

ユーザ行動に応じた情報を提供する AIoT型スマートスピーカの提案

佐々木 一郎¹ 藤田 茂²

1. はじめに

センサ, 家電をはじめとした多種多様なモノを, インターネットを介し通信を行うことで, 端末間での情報交換や制御を行う IoT (Internet of Things) に基づくサービスが注目されている. 主なサービスとして, 環境センシング, 遠隔監視, 遠隔制御などといったものが挙げられ, 様々な分野において活用されている.

スマートスピーカはその1つであり, ユーザの音声入力を基に, 対話型音声アシスタントから様々なサービスを提供することができる. 市販されているスマートスピーカの代表例として, GoogleHome[1] や AmazonEcho[2], Clova[3] が挙げられる.

アシスタントは, 設定された命令のみを実行するものもあれば, 自然言語処理を用いたスマートスピーカ独自の AI アシスタントなど, 種類は多岐にわたる. クラウド上に保存されている音楽の再生や, チャットアプリを通じた他者とのチャット・通話, 家電の操作などの様々なサービスが存在する. これらのサービスは, スマートスピーカ独自のものではなく, 他の IoT デバイスと連携して提供されることが多い.

しかし, スマートスピーカには以下のような欠点がある

- サービス提供のトリガーとして音声入力が必要なため, 他者との会話中に誤反応することがある
- 求めるサービスに応じて, ユーザが指示をしなければならない

これらの課題を解決するために, エージェント型 IoT (AIoT) システムを用いたスマートスピーカの提案を行う. AIoT システムとは, IoT デバイスのインテリジェント化を行うことにより, デバイスが相互に協調・連携を行うことで, デバイスが得られる情報に応じて自律的なサービス提供を行う機構である.

本稿では, ユーザの行動と身の回りの環境情報を取得することで, ユーザの状況に応じたサービスを提供するスマートスピーカの提案を行う.

2. 関連研究

MMDAgent を利用した家電の制御に関する研究がある [4]. MMDAgent とは, 3D モデルに話しかけるような対話が可能, 音声インタラクションシステム構築ツールである. この研究では, Android 端末を用いて MMDAgent と音声対話を行うことにより, 家電制御を行っている. MMDAgent と別にサーバを設け, このサーバを介して MMDAgent は実行可能な家電操作命令の一覧を取得する. その一覧を基に, ユーザの与えた家電制御命令と, 端末から取得できるユーザの状況を用いて, 実行可能な家電操作命令を動的に変更することができる. この研究では, MMDAgent の対話シナリオが記述される FST ファイルの動的生成を主としている. 本稿では, ユーザの命令を基にデバイスに搭載されるエージェントがセンサの操作を行うため, エージェントが直接動的に命令を実行する.

異種 IoT のデバイス協調制御手法 [5][6][7] では, 知的に行動するエージェントが各 IoT デバイスを管理・制御する. この研究における機構では, エージェントは動作制御の基本となるプラン知識に基づき, 行動ポリシーを切り替えることで, 自律的な振り舞いを実現する. これと合わせて, それぞれのデバイスが持つデバイスステータスに基づき, エージェント同士が用途に応じて連携動作を行うために, 各デバイスは内部状態と外部環境に応じて協調動作を行う. この研究では, 各デバイスエージェントが自律的に協調することによってシステムを実装している. 本稿はこの AIoT 機構に倣ってスマートスピーカを再現しているが, ユーザからの入力を基にエージェントが行動を行う.

3. 提案手法

市販されているスマートスピーカは, スマートスピーカとしての単体のデバイスと, アシスタントの役割を担うク

¹ 千葉工業大学 情報科学研究科 情報科学専攻, Graduate School of Computer Science, Chiba Institute of Technology JAPAN

² 千葉工業大学 情報科学部 情報工学科, Department of Computer Science, Chiba Institute of Technology JAPAN

クラウドサービスによって構成されるものが多い。本稿で提案するスマートスピーカはそれらとは異なり，AIoT[6]の考えに基づいて，RaspberryPiを用いた複数のデバイスから構成されるものとする。各デバイスにはセンサとエージェントを搭載する，エージェントはセンサに対応したものとなっている。表1にセンサと対応するエージェントを搭載する。

表1 デバイス，センサの仕様と対応するエージェント

デバイス	
使用機器	Raspberry Pi3
OS	Raspbian
CPU	ARM Cortex-A53 1.2GHz
メモリ	1GB
サーバデバイス	
使用機器	MacBook Pro
OS	macOS High Sierra
CPU	3.1GHz intel Core i5
メモリ	16GB
画像取得	
使用機器	PiCamera
エージェント	CDA (Camera Device Agent)
音声入出力	
使用機器	GoogleAIY VoiceKit
エージェント	MDA (Mic Device Agent)
サーバ管理	
エージェント	SA (Server Agent)

AIoT フレームワークに基づいたスマートスピーカの設計を図1に示す。

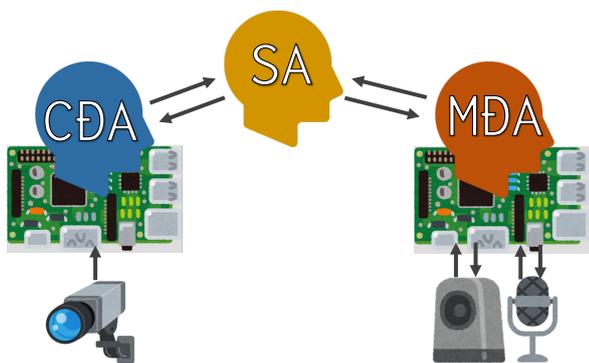


図1 システム概要図

3.1 サービス提供の仕組み

スマートスピーカによるサービス提供の仕組みについて説明する。サービスの提供は，ユーザの発話内容の中に含まれるサービスと紐づけられたキーワードによって行われる。サービス提供には各種 API を用いる。本稿で扱うサービスは以下のものとする。

サービス提供時の各デバイスの行動について説明する。

表2 サービス一覧

キーワード	サービス内容
天気	周辺地域の天気予報の提供
電車	最寄駅の次の電車の時刻提供
撮影	CDA により写真撮影

- SA (Server Agent)

SA は各エージェントのメッセージの受信を行い，受信されたメッセージを他の全てのエージェントへ送信する。SA を介することで，全てのエージェントは，他のエージェントが送信したメッセージを取得できる。また，エージェントから送信されたメッセージのタイムスタンプやユーザ識別タグを認識し，これらをログファイルに保存する。

- CDA (Camera Device Agent)

CDA はカメラから得た画像を基に，人物顔検出を行う。顔が検出されている間は，ユーザが在室しているものとして扱う。人物顔検出は OpenCV の Haar-like 特徴分類器を用いる。CDA はユーザ識別を行わない。現在検出されている顔の数とタイムスタンプを SA へ送信する。

MDA に撮影サービスを提供するよう要求された場合，CDA は顔検出モードから撮影モードへ移行する。撮影モードでは，MDA の合図に合わせて撮影を行う。処理の流れを Algorithm1 に示す。

Algorithm 1 Camera Device Agent

```

カメラ設定の初期化
while True do
  人物顔検出の実行
  if MDA による撮影サービスの呼び出し then
    人物顔検出の一時停止
    5秒待機
    撮影
    画像の保存
  end if
end while

```

- MDA (Mic Device Agent)

MDA は発話をトリガーとしたスマートスピーカとして振る舞う。MDA は発話により起動し，CDA から得られるユーザの在室状況を基に，サービスの提供を行う。入力された音声の解析を docomoAPI，音声出力を AquesTalk を用いて行う。MDA によって取得されたユーザの発話は SA に送信される。

MDA は，発話内容からキーワードを確認し，対応したサービスを提供するメソッドを呼び出す。メソッドは引数に発話内容を持ち，それぞれのメソッドで字句解析を行い，ユーザからの細かい指示にも答えられるよう，オプションの設定を行う。処理の流れを Algorithm2 に示す。

Algorithm 2 Mic Device Agent

```
オーディオ設定
message の初期化
while True do
  マイクによる音声録音
  message ← API を用いた音声解析
  if message がトリガーワードを含む then
    マイクによる音声録音
    message ← API を用いた音声解析
    if message が "天気" を含む then
      現在の天気予報の提供
    end if
    if message が "電車" を含む then
      最寄駅の電車時刻の提供
    end if
    if message が "撮影" を含む then
      CDA によりカメラの撮影
    end if
  end if
end while
```

3.2 ユーザのモデル化

MDA は、図 2、CDA による人物顔検出を行なった上でサービスの提供を行う。

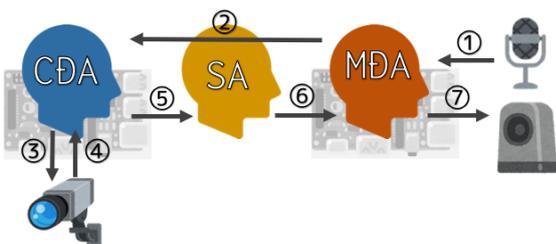


図 2 スマートスピーカの動作

- 1 MDA がマイクから音声入力を検知する
- 2 SA を介して CDA へ発話内容が共有される
- 3 CDA はカメラから現在の画像を取得する
- 4 取得した画像を基に、CDA は人物顔検出を行う
- 5 検出結果が SA を介して MDA へ共有される
- 6 MDA がサービスの提供を始める

CDA は、人物顔検出によって検出された顔画像を保存する。SA は MDA、CDA と共有されるユーザ情報の取得時刻を用いて、ユーザがもっともスマートスピーカの使用を行わない時間を求める。(以下、待機時間) 待機時間では、CDA は保存した顔写真を基に、ユーザの識別・学習を行う。CDA はこれらの学習結果から SA、MDA と情報を共有する際に、顔画像にユーザタグを付与する。これにより、スマートスピーカを利用するユーザを追跡できるようにする。

ユーザタグがついた情報は、SA によって待機時間に処理される。SA はユーザからの命令内容と、命令された時間帯を基に命令を評価し、時刻によってユーザの求める情報

を推測する。これを基に、推測された時刻にユーザが求めるであろう情報を先に提供する。その際、スマートスピーカはユーザからのフィードバックを求め、命令の再評価を行う。

4. 評価実験

第 3 章で提案したスマートスピーカを実装し、各エージェントが正しく動作しているかを評価する。

4.1 実験環境

実験では表 1 にあるとおり、AIoT システムによる 3 つのエージェントを用いて、スマートスピーカを実現する。提供するサービスは表 2 の通りである。複数ユーザによる仕様に実験を行い、ユーザ行動に応じた情報を提供できるかを確認する。

4.2 シナリオ

- 1 サービス提供の確認
- 2 複数ユーザでの動作確認
- 3 ユーザが命令をする前でのサービス提供

5. おわりに

本稿では、複数の RaspberryPi を用いて AIoT システムを実装し、スマートスピーカとしてユーザにサービス提供を行うシステムを提案した。今後はシステムの開発を行い、スマートスピーカのもつ課題を解決するために実験を進めていく。

参考文献

- [1] “GoogleHome.” https://store.google.com/product/google_home. (accessed 2018/5/12).
- [2] “AmazonEcho.” <https://www.amazon.com/b/?ie=UTF8&node=9818047011>. (accessed 2018/5/12).
- [3] “Clova.” <https://clova.line.me>. (accessed 2018/5/12).
- [4] 岩田剛士, 金鎔煥, 山本大介, 片山喜章, “MMDAgent を利用した家電制御インタフェースのための FST ファイルの動的生成とそれによる制御対象家電の動的制御について,” Tech. Rep. 51, feb 2018.
- [5] 横山真悟, 加藤匠, 高橋秀幸, 木下哲男, “エージェント型 IoT システムの自律適応制御手法,” 第 78 回全国大会講演論文集, vol. 2016, pp. 465–466, mar 2016.
- [6] 横山真悟, 加藤匠, 高橋秀幸, 木下哲男, “マルチエージェントに基づく異種 IoT デバイスの協同制御手法,” 第 79 回全国大会講演論文集, vol. 2017, pp. 213–214, mar 2017.
- [7] 片山健太, 横山真悟, 加藤匠, 高橋秀幸, 横田信英, 杉安和也, 木下哲男, “避難行動支援向けエージェント型 IoT システムの検討,” 第 79 回全国大会講演論文集, vol. 2017, pp. 593–594, mar 2017.