

スピーカと複数マイクロフォンアレイを用いた 宅内における住人存在検知手法の検討

米岡 尚樹¹ 荒川 豊¹ 安本 慶一¹

概要：近年，Amazon Echo や Google Home のような市販のスマートスピーカが広く普及している．これらのスマートスピーカは音声を利用したインターフェイスを介して住人から命令を受け，住人の生活を消費行動の代行や家電の制御などといった様々な形で支援する．その一方で，スマートスピーカが住人の発する声と機械の発する声を区別できないことを悪用し，宅内に設置された他のデバイスから音声を流すことによって，スマートスピーカから住人の個人情報盗む，スマートスピーカに住人の望まない商品を買わせるなどの操作をできてしまうことが報告されている．さらに，人間には聞こえない超音波を使って同様の攻撃を行う DolphinAttack という攻撃も報告されている．そのため，スマートスピーカに対して人間が命令しているのか否かを何らかの方法で判定することが望まれている．そこで本研究では，ソナー音を発信するスピーカとそれを受信する複数のマイクロフォンアレイを組み合わせて，スマートスピーカの周囲に住人が存在するかどうかを検知するシステムを提案する．提案システムでは，直交周波数多重分割方式 (OFDM) で生成したソナー音をスピーカから周囲の全方向に発信し，その反射波をスピーカの真上に設置した 8ch マイクロフォンアレイで受信する．評価として，搬送するパワーを比較する実験を行った結果，約 0.5 秒間のパワー計測によって人の有無を十分に検知できる可能性が示された．

Estimating User's Existence by Reverberation Analysis with a Smart Speaker and Microphone Array

NAOKI YONEOKA¹ YUTAKA ARAKAWA¹ KEIICHI YASUMOTO¹

1. はじめに

近年，音声インターフェイスを搭載したスマートスピーカが Google や Amazon といった企業を筆頭に多く製造され，2018 年第 2 四半期における全世界での出荷数は 1680 万台に達している [1]．現在のペースで市場が拡大していけば，2020 年には 9000 万台のスマートスピーカが全世界の家庭に普及すると予測されている [2]．

スマートスピーカは無指向性のスピーカとマルチチャンネルのマイクロフォンアレイから成る装置であり，電源と接続されている限り常に周囲の音を受信する．これはウェイクワードと呼ばれる言葉 (Amazon Echo であれば Alexa，Google Home であれば OK，Google を指す) をユーザか

ら受信するためであり，スマートスピーカはウェイクワードの受信をトリガとしてサービスの提供を開始する．ウェイクワード以降のユーザ音声はインターネットを通じてクラウドへ送られ，音声認識アルゴリズムによって意味内容を解析される．スマートスピーカはクラウドで分析されたユーザの言葉に応じた回答をユーザへ音声でフィードバックし，ユーザの要求に答える形でサービスを提供する．スマートスピーカが提供可能なサービスは，目覚ましの設定やニュースの読み上げ，家電制御やインターネットショッピングなど多岐にわたり，手軽なインターフェイスで多様にユーザの生活を支援する．

一方で，スマートスピーカは常にインターネットと接続されており，常に周囲の音を拾っているがゆえに，ユーザの意図しない動作を引き起こしやすいという問題がある．実際の例として，TV 番組におけるウェイクワードの発言

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

を受け、視聴者のスマートスピーカがドールハウスを注文した事件 [3] が過去に発生している。音声インターフェイスが抱えるこの問題に対しては、これまでに様々な研究者が可能な攻撃方法を提唱している。Roy ら [4] は、マイクがもつ高周波数特性の非線形性を利用することにより、非可聴領域の超音波で話しかけてもスマートデバイス进行操作できることを示した。この手法ではスピーカを二つ用いる必要があったが、Zhang ら [5] が同様の手法を拡張したことで単一のマイクによる実現が可能となり、より汎用性の高い攻撃方法となった。このような超音波による攻撃は合成音声によって行われるが、Google や Microsoft が提供する音声認識アルゴリズムでは個人の音声を識別できるため、合成音声を受け付けられないよう設定することは可能である。しかし、Diao ら [6] は、ユーザのボイスアシスタント機能に対する発言内容がスマートフォンやクラウドに保存されていることを利用し、それを任意に再生することでユーザの意図しない動作を引き起こせることを示した。音声インターフェイスがこの種の攻撃から自衛するためには、ユーザの音声ユーザから発せられたものかどうかを識別する必要がある。

しかしながら、音声スピーカから発せられたものか、人体から発生されたものかを音響的に識別することは困難である。そこで我々は、まず、スマートスピーカの周囲に人間がいるかどうかを認識できないかと考えた。その手法として、カメラや人感センサなどの新たな装置をスマートスピーカに追加することが容易に考えられる。しかしながら、普及の観点から考えると、追加デバイスを装着することは好ましくない。

本稿では、スマートスピーカに搭載されているスピーカとマイクロフォンアレイのみで、スマートスピーカの周辺に人が存在しているかどうかを認識するシステムを提案する。提案システムでは、スマートスピーカのスピーカからソナー音を周囲の全方向に発信し、その反響音の解析によって、人の存在を判定する。このとき、ソナー音は直交周波数多重分割方式 (OFDM) を用いて生成する。これにより、受信信号からソナー音の受信時刻を正確に計算できる。受信信号にはスピーカから直接到来するソナー音と、部屋や人に反響してから到来するソナー音が混在しているが、直接音の受信タイミングが一定であることから、信号の減算によって直接音を消去できる。残る反響音からは、各マイクロフォン素子が受信した音の平均パワーを求めることができる。この平均パワーは平均計算に用いた受信信号の時間長依存性をもつため、それを比較すると、人の有無によって反響音の大きさに違いがあることから、人のいる場合とない場合とでは平均パワーの大きさが異なる。そのため、平均パワーの値に一定の閾値を設けることで、人の有無を判定できる。

提案システムを評価するため、ソニー株式会社製のス

スマートスピーカ LF-S50G と株式会社システムインフロンティア製の 8ch マイクロフォンアレイ TAMAGO-03 を用い、擬似的なスマートスピーカを作成した。そして、その装置を部屋の中心に設置した実験環境を構築し、部屋に人がいる場合とない場合とで計測を行った。両方の場合において 30 秒間ソナー音の送受信を行い、受信音に提案手法を適用して分析を行った結果、人のいる場合とない場合とで平均パワーの大きさに明らかな違いが見られ、両者は先行研究に比べて十分に短い時間長で分離可能であることがわかった。

以降では、第 2 章に関連研究を示し、第 3 章で提案システムについての詳細を説明する。第 4 章では実験環境と実験条件について示し、第 5 章では実験で得られたデータの分析について、その方法と実際の分析結果、および考察を示す。おわりに、第 6 章で結論を述べる。

2. 関連研究

宅内の人を検知する研究は、屋内位置推定の分野でこれまでに多く行われている。Bocca ら [7] は宅内に 30 個以上のセンサを設置することにより、電波を用いたセンサ間の通信ネットワークで宅内に存在する複数の人間を検知した。この手法では、90%以上の対象を誤差 1m 以内の精度で検知できている。Adib ら [8] も電波を用い、縦横の大きさ 1m × 2m のアンテナアレイを宅内に 2 つ設置することで複数の人間を検知した。この手法では、99%以上の対象を誤差 1m 以内の精度で検知できている。これらの研究は電波を用いて宅内にいる複数の人間を高い精度で検知できることを示すものであるが、いずれも自作の特殊なセンサを複数設置する、あるいは自作の大きなアンテナを宅内に設置する必要があるといった点で汎用性に乏しい。

一方で、電波ではなく音波を用いることによって、スマートデバイスが持つ機能を利用して簡便に実装できる人検知手法が Nandakumar ら [9] によって提案されている。この手法では、発信信号としてスマートデバイスでも再生可能なソナー音を用い、スマートフォンが持つ 2 つのマイクでそれを受信することによって人検知を実現する。しかし、マイクを 2 つしか用いないことから複数の人を検知できないことが問題となっている。

人の検知に焦点を当てた研究としては、Alanwar ら [10] のソナー音を用いたものが挙げられる。Alanwar らは宅内のスマートスピーカにウェイクワードの話しかけがあった場合を想定し、宅内に人がいる場合とない場合のそれぞれにおいて、スマートスピーカのマイクを模した 8ch マイクロフォンアレイで無指向性スピーカから宅内に発信したソナー音を受信した。ここで受信したソナー音から様々な統計量を計算し、機械学習アルゴリズムを用いて統計量のクラスタリングを行うことにより、人の有無を推定した。この手法では、16 秒間の受信音から計 50 個の統計量を用

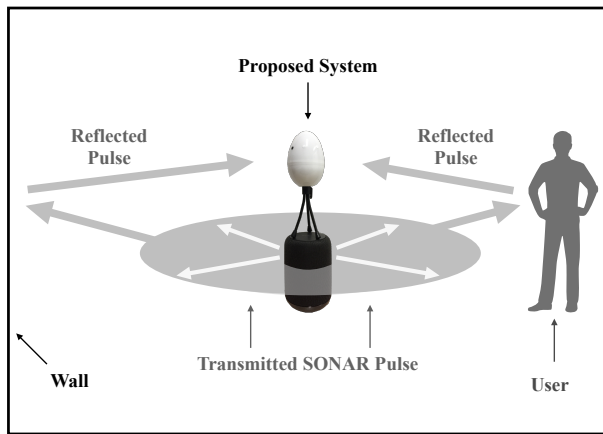


図 1 提案システムの概要

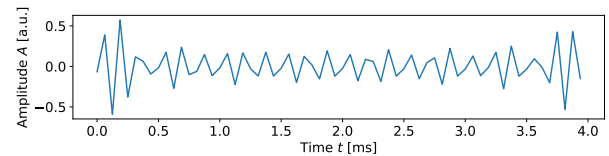


図 4 使用したソナー音

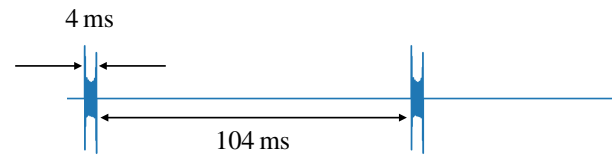


図 5 ソナー音の長さや発信間隔



図 2 使用したスピーカ



図 3 使用したマイクアレイ

いてクラスタリングを行った結果、93.13%の精度で人の有無を推定できている。しかし、クラスタリングに用いる受信音の長さが 16 秒より長すぎても短すぎても精度は改善せず、少なくとも測定開始から 16 秒以上経たなければ人の有無を検知できない点が問題点として挙げられている。そのため、より短い時間長の受信音を用いて人の検知ができる手法が必要だと考えられる。

そこで本研究では、宅内のスマートスピーカに周囲の人の有無を検知させることを目的として、宅内に発信したソナー音をマイクで受信した結果を用い、短い時間での人検知方法について検討する。

3. 提案システム

本項では、ソナー音を用いた人検知のために提案するシステムについて詳細を述べる。始めにシステムを構成する機器について述べ、次にスピーカから発信するソナー音について説明する。図 1 に提案システムの概要図を示す。

3.1 構成機器

本研究では、無指向性スピーカの上にマルチチャンネルマイクロフォンアレイを乗せることで、擬似的なスマートスピーカを構成し、提案システムとする。無指向性スピーカとマルチチャンネルマイクロフォンアレイを組み合わせた構成は、実際に市販されている Amazon Echo や Google

Home などのスマートスピーカと同様のものである。ただし、スピーカの向きやマイクロホンアレイのマイクの数の観点から、今回は下記に示す機器構成とした。

図 2 と図 3 に、提案システムを構成するスピーカとマイクロフォンアレイを示す。スピーカは、横方向に音を出す形態が望ましいため、円筒形で周囲にスピーカを備えるソニー株式会社の LF-S50G を選択した。しかしながら、LF-S50G は、マイクを 2 つしか搭載していない (Amazon Echo には 7 つのマイクロホンアレイが搭載されている)。そこで、マイクロフォンアレイは、株式会社システムインフロンティアの TAMAGO-03 を用いた。TAMAGO-03 は水平な同一平面上において、方位角方向に等間隔に配置された 8 つのマイクロフォン素子を持つマイクロフォンアレイである。サンプリング周波数は 16 kHz で、8 つのチャンネルは全て同期サンプリング可能である。

3.2 ソナー音

本提案システムでは、図 4 に示したソナー音を用いる。このソナー音は 6000Hz から 7750Hz までの、250Hz の間隔でとった 8 つの同パワーかつ同位相の正弦波を加算することによって作成した。この構成は直交周波数多重分割方式 (OFDM) に倣ったものであり、その方式で生成したソナー音を用いると受信信号との相互相関が強く出ることがわかっているため、受信信号からソナー音の受信時刻を推定することが容易となる [9]。ソナー音のサンプリング周波数は TAMAGO-03 のサンプリング周波数に合わせ、16 kHz とした。また、ソナー音一つあたりの時間長は 4 ms である。図 5 に示すように、提案システムの動作中、ソナー音は一定の時間間隔 104 ms で連続的に発信する。

4. 実験方法

本章では、提案システムを用いた人検知のために行った実験について詳細を述べる。始めに実験環境として実験に用いた部屋について述べ、次に実験内容として実験条件の設定パターンについて説明する。



図 6 実験に利用した部屋。
中心に提案システムを置いている。

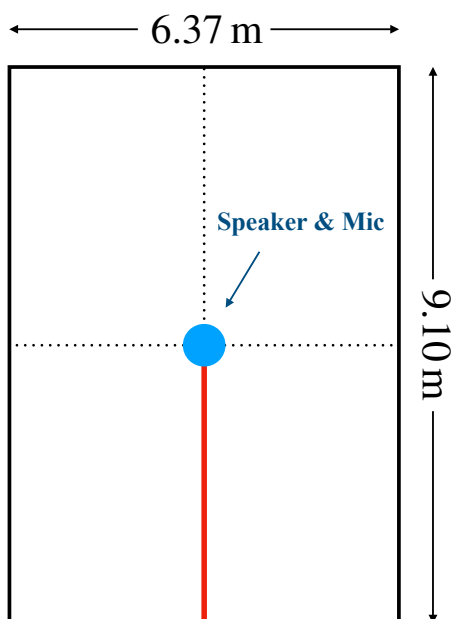


図 7 実験に利用した部屋の寸法
中心の青丸は提案システムの位置，
赤線はマイクの ch1 が向く方向を表す。

4.1 実験環境

宅内での人検知のため，本研究では実際の宅内と同等の広さを持ち，家具等の障害物が任意に配置され，エアコンが動作している屋内の部屋において実験を行った。

図 6 に実験で用いた部屋の写真を示す。部屋の寸法は図 7 に示す通りで，部屋の中心に提案システムを構成するスピーカとマイクを置いた。

4.2 実験条件

本節では，実験における人の配置方法を，実験状況の設定パターンとして説明する。

本実験は，提案システムを用いて，部屋に人がいない場合，部屋に人が 1 人いる場合，部屋に人が 2 人いる場合のそれぞれについて行った。また，人がスピーカの近くにい

る場合と遠くにいる場合のそれぞれについても実験を行った。人の配置パターンはそれぞれパターン A,B,C,D,E と名付け，その概要を図 8 に示す。また，実際の配置例としてパターン B とパターン C の場合の写真を図 9 と図 10 に示す。なお，図 8 中の赤線は図 3 に示した中で黒く塗られたマイクロフォン素子 ch1 の正面に当たる方向である。

全てのパターンにおいて実験は 30 秒間行い，その間ソナー音は一定の間隔で発信し続けた。以下ではそれぞれのパターンについて説明する。

パターン A 部屋に人がいない場合

部屋に人がいない場合の実験状況である。図 8(a) に示すように，部屋に人がいない状態を 30 秒間保ち，ソナー音の発信と受信を行った。

パターン B 部屋の中心から 0.5 m 離れて 1 人いる場合

部屋に人を 1 人入れ，図 8(b) に示すように，マイクロフォン素子 ch1 の正面方向へスピーカから 0.5 m 離れた位置に配置した。図 9 に示した配置例からわかるように，この距離はスピーカにできるだけ近づいた時の距離である。後述する分析方法では，ソナー音を送受信する 30 秒間における部屋の環境が定常だとみなして計算を行うため，部屋に入れた人には 30 秒間直立不動の状態を保たせたまま，ソナー音の発信と受信を行った。

パターン C 部屋の中心から 2.5 m 離れて 1 人いる場合

部屋に人を 1 人入れ，図 8(c) に示すように，マイクロフォン素子 ch1 の正面方向へスピーカから 2.5 m 離れた位置に配置した。図 10 に示した配置例からわかるように，この距離はスピーカからできるだけ離れた時の距離である。そこでパターン B と同様に 30 秒間直立不動の状態を保たせたまま，ソナー音の発信と受信を行った。

パターン D 部屋の中心から 0.5 m 離れて 2 人いる場合

部屋に人を 2 人入れ，図 8(d) に示すように，一人はマイクロフォン素子 ch1 の正面方向へスピーカから 0.5 m 離れた位置に配置し。もう一人は ch7 の正面方向へスピーカから 0.5 m 離れた位置に配置した。そこで二人とも 30 秒間直立不動の状態を保たせたまま，ソナー音の発信と受信を行った。

パターン E 部屋の中心から 2.5 m 離れて 2 人いる場合

部屋に人を 2 人入れ，図 8(e) に示すように，一人はマイクロフォン素子 ch1 の正面方向へスピーカから 2.5 m 離れた位置に配置し。もう一人は ch7 の正面方向へスピーカから 2.5 m 離れた位置に配置した。そこでパターン D と同様に二人とも 30 秒間直立不動の状態を保たせたまま，ソナー音の発信と受信を行った。

5. 分析

本章では，実験で得られたデータの分析方法について述

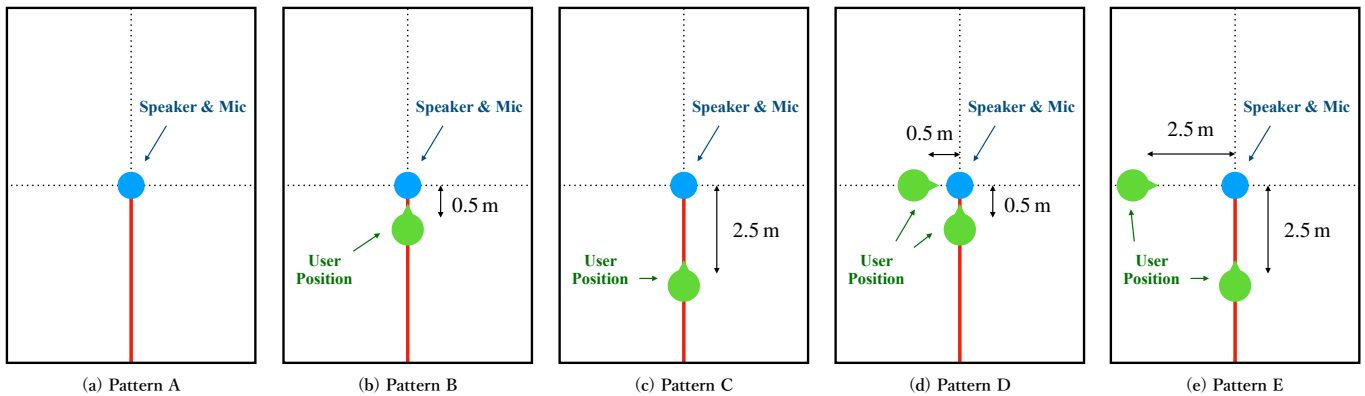


図 8 実験パターン



図 9 パターン B の配置例



図 10 パターン C の配置例

べる．始めにデータの処理と分析の方法について詳細を説明し，次に実際の分析結果と考察内容を示す．

5.1 方法

データの分析は次の 4 つのステップを踏んで行った．

- (1) 受信音とソナー音との相互相関の計算
- (2) 相互相関のフィルタリングと切り分け
- (3) ベースラインの設定と差の計算
- (4) 相互相関の平均パワーの計算と比較

ステップ 1 から 4 まで各パターンについて独立に計算を行い，ステップ 4 の最後に各パターンで得られた相互相関の平均パワーをパターン間で比較し，これを結果とした．以下では，それぞれのステップについて詳細を説明する．

ステップ 1 受信音とソナー音との相互相関の計算

マイクロフォンアレイで受信した音には無指向性スピーカから発信されたソナー音が含まれるが，宅内のように音の攪乱要因が多くノイズが乗りやすい環境では，受信音からソナー音の受信時刻を直接計算するこ

とは難しい．そのため，受信音とソナー音との相互相関をとることによって，ソナー音の受信時刻の計算を容易にする．ここで時刻を t とし，受信音を $x(t)$ ，ソナー音を $s(t)$ とすると，時刻 t における受信音とソナー音との相互相関 $\phi(t)$ は

$$\phi(t) = \frac{1}{N_s} \sum_{t'=0}^{N_s-1} x(t)s(t-t')$$

と表せる．ここで N_s はソナー音のサンプル数である．受信音の時刻 t においてソナー音と相関の高い波形が存在すると $\phi(t)$ は大きな値を示すため， $\phi(t)$ の極値からソナー音の受信時刻を計算できる．

ステップ 2 相互相関のフィルタリングと切り分け

受信音とソナー音の相互相関はソナー音受信時刻の計算に特化しているが，人検知に必要と考えられる物理量を計算するためには，さらにノイズなどのソナー音と無関係な音の影響を取り除く必要がある．本実験では 6000Hz 以上の 8 種類の周波数帯を用いるが，その中で最も大きな音量で受信できた周波数帯を最もロスが少なかった周波数帯として，ここでは人検知にで用いるために注目する周波数帯とみなす．そこで相互相関をバンドパスフィルタにかけることで，注目する周波数帯以外のノイズを落とす．さらに約 30 秒間の時系列データである相互相関を，一回のソナー音受信ごとに切り分ける．従って，これ以降相互相関は 1 回分のソナー音から得られた相互相関を 1 処理単位とし，

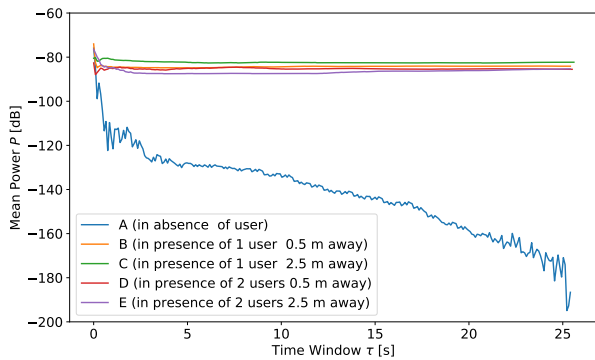


図 11 各パターンにおける平均パワーと受信音の時間長との関係
 縦軸は平均パワー P [dB] を、横軸は P の計算に使用した受信音の時間長 τ [s] を表す。

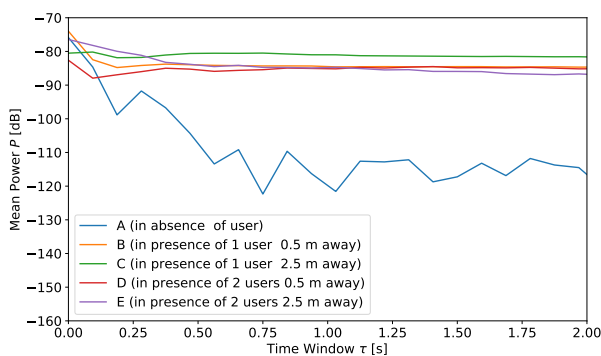


図 12 図 11 を領域 $0 \leq \tau \leq 2$ について拡大した図

その集合として取り扱う。1 処理単位の相互相関はソナー音の発信間隔程度の長さをもつ時系列データである。

ステップ 3 ベースラインの設定と差の計算

マイクロフォンアレイは受信音として、スピーカから到来する直接音と、スピーカから発信され宅内の環境に反響してから到来する間接音とを、どちらも含んでいる。人がいる場合といない場合とでは、ソナー音が人に反響してから到来する間接音の有無が関わり、直接音には違いが現れないと考えられる。そこで、人がいない場合、即ちパターン A で全処理単位の相互相関を 8 つのチャンネルそれぞれで平均し、これを各チャンネルの相互相関のベースラインとして設定する。このベースラインを用い、各パターンで相互相関とベースラインとの差を平均することによって、相互相関から直接音の影響と部屋の音響特性を打ち消す。

ステップ 4 相互相関の平均パワーの計算と比較

ベースラインとの差をとった相互相関は、処理単位ごとに位相差等に由来するゆらぎが存在すると考えられる。そのためこれをいくつかの処理単位で平均してゆらぎの影響を消去する。このようにして加工した相互相関を二乗してチャンネルごとに時間平均をとり、さらに全チャンネルに渡って平均することで、各マイク

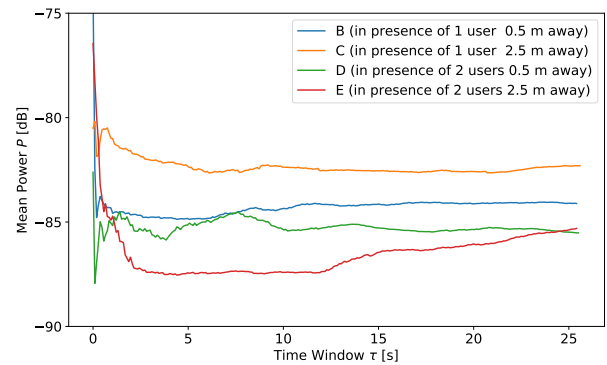


図 13 図 11 からパターン A の結果を除いた関係

ロフォン素子が受信したソナー音による搬送パワーが得られる。この値はゆらぎ消去の際に平均した処理単位の数によって変わってくるため、計算に利用する処理単位の数、すなわち計算に利用する受信音の時間長の関数として計算する。最終的に、それぞれのパターンで得られた平均パワーと計算に用いた受信音の時間長との関係を、パターン間で比較する。

5.2 結果

図 11 に、平均パワー P とその計算に利用した受信音の時間長 τ との関係を、各実験パターンごとに示す。図からわかるように、パターン A、即ち人がいない場合の平均パワーは、他のパターン、即ち人がいる場合の平均パワーと比べて明らかに小さい。図 12 に、図 11 の τ が小さい領域を拡大した図を示す。この図からわかるように、パターン A の P は τ が 1s に満たない時点から他のパターンにおける P と十分に分離可能である。

図 13 に、図 12 からパターン A の結果を除いた図を示す。パターン B,C は人が 1 人いる場合、パターン D,E は人が 2 人いる場合の平均パワーを表すが、人の配置距離によっては人数ごとに平均パワーを分離できることがわかる。しかし、パターン E の結果からわかるように平均パワーが τ に対して一定値に近づくとは言えず、一定値に近づくとしてもその値と人の配置距離との間に関係性は認められない。

5.3 考察

人がいる場合といない場合の P は τ が 1s に満たない時点で十分に分離可能であることから、適切に平均パワーの閾値を設けることによって、ソナー音の発信開始から 1s 以内には人の有無を識別できると考えられる。図 11 を参考に閾値を設定することを考えると、この場合は平均パワーの閾値を -100 dB に設定することで、ソナー音の発信開始から 0.5s 後には人検知ができる可能性がある。受信音の生データから得た統計量をクラスタリングして人検知を試みた Alanwar ら [10] の手法において、人検知までに必

要な時間はソナー音の発信開始から 16s であった．これに対し，本研究で示した平均パワーに着目する方法は，従来手法よりも速やかに人検知をできる可能性があり，平均パワーの大きさは迅速な人検知に有効であると考えられる．

一方で，平均パワーと受信音時間長との関係が，人数と距離の変化に対して単調ではない変化をすることから，平均パワーを用いて人数や距離の推定を同時に行うことは難しいと考えられる．ただ，図 13 において人数差だけに着目した場合，人が 1 人いる場合と 2 人いる場合の P は $\tau > 10$ で十分に分離できるとみなせるため，ソナー音の発信開始から十分時間が経てば人数を識別できる可能性がある．しかし P の τ 依存性に関しては，パターン E で得られた結果から， τ がさらに大きい場合の P の挙動が単調でない可能性が示唆されているため，慎重な検討が必要である．

人がいるかいないかが識別できれば，人がいない時にウェイクワードの呼びかけがあった場合は，その声を無視するといった機能を作れるようになると考えられる．さらに，人の存在検知に加えて周囲の人の位置や人数がわかれば，人がいない場所からの音を無視しつつ，人がいる場所からの音を聴くといった機能を作れる可能性がある．このような機能の実現に向けて，今後スマートスピーカが人の位置や人数を把握できるようにするためには，部屋に配置する人数や距離のパターンを増やして P の τ 依存性に関する法則を見出すことや，各チャンネルが受信したエネルギーに到来角推定の手法を適用して位置推定のアプローチを増やすことが必要になってくると考えられる．

6. おわりに

本研究では，スマートスピーカに対して音声の呼びかけがあった際に，スマートスピーカが周囲に人がいるかどうかを検知するための手法を提案した．提案手法では，スピーカから発信したソナー音をマルチチャンネルマイクロフォンアレイで受信し，そこからスピーカより到来する直接音の影響を消去したのち，残る宅内の反響音について平均パワーを計算して，これを人検知の指標とした．平均パワーはソナー音の発信開始から逐次計算し，平均パワーが事前に設定した閾値を下回るか否かを見ることによって，ソナー音の発信開始からごく短い時間で人検知ができる可能性があることを示した．提案手法による人検知の速度は従来手法に比べて十分速やかであり，平均パワーの大きさから人の有無を判定する提案手法は迅速な人検知に有効であることがわかった．

謝辞

本研究の一部は，科学研究費補助金国際共同研究加速基金(国際共同研究強化)(15KK0011)の研究助成に基づくものである．ここに記して謝意を示す．

参考文献

- [1] Canalys : Global smart speaker shipments grew 187% year on year in Q2 2018, with China the fastest-growing market(online), 入手先 <https://www.canalys.com/static/press_release/2018/Press-release-160818-global-smart-speaker-shipments-grew.pdf> (2018.11.05).
- [2] 総務省 : 平成 30 年度 情報通信白書 (online), 入手先 <<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/pdf/n1100000.pdf>> (2018.11.05).
- [3] TheVerge : Amazon 's Alexa started ordering people dollhouses after hearing its name on TV(online). 入手先 <<https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse>> (2018.11.05).
- [4] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury : Backdoor: Making microphones hear inaudible sounds. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2-14(2017).
- [5] Guoming Zhang, Chen Yan, Xiaoyu Ji, Taimin Zhang, Tianchen Zhang, and Wenyuan Xu : Dolphinattack: Inaudible voice commands. *CoRR*, abs/1708.09537(2017).
- [6] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang : Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*. ACM, 63-74(2014).
- [7] Maurizio Bocca, Ossi Kaltiokallio, Neal Patwari, and Suresh Venkatasubramanian : Multiple target tracking with RF sensor networks. *Mobile Computing, IEEE Transactions*(2013).
- [8] Fadel Adib, Zach Kabelac, and Dina Katabi : Multi-Person Localization via RF Body Reflections, In *Proceedings of USENIX NSDI*(2015).
- [9] Rajalakshmi Nandakumar, Alex Takakuwa, Tadayoshi Kohno, and Shyamnath Gollakota. CovertBand: Activity Information Leakage using Music. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3, 87(2017).
- [10] Amr Alanwar, Bharathan Balaji, Yuan Tian, Shuo Yang, and Mani Srivastava. EchoSafe: Sonar-based Verifiable Interaction with Intelligent Digital Agents. In *Proceedings of the 1st ACM Workshop on the Internet of Safe Things (SafeThings'17)*, Rasit Eskicioglu (Ed.). ACM, New York, NY, USA, 38-43(2017).