anssi Kanervisto, Ville Hautamäki, Tomi Kinnunen, Junichi Yamagishi “An initial investigation on optimizing tandem speaker verification and countermeasure systems using reinforcement learning” Proc. Odyssey 2020 The Speaker and Language Recognition Workshop, 151-158, 2020
[6] T. Ohki, V. Gupta and M. Nishigaki, "Efficient Spoofing Attack Detection against Unknown Sample using End-to-End Anomaly Detection," 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 2019, pp. 224-230
[7] Yumo Ouchi, Ryosuke Okudera, Yuya Shiomi. Kota Uehara, Ayaka Sugimoto, Tetsushi Ohki, and Masakatsu Nishigaki., "Study on Possibility of Estimating Smartphone Inputs from Tap Sounds," 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Auckland, New Zealand, 2020, pp. 1425-1429.

５．主な発表論文等

〔雑誌論文〕　計10件（うち査読付論文　10件／うち国際共著　8件／うちオープンアクセス　10件）

１．著者名 Nautsch Andreas、Wang Xin、Evans Nicholas、Kinnunen Tomi H.、Vestman Ville、Todisco Massimiliano、Delgado Hector、Sahidullah Md、Yamagishi Junichi、Lee Kong Aik	４．巻 3
２．論文標題 ASVspoof 2019: Spoofing Countermeasures for the Detection of Synthesized, Converted and Replayed Speech	５．発行年 2021年
３．雑誌名 IEEE Transactions on Biometrics, Behavior, and Identity Science	６．最初と最後の頁 252～265
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TBIOM.2021.3059479	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Wang Xin、Yamagishi Junichi、Todisco Massimiliano、Delgado Hector、Nautsch Andreas、Evans Nicholas、Sahidullah Md、Vestman Ville、Kinnunen Tomi、Lee Kong Aik、Juvela Lauri、Alku Paavo、Peng Yu-Huai、Hwang Hsin-Te、Tsao Yu、Wang Hsin-Min、Maguer Sebastien Le、Becker Markus、Henderson Fergus et al	４．巻 64
２．論文標題 ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech	５．発行年 2020年
３．雑誌名 Computer Speech & Language	６．最初と最後の頁 101114～101114
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.csl.2020.101114	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Kinnunen Tomi、Delgado Hector、Evans Nicholas、Lee Kong Aik、Vestman Ville、Nautsch Andreas、Todisco Massimiliano、Wang Xin、Sahidullah Md、Yamagishi Junichi、Reynolds Douglas A.	４．巻 28
２．論文標題 Tandem Assessment of Spoofing Countermeasures and Automatic Speaker Verification: Fundamentals	５．発行年 2020年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 2195～2210
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/taslp.2020.3009494	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Kanervisto Anssi、Hautamaki Ville、Kinnunen Tomi、Yamagishi Junichi	４．巻 －
２．論文標題 An Initial Investigation on Optimizing Tandem Speaker Verification and Countermeasure Systems Using Reinforcement Learning	５．発行年 2020年
３．雑誌名 ISCA The Speaker and Language Recognition Workshop Odyssey 2020	６．最初と最後の頁 151～158
掲載論文のDOI（デジタルオブジェクト識別子） 10.21437/Odyssey.2020-22	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Yumo Ouchi, Ryosuke Okudera, Yuya Shiomi, Kota Uehara, Ayaka Sugimoto, Tetsushi Ohki, Masakatsu Nishigaki	4．巻 －
2．論文標題 Study on Possibility of Estimating Smartphone Inputs from Tap Sounds	5．発行年 2020年
3．雑誌名 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)	6．最初と最後の頁 1425-1429
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 C. Lai, A. Abad, K. Richmond, J. Yamagishi, N. Dehak and S. King	4．巻 -
2．論文標題 Attentive Filtering Networks for Audio Replay Attack Detection	5．発行年 2019年
3．雑誌名 Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	6．最初と最後の頁 6316-6320
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICASSP.2019.8682640	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Massimiliano Todisco, Xin Wang, Ville Vestman, Md. Sahidullah, Hector Delgado, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Tomi H. Kinnunen, Kong Aik Lee	4．巻 -
2．論文標題 ASVspoof 2019: Future Horizons in Spoofed and Fake Audio Detection	5．発行年 2019年
3．雑誌名 Proc. Interspeech 2019	6．最初と最後の頁 1008-1012
掲載論文のDOI（デジタルオブジェクト識別子） 10.21437/Interspeech.2019-2249	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Ohki Tetsushi、Gupta Vishu、Nishigaki Masakatsu	4．巻 -
2．論文標題 Efficient Spoofing Attack Detection against Unknown Sample using End-to-End Anomaly Detection	5．発行年 2019年
3．雑誌名 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)	6．最初と最後の頁 224-230
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/APSIPAASC47483.2019.9023183	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Massimiliano Todisco, Hector Delgado, Kong Aik Lee, Md Sahidullah, Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi	4．巻 1
2．論文標題 Integrated Presentation Attack Detection and Automatic Speaker Verification: Common Features and Gaussian Back-end Fusion	5．発行年 2018年
3．雑誌名 Proc. Interspeech 2018	6．最初と最後の頁 77-81
掲載論文のDOI（デジタルオブジェクト識別子） 10.21437/Interspeech.2018-2289	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Fuming Fang, Junichi Yamagishi, Isao Echizen, Md Sahidullah, Tomi Kinnunen	4．巻 1
2．論文標題 Transforming acoustic characteristics to deceive playback spoofing countermeasures of speaker verification systems	5．発行年 2018年
3．雑誌名 2018 IEEE International Workshop on Information Forensics and Security (WIFS)	6．最初と最後の頁 1-9
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/WIFS.2018.8630764	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計8件（うち招待講演　4件／うち国際学会　1件）

1．発表者名 山岸順一
2．発表標題 深層生成モデルによるメディア生成とフェイク検知
3．学会等名 第23回情報論的学習理論ワークショップ（IBIS2020）（招待講演）
4．発表年 2020年

1．発表者名 藤垣成汰朗，成田惇，塩見裕哉，菅沼弥生，西垣正勝，大木哲史
2．発表標題 ディープフェイク画像からの個人再識別化に関する検討
3．学会等名 暗号と情報セキュリティシンポジウム2021
4．発表年 2021年

１．発表者名
大内結雲，奥寺瞭介，塩見祐哉，大木哲史，西垣正勝

２．発表標題
スマートフォンのタップ音からの入力内容推測可能性に関する研究(その2)

３．学会等名
暗号と情報セキュリティシンポジウム2021

４．発表年
2021年

１．発表者名
山岸順一

２．発表標題
話者照合の生体検知チャレンジ「ASVspoof 2019」の概要と今後の展望

３．学会等名
第9回バイオメトリクスと認識・認証シンポジウム（招待講演）

４．発表年
2019年

１．発表者名
山岸 順一

２．発表標題
フェイク動画問題: メディア解析技術によるアプローチ

３．学会等名
JST/CRDS 公開ワークショップ 「意思決定のための情報科学 ～情報氾濫・フェイク・分断に立ち向かうことは可能か～」 2019年7月25日（招待講演）

４．発表年
2019年

１．発表者名
Junichi Yamagishi

２．発表標題
Speaker Identity Cloning and Protection

３．学会等名
AFEKA SPEECH PROCESSING CONFERENCE 2019: 10-YEAR ANNIVERSARY CONFERENCE（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Vo Ngoc Khoi Nguyen，西垣正勝，大木哲史

２．発表標題
生体認証を回避する物理的なAdversarial Exampleの検討

３．学会等名
第82回情報処理学会全国大会

４．発表年
2020年

１．発表者名
大内結雲，奥寺瞭介，塩見裕哉，上原航汰，杉本彩歌，大木哲史，西垣正勝

２．発表標題
スマートフォンのタップ音からの入力内容推測可能性に関する研究

３．学会等名
暗号と情報セキュリティシンポジウム2020，

４．発表年
2020年

〔図書〕　計1件

１．著者名	４．発行年
Md Sahidullah, Hector Delgado, Massimiliano Todisco, Tomi Kinnunen, Nicholas Evans, Junichi Yamagishi, and Kong-Aik Lee	2019年
２．出版社	５．総ページ数
Springer	41
３．書名	
Introduction to Voice Presentation Attack Detection and Recent Advances (Chapter 15, Handbook of Biometric Anti-Spoofing, 2nd edition)	

〔産業財産権〕

〔その他〕

ASVspoof website
https://www.asvspoof.org

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	大木　哲史 (Ohki Tetsushi) (80537407)	静岡大学・情報学部・講師 (13801)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
フランス	Eurecom	ロレーヌ大学	Avignon University	他1機関
フィンランド	University of East Finland	Aalto University		
中国	中国科学技術大学	iFlytek Research		
米国	MIT	Google Inc	Johns Hopkins University	
英国	University of Edinburgh			
米国	Johns Hopkins University			