Vol.2020-HCI-190 No.21
Vol.2020-UBI-68 No.21
2020/12/9

鼻に対するアクティブ音響センシングを用いた個人識別手法の提案

磯部 海斗[1]　村尾 和哉[1,2]

概要：スマートグラスや VR ゴーグルが普及しつつあるが，多くの製品では電源 ON の後にユーザ認証を経ないままホーム画面が表示される．またデバイス使用中にアプリのログイン時や決済時にパスワードなどの入力を求められることがあるが，付属のタッチパッドを用いるか，手持ちのコントローラで画面内のソフトウェアキーボードから入力することが一般的である．このほかディスプレイを搭載しない眼鏡型デバイスもあるが，本人認証は行われないことが一般的である．本研究では，鼻当て部分を伴う眼鏡型デバイス使用時を想定して，鼻当て部分にマイクとスピーカを搭載して，アクティブ音響センシング技術を適用させることで，自動的かつ定期的に個人識別を行う手法を提案する．提案手法では，眼鏡型デバイスの鼻当て部分のスピーカから音響信号を流し，接触している鼻を通過した音響信号をマイクから取得して，その音響信号の周波数特性を解析することで個人を認識する．提案手法は，眼鏡型デバイスの鼻当て部分は認識したいユーザの鼻と常時自然に接触することと，ユーザの鼻の形状が固有の音響特性をもつことを利用している．本研究では，マイクとスピーカを搭載した眼鏡型デバイスのプロトタイプを実装し，提案手法の有効性を評価する実験を 3 人の被験者で行った．その結果，F 値 0.86〜1.00 の精度で個人を認識できることを確認した．

1. はじめに

スマートフォンやタブレット，ノート PC では顔認証や指紋認証などの生体認証を搭載した製品が多く登場している．生体認証は端末ログイン時やアプリのログイン時，決済時などのさまざまな場面で用いられ，パスワード認証と比較して素早く認証処理を行うことができる．また，Google Glass や Oculus Quest 2 のようなスマートグラス・ヘッドマウントディスプレイ（以下，眼鏡型デバイス）が普及しつつある．しかし，それらの端末の多くは眼鏡型であるという形状の制約からキーボードは搭載されておらず，簡易なタッチパネルか手持ちデバイスによってソフトウェアキーボードで入力が一般的である．眼鏡型デバイスの多くは電源 ON の後にユーザ認証を経ないままホーム画面が表示され，アプリのログイン時や決済時などに他の端末と連携するか，画面内に表示されるソフトウェアキーボードを使用してパスワード入力を求められることがある．眼鏡型デバイスは基本的に共用することは少ないが，悪意のある第三者にデバイスを入手されると個人情報やプライバシに関する情報を閲覧される可能性がある．今後普及していくと考えられる眼鏡型デバイスでは使用開始時および使用中定期的にユーザの負担なく認証する手段が求められる．

本研究では眼鏡型デバイス装着者を識別する手法として，装着者の鼻に対してアクティブ音響センシング技術を適用させることで個人を識別する手法を提案する．アクティブ音響センシング技術とは物体がもつ固有の音響特性を利用した物体認識手法であり，スピーカを用いて特定の音響信号を物体に伝搬させ，マイクで得たその応答の周波数特性を解析することで，物体の種類や状態を認識するものである．提案手法では，眼鏡型デバイスの鼻当て部分にスピーカとマイクを取り付け，スピーカから音響信号を流し，装着者の鼻を伝搬した振動を鼻当て部分のマイクから取得し，その応答の周波数特性を解析することで，装着者を認識する．このように，眼鏡型デバイスの鼻当て部分は認識したいユーザの鼻と常時自然に接触することと，ユーザの鼻の形状が固有の音響特性をもつことを利用している．

2. 関連研究

本節では，生体情報を用いた個人識別に関する研究および音響センシングを用いた物体認識に関する研究を紹介する．

2.1 生体情報を用いた個人識別

生体情報を用いた個人識別手法は数多く存在する．Bar-

[1] 立命館大学情報理工学部
[2] 科学技術振興機構さきがけ

ⓒ 2020 Information Processing Society of Japan

ros Barbosa らは爪表面を利用した生体認証を提案している [1]．爪の生え変わりを利用することにより，廃棄可能性を満たした認証方式となっており，通常のカメラで撮影した爪画像をそのまま認証に利用している．この認証方法は眼鏡型デバイスにカメラが搭載されている場合は行うことができるが，カメラに爪を接近させる必要があり不自然な動きとなる．また，カメラに対して正しい距離と位置で接近させる必要があるという問題点もある．Stefan らは頭蓋骨の骨伝導を利用した生体認証システムを行った [2]．人の頭部の構造である頭蓋骨，軟骨，流体，および組成がユーザごとに異なるため，音の伝わり方が異なることを利用している．頭部の側面にホワイトノイズを骨伝導スピーカから再生し，マイクで信号を受け取っている．頭蓋骨特性周波数応答とメル周波数ケプストラルの組合せを用いて，個人を 97%の精度で識別することに成功した．ユーザの顔を登録し，カメラによってユーザそれぞれの顔を識別する個人認証 [3] もあるが，眼鏡型デバイスでは顔全体をカメラで写すことはデバイスの構造上困難である．白川らは他の生体認証と比べ非常に高い認証精度を持つ虹彩認証に対し，目の周辺の分割画像を利用した認証と虹彩認証と統合し，認証精度が向上し，有効性を確認した [4]．虹彩認証を搭載するときの問題点として，カメラを搭載するコストやカメラによるユーザの心理的負担が大きくなる問題点がある．これらに対して提案手法は，対象場面においてユーザが必ず接触することを利用し，かつ人の非可聴域の音の出力を行っているため，心理的負担を生じさせることもない．

2.2 音響センシングを用いた認識手法

ある動作などを行うことによって発生した音を取得して分析，解析を行うパッシブ音響センシングを用いた技術はこれまでに多く提案されている．Brian らの研究では，指先の動作を行った際に生成する骨伝導音を用いることで，指先のジェスチャを認識している [5]．また，Chris らは質感ある素材の表面に指の爪をドラッグしたときに発生するサウンドを利用して，ジェスチャ入力表面を作成した [6]．Chris らはほかにも，身体の機械的振動を分析することにより，腕と手の指のタップ位置の推定を可能とした [7]．Roderick らは，触覚入力によって制御される手持ち式の対話型デバイスを用いて，これらの表面をなでたり引っ掻いたりすることによって発生する振動信号を分類することができる．これは，相互作用を制御するための機器として非常に安価に製造することができる [8]．このように音響センシングでは，ユーザと物体のインタラクションに多用な可能性を見出すことができる．

スピーカとマイクを使用したアクティブ音響センシングを用いた研究は以前にも提案されている．大野らはアクティブ音響センシングを用いたタッチ認識技術を提案している [9][10]．これは，既存物体にコンタクトマイクとコンタクトスピーカを取り付け，物体の触り方の違いを取得することで把持状態を認識する手法である．河田らは，把持状態の認識を，物体ではなくユーザにセンサを取り付けることにより行った [11]．また，Adiyan らは低周波の超音波で皮膚表面を共振させ，身体上の別地点で受け取ことにより，身体上の接触位置やジェスチャを推測する手法を提案している [12]．位置推定にもアクティブ音響センシングは用いられている．例えば，Kim らは，携帯電話に内蔵されているスピーカの音と振動モーターからの振動を利用して，加速度センサとマイクを用いて，応答を分析することで携帯電話の位置を把握する手法を提案している [13]．また，岩瀬らは，日常生活を支援するために，実空間内にある物体情報を取得する手法を述べている [14]．このようにアクティブ音響センシングを用いた認識では，さまざまな情報の取得を可能とする．本研究では，アクティブ音響センシングを眼鏡の鼻当て部分に適用して装着者を認識する手法を提案する．

3. 提案手法

提案手法は，眼鏡型デバイス装着中に個人を識別する場面において，鼻にアクティブ音響センシング技術を適用することで個人を識別する．提案手法は，ユーザが装着する眼鏡型デバイスの鼻当て部分に小型の圧電素子をスピーカおよびマイクとして取り付け，スピーカから音響信号を流し，鼻の振動・応答をマイクで取得し，その応答の周波数特性を解析することで個人を識別する．

3.1 想定環境

本研究では，鼻当て部分を有して装着中は鼻当て部分が装着者の鼻に接触し続ける眼鏡型デバイスの使用を想定する．ユーザがそのデバイスを装着すると鼻当て部分が装着者の鼻に接触することで個人を識別する．そのため，デバイスを装着すると自動的に認証が行われ，その後デバイス使用中も定期的に個人識別が行われるため，アプリのログインや web サービス利用時などユーザ個人の識別が必要な場面においても，ユーザは認証のための操作を行う必要はない．ユーザは特別な操作を行うことなく，デバイスを装着しているだけで個人を識別できる．指紋認証のように識別子（指）を読み取り機に読み取らせる必要や，音声認識のように発話する必要もない．

なお，本稿では事前に複数のユーザからデバイス装着時のデータを収集して，ユーザごとにラベルを付与したデータを用いて機械学習モデルを構築し，入力データのユーザが事前に学習したユーザ群のうちの誰であるかを推定する．

3.2 提案手法の概要

提案手法の概要を**図 1** に示す．眼鏡デバイスの左右の鼻当て部分に圧電素子 1 個ずつ組み込み，一方をスピー

ⓒ 2020 Information Processing Society of Japan

Vol.2020-HCI-190 No.21
Vol.2020-UBI-68 No.21
2020/12/9

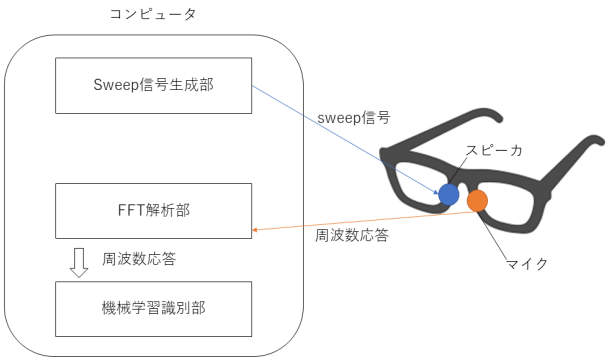


図 1 提案手法の概要

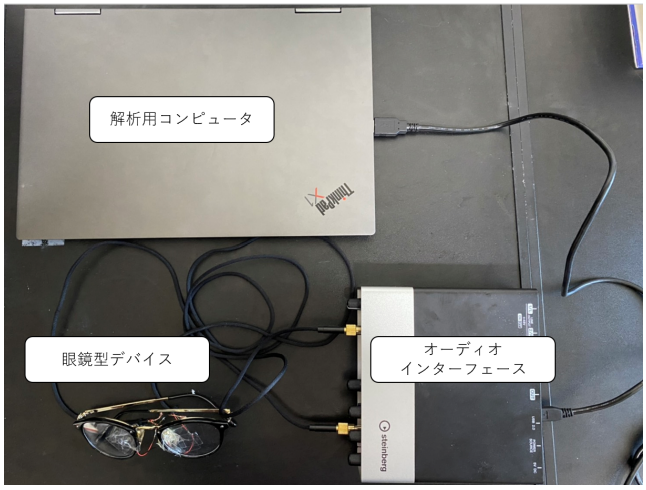


図 2 プロトタイプシステム全体図

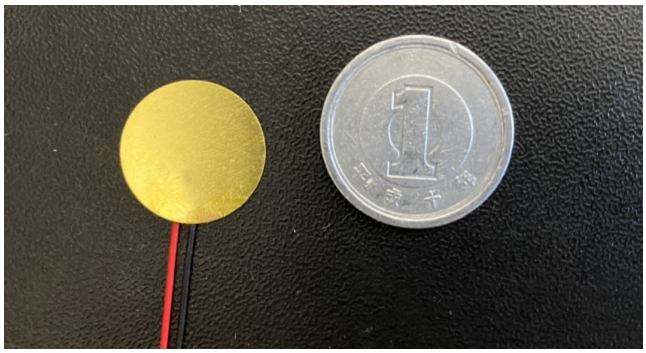


図 3 使用した圧電素子

カ，もう一方をマイクとして使用する．Sweep 信号生成部では，Sweep 信号（Chirp 信号とも呼ばれる）と呼ばれる，周波数が一定時間内に線形に増加する正弦波の繰り返し波形をスピーカから流し，鼻を通過した応答信号をマイクで音データとして取得する．取得した音データを FFT 解析部において FFT（Fast Fourier Transform：高速フーリエ変換）し，周波数パワースペクトルを取得する．機械学習識別部では事前に複数のユーザから収集した周波数パワースペクトルを用いて識別モデルを構築しておき，入力された周波数スペクトルが誰のデータであるかを認識する．

3.3 Sweep 信号生成部

Sweep 信号生成部分では，0.2 秒のあいだに 20,001Hz から 40,000Hz まで直線的に周波数が変化する正弦波を Sweep 信号として生成する．この周波数帯域は，先行研究 [9] の論文を参考に決定した．人の可聴域は年齢にもよるが，おおよそ 10Hz から 20,000Hz までであるため，スピーカから発せられる音は人間には聞こえない．生成した Sweep 信号のサンプリング周波数は 44.1kHz である．

3.4 FFT 解析部

サンプリング周波数 44.1kHz，量子化ビット数 16bit でマイクから音データを取得する．長さ 1 秒のウィンドウサイズで切り出した波形に対して FFT を行い，周波数パワースペクトルを得る．得られたパワースペクトルのうち，20,001Hz から 40,000Hz の領域の 20,000 点のデータを抽出して特徴量ベクトルとする．

3.5 機械学習識別部

最後に機械学習部分では，得られた 20,000 次元の特徴量ベクトルを入力として Randam Forest 用いて個人識別を行う．事前に複数人から取得された特徴量ベクトルにユーザ名をクラスラベルとして付与しておき，Random Forest を学習しておく．

3.6 ハードウェア

プロトタイプシステムは 2 つの圧電素子を搭載した眼鏡および信号処理と機械学習を行うコンピュータによって構成される．プロトタイプシステムの構成を**図 2** に示す．Sweep 信号の生成，FFT 解析，機械学習を行うコンピュータとして，Lenovo ThinkPad X1 Carbon (Intel Core i7-10510U, RAM:16GB, OS Windows 10) を使用した．音の入出力は USB 接続のオーディオインタフェース Steinberg 社の UR22mk を介して行う．

使用した圧電素子は uxcell 社製で直径 15mm，厚さ 0.mm であり，写真を**図 3** に示す．この圧電素子を 2 枚使用して，一般的な眼鏡に取り付けた様子を**図 4** に示す．左側の鼻当てにマイクとして使用する圧電素子を，右側の鼻当てにスピーカとして使用する圧電素子を取り付けた．圧電素子は鼻当てにセロハンテープで固定して取り付けた．実装した眼鏡を装着した様子を**図 5** に示す．有線接続のためケーブルが眼鏡の下から出ているが，配線方法については今後の課題とする．

4. 評価実験

提案手法の有効性を評価するために，プロトタイプデバイスを被験者に装着してもらい個人識別精度を取得する実験を行った．

ⓒ 2020 Information Processing Society of Japan

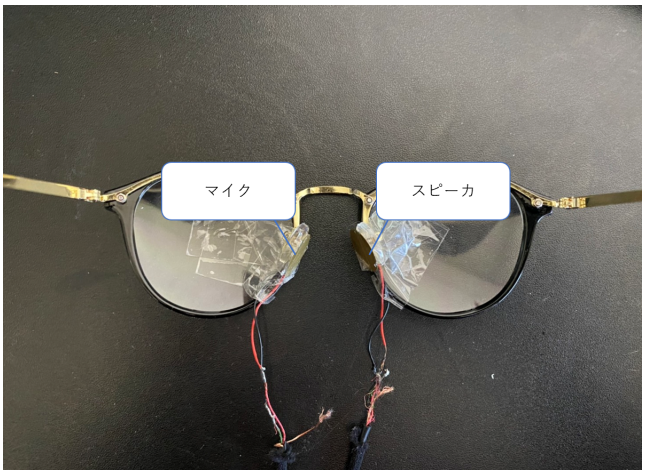


図 4 センサが取り付けた眼鏡



図 5 実装した眼鏡型デバイスを装着した様子

4.1 実験環境

被験者 3 人（20 代，男性，A∼C）にプロトタイプデバイスを装着してもらい，以下の手順でデータを採取した．被験者はデバイスを自然な状態で装着し，10 秒間装着する試行を 1 セッションとし，各被験者ごとに 10 セッション行った．各セッション間では，デバイスを一度取り外してもらい，次のセッションに移行するときに再度装着してもった．そのため，セッションごとに眼鏡の鼻当て部分と被験者の鼻の接触状態は変化している．1 セッション 10 秒のデータに対して FFT を行い，800 サンプル（10 セッションで 8,000 サンプル）のパワースペクトルを取得した．機械学習の分類器には Randam Forest を使用し，leave-one-session-out の 10 分割交差検証によって個人識別の精度を評価した．

表 1 各被験者の認識結果

入力＼出力	A	B	C	適合率	再現率	F 値
A	647	2	151	0.915	0.809	0.859
B	0	798	2	0.994	0.998	0.996
C	60	3	737	0.828	0.921	0.872
平均				0.912	0.909	0.909

4.2 結果と考察

個人識別の結果を**表 1** に示す．被験者 A，B，C の F 値はそれぞれ 0.859，0.996，0.872 となり，平均 0.909 となった．結果より，各被験者を平均 90%の精度で識別することができ，提案手法が個人識別を行うために有効であることを確認した．

セッションごとにデバイスの取り外しと装着を行うことで鼻当て部分と鼻の接触に変化が生じるが，高い精度で個人を識別できたことから，人の鼻がもつ音響特性は個人の識別能力を有すると考えられる．今回の実験では 1 種類の眼鏡のみを用いたが，他の異なる素材の眼鏡でも提案手法は有効に働くのではないかと考える．今回は眼鏡型デバイスの装着の場面を想定して実験を行ったが，VR ゴーグルのような形状でも個人を識別できる可能性があると考える．詳細な評価は今後の課題とする．また，実験で被験者に自然に静止した状態でデータを採取したが，日常的な動作や会話，作業をしながらの状態における性能も今後評価したい．

5. おわりに

本研究では，眼鏡型デバイスのための個人識別手法として，ユーザの鼻に対してアクティブ音響センシング技術を適用することで，個人識別を行う手法を提案した．提案手法は，眼鏡型デバイスの鼻当て部分にスピーカとマイクを取り付け，デバイス装着中は常時ユーザの鼻が鼻当てに接触していることを利用し，鼻を伝搬した音響信号を取得して，その音響信号の周波数特性を解析することで，個人を認識した．マイクとスピーカを搭載したプロトタイプデバイスを実装した．そして，提案手法の有効性を評価する実験を 3 人の被験者に対して行い，平均 F 値 0.9 以上の精度で個人を認識できることを確認した．今後は，今回の実験で使用した眼鏡とは異なる形状および材質の眼鏡を用いた場合の評価や，VR ゴーグルを用いた場合の評価，様々な作業を行っている環境下での評価を実施する予定である．

参考文献

[1] Igor Barros Barbosa, Theoharis Theoharis, Ali E. Abdallah, On the Use of Fingernail Images as Transient Biometric Identifiers, Machine Vision and Applications, Vol. 27, Issue 1, pp. 65–76 (2016).
[2] Stefan Schneegass, Youssef Oualil, Andreas Bulling, SkullConduct: Biometric User Identification on Eyewear

ⓒ 2020 Information Processing Society of Japan

Computers Using Bone Conduction Through the Skull, In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16) (2016).
[3] 佐藤俊雄, 顔による個人認証, 生体医工学 (2006).
[4] 白川功治 吉浦 裕, 市野将嗣, 虹彩認証および目の周辺の分割画像を用いた個人認証, 情報処理学会論文誌 (2018).
[5] Brian Amento, Will Hill, Loren Terveen, The sound of one hand: a wrist-mounted bio-acoustic fingertip gesture interface. In Proceedings of the 2002 ACM SIGCHI International Conference on Human Factors in Computing Systems (CHI' 02), pp. 724–725 (2002).
[6] Chris Harrison, Scott E. Hudson. Scratch Input, Creating Large, Inexpensive, Unpowered and Mobile Finger Input Surfaces. In Proceedings of the 21th annual ACM symposium on User interface software and technology (UIST' 08), pp. 205-–208 (2008).
[7] Chris Harrison, Desney Tan, Dan Morris, Skinput: appropriating the body as an input surface. In Proceedings of the 2010 ACM SIGCHI International Conference on Human Factors in Computing Systems (CHI' 1), pp. 453–462 (2010).
[8] Roderick Murray-Smith, John Williamson, Stephen Hughes, Torben Quaade, Stane: synthesized surfaces for tactile input. In Proceedings of the 2008 ACM SIGCHI International Conference on Human Factors in Computing Systems (CHI' 08), pp. 1299–1302 (2008).
[9] 大野 誠, 志築文太郎, 田中二郎, アクティブ音響センシングを用いた把持状態認識, 一般社団法人情報処理学会インタラクション' 13, pp. 5663 (2013).
[10] Makoto Ono, Buntarou Shizuki, Jiro Tanaka, Touch and Activate: Adding Interactivity to Existing Objects using Active Acoustic Sensing. In Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST' 13), pp. 31-–40 (2013).
[11] 河田大史, 寺田 努, 塚本昌彦, 振動のアクティブセンシングによるウェアラブル所持物体認識システム, 情報処理学会マルチメディア, 分散, 協調とモバイルシンポジウム DICOMO' 15 論文集, pp. 400-–406 (2015).
[12] Adiyan Mujibiya, Xiang Cao, Desney S. Tan, Dan Morris, Shwetak N. Patel, Jun Rekimoto, The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces (ITS' 13), pp. 189-–198 (2013).
[13] Kai Kunze, Paul Lukowicz, Symbolic Object Localization Through Active Sampling of Acceleration and Sound Signatures. In proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp' 07), pp. 163-–180 (2007).
[14] 岩瀬大輝, 伊藤雄一, 秦 秀彦, 山下真由, 尾上孝雄, アクティブ音響センシングによる日常物体識別と位置推定, 一般社団法人情報処理学会インタラクション' 18, pp. 6271 (2018).

ⓒ 2020 Information Processing Society of Japan