

アクティブ音響センシングを用いた把持状態認識

大野 誠¹ 志築 文太郎² 田中 二郎²

概要: これまでに、物体の把持状態を認識し、HCI への応用を試みる研究が数多く行われている。一方で、それらの研究は大量のセンサや特殊なハードウェア構成を必要とするため、複雑もしくは高コストである。本稿では、アクティブ音響センシングにより、手軽かつ安価に物体の把持状態認識を行う手法を示す。本手法の特徴は1組のスピーカとマイクを物体に貼り付けることによって、物体を把持する手の姿勢、及び把持する力の認識を可能とする点にある。携帯情報端末の操作に本手法を用いることを想定した実験を行った結果、7種類の把持姿勢の認識精度が per-user test において 90~99%、cross-user test において 66%となった。また、3段階の把持力の認識精度が per-user test において 95~100%、cross-user test において 81%となった。

Grasp Recognition using Active Acoustic Sensing

MAKOTO ONO¹ BUNTAROU SHIZUKI² JIRO TANAKA²

Abstract: This paper presents a simple and low-cost grasp recognition technique using active acoustic sensing. The technique requires a pair of a speaker and a microphone as its equipment for sensing. We conducted an experiment to measure the recognition rate of our technique. As its results, per-user recognition accuracy with seven grasping postures and with three grasping strengths were 90-99% and 95-100%, respectively. Cross-user recognition accuracy with the seven grasping postures and with the three grasping strengths were 66% and 81%, respectively.

1. はじめに

物体を掴む手の姿勢及び把持する力（以降、把持状態）が場合に応じて異なることに着目し、これを認識してHCIへの応用を試みる研究が数多く行われている。例えば携帯情報端末（以降、端末）を使用する際、電話することを目的とした場合と、メール作成を目的とした場合によって把持状態は異なる。この例における把持状態の差は主に操作の目的に起因するが、その他にも様々な要素が把持状態に影響を与える。Wimmerらは把持状態に影響する要素を、Goal, Relationship, Anatomy, Setting, Propertiesの5種類に分類している [17]。これらの要素は対象の把持状態が決定されるまでに至るコンテキストを表している。ユーザ

の置かれている暗黙的な状況をコンピュータが理解することにより、行動を支援するコンテキストウェアネスについての研究が数多く行われているが、把持状態に含まれるコンテキストについても同様の応用が期待される。また、物体を意識的に握り締める等の、明示的な把持状態の変更による操作も可能であるため、把持状態の認識はさまざまな面においてHCIの発展に寄与するものと考えられる。例えば、Songらの研究では静電容量センサを用いて把持状態認識可能なタッチペンを実装している [14]。これは、スケッチ技法によって異なるペンの把持姿勢を、それらの技法に対応したペイントツールに割り当てている。そのため、ユーザはペイント操作を行う際に、実世界に近いペンの使用感を得られる。

一方で、コンピュータに物体の把持状態を認識させることは難しく、様々な課題を抱えている。その1つとしてセンシングの敷居の高さがある。手の接触を認識する技術は、タッチパネルの普及により、容易に使われる。しかし物

¹ 筑波大学 情報学群 情報メディア創成学類
School of Infomatics, College of Media Arts, Science and Technology, University of Tsukuba

² 筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

体の把持状態認識の既存手法は大量のセンサや特殊なハードウェア構成を必要とするため、複雑もしくは高コストである。この問題に対する解決策として、導電性の物体に対して周波数を掃引させた電流を流すことにより、単一の静電容量センシング機構による把持状態認識を行った例もある [13]。しかしながら、この手法は一般に導電性の物体のみを把持状態認識対象とする。

そこで、我々は、アクティブ音響センシングにより、手軽かつ安価に物体の把持状態認識を行う手法を示す。本手法の特徴は1組のスピーカとマイクを物体に貼り付けることによって、物体を把持する姿勢、及び物体を把持する力の認識を可能とする点にある。そのため、把持対象物体をデバイスとして一から作成することなく、既存の物体に対しても把持状態認識機能を追加することが可能である。また、センシングは音響特徴に基づくため、物体の導電性を要求しない。これらは、把持状態認識を利用するシステムの開発においてプロトタイピングを容易にする。

本稿では、まず提案手法のセンシング原理、及び実装方法を述べる。また、提案手法の適用例を示すものとして、携帯情報端末における把持状態認識を行う。また、その認識精度実験から本手法の有用性を議論する。

2. 関連研究

2.1 把持状態の認識

今日までに、様々な手法による把持状態認識が試みられている。

最も多く見られる手法は静電容量センサを使用したものである [2, 10, 13, 15, 18]。これらの手法は、数多くの静電容量センサを用いた専用の回路により把持状態認識を実現している。例えば Kim らは、64 個の静電容量センサを端末に組み込むことにより、端末の把持状態の認識を行った [10]。そのため、把持状態を高精度に認識できるものの高コストを要する。また、Sato らの Touché [13] は導電性物質に周波数を掃引させた電流を流すことにより、単一の静電容量センシング機構による多様なタッチ状態の認識を行った。しかし、これは一般に導電性物質に対してのみ適用可能となるため、非導電性物質に対しては導電インクや導電テープを使用したコーティングを必要とする。

また、静電容量センサ以外のセンサを使用した手法、及び端末に内蔵されたセンサを使用した手法も提案されている。Wimmer らは光ファイバと画像処理によって把持状態認識が可能な面のプロトタイピングを行う手法を提案している [16]。面に大量の光ファイバを組み込み、それらの末端を束として収束させたものを面画像として捉え、画像処理を施すことにより把持状態認識を実現した。Goel らは端末に内蔵された、ジャイロセンサと振動モータとスワイプ軌跡の組み合わせにより端末の把持状態認識を行った [5]。一般的なスマートフォンのみによる、追加のデバイスを必

要としない把持状態認識手法を提案しているが、認識に際してユーザに画面のスワイプ操作を要求する。

これらの認識手法に対して、我々の手法は音響に基づくため、導電性及び非導電性の両者を含む様々な物体に対して把持状態認識を行える。また、認識に必要なセンサは1組のスピーカとマイクのみであるため、手軽かつ安価である。

2.2 アクティブ音響センシング

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システムはこれまでにいくつか提案されている。

Gupta らの SoundWave [6] はラップトップ型 PC に搭載されたスピーカとマイクを使用して、PC の周囲における in-air ジェスチャーの認識を行った。SoundWave は、認識のために高周波信号を空気中に出力し、ドップラー効果を観測する。一方、本手法は高周波信号を振動として固体に直接出力し、その振動特性を観測する。

Collins は1つのピエゾトランスミッタと2つのコンタクトマイクを用いてガラス等の平面をタッチパネルにする手法を提案した [4]。これは、面の特定の位置をタッチした時に生じる振動特性の変化をアクティブ音響システムによって計測し、タッチ位置を推定する。本研究では振動特性の変化をタッチ位置ではなく、把持状態の認識に使用する点が異なる。

2.3 オーディオ入出力端子

本手法が認識に使用する特徴量は音響信号であるため、端末や PC に設けられたオーディオ入出力端子を介したデータの入出力が可能である。結果として、センサ以外に特別なハードウェアを必要とせず認識システムを端末内や PC 内に閉じて構築することが可能である。このように、オーディオ入出力端子をセンサ情報取得のためのインタフェースとして用いる研究はこれまでにいくつか提案されている。

Kuo らは、端末のオーディオ入出力端子の可能性について探求した。端末の端子を使った通信方法として、アナログ、デジタル、シリアルの例を示すことにより、信号入出力インタフェースとしてのデザイン空間について述べている [11]。Hwang らの MicPen [8] はペン先のスクラッチノイズを iPad のマイク入力端子を介して取得することにより、感圧ペンを実装した。巻口らは端末のオーディオ入出力端子を使用した脈波測定装置を示した [19]。これらの研究は、端末のオーディオ入出力端子を使用することによって、低コスト化に貢献できることを示している。

3. センシング原理

物体はその構造の特徴を表す指標のひとつとして振動特性を持つ。この特性は振動モード、固有振動数、モード減

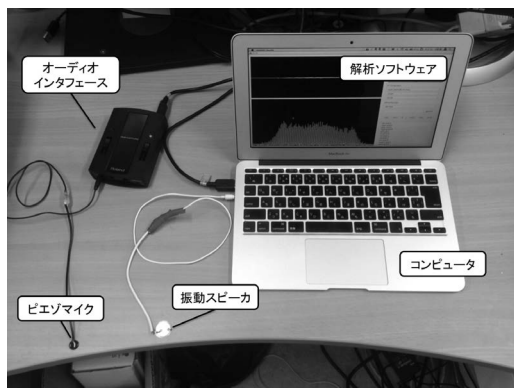


図 1 プロトタイプシステム
Fig. 1 Our Prototype System.

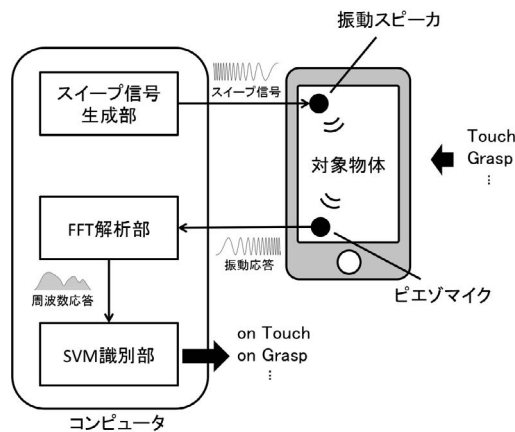


図 2 システム構成
Fig. 2 System Configuration.

衰比によって表現され、物体の形状、材質、境界条件に依存する。これは、建築や機械の分野において、構造物の耐久度の評価や制振技術として応用されている。

境界条件は物体の振動特性に影響を与える要因のひとつであり、物体に対する荷重や拘束の条件によって定義される。物体を把持する場合、把持姿勢や把持力によって物体を支える点と力は異なり、これに伴って拘束条件や荷重条件が変化する。これらの条件の変化は振動特性の変化として表れる。よって、物体の形状や材質に変化がない場合、物体の把持状態の変化は、振動特性の変化として観測可能である。

本研究ではこの現象を利用して、振動特性を解析し、その解析結果から把持状態を推測する。振動特性の解析には、実現象の観測に基づく周波数応答解析を用いる。これはアクチュエータによって物体を様々な周波数で振動させ、センサによって振動の周波数応答を取得する、実現象の観測に基づく手法である。得られた周波数応答から具体的な把持状態を求めるには、よく有限要素法等の理論モード解析が用いられる。しかしながら、これには物体の元の形状や材質、把持以外の影響による境界条件についての情報を要する。そのため、物体の構造情報について既知でなければならず、計算コストも高い。

これに対して、本手法では認識したい把持状態をラベルとし、周波数応答を特徴量とした機械学習を用いる。これにより、学習フェーズを必要とするものの、理論モード解析を用いるよりも計算量を抑えつつ、物体の構造情報が未知な物体に対しても大まかに把持状態を推測できる。

4. 実装

以上のセンシング原理に基づき、プロトタイプシステムを実装した。実装したプロトタイプシステムの全体図を図1に、システム構成を図2に示す。

プロトタイプシステムは、物体を振動させる振動スピーカ、振動応答を取得するピエゾマイク、信号の入出力と解析を行うコンピュータ及びソフトウェアからなる。以下の

節ではプロトタイプシステムの詳細な実装について述べる。

4.1 ハードウェア 振動スピーカ

振動スピーカはそれを物体に貼り付けることにより、物体自体を振動させて音を出力するスピーカである。一般的なスピーカと同様に音響信号を再生するため、振動の細かい制御が可能である。本システムでは、振動スピーカとしてスライブ社の OMR20F10H-BP-310 を用いた (図3左)。これは厚さ 0.3mm、直径 21mm の円盤型の圧電式のスピーカであり、その薄さから様々な部分に貼り付けることが可能である。このスピーカの再生可能周波数は 1kHz~40kHz 以上である。

ピエゾマイク

ピエゾマイクは固体音を検出するためのマイクである。これを振動応答の取得に用いる。本システムでは、ピエゾマイクとして SHADOW 社の SH710 を使用した (図3中央)。その形状は厚さ 2mm、直径 12mm の円盤型である。

コンピュータ

振動スピーカを駆動させるための音響信号の出力、ピエゾマイクによって取得した振動応答の入力、及び機械学習を行うコンピュータとして、Apple 社の MacBook Air (CPU: Intel Core 2 Duo 1.4GHz, RAM: 2GB) を使用した。振動スピーカへの信号の出力を本体のヘッドフォンポートから出力する。また、ピエゾマイクからの入力を USB 接続のオーディオインタフェース (Roland Duo-Capture) を介して行なう。

4.2 解析ソフトウェア

解析ソフトウェアを C++ 言語を用いて作成した。これは、スイープ信号生成部、FFT 解析部、SVM 識別部という 3 つのモジュールからなる。

スイープ信号生成部

スイープ信号生成部は振動スピーカを駆動させるための

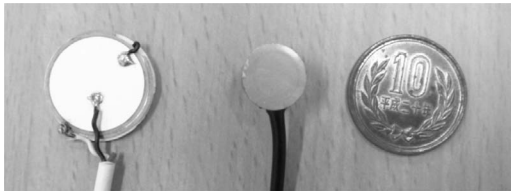


図 3 振動スピーカとピエゾマイク

Fig. 3 Vibration Speaker and Piezoelectric Microphone.

信号を生成する。周波数応答を得るには幅広い周波数成分を持つ信号によって物体を振動させる必要がある。そのような特性をもつ信号として、インパルス信号やスイープ信号等があるが、本システムでは帯域の範囲を指定できるスイープ信号を使用した。スイープ信号は30ms間に18.5kHzから21.5kHzまで掃引するものとした。振動スピーカによって物体を振動させる際、音の発生を伴うが、この周波数帯は人間の耳に聞こえにくい[3]。また、物体を触った時に生じるノイズ（多くは1kHz未満）等の影響も受けにくい。なお、出力の際のサンプリング周波数を44.1kHzとしたため、スイープ信号の周波数分解能は約2.3Hzとなる。

FFT 解析部

FFT解析部はピエゾマイクから取得した振動波形をFFTにより、共振スペクトルへ変換する。取得の際のサンプリング周波数を出力と同じく44.1kHzとした。FFTを行う際のフレームサイズを8192(約186ms)とし、窓関数としてハニング窓を使用した。この変換により、4096点の周波数エネルギーが得られる。標本化定理により、復元可能な最大周波数は22.05kHzとなるため、この解析における周波数分解能は約5.4Hzとなる。

SVM 識別部

SVM識別部は機械学習手法のひとつであるSupport Vector Machine (SVM)を用いて、振動特性と把持状態の関数の学習とリアルタイム識別を行う。学習に使用する特徴量として、FFT解析部によって得られた共振スペクトルのうち、出力するスイープ信号と同じ周波数帯である18.5kHz~21.5kHzの部分(557次元)を用いた。なお、実装に際して、SVMのライブラリであるLIBSVM[1]を使用した。システムの適用対象やマイクとスピーカの設置箇所によって共振スペクトルの変化の仕方が大きく異なるため、本システムではハードウェアセットアップの変更の際にグリッドサーチによるパラメータ最適化を行う。

5. 携帯情報端末の把持状態認識 (動作確認実験)

本手法の動作確認実験として、携帯情報端末の把持状態認識を行った。端末の把持状態は、把持状態認識の研究における認識対象として多く見られる[2,5,10,15,18]ため、認識対象とする把持状態をこれらの先行研究を参考に決定した。動作確認に用いる端末としてApple社のiPhone 4S

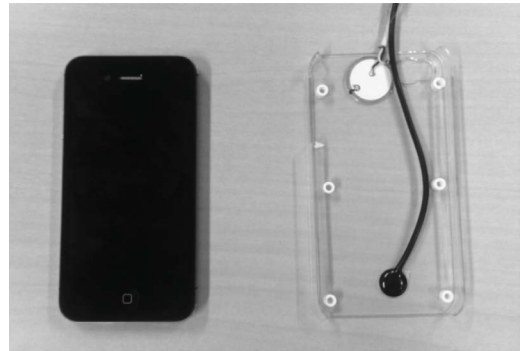


図 4 iPhone 4S とシステムを設置したハードケース

Fig. 4 iPhone 4S and Hard Case with Our System.

(図4左)を使用した。また、薄い構造物の方が高い周波数において振動しやすいという点、及び手がセンサに直接触れないという点から、スピーカとマイクを端末本体ではなくプラスチック製のハードケースの内側に貼りつけた(図4右)。なお、このケースを端末本体に直接取り付けの場合、センサ部分の厚みによって端末本体が安定しない。そのため、ケースの4隅と底面両サイドにプラスチックビーズ(直径、高さ5mm)を貼り付けることにより端末本体とケースの間に隙間を確保した。

以上の構成を用いて、2つの動作確認実験を行った。

5.1 把持姿勢の認識

スマートフォンを始めとした、近年の端末は、電話、カメラ、メール等、様々なアプリケーションを含む。そのうち、いくつかのアプリケーションは特有の把持姿勢を伴う。例えば電話の場合、端末を耳に当てようとするため、親指が端末の側面に沿った姿勢になりがちである。また、カメラの場合には、端末の4隅を摘むような姿勢になりがちである。したがって、これらの把持姿勢を認識できれば、端末の把持姿勢に合わせてアプリケーションを起動するという自然なインタラクションが可能になる。また、右手、左手、両手といった持ち手の違いによって、指の可動範囲や操作の仕方が異なる。したがって、これらを認識することによりそれぞれの持ち手に最適化したUIへ動的に変更すれば、操作性の向上が期待される。

これらのことから、本節ではアプリケーション間に生じる端末の把持姿勢の違い及び持ち手の認識を行う。認識対象とする姿勢は、[10]を参考にした上で、図5に示す把持姿勢カテゴリの認識セット(7把持姿勢)とした。この認識セットは4種類のアプリケーション(call, sms, camera, game)と3種類の持ち手(left, right, both)の組み合わせによって構成される。

著者1名が把持姿勢カテゴリに含まれる7種類の把持姿勢で端末を把持したところ、結果として、図6に示されるように、各把持姿勢に応じて異なる形状の共振スペクトルが得られた。また、これら7種類の把持姿勢に対する学

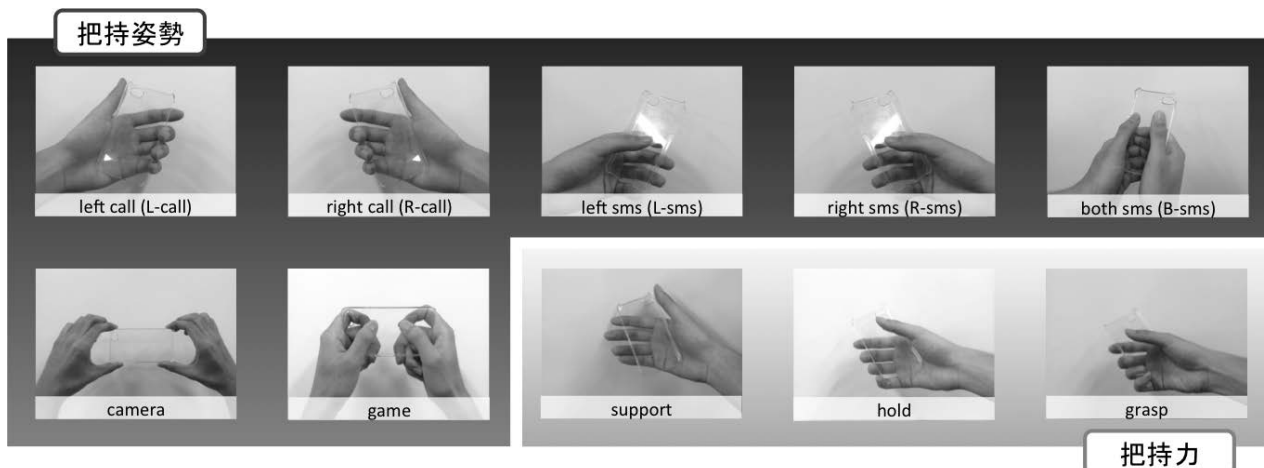


図 5 認識カテゴリと認識セット

Fig. 5 Recognition Category and Set.

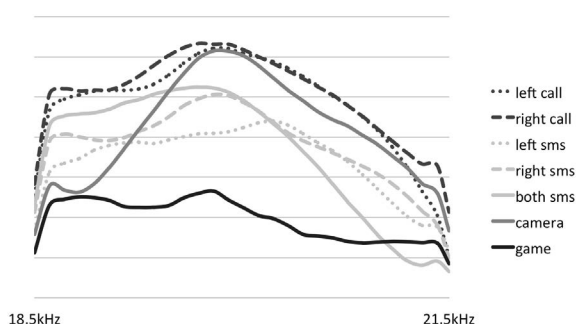


図 6 7 把持姿勢の共振スペクトル

Fig. 6 Resonance Spectra on 7 Grasp Postures.

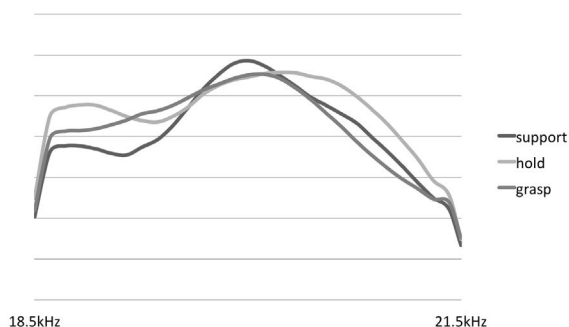


図 7 3 把持力の共振スペクトル

Fig. 7 Resonance Spectra on 3 Grasp Strengths.

習を行ったところ、リアルタイムに認識が行えることが分かった。

5.2 把持力の認識

スマートフォンなどのタッチパネルのみを入力デバイスとして持つ端末は、いくつかのマルチタッチジェスチャが使用可能であるものの、片手把持状態において操作されることが好まれるため [9, 12], 親指のみを使用したシングルタッチ操作になりがちである。この場合、入力チャンネルが制限される。これに対して、端末に対する把持力は片手

操作における入力チャンネルとなり得る。例えば、地図等のビューワ系アプリケーションを使用する際に、パンやズームのために 2 本指によるピンチ操作が要求される。この操作は片手で端末を把持し、もう一方の手でタッチすることを想定した設計になっているため、片手で把持と操作を同時に行うことは難しい。把持力を認識できれば、強く握った時の上下スクロールをパンとズームに割り当てることにより、片手操作における操作性の向上が期待される。

このような入力チャンネルの拡張を目的として、端末の把持力の認識実験を行った。認識対象は図 5 に示す把持力カテゴリの認識セット (3 把持力) とした。前節と同様に著者 1 名が 3 段階の把持力で端末を把持したところ、共振スペクトルとして図 7 に示すように、各把持力に応じて異なる形状の共振スペクトルが得られた。また、リアルタイムに把持力を認識できることが分かった。

6. 評価実験

本手法による把持状態認識の実用性の検証を目的として、認識精度を求める評価実験を行った。実験では 6 人 (男性 5 人, 女性 1 人, 21~24 歳) の被験者から様々な端末把持状態における振動特性データの収集を行った。被験者の内、4 人はスマートフォンを、2 人はフィーチャーフォンを普段から使用していた。実験期間は 1 日、1 人あたりの所要時間は 20~30 分であった。

6.1 実験機器

把持する端末として、5 節に述べたケースを取り付けた iPhone 4S を用いた。

6.2 実験内容

実験では 5 節に述べた把持姿勢と把持力のカテゴリについて独立に次のタスクを行う。



図 8 評価実験の様子
Fig. 8 Experiments.

まず、実験者は各カテゴリの認識セットのうち、1つの把持状態の写真とその解説を被験者に提示する(図8)。その後、被験者は提示された写真と同様の把持を行い、把持の完了後、実験者に合図を送る。実験者は、被験者の合図と共に、10フレームの共振スペクトルを取得する。データの取得後、認識セット内の異なる把持状態の写真を提示する。全ての把持状態の写真が12回ずつ提示されるまで同様のタスクを繰り返す。なお、提示する写真の順序はランダムかつ同じものが連続しない順序とした。

このタスクによって得られるデータの数は、把持姿勢について5040フレーム(10フレーム×7把持姿勢×12セット×6被験者)、把持力について2160フレーム(10フレーム×3把持力×12セット×6被験者)となる。

6.3 実験結果

以上の実験で得られたデータに対して、Weka Machine Learning Toolkit [7]を用いた交差検定を行い、認識精度を評価した。機械学習としてプロトタイプシステムと同様にLIBSVM (RBF kernel, $c=32.0$ $\gamma=0.0078125$)を使用した。被験者間に生じる把持状態の違いを考慮して、学習とテストに同一の被験者のデータのみ使用する per-user test と、全てのデータを交差させる cross-user test を実施した。

6.3.1 Per-user test

各被験者のデータごとに12-fold交差検定を行い、個人使用環境における認識精度を評価した。検定結果を図9に示す。

結果として、7把持姿勢の認識精度は全ての被験者において90%以上(90.5~99.5%)となった。また、3把持力の認識精度は全ての被験者において95%以上(95.5~100%)となった。

これらのことは、本手法が、端末のような個人での使用を主目的とする物体に対して、把持状態認識手法としての実用可能性があることを示している。

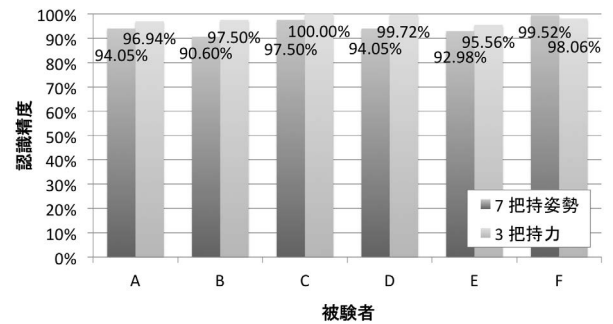


図 9 Per-User Test における 7 把持姿勢の認識精度

Fig. 9 Per-User Recognition Accuracy for 7 Grasp Postures.

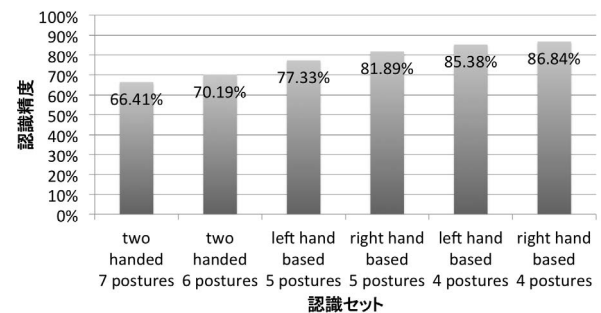


図 10 Cross-User Test における把持姿勢の認識精度

Fig. 10 Cross-User Recognition Accuracy for Grasp Postures.

6.3.2 Cross-user test

実験で収集した全てのデータを使用して、6-fold交差検定を行い、複数人使用環境における認識精度を評価した。

結果として、7把持姿勢の認識精度は66.4%となった。また、3把持力の認識精度は81.4%となった。両者の結果から、複数人使用環境における認識精度は前節の個人使用環境のものより低下した。特に、把持姿勢の認識精度の低下はより顕著である。

把持姿勢の検定における混同行列を表1に示す。混同行列からは、左右の持ち手による把持パターンを持つ把持姿勢(L/R-call, L/R-sms)において、本来認識されるべき手と異なる手として誤認識される割合が高くなっている。また、L-callとB-smsについては他の把持姿勢に比べて高いfalse positive rate (FP rate)を示している。一方でcameraのみ90%以上の認識精度となっている。

これらの結果を受けて、認識精度に悪影響を及ぼしていると思われる把持姿勢(L/R-call, L/R-sms, B-sms)を認識対象から取り除くことにより、精度がどのように向上するか調査した。図10に認識セットを減らしていった時の認識精度の変化を示す。two handed 7 posturesは実験で収集した全ての把持姿勢を含む。これを本調査における基準のセットとした。two handed 6 posturesはB-smsを基準セットから除外した認識セットである。left/right hand based 5 posturesは、左/右手による把持パターンを持つ把持姿勢について、把持する手をどちらか片方に限定

表 1 Cross-User Test における 7 把持姿勢認識の混同行列

Table 1 Confusion Matrix on Cross-User Test for 7 Grasp Postures Recognition.

Classified as →	L-call	R-call	L-sms	R-sms	B-sms	camera	game	TP Rate	FP Rate
L-call	424	100	29	58	67	32	10	0.589	0.081
R-call	148	502	10	6	36	16	2	0.697	0.06
L-sms	22	20	399	93	118	10	58	0.554	0.066
R-sms	86	38	103	395	56	5	37	0.549	0.058
B-sms	61	19	49	65	497	3	26	0.69	0.081
camera	10	32	0	0	17	661	0	0.918	0.015
game	23	51	95	28	54	0	469	0.651	0.056

した認識セット (L-call, L-sms, B-sms, camera, game / R-call, R-sms, B-sms, camera, game) である。これらの認識セットからさらに B-sms を除外したものが left/right hand based 4 postures となる。

この結果から、認識対象とする把持姿勢の数が減るに従い、精度が向上していることが分かる。また、左/右手による把持パターンを持つ把持姿勢について、把持する手を片方に限定することにより、85%以上の認識精度が得られることを示している。

7. 議論

7.1 実環境利用の想定

本論文における評価実験には、全て座った状態のデータを用いた。しかし、実環境での利用を考えた場合、立ち状態や歩行状態等の様々な体勢においてもロバストに認識する必要がある。端末へ加わる荷重分布が体勢間で異なる場合、認識精度への影響が懸念される。

また、全ての実験は屋内の同じ部屋にて行った。このとき、周波数応答の測定に使用する高周波領域でのノイズはほとんど見られなかった。しかしながら、屋外や電車内等での利用を考えた場合、そのような帯域の信号を発信する機器等に遭遇し、精度が低下する恐れもある。

これらの点から、今後は実環境における動作テスト及び認識精度の評価を行なう。そのためには、PC との連携を必要としないスタンドアロンシステムが必要となる。最近の端末には高いパフォーマンスを持つプロセッサが搭載されている。また、ヘッドセットによる利用を想定した外部オーディオ入出力環境も備わっている。本手法におけるシステム要件は、1組のマイクとスピーカに加えて、機械学習を実装可能な計算機環境と、オーディオ入出力環境のみであることから、端末によるスタンドアロンシステムの実装は十分に可能であると考えられる。今後は端末を用いたスタンドアロンシステムを実装し、実環境における実用可能性を調査したい。

7.2 適用可能範囲の探索

本論文では端末を対象とした適用例について述べたが、

原理的には、他の物体に対しても適用可能である。しかしながら、物体の材質や形状によっては適用できない場合もある。例えば、高い自由度で物体が変形する場合、把持状態と振動特性の対応付けは困難である。振動を吸収する物体についても同様である。物体の大きさによっては、振動スピーカにアンプを接続し、より大きなエネルギーで振動させる必要がある。また、部分的な振動が強くなるため、把持状態認識可能な部位が振動スピーカの周囲に限定される。今後は様々な物体や条件に対して本手法の適用を試みるにより、適用可能範囲の探索を行いたい。

7.3 認識精度向上の工夫

我々のプロトタイプシステムは出力するスイープ信号及び観測する周波数帯域として 18.5kHz から 21.5kHz を使用した。人間に耳障りとならず、ノイズに対してロバストであることから下限値を 18.5kHz としたが、上限値については多くのデジタルオーディオ入出力機器がサンプリング周波数として 44.1kHz (復元可能周波数の上限が 22.05kHz) を採用していることによる。一方で、44.1kHz より高いサンプリング周波数が利用可能になった機器も普及してきている。それらを使用した場合、22.05kHz 以上の帯域の共振スペクトルを観測可能となる。その場合、機械学習に使用する特徴量が改善されるため、認識精度の向上が期待される。

8. 終わりに

我々は、本論文において1組の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により、安価かつ手軽に物体の把持状態認識を行う手法を提案し、プロトタイプシステムの実装を行った。携帯情報端末の操作に本手法を用いることを想定した実験を行った結果、7種類の把持姿勢の認識精度が per-user test において 90~99%、cross-user test において 66%となった。また、3段階の把持力の認識精度が per-user test において 95~100%、cross-user test において 81%となった。

今後は様々な物体に本手法を適用し、適用可能範囲の探索を行う。また、システムのスタンドアロン化を行い、実

環境利用における評価も行うことを考えている。

参考文献

- [1] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, pp. 27:1–27:27, May 2011.
- [2] Lung-Pan Cheng, Fang-I Hsiao, Yen-Ting Liu, and Mike Y. Chen. irotate grasp: automatic screen rotation based on grasp of mobile devices. In *Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST Adjunct Proceedings '12, pp. 15–16, New York, NY, USA, 2012. ACM.
- [3] D'Ambrose Christopher. Frequency range of human hearing. In *The Physics Factbook*.
- [4] Tim Collins. Active acoustic touch interface. *Electronics Letters*, Vol. 45, No. 20, pp. 1055–1056, 24 2009.
- [5] Mayank Goel, Jacob Wobbrock, and Shwetak Patel. Gripsense: using built-in sensors to detect hand posture and pressure on commodity mobile phones. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pp. 545–554, New York, NY, USA, 2012. ACM.
- [6] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave: using the doppler effect to sense gestures. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pp. 1911–1914, New York, NY, USA, 2012. ACM.
- [7] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, Vol. 11, No. 1, pp. 10–18, November 2009.
- [8] Sungjae Hwang, Andrea Bianchi, and Kwangyun Wohn. Micpen: pressure-sensitive pen interaction using microphone with standard touchscreen. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*, CHI EA '12, pp. 1847–1852, New York, NY, USA, 2012. ACM.
- [9] Amy K. Karlson and Benjamin B. Bederson. Understanding single-handed mobile device interaction. Technical report, Department of Computer Science, University of Maryland, 2006.
- [10] Kee-Eung Kim, Wook Chang, Sung-Jung Cho, Junghyun Shim, Hyunjeong Lee, Joonah Park, Youngbeom Lee, and Sangryong Kim. Hand grip pattern recognition for mobile user interfaces. In *Proceedings of the 18th conference on Innovative applications of artificial intelligence - Volume 2*, IAAI'06, pp. 1789–1794. AAAI Press, 2006.
- [11] Ye-Sheng Kuo, Sonal Verma, Thomas Schmid, and Prabal Dutta. Hijacking power and bandwidth from the mobile phone's audio interface. In *Proceedings of the First ACM Symposium on Computing for Development*, ACM DEV '10, pp. 24:1–24:10, New York, NY, USA, 2010. ACM.
- [12] Pekka Parhi, Amy K. Karlson, and Benjamin B. Bederson. Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, MobileHCI '06, pp. 203–210, New York, NY, USA, 2006. ACM.
- [13] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pp. 483–492, New York, NY, USA, 2012. ACM.
- [14] Hyunyoung Song, Hrvoje Benko, Francois Guimbretiere, Shahram Izadi, Xiang Cao, and Ken Hinckley. Grips and gestures on a multi-touch pen. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pp. 1323–1332, New York, NY, USA, 2011. ACM.
- [15] Brandon T. Taylor and V. Michael Bove, Jr. Graspables: grasp-recognition as a user interface. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pp. 917–926, New York, NY, USA, 2009. ACM.
- [16] Raphael Wimmer. Flyeye: grasp-sensitive surfaces using optical fiber. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, TEI '10, pp. 245–248, New York, NY, USA, 2010. ACM.
- [17] Raphael Wimmer. Grasp sensing for human-computer interaction. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, TEI '11, pp. 221–228, New York, NY, USA, 2011. ACM.
- [18] Raphael Wimmer and Sebastian Boring. Handsense: discriminating different ways of grasping and holding a tangible user interface. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, TEI '09, pp. 359–362, New York, NY, USA, 2009. ACM.
- [19] 巻口誉宗, 吉川浩, 棟方渚, 小野哲雄. スマートフォンの音声入出力端子をインターフェースとする脈波測定装置の実装と評価. *インタラクション '12*, pp. 593–598. 一般社団法人情報処理学会, 2012.