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　元 年　 ６ 月　 ４ 日現在

機関番号：２２６０４
研究種目：若手研究(B)
研究期間：2016～2018
課題番号：１６Ｋ１６０８８

研究課題名（和文）セキュアな話者照合のためのポップノイズバランスを考慮したコーパス構築

研究課題名（英文）Corpus development considering pop-noise balance for robust speaker verification systems

研究代表者
塩田　さやか（Shiota, Sayaka）

首都大学東京・システムデザイン研究科・助教

研究者番号：９０７０５０３９

交付決定額（研究期間全体）：（直接経費）　　3,000,000円

研究成果の概要（和文）：信頼性の高い話者照合システムを構築するために、どのような発話内容ならスピーカ再生によるなりすまし攻撃を防ぎ高い照合性能を維持できるかを調査する必要があった。そのため本研究では再生音声と実発話の根本的な違いである呼気を検出するポップノイズ検出システムを構築した。また、発話内容の調査のために必要な音声コーパスの構築を行った。コーパスの性能を評価するためにポップノイズバランスを考慮した文章としなかった文章で照合精度を比較した結果、特にバランスを考慮した文章において照合エラー率が低くなり、話者照合システムの安全性が向上したことを確認した。

研究成果の学術的意義や社会的意義
近年、音声対話システムやスマートスピーカーの普及により音声を入出力インタフェースとして機械と会話や命令をする機会が増えてきている。特に電話の発信やスケジュールの確認など個人的な情報を発声する場合、これを第三者が盗聴してなりすまし音声に使って個人情報を抜き出したり誤った操作を行わせることができてしまうという問題がある。そのためなりすまし音声の検出は重要課題であり、近年話者照合という声を使った生体認証技術に関する分野においても国内外、多くの研究機関によって研究が行われている。なりすまし検出の精度が向上し普及することで機械との音声対話の安全性も大幅に向上することが期待できる。

研究成果の概要（英文）：This research focuses on the corpus development considering pop-noise balance for speaker verification systems. Recently, spoofing against speaker verification systems became serious problems, and many research groups start to research anti-spoofing countermeasures. We focus on a voice liveness detection (VLD) approach which is regarded as one of the anti-spoofing countermeasure. The voice liveness detection approach is to detect the inputs came from a genuine human or a loudspeaker. To improve the robustness of the VLD, the pop-noise detection has been proposed. It reported that the accuracies depend on the sentences. Therefore, the pop-noise balanced sentences are designed, and the new database is recorded by using the designed sentences. From the experimental results, the designed sentence can improved the performance of the VLD systems. It means that the speaker verification systems can protect from the spoofing attacks by using the VLD systems.

研究分野：音声信号処理

キーワード：話者認識　話者照合　声の生体検知　なりすまし検出　ポップノイズ検出　音素リスト

１．研究開始当初の背景

ユビキタス社会の到来により，建物への出入や携帯電話，パソコンなどのセキュリティとして指紋や静脈など様々な生体情報を用いた生体認証を利用することが普及してきている．生体認証の中でも，マイクが一つあれば導入が可能であるという点から，音声を用いた生体認証である話者照合が注目を浴びてきている．また，次に述べる二つの要因が後押しし，今後ますます普及することが期待されている．要因の一つ目は，話者照合システムの識別性能が飛躍的に向上し指紋や顔画像と同等の精度が得られるようになってきたことであり，もう一つはスマートフォンやスマートスピーカーなどの音声対話システムの普及により，機械とのやり取りを行うインタフェースとして声を使う機会が増えてきていることである．

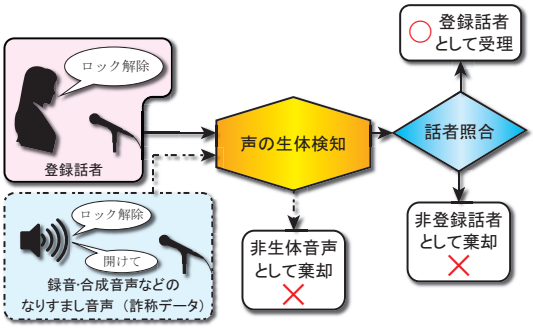

図 1　声の生体検知と話者照合システム

しかしながら，話者照合の性能が向上する一方で，音声合成や声質変換といった特定の人物の声を本人に似せてつくる技術も品質が飛躍的に向上してきた．また，携帯電話や IC レコーダなどが簡単に手に入ることから音を録音すること自体も簡単になったため，たとえ最先端の話者照合システムを用いたとしても合成音声や再生音声を使ったなりすまし攻撃（詐称攻撃）が可能であったことが報告されている．この詐称攻撃に関する問題は声だけに限らず，顔画像や指紋・虹彩などを用いる生体認証技術について重大な問題となっている．これまでに提案されてきた音声のなりすまし攻撃に関する対策法は，登録話者の音声と登録話者に似せた合成音声を識別するために様々な音響的特徴量を用いた対処法が主流であった．しかし，これらの検出法は現在の音声合成・声質変換のメカニズムを前提とした対処法であり，なりすまし音声による詐称攻撃に対する根本的な解決法には至っていなかった．そこで研究代表者は問題の根本的解決のために，入力された音声が本当に生きている人間から発せられた音声かスピーカーで再生された音声なのかを判断する枠組みである声の生体検知を世界に先駆けて提案してきた．声の生体検知と話者照合システムの利用イメージは図 1 に示すとおりである．

人間がマイクに向かって話しかける時にマイク内部に息が入りこむことで発生してしまうボフッやバフッといったノイズをポップノイズと呼ぶ．声の生体検知ではこのポップノイズを検出することで生体の判定を行っている．現在の検出方法は発話内にポップノイズを含むか否かで入力音声の発声元が生きている人間なのか否かを判定するという非常にシンプルな方法を用いている．しかし実際には，詐称者が何に着目して生体検知が行われているとわかっていても簡単には再現できない検知方法を実現する必要がある．ポップノイズは“p”や“t”といった破裂音を発声するときに起こりやすく，“a”や“u”といった音素には起こりにくいという音素依存性がある．そこで，発話内のポップノイズの有無を判定するだけでなくポップノイズが発生する箇所の音素や音響的特徴についても着目する必要がある．しかしながら，声の生体検知に関する研究は研究代表者らが世界に先駆けて始めた研究であり重要な課題であるという認識が高まる一方，ポップノイズを陽に含むデータベースというものは研究代表者が構築したものしかなく，圧倒的にデータ量が少ないという問題がある．

２．研究の目的

声の生体検知では，人間の息がマイク内に入ることで録音された音声に乗ってしまうポップノイズを生体固有の情報として用いている．声の生体検知は画期的な解決策である一方，データが圧倒的に足りないという問題がある．これは，これまでに公開されてきた様々な音声コーパスはポップノイズのようなノイズは雑音としてみなしているためになるべく音声データに含まれないように設計して収録を行っているからである．これまでに研究代表者は独自収録を行った小規模のデータを用いた実験を行い，ポップノイズを含む音声を用いた場合の声の生体検知の有効性を示してきた．しかし，データ量が少ないために発表論文としてのインパクトも少なく，また分析するための発話内容なども不十分であるという問題があった．そこで本研究では，分析が明確にするためにポップノイズバランスを考慮した汎用性の高い音声コーパスの構築を目的とした．ポップノイズを陽に捉えた文章設計及びコーパスの構築どちらも高い新規性を持っており，分野に対しても貢献度が高いといえる．

３．研究の方法

本研究の研究計画は大きく分けるとプロンプト文セットの設計と音声収録および評価に分けられる．研究の流れとしてはまずプロンプト文セット設計のために図 2 に示すようなポップノイズ検出法を用いることでポップノイズが含まれる音声データからポップノイズの入りやすい音素の傾向，入りにくい音素の傾向を調査する．設計されたプロンプト文セットを用いた音声

収録を行い実際にポップノイズが含まれる割合が設計通りか，話者照合や声の生体検知にも使用可能なコーパスとなるかを評価しその都度プロンプト文セットの修正および改善，収録したデータの分析を行いつつ収録を重ね，最終的にはSpoofing-challengeなどで配布可能なコーパスとなるようコーパスの整備を行いまとめるよう研究を進める．

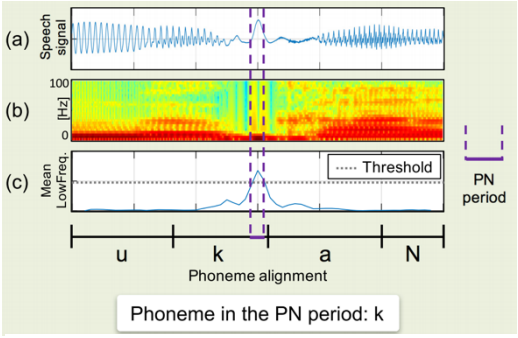

図 3 ポップノイズ検出区間検出手順

４．研究成果

本研究計画を通じて得られた研究成果をまとめる．図 3 に様々な手法を用いて行った話者照合実験の等価エラー率を示している．本研究では，ポップノイズの入りやすい音素と入りにくい音素について分析を行い，それぞれの特徴を積極的に活用することを計画していた．図に示すEPNおよびHPN phoneme detectionと記載されている手法がそれぞれの音素の特徴を活用した手法になっている．また，本研究によって構築したデータベースは音素バランスを調整した文章とそうでない文章を用意しておりそれぞれを比較することで提案手法およびデータベースの有効性を示す必要があった．そこでさらに従来プロンプト文，提案プロンプト文それぞれを用いた実験結果を青色と赤色の棒グラフによって示した．左の 2 項目がなりすまし攻撃がない場合とある場合で話者照合システムの受ける影響を示している．また，(A)から(D)がなりすまし検出を行ったあとに話者照合を行うという図1のシステムを実行した場合の等価エラー率を示している．はじめに，なりすまし攻撃なしの EER となりすまし攻撃ありの EER を比較する．従来プロンプト文と提案プロンプト文のどちらにおいてもなりすまし攻撃なしの EER に比べて，攻撃ありの EER の方が高くなっている．このことから，話者照合システムは登録話者の音声を録音再生するなりすまし攻撃に対して脆弱であることが確認できる．次になりすまし攻撃ありの EER とポップノイズ検出による EER を比較する．従来プロンプト文では，攻撃ありの EER とポップノイズ検出（A）による EER では変化がない．一方で，提案プロンプト文ではポップノイズ検出により EER が約 0.27 ポイント改善した．これは，従来プロンプト文では実発話と再生音声間で生じなかったポップノイズ発生の差が，提案プロンプト文では生じたため，ポップノイズ検出により再生音声を棄却することができたためである．次にEPN 音素検出（B）およびHPN 音素検出（C）による EER に着目する．従来プロンプト文を用いたとき，EPN 音素検出の EER が他の手法と比べて最も低い EER が得られ，なりすまし攻撃ありの EER から約 0.13 ポイント改善した．一方，HPN 音素検出による EER はベースラインであるなりすまし攻撃ありの EER よりも悪化した．これは従来プロンプト文に入っている音素と HPN 音素検出に用いた音素リストが合わず，再生音声だけでなく実発話まで棄却してしまうなど正しく生体検知できなかったためである．提案プロンプト文でも EPN 音素検出の EER が他手法の中で最も低い EER が得られ，なりすまし攻撃ありのものと比較して約 0.25 ポイント改善した．また HPN 音素検出による EER も，EPN 音素検出に比べて改善は少ないものの，ポップノイズ検出による EER よりも低下した．これはプロンプト文と音素リストが合っていたため，実発話を棄却しすぎることなく，話者照合に影響を与える再生音声を棄却できたことを示している．また，なりすまし攻撃ありの EER から声の生体検知によって最も改善された EER は，従来プロンプト文で約 0.13 ポイント，提案プロンプト文で約 0.25 ポイントであることから，提案プロンプト文を用いることで，声の生体検知と話者照合を統合したシステムがより頑健になるといえる．最後に EPN-HPN 音素検出（D）のときの EER を見ると，なりすまし攻撃ありの EER と比較して，従来プロンプト文では約 0.08 ポイントしか改善していないのに対し，提案プロンプト文では約 0.25 ポイント改善している．以上より，ポップノイズの発生頻度を考慮したプロンプト文を用いることで，声の生体検知および話者照合のなりすまし攻撃に対する頑健性が向上するといえる．

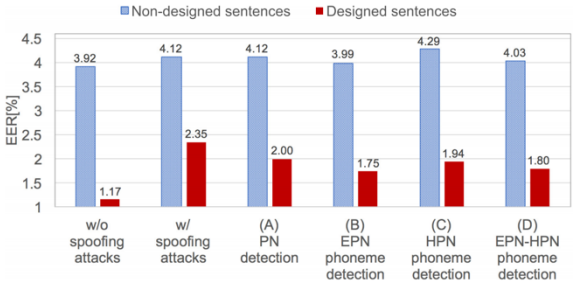

図 2 なりすまし検出と話者照合結果

これらの結果から助成金を用いて構築したコーパスの有効性が示された．本研究の目的の一つであったコーパス公開に関しては個人情報保護の観点及び生体情報を扱うという観点から限定的なものとなってしまったが，問題提起としては十分の成果があった．

５．主な発表論文等
〔雑誌論文〕（計　１件）
望月 紫穂野，塩田 さやか，貴家 仁志，“話者照合のための音素情報を考慮したポップノイズ検出法による声の生体検知，”電子情報通信学会 論文誌，査読あり，vol.J101-D，no.3，pp.588-596，2018年3月．DOI:10.14923/transinfj.2017PDP0017
〔学会発表〕（計　32件）
① Sayaka SHIOTA, Shinnosuke Takamichi, Tomoko Matsui, “DATA AUGMENTATION WITH MOMENT-MATCHING NETWORKS FOR I-VECTOR BASED SPEAKER VERIFICATION,” Proc. APSIPA Annual Summit and Conference, November, 2018.
② Haruna MIYAMOTO, Sayaka SHIOTA, and Hitoshi KIYA, “Non-linear Harmonic Generation Based Blind Bandwidth Extension Considering Aliasing Artifacts,” Proc. APSIPA Annual Summit and Conference, November, 2018.
③ Shihono MOCHIZUKI, Sayaka SHIOTA, and Hitoshi KIYA, “Voice liveness detection using phoneme-based pop-noise detector for speaker verification,” Proc. The Speaker and Language Recognition Workshop Odyssey, June, 2018.
④ Ryosuke NAKANISHI, Sayaka SHIOTA, and Hitoshi KIYA, “Ensemble Based Speaker Verification Using Adapted Score Fusion in Noisy Reverberant Environments,” Proc. APSIPA Annual Summit and Conference, December, 2016.
⑤ Shihono MOCHIZUKI, Sayaka SHIOTA, and Hitoshi KIYA, “Voice liveness detection based on pop-noise detector with phoneme information for speaker verification,” Proc. 5th Joint Meeting of Acoustical Society of America and Acoustical Society of Japan, November, 2016.
⑥ Shihono MOCHIZUKI, Sayaka SHIOTA, and Hitoshi KIYA, “A study of sentence design based on pop-noise balance for voice liveness detection,” Proc. Workshop on Community-centric Systems as Interdisciplinary Study, August, 2016.
他、国内学会22件

６．研究組織
(2)研究協力者
研究協力者氏名：貴家仁志
ローマ字氏名：KIYA, Hitoshi

研究協力者氏名：小野順貴
ローマ字氏名：ONO, Nobutaka