

修士論文

スマートスピーカへの偽音声攻撃に対する
防御に向けたソナー音に基づく人検知手法

米岡 尚樹

2019年3月15日

奈良先端科学技術大学院大学
情報科学研究科

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

米岡 尚樹

審査委員：

安本 慶一 教授 (主指導教員)

林 優一 教授 (副指導教員)

荒川 豊 准教授 (副指導教員)

藤本 まなと 助教 (副指導教員)

スマートスピーカへの偽音声攻撃に対する 防御に向けたソナー音に基づく人検知手法*

米岡 尚樹

内容梗概

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を盗む, 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, スマートスピーカの周囲に人がいないにも関わらずスマートスピーカに対し命令がなされるといった状況を検知しユーザに警告することを目指し, スマートスピーカの機能 (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) だけを用いて, 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. 受信信号にはスピーカから直接到来する直接音と, 部屋や人に反射してから到来する反響音が混在しているため, ソナー音との相互相関関数を減算する

*奈良先端科学技術大学院大学 情報科学研究科 修士論文, 2019 年 3 月 15 日.

ことによって直接音を消去する。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、提案システムの動作開始からの経過時間依存性を比較すると、人の有無によって反響音に違いがあることから、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。

提案手法による人検知の可能性を確かめるため、二種類の異なる部屋において、無人を含む複数の人配置パターンでソナー反響音データを収集した。収集したデータを用いて計算した平均パワーを配置パターン間で比較したところ、いずれの部屋においても人がいる時といない時とで平均パワーの大きさに十分な差があり、人を検知できることが分かった。また、平均パワーと提案システムの動作開始からの経過時間との関係を確認したところ、約2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ことが分かった。

キーワード

セキュリティ, 音声アシスタント, スマートスピーカ, スマートホーム, IoT

Sonar-based Detection of Surrounding Users for Defending against False Voice Command Attack on Smart Speaker*

Naoki Yoneoka

Abstract

Recently, smart speakers like Amazon Echo and Google Home have been spreading widely. These devices support users' life through voice interface by receiving voice commands to operate appliances and order goods from online shops. Meanwhile, it is reported that smart speakers are vulnerable to some malicious attacks that can steal personal information and/or order unnecessary goods by uttering voice from a device nearby the speaker, exploiting the fact that the smart speakers cannot distinguish human voice from machine voice. A new type of attack called DolphinAttack which utters ultrasonic voice inaudible to human is also reported. Therefore, a method to identify which of human or machine is sending voice commands to a smart speaker is desired. In this thesis, to detect an unnatural state that voice commands are sent to smart speakers despite the absence of users and warn users, we propose a system consisting of a speaker and a microphone array built in smart speakers to detect the existence of a human nearby. In our proposed system, the speaker emits sonar pulse generated based on Orthogonal Frequency Division Multiplexing (OFDM) in all directions and the microphone array with 8 channels attached on top of the speaker receives the reverberated sound.

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, March 15, 2019.

Since the received sound includes both the direct sound from the speaker and the reverberated sound via an indoor environment and users, we can cancel the direct sound by subtracting the cross-correlation function with the sonar pulse, and calculate the mean power of received sound in each microphone from the reverberated sound. Comparing the mean power in the case of users present with the case of no users, we can classify the mean power into cases of users present or absent by setting a certain threshold, since there should be some difference in the reverberated sound depending on the users' presence/absence.

In order to evaluate the feasibility of the proposed system, we collected the reverberated sound data in two different rooms with several patterns (one person and two persons with different positions, and no person). As a result of comparison of the mean power calculated from the collected data between the patterns, there is a difference enough to distinguish the mean power depending on presence/absence of users in each room. Moreover, we found that the proposed system can judge users' presence within 2 seconds after emitting the sonar sound.

Keywords:

Security, Voice Assistant, Smart Speaker, Smart Home, IoT

目次

1. 序論	1
2. 想定環境	4
2.1 スマートスピーカ	4
2.1.1 スマートスピーカ概念と定義	4
2.1.2 市販されているスマートスピーカの例	5
2.2 脅威モデル	6
3. 提案システム	7
3.1 目的	8
3.2 システム構成	8
3.3 構成機器	8
3.4 ソナー音の仕様	11
3.5 人検知手法	13
3.5.1 概要	13
3.5.2 人検知ステップ	13
4. 提案システムによる人検知可能性の検証	16
4.1 実験目的と実験方法	16
4.1.1 実験目的	16
4.1.2 実験環境	16
4.1.3 実験条件	18
4.2 実験結果	22
4.3 考察	24
5. 人検知判定基準の検証	25
5.1 実験目的と実験方法	25
5.1.1 実験目的	25
5.1.2 実験環境	25

5.1.3 実験条件	27
5.2 実験結果	27
5.3 考察	31
6. 関連研究	33
6.1 電波を用いた人検知	33
6.2 音波を用いた人検知	33
6.3 スマートスピーカを用いた人検知	34
7. 結論	35
謝辞	37
参考文献	38

図 目 次

1	提案システムの概要	7
2	使用したスピーカ	10
3	使用したマイクロフォンアレイ	10
4	使用したソナー音	12
5	ソナー音の長さとの発信間隔	12
6	人検知可能性の検証実験に利用した部屋	17
7	人検知可能性の検証実験に利用した部屋の寸法	17
8	人検知可能性の検証実験パターン	20
9	パターン B の配置例	21
10	パターン C の配置例	21
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	22
12	図 11 を領域 $0 \leq \tau \leq 2$ について拡大した図	23
13	図 11 からパターン A の結果を除いた関係	23
14	人検知判定基準の検証実験に利用した部屋	26
15	人検知判定基準の検証実験に利用した部屋の寸法	26
16	$\tau = 0.1 \text{ s}$ における平均パワーと人-システム間距離との関係	28
17	$\tau = 1.5 \text{ s}$ における平均パワーと人-システム間距離との関係	29
18	$\tau = 20 \text{ s}$ における平均パワーと人-システム間距離との関係	29
19	平均パワーの閾値に依る最大検知可能距離と受信音の時間長との 関係	30

表 目 次

1. 序論

近年、音声ユーザ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、ユーザの要求に応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。音声ユーザ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 [4, 5, 6, 7, 8]。Roy ら [9] は、マイクロフォン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 [10] が同様の手法を拡張したことで単一のマイクロフォンによる実現が可能となり、DolphinAttack と名付けられたこの手法は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ないよう設定することは可能である。しかし、Diao ら [11] 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。音声ユーザ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、音声ユーザから発せられたものか、人体から発生されたものかを音声の特徴から識別することは困難である。従って、スマートスピーカが自身への攻撃から自衛するためには、音声の特徴を用いる以外の手法を用いることにより、音声ユーザから発せられたものか否かを識別する必要がある。

スマートスピーカが攻撃を受けている状況の一つとして、部屋に人がいないにも関わらずウェイクワードの話しかけがなされるといった状況が挙げられる。こうした状況をスマートスピーカが不自然だと判断して自衛を行うためには、スマートスピーカの周囲にいる人の有無を検知できることが必要だと考えられる。さらに DolphinAttack のような攻撃を受けて話しかけ音声の音源位置と住人の位置が異なるような状況をスマートスピーカが不自然だと判断するためには、人の有無だけでなく、人の位置も検知できることが必要になると考えられる。人の有無を検知し、将来的には人の位置も検知できるようにすることを考えると、容易な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挙げられる。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。

本論文では、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おける人の有無を検知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，人のいる場合とい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．

提案手法による人検知の可能性を確かめるため，ソニー株式会社製のスマートスピーカ 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 を組み合わせ，擬似的なスマートスピーカを作成し提案システムとした．これを用いて，二種類の異なる部屋において無人を含む複数の人配置パターンで計測を行った．全ての人配置パターンにおいて 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，いずれの部屋においても人がいる時といない時とで平均パワーの大きさに十分な差があり，人を検知できることが分かった．また，平均パワーと提案システムの動作開始からの経過時間との関係を確認したところ，先行研究に比べて十分に短い約 2 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ことが分かった．

以降では，第 2 章で本研究において想定する脅威モデルをスマートスピーカの説明と合わせて設定し，第 3 章で提案システムの詳細と人検知の手法を説明する．第 4 章では提案システムによる人検知可能性の検証実験について述べ，第 5 章では人検知判定基準の検証実験について述べる．第 6 章では関連研究を取り上げ，最後に，第 7 章で結論を述べる．

2. 想定環境

本章では、本研究の対象として想定する環境について述べる．はじめにスマートスピーカ概念と定義を述べ，実際に市販されているスマートスピーカに言及し，次にスマートスピーカが攻撃を受ける環境を想定した脅威モデルを定義する．

2.1 スマートスピーカ

2.1.1 スマートスピーカ概念と定義

スマートスピーカとは，音声ユーザインターフェイスとアシスタント機能を持つスピーカである．ここで音声ユーザインターフェイスとは，ユーザが声でコンピュータや端末をコントロールできるようにする仕組みを指し [12]，アシスタント機能とは，ユーザの要求を理解してそれに応える機能を指す．これらの機能は統合して音声アシスタント，もしくは AI アシスタントとも呼ばれる．音声アシスタントは，Apple Inc. が提供する Siri や Microsoft Corporation が提供する Cortana，Google LLC が提供する Google Assistant などのアプリケーションとしてスマートフォンやコンピュータに実装され，デバイスに話しかけることでユーザが要求を満足できるインターフェイスを実現している．スマートスピーカは，こうした音声アシスタント機能をスピーカに実装したものである．スマートスピーカはスマートフォンとは異なり，ユーザからの話しかけがマイクロフォンから離れた場所で行われることが多い．従って，スマートスピーカに搭載されている音声アシスタント機能は，マイクロフォンから離れた場所での発話に対する音声認識技術である遠隔音声処理 (FFVIP: Far-Field Voice Input Processing) に対応している [12]．

スマートスピーカを構成する主な要素は，無指向性のスピーカとマイクロフォンアレイである．スマートスピーカはウェイクワードと呼ばれる単語をユーザから話しかけられた場合に，それをトリガとしてアシスタント機能等サービスの提供を開始するため，マイクロフォンアレイは電源と接続されている限り常に周囲の音を受信する．マイクロフォンアレイがウェイクワードを含む音声を受信すると，ウェイクワードを含むユーザ音声はインターネットを通じてクラウドへ送ら

れる。クラウドに送られた音声は、クラウドが提供する自然言語理解のアルゴリズムによって意味内容を分析される。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各種サービスを提供する。

2.1.2 市販されているスマートスピーカの例

Google Home

Google Home[13] は、Google LLC が開発したスマートスピーカである。音声アシスタントとして Google Assistant を搭載しており、標準のウェイクワードとして”OK, Google”を利用可能である。水平方向中心全方位角に発音できるスピーカを搭載する通常モデルと、主として上向きに発音できるスピーカを搭載するモデル”mini”があり、共にマイクロフォンアレイは2chである。アプリケーションの利用や対応家電との連携により、インターネット上の情報検索、オンラインショッピング、音楽やラジオの再生、ニュースや天気予報の読み上げ、テレビの操作、エアコンや照明の操作などといった機能を音声ユーザインターフェイスから利用できる。遠隔音声処理のために独自に開発した大規模なデータセットを用いて訓練した DNN モデルを搭載しており、最大 8m × 10m の大きさを持つ部屋で音声アシスタントを利用可能である [14]。

Clova デバイス

Clova デバイス [15] は、LINE Corporation が開発したスマートスピーカである。音声アシスタントとして LINE Clova を搭載しており、標準のウェイクワードとして”Clova”や”Jessica”などを利用可能である。水平方向中心全方位角に発音できるスピーカを搭載し、マイクロフォンアレイが2chのモデル”Clova Friends””Clova Friends mini”や、4chのモデル”Clova WAVE”がある。スキルと呼ばれるアプリケーションの利用により、Google Homeと同様の機能に加え、LINE アプリケーションとの連携によって通話も可能である。

Amazon Echo

Amazon Echo[16] は, Amazon が開発したスマートスピーカである. 音声アシスタントとして Amazon Alexa を搭載しており, 標準のウェイクワードとして”Alexa”を利用可能である. 水平方向中心全方位角に発音できるスピーカを搭載し, マイクロフォンアレイは7ch である. ”Echo dot”や”Echo”といったモデルがあり, Google Home と同様の機能を利用可能である. これらの機能に加え, 温度センサを搭載したモデル”Echo Plus”は室温情報を提供できる. また, 7つのマイクロフォンを利用することで声の方向を特定することもできる.

2.2 脅威モデル

本研究では, スマートテレビやスマートフォンといった遠隔操作によって音声を発せられるスマートデバイスが宅内の一室に存在することで, デバイスの周囲に人が存在しないにも関わらず, デバイスが第三者の攻撃による悪意ある遠隔操作を受けて, 同じ部屋に存在するスマートスピーカに対するウェイクワードの話しかけができてしまうような状況を想定する. この時, 攻撃を受けたスマートデバイスやスマートスピーカが置かれた部屋には, スマートデバイスやスマートスピーカの他にも家具が任意に置かれており, 話しかけ音声の他にもエアコンの動作音などといった背景雑音があるものとする. 部屋には人がいてもいなくてもよく, 攻撃者による音声は部屋に人がいればその人によって容易に検知可能であるものとする. また, 攻撃者は攻撃対象のデバイスや, それに搭載されるソフトウェア, あるいはデバイスがバックエンドとして利用するクラウドサービスの開発者ではなく, 攻撃が音声を利用したスマートスピーカに対する話しかけ以外の形を取ることはないものとする. さらに, 攻撃者は本研究で提案する手法を知る者ではなく, 提案手法が用いるソナー音に対するジャミングやソナー音の偽装等によって人検知を妨害することはないものとする.

3. 提案システム

本章では，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．始めに提案システムの目的を述べ，次にシステムを構成する機器とスピーカから発信するソナー音について説明し，最後に受信音の分析に基づく人検知手法を説明する．図 1 に提案システムの概要図を示す．

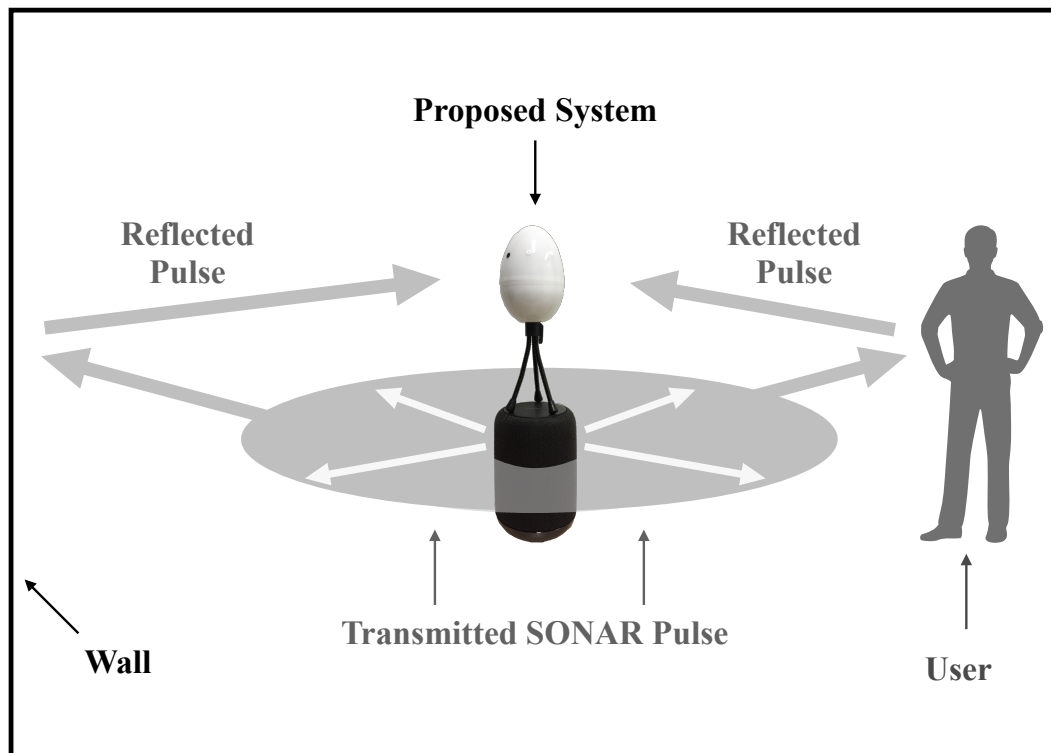


図 1 提案システムの概要

3.1 目的

第三者によるスマートスピーカへの攻撃に対応するためには，スマートスピーカが (1) 部屋内における人の有無と (2) 部屋内に存在する人の位置を検知する必要がある．

スマートスピーカが (1) を検知できると，部屋に人がいないにも関わらずウェイクワードの話しかけがなされるといった状況を不自然だと判断できるため，スマートスピーカは住人不在の場合における攻撃に対応できるようになる．さらにスマートスピーカが (2) を検知できると，住人が認識できないような音声を用いて住人によるものではないウェイクワードの話しかけがなされるといった状況を，音声の音源位置と住人の位置が異なることから不自然だと判断できるため，スマートスピーカは住人がいる場合においても DolphinAttack のような攻撃に対応できるようになる．

本研究では特に (1) に注目し，スマートスピーカが部屋内における人の有無を検知できるようなシステムを提案する．

3.2 システム構成

本研究では，無指向性スピーカの上にマルチチャ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ネルマイクロフォンアレイを組み合わせた構成とそれらの仕様は，第 2.1.2 項で述べた市販のスマートスピーカと同様のものである．市販のスマートスピーカがマイクロフォンアレイから得た音声データをそのまま用いることは難しかったため，このような構成とした．

3.3 構成機器

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．将来的に提案システムを 4.1 節における (2) の検知にも発展させて市販のスマートスピーカに組み込むことを鑑み，スピーカとマイクロフォンアレイは二次

元位置推定にも応用可能なものを利用した．その観点から考えると，スピーカは多くの市販のスマートスピーカと同じく水平方向を中心として周囲の全方位角方向に音を出す形態が望ましいため，円筒形で側面に発音構造を備えるソニー株式会社の LF-S50G を選択した．同様の観点から，マイクロフォンアレイは Amazon Echo のようにチャンネル数が多いものが望ましいが，LF-S50G に搭載されたマイクロフォンは 2 つしかない．そこで，マイクロフォンを 8 つ搭載した株式会社システムインフロンティアの TAMAGO-03 をマイクロフォンアレイとして用いた．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．



図 2 使用したスピーカ



図 3 使用したマイクロフォンアレイ

3.4 ソナー音の仕様

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 [17]．

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．この TAMAGO-03 の仕様によりソナー音の周波数は可聴域から選択したが，提案システムの実装環境に応じて OFDM のサブキャリアを適切に選択できるため，サンプリング周波数が高いマイクロフォンアレイを用いればソナー音の周波数を非可聴域から選ぶことができる．ソナー音一つあたりの時間長は 4 ms であり，そのサンプル数は $16 \text{ kHz} \times 4 \text{ ms} = 64$ 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．

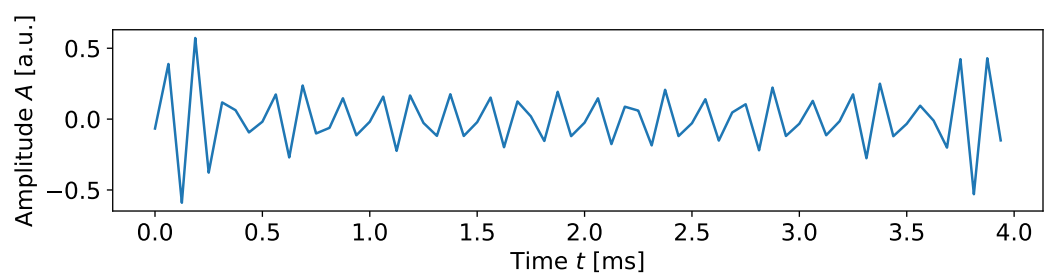


図 4 使用したソナー音

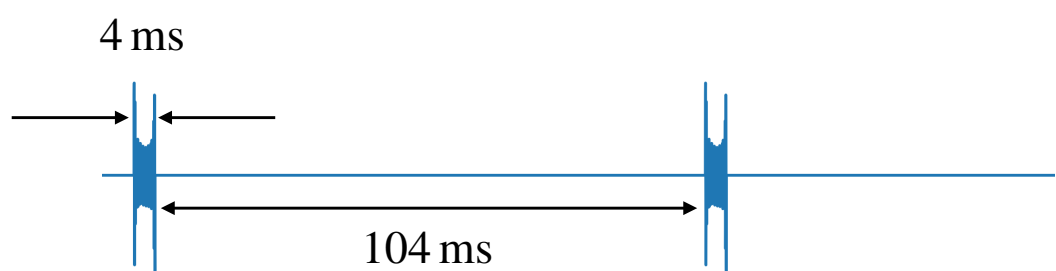


図 5 ソナー音の長さと発信間隔

3.5 人検知手法

本提案システムでは，スピーカから発信され部屋の環境で反射してマイクロフォンアレイに返ってきた反響音を分析することで人検知を行う．本項では，反響音の分析方法について説明する．

3.5.1 概要

人検知は次の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処理単位ごとへの切り分け
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．

3.5.2 人検知ステップ

本項では，人検知の各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

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

と表せる．ここで $N_s = 64$ 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．従って， $\phi(t)$ 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処理単位ごとへの切り分け

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に注目し，その周波数帯を中心として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音を落とす．さらに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，以降の相互相関は 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，その集合として取り扱う．この操作により，今後ソナー音受信時刻を用いることはなくなるため，室温変化等に由来するソナー音受信時刻の変化は最終的な人検知に影響を及ぼさない．1 処理単位の相互相関はソナー音の発信間隔に等しい長さをもつ時系列データであるが，隣接する処理単位間での影響を除くために各処理単位の末尾を省き，最終的な 1 処理単位あたりの長さは 1500 サンプル，即ち 93.75 ms としている．

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が受信した音は，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，人がいない場合について相互相関を 8 つのチャンネルそれぞれで全処理単位に渡って平均し，これを各チャンネルの相互相関のベースラインとして取り扱う．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のうち，いくつかの処理単位を平均する．それを二乗してから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．

4. 提案システムによる人検知可能性の検証

本章では，提案システムを用いた人検知が可能かどうかを検証するための実験について詳細を述べる．始めに実験目的と実験方法を説明し，次に実際の実験結果と考察を述べる．

4.1 実験目的と実験方法

4.1.1 実験目的

本実験の目的は，提案システムによる人検知の可能性を検証することである．本実験によって，提案システムにおける人検知手法がどのような条件下でどこまで適用可能なかを簡単に確かめる．従って，本実験における実験条件として，部屋にいる人数を変えた場合と，部屋にいる人と提案システムとの距離を変えた場合を設定し，そうした条件が人検知にどういった影響を及ぼし得るのかを検証する．

4.1.2 実験環境

本実験では，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，エアコンが動作する屋内の部屋を用いた．

図 6 に実験で用いた部屋の写真を示す．部屋の中心には提案システムを置いた．部屋の寸法は図 7 に示す通りである．図 7 において，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．



図 6 人検知可能性の検証実験に利用した部屋

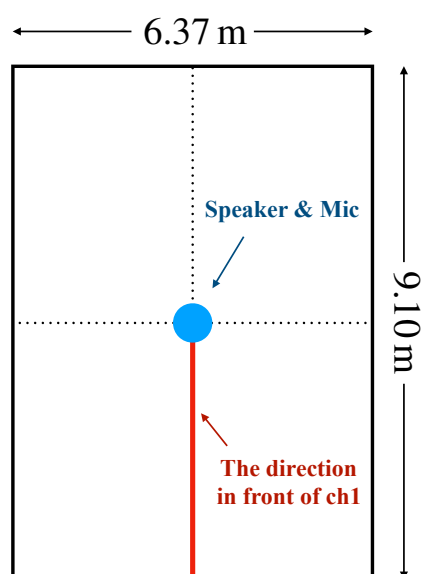


図 7 人検知可能性の検証実験に利用した部屋の寸法

4.1.3 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。人の配置パターンをそれぞれパターン 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 0.5m 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 0.5m 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 2.5m 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 2.5m 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．

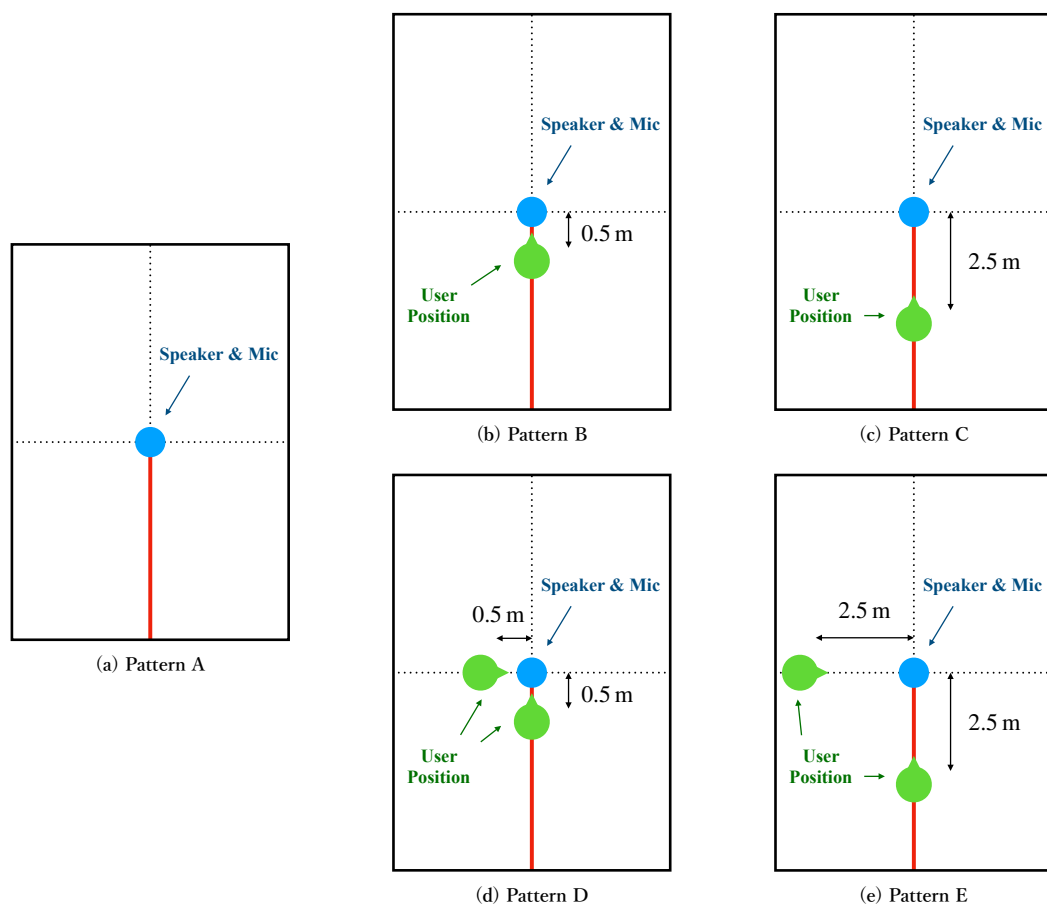


図 8 人検知可能性の検証実験パターン



図 9 パターン B の配置例



図 10 パターン C の配置例

4.2 実験結果

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。図において、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。図からわかるように、パターン A の P は τ が 2.0 s 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。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。

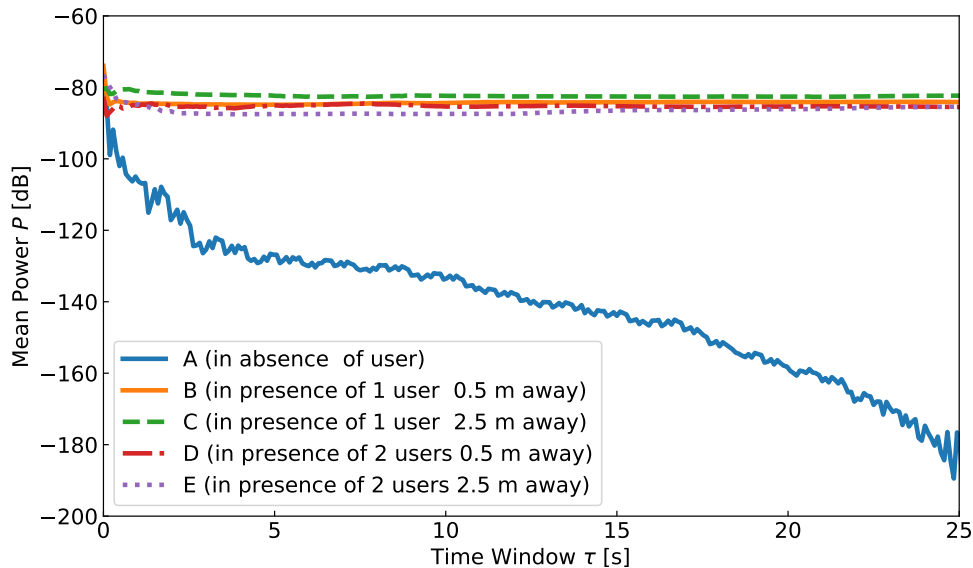


図 11 各パターンにおける平均パワーと受信音の時間長との関係

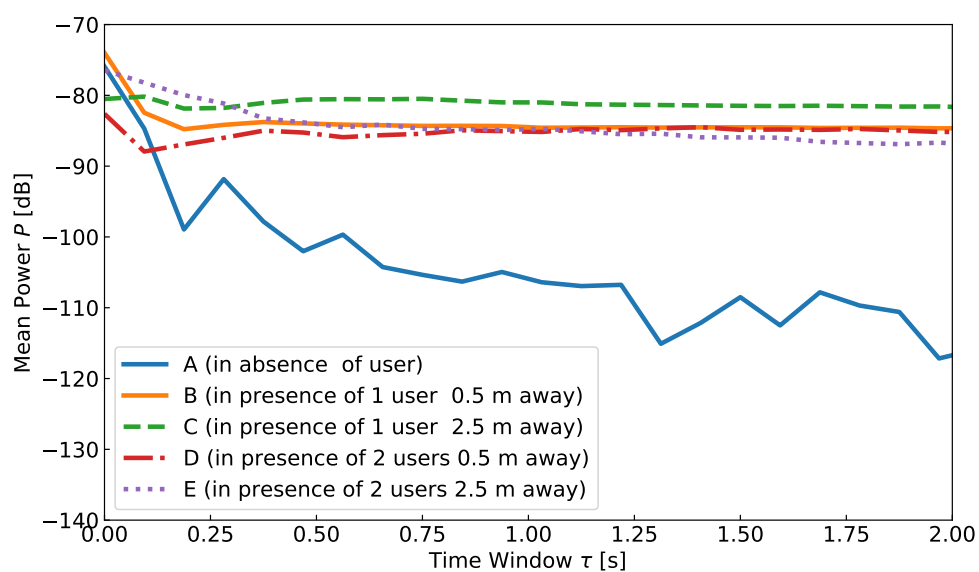


図 12 図 11 を領域 $0 \leq \tau \leq 2$ について拡大した図

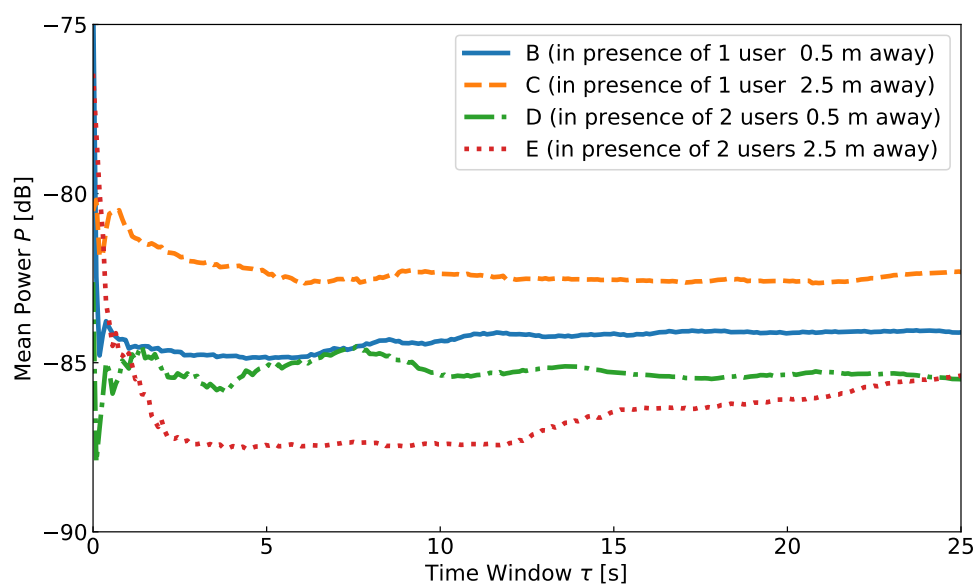


図 13 図 11 からパターン A の結果を除いた関係

4.3 考察

人がいる場合といない場合の P は τ が 2.0 s に満たない時点で十分に分離可能であり、 τ が提案システムの動作開始からの経過時間を表すことから、適切に平均パワーの閾値を設けることによって、提案システムの動作開始から 2.0 s 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、例として平均パワーの閾値を -110 dB に設定することで、提案システムの動作開始から 2.0 s 後には人を検知できることがわかる。ただし、本実験の実験パターンからは人がいる時の P と τ との関係にはっきりとした法則性を見出せないため、閾値の決め方に妥当性があるか否かの判断は難しい。平均パワーの閾値設定について、その妥当性は次章で議論する。

図 13 より、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行うことは難しいと考えられる。従って、提案システムは第 3.1 節で述べた目的の通り、部屋に人がいるかどうかを検出することに特化したものである。

5. 人検知判定基準の検証

本章では、提案システムを用いて人がいるかどうかを判定するにあたり、どのような基準を設けるべきかを検証するための実験について詳細を述べる。

5.1 実験目的と実験方法

5.1.1 実験目的

本実験の目的は、人検知を判定する基準となる平均パワー P の閾値 P_{thre} の決め方を検証することである。第4章の結果から、提案システムは平均パワー P の閾値 P_{thre} を適当に定めることによって人を検知できることがわかった。しかし、第4章の結果のみでは P と提案システムの動作開始からの経過時間 τ との関係が実験条件を変えた場合にどう変わるかを把握することは難しく、人がいるにも関わらず P が閾値 P_{thre} を下回ってしまうような実験条件が存在しないとは限らない。そこで、本実験では第4章とは異なる実験環境において、細かく実験条件を変えることによって、 P と τ との関係がどのように変わるのかを確かめる。その結果を用いて、提案システムを動作させたい環境において人検知を行うにあたり、適切に閾値 P_{thre} を設定する方法について検証する。

5.1.2 実験環境

本実験では、第4章の場合と同様に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、エアコンが動作する屋内の部屋を用いた。ただし、本実験で用いた部屋は第4章の場合と異なるものである。

図14に実験で用いた部屋の写真を示す。実際の宅内で用いられる状況に近づける目的で、部屋を中心ではない位置に提案システムを置いた。部屋の寸法は図15に示す通りである。図15において、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。



図 14 人検知判定基準の検証実験に利用した部屋

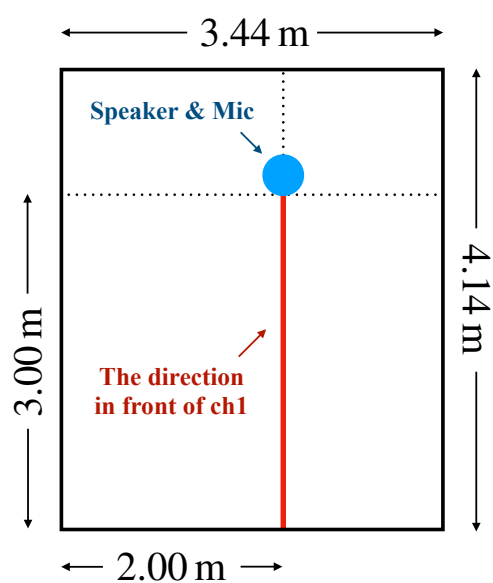


図 15 人検知判定基準の検証実験に利用した部屋の寸法

5.1.3 実験条件

本実験では、部屋にいる人と提案システムとの距離を変え、それぞれの距離において 30 秒間のソナー音発信による平均パワーの計測を行った。ここで部屋にいる人の数は 1 人とし、図 15 に実線で示したマイクロフォン素子 ch1 の正面に当たる一直線上にいるものとした。また、距離はマイクロフォン素子 ch1 からの直線距離 d とし、 d は 0 cm から 250 cm までの 10 cm ずつの値それぞれについて計測した。

5.2 実験結果

平均パワー P と人-システム間距離 d との関係を、 τ が 0.1 s, 1.5 s, 20 s の場合それぞれについて図 16, 図 17, 図 18 に示す。それぞれの図において、丸点はある時の $P-d$ 関係を、破線は丸点を一次関数でフィッティングした回帰直線を、実線は人がいない時の平均パワーを表す。

図 16 からわかるように、 $\tau = 0.1$ s における人がいない時の平均パワーは、人が $d \geq 70$ cm にいる時の平均パワーよりも大きいため、この時点で平均パワーに閾値を定めて人の有無により平均パワーの値を分離することは不可能である。一方で、図 17 からわかるように、 $\tau = 1.5$ s における人がいない時の平均パワーは、人がどの距離にいる場合の平均パワーよりも小さいため、閾値 P_{thre} を -105 dB から -110 dB の間に設定すれば $\tau = 1.5$ s の時点で迅速に人検知が可能であるように見える。しかし、同図の回帰直線が表す平均パワーは $d \geq 250$ cm において人がいない時の平均パワーに達することを示唆しており、 $\tau = 1.5$ s の時点で閾値 P_{thre} を前述のように設定してしまうと、提案システムから 250 cm よりも遠い位置に人がいる場合にシステムが住人は不在であると誤判定を起こしてしまうと考えられる。また、図 18 からわかるように、 $\tau = 20$ s においては回帰直線が表す平均パワーが人がいない時の平均パワーに達するまで d にある程度の大きさを要すると考えられるため、 d が大きい場合にも対応できるような閾値 P_{thre} を設定でき得るが、提案システムの動作開始から 20 秒待たなければ人検知ができないという欠点が残る。

そこで、或る τ における回帰直線が表す平均パワーが、或る閾値 P_{thre} に達するときの人-システム間距離 d を、その d 以内の距離にいる人を検知できる最大距離という意味で最大検知可能距離 d_{max} と呼ぶこととし、この d_{max} と τ との関係を異なる P_{thre} ごとに表したものを図 19 に示す。実線は P_{thre} を或る τ における人がいない時の平均パワー $P_{ab}(\tau)$ とした場合の $d_{max} - \tau$ 関係を表す。それ以外のグラフは、 P_{thre} を或る値に設定したときの $d_{max} - \tau$ 関係を表す。各グラフに対応する P_{thre} の値は凡例の通りである。閾値を或る値とした時に検知できる距離は、その値に対応する d_{max} 以下の d 、かつ実線で表した d_{max} 以下の d でなければならない。すなわち、図中の実線が表す d_{max} よりも値が小さい領域と、設定したい閾値に対応するグラフが表す d_{max} よりも値が小さい領域との積集合が、設定したい閾値で人検知できる距離と τ との関係を表す。

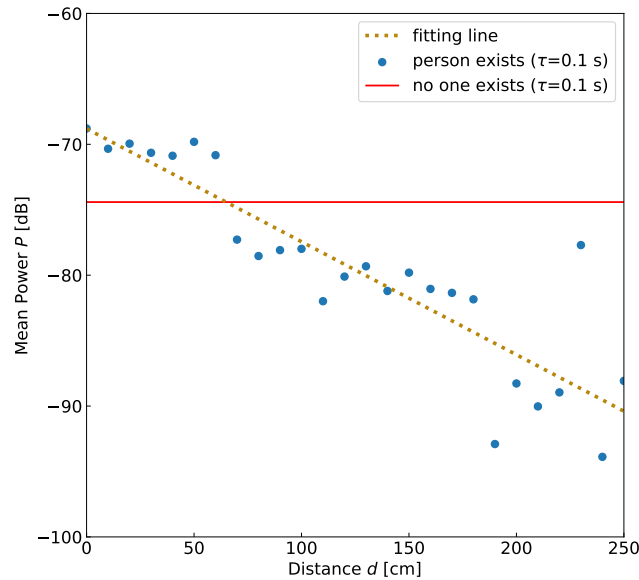


図 16 $\tau = 0.1 \text{ s}$ における平均パワーと人-システム間距離との関係

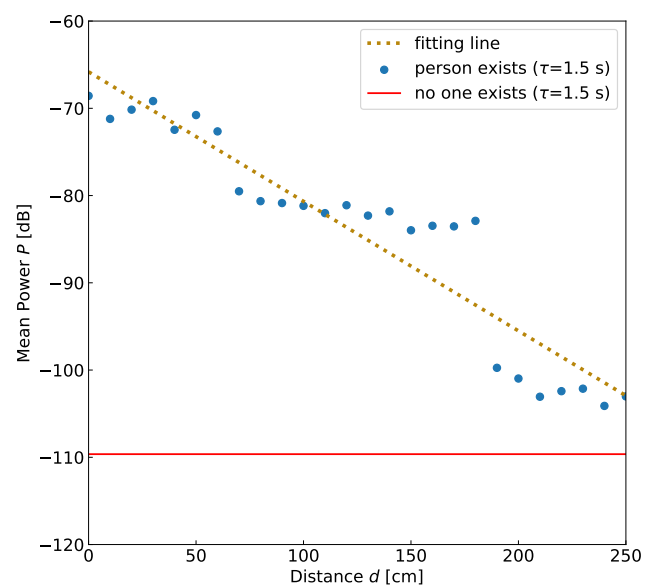


図 17 $\tau = 1.5$ s における平均パワーと人-システム間距離との関係

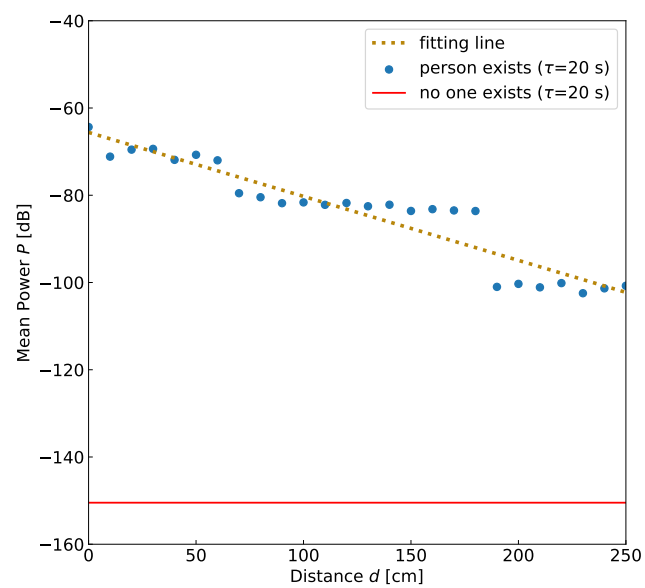


図 18 $\tau = 20$ s における平均パワーと人-システム間距離との関係

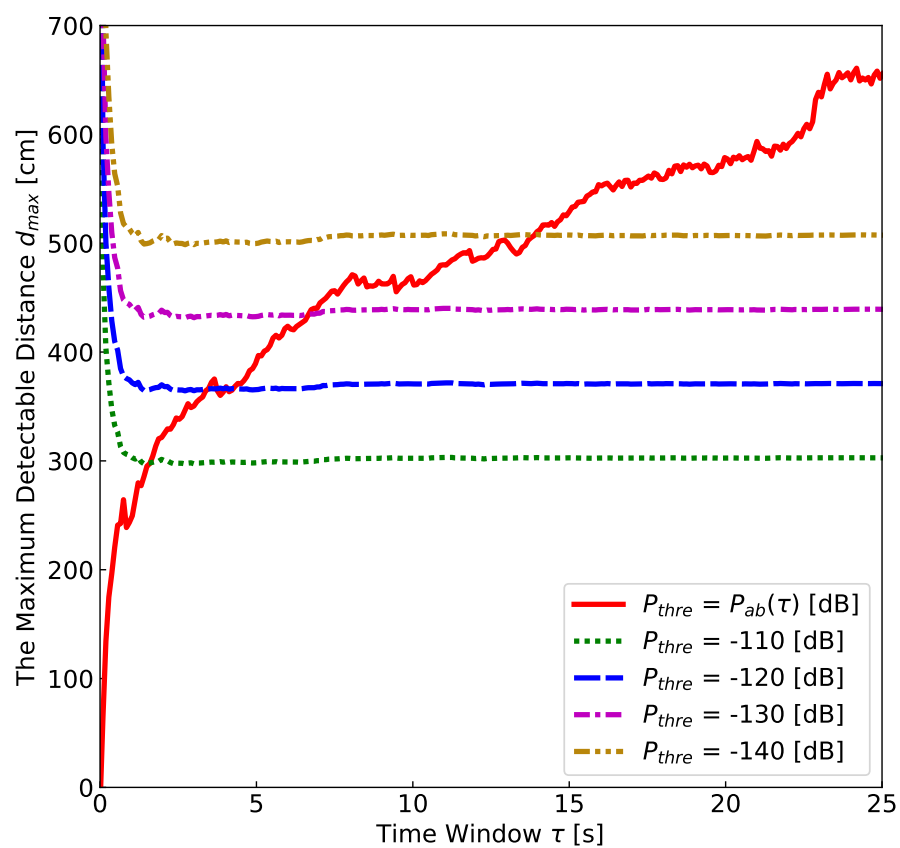


図 19 平均パワーの閾値に依る最大検知可能距離と受信音の時間長との関係

5.3 考察

図 19 において, $P_{thre} = -110$ dB とした場合を考える. この場合に対応する $d_{max} - \tau$ 関係は点線のグラフであり, 点線のグラフと実線のグラフが交わる時刻は 1.6 s である. 従って, $P_{thre} = -110$ dB とした場合の最大検知可能距離とシステム起動からの経過時間との関係は, $\tau < 1.6$ s のときは図における実線のグラフに相当し, $\tau \leq 1.6$ s のときは図における点線のグラフに相当する. 図から明らかなように, $P_{thre} = -110$ dB とした場合の最大検知可能距離 d_{max} は $\tau < 1.6$ s においてほぼ上昇し続け, $\tau \leq 1.6$ s においてほとんど不変となり, その値は概ね 300 cm である. よって, 閾値を -110 dB とした場合, 提案システムは音声の呼びかけを受けてから 1.6 秒後に概ね 300 cm まで離れた距離にいる人の存在を検知できるようになることがわかる. 同様にして, 閾値を -120 dB とすれば 3.5 秒後に概ね 370 cm まで, 閾値を -130 dB とすれば 6.7 秒後に概ね 440 cm まで, 閾値を -140 dB とすれば 14 秒後に概ね 510 cm までの範囲で人検知が可能となることがわかる.

実際の宅内において提案システムを利用することを考えると, 提案システムは利用される部屋の全範囲で人検知できる必要がある. 政府機関の調査によれば, 日本における 1 人当たり居住室の畳数は 13.54 畳である [18]. ここで, 居住室とは, 居間, 茶の間, 寝室, 客間, 書斎, 応接間, 仏間, 食事室など居住用の室をいう [19]. 日本の平均世帯人員は 2.31 人であり [20], 誘導居住面積水準では 2 人暮らしにおける世帯人員一人当たりの延べ床面積を約 13 畳と定めているため [19], 13.54 畳は日本において豊かな住生活を実現しうる一般的な居住室の広さであると言える. 13.54 畳は 20.97 m^2 に相当し, 正方形で換算すると一辺の長さは 4.58 m となる. 従って部屋の中央に提案システムを設置することを考えると, 提案システムは少なくとも自身から半径 230 cm 以内の範囲において人検知ができればよく, 閾値は 300 cm までの人検知が可能な値 -110 dB に設定すれば十分であると言える. このことから, 提案システムは日本の一般的な住宅において, 音声の呼びかけを受けてから 1.6 秒後には十分に人検知を行えることがわかる. この結果に対し, 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 [21] の手法では, 音声の呼びかけを受けてから十分な人検知までに 16

秒かかっていた。従って、本研究で提案したシステムは従来手法よりも速やかに人を検知できると考えられ、提案システムは迅速な人検知に有効であると言える。また、本実験で利用した部屋は第4章とは異なるものであったが、前述の通り十分な人検知を行うことができた。このことから、提案システムは運用する環境に依存することなく人検知が行える頑強なシステムであると言える。

6. 関連研究

本章では、人検知の関連研究について述べる。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。そのため、はじめに屋内位置推定を目的とした人検知手法に関する研究について、電波を利用したものと音波を利用したものに分けて述べる。その後、本研究と類似した環境を脅威モデルとして設定した研究を紹介する。

6.1 電波を用いた人検知

Bocca ら [22] 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 1 m 以内の精度で検知できている。Adib ら [23] も電波を用い、縦横の大きさ 1 m × 2 m 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 1 m 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。

6.2 音波を用いた人検知

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 [17] 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 つのマイクロフォン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、マイクロフォンを 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。

6.3 スマートスピーカを用いた人検知

人の検知に焦点を当てた研究としては、Alanwar ら [21] 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い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ロフォンを模した 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。

7. 結論

本研究では、機械からの偽音声によるスマートスピーカへの攻撃から防御するため、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システムを提案した。提案システムでは、スピーカから発信したソナー音をマルチチャネルマイクロフォンアレイで受信し、そこから計算したソナー音との相互相関関数によって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2秒以内に人を検知できることがわかった。

提案システム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。また、提案システムが複数の環境下で人検知できたことから、提案システムは運用環境に依存しない頑強なシステムであることもわかった。

第3.1節で言及したように、本研究はスマートスピーカが第三者による攻撃から自衛するために必要だと考えられるアプローチの中でも、特に人の有無を検知することに着目して行ったものである。人の有無を検知することの応用先はセキュリティに限定されるものではなく、他にも様々な利用法が考えられる。その例として、周囲のユーザ検知をトリガとしてスマートスピーカからユーザに対する能動的な呼びかけを行う機能が挙げられる。現在市販されているスマートスピーカは利用のためにユーザから適切な命令を与える必要があり、スマートスピーカが実行可能な機能を知らないユーザや、適切な命令文を考えることが難しいユーザにとっては利用しづらいものとなっている。そこでスマートスピーカからユーザに話しかけることによって、スマートスピーカ側から利用できる機能を提示したり、命令文の選択肢を簡潔にしたりと、ユーザがスマートスピーカに対し利便性を感じやすい設計を創出できるようになると考えられる。

また、本研究で利用したマイクロフォンアレイは人の位置を検知するアプローチにも発展させることが可能である。本研究を発展させて周囲の人の有無に加え人の位置も検知することができれば、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と同時に，セキュリティ面以外にもスマートスピーカからユーザに向けて適切な方向に話しかけることができるようになるとも考えられる．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も把握できるようにするためには，屋内位置推定の手法を用い，マイクロフォンアレイの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すると行ったアプローチが必要となる．そうしたアプローチを適用するためにも，距離分解能に優れる FMCW レーダの手法に倣い，ソナー音をチャープ信号とした場合についても今後評価すべきだと考えられる．

謝辞

本研究を進めるにあたり，奈良先端科学技術大学院大学 情報科学研究科 安本 慶一 教授には，研究の全般にわたって多大なるご指導とご助言を賜りました．また，充実した研究環境の整備など，研究活動を手厚くご支援いただきました．これまで格別のご厚誼にあずかり，謝意を表すと共に，心より厚く御礼申し上げます．

奈良先端科学技術大学院大学 情報科学研究科 林 優一 教授には，ご多忙の中にも関わらず，論文審査員を引き受けていただき，謝意を表すと共に，心より厚く御礼申し上げます．

奈良先端科学技術大学院大学 情報科学研究科 荒川 豊 准教授には，多角的な視点から様々な知見をご指導いただきました．情報科学分野における研究の進め方を考えるにあたって，ご教授いただいた概念や考え方はその指針となりました．これまでのご芳情に謝意を表すと共に，心より厚く御礼申し上げます．

奈良先端科学技術大学院大学 情報科学研究科 藤本 まなと 助教には，電波工学の専門的な知見に関してご助言を賜りました．音波を用いた本研究を進めるために不可欠であった電波の知識について，アンテナ工学や高周波工学などの様々な見地からご教授いただきました．心より感謝申し上げます．

奈良先端科学技術大学院大学 情報科学研究科 諏訪 博彦 助教には，データ分析に関して専門的な見地からご指導を賜り，研究について本質的なご助言も多くいただきました．また，各種催事に際しても専門的な見地から多くご助言を賜りました．心より感謝申し上げます．

金岡恵事務補佐員，山内奈緒事務補佐員，尾川恵理事務補佐員，南あずさ事務補佐員には，学会や出張に関する事務処理を始めとし，様々な面で研究生活を支えていただきましたこと，謹んで感謝申し上げます．

最後に，ユビキタスコンピューティングシステム研究室の学生の皆様には，研究に関して多くのご助言をいただいたほか，公私ともにお世話になりました．心より感謝申し上げます．

参考文献

- [1] Canalys. Global smart speaker shipments grew 187% year on year in q2 2018, with china the fastest-growing market. https://www.canalys.com/static/press_release/2018/Press-release-160818-global-smart-speaker-shipments-grew.pdf. accessed: 11 Jan. 2019.
- [2] 総務省. 情報通信白書〈平成 30 年版〉人口減少時代の ICT による持続的成長. 日経印刷株式会社, 2018.
- [3] Andrew Liptak. Amazon’s alexa started ordering people dollhouses after hearing its name on tv. <https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse>. accessed: 11 Jan. 2019.
- [4] Chaouki Kasmi and Jose Lopes Esteves. Iemi threats for information security: Remote command injection on modern smartphones. *IEEE Transactions on Electromagnetic Compatibility*, Vol. 57, No. 6, pp. 1752–1755, IEEE, 2015.
- [5] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. Cocaine noodles: exploiting the gap between human and machine speech recognition. *WOOT*, Vol. 15, pp. 10–11, 2015.
- [6] Yeongjin Jang, Chengyu Song, Simon P Chung, Tielei Wang, and Wenke Lee. A11y attacks: Exploiting accessibility in operating systems. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 103–115. ACM, 2014.
- [7] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. Hidden voice commands. In *USENIX Security Symposium*, pp. 513–530, 2016.

- [8] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. Inaudible voice commands: The long-range attack and defense. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*, pp. 547–560. USENIX Association, 2018.
- [9] 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*, pp. 2–14, 2017.
- [10] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyan Xu. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 103–117, 2017.
- [11] 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*, pp. 63–74, 2014.
- [12] Amazon. 音声ユーザーインターフェース (vui) とは何か. <https://developer.amazon.com/ja/alexa-skills-kit/vui>. accessed: 17 Jan. 2019.
- [13] Google. Google home - スマート スピーカー&ホーム アシスタント - google ストア. https://store.google.com/jp/product/google_home. accessed: 20 Jan. 2019.
- [14] Chanwoo Kim, Ananya Misra, Kean Chin, Thad Hughes, Arun Narayanan, Tara Sainath, and Michiel Bacchiani. Generation of large-scale simulated utterances in virtual rooms to train deep-neural networks for far-field speech recognition in google home. *Proc. INTERSPEECH. ISCA*, 2017.

- [15] LINE. Line clova 公式サイト. <https://clova.line.me>. accessed: 20 Jan. 2019.
- [16] Amazon. Amazon.co.jp: Echo & alexa: Amazon デバイス・アクセサリ. <https://www.amazon.co.jp/b/?node=5364343051>. accessed: 20 Jan. 2019.
- [17] Rajalakshmi Nandakumar, Alex Takakuwa, Tadayoshi Kohno, and Shyam-nath Gollakota. Covertband: Activity information leakage using music. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 1, No. 3, p. 87, 2017.
- [18] 総務省統計局. 統計局ホームページ/日本の住宅・土地—平成 25 年住宅・土地統計調査の解説—結果の解説. <https://www.stat.go.jp/data/jyutaku/2013/pdf/nihon00.pdf>. accessed: 26 Jan. 2019.
- [19] 総務省統計局. 統計局ホームページ/平成 25 年住宅・土地統計調査 用語の解説. https://www.stat.go.jp/data/jyutaku/2013/pdf/giy14_0.pdf. accessed: 26 Jan. 2019.
- [20] 総務省統計局. 統計局ホームページ/地方別、世帯人員別の世帯数分布表 (100 万百分比) 平成 30 年 (2018 年)1 月分 地方, 世帯人員別の世帯数分布 (100 万百分比) . https://www.stat.go.jp/data/kakei/setai_bunpu.html. accessed: 26 Jan. 2019.
- [21] 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*, pp. 38–43, 2017.
- [22] Maurizio Bocca, Ossi Kaltiokallio, Neal Patwari, and Suresh Venkatasubramanian. Multiple target tracking with rf sensor networks. *IEEE Transactions on Mobile Computing*, Vol. 13, No. 8, pp. 1787–1800, 2014.

- [23] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-person localization via rf body reflections. In *NSDI*, pp. 279–292, 2015.