音声の匿名化によるプライバシー保護を可能にする個人用防御音生成法の検討

代表研究者	橋　本　佳	名古屋工業大学　大学院工学研究科　特任助教
共同研究者	山　岸　順　一	国立情報学研究所　コンテンツ科学研究系　准教授
共同研究者	越　前　功	国立情報学研究所　コンテンツ科学研究系　教授

1 はじめに

情報通信基盤の整備とスマートフォンやタブレットなどの携帯端末の普及により，音声や動画を含む様々な情報をインターネット上で容易に共有，収集することが可能となった．また，近年の音声情報処理技術の高度化に伴い，音声認識による音声入力など，音声情報処理技術を利用した様々な有益なサービスが提供されるようになった．Facebook によるスマートフォンを経由した周辺音の収集解析サービス，常時音声入力を受け付けて情報検索や音楽再生を行う Amazon Echo [1]など，広範囲の音声を収集，分析するようなサービスも提供され始めた．しかし，その一方で，携帯デバイス等で収集した様々な情報が本人の意思によらずに SNS（Social Networking Service）などを通してインターネット上で共有されることによるプライバシーの侵害が社会問題となりつつある．これらの許可無く共有された情報が解析されることで，いつどこにいたのかなどの個人情報が流出する恐れがある．さらに，共有された情報から個人が特定されることで，プライバシーの侵害の問題はより深刻なものとなる．

音声情報処理技術の一つである話者認識技術[2]は，音声から発話者を特定する技術であり，音声による気軽な個人認証を実現する技術として注目を集めている．話者認識の性能は著しく向上しており，人間の認識性能よりも高い性能を示すまでになった．話者認識技術は個人認証に利用可能であり，既に実用化が進められている．個人認証の他にも，話者情報を付与した自動議事録作成や，話者情報を対象とした音声データの検索サービスなど，様々な応用が期待されている．しかし，話者認識技術が悪用された場合，個人情報などの重要な情報が流出する危険がある．このため，話者認識技術が悪用されることに対する防御技術が必要である．

本研究では，話者認識技術によって本人の意志によらずに収集された音声から個人が特定されるというプライバシーの侵害の問題に対し，人々が自身の意志でプライバシーを保護することを可能にする音声の匿名化技術を確立することを目的とする．これまでにも，なんらかの音を発生させることで発話内容を他者に認識されないようにする音声に関するプライバシー保護技術が提案されてきている[3][4][5]．これらの手法は，発話内容が周囲の人間に聞き取られることによる情報漏洩やプライバシーの侵害を防ぐことを目的としているが，その目的のため，発生した音によって利用者以外の周囲の人間のコミュニケーションが阻害される恐れがある．そこで本研究では，音声の個人性のみに注目し，人間のコミュニケーションを阻害せずに自動話者照合システムの照合性能を低下させるプライバシー保護音（防御音）を提案する．

2 関連研究

音声から個人を特定する話者認識技術は，評価データベースの整備や米国立標準技術研究所（National Institute of Standards and Technology; NIST）[6]が主催する話者認識タスクの競争型評価ワークショップ NIST Speaker Recognition Evaluation（SRE）が開催されたことに伴い，急速な発展を遂げている．NIST SRE においては，世界各国の多くの研究機関が参加しており，また，音声情報処理分野の代表的な国際会議においても，一定数以上の話者認識に関する研究成果が発表されるなど，話者認識技術に対して強い関心が向けられていることがうかがえる．

一方で，音声や顔画像などのマルチメディア情報から個人が特定されることによるプライバシーの侵害に対し，マルチメディア情報に対する匿名化技術に関する研究が近年注目を集め始めている．欧米の複数の研究機関によるマルチメディア情報に対する匿名化技術に関する研究プロジェクト COST Action IC1206 [7]が立ち上げられ，また，国際会議 MIPRO 2014 [8]においては，マルチメディア情報に対する匿名化技術に関するセッションが設けられた．

音声は発話内容などの言語情報と発話者の性別などの非言語情報の両方を含んでおり，これらを対象とした様々なプライバシー保護技術が提案されてきている．Parthasarathi らは言語情報を保護する技術として新たな音響特徴表現を提案した[9][10]．提案する表現形式は個人を特定することは可能であるが，言語情報は保たれていない．このため，提案する表現形式で音声データを保存することで，個人性は保持しつつ発話内容に含まれる個人情報が保護されたデータが保存できる．山本らは音声に含まれる個人性と発話内容の両方を保護する手法として，音声を含むマルチメディアデータから音声を除去する技術を提案した[11]．これらの技術は音声に含まれる個人情報を保護することは可能であるが，収録済みの音声データに対して適用される技術であるため，第三者が許可無く音声を収集した場合には適用することが困難である．

Jin らは声質変換技術を利用し，音声の声質を変換することで話者認識技術による個人の特定を防ぐ方法を提案した[12][13]．声質は変化するものの音声は自然であり，言語情報は損なわれない．このため，個人を特定されることを防ぎながら発話内容を伝えることが可能となる． しかし，声質が本人の声質から大きく変化するため，人間同士の会話においては不自然さを感じさせる恐れがある．また，あらかじめ何らかのデバイスに音声を入力し変換処理を行う必要があるため，人間同士の会話がデバイスを介して行われることとなり，やはり会話が不自然なものとなる．

特定の音を発生させることによって周囲の人間に発話内容を聞き取られないようにする手法が提案されている．この技術は音声が持つ言語情報を対象としたプライバシー保護技術であり，また，収録前の音声に適用可能な手法である．病院や銀行などでの重要な個人情報を含む会話が周囲の人に聞き取られないようにするために利用可能であり，ヤマハ株式会社のスピーチプライバシーシステム[14]など，既に実用化されている．しかし，周囲の人間に発話内容を聞き取られないような音を発生させることとなるため，周囲の人間の会話を阻害してしまう恐れがある．このため，適切なシチュエーションで利用する必要がある．

3 プライバシー保護音

第三者が本人の許可無く音声などのマルチメディア情報をインターネット上で共有される，さらに，共有された情報から個人が特定されるというプライバシー侵害の問題は今後より深刻なものとなると予想される．近年の自動話者照合システムは人間よりも高い性能を示しており，自動話者照合システムが悪用された場合，音声のみから個人を特定される恐れがあり，プライバシーの侵害は一層深刻なものとなる．このため，本研究では自動話者照合システムによって個人が特定されることを防ぐ防御方法を検討する．本研究では，自動話者照合システムによる個人の特定のみを対象とする．また，本研究では，本人の意志によらず，第三者によって音声が収集されることを想定する．このため，音声が収録された後に何らかの処理を加えることはできないという状況を想定する．

自動話者照合システムの性能を低下させることは，大音量のノイズを発生させる，ボイスチェンジャーなどを用いて声質を大きく変えるなどによって，比較的容易に実現可能である．しかし，大音量のノイズを発生させた場合，プライバシーを保護することは可能かもしれないが，実環境における人間同士の会話において発話内容を聞き取ることが困難になる恐れがある．また，ボイスチェンジャーなどを用いて声質を変えた場合，発話内容を聞き取ることは可能だが，利用者の声質が大きく変わるため会話が不自然なものとなる恐れがある．また，何らかのデバイスが必要となり，普段の会話とは異なる状況を作り出し，不自然なものとなる．

これらの問題を解決するために，本研究では人間のコミュニケーションを阻害することなく自動話者照合システムによって個人が特定されることを防ぐ「プライバシー保護音」を提案する．図 1 にプライバシー保護音の概要を示す．プライバシー保護音は，人間の会話における自然性や発話内容の聞き取りを損なわずに自動話者照合システムの性能を低下させる音であり，会話をする際にプライバシー保護音を発生させることで，第三者が音声を許可無く音声を収集したとしても，自動話者照合システムによって個人が特定されることを防ぐことを可能とする．また，話者性のみに注目することで，周囲の人間の会話を阻害することなく利用することを可能とする．

本研究では，発話内容ではなく音声の個人性のみに注目し，自動話者照合システムの照合性能を低下させることを目的としており，人間の会話を阻害することなく利用することが可能なプライバシー保護技術であるという点で従来研究と大きく異なる．また，収録済みの音声データを対象とした技術ではなく，収録前の実空間の音声を対象とした技術である点が従来研究と大きく異なる．さらに，提案法では，プライバシー保

護音を発生させる以外は自然な会話と同じ状況となり，人間のコミュニケーションを阻害しないという点が利点として挙げられる．

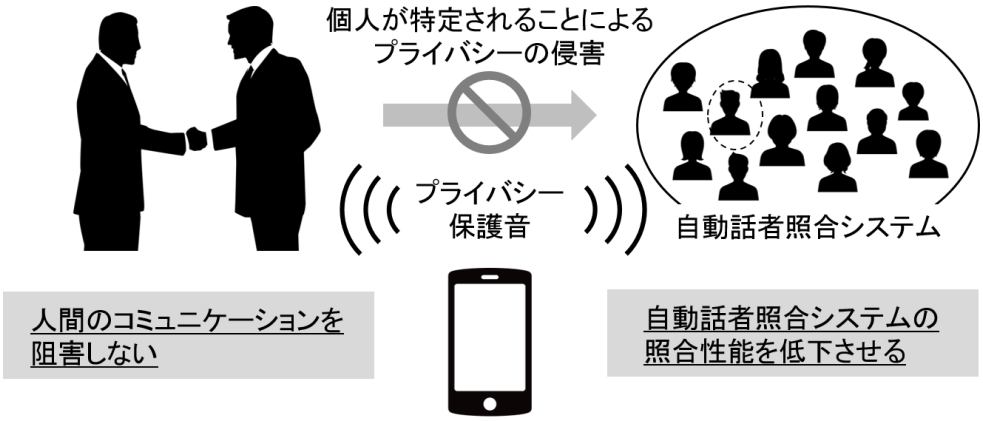

図 1 プライバシー保護音の概要

4 評価実験

人間同士のコミュニケーションを阻害することなく自動話者照合システムの照合性能を低下させるプライバシー保護音の初期的検討として，どのような音がプライバシー保護音として適切であるかを実験的に検討した．本実験では，ホワイトノイズを基本とした様々な音をテストデータに重畳し，その際の自動話者照合システムの照合性能と人間の聞き取り精度に関する客観評価について比較した．ホワイトノイズの他にもピンクノイズや ICRA ノイズ[15]を用いた実験も行ったが,実験結果はノイズの種類によらず近い傾向を示したため，ここではホワイトノイズを用いた実験結果のみを示す．ここでは特に，SNR，周波数がどのように影響を与えるかに注目し検討を行った．

4-1 実験条件

話者照合システムは ALIZE 3.0[16]を用いて構築し，話者照合手法としては GMM-UBM に基づく手法[17]を用いた．また，音声データベースとしては TIMIT データベース[18]を用いた．Universal Background Model (UBM) の学習データとしては，男性 326 人，女性 136 人，計 462 人，各話者 10 発話を用いた．登録話者は UBM の学習データに含まれない男性話者 112 人，女性話者 56 人，計 168 人であり，登録データとして各話者 8 発話，テストデータとして各話者 2 発話を用いた．音声データのサンプリング周波数は 16kHz であり，音声の分析フレームは 25ms，フレームシフトは 10ms とした．また，音響特徴量として MFCC19 次元とエネルギー，および，その 1 次，2 次動的特徴量を用い，計 60 次元の音響特徴量は各発話において平均 0 分散 1 となるように正規化した．

話者照合の評価尺度としては等価エラー率 (Equal Error Rate; EER) を用い，人間の聞き取り精度の評価尺度として客観評価手法である Short-Time Objective Intelligibility (STOI) [19][20] のスコアを用いた.EER は誤棄却率と誤受理率が等しくなるエラー率であり,EER が小さいほど話者照合性能が低いといえる．STOI はクリーン音声と雑音付き音声を入力として，0.0 から 1.0 までのスコアを出力する客観評価手法であり，STOI が大きいほど聞き取り精度が高いことを示す．

4-2 SNR に関する比較実験

SNR の違いによる EER と STOI への影響を調査するために，SNR が 10dB，0dB となるようにホワイトノイズをテストデータに重畳し，EER と STOI を比較する．図 2 に，ホワイトノイズを重畳していない場合，SNR が 10dB，0dB となるようにホワイトノイズを重畳した場合の EER と STOI をそれぞれ示す．図より，SNR が小さくなるにつれて EER は大きく増加していき，STOI は大きく減少していくことがわかる．つまり，SNR が小さくなるにつれて自動話者照合システムの照合性能は低下するが，人間の聞き取り精度も低下するため，実

環境における人間の会話を阻害することとなる．このことから，実環境における人間の会話を阻害せずに話者照合の性能を低下させるためには，大音量のノイズを加えるだけでは適切ではないといえる．

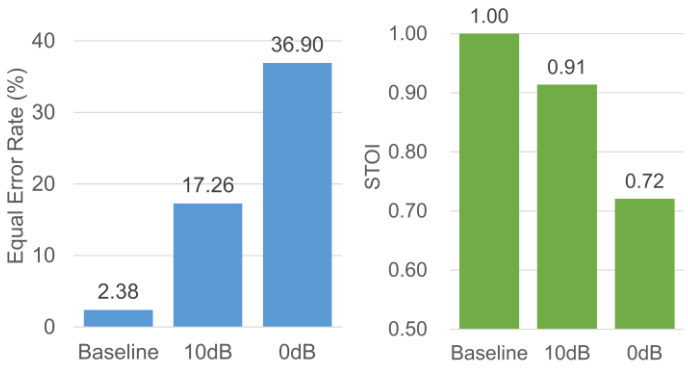

図 2 ホワイトノイズを重畳した場合の EER と STOI

4-3 バンドパスフィルタの周波数帯域に関する比較実験

自動話者照合システムの照合性能と人間の聞き取り精度に対して強く影響を及ぼす周波数を調査するために，ホワイトノイズに対して様々な周波数帯域のバンドパスフィルタを適用し，周波数帯域が制限されたホワイトノイズを用いて EER と STOI を比較する．表 1に本実験で用いたバンドパスフィルタの周波数帯域を示す．表に示した計 35 種類のバンドパスフィルタを適用したホワイトノイズを，それぞれ SNR が 10dB および 0dB となるようにテストデータに重畳し，EER および STOI を評価した．**図 3**，**図 4** に SNR が 10dB，0dB の EER の結果を，**図 5**，**図 6** に SNR が 10dB，0dB の STOI の結果をそれぞれ示す．

図 3，**図 4** より，1kHz 単位のバンドパスフィルタを適用した際の EER の結果に注目すると，Baseline からの EER の変化は大きくないが，SNR によらず，4-5kHz，5-6kHz のバンドパスフィルタを適用した際に，EER が大きくなる．一方で，**図 5**，**図 6** より，1kHz 単位のバンドパスフィルタを適用した際の STOI の結果に注目すると，SNR によらず，0-1kHz のバンドパスフィルタを適用した際に STOI が最も小さくなり，バンドパスフィルタの周波数帯域が高周波数になるにつれて STOI が大きくなる．バンドパスフィルタの周波数帯域幅が異なる場合においても，4-5kHz，5-6kHz を含む場合に EER が大きくなり，0-1kHz を含む場合に STOI が大きくなっている．

バンドパスフィルタの周波数帯域幅の違いに注目すると，周波数帯域幅が広くなるにつれて EER は大きくなる傾向がみられる．このような傾向は SNR が 0dB の際により顕著である．しかし，EER が最大となるのは SNR が 10dB の時には 1-6kHz，SNR が 0dB の時には 0-6kHz とした時であり，5kHz 単位や 6kHz 単位としたときに最も EER を大きくする．STOI については，周波数帯域幅が広くなるにつれて小さくなる傾向がみられるが，周波数帯域幅よりもどの帯域を含むかがより大きく影響を与えている．

これらの結果から，自動話者照合システムの照合性能を表す EER と人間の聞き取り精度を表す STOI において，その性能に最も影響を与える周波数やバンドパスフィルタの周波数帯域幅が異なることがわかる．このような，周波数やバンドパスフィルタの周波数帯域幅が各評価指標に与える影響が異なるという特性を利用することによって，人間の聞き取り精度を大きく損なうことなく，自動話者照合システムの照合性能を低下させることが可能となる．

表 1 実験で用いたバンドパスフィルタ

	バンドパスフィルタの周波数帯域［kHz］
1 kHz 単位	0-1，1-2，2-3，3-4，4-5，5-6，6-7，7-8
2 kHz 単位	0-2，1-3，2-4，3-5，4-6，5-7，6-8
3 kHz 単位	0-3，1-4，2-5，3-6，4-7，5-8
4 kHz 単位	0-4，1-5，2-6，3-7，4-8
5 kHz 単位	0-5，1-6，2-7，3-8
6 kHz 単位	0-6，1-7，2-8
7 kHz 単位	0-7，1-8

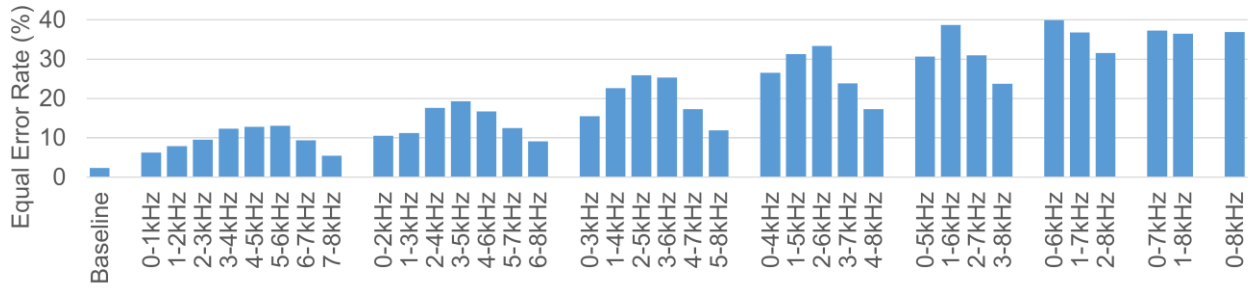

図 3 バンドパスフィルタが適用されたホワイトノイズを用いた場合の EER（SNR: 10 dB）

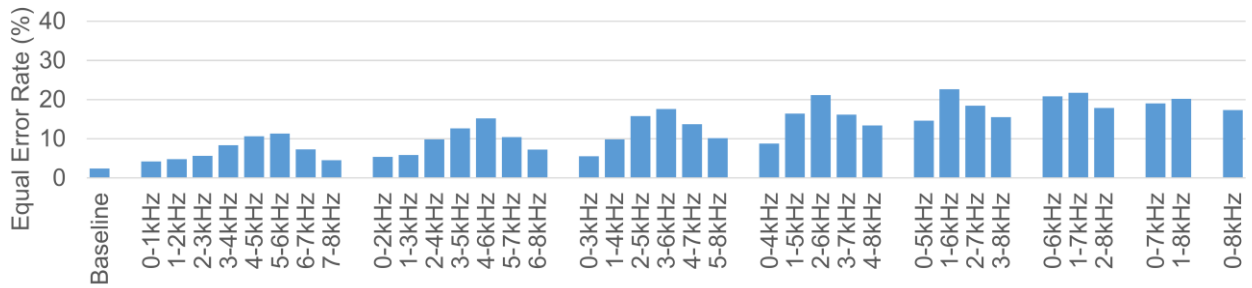

図 4 バンドパスフィルタが適用されたホワイトノイズを用いた場合の EER（SNR: 0 dB）

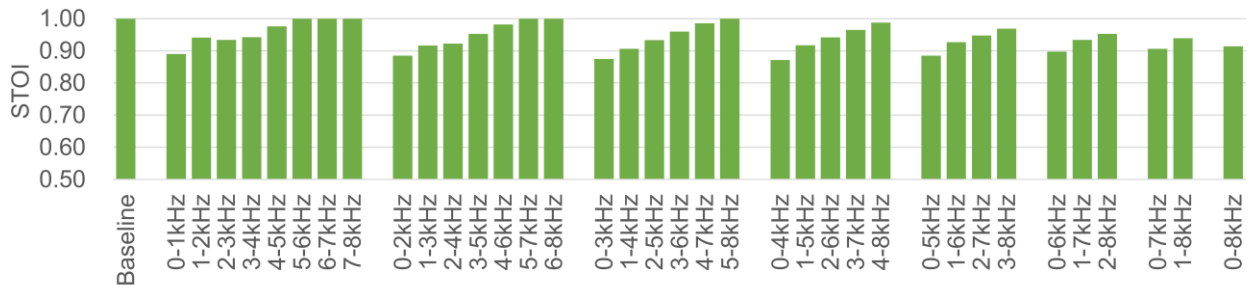

図 5 バンドパスフィルタが適用されたホワイトノイズを用いた場合の STOI（SNR: 10 dB）

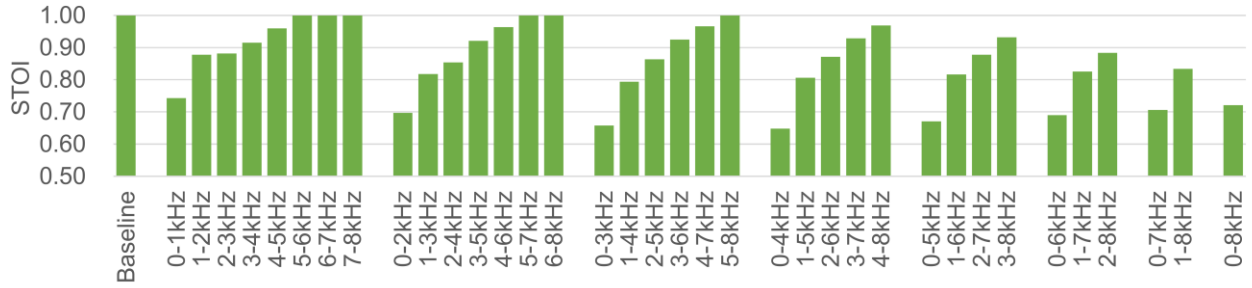

図 6 バンドパスフィルタが適用されたホワイトノイズを用いた場合の STOI（SNR: 0 dB）

4-4 高 STOI における EER の比較

人間の聞き取り精度を大きく損なうことなく，自動話者照合システムの照合性能を低下させるためには，

STOI を大きく低下させることなく，EER を大きくするような音が望ましい．**図 3 – 図 6**より，EER を最も大きくするバンドパスフィルタ 0-6kHz，SNR 0dB という条件では，STOI が 0.69 と大きく低下しており，人間の聞き取り精度を大きく損なってしまっているといえる．そこで，STOI が高い条件に限定し，EER を比較する．

表 **2**，表 **3** に，STOI が 0.93 以上となるバンドパスフィルタに限定した中で EER が最も大きくなる 3 つの条件での EER と STOI を示す．表より，SNR を 10dB，0dB としたときの EER の差は小さく，適切なバンドパスフィルタを利用することによって，SNR が 10dB とした場合においても効率良く EER を大きくすることができる．つまり，小さな音量の音で人間の聞き取り精度を大きく損なうことなく，自動話者照合システムの照合性能を低下させることができるといえる．

表 2 高 STOI での EER の比較（SNR: 10 dB）

	1-7 kHz	2-6 kHz	1-8 kHz
EER	21.73	21.13	20.20
STOI	0.934	0.942	0.940

表 3 高 STOI での EER の比較（SNR: 0 dB）

	3-8 kHz	4-8 kHz	4-7 kHz
EER	23.71	17.26	17.26
STOI	0.932	0.969	0.967

4-5 テスト話者の比較

テスト話者毎のプライバシー保護音の効果を比較するため，各テスト話者の EER を比較した．ここでは，SNR 10dB の条件において最も EER が大きくなる 1-6kHz のバンドパスフィルタを適用したノイズを用いた．話者照合における閾値はテスト話者毎にチューニングした．この結果，テスト話者毎の EER は 0.0-61.1%と幅広く分布しており，プライバシー保護音としての効果がテスト話者に強く依存しているといえる．このため，あらゆる話者の話者照合性能を低下させるためにはテスト話者の声質を考慮したプライバシー保護音を開発する必要があるといえる．

5 UBM に基づくプライバシー保護音

ここまでの検討結果から，話者照合性能と人間の聞き取り精度に強い影響を与えるノイズの周波数帯域が異なることを示し，適切な周波数帯域にノイズを重畳することによって，人間の聞き取り精度への影響を抑えながら話者照合性能を大きく低下させることを示した．しかし，話者によってその効果が大きく異なり，あらゆる話者の話者照合性能を低下させるには至っていない．ここまでの検討では，ノイズとしてホワイトノイズを用いており，人間の音声の音響的特徴を考慮したものではなかった．そこで，人間の音声の音響的特徴を考慮したプライバシー保護音として，UBM（Universal Background Model）に基づくプライバシー保護音を提案する．

近年のガウス混合モデル（Gaussian Mixture Model; GMM）に基づく話者照合システムにおいては，多種多様な話者の音声データから構築される UBM を基盤とし，各話者の特徴をモデル化する手法が広く用いられている[17]．UBM は不特定話者の音声をモデル化したものであり，人間の音声の周波数特徴を適切に表現している．そこで，本研究では UBM に基づいて音声を生成することによって，人間の音声の音響的特徴に近いプライバシー保護音を生成する．

まず，多量の話者の音声データから GMM を学習し，UBM を構築する．プライバシー保護音を生成する際には，学習した UBM に基づいて音響特徴量ベクトルをサンプリングし，生成された音響特徴量ベクトルからプライバシー保護音を生成する．生成されるプライバシー保護音は UBM に従っているため，人間の音声に近い音響的特徴を持つが，サンプリングによって生成されるため言語的情報は含まないノイズが生成される．このため，人間の音声に重畳する際にホワイトノイズよりも強い影響を与えるノイズになることが期待される．

6 評価実験

UBM に基づくプライバシー保護音の有効性を評価するため，ホワイトノイズおよび UBM に基づくプライバシー保護音を重畳した際の話者照合性能と人間の聞き取り精度を比較評価した．

6-1 実験条件

話者照合手法としては GMM-UBM に基づく手法[17]を用いた．また，音声データベースとしては TIMIT データベース[18]を用いた．UBM の学習データとしては，男性 326 人，女性 136 人，計 462 人，各話者 10 発話を用いた．登録話者は UBM の学習データに含まれない男性話者 112 人，女性話者 56 人，計 168 人であり，登録データとして各話者 8 発話，テストデータとして各話者 2 発話を用いた．音声データのサンプリング周波数は 16kHz であり,音声の分析フレームは 25ms,フレームシフトは 10ms とした.また,音響特徴量として MFCC19 次元とエネルギー，および，その 1 次，2 次動的特徴量を用いた．プライバシー保護音の生成に利用する UBM は，話者照合システム構築時に用いた音声データから学習した．音声の分析フレームは 25ms，フレームシフトは 5ms とし，音響特徴量は 24 次メルケプストラムを用いた．GMM の混合数は 256 とした．

話者照合の評価尺度としては当該話者モデルを用いた際の対数尤度比（Log Likelihood Ratio; LLR）を用い，人間の聞き取り精度の評価尺度として Short-Time Objective Intelligibility（STOI）[19][20]を用いた．LLR は話者照合時に用いるスコアであり，当該話者モデルを用いた際の LLR が小さくなるほど本人として照合されにくくなる．STOI はクリーン音声と雑音付き音声を入力として，0.0 から 1.0 までのスコアを出力する聞き取り精度の客観評価手法であり，STOI のスコアが大きいほど聞き取り精度が高いことを示す．

本実験ではホワイトノイズ(White)，UBM 用の音声データの平均を用いて作成したノイズ(Mean)，UBM に基づくノイズ(UBM)，および UBM と White の混合ノイズ(UBM+White)，UBM と Mean の混合ノイズ(UBM+Mean)について評価する．混合ノイズは UBM に対して SNR が 10dB となるように White と Mean を混合した．ノイズは周波数帯域を制限して重畳することで話者照合性能が大きく低下することが示されている．そこで，本実験では周波数帯域を 2-7kHz に制限するバンドパスフィルタを適用したノイズを用いた．テストデータに対して SNR が 10dB となるようにバンドパスフィルタを適用したノイズを重畳した．

6-2 実験結果

図 **7** に LLR と STOI をそれぞれ示す．ただし，クリーン環境における LLR は 0.995，STOI は 1.0 である．まず，White，Mean，UBM に注目すると，Mean と UBM は White よりも STOI が小さいことわかる．これは，Mean と UBM は人間の音声に近い周波数の特徴を持つため，低周波数に大きいパワーを持ち，聞き取り精度に対する影響が大きかったためと考えられる．一方，Mean は White よりも LLR が大きく，話者照合性能に対する影響は White よりも小さかった．これに対し，UBM は White よりも LLR が小さく，話者照合性能を White よりも低下させている．この結果から，音声データの平均を用いるだけでは十分ではなく，適切に音声データをモデル化した UBM を用いてサンプリングすることによって，テストデータに含まれる話者性により強い影響を与えることができるといえる．

次に，White，Mean，UBM と混合ノイズ UBM+White と UBM+Mean を比較すると，混合ノイズは LLR を UBM からさらに低下させており，UBM よりも話者照合性能を低下させることを示した．全周波数に一定の影響を与える White や Mean のようなノイズと周波数に対して非定常な影響を与える UBM のように異なる特性を持つノイズを混合することで，話者照合性能をより低下させることを可能にした．また，混合ノイズの STOI は UBM の STOI よりも高い値を示しており，混合前からの劣化はみられなかった．このため，混合ノイズは人間の聞き取り精度への影響を抑えながら話者照合性能を低下させたといえる．

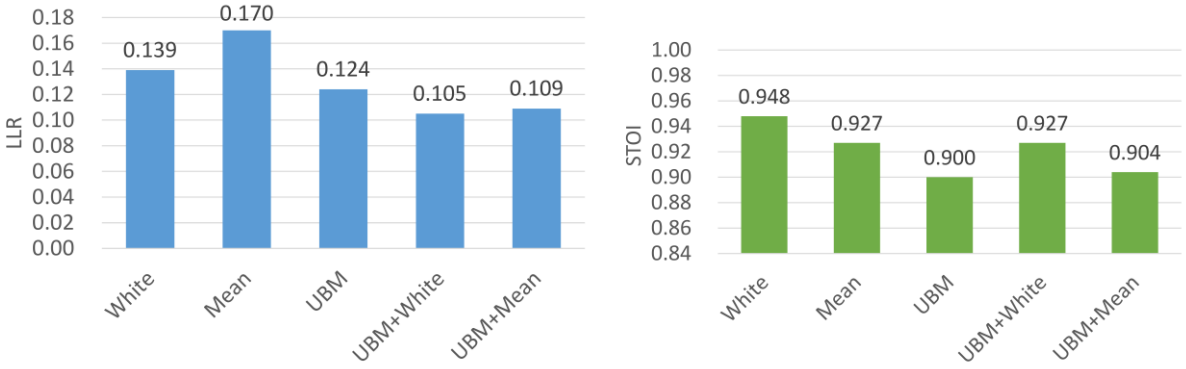

図 7 Log Likelihood Ratio (LLR) (左), Short-Time Objective Intelligibility (STOI) (右)

6 まとめ

本研究では，人間のコミュニケーションを阻害することなく自動話者照合システムの照合性能を低下させるプライバシー保護音の検討を行った．本人の意思によらずに収集された音声データがインターネット上で共有され，さらに，話者照合技術が悪用され音声から個人が特定されるというプライバシーの侵害の問題に対し，人間のコミュニケーションを阻害することなくプライバシーを保護することが可能な技術としてプライバシー保護音を提案した．プライバシー保護音は，人間のコミュニケーションの質を低下させることなくプライバシーを保護することを目的としており，本研究では，どのような音が人間の聞き取り精度と自動話者照合システムの照合性能に影響を与えるかを評価することで，プライバシー保護音として適切な音について検討した．実験結果から，自動話者照合システムの性能を表す EER に最も影響を与える周波数は 4-5kHz，5-6kHz 周辺であり， 1-7kHz などの周波数帯域を制限した音を重畳することによって EER を増加させ，自動話者照合システムの性能を低下させることを示した．また，人間の聞き取り精度を表す客観評価尺度である STOI に最も影響を与える周波数は 0-1kHz などの低い周波数であり，高周波数であるほど STOI に与える影響は小さいことを示した．これらの結果から，周波数によって人間の発話内容の聞き取り精度に与える影響と自動話者照合システムの照合性能に与える影響が異なり，この違いを考慮することによって，人間の聞き取り精度を損なうことなく自動話者照合システムの性能を低下させることが可能であることが示された．さらに，人間の音声の音響的特徴を考慮した UBM に基づくプライバシー保護音を提案した．UBM は不特定話者の音声をモデル化したものであり，人間の音声の周波数特徴を適切に表現している．UBM に従って音響的特徴をサンプリングしているため，人間の音声に近い音響的特徴を持つが，サンプリングによって生成されるため言語的情報は含まないノイズが生成される．実験結果から，UBM に基づくプライバシー保護音はホワイトノイズから話者照合性能をさらに低下させることを示した．また，ホワイトノイズのような全周波数に一定の影響を与えるノイズと UBM に基づくノイズのような周波数に対して非定常な影響を与えるノイズを混合することでさらに話者照合性能を低下させた．

本研究では，どのような条件でどの程度の効果が得られるかということが，実験的に示されたのみであり，実用化という観点からはさらなる改善が必要である．今回の実験では人間の聞き取り精度は客観評価による評価のみであったが，今後は実際の利用状況を考慮した主観評価実験による聞き取り精度や不快度の評価を行う必要がある．聞き取り精度の主観評価実験には，人間の単語予測による影響を軽減するために意味を持たない文（無意味文）を用いた評価を行う必要があり，準備を進めている．また，本研究では UBM を用いることで人間の音声の音響的特徴を考慮したプライバシー保護音を生成したが，さらに利用者の詳細な声質を考慮することで効率良く照合性能を低下させることが期待される．既に幾つかの利用者の声質を考慮したプライバシー保護音生成法について検討を進めている．

【参考文献】

1. Amazon Echo, http://www.amazon.com/Amazon-SK705DI-Echo/dp/B00X4WHP5E
2. 越仲孝文，篠田浩一，“話者認識の国際動向，” 日本音響学会誌, vol.69, no.7, pp.342-348, 2013.

3. 佐藤洋，清水寧，“スピーチプライバシー研究の歴史と近年の動向，” 日本音響学会誌，vol.64，no.8，pp.475-480, 2008.
4. 藤原舞，山川高史，秦雅人，清水寧，“薬局におけるサウンドマスキング評価方法の実験的検討，” 日本音響学会 2011 年秋季研究発表会講演論文集, pp.1127-1130, 2011.
5. 赤木正人，入江佳洋，“音情景理解を応用した音声プライバシー保護，” 信学技報, pp.19-24, 2011.
6. “NIST Speaker Recognition Evaluation,” http://www.nist.gov/itl/iad/mig/sre.cfm
7. “COST Action IC1206,” http://costic1206.uvigo.es/
8. “MIPRO 2014,” http://www.mipro.hr/MIPRO2014.BiForD/ELink.aspx
9. S.H.K. Parthasarathi, M. Magimai-Doss, H. Bourlard, and D. Gatica-Perez, “Evaluating the robustness of privacy-sensitive audio features for speech detection in personal audio log scenarios,” Proceedings of ICASSP 2010, pp.4474-4477, 2010.
10. S.H.K. Parthasarathi, H. Bourlard, and D. Gatica-Perez, “Wordless sounds: robust speaker diarization using privacy-preserving audio representations,” IEEE Transactions on Audio, Speech, and Language Processing, vol.21, no.1, pp.85-98, 2013.
11. K. Yamamoto, M. Tsuchiya, and S. Nakagawa, “Privacy protection for speech signals,” Procedia Social and Behavioral Science, vol.2, no.1, pp.153-160, 2010.
12. Q. Jin, A.R. Toth, T. Schultz, and A.W. Black, “Speaker de-identification by voice transformation,” Proceedings of ICASSP 2009, pp.3909-3912, 2009.
13. Q. Jin, A.R. Toth, T. Schultz, and A.W. Black, “Speaker de-identification via voice transformation,” Proceedings of ASRU 2009, pp.529-533, 2009.
14. “ヤマハスピーチプライバシーシステム,” http://www.yamaha.co.jp/acoust/speechprivacy/
15. W.A. Dreschler, “ICRA Noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment,” Audiology, vol.40, pp.148-157, 2001.
16. A. Larcher, “ALIZE 3.0 -Open source toolkit for state-of-the-art speaker recognition,” Proceedings of Interspeech 2013, pp.2768-2772, 2013.
17. D. Reynolds, T.F. Quatieri, and R.B. Dunn, “Speaker verification using adapted Gaussian mixture models," Digital Signal Processing, vol.10, pp.19-41, 2000.
18. J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, “TIMIT acoustic-phonetic continuous speech corpus,” https://catalog.ldc.upenn.edu/LDC93S1
19. C.H. Taal, R.C. Hendriks, R. Heusdens, and J. Jensen, “A short-time objective intelligibility measure for time-frequency weighted noisy speech,” Proceedings of ICASSP 2010, pp.4214-4217, 2010.
20. C.H. Taal, R.C. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” IEEE Transactions on Audio, Speech, and Language Processing, vol.19, no.7, pp.2125-2136, 2011.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Privacy-preserving sound to degrade automatic speaker verification performance	ICASSP 2016	2016 年 3 月
話者照合性能を低下させる UBM に基づくプライバシー保護音の検討	日本音響学会 2016 年春季研究発表会	2016 年 3 月
自動話者照合システムの性能を低下させるプライバシー保護音の検討	日本音響学会 2015 年秋季研究発表会	2015 年 9 月
自動話者照合システムの性能を低下させるプライバシープリザービングサウンドの検討	電子情報通信学会技術研究報告	2015 年 7 月