

反響音の組み合わせを用いたスマートフォンの置き場の材質推定

長谷川 達人†
東京医療保健大学
医療保健学部医療情報学科†

平橋 智史††
金沢大学大学院
自然科学研究科††

越野 亮†††
石川工業高等専門学校
電子情報工学科†††

1. はじめに

本研究ではコンテキストウェアネス研究の一環として、スマートフォンの置き場所推定手法を提案する。Lopez-Tovar らの研究[1]で、スマートフォンがポケットにあるときにはバイブで、机上にあるときには LED 点灯で通知して欲しいと思う人が多いことが示されている。机のような硬い材質の上と、布団のような柔らかい材質の上に置かれた場合でも、通知の聞こえ方やバイブの振動音が変わる。このように、スマートフォンの置き場所や、その材質に応じて通知手法が変更されることが望ましい。

停止時に加速度はほぼ変動しないため、加速度センサを用いた識別手法は本研究には適さない。そこで、スマートフォンから発した音の反響を用いた置き場所推定手法を提案する。スマートフォンが置かれることが多いと思われる鞆の中、ポケットの中、机の上、布団の上、スマートフォンスタンドの上、手持ちを対象に、提案手法を評価する。

2. 提案手法

提案する処理手順では、図1のように対象物の上に置いたスマートフォンから特定の周波数の音を鳴らし、同時に周囲の音を記録する。記録された音を一定のウィンドウに区切り、フーリエ変換する。変換後、対象物の特性を示す特徴量を抽出し、機械学習によって対象物を識別する。反響音をもつ情報を解析することで、対象物の材質など高次の情報が識別可能であると考えられる。観測する反響音は、対象物に限らず壁や周囲の物体からの反響を含んでいるが、最も近くにある対象物の影響を強く受けると考えられる。一般的なスマートフォンではサンプリング周波数 44.1kHz が上限となるため、本研究では可聴域のビーブ音を音源として採用する。

事前検証[2]では 1024Hz のビーブ音を音源として提案手法の有効性を確認している。18 種類の対象物で実験を行ったところ、(1)カバンやポケットの中、(2)柔らかいものの上、(3)硬いものの上、(4)ゴムマットの上、(5)スマートフォンスタンドの上のグループを高い精度で識別できる可能性が示唆された。また、端末の向きや周囲環境の影響が大きいことも示した。

本稿では、より詳細な分析を行う。事前検証では音量の大きさ、周囲の雑音を一定にして実

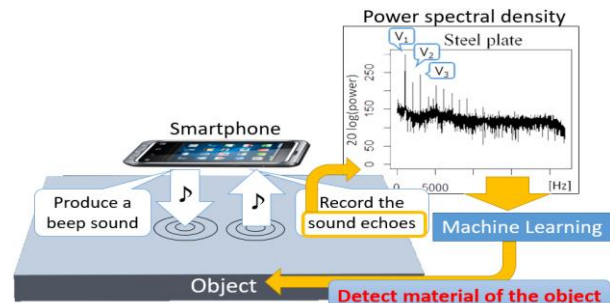


図1. 提案手法の概要

験を行った。一方、Diaconita らの研究[3]で、ホワイトノイズを用いた手法も提案されており、音の種類の変更や複数音の組み合わせを行うことで推定精度が改善できる可能性がある。そこで、様々な音の種類や大きさ、組み合わせ、雑音の影響も踏まえて再実験を行い、スマートフォンの置き場所推定に有効な手法を検討する。関連研究[3]のホワイトノイズとの比較も行い、本研究の有用性を示す。

3. 実験方法

提案手法を評価するため、反響音を計測する実験を実施した。実験では、一回の計測で 3 秒間ビーブ音を鳴らし、同時に周囲音を記録する。記録したデータから 2000 サンプルシフトしつつ FFT 用の 4096 サンプルを 50 個抽出する。FFT 結果から特徴量を抽出し、1 インスタンスとする。即ち、約 0.1 秒の音データを 1 インスタンスとする。一回の計測で得られる 50 インスタンスを 1 データセットと呼ぶ。今回、12 種類(ズボン前ポケット、鞆、タオル、クッション、銅板、木の板、ゴムマット 2mm & 5mm、スマホスタンド置型 & はさみ型、手持ち 2 名)の対象物に対して、スマートフォン表向き・裏向き、周囲に物が無い状態・ある状態、周囲に雑音がない状態・ある状態で実験を行う。4800 インスタンスが生成される実験を 1 セットとして、音量 3 種類(Android の 3(Low), 8(Middle), 13(High)), 音源 10 種類(周波数 200Hz, 300Hz, 500Hz, 750Hz, 1kHz, 2kHz, 2.5kHz, 3kHz, 5kHz とホワイトノイズ(WN))で計測した。騒音の有無に関しては、小野測器社製の高機能型騒音計 LA-5560 を用いて、実験環境の観測対象周辺にて騒音の計測を行い、無音状態を雑音なし(25dB)、スピーカから NHK CREATIVE LIBRARY の環境音「街角 2」を流した状態を雑音あり(50dB)とした。

抽出する特徴量は、事前検証[2]にて重要度が高かった P1 から P10(パワースペクトルの基本周波数と高調波成分の各ピーク値)と PR12 から PR 35(ピーク間の比)の計 22 次元を採用した。ホワイトノイズの場合、基本周波数と高調波成分が特

Determining a smartphone's placement by material detection using combined sound echoes

† Tatsuhito Hasegawa: Division of Healthcare Informatics, Faculty of Healthcare, Tokyo Healthcare University

†† Satoshi Hirahashi: Graduate School of Natural Science & Technology, Kanazawa University

††† Makoto Koshino: Department of Electronics and Information Engineering, Ishikawa National College of Technology

表 1. 対象物 12 種類に対する音源別の推定精度比較 (Accuracy)

Vol.\Freq. [Hz]	200	300	500	750	1000	2000	2500	3000	5000	WN
Low	49.1%	45.4%	52.2%	61.3%	54.3%	39.4%	35.6%	39.8%	30.0%	52.6%
Middle	46.6%	53.2%	62.9%	61.6%	51.2%	42.6%	34.8%	35.7%	23.5%	57.6%
High	54.4%	66.2%	59.5%	51.8%	49.9%	38.2%	29.6%	30.5%	14.7%	56.6%

表 2. チューニング後の提案手法 (Middle750Hz & High300Hz の組み合わせ) の推定結果クロス表

Pred. \ Real	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	Precision
(a) bag	86.5%	0.3%	0.8%	0.3%	11.5%	0.0%	0.3%	0.0%	0.3%	0.5%	0.5%	0.0%	85.9%
(b) pocket	0.3%	96.0%	0.3%	0.5%	0.0%	0.0%	0.0%	0.0%	1.0%	0.5%	0.0%	0.0%	97.5%
(c) cushion	1.0%	1.0%	93.0%	1.5%	0.0%	0.0%	0.3%	0.5%	0.0%	1.3%	0.8%	6.8%	87.7%
(d) towel	1.0%	0.3%	1.0%	82.5%	2.0%	0.5%	0.3%	0.0%	0.8%	1.0%	0.5%	0.3%	91.7%
(e) gommat2	0.5%	0.0%	0.3%	12.8%	37.0%	6.5%	24.0%	2.3%	0.5%	0.3%	0.3%	0.5%	43.7%
(f) gommat5	0.3%	0.0%	0.0%	0.8%	22.3%	75.8%	14.3%	15.0%	0.5%	0.5%	0.8%	1.0%	57.8%
(g) copper	0.0%	0.0%	0.3%	0.0%	23.3%	13.3%	58.0%	2.0%	0.3%	11.5%	0.3%	0.8%	53.0%
(h) wood	9.0%	0.3%	0.5%	0.0%	1.5%	1.3%	1.3%	77.8%	0.0%	0.0%	0.0%	1.8%	83.4%
(i) personA	0.3%	1.5%	0.3%	0.0%	1.0%	0.5%	0.0%	0.5%	55.3%	19.3%	0.0%	1.0%	69.5%
(j) personB	0.5%	0.0%	0.8%	0.3%	0.3%	0.8%	0.5%	0.5%	29.0%	54.5%	0.8%	0.3%	61.9%
(k) smaA	0.5%	0.8%	2.0%	0.3%	0.5%	0.5%	1.0%	0.5%	6.0%	1.0%	88.8%	12.3%	77.9%
(l) smaB	0.3%	0.0%	1.0%	1.3%	0.8%	1.0%	0.3%	1.0%	6.5%	9.8%	7.5%	75.5%	72.1%
Recall	86.5%	96.0%	93.0%	82.5%	37.0%	75.8%	58.0%	77.8%	55.3%	54.5%	88.8%	75.5%	73.4%

定できないため、音声認識分野でよく用いられるメル周波数ケプストラム係数(12次元), と線形予測符号係数(33次元)を追加採用した。

全データで 10-fold CV を実施すると、同じ環境で計測したデータが学習データに含まれ、推定精度がほぼ 100% となってしまう。そこで、1 回の計測で生成される 1 データセットをテストデータ、残りを学習データとする、Leave One Dataset Out Cross-Validation (LODO-CV) にて、提案手法を評価する。

4. 実験結果

音源別の推定精度 (Accuracy) を表 1 に示す。学習アルゴリズムはランダムフォレストを用い、木の数を 50、特徴量選択数を特徴次元数の平方根とした。音量 High, 周波数 300Hz (High300Hz) の音源で 66.2% と最も高い推定精度になった。関連研究で提案されているホワイトノイズ (Middle WN で 57.6%) や事前研究の 1024Hz 付近 (Low1kHz で 54.3%) よりも精度が 10% 程度高くなった。音源周波数は 300Hz から 750Hz の低周波数帯で精度が高くなる傾向が見られた。音量が大きいくほど、精度が高いようにも見えるが、今回の結果からは読み解けなかった。音源を 2 種類組み合わせの場合 (30 種類組み合わせ 465 パターン) の精度を確認したところ、Middle750Hz & High300Hz の組み合わせで最も高い 70.9% の精度となった。パラメータチューニングの結果、木の数 500、特徴量選択数 20 で 73.4% の精度となった。

チューニング後の推定結果のクロス表を表 2 に示す。誤りパターンを見ると、(e) (f) ゴムマット 2mm&5mm と (g) 銅板間、(i) (j) 手持ち 2 名間で誤りが多く発生している。ゴムマットの厚みや手持ち 2 名間の差までは識別できないものと思われる。似たものをグループ化 (a&b, c&d, e&f, g&h, i&j, k&l を合併) すると、6 種類の置き場所を 82.0% の精度で識別できた。それでも、ゴムマットと銅板で誤りが多く、硬いものとゴムマットは識別が難しいという結果になった。

表 2 の 12 種類に対する推定結果を、状態毎に分離し精度を確認したところ、表向きのみで 76.8%、裏向きのみで 70.0%、周囲に物がない状態のみで 70.9%、ある状態のみで 75.8%、雑音なしのみで 66.1%、雑音ありのみで 80.7% となった。表裏向きや、周囲の物の有無の間ではそれぞれ 5% 程度の精度差となったが、雑音ありの場合が雑音なしの場合に比べ 15% 程度精度が高くなるという結果になった。

5. おわりに

本研究では、音の反響を用いたスマートフォン停止時の置き場所を推定する手法を開発し推定精度の評価を行った。実験の結果、事前研究で用いた 1024Hz のビーブ音や、関連研究で用いたホワイトノイズよりも、High300Hz の音源を用いることで 10% 程度精度が向上することを明らかにした。また、音源を複数組み合わせる手法を新たに提案し、精度の高くなる音源の組み合わせを分析した結果、Middle750Hz と High300Hz の組み合わせで精度が最も高くなり、パラメータチューニングの結果、最終的に 12 種類の置き場所を 73.4% の精度で識別することに成功した。単一の音源 (High300Hz) を用いる場合に比べて、さらに 7% 程度推定精度の向上に成功した。今後は学習データにない置き場所に対する対応や、スマートフォンの機種の影響などを考察していく。

参考文献

- [1] H.Lopez-Tovar, et al., Managing Smartphone Interruptions Through Adaptive Modes and Modulation of Notifications, Proc. of the IUI2015, pp.296-299, 2015.
- [2] T.Hasegawa, et al., Determining a Smartphone's Placement by Material Detection Using Harmonics Produced in Sound Echoes, Proc. of the MobiQuitous2016, pp.246-253, 2016.
- [3] I.Diaconita, et al., Do you hear what I hear? Using acoustic probing to detect smartphone locations, Proc. of the PERCOM Workshops 2014, pp.1-9, 2014.