Vol.2016-MBL-81 No.19
Vol.2016-ITS-67 No.19
2016/12/9

マイクの部分遮蔽を用いた超音波によるジェスチャ認識手法

渡邉 拓貴[1,2,a)]　寺田 努[1,3,b)]　塚本 昌彦[1,c)]

概要：ジェスチャ認識は，コンピュータとのインタラクション手法として広く用いられるようになってきている．特に，画面が小さく身体に常にセンサを装着しているような，モバイル/ウェアラブルコンピューティング環境では，ジェスチャによる操作が効果的であるといえる．しかし，従来手法ではデバイスに追加のハードウェアが必要であったり，識別できるジェスチャに制限がある．そこで本研究では，デバイスに搭載されているマイクとスピーカを用いた，超音波によるジェスチャ認識手法を提案する．スピーカからは超音波を発信し，手などの対象から反射された超音波をマイクで取得する．対象の動きにより発生するドップラー効果や，距離によって変化する反射音の音量を用いて，ジェスチャを識別する．さらに，マイク/スピーカ部に取得/発信する音響特徴が変化するようなカバーを装着し，従来の超音波による手法では困難であったジェスチャの識別を可能にする．本手法では，デバイス内蔵のマイクとスピーカを利用するため，ジェスチャ認識のための追加のデバイスが必要ない．また，3D プリントされたカバーを用いて音響特徴を変化させるため，ジェスチャ識別能力拡張のために電力を消費する追加のデバイスが必要ない．本研究では，被験者 3 名に対して左右からのスワイプジェスチャを識別したところ，最も認識率の良いカバーを用いた場合，平均 92.7%の認識率で識別できた．

Hiroki Watanabe[1,2,a)]　Tsutomu Terada[1,3,b)]　Masahiko Tsukamoto[1,c)]

1. はじめに

近年のコンピュータ小型化に伴い，スマートフォンやスマートウォッチ等の小型デバイスを用いた，モバイル/ウェアラブルコンピューティング環境に注目が集まっている．これら小型のデバイスを操作するのに，主に用いられる手法はタッチ操作である．タッチ操作はマルチタッチによる様々な操作や，直観的な操作が可能である．しかし，小型デバイスの画面上でのタッチ操作は，テキスト入力時にはソフトウェアキーボード上の小さなボタンを正確に選択する必要がある．また，タッチ操作の大きな問題の一つとして，操作を行う指自体で画面を覆ってしまうという問題点がある．さらに，タッチ操作時には画面に直接触れる必要があるが，料理中や機械の作業中のような手が濡れている/汚れている場合や，タッチパネルに反応しない手袋を装着している場合のように，画面に触れたくない/触れても反応しない状況があり，空中でのユーザのジェスチャが取得できれば，より便利である．この問題の解決手法として，赤外線距離センサをデバイス周囲に配置し，画面外でのジェスチャ認識を可能にした手法 [3], [10], [12] や，カメラを用いた手法がある [9], [16], [19]．しかし，これらの手法ではジェスチャ認識用に電力を消費する追加デバイスが必要であったり，カメラベースの認識は周囲環境の明るさに左右されやすい，一般的に計算コストが大きい等の問題がある．デバイス内蔵のセンサを用いたジェスチャ認識手法として，マイクとスピーカによる超音波を用いたジェスチャ認識手法 [6], [14] が提案されている．しかし従来の手法では，ドップラー効果やエコーによってマイクへの物体の近づき/遠ざかりは識別できても，それが左右どちらからのものかまでを識別することは困難であり，この情報を識別するには複数のマイクを用いて，到達時間差を計算することにより算出することが一般的である．

そこで本研究では，文献 [11], [13], [19] 等で用いられているような，センサ部に物理的な機構を加えることで取得できる情報量を増やす手法を，マイク/スピーカに適応する．具体的には，デバイスのマイク部分に，片側からの音

1 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University
2 日本学術振興会
Japan Society for the Promotion of Science
3 科学技術振興機構さきがけ
PRESTO, Japan Science and Technology Agency
a) hiroki.watanabe@stu.kobe-u.ac.jp
b) tsutomu@eedept.kobe-u.ac.jp
c) tuka@kobe-u.ac.jp

は通常通り取得でき，もう片側からの音は取得しにくいような機構のカバーを装着することで，一つのマイクで取得できる情報量を増やすことを目的とする．従来マイクに対して左右対象である動きは一つのマイクでの識別は困難であったが，提案手法によってこれらが区別できる．

本研究では，提案手法のプロトタイプカバーを 5 種類作製し，それぞれの性能比較のための評価実験を行った．3 名の被験者に対し，左からのスワイプ，右からのスワイプの 2 つのジェスチャを識別したところ，最も効果の高いカバーの場合，92.7%の認識率であった．

以降，2 章で本研究に関連する研究について述べ，3 章で提案手法について述べる．4 章で実装を紹介し，5 章で評価実験について示す．最後に 6 章でまとめを行う．

2. 関連研究

2.1 様々なセンサによるインタラクション

Toffee[18] では，PC やスマートフォンの四隅に取り付けたピエゾ素子によって，ユーザのテーブル上でのタップの場所を識別できる．Skinput[7] では，腕に装着したマイクアレイによって，腕のどこをタップしたのかを識別する．これらの研究では，複数のマイクを用いて発生した音を受動的に受け取り，タップ等の音の発生場所を特定する．

石川らは，光学距離センサを 4 つ搭載したデバイスを作製し，センサの反応する順番によってコマンドを割り当て，PC でのタッチレスジェスチャを可能にしている [8]．Skin Buttons[12] では，スマートウォッチの両サイドに近接センサと小型プロジェクタを搭載したデバイスを装着することで，皮膚上へのボタン投影と選択を可能にしている．SideSight[3]，HoverFlow[10] では，スマートフォンの周囲に赤外線距離センサを装着することで，デバイス周囲での指の位置やハンドジェスチャを取得できる．これらの研究では，距離センサをデバイス周辺に取り付け，インタラクション手法を拡張している．

Song らは，デバイス内蔵のカメラを用いて空中でのジェスチャを取得している [16]．Digits[9] では，手首に装着した赤外線カメラによってユーザの手の 3D 形状を認識できる．Surround-See[19] では，スマートフォンのカメラ部分に取り付けた全方位鏡によって，カメラの視野を広げた全方位でのジェスチャ認識を可能にする．これらの研究では，カメラベースでジェスチャを識別する．

これらの研究に対して本研究では，デバイス内蔵のマイクとスピーカのみを用いて，能動的に超音波を発信し，その反射音を利用する．さらに，3D プリントされたカバーを用いることで，取得できる音響特徴を変化させるため，電力を消費する追加デバイスが必要ない．

2.2 超音波を用いたインタラクション

Doplink[2]，SurfaceLink[5]，AirLInk[4] では，スマートフォン同士やスマートフォンと PC 間で，超音波によるペアリングやファイル転送等を可能にする．これらの研究では，デバイス間のインタラクションについて議論されており，本研究とは異なる．

Chirp microsystems[1] は，チップ上に実装された超音波レンジファインダによって，空中でのジェスチャを取得できる．しかし，この手法ではデバイスに特別なチップを搭載する必要がある．

Tarzia らは, PC の前にユーザが存在するか否かを, PC に内蔵されたスピーカとマイクによるソナーで取得した [17]．SoundWave[6] では，PC から超音波を発信し，ハンドジェスチャによって生じた反射波のドップラー効果を利用して，ジェスチャを識別している．FingerIO[14] では，スマートフォン/ウォッチ型デバイスのスピーカから超音波を発信し，指先からの反射波の，二つのマイクへの到達時間差を用いて 2 次元のトラッキングを可能にしている．これらの研究では，ドップラー効果によりデバイスへの近づき/遠ざかりは認識できても，それがどの方向からのジェスチャかまでは識別できない．また，これを識別するには二つ以上のマイクが必要であるが，ステレオマイクを搭載したスマートフォン/ウォッチは一般的ではない．

Touch & Activate[15] では，物体に一対のコンタクトマイクとコンタクトスピーカを装着することでタッチ入力を認識し，既存物体へのインタラクティブ性を付与している．Acoustruments[11] では，スマートフォンのスピーカからマイクへと続く管のようなアタッチメントを装着し，スピーカから超音波を発信する．ユーザがアタッチメントを触れる位置によって伝播の特性が変化し，ユーザの操作が識別できる．これらの研究では，物体へのタッチインタラクションについて着目しており，本研究の目的とする空中でのジェスチャ認識とは異なる．

3. 提案手法

本研究のシステム構成を図 1 に示す．本研究では，超音波をスピーカから発信し，対象に反射した音をマイクで取得し解析することにより，ジェスチャを認識する．対象がスピーカ/マイクに近づく/遠ざかる動きにより，ドップラー効果による周波数の遷移や，受信できる超音波音量の変化が発生するため，これらの特徴を利用してユーザのジェスチャを識別する．本研究の想定環境では，超音波の発信器と受信器として，スマートフォン等の既存のデバイスに内蔵されているマイクとスピーカを用い，ユーザの手によってジェスチャ入力を行う．発信する超音波の周波数は 20kHz とした．超音波は人には聞こえず，周波数解析することで環境音との分離も容易に可能である．

スピーカから 20kHz の超音波を発信し，マイクでは同時に音声信号を取得する．音声取得のためのサンプリング周波数は 44.1kHz とした．取得した音声データは環境音と超

Vol.2016-MBL-81 No.19
Vol.2016-ITS-67 No.19
2016/12/9

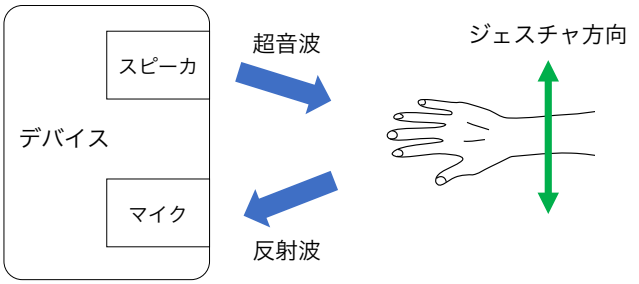

図 1 システム構成

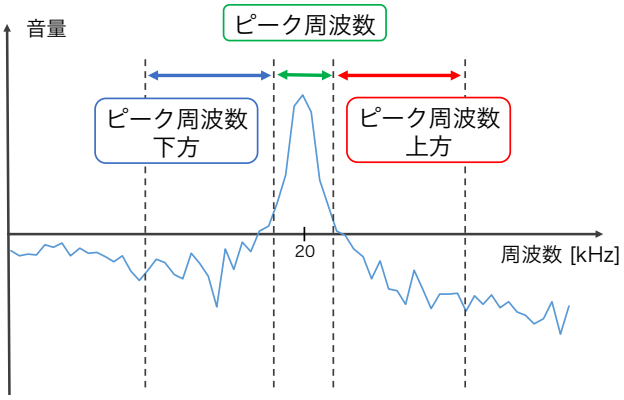

図 2 ピーク周波数周辺の周波数スペクトル

音波が混在しているため，高速フーリエ変換 (FFT: Fast Fourier Transform) を行い，20kHz 周囲の周波数スペクトルを抽出する．FFT 計算のためのウインドウサイズは 2048 とした．文献 [6] より，PC の前でジェスチャ入力のためにユーザが手を動かした時の速度は最大でも 3.9m/sec 程度であることが分かっている．そのため，ドップラー効果で生じる周波数のずれは最大でも約 218Hz だといえる．本研究でのサンプリング周波数と FFT のウインドウサイズを考慮すると，周波数分解能は 1 ビン約 21.5Hz(44100/2048) であるので，周波数ピーク値の前後 10 ビン程度を観察すればジェスチャによる周波数変化を取得できると考えられる．本研究では，余裕を持って周波数ピーク値の前後 15 ビンずつを観測した．図 2 にピーク周波数周辺の周波数スペクトルを対数表示で示す．この図に示すように，本研究ではピーク周波数の前後 2 サンプルの範囲をピーク周波数，ピーク周波数下方 15 サンプルの範囲をピーク周波数下方，ピーク周波数上方の 15 サンプルをピーク周波数上方と定義する．

図 3 に，音源に対し右から手を通過させた場合と，左から手を通過させた場合のピーク周波数周辺の音量の時間変化について示す．青色の線は，ピーク周波数下方の音量を示し，手がマイクから離れる動きをした時に増加する．赤色の線は，ピーク周波数上方の音量であり，手がマイクへと近づく動きをした時に増加する．図 3 のカバー無しに示すように，対象が近づく/遠ざかる動きはピーク周波数上方と下方の変化から取得することができるが，左右どちらからのジェスチャでも波形の変化が似ているため，どの方向から近づいているかを識別することは難しい．そこで本研究では，マイクの一部を遮蔽するようなカバーを装着することで，左右からのマイクへの入力音量に差をつける．例えば，図 3 カバーありのようにマイク右側を遮蔽し，音を取得しにくくした場合に右から左へと手を動かすと，マイクへと接近するまでのドップラー効果の音量は小さく，マイクを通過した後のドップラー効果の音量は大きく反応するため，先に小さく赤色の線が反応し，次に大きく青色の線が反応すると考えられる．また，逆側からのジェスチャでは大小関係が入れ替わると考えられる．

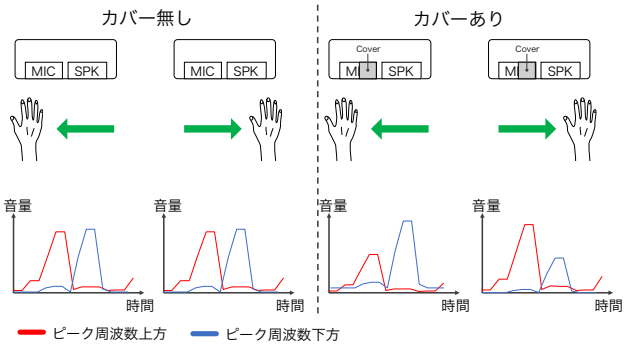

図 3 音量の時間変化の一例

ジェスチャ認識には，ピーク周波数下方の音量，ピーク周波数上方の音量を用いる．図 3 カバーありに示すような，マイク右側を覆うデバイス構成の場合，右から左へのスワイプでは，ピーク周波数上方が小さく反応してから，ピーク周波数下方が大きく反応する．また，左から右へのスワイプでは，ピーク周波数上方が大きく反応してからピーク周波数下方が小さく反応する．従って，ピーク周波数上方の音量，下方の音量が，一定時間内にしきい値を超えた時のみシステムは判定を行い，これらの大小関係によりジェスチャの方向を識別する．本研究では，予備実験からしきい値を設定し，ピーク周波数上方と下方の反応は，1s 以内であれば有効とした．なお，デバイスとマイクカバーの構成によって，ジェスチャの向きとこれらの大小関係の対応は変化する．

4. 実装

提案手法を実装したプロトタイプデバイスを作製した．本研究では，スマートフォンを対象のデバイスとし，スマートフォン用のカバーとして装着できるようなマイク/スピーカカバーを実装した．用いたスマートフォンは Apple 社の iPhone5 である．考案したマイクカバー機構の 3D モデルを図 4 に，実際に出力したカバーを図 5 に示す．3D モデルの作成には Autodesk 社の 123D Design を用いた．また，使用した 3D プリンタは PP3DP 社の UP Plus2 であり，フィラメントには ABS 樹脂を用いた．カバー形状によっては，一度では出力が難しい形状があったため，部

Vol.2016-MBL-81 No.19
Vol.2016-ITS-67 No.19
2016/12/9

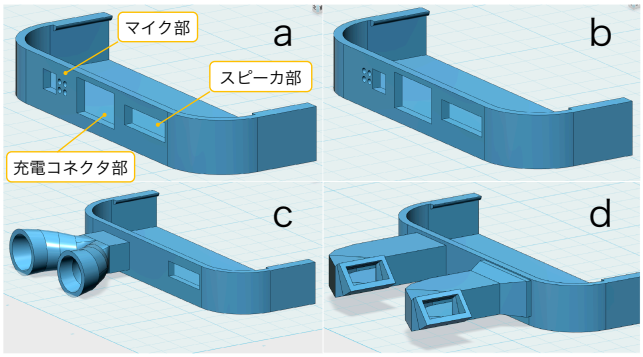

図 4 3D モデル

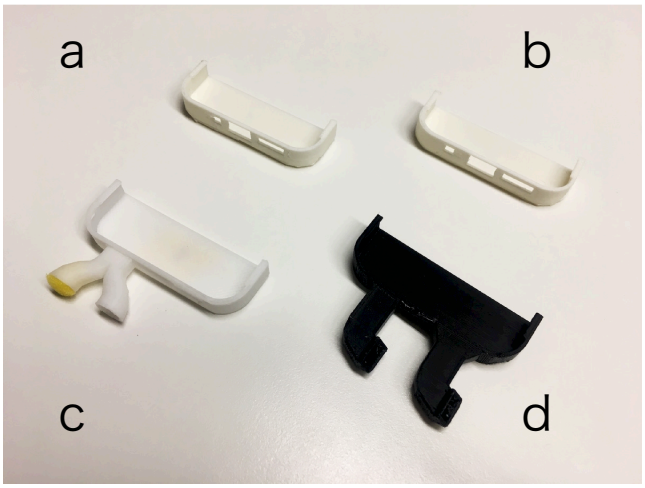

図 5 出力されたスマートフォンカバー

品ごとに出力し，ABS 用の接着剤で接着した．

図 4，5 に示すように，提案手法に最適な形状を求めるため，以下の 3 つの方針に従って，3 つのタイプのカバーを作製した．

- 部分遮蔽タイプ
  マイク部分の左右で開放度を変更した．片側は遮蔽が無いが，もう片側はカバーで遮蔽した上で，直径 1mm 程度の穴を 4 つ開けた (図 4，5 の a，b)．
- 管タイプ
  マイク部分から 2 又に分かれた管を取り付ける．片側の管を吸音性のある物質で塞ぐことで，左右の管から得られる特性に変化をつける．本研究ではスポンジを管に詰めた (図 4，5 の c)．
- 指向性タイプ
  マイクとスピーカに同じ向きの指向性を与えることで，片側からの動きのみに敏感になるようにした (図 4，5 の d)．

超音波発信/取得用のアプリケーションは，openFrameworks を利用して開発した．アプリケーション画面を図 6 に示す．図上部はピーク周波数付近の周波数スペクトルを表しており，図下部は周波数下方，ピーク周波数，周波数上方それぞれの音量の和の時間変化を，それぞれ青色，緑

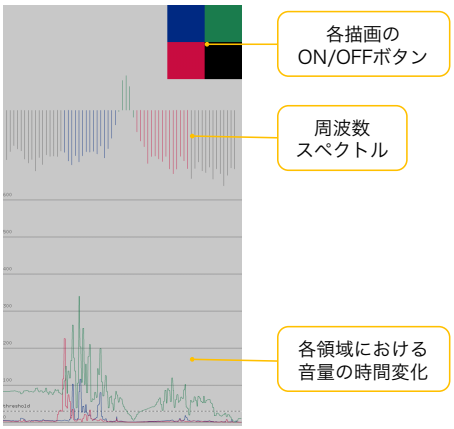

図 6 実装したアプリケーション画面

色，赤色で示している．なお，ピーク周波数音量は他の 2 つに比べて値が大きいため，画面内に収めるために，他 2 つとは異なるスケールで描画している．アプリケーション開発に用いた PC は Macbook Pro (CPU: Intel Core i7 3.1Ghz, RAM: 16GB) である．

5. 評価実験

実装したマイクカバーを用いて，評価実験を行った．本研究では，料理中や機械の作業中等のような手が汚れており，デバイスが直接触れない状況で，電子書籍のようにページをめくっていくレシピ/マニュアルを見ているような環境を想定する．従って，スマートフォンは縦に見る状態で，対象とする操作は左右のスワイプのみとした．図 7 に示すように，机の上にカバーを装着したデバイスを置き，デバイスを中心に，左右 10cm 程度の間を，利き手で平行に移動するようなジェスチャを行う．左から右，右から左へのジェスチャを，各カバーにつきそれぞれ 10 回ずつ行う．使用したカバーは，4 章で実装した 5 種類 (a，b，c 右側遮蔽，c 左側遮蔽，d) である．被験者は 24 歳から 27 歳の男性 3 名であり，全員右利きであった．

実験結果を表 1 に示す．この表に示すように，本研究で用いたカバーの中では，カバー d の認識率がもっとも良い結果となった．カバー d では，デバイスの右斜め上にマイクとスピーカの指向性を持たせることで，右から近付く動作と右へと離れていく動作の時に大きな反応が確認され，他のカバーよりも正確に識別することができた．

被験者 B，C の実験時には，左右どちらからのジェスチャでも周波数下方が大きく反応する傾向が見られ，マイクカバーにより左右からの取得音量に差を与える手法が効果的に働いていない場合が見られた．この原因としては，今回用いた認識手法では，ジェスチャが速ければより大きなドップラー効果が発生し，その分反応する音量も大きくなるため，これらの被験者のスワイプ動作中盤から終盤にかけての速度が大きく，マイクから離れる動作の方が強調されていたためであると考える．この問題の解決としては，

図 7 実験の様子

表 1 各ジェスチャの認識正解数

カバーの種類		a	b	c 右	c 左	d
被験者 A	左スワイプ	8	6	7	3	10
	右スワイプ	1	9	5	6	10
被験者 B	左スワイプ	5	3	10	0	5
	右スワイプ	0	10	0	8	10
被験者 C	左スワイプ	10	0	10	0	10
	右スワイプ	0	4	8	4	10
認識率 [%]		40.0	53.3	66.7	25.0	92.7

速度を考慮した認識アルゴリズムによる認識や，被験者各自のジェスチャにも個性があるため，練習フェーズのデータから個人にキャリブレーションしたような認識を考えている．しかし，本研究の認識アルゴリズムでも，カバー d においては被験者 A，B，C ともに比較的良い識別率が得られた．

以上より，今回実験したカバーの中では，カバー d のようにマイクとスピーカに指向性を付与するものが最適であったといえる．

5.1 考察

本研究では，左右のジェスチャの違いのみを識別したが，提案手法ではより多くのジェスチャを識別できる可能性がある．例えば，物体がマイクの近くで静止している時には取得できるピーク周波数の音量が増加するという特徴を利用し，スワイプの後にマイク付近で手を静止させることで，静止している間はスクロールを続けることや，マイクへの近づきと遠ざかりが同時に起こる，手首を回転させるローテーションのようなジェスチャや，速いスワイプと遅いスワイプを識別することなどを考えている．

本研究では，実験用のデバイスとして iPhone5 のみを使用したが，その他のスマートフォンや，スマートウォッチ，ヘッドマウントディスプレイのようなウェアラブルデバイスにおいてもデバイス内蔵のマイクとスピーカを用いたジェスチャ認識の可能性について調査を行う予定である．

今回用いたしきい値は予備実験からあらかじめ設定したものであるが，現在のピーク周波数音量からの割合で設定することにより，デバイスからの音量を変化させた際にも対応できる柔軟なしきい値の設定を行う予定である．

本研究の評価では，騒音が無く，周囲に人の動きの無い実験室環境で実験を行ったが，周囲雑音による誤認識や，周囲の他人の動きによる誤反応についても考慮していく必要がある．

本研究の被験者は 3 人とも右利きであったため，カバー d の右上への指向性と手のひらが直交するような形となり効果的に超音波が反射されたが，左利きの場合は超音波の指向性と手のひらが水平になるため，右利きよりも認識精度が落ちると想定される．効き手による影響が少ないカバーの形を今後調査していきたい．

6. おわりに

本研究では，既存デバイスのマイク/スピーカ部に取得/発信する音響特性が変化するようなカバーを装着し，従来の超音波による手法では困難であったジェスチャの識別を可能にする手法を提案した．5 種類のプロトタイプカバーを実装し，左右のスワイプの空中ジェスチャを識別したところ，最も効果的なカバーでは，平均 92.7%の精度で認識できた．

謝辞 本研究の一部は，日本学術振興会特別研究員奨励費 (15J04608) および科学技術振興機構戦略的創造研究推進事業 (さきがけ) および文部科学省科学研究費補助金挑戦的萌芽研究 (25540084) によるものである．ここに記して謝意を表す．

参考文献

[1] : Chirp Microsystems Technology, (online), available from ⟨http://www.chirpmicro.com/technology.html⟩ (accessed 2016-11-13).
[2] Aumi, M. T. I., Gupta, S., Goel, M., Larson, E. and Patel, S.: DopLink: using the doppler effect for multi-device interaction, *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM, pp. 583–586 (2013).
[3] Butler, A., Izadi, S. and Hodges, S.: SideSight: multi-touch interaction around small devices, *Proceedings of the 21st annual ACM symposium on User interface software and technology*, ACM, pp. 201–204 (2008).
[4] Chen, K.-Y., Ashbrook, D., Goel, M., Lee, S.-H. and Patel, S.: AirLink: sharing files between multiple devices using in-air gestures, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, pp. 565–569 (2014).
[5] Goel, M., Lee, B., Islam Aumi, M. T., Patel, S., Borriello, G., Hibino, S. and Begole, B.: SurfaceLink: using inertial and acoustic sensing to enable multi-device interaction on a surface, *Proceedings of the 32nd annual*

ACM conference on Human factors in computing systems, ACM, pp. 1387–1396 (2014).

[6] Gupta, S., Morris, D., Patel, S. and Tan, D.: Soundwave: using the doppler effect to sense gestures, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 1911–1914 (2012).

[7] Harrison, C., Tan, D. and Morris, D.: Skinput: appropriating the body as an input surface, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 453–462 (2010).

[8] Ishikawa, T., Horry, Y. and Hoshino, T.: Touchless input device and gesture commands, *2005 Digest of Technical Papers. International Conference on Consumer Electronics, 2005. ICCE.*, IEEE, pp. 205–206 (2005).

[9] Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I. and Olivier, P.: Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor, *Proceedings of the 25th annual ACM symposium on User interface software and technology*, ACM, pp. 167–176 (2012).

[10] Kratz, S. and Rohs, M.: HoverFlow: expanding the design space of around-device interaction, *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ACM, p. 4 (2009).

[11] Laput, G., Brockmeyer, E., Hudson, S. E. and Harrison, C.: Acoustruments: Passive, acoustically-driven, interactive controls for handheld devices, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, pp. 2161–2170 (2015).

[12] Laput, G., Xiao, R., Chen, X., Hudson, S. E. and Harrison, C.: Skin buttons: cheap, small, low-powered and clickable fixed-icon laser projectors, *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM, pp. 389–394 (2014).

[13] Manabe, H.: Multi-touch gesture recognition by single photoreflector, *Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology*, ACM, pp. 15–16 (2013).

[14] Nandakumar, R., Iyer, V., Tan, D. and Gollakota, S.: FingerIO: Using Active Sonar for Fine-Grained Finger Tracking, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, pp. 1515–1525 (2016).

[15] Ono, M., Shizuki, B. and Tanaka, J.: Touch & activate: adding interactivity to existing objects using active acoustic sensing, *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM, pp. 31–40 (2013).

[16] Song, J., Sörös, G., Pece, F., Fanello, S. R., Izadi, S., Keskin, C. and Hilliges, O.: In-air gestures around unmodified mobile devices, *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM, pp. 319–329 (2014).

[17] Tarzia, S. P., Dick, R. P., Dinda, P. A. and Memik, G.: Sonar-based measurement of user presence and attention, *Proceedings of the 11th international conference on Ubiquitous computing*, ACM, pp. 89–92 (2009).

[18] Xiao, R., Lew, G., Marsanico, J., Hariharan, D., Hudson, S. and Harrison, C.: Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation, *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, ACM, pp. 67–76 (2014).

[19] Yang, X.-D., Hasan, K., Bruce, N. and Irani, P.: Surround-see: enabling peripheral vision on smartphones during active use, *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM, pp. 291–300 (2013).