人と機械による協調型エコーロケーションの提案

渡邉 拓貴[1,a)]　角谷 美和[2,b)]　寺田 努[3,c)]

概要：エコーロケーションとは音を用いて周囲の状況を認識する技術であり，一部の視覚障害者は舌打ち音等の可聴音によりこれを実現できる．エコーロケーションにおいて超音波を用いると，より詳細な情報の取得等の利点があることが先行研究より報告されているが，通常の人間の能力では，超音波の発信/取得は困難である．本研究では，ウェアラブルコンピューティング環境であれば，デバイスにより超音波の発信と反射音の可聴化が可能であることに着目した．さらに，超音波可聴化のためにデバイス内で信号を処理するため，可聴化と同時に機械学習でも認識が可能である．従って本研究では，超音波を可聴化した音を用いた人による認識と，機械学習による認識を組み合わせた，協調型エコーロケーションを提案する．具体的には，スピーカから超音波信号を発信し，反射音をマイクで検出する．人は，低周波に可聴化された反射波の音色で物体の違いを認識する．機械は，反射波から得られる音響特徴量により認識を行い，ユーザの指定した物体を認識した際に振動で物体の存在を通知する．提案手法により，ユーザは自身の耳で得られる情報と，機械により得られる情報を組み合わせた物体の探索が可能になる．協調型エコーロケーションの実現に向けて，本稿ではエコーロケーションにおける機械学習の有効性について調査した．プロトタイプを作製し評価を行った結果，機械学習による物体の認識精度は，6 種類の物体に対して平均 92.5% であった．また．機械学習の有無による物体の探索行動の変化を 10 人の被験者に対して評価した結果，機械学習により探索時間が平均 71.5 秒から 45.9 秒に減少した．さらに，機械学習を加えることで，エコーロケーション時の精神的負荷が有意に減少した．

1. はじめに

視覚障害者の数は全世界で 22 億人と言われており，彼らへの支援環境が必要である [1]．視覚障害者を支援するための設備として，すでに街中には至る所に誘導ブロックや点字が配置されているが，これらの手がかりは触るまでその場所を知ることはできず，改善の余地があるといえる．近年の機器の発展により，小型化したコンピュータを常に持ち運んだり，身につけたりする環境が整ってきたことで，IT 技術によって視覚障害者を支援するための取り組みが盛んに研究されている．これらの支援技術の多くはカメラで得られた画像を解析することで周囲の状況を把握するものであるが，カメラを用いた手法では夜間など暗い場所での利用は難しいなど，環境光に大きく左右される問題がある．また，物体までの距離を測定できるデプスセンサを用いる手法も考えられるが，屋外などの赤外線の外乱がある場合には利用するのが難しい．

[1] 北海道大学
[2] 日本学術振興会
[3] 神戸大学
[a)] hiroki.watanabe@ist.hokudai.ac.jp
[b)] miwa1804@gmail.com
[c)] tsutomu@eedept.kobe-u.ac.jp

一方，音により周囲状況を把握する方法も考えられる．イルカやコウモリ等の一部の動物は，自身が発する音波が帰ってくるまでの時間を利用して，対象との距離を測定するエコーロケーションを用いて周囲状況を把握している．一部の視覚障害者は，自身で発信する可聴音によって，このエコーロケーションを実現可能である [2]．さらに，エコーロケーションは視覚障害者の自立感を向上させることが先行研究 [3] より報告されており，視覚障害者自身の聴覚を用いた周囲状況認識には価値があるといえる．

エコーロケーションでは，一般的に舌打ち音等の可聴音が用いられるが，先行研究 [4] より，エコーロケーションにおいて可聴音の代わりに超音波を発信しその反射波を可聴化すれば，より詳細な物体情報の取得，環境音からの容易な分離等の利点が報告されている．また，周囲の人へエコー音が聞こえないことも利点だと考えられる．しかし，人従来の能力では超音波を発信/取得することは困難である．

本研究では，ウェアラブルコンピューティング環境を活用すれば，リアルタイムの超音波の発信/可聴化が可能であることに着目した．特に近年はイヤホン型のコンピュータであるヒアラブルデバイスを用いて外界音を操作する研究が行われており，音声提示デバイスの常時装着環境が整ってきているといえる [5]．また，超音波を可聴化する際に

は，取得した信号が可聴化処理のためにデバイスを経由するため，機械学習を用いて物体を認識することも考えられる．先行研究では，コウモリの “音で見る” 感覚の理解と，人間への適用を目標としている [4], [6]．そのため，超音波反響音を可聴化し，人間に聞かせることで評価しており，機械による認識については考慮されていなかった．人によるエコーロケーションと，機械学習の認識を組み合わせることで，エコーロケーション使用時の人への負荷の低減や，エコーロケーションの学習支援への応用が考えられる．

以上より，本研究では，超音波の可聴化音を用いた人による認識と，機械で超音波信号を処理した機械による認識を組み合わせた，人と機械による協調型エコーロケーションを提案する．本研究では，対象とするユーザはエコーロケーションを利用可能な人，またはエコーロケーションを学習中の人と想定する．従って，ユーザが超音波エコーロケーションを行うために，超音波スピーカとマイクをすでに装着している環境を想定する．また，ユーザは探したい物体をシステムに設定する．ユーザが身に付けた超音波スピーカからは超音波信号を発信し，物体から反射した音をマイクで取得する．人による認識のために，反射波の周波数を下げ，ユーザに聞き取れるように可聴化する．同時に，機械による認識のために，反射音から周波数スペクトルを計算し，機械学習による認識を行う．機械学習の結果がユーザの事前に指定した物と一致すれば，デバイスが音/振動でユーザに物体の存在を提示する．ユーザはこれらの情報を同時に受け取り，物体探索できる．提案手法では，従来の視覚障害者支援システムのように機械学習の結果に頼るだけでなく，可聴化した音によりユーザ自身の耳でも周囲のセンシングを行う点が特徴である．前述したように，視覚障害者が自身の耳で環境を認識することには意義があり，視覚障害者の自立にとって重要だと考えられる [3]．従って，ユーザの自立心を保ちながら，どのように機械が補助するべきか考慮することが重要である．さらに，本研究では超音波エコーロケーションを行う環境を想定するため，提案手法では機械学習のための追加のデバイスが必要ない．

機械学習をエコーロケーションに組み込む手法は，我々の知る限り本研究が初めてである．従って，本稿では協調型エコーロケーション実現の初歩として，エコーロケーションにおける機械学習の有効性と，協調型エコーロケーションにおける機械学習の設計指針を明らかにすることを目的とする．

提案手法の有効性を確かめるためにプロトタイプを実装し，6 種類の物体に対する認識精度を評価した結果，平均で 92.5%の精度であった．さらに，可聴化と機械学習を組み合わせたシステムを構築し，物体発見実験を行った．実験の結果，人による認識のみの場合，物体探索に平均 71.5 秒かかっていたところ，機械学習を加えると平均 45.9 秒で発見できた．また，機械学習を加えることで，精神的負荷の尺度を示す NASA Task Load Index (NASA-TLX) が有意に減少した．さらに，評価結果に基づき，超音波エコーロケーションにおける機械学習の設計指針を議論した．

2. 関連研究

2.1 エコーロケーション

一部の視覚障害者は，自身の発するクリック音から周囲の状況を認識する，“音で見る” 技術を発達させている．可聴音を用いたエコーロケーションについては多く研究されている [2], [3], [7]．一方，コウモリのエコーロケーションに着想を得て，超音波を用いたエコーロケーションを人間に適用しようとする研究も複数行われている [4], [6], [8], [9]．Sohl-Dickstein らは，超音波を発信し，反射波をリアルタイムに可聴化するシステムを提案した [8]．このシステムにより，物体の有無や物体との距離を超音波で認識できる．Sumiya らは，超音波反響音を可聴化することによって，物体のテクスチャまで弁別できることを示した [6]．被験者は訓練することで，物体からの反響音のみで，その物体の形状やテクスチャの種類の弁別が可能になる．

上記のようにエコーロケーションに関する研究は多く行われているが，エコーロケーションの分野では，人の “音で見る” 感覚の理解を目指しており，ウェアラブル環境でリアルタイムにシステムとして実現したものは少ない．文献 [8] では，リアルタイムの超音波エコーロケーションを実現しているが，物体との距離や物体の有無の認識に着目している．本研究では先行研究では着目されていない，物体の種類の認識に着目する．さらに，超音波エコーロケーションと機械学習を組み合わせることで，人と機械の協調型エコーロケーションシステムを実現する．

2.2 視覚障害者の支援システム

近年のデバイスの発展により，IT 技術を活用した視覚障害者支援システムが数多く研究されている．Brock らは，Microsoft Kinect を利用して物体を検出し，音によりその存在を提示するシステムを提案した [10]．物体の位置や物体との距離によって，提示音の音量やピッチが変化することで，ユーザは物体の位置を把握できる．Kayukawa らは，視覚障害者の歩行を支援するスーツケース型システムを提案した [11]．Virtual Paving は，振動と音声によって，誘導ブロックが利用できない場所でユーザを誘導するシステムである [12]．日常品の買い出しは必要不可欠な活動であるが，視覚障害者にとっては困難であり，食料品店での買い物をサポートするシステムが提案されている [13]．Acoussist は，車から発信される超音波を検知することにより，視覚障害者の道路横断を支援するシステムである [14]．

上記のように，多くの視覚障害者支援システムが研究されているが，これらはいずれも機械が主に判断を行ってい

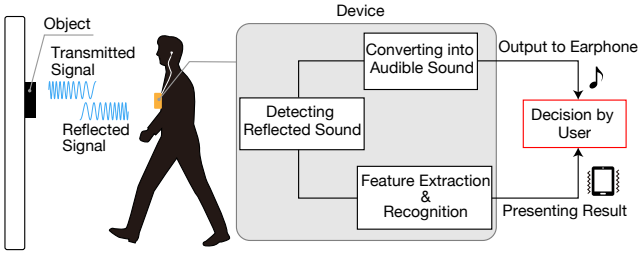

図 1: システム構成

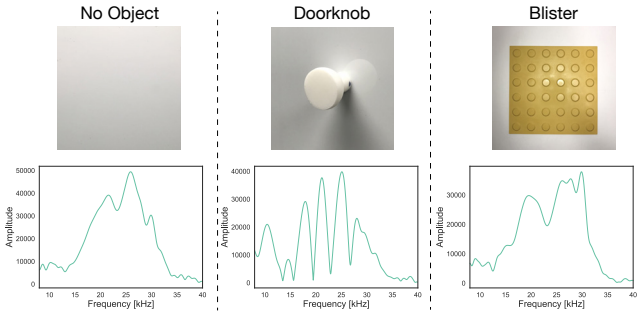

図 2: 各物体に対応する周波数スペクトル

るものである．本研究では，エコーロケーション（人による認識）と機械による認識を組み合わせる点で異なる．

2.3 音による物体認識手法

音を用いた物体検出/認識の研究も多く行われている．Komatsu らは，超音波の反射音の周波数特性から，対象物体の特性を判別する検討をした [15]．ObstacleWatch は，音響信号の反射を利用して障害物を検知する手法である [16]．Remaggi らは，音の反射波の周波数領域での減衰率から，反射した対象の材質を推定する手法を提案した [17]．Mao らは，市販のスマートフォンを用いた，音波によるイメージング手法を提案した [18]．ユーザは仮想のセンサアレイを模して，あらかじめ設定された軌道に沿ってスマートフォンを動かし，システムはその反射音を用いて対象物を再現する．

上記の研究では，機械によって物体を認識することに着目しているが，本研究では可聴化した音を用いた人による認識と，機械による認識を組み合わせる点が異なる．

3. 提案手法

図 1 に本研究のシステム構成を示す．本研究のシステムは，人間による認識のための処理と，機械による認識のための処理の 2 つの部分から構成されている．まず，スピーカから超音波信号を送信し，物体からの反射音をマイクで検出する．人間側の処理では周波数を落として可聴化し，ユーザにイヤホンを通して可聴化した音を提示する．機械側の処理では，得られた信号に対して機械学習による分類を適用し，ユーザの設定した対象物が認識された際にユーザに通知する．本研究ではタブレットの振動で提示する．**図 2** に各物体からの反射波から得られた周波数スペクトルの一例を示す．この図に示すように，物体によって得られる周波数スペクトルに違いが確認できる．人間は音色でこの違いを，機械は特徴量により周波数スペクトルの違いを，それぞれ認識する．最終的に，ユーザは自身の耳から得られる音と，機械の認識結果である振動を組み合わせ，物体の認識を行う．

3.1 信号の発信と検出

スピーカからは超音波のスイープ信号が発信され，マイクでは物体からの反射波を検出する．信号の発信/録音のサンプリング周波数は 96kHz とした．本研究で用いた信号は 40kHz から 20kHz へと 1ms で下降するスイープ信号である．スイープ信号はコウモリが物体を識別する際に用いられる信号であり，先行研究でもスイープ信号が用いられている [6]．先行研究ではコウモリの感覚の人の理解という観点のため 8k–40kHz の範囲の信号を利用していたが，本研究ではウェアラブルデバイスとしての利用を想定する．デバイス利用中に常に可聴音が発生するのはユーザにとっても周囲の人にとっても不快となる可能性があるため，可聴音の領域を含まないよう本研究では 20k–40kHz の周波数範囲を選択した．

マイクで得られる信号には，スピーカからマイクへと直接届く音波（直接波），対象物体から反射してマイクへと入力される音波，周囲の物体に反射し，異なる経路を通る反射波（マルチパス）等も含まれる．したがって，対象物体からの反射波のみを検出するために，マイクからの受信信号と発信信号との相互相関値を計算する．**図 3** に壁から 50cm 離れた距離から信号を発信した時に得られた相互相関値の包絡線を示す．相互相関値は，受信信号に送信信号と同じ波形が得られた時に大きな値を示す．図 3 では，最初に直接波を表す大きなピークが確認され，その後にも複数のピークが検出されている．予備的な検討より，本研究では直接波以降のピーク値のうち最も大きな相互相関値を対象からの反射波とし，その他のピークはマルチパスによる反射波とした．求めた 1 次反射波の位置から，128 サンプルを抽出する．送信波のサンプル数は 96（96,000 Hz × 1 ms）であるが，余裕を持って 128 サンプルとした．

3.2 超音波の可聴化

3.2.1 変換手法

前節までに切り出した超音波反射音の周波数を下げることで，人に知覚できる音に変換する．可聴化手法としては，先行研究でも用いられている，波形情報を崩さないタイムストレッチを選択した [6], [8]．具体的には，検出された反射波は通常の $1/m$ の速度でユーザに提示される．m は可聴化倍率である．これにより，信号は時間軸上で m 倍に拡

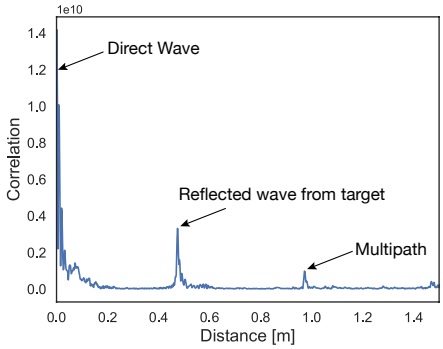

図 3: 相互相関値

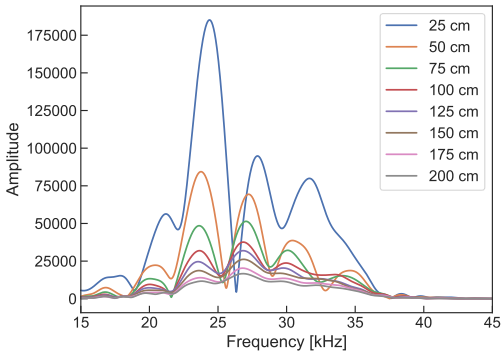

図 4: 距離による周波数スペクトルの変化

大され，周波数は可聴範囲に下げられる．本研究では事前調査に基づき，可聴化音の聞き分け性能が高かった $m = 20$ とした．

3.2.2 ミニチュアダミーヘッド

ダミーヘッドとは人間の頭部を模したものであり，これを用いて録音することで人の頭部伝達関数（HRTF: Head-Related Transfer Function）を考慮した録音ができる．ミニチュアダミーヘッド（MDH: Miniature Dummy Head）とはダミーヘッドを小さくしたものであり，スケールに応じた超音波の観測に対して用いられる．MDH を用いて録音した超音波を可聴化すると，ユーザは可聴音によって得たエコー音と同じような感覚で超音波反射音を聞けることが報告されている [4]．本研究では，先行研究 [4], [6] で用いられたものと同様の 1/7 サイズの MDH を用いた．

3.3 機械学習

3.3.1 前処理

3.1 節で抽出した波形の不連続性を解消するため，ハン窓をかける．さらに，抽出した 128 サンプルのデータにゼロを加える（ゼロパディング）ことで，8,192 サンプルとし，高速フーリエ変換（FFT: Fast Fourier Transform）を適用することで周波数領域の情報を得る．ゼロパディングにより周波数分解能は約 11.7Hz（96,000 / 8,192）となり，より詳細な特徴が得られると考えた．

3.3.2 特徴量抽出と認識

図 2 に示すように，反射の違いは周波数スペクトルに現れる．従って，本研究では特徴量として，LFCC（Linear-Frequency Cepstral Coefficients）を利用した．LFCC は，音声認識によく用いられる MFCC（Mel-Frequency Cepstral Coefficients）の線形版である．MFCC では人の声の特徴を考慮し，周波数スペクトルに対して低周波ほど細かいメルフィルタバンクを用いる．一方，LFCC では等間隔に設置されたフィルタバンクを用いる．本研究では人の声を対象としていないため，対象領域から等しく特徴を抽出できる LFCC が適していると考えた．LFCC のフィルタバンク数は 20 とした．抽出した 20 次元の特徴量のうち，直流成分を示す 1 次元目を除いた 19 次元を特徴として用いる．

さらにその他のスペクトルの特徴量として，周波数スペクトルの平均，分散，スペクトラル重心，ロールオフ，フラットネス，歪度，尖度，帯域幅，エントロピーの 9 個の特徴量を算出する．上記の特徴量をマイクの左/右チャネルそれぞれに対して計算し，合計 56 個の特徴量を得る．

分類器としては，ウェアラブルデバイス上でも実装が容易かつ高性能な必要がある．本研究では Support Vector Machine (SVM) を選択した．

3.3.3 距離に応じた学習モデルの切り替え

音圧は距離とともに減衰することが知られており，同じ物体でも距離によって得られる周波数特性が異なる可能性がある．距離によって得られる周波数特性の変化の一例を**図 4** に示す．この図に示すように，周波数スペクトルのおおよそのピーク/ノッチの位置はどの距離でも近傍の周波数に存在してはいるが，距離によって少しずつずれていることが確認できる．また，ピークとノッチの凹凸も距離が大きくなるに従って鈍っていることが確認できる．従って，同じ学習モデルで全ての距離のデータを認識すると認識率の低下が懸念される．

そこで，本研究では算出された物体との距離に従って学習モデルを切り替えることにより認識精度の向上を行う．物体との距離は以下の式から導出できる．

$$d = \frac{c(t_1 - t_0)}{2} \tag{1}$$

ただし，d は物体とデバイスとの距離，c は音速，t_0 はスピーカの送信時刻，t_1 は反射波の取得時刻である．本研究では 25cm から 200cm まで 25cm 間隔ごとにデータを取得し，8 つのモデルを構築した．算出された物体までの距離から最も近いモデルを利用することで，認識率を向上させる．

4. 実装

図 5 に示すように，提案手法のプロトタイプを実装した．プロトタイプは，タブレット（Huawei MediaPad M5），オーディオインタフェース（Zoom U-24），スピー

図 5: 実装したデバイス: (a) 全体図，(b) 把持部分

カ（Fostex FT200D），マイク（Countryman B6），ヘッドフォン（Bose QC35）から構成されている．Bose QC35 にはノイズキャンセリング機能が備わっているが，本研究では用いていない．スピーカからマイクへの直接音を抑えるために，スピーカとマイクの間にエラストマー樹脂を挟んだ．スピーカの周波数特性は 1k–50kHz であり，超音波信号を発信するのに十分な性能である．マイクの周波数特性は 30–20kHz[*1]であるが，先行研究 [6] においてもこのマイクが使用されていること，及び予備実験から超音波領域の音も取得可能であることが確認できたため，このマイクを使用した．MDH は 3D プリンタにより作製した．MDH の 3D モデルは先行研究 [4], [6] で用いられたものと同様であり，一般的に用いられるダミーヘッドの 1/7 のサイズとなっている．2 つのマイクは MDH の両耳の鼓膜に当たる位置にそれぞれ設置されている．MDH のサイズは，高さ 6.5cm，幅 5.2cm，奥行き 2.5cm である．人体に近い柔らかさを実現するために，素材にはシリコンゴム素材（ショア硬度：35）を用いた．ヘッドフォンとマイクはオーディオインタフェースを介してタブレットと接続される．スピーカにはマイコン（Sony Spresense）が接続されており，SD カードに保存した音源を再生する．Spresense は 96kHz 以上の再生にも対応しており，超音波信号を発信することが可能である．また，アンプも内蔵されているため，スピーカから十分な音圧の信号を発信することが可能である．

また，3 章に示す手法を実現するために，Android アプリケーションを実装した．なお，本稿では各種設定（可聴化の有無，機械学習の有無，可聴化倍率の変更等）を固定したが，アプリケーションではこれらの設定の変更が可能であり，ユーザは自分好みに変更することが可能である．機械学習には 3.3.2 節に示すように SVM を利用した．

5. 評価

5.1 機械学習による物体認識

本研究で認識すべき対象物体として，触れる前にわかれば便利と考えられる物体を選定した．点字ブロックや点字は，視覚障害者への情報提示として広く用いられているが，これら自体がどこに存在するかは実際に触れてみるまで知

[*1] https://countryman.com/product/b6-omnidirectional-lavalier/

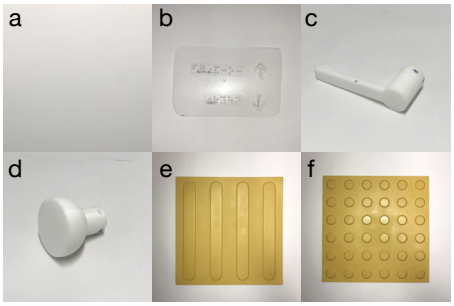

図 6: 評価した物体: (a) 物体なし，(b) 点字，(c) ドアハンドル，(d) ドアノブ，(e) 線状ブロック，(f) 点状ブロック

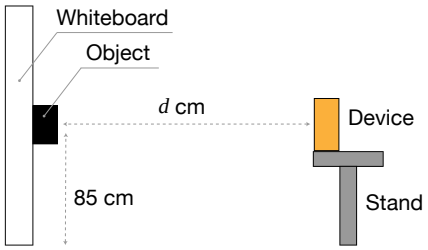

図 7: 実験環境 (側面図)

ることができない．また，日常生活において，ドアの開閉に用いられるドアハンドルやドアノブの操作は欠かせないが，これらがどこに存在しているかも触れるまで知ることができない．手探りでこれらの物体を探索することも可能ではあるが，衛生面や社会的受容性の面からも，不用意に未知の物体に触れるのは望ましくない．従って，非接触でこれらを識別できれば有用だといえる．以上を考慮し，**図 6** に示す，物体なし，点字，ドアハンドル，ドアノブ，線状ブロック，点状ブロックの合計 6 種類を認識すべき物体として設定した．ドアハンドルとドアノブは 3D プリンタ（UP Plus 2）により作製した．各物体の高さは，点字が約 0.3mm，ドアハンドルが約 4cm，ドアノブが約 5cm，線状/点状ブロックが約 3mm である．

実験環境を**図 7** に示す．図 7 に示すホワイトボードに順番に物体を貼り付け，実装したプロトタイプデバイスにより超音波信号を発信し，反響音を取得する．デバイスとホワイトボードとの距離 d は 25，50，75，100，125，150，175，200cm の 8 種類とし，それぞれの距離において 30 回の測定を行なった．これを 1 セットとする．物体とデバイスを設置しなおして，もう 1 セット測定した．結果として，2,880 個の測定データ（6 物体 ×8 距離 ×30 測定 ×2 セット）を得た．取得したデータに対して 3.3.2 節に示す手法で特徴量を抽出し，認識精度を算出する．評価には leave-one-set-out cross-validation を利用した．

表 1 に各距離での認識率を示す．距離による学習モデル切り替え無しの場合，25cm の際の学習モデルを用いて他の距離のデータを評価した．この表が示すように，モデル切替ありの場合，100cm を除く全ての距離で 90%以

表 1: 各距離での認識精度 [%]

距離 [cm]	25	50	75	100	125	150	175	200	平均
切替無し	97.5	49.4	23.1	25.0	25.0	21.4	16.7	24.4	24.4
切替あり	97.5	91.7	92.2	76.4	90.8	100	100	91.7	92.5

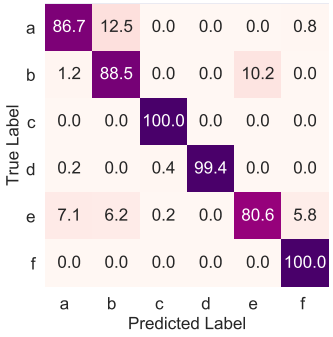

図 8: 混同行列 [%]: (a): 物体なし，(b) 点字，(c) ドアハンドル，(d) ドアノブ，(e) 線状ブロック，(f) 点状ブロック

上の認識精度が得られ，平均で 92.5%の精度であった．一方，学習モデル切替無しの場合，学習モデルに用いた距離（25cm）とテストデータとの距離が離れるに従って，認識精度の低下が確認できた．モデル切替ありの場合でも，距離が 100cm の際の認識精度は 75.6%となり，他の距離と比較すると低くなった．これは，マルチパスの影響だと考えられる．つまり，デバイス周囲 100cm の距離の別の物体からの反射波が，対象物体からの反射波に重なり，対象物体から得られる反射波本来の周波数特性とは異なったためだと考えられる．**図 8** に，学習モデル切替ありの場合の，全距離での結果をまとめた混同行列を示す．この図に示すように，線状ブロックを除き，全ての物体で認識精度は 86%以上であった．また，物体なしは点字と，線状ブロックは物体なし，点字，点状ブロックと混同しやすいことが確認できた．一方，ドアハンドルとドアノブの認識精度は 99%以上であった．これは，物体の高さによるものと考えられる．本研究で用いたドアハンドルとドアノブはそれぞれ約 4cm，5cm となっており，他の物体と比較すると大きい．物体の高さよって周波数スペクトルに現れるノッチの深さが大きくなることが先行研究からも報告されており，この特徴的な周波数スペクトルによって他の物体よりも識別が可能であったと考えられる [6]．

以上の結果より，反射音によって物体を識別すること及び，物体との距離によって学習モデルを切り替える手法は有効だと考えられる．

5.2 人と機械による認識を組み合わせた物体発見

本節では，超音波エコーロケーションに機械学習を組み合わせた際の影響や効果の調査を目的とする．実験環境を**図 9** に示す．本実験では，部屋へ入る/部屋から出る場面を想定し，前節で用いた物体のうち，ドアノブを探索対象の

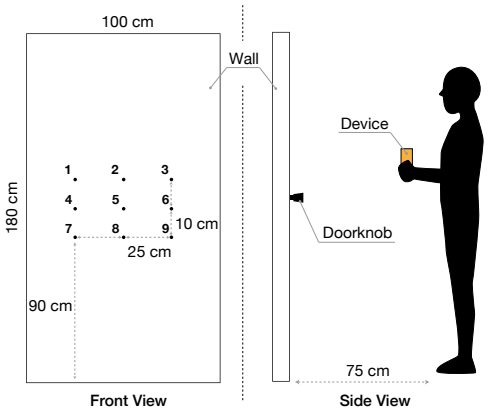

図 9: 実験環境: (左) 正面図，(右) 側面図

物体として選択した．図 9 に示す，ドアを想定した 100cm × 180cm の領域のうち，高さ 90–110cm の範囲の 9 箇所のいずれか一つにドアノブを設置する．この範囲は，ドアノブが存在すると考えられる高さを考慮して決定した．ドアノブは前節で用いたものと同様であり，機械学習には前節で取得したデータで，検出距離ごとに学習モデルが切り替わるものを利用した．

5.2.1 信号発信間隔

先行研究 [7] では，エコーロケーション習熟者の信号発信間隔が約 600ms であることが報告されているが，超音波エコーロケーションにとってもその間隔が最適であるかは明らかでない．人によるエコーロケーションでは舌打ち音を用いるため信号発信の間隔に限界があるが，システムであれば人よりも短い間隔での信号発信も可能である．従って，本節では超音波エコーロケーションの探索において適した信号発信間隔を調査した．

本実験ではドアノブの位置を特定するタスクを想定し，壁に貼り付けたドアノブを，デバイスを用いて被験者に探索してもらう．被験者は 20 代の健常者の男性 10 名であり，全員エコーロケーションの初心者であった．被験者は図 9 に示す環境で，図 5 に示すデバイスのスピーカ部を手に持って，壁に向かって平行にスキャンするようにデバイスを動かして物体の検出を行う．まず，被験者にはデバイスを用いて物体を検出する練習を，目を開けた状態で 5–10 分程度行ってもらう．被験者と壁との距離は約 75cm とした．被験者自身が習熟したと感じた段階で練習を終了し，図 9 に示す領域のうち，ランダムな箇所にドアノブを設置して，被験者は目を閉じた状態でこれを探索する．被験者はドアノブを発見したと思ったら実験者に報告する．実験者がドアノブの位置と，デバイスを壁に投影した位置とを確認し，約 5cm 以内であれば正解として実験を終了する．それ以外の場合は正解できるまで実験を続けた．

被験者はこの試行を，以下の 3 つの信号発信間隔について行った．

- 600ms: 先行研究で報告されているエコーロケーショ

表 2: 被験者が選択した信号発信間隔 [ms]

被験者	A	B	C	D	E	F	G	H	I	J
音	200	200	600	200	400	200	400	200	200/400	200
音+振動	200	200	200	200	200/400	200/400	400	200	200/400	200

ン習熟者の信号発信間隔 [7]

- 400ms: 600ms と 200ms の中間
- 200ms: 本研究の実装で実現可能な最短の信号発信間隔

カウンターバランスを取るために，被験者ごとに試行する信号発信間隔の順番が異なるようにした．被験者は上記の操作を，まず可聴化した音のみ（音条件）で行い，次に可聴化した音と機械学習結果の振動（音+振動条件）を用いて行った．なお，本研究のシステムは超音波エコーロケーションを前提としているため，超音波エコーロケーション自体に慣れてもらうために，全被験者に対してまずは音条件で試行し，その後に音+振動条件で実験を行った．各手法の試行後に，どの信号発信間隔が適切と感じたか，被験者に回答してもらった．

表 2 に，各被験者が選択した信号発信間隔を示す．なお，200/400 は 200ms も 400ms も同じように感じたことを意味する．音条件の場合，最も選択された信号発信間隔は 200ms であり，10 人中 7 人が選択した．この理由として，“短い信号発信間隔によって，物体をスキャンするスピードが上がった” というコメントが得られた．一方，被験者 A と D は 200ms を選択したが，“短すぎる信号発信間隔は不快だった” とコメントした．200ms 以外を選択した被験者 C，E，G からは，“信号発信間隔が短すぎると，音を聞き分けるのが難しかった” という意見が得られた．

音+振動の場合においても，最も選択された信号発信間隔は 200ms であり，10 人中 9 人が選択した．音のみの場合よりも多くの被験者が 200ms を好んでおり，この理由として，“音と振動が短い間隔で提示されることで，デバイスを少しずつ動かしてもドアノブを見つけやすくなった” というコメントが得られた．

以上より，実験環境統制のため，以降の評価では信号発信間隔を 200ms に設定した．なお，実利用の際には，本システムは信号発信間隔の変更が可能であり，ユーザごとに好みの信号発信間隔を設定することが可能である．また，被験者から得られた意見にもあるように，信号発信間隔が短いと得られる情報が増える一方，ユーザにとって不快となることが報告された．この点については 6.2 節で議論する．

5.2.2 物体検出

前節で求めた信号発信間隔（200ms）を用いて，人による認識と，人と機械による認識を組み合わせた場合のユーザの行動を評価した．本節の実験環境は図 9 と同様であり，

表 3: 各被験者の誤答数と (強制終了数)

被験者	A	B	C	D	E	F	G	H	I	J
音	0	0	1	0	0	0	5 (1)	0	0	0
音 + 振動	0	0	0	0	0	0	5	1	0	0 (1)

被験者は前節と同じである．つまり，被験者は提案手法の利用方法の練習が済んだ状態である．被験者にはドアノブを，目隠しした状態でデバイスを用いて発見してもらった．前節同様に，音条件と音+振動条件の 2 通りを比較した．なお，本研究は超音波エコーロケーションが行われているという環境を想定するため，音条件がベースラインとなる．音条件と音+振動条件の実験の順番は被験者全体でカウンターバランスが取れるように設定した．被験者は各手法において 3 回ずつ実験を行った．ドアノブの設置位置による影響が少なくなるように，3 回の設置位置は図 9 の縦の 3 つのライン（左，中，右）及び横の 3 つのライン（上，中，下）の全てを含むように設定した．例えば，被験者 A の実験でのドアノブ設置位置は，音条件の場合 1, 5, 9 で，音+振動条件の場合，2, 4, 9 のような配置となる．実験のスタート位置は，毎回領域の中心（図 9 中の 5）からとした．以上より，合計で 60 個（3 試行 × 2 手法 × 10 被験者）の測定データを得た．本研究では，機械学習の有無で被験者の探索行動が変化するかどうかを調査するため，デバイス下部（スピーカの中心から 5cm 下）にレーザーポインタを設置し，録画映像から画像処理によって壁に投影されたレーザーポインタの軌跡を取得し，探索行動の軌跡とした．被験者は対象の物体を発見したと判断したらその旨を実験者に対して述べる．実験者は回答時点のデバイスを壁に投影した位置と，ドアノブの位置とが約 5cm 以内に存在すれば正解と判断し，実験を終了する．デバイスの投影位置と物体が 5cm 以上離れている場合誤答数としてカウントし，被験者は実験を続行する．300 秒以上正答できなかった場合，実験を強制的に終了した．音条件，音+振動条件の各実験後に，被験者には NASA-TLX に回答してもらう．また両手法の終了時には自由記述のアンケートにも回答してもらった．

5.2.2.1 誤答数

表 3 に各被験者の誤答数を示す．この表に示す通り，被験者 G をのぞいて機械学習の有無に関わらず誤答数は少なかった．これは，被験者が確信が得られるまで時間をかけて探索を行ったためだと考えられる．一方，被験者 G は “物体がない場所でも音の変化を感じた” と述べており，誤答数が多くなった．

5.2.2.2 物体発見に要した時間

図 10 に機械学習の有無による物体発見までの時間を示す．この図に示すように，機械学習によって物体発見までの平均時間は 71.5 秒から 45.9 秒へと減少した．しかし，検出時間について t 検定を行った結果，機械学習の有無で

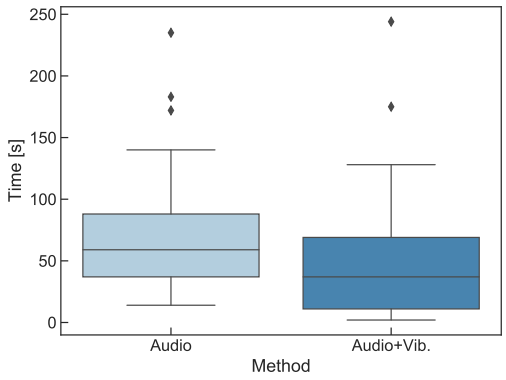

図 10: 物体発見に要した時間

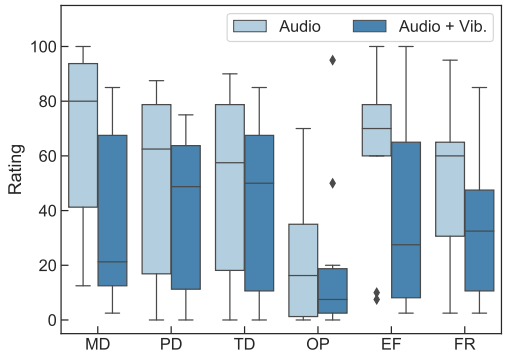

図 11: NASA-TLX の比較

有意な差は確認できなかった ($t\,(27) = 1.92$, $p = .065$).

5.2.2.3 NASA-TLX

NASA-TLX の評価には，6 つの下位尺度の単純平均値である Raw TLX (RTLX) を用いた [19]．得られた RTLX は，音条件と音+振動条件でそれぞれ 50.6 と 35.6 であった．各手法の得点について t 検定を行った結果，両手法間の平均の差は有意であり ($t\,(9) = 2.93$, $p < .05$)，音条件の平均が音+振動条件よりも有意に大きかった．**図 11** に，機械学習の有無による NASA-TLX の各尺度の変化を示す．値が低いほど精神的負荷が少ないことを示す．平均値で比較すると，機械学習を加えることで全ての指標において精神的負荷が低減していることが確認できた．特に，身体的要求（MD: Mental Demand），努力（EF: Effort），不満（FR: Frustration）が 6 つの尺度の中では大きく減少しており，MD は 43.3%，EF は 36.6%，FR は 30.7%減少した．一方，身体的要求（PD: Physical Demand），時間切迫感（TP: Time Pressure），作業達成度（OP: Own Performance）は上記の 3 指標と比較すると減少幅が少なく，PD は 20.2%，TP は 16.8%，OP は 19.2%の減少にとどまった．注目すべき点として，OP は機械の有無に関わらず低い値を示していた．つまり，音のみでも被験者にとって達成感は高かったといえる．

5.2.2.4 物体探索行動

デバイスの移動軌跡の一例として，被験者 C の 3 回目の試行の軌跡を**図 12** に示す．なお，図中の点線の円はドアノブの位置を示す．音条件の場合は，物体の箇所を通り過ぎて一度全体をスキャンした後に，被験者が怪しく感じた箇所に戻ってきている．その後左右に往復して他の箇所との違いを確かめた後に物体の位置を特定していることがわかる．一方，音+振動条件の場合，全体をスキャンする前に被験者が気になった部分で止まり，かつ機械からの振動があれば確信を得て物体の位置を特定しており，デバイスの移動距離及び探索時間を削減できていた．以上より，機械学習を組み合わせることで物体発見までの探索距離と時間を削減できていることが確認できる．

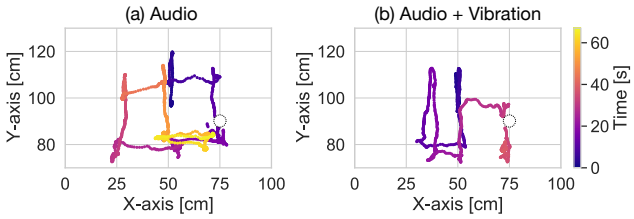

図 12: 探索軌跡: (a) 音条件，(b) 音+振動条件

6. 考察

6.1 機械学習の有効性

実験結果に示すように，機械学習による反射波を用いた物体の認識は可能であることがわかった．また，統計的な有意差は確認されなかったが，機械学習により物体発見までの平均時間が減少した．さらに，NASA-TLX の結果より，音条件と比較して，音+振動条件では精神的負荷が有意に低いことが確認できた．これは，図 12 に示すように，音のみでは確信が得られなかった場合に，機械による認識の振動があることで，確信を持って物体の検出を行えたことに起因する．つまり，機械が人間の判断の補助をしていると考えられる．

被験者からのアンケートでは，10 人中 9 人の被験者が，機械学習ありの方が無しよりも良かったと回答していた．機械学習ありの方が良いと回答した被験者からは，“音だけでは確信を持てない場合に，機械による振動があることで物体があることの確信を持てた” という回答が得られた．一方，機械無しの方が良いと回答した被験者 A は，“機械の振動があるとかえって注意が散漫になった” と回答している．これは，機械の認識結果による振動に集中するあまり，耳から聞こえてくる音の変化に集中することが難しかったためだと考えられる．この解決方法として，音による探索をメインに，振動はオプションとして設定しておき，ユーザが欲している時だけ機械による探索を有効にするという対策が考えられる．また，音提示の有無も切り替えられる必要がある．視覚障害者にとってエコー音以外の音の情報も重要なため，ユーザが欲したときにだけ可聴化音を提示し，それ以外は可聴化しない方が良いといえる．

本研究では，機械の有無による変化を観測するためにシステムの設定を統一したが，提案システムは様々なパラメータ（信号発信間隔，機械学習の有無，可聴化の有無，可聴化倍率）を被験者ごとに好みの設定に変更することが可能であり，個人ごとの特性にあった設定にすることで，これらの問題は解決できると考える．以上より，超音波エコーロケーションに機械学習を導入することには効果があると考える．

6.2 人と機械の認識の組み合わせ

5.2.1 節に示すように，被験者からのコメントとして，“200ms の間隔で常に可聴音が発生していると不快だった”という意見がある一方で，機械学習については多くの被験者が 200ms 間隔を好んだ．これらの意見から，可聴化と機械でユーザに提示する間隔をそれぞれに最適なものにすることが良いと考えられる．つまり，信号自体は 200ms 間隔で発信するが，可聴化は 200–600ms 程度の各ユーザの好みの間隔で，機械は 200ms 間隔で認識を行い，ユーザに提示することが考えられる．ユーザが音への違和感を感じ，より短い間隔で可聴音が欲しい場合には，可聴音の提示間隔も動的に短く変更するなどの対応が考えられる．

図 11 に示すように，NASA-TLX の値において，OP は機械の有無に関わらず低い値（高い達成感）の被験者が多かった．被験者 E からは，“音のみの方が自力でやったという思いが強く，達成感がある”というコメントが得られた．さらに，先行研究からもエコーロケーションは視覚障害者の自立感を高めることが報告されている [3]．また，被験者 J の実験では，音条件と音+振動条件で同じ箇所の物体探索があったが，音条件の場合は 69s，音+振動条件の場合は 300s（強制終了）と，機械学習を加えた際に大幅な時間ロスが見られており，元々音だけで発見できていたものが，機械を加えることで逆に発見が遅くなった．これに対して，被験者 J は “機械ありの場合，途中でいつの間にか機械だけに頼る認識になっており，音を利用しておらず，余計に時間がかかった” とコメントしている．以上より，エコーロケーションにおいて音を用いることは重要であり，機械だけの判断に頼るのではなく，人による認識も組み合わせることが提案手法において重要だと考える．

6.3 機械学習の戦略

被験者 I，J からは，“物体がない場所で機械の誤認識（振動）があり，その付近を探索して時間がかかってしまった”という意見が得られた．実環境での利用を想定すると，誤認識はさらに発生すると想定され，機械学習が人の判断を混乱させる可能性が考えられる．従って，本システムにおける機械学習の戦略としては，適合率を重視すべきと考える．つまり，物体の検出を多少ミスしても，機械が探索対象の物体だと判断した時には，その確率は非常に高いという状態が望ましいと考える．また，本研究では機械の毎回の認識結果をそのまま用いていたが，直近数回の認識結果を用いて多数決をとった上でユーザに認識結果を提示するという方法も考えられる．

さらに，本研究では，機械学習の結果提示のための振動は 100ms の一定強度の振動にしていたが，被験者 B からは “振動にもパターンがあれば良い” という意見が得られた．従って，機械学習の認識結果の信頼度によって振動にパターンを与えることも効果的だと考えられる．例えば，機械学習の信頼度が高い時には強く長い振動で，信頼度が低い時には弱い振動などの使い分けが考えられる．

6.4 協調型エコーロケーション実現に向けて

我々の知る限り，エコーロケーションに機械学習を導入した手法は本研究が初めてのため，本稿では，まず協調型エコーロケーションへの第一歩として，エコーロケーションにおける機械学習の有効性と，エコーロケーションにおける機械学習の設計指針を明らかにすることに着目した．協調型エコーロケーション実現のためには，実環境を想定したさらなる調査が必要である．例えば，視覚障害の方に提案システムを使ってもらう，長期間利用による影響，より広い/混雑した実環境に近い環境での利用等が考えられる．本稿での結果は，今後の研究の方向性を示すものと考える．

6.5 発信信号の動的な変更

本研究では，信号の種類と周波数，信号発信間隔を固定して実験を行ったが，超音波エコーロケーションを用いるコウモリは，対象との距離や反響音から得られる情報によって，発信する信号の種類や間隔を動的に変化し，得たい情報に最適な信号を選択しているといわれている [20]．このようなコウモリの生態を参考に，提案手法でも反響音から得られた情報によって送信する信号を動的に変化させ，人/機械にとってより精度の高い認識が実現できる可能性がある．発信信号の動的な変更についての調査は今後の課題としたい．

6.6 物体検出に要する時間

本研究の方法では，機械学習を用いても物体の発見までに平均 45.9 秒の時間を要しており，実環境で用いるには時間がかかりすぎだといえる．この原因として，被験者がエコーロケーションの初心者であったため，必要以上の時間を要した可能性がある．エコーロケーションの習熟者であれば，可聴化のみの場合でも物体発見までの時間を短縮でき，さらに機械学習を加えることで探索時間や精神的負荷も低減可能であると考えられる．しかし，エコーロケーションを実行可能な視覚障害者は多くないため，被験者とするのは困難であった．提案システムは，エコーロケー

ションの練習にも利用可能と考えられる．従って，提案システムが，今後エコーロケーション初心者のエコーロケーションへの挑戦を促すことを期待する．エコーロケーションの訓練も含めた長期間の実験は今後の課題である．

7. おわりに

本研究では，物体へ超音波を発信し，反射波をリアルタイムに可聴化する超音波エコーロケーションと，機械学習により物体を認識する手法を組み合わせた，人と機械の協調型エコーロケーションを提案した．プロトタイプデバイスを実装し，評価を行った結果，機械学習による物体の認識は，6 物体に対して平均 92.5%の認識精度であった．また，物体の探索を対象として，可聴化のみによるエコーロケーションと可聴化と機械を組み合わせたエコーロケーションを比較した結果，物体発見までの時間は平均 71.5 秒から 45.9 秒へと減少した．また，機械学習により精神的負荷が有意に減少することが確認できた．実験結果より，エコーロケーションにおける機械学習の有効性と設計指針を議論した．本稿で得られた知見が今後の協調型エコーロケーションの研究の方向性を示すと考える．

謝辞 本研究は JSPS 科研費 21K11973，JST さきがけ（JPMJPR2138），JST CREST（JPMJCR18A3）の助成を受けたものである．

参考文献

[1] World Health Organization: World Report on Vision (2019). `https://www.who.int/publications/i/item/9789241516570`.

[2] Thaler, L., De Vos, H., Kish, D., Antoniou, M., Baker, C. and Hornikx, M.: Human Click-based Echolocation of Distance: Superfine Acuity and Dynamic Clicking Behaviour, *Journal of the Association for Research in Otolaryngology*, Vol. 20, No. 5, pp. 499–510 (2019).

[3] Norman, L. J., Dodsworth, C., Foresteire, D. and Thaler, L.: Human Click-based Echolocation: Effects of Blindness and Age, and Real-Life Implications in a 10-week Training Program, *PLOS ONE*, Vol. 16, No. 6, pp. 1–34 (2021).

[4] Uchibori, S., Sarumaru, Y., Ashihara, K., Ohta, T. and Hiryu, S.: Experimental Evaluation of Binaural Recording System using a Miniature Dummy Head, *Acoustical Science and Technology*, Vol. 36, No. 1, pp. 42–45 (2015).

[5] Watanabe, H. and Terada, T.: Manipulatable Auditory Perception in Wearable Computing, *Proc. of the Augmented Humans International Conference*, AHs '20, No. 10, pp. 1–12 (2020).

[6] Sumiya, M., Ashihara, K., Yoshino, K., Gogami, M., Nagatani, Y., Kobayasi, K. I., Watanabe, Y. and Hiryu, S.: Bat-Inspired Signal Design for Target Discrimination in Human Echolocation, *The Journal of the Acoustical Society of America*, Vol. 145, No. 4, pp. 2221–2236 (2019).

[7] Thaler, L., Vos, R. D., Kish, D., Antoniou, M., Baker, C. J. and Hornikx, M.: Human Echolocators Adjust Loudness and Number of Clicks for Detection of Reflectors at Various Azimuth Angles, *Proc. of the Royal Society B*, Vol. 285 (2018).

[8] Sohl-Dickstein, J., Teng, S., Gaub, B. M., Rodgers, C. C., Li, C., DeWeese, M. R. and Harper, N. S.: A Device for Human Ultrasonic Echolocation, *IEEE Transactions on Biomedical Engineering*, Vol. 62, No. 6, pp. 1526–1534 (2015).

[9] Sarumaru, Y., Sumiya, M., Banda, T., Ashihara, K., Kobayashi, K. and Hiryu, S.: Human Echo Perception by using Miniature Dummy Head, *Auditory Research Meeting*, The Acoustical Society of Japan, pp. 57–62 (2015).

[10] Brock, M. and Kristensson, P. O.: Supporting Blind Navigation Using Depth Sensing and Sonification, *Proc. of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pp. 255–258 (2013).

[11] Kayukawa, S., Ishihara, T., Takagi, H., Morishima, S. and Asakawa, C.: Guiding Blind Pedestrians in Public Spaces by Understanding Walking Behavior of Nearby Pedestrians, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 3 (2020).

[12] Xu, S., Yang, C., Ge, W., Yu, C. and Shi, Y.: Virtual Paving: Rendering a Smooth Path for People with Visual Impairment through Vibrotactile and Audio Feedback, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 3 (2020).

[13] Boldu, R., Matthies, D. J., Zhang, H. and Nanayakkara, S.: AiSee: An Assistive Wearable Device to Support Visually Impaired Grocery Shoppers, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 4 (2020).

[14] Jin, W., Xiao, M., Zhu, H., Deb, S., Kan, C. and Li, M.: Acoussist: An Acoustic Assisting Tool for People with Visual Impairments to Cross Uncontrolled Streets, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 4 (2020).

[15] Komatsu, T. and Akita, J.-i.: Power Spectrum Analysis of Reflected Waves with Ultrasonic Sensors Indicates "What the Target Is", *Proc. of the 13th ACM Conference on Embedded Networked Sensor Systems*, SenSys '15, p. 387–388 (2015).

[16] Wang, Z., Tan, S., Zhang, L. and Yang, J.: ObstacleWatch: Acoustic-based Obstacle Collision Detection for Pedestrian Using Smartphone, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 2, No. 4, pp. 194:1–194:22 (2018).

[17] Remaggi, L., Kim, H., Jackson, P. J. B. and Hilton, A.: An Audio-Visual Method for Room Boundary Estimation and Material Recognition, *Proc. of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia*, AVSU '18, pp. 3–9 (2018).

[18] Mao, W., Wang, M. and Qiu, L.: AIM: Acoustic Imaging on a Mobile, *Proc. of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '18, p. 468–481 (2018).

[19] Hart, S. G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later, *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, pp. 904 – 908 (2006).

[20] Fujioka, E., Mantani, S., Hiryu, S., Riquimaroux, H. and Watanabe, Y.: Echolocation and Flight Strategy of Japanese House Bats during Natural Foraging, Revealed by a Microphone Array System, *The Journal of the Acoustical Society of America*, Vol. 129, No. 2, pp. 1081–1088 (2011).