

音響センシングによる宅内スマートフォン被覆状態分類の精度向上に向けた改善

西 陽也^{1,a)} 石田 繁巳² 村上 友規³ 大槻 信也³

概要: 多くのスマートフォンユーザは宅内でスマートフォンを紛失した際にスマートフォンから音を鳴らし、その音を頼りにユーザが歩いて探す。しかしこの方法はユーザの感覚に頼っており、音が反響しやすい宅内では音源定位が難しく、発見までに時間と労力が必要になる。筆者らは、スマートスピーカを用いてスマートフォンの周辺状況を推定することでスマートフォンの発見を手助けする宅内スマートフォン搜索支援システムの開発を行っている。先行研究において、周辺状況の1つである被覆状態を音響センシングによって分類する手法を示したが、平均分類精度 0.66 と実環境での利用に向けて精度の向上が課題であった。そこで、本稿では被覆物体ごとに様々な周波数における調波構造が変化することに着目し、スイープ音を用いて抽出した特徴量を用いることで被覆状態分類精度を向上させる手法を提案する。先行研究と同様の実環境で収集したスイープ音データから提案する特徴量を抽出することで、被覆状態の平均分類精度 0.94 と大幅に精度を向上させることができた。

1. はじめに

近年、宅内でスマートフォンを紛失するユーザが増加している。スマートフォンを紛失した場合、ユーザの多くは電話をかけるなどしてスマートフォンから音を発し、その音源位置を特定するように宅内を歩いて探す [1]。しかし、この方法はユーザの聴覚に頼っている点や宅内で音が反響する点などから、スマートフォンを見つけるまでに時間と労力がかかる。

筆者らは、スマートスピーカを用いた音響センシングによってスマートフォンの周辺状況を推定し、推定された周辺状況をユーザに提供することでスマートフォンの発見を手助けする宅内スマートフォン搜索支援システムを提案している [2]。図 1 に示すように、筆者らの提案する宅内スマートフォン搜索支援システムは、スマートフォンの周辺状況として被覆状態、接触物体、スマートフォンが存在する部屋の3つを同時かつ独立に推定する。被覆状態とはスマートフォンが接触している物体によってどのように覆われているかを定義した状態である。スマートフォン搜索支

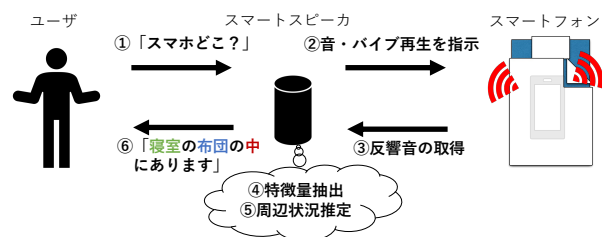


図 1: 提案システム概要図 [2]

援を必要十分に行う粒度として、本研究では被覆状態を非被覆、単一被覆、複数被覆の3つに分類する。非被覆、単一被覆、複数被覆は、それぞれ、スマートフォンが物体の上にとただ置かれているだけで何にも覆われていない状態、スマートフォンが物体の中にあるようなスマートフォンの上面と下面で同一の物体に覆われている状態、スマートフォンが物体の上に置かれており、さらにスマートフォンの上面が別の物体によって覆われている状態である。

先行研究において筆者らは被覆状態を分類する手法を示し、初期的評価によって実現可能性を示した [2]。この手法では、スマートフォンから再生された音が被覆状態によって異なる影響を受けることに着目し、スマートフォンで再生した音をスマートフォンから離れた位置にあるスマートスピーカで取得して分析することで被覆状態を分類する。音への影響は周辺物体やその状態によって定まり、主に倍音成分に表れることから、倍音成分を取り出すことで被覆状態を分類する。初期的評価では、iOS 端末に標準で搭載されている「探す」アプリのサウンド再生機能に使用され

¹ 公立はこだて未来大学大学院システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate

² 公立はこだて未来大学システム情報科学部
School of Systems Information Science, Future University Hakodate

³ 日本電信電話株式会社 アクセスサービスシステム研究所
Access Network Service Systems Laboratories, Nippon Telegraph and Telephone Corporation

a) g2123046@fun.ac.jp

る音を用い、基音と倍音のパワー比、すなわち調波構造を特徴量として抽出して教師あり学習により被覆状態を分類した。

しかし、初期的評価の平均分類精度は 0.66 であり、実環境での利用に向けて精度向上が必要である。分類精度が低かった要因として、単一被覆と複数被覆で誤認するケースが多く、単一被覆と複数被覆の分類精度が低下したことが挙げられる。初期的評価に用いた音は 1571 Hz を基音としているもののいくつかの倍音を含んでいる。被覆状態によって変化する倍音成分の大きさは元から含まれている倍音成分に比べて小さいため、単一被覆と複数被覆での調波構造の違いを抽出することが困難であったと考えられる。

これに対し、本稿では純音を用いることで被覆状態によって変化する倍音成分を特徴量として抽出し、被覆状態分類性能を向上させる手法を提案する。純音を用いた場合に観測される倍音成分は主に周辺物体の影響で生じたものであり、被覆状態による影響を多く含んでいることから被覆状態の分類性能を向上できる。被覆状態の影響を受けやすい純音の周波数は被覆物体や被覆状態によって異なる。実環境ではどのような物体に覆われているかは未知であるため、様々な周波数の純音で調波構造を特徴量として抽出する。

具体的には、周波数が時間とともに増加するスイープ音をスマートフォンから再生し、一定時間幅の window に区切って調波構造を抽出する。そして、各 window の調波構造を並べたベクトルを特徴量ベクトルとして教師あり学習により被覆状態を分類する。

提案手法の評価に向けて初期的評価と同様の環境でスイープ音を用いた被覆状態データを収集し、被覆状態分類性能を評価した。その結果、単一被覆と複数被覆を明確に分類し、平均分類精度 0.944 と着信音を用いた初期的評価に比べて大幅に精度を向上することができた。

本稿の構成は以下の通りである。2 章では音響センシングに関する研究や音波による材質推定に関する研究について述べる。3 章では被覆状態分類手法の概要と初期的評価での課題を示す。4 章では被覆状態分類の精度を向上させるスイープ音を用いた被覆状態特徴量抽出手法を示し、5 章でスイープ音による被覆状態分類の評価を行う。最後に 6 章でまとめとする。

2. 関連研究

本研究は、音響センシングに関する研究と音波による材質推定に関する研究に関連している。本章ではこれらの研究について述べる。

2.1 音響センシングに関する研究

様々な対象に対して、物体識別や状態推定のための音響センシングに関する研究が報告されている [3-8]。岩瀬らは、宅内に存在するような日常物体を対象にしたアクティブ

音響センシングによる物体識別の研究を報告している [3]。この研究ではアクティブ音響センシングによって装置上に置かれた日常的に使用する 26 個の物体を 98.9%の精度で識別している。使用音には 1.0 秒間に周波数が 20000 Hz から 40000 Hz まで一定の速さで上昇するスイープ音を使用している。特徴量として取得する周波数特性は 20000 Hz から 40000 Hz に含まれる 107 個の周波数における振幅スペクトルの最大値を用いる。物体の周波数特性を取得した特徴量を用いることで、SVM (Support Vector Machine) による評価で 26 個の物体を 98.9%の精度で識別した。この研究によって、複数周波数における調波構造の変動から物体を識別できることが示された。しかし、この研究では piezo 素子を用いて 20000 Hz 以上のスイープ音を再生・録音しているが、スマートフォンのスピーカーとスマートスピーカーに内蔵されているマイクでは 20000 Hz 以上の音を再生・録音できない。そこで本研究では、スマートフォンで十分再生できる周波数帯域のスイープ音を用いて被覆物体による調波構造の変動を抽出することを目指す。

2.2 音波による材質推定に関する研究

音波を用いて接触物体などの材質を推定する研究がいくつか報告されている [9,10]。長谷川らは、スマートフォンの置き場推定に向けて、1024 Hz の純音を用いて 6 種類の材質を 89.2%の精度で識別する手法を報告している [9]。この手法では、スマートフォンの向きや周囲に置かれている別の物体による識別への影響は小さいことが明らかになった。田中らは、ギターの違いによる音色の定量的な評価を報告している [10]。この報告では、ギターの開放弦の撥弦音に含まれる成分を音階論に基づき 12 平均律を用いて複数の周波数における基音、倍音による定量的な評価により、高級手工ギターと量産型ギターによる音の響き方による違いを分類している。田中らによると、ギターの音は、様々な高さの周波数が合成されたもので倍音の含まれ方により音色が変わる。倍音の含まれ方はギターに使用される木材の材質や形状によって変化する音のこもり方に関係すると考えられる。そこで複数の周波数における基音と倍音を用いることで各被覆状態特有の周波数変化を抽出できると考える。

物体の周波数特性は、振動モード、固有振動数、モード減衰率によって表現され、物体の形状、材質、物体への荷重や高速による物体表面の状態の 3 つの要因によって変化する [3]。そのため、被覆状態によって変化する固有振動数やモード減衰率の変化を複数周波数における基音と倍音の比として特徴量に追加することで単一被覆と複数被覆の分類が可能であると考えられる。

3. 音響センシングによるスマートフォン被覆状態分類手法

本章では先行研究 [2] で示した音響センシングによる被

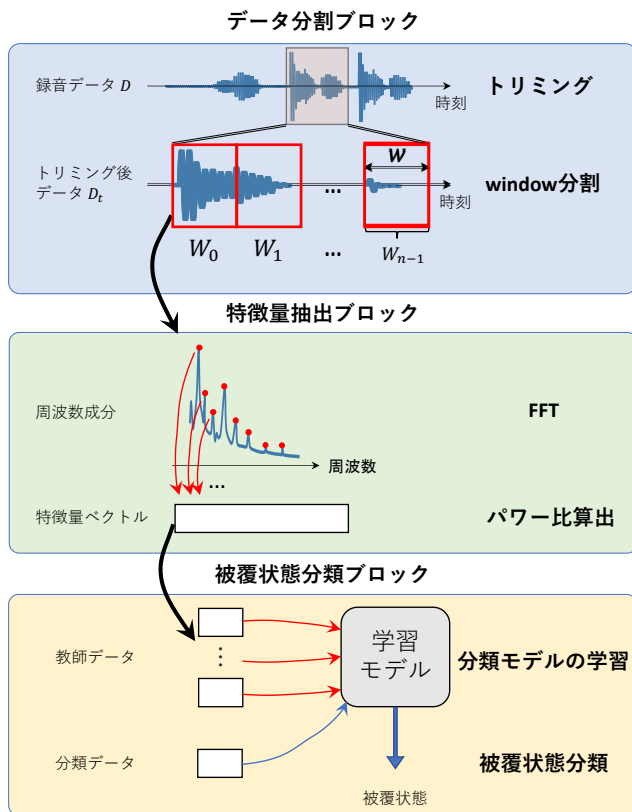


図 2: 被覆状態分類手法の概要

覆状態分類手法の初期的評価の概要と課題について述べる。

3.1 概要

本手法ではスマートフォンから再生した音をスマートフォンから離れた位置にあるスマートスピーカ内蔵のマイクで収集し、その音の調波構造によってスマートフォンの被覆状態を分類する。

図 2 に音響センシングによるスマートフォン被覆状態分類手法の概要を示す。本手法はデータ分割ブロック、特徴量抽出ブロック、被覆状態分類ブロックの 3 つで構成されている。

データ分割ブロックでは、録音データから 1 コール分の着信音データをトリミングし、一定の window 幅ごとに window として分割する。

特徴量抽出ブロックでは、FFT (fast Fourier transform) によって時間領域の着信音データを window ごとに周波数領域に変換し、各 window の調波構造から特徴量ベクトルを生成する。本手法では、音源となるスマートフォンとマイクの間距離が不明であることを想定しており、各周波数の振幅スペクトルは距離とともに減衰する。そのため、調波構造をそのまま特徴量とするのではなく、基音と倍音のパワー比を特徴量として用いることで、音源とマイクの距離による影響を軽減する。基音のパワーを P_1 [dB]、倍音のパワーを P_2, P_3, \dots, P_k [dB] とすると、基音と各倍音のパワー比は式 (1)~(3) のように算出する。

$$PR_2 = P_1 - P_2 \quad (1)$$

$$PR_3 = P_1 - P_3 \quad (2)$$

⋮

$$PR_k = P_1 - P_k \quad (3)$$

これらを並べた (PR_2, \dots, PR_k) を特徴量ベクトルとする。

被覆状態分類ブロックでは、特徴量抽出ブロックで生成した特徴量ベクトルを用いて分類モデルの学習と分類モデルによる被覆状態分類を行う。初期的評価では、再生音として基音 1571 Hz といくつかの倍音を含む着信音、学習アルゴリズムとして Random Forest を使用した。

3.2 課題

筆者が行った被覆状態分類手法の初期的評価では平均分類精度が 0.66 であり、精度向上に向けて単一被覆と複数被覆における被覆状態固有の変化を取得することが課題である。

図 3 は、初期的評価で得た特徴量ベクトルを 2 次元空間上にプロットした結果を示している。360 次元の特徴量ベクトルを t-SNE (t-Distributed Stochastic Neighbor

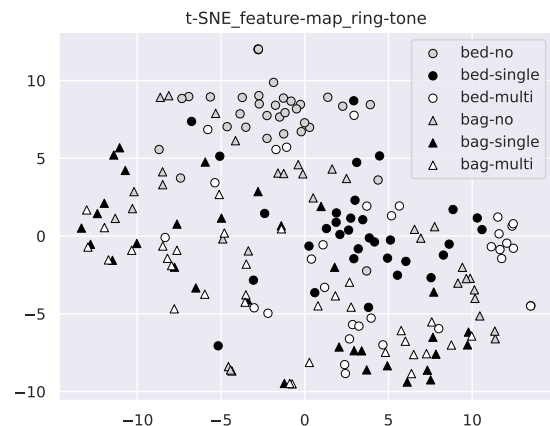


図 3: t-SNE による着信音データの特徴量空間のグラフ

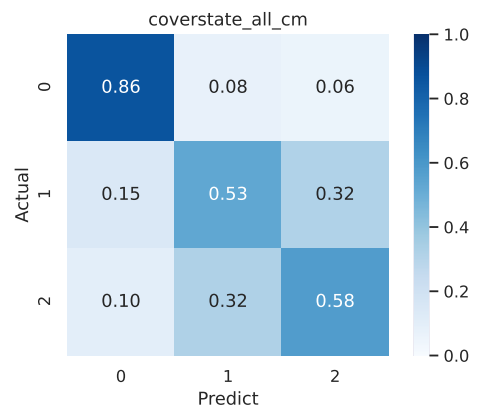


図 4: 初期的評価の評価結果 [2]

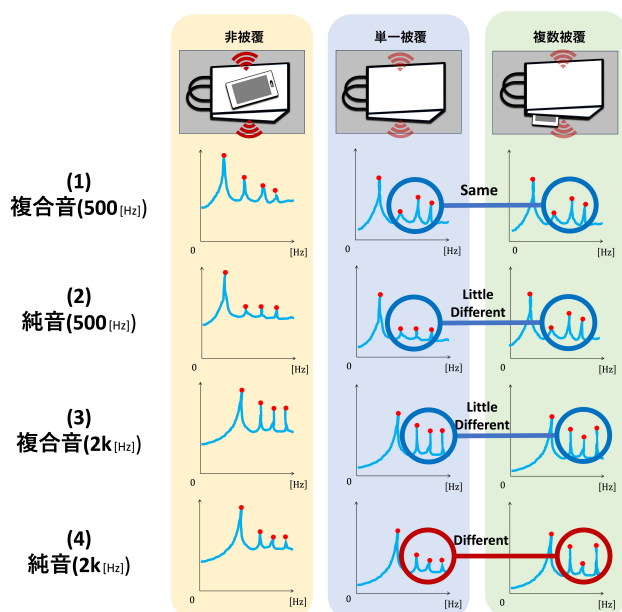


図 5: キーアイデア概要図

Embedding) によって2次元に次元削減した上でプロットしており、被覆状態と被覆物体によって異なるマークを用いている。図3より、各被覆状態の特徴量ベクトルが特徴量空間内に点在していることがわかる。被覆状態固有の特徴を十分に抽出できておらず、被覆状態以外の影響によって特徴量がばらついていると考えられる。

図4に、初期的評価で行った10分割交差検証結果の混同行列を示す。図4より、非被覆状態は精度0.86で分類できているが、単一被覆と複数被覆では0.53, 0.58と分類精度が低いことがわかる。単一被覆を複数被覆、複数被覆を単一被覆と予測するケースが多かったためであり、単一被覆と複数被覆の誤分類を削減する必要がある。

分類精度が低かった原因として、特徴量として用いる調波構造に着信音に元から含まれている倍音成分を含んでいることが挙げられる。被覆状態によって変化する倍音成分は着信音に元から含まれている倍音成分に比べて小さい。このため、被覆状態によって変化する倍音成分が元から含まれる倍音成分に埋もれてしまい、被覆状態によって変化する調波構造変動のみを抽出することが困難となる。

4. スイープ音による被覆状態特徴量抽出手法

前章で示した被覆状態分類手法の課題を解決し、被覆状態分類精度を向上させるため、本章では被覆状態分類のための新たな特徴量抽出手法として、スイープ音による被覆状態特徴量抽出手法を示す。

4.1 キーアイデア

本手法のキーアイデアは、様々な周波数の純音でデータを収集し、被覆物体・被覆状態によって変化する調波構造を特徴量として抽出することである。倍音を含まない純音を用いることで、主に周辺物体・周辺状況の影響で生じる

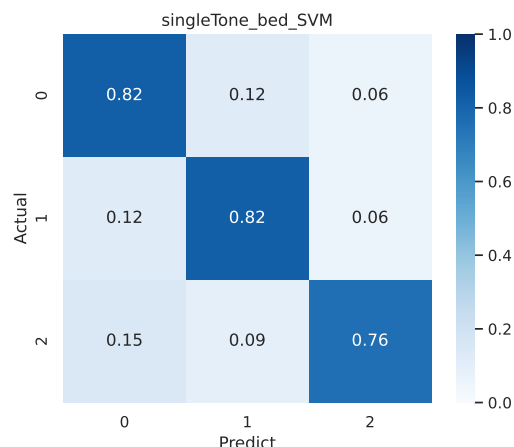


図 6: 440 Hz 純音による評価結果

倍音成分を調波構造という特徴量として抽出する。被覆物体によって固有振動数は変化し、様々な周波数帯域での吸音・透過作用が異なるものとなることから、被覆物体やその状態の差は観測した音の調波構造の差として表れる。そこで、様々な周波数での調波構造を特徴量とすることで被覆状態の分類精度を向上させる。

図5にキーアイデアの概要図を示す。(1) スマートフォンから500 Hzの複合音を再生し、その音を観測した場合、単一被覆と複数被覆間で調波構造には大きな差が見られず、これら2つの状態の区別は困難である。そこで、(2) 500 Hzの純音を用いると、元の音に倍音が含まれていないことから調波構造の小さな変動から被覆状態の区別が可能となる。単一被覆と複数被覆での調波構造の差が大きく表れる周波数は被覆物体やその状態によって異なる。このため、500 Hzで単一被覆と複数被覆間での差がわずかである場合でも、(3), (4)のように2 kHzを用いる場合に大きな差が表れる、ということが起こる。図5では500 Hzと2 kHzという2つの周波数での調波構造を考えているが、実際には様々な周波数での調波構造を抽出して単一被覆と複数被覆を区別する。

純音によって被覆状態分類性能が向上することは予備的実験によって確認している。図6に、純音による単一被覆と複数被覆の分類精度向上の実現可能性を調べた評価結果の混同行列を示す。440 Hzの純音を1秒間再生し、被覆状態を変えて各状態33試行、合計99試行のデータを収集した。1試行あたり12個の倍音と基音のパワー比を20 window分、合計240次元を特徴量として被覆状態分類を行った。図6に示すとおり、純音を用いることで単一被覆と複数被覆を約8割程度の精度で分類できた。

どのような周波数で被覆状態による調波構造の変化を観測しやすいかは被覆物体やその状態によって異なる。そこで、周波数が時間とともに増加するスイープ音をスマートフォンから再生し、一定時間幅のwindowに分割することで周波数ごとの調波構造を抽出する。

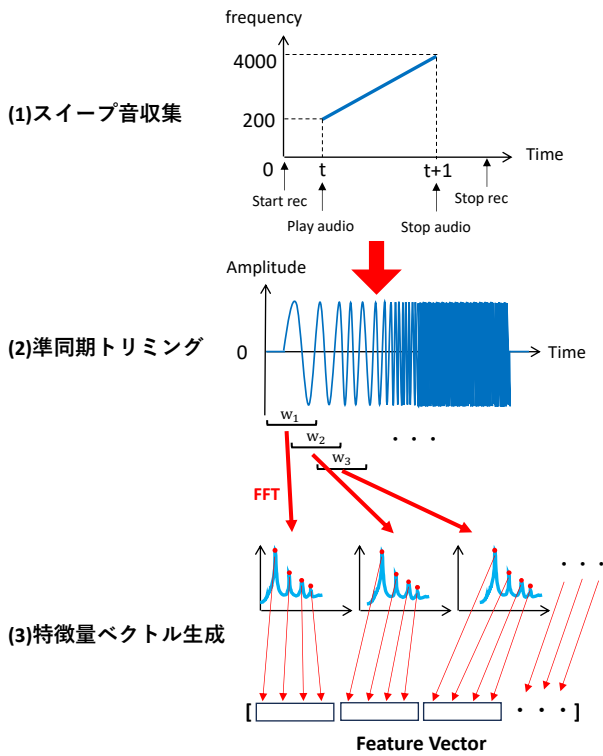


図 7: 特徴量抽出手法概要図

4.2 概要

図 7 にスイープ音による被覆状態特徴量抽出手法の概要を示す。本手法では、(1) スイープ音収集ステップ、(2) 準同期トリミングステップ、(3) 特徴量ベクトル生成ステップの 3 つのステップで、図 2 に示した被覆状態分類手法の被覆状態分類ブロックで用いる特徴量を抽出する。まず、(1) のスイープ音収集ステップで、一般的なスマートフォンで再生可能な 200 Hz から 4000 Hz のスイープ音を再生し、スマートスピーカ内蔵マイクで録音する。次いで、(2) の準同期トリミングステップでは、録音と再生タイミングなどによる同期ずれを許容するためにおおよその同期を行ってトリミングをし、window に分割する。最後に、(3) の特徴量ベクトル生成ステップにおいて、各 window で FFT を行い、各 window の基音と倍音のパワー比を得て特徴量ベクトルを生成する。生成した特徴量ベクトルを用い、図 2 の被覆状態分類ブロックによって被覆状態进行分类する。

以降では各ステップについて詳述する。

4.3 スイープ音収集ステップ

スイープ音収集ステップでは、複数周波数における音データ取得のために、スイープ音の再生・録音を行う。スイープ音は一般的なスマートフォンのスピーカで再生できる 200 Hz から 4000 Hz までの周波数範囲を一定の速さで上昇する純音を用いる。本稿ではスイープ音の再生時間は 1 秒間とした。本手法では音の再生と録音を別々のデバイスで行うことを想定しているため、録音と再生のタイミングを厳密に同期することが難しい。そこで、図 7 に示すと

おり録音開始から t 秒後にスイープ音の再生と再生停止を確認してから録音の停止を行い、準同期トリミングステップでスイープ音部分をトリミングした。

4.4 準同期トリミングステップ

準同期トリミングステップでは、window 幅、オーバーラップ率を設定し、収集したスイープ音データをオーバーラップをさせながら一定時間幅の window に分割する。スイープ音の再生時間を T 、スイープ音の周波数帯域幅を B 、サンプリング周波数を f_s 、ウィンドウ幅を T_w 、オーバーラップ率を γ とすると、1 window あたりの周波数帯域幅 B_w 、隣接 window との周波数差 ΔB_w は、それぞれ式 (4)、(5) で表される。

$$B_w = \frac{T_w B}{T f_s} \quad (4)$$

$$\Delta B_w = (1 - \gamma) \frac{T_w B}{T f_s} \quad (5)$$

式 (4)、(5) より、window 幅 T_w を小さく、オーバーラップ率 γ を大きくすれば周波数帯域幅 B_w 、隣接 window との周波数差 ΔB_w を小さくでき、周波数分解能を高くすることができる。しかし、実環境では、収集データを完全に同期できないため、1 window あたりの周波数帯域幅と隣接 window との周波数帯域差が小さいと、データ間で各 window で扱う周波数帯域にずれが生じてしまう。このため、周波数分解能と周波数の同期にはトレードオフがあると言える。window 幅とオーバーラップ率の設定の影響は評価において検証する。

4.5 特徴量ベクトル生成ステップ

特徴量ベクトル生成ステップでは、初期的評価と同様に window ごとに FFT を行い、window 内で振幅スペクトルが最大である音を基音とし、その基音の倍音のパワー比を特徴量として特徴量ベクトルを生成する。本手法では、初期的評価と異なり、window ごとに含まれる周波数帯域が異なるため、各 window で基音と倍音が異なる。

本研究では、スマートスピーカ内蔵マイクによる録音を想定しており、サンプリング周波数を 44100 Hz に設定しているためナイキスト周波数は 22050 Hz である。スイープ音の最大周波数 4000 Hz では 4 次倍音 20000 Hz までしか扱うことができないため、すべての window で 4 次倍音までを抽出して特徴量ベクトルを生成する。

5. 評価

本章では、スイープ音のデータ収集実験とスイープ音による被覆状態分類性能の評価、トリミングと学習モデルによる影響について述べる。まず、スイープ音を用いたデータ収集実験について述べる。その後、スイープ音による被覆状態分類の性能評価およびトリミング、学習モデルによ

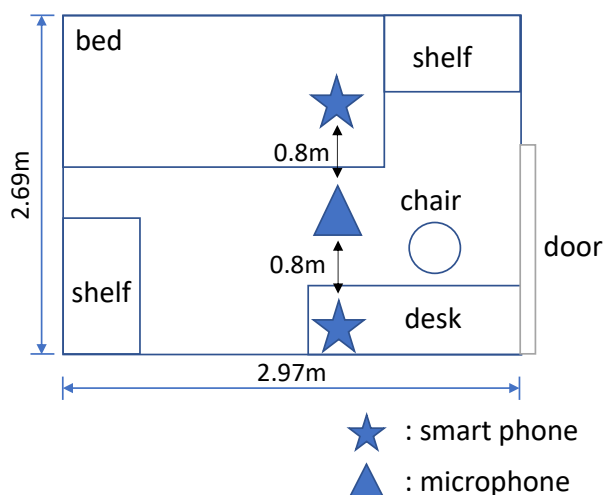


図 8: 実験部屋の間取り図 [2]

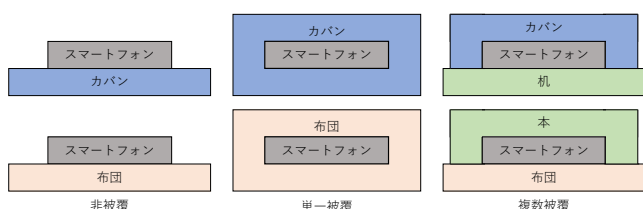


図 9: 被覆状態イメージ図 [2]

る影響について述べる。本研究では、10 分割交差検証を 100 回行い混同行列を生成し、被覆状態ごとの F 値を分類精度として評価する。

5.1 実験環境

データ収集実験は初期的評価と評価結果を比較するために、初期的評価でデータを収集した環境と同様の環境で行った。図 8 に実験に使用した部屋の間取り図を示す。実験に使用した部屋は 1LDK 住宅の寝室で、図 8 の三角印の位置にマイクロフォン、マイクロフォンから約 0.8m 離れた 2 箇所の星印のいずれかにスマートフォンと被覆物体を設置した。初期的評価と同様に一般的なスマートスピーカー内蔵マイクを想定して、サンプリング周波数は 44100 Hz、量子化ビット数は 16 bit とした。マイクロフォンは audio-technica AT2050、スマートフォンは ASUS Zenfone 8、オーディオレコーダは ZOOM H6 を使用した。AT2050 マイクロフォンは指向性を持たせずに全方向設定とした。実験に使用した被覆物体は初期的評価と同様に布団とカバンの 2 つを使用した。

環境ノイズ低減のために以下の環境でデータを収集した。

- ドア、窓、カーテンは閉め切る
- 浴室の 24 時間換気は消す
- エアコンは停止する
- 家具など部屋内の配置は変更しない
- 屋外から街宣車や人の話し声などが聞こえる場合は実験を一時中断する

同じパターン of データを連続で収集することを防ぐために、1 試行ごとに被覆状態を変えてデータを収集し、99 試行終わった段階で被覆物体を変えてデータを収集した。被覆状態は非被覆、単一被覆、複数被覆の順でデータを収集した。初期的評価ではスマートフォンの向きも変えながらデータを収集した。しかし、関連研究 [9] や初期的評価の評価結果でもスマートフォンの向きによる分類精度への影響は見られなかったため、本実験ではスマートフォンの向きをマイク方向にしてデータを収集した。

データの収集手順を以下に示す。

- (1) 録音を開始する
- (2) 布団の被覆状態を変えてスマートフォンを設置する
- (3) 1 秒間スイープ音を再生する
- (4) (2)(3) を 99 試行行う
- (5) カバンの被覆状態を変えてスマートフォンを設置する
- (6) 1 秒間スイープ音を再生する
- (7) (4)(5) を 99 試行行う
- (8) 録音停止する

図 9 に、各被覆物体・被覆状態の設置イメージ図を示す。被覆物体が布団の場合、非被覆では布団の上にスマートフォンのみを置いた。被覆では非被覆と同様に設置した後、スマートフォンを覆うように布団をかけた。複数被覆では非被覆と同様に設置した後、スマートフォンを覆うように上から本を置いた。被覆物体がカバンの場合、非被覆では机の上にカバン、スマートフォンの順に設置した。単一被覆ではカバンの中にスマートフォンを入れてカバンの開口部を閉じた。複数被覆では机の上にスマートフォン、カバンの順に設置した。

5.2 被覆状態分類性能

被覆状態分類性能を評価するために、トリミングには window 幅を 4096、オーバーラップ率を 0.50、窓関数をブラックマン窓に設定し、初期的評価と同様に学習モデルには Random Forest を用いた。

図 10 に評価に用いる被覆物体データごとの混同行列を示す。図 10 に示す混同行列のラベルは各被覆状態を示しており、0 が非被覆、1 が単一被覆、2 が複数被覆である。すべてのデータを用いた評価の分類精度は、平均で 0.925、非被覆で 0.98、単一被覆で 0.91、複数被覆で 0.88 であった。布団のデータを用いた評価の分類精度は、平均で 0.991、非被覆で 1.00、単一被覆で 1.00、複数被覆で 0.97 であった。カバンのデータを用いた評価の分類精度は、平均で 0.917、非被覆で 0.97、単一被覆で 0.90、複数被覆で 0.88 であった。図 4 に示す初期的評価での着信音による評価に比べて、いずれの被覆物体でも分類精度は大幅に向上した。単一被覆と複数被覆においても分類精度約 0.90 を超えており、初期的評の課題であった単一被覆と複数被覆の分類について解決できた。

図 11 に被覆状態分類の性能評価に用いる特徴量ベクト

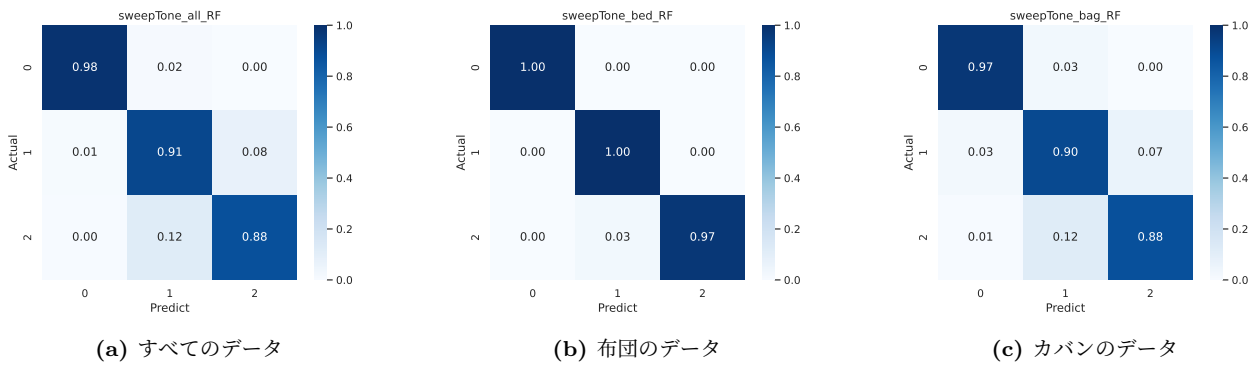


図 10: スイープ音データの混同行列

表 1: トリミングパラメータごとの分類精度

| window 幅 | 1024 | | | 2048 | | | 4096 | | | 8192 | | |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| オーバーラップ率 | 0 | 0.50 | 0.75 | 0 | 0.50 | 0.75 | 0 | 0.50 | 0.75 | 0 | 0.50 | 0.75 |
| 非被覆 | 0.98 | 0.98 | 0.98 | 0.98 | 0.97 | 0.98 | 0.98 | 0.98 | 0.96 | 0.96 | 0.94 | 0.97 |
| 単一被覆 | 0.82 | 0.90 | 0.89 | 0.90 | 0.90 | 0.96 | 0.86 | 0.92 | 0.87 | 0.78 | 0.88 | 0.86 |
| 複数被覆 | 0.84 | 0.92 | 0.91 | 0.91 | 0.93 | 0.91 | 0.86 | 0.94 | 0.89 | 0.80 | 0.87 | 0.86 |
| 平均分類精度 | 0.884 | 0.934 | 0.929 | 0.929 | 0.934 | 0.929 | 0.899 | 0.944 | 0.909 | 0.848 | 0.899 | 0.894 |

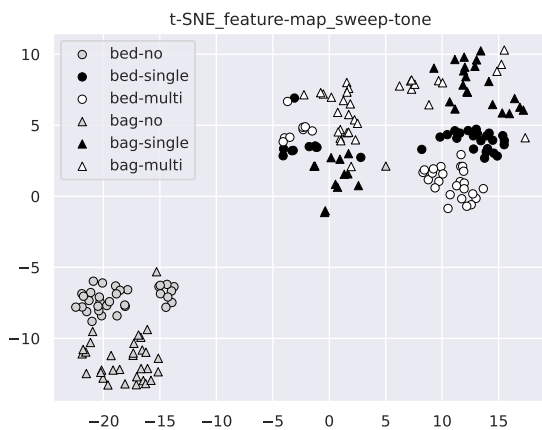


図 11: t-SNE による特徴量空間のグラフ

ルを2次元空間上にプロットした結果を示す。84次元の特徴量ベクトルをt-SNEによって2次元に次元削減した上でプロットしており、被覆状態と被覆物体によって異なるマークを用いている。図11より、非被覆状態は他の被覆状態と明らかに離れており、分類が容易であると推察される。単一被覆、複数被覆も被覆物体でグループが分かれているものの、被覆状態ごとにみればグループ化できており、分類が容易であると推察される。これは、スイープ音に複数の周波数が含まれていることで、各被覆状態で固有の変化が現れる周波数帯の音を特徴量として取得することができていることが要因と考えられる。単一被覆と複数被覆間の分類においても被覆物体による固有の変化を取得できていることが示された。したがって、音響センシングによる被覆状態分類にはスイープ音によるデータの有効性が

示された。

5.3 トリミングの影響

トリミングによる影響を分析するために、4つのwindow幅と3つのオーバーラップ率の組み合わせを変えて評価した。window幅は1024, 2048, 4096, 8192で評価した。オーバーラップ率は、0, 0.50, 0.75の場合を用いて評価した。評価にはスイープ音によるデータを、窓関数にはブラックマン窓を、学習アルゴリズムにはSVMを用いた。

表1に布団とカバンのすべてのデータをwindow幅とオーバーラップ率ごとにSVMで学習・評価した結果を示す。評価の結果、最も分類精度が高かった組み合わせは平均分類精度0.944でwindow幅4096とオーバーラップ率0.50の組み合わせであった。したがって、本実験環境ではwindow幅4096, オーバーラップ率0.50の組み合わせがwindow幅とオーバーラップ率のトレードオフを最小化し、分類精度を高くできることが示された。

5.4 学習モデルの影響

学習モデルによる影響を分析するために、初期的評価で用いたRandom Forestの他に、SVM, ロジスティック回帰を用いて評価した。トリミングではwindow幅を4096, オーバーラップ率を0.50, 窓関数をブラックマン窓に設定した。

図12に学習モデルごとの評価結果を示す。図12に示す混同行列のラベルは各被覆状態を示しており、0が非被覆, 1が単一被覆, 2が複数被覆である。Random Forestを用いた評価の分類精度は、平均で0.924, 非被覆で0.99, 単一被覆で0.89, 複数被覆で0.90であった。SVMを用い

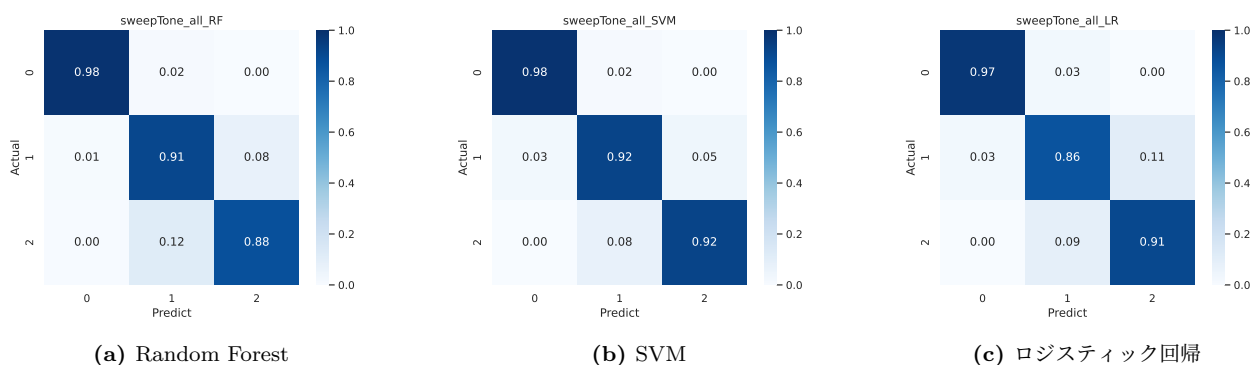


図 12: 学習モデルごとの混同行列

た評価の分類精度は、平均で 0.944、非被覆で 0.98、単一被覆で 0.92、複数被覆で 0.94 であった。ロジスティック回帰を用いた評価の分類精度は、平均で 0.914、非被覆で 0.97、単一被覆で 0.87、複数被覆で 0.90 であった。SVM を用いた評価が他の学習モデルによる評価に比べて、全ての被覆状態で分類精度が高かった。SVM による評価の分類精度が高かった要因として、3つのうち2つの被覆状態間ごとの周波数特性の違いを学習できたことが考えられる。これは SVM が 3 値分類を 2 値分類を 3 回行うことで求めており、各被覆状態間ごとに分類の基準となる境界線を引くことができるからであると推測される。したがって、被覆状態分類には SVM を用いた学習が最も分類精度が高くなることが示された。

6. おわりに

本稿では、宅内スマートフォン検索支援システムの実現に向けて、音響センシングによる宅内スマートフォン被覆状態分類精度向上に向けた被覆状態特徴量抽出手法を提案した。本稿の目的は筆者らが行った被覆状態分類の初期的評価 [2] において 0.66 であった被覆状態分類精度を向上させることであった。そこで本研究では被覆物体ごとに様々な周波数における調波構造が変化することに着目し、複数周波数における調波構造変動を抽出するため、スイープ音を用いた被覆状態特徴量抽出手法を提案した。トリミングによる影響分析では、トリミング時に設定する window 幅とオーバーラップ率のトレードオフを最小化するパラメータを明らかにした。学習モデルによる影響分析では、Random Forest の他に SVM やロジスティック回帰を用いて評価し、評価結果の比較によって被覆状態分類に最適な学習アルゴリズムを明らかにした。評価実験より、スイープ音によるデータをオーバーラップ率 0.50 でブラックマン窓を用いてトリミングを行い、SVM を用いて学習することで、平均分類精度 0.944 と提案手法の有用性が示された。

今後の展望として、宅内スマートフォン検索支援システムの構築に向けて、接触物体、スマートフォンが存在する部屋の推定を行う予定である。

謝辞 本稿で示した研究の一部は、JSPS 科研費 (JP21K11847) および東北大学電気通信研究所共同プロジェクト研究の助成で行われた。

参考文献

- [1] TrackR: 探し物に関する調査. <https://prtimes.jp/main/html/rd/p/000000006.000022312.html> (2017).
- [2] 西 陽也, 石田繁巳, 村上友規, 大槻信也: 宅内でのスマートフォン検索に向けた音響センシングによる被覆状態分類手法, 情報処理学会研究報告, MBL 研究会, Vol. 2023-MBL-106, pp. 1-7 (2023).
- [3] 岩瀬大輝, 伊藤雄一, 秦 秀彦, 尾上孝雄: SenseSurface: アクティブ音響センシングによる物体識別と位置推定, 情報処理学会論文誌, Vol. 60, No. 10, pp. 1869-1880 (2019).
- [4] Joya, T., Ishida, S., Mitsukude, Y. and Arakawa, Y.: Design of Room-Layout Estimator Using Smart Speaker, *Mobile and Ubiquitous Systems: Computing, Networking and Services. MobiQuitous 2021*, Vol. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST), Springer, pp. 24-39 (2021).
- [5] Song, X., Yang, B., Yang, G., Chen, R., Forno, E., Chen, W. and Gao, W.: SpiroSonic: Monitoring Human Lung Function via Acoustic Sensing on Commodity Smartphones, *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, Vol. 2020-MobiCom-26, pp. 1-14 (2020).
- [6] Lu, L., Yu, J., Chen, Y., Liu, H., Zhu, Y., Kong, L. and Li, M.: Lip Reading-Based User Authentication through Acoustic Sensing on Smartphones, *IEEE/ACM transactions on networking*, Vol. 27, No. 1, pp. 447-460 (2019).
- [7] Yu, J., Lu, L., Chen, Y., Zhu, Y. and Kong, L.: An Indirect Eavesdropping Attack of Keystrokes on Touch Screen through Acoustic Sensing, *IEEE Transactions on Mobile Computing*, Vol. 20, No. 2, pp. 337-351 (2019).
- [8] 金澤玲実, 石田繁巳, 白石 陽: スマートフォン内蔵マイクを用いた歩道路面状況推定における汎用性の向上に向けた改善, 第 30 回マルチメディア通信と分散処理ワークショップ論文集, pp. 105-111 (2022).
- [9] 長谷川達人, 平橋智史, 越野 亮: 音の反響を用いたスマートフォン置き場の材質推定, マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, Vol. 2016 DICOMO, pp. 1691-1698 (2016).
- [10] Tanaka, H. and Itako, K.: ギターの音色の FFT 解析を用いた新しい評価方法 (オクターブ係数) の検討, *Journal of Advanced Science*, Vol. 33, p. 33101 (2021).