

## 科学研究費助成事業 研究成果報告書

令和 3 年 6 月 10 日現在

機関番号：12102

研究種目：研究活動スタート支援

研究期間：2018～2020

課題番号：18H06473・19K21541

研究課題名(和文)非言語的情報を活用したスマートアシスタントシステムの開発

研究課題名(英文)Development of Smart Assistant System Using Nonverbal Information

研究代表者

川口 一画(kawaguchi, ikkaku)

筑波大学・システム情報系・助教

研究者番号：80826975

交付決定額(研究期間全体):(直接経費) 2,300,000円

研究成果の概要(和文):本研究では、音声のみを用いる既存のスマートスピーカにおいて発生するウェイクアップワードの制約を排除し、利用者が必要とする適切なタイミングでコマンド認識を開始する手法の実現を目的とした。そのための方策として、人間同士のコミュニケーションにおいて対話の開始に用いられる「注視」に着目し、それを利用可能なスマートスピーカの開発を行った。開発したシステムを用いた評価実験の結果より、注視の入出力を用いることでシステムの操作性や印象が向上することが示された。

研究成果の学術的意義や社会的意義

本研究では、人間が普段対話の開始に用いている注視の入出力をシステムに取り入れることで、スマートスピーカの操作性や印象が向上することを示した。この際、システムがユーザの注視を検出するだけでは効果が低く、ユーザの注視に対してシステム側からも注視を返し相互注視が成立する場合に最も効果が高くなるという知見が得られた。この知見は、本研究の対象としたスマートスピーカだけでなく、近年増加している注視や顔検出を用いるインタフェースの操作性や印象の向上につながるものである。

研究成果の概要(英文): In this study, we aimed to develop a method that eliminates the wake-up word constraint that occurs in existing smart speakers, and starts command recognition at the appropriate timing required by the user. To achieve this, we focused on "gaze", which is used to initiate conversation in human communication, and developed the smart speaker system that can use gaze. The results of evaluation experiments using the developed system showed that the usability and impression of the system were improved by using the input and output of gaze.

研究分野：ヒューマンコンピュータインタラクション

キーワード：スマートアシスタント 音声インタフェース 非言語的情報 コミュニケーションロボット

### 1. 研究開始当初の背景

近年普及が進む Google home 等のスマートアシスタントは、音声を用いてハンズフリーで情報検索等を実行できるという特性を活かし、利用者の日常生活をさりげなくサポートすることを目的としている。ただし、現在市販されているスマートアシスタントの多くは、コマンドの開始をウェイクアップワード（「OK Google」等）で明示的に示さなければならないという制約があった。これに対して、より自然な対話型インタフェースの実現を目指し、周囲の音声を常時クラウドに送信して機械学習により文脈を理解し、ウェイクアップワード無しに必要なコマンドを実行する手法が検討されている。ただし、音声を常時クラウドに送信し続けることによるプライバシーの課題や、複数人で会話をしている場合に適切なタイミングで必要な処理を判断・実行することは依然困難であるという課題が示されていた。

本提案ではこれらの課題に対し、利用者とのインタラクションにおいて、音声だけでなく視線や身振り等の非言語的情報の入出力を活用する手法を提案する（図 1）。非言語的情報は、人間同士の会話において発話意図の提示や話者交替に関わり、システムがコマンド開始を予期する上でも有効な手がかりになると考えられる。非言語的情報を活用しクラウドへ音声を常時送信することなくコマンドの開始を予期することが出来れば、プライバシーの課題改善につながる。また、より多くの手がかりを活用することで複数人会話を含むより複雑な状況への対応が可能になる。



図 1 提案手法の概要

### 2. 研究の目的

本研究では、音声だけでなく視線等の非言語的情報の入出力を活用することにより、音声をクラウドへ常時送信することなく、複数人会話を含むより複雑な状況でもコマンドの開始を予期可能なスマートアシスタントを実現することを目的とする。

### 3. 研究の方法

#### 3. 1. 提案システムの実装

本研究では、利用者のコマンド開始を予期するための非言語的情報として、「注視」に着目した。社会学的知見によれば、注視は人間同士のコミュニケーションにおいて対話の開始や進行に関わるとされ、特に「相互注視」が対話開始の合図となることが示されている。そこで本研究においては、注視の入出力を利用可能なスマートスピーカの開発を行った。実装したシステムの外観を図 2 に示す。

提案システムは、Linux ベースの OS である OpenWrt で制御される制御基板 (Seeed 社 ReSpeaker Core)、顔・視線の検出と角度推定等をデバイス単体で実行可能なカメラモジュール (OMRON 社 HVC-P2)、Mic Array、スピーカ、およびモータと LED により視線を提示するロボット頭部より構成した。スマートスピーカの核となる音声アシスタント機能の実装には Amazon 社の Alexa Voice Service (AVS) を用いた。

実装したシステムを用いて、単独の利用者を対象とした評価実験、複数名の利用者を対象とした評価実験を実施した（詳細は 4. 1 節、4. 2 節参照）。また、研究を進める中で、提案システムを応用したスマートスピーカの新たなインタラクション手法として、注視の入出力機能を活用した音声通話機能の着想を得た。そのため、当該機能を実現するためのシステム実装を行い、評価を実施した（詳細は 4. 3 節参照）。



図 2 提案システムの外観

#### 4. 研究成果

##### 4. 1. 単独の利用者を対象とした評価実験

提案システムの効果を検証するため、1名の実験参加者を対象とした評価実験を実施した。実験では、着席状態でタスクリストに沿ってシステムとインタラクションを行う実験タスクを設定した。評価にあたっては、注視を用いないウェイクワード条件、注視の入力のみを用いる注視検出条件、および注視の入出力を用いる相互注視条件の比較を行った。

実験結果より、相互注視条件においてウェイクワード条件と注視検出条件よりもシステムの操作性および印象が有意に向上することが示された（図3）。この結果より、注視を用いるシステムを設計する上では、システムが利用者の注視を検出して利用するだけでなく、利用者の注視に対してシステム側からも注視を返し相互注視を成立させることが有効であるという知見が得られた。実験結果に基づき執筆した論文は、国内の査読付き論文誌（ヒューマンインタフェース学会論文誌）へ採録され、さらに第20回ヒューマンインタフェース学会論文賞を受賞した。

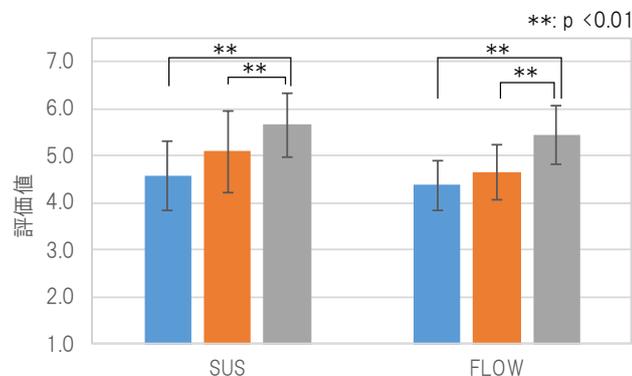


図 3 各条件に対するアンケートの結果. SUS はシステムの操作性, FLOW はタスクへの興味関心・集中度の評価値を示す.

##### 4. 2. 複数数の利用者を対象とした評価実験

次に、提案システムを複数人で利用する場合の効果を検証するため、2名の実験参加者を対象とした評価実験を実施した。なお、当該実験はストックホルム大学との共同研究として実施した。実験の準備に当たって、ハードウェアの実装・調整・およびインタラクション設計を申請者が、実際に現地での実験設定および実験実施をストックホルム大学側で行った。実験では2名の実験参加者を対象とし、システムを利用して調べ物を行いながら参加者間で相談して旅行先を決めるタスクを実施した。評価にあたっては、3. 2節の実験と同様に、ウェイクワード条件、注視検出条件、および相互注視条件の比較を行った。

実験結果より、2名で利用する場合には注視検出条件と相互注視条件で誤検出が発生し、ウェイクワード条件よりもコマンドの実行にかかる時間が長くなることが示された。また、ビデオ解析によりその原因となる状況の分析を行い、注視を用いるシステムを設計するための設計指針を得た。実験結果に基づき執筆した共著論文は HCI 分野のトップカンファレンスの一つである CSCW2019 に採録され、さらに Honorable mention を受賞した。

#### 4. 3. 注視の入出力機能を活用した音声通話機能の実装および評価実験

提案システムを応用したスマートスピーカの新たなインタラクション手法として、注視の入出力機能を活用した音声通話機能の着想を得た。これは、Amazon Echo等のスマートスピーカで利用可能なデバイス間での音声通話の接続のために、利用者の注視を検出・伝達する手法である。ここでは、2か所の遠隔地に注視の入出力を利用可能なデバイスを設置し、デバイスを介してそれぞれの利用者の注視を相互に伝達しあう。そして一方の利用者がもう一方の利用者に注視を向け、それに対して受け手が注視を返して相互注視が成立した場合に、自動的に音声通話を開始する。

実装した音声通話システムの効果を検証するための評価実験を実施した。実験では、ドキュメント作成タスク中にシステムを介して音声通話を発信、および受信するよう設定した。そして注視の入出力を用いて相互注視により音声通話を接続する注視提示条件と、GUIを用いて音声通話の発信・受信を行う条件の比較をおこなった。実験結果より、注視提示条件では相手の存在感がより強く感じられること、および通話の義務感が低下することが示された。この実験結果については、現在国際会議に共著論文を投稿中である。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 2件/うちオープンアクセス 1件）

1. 著者名 McMillan, D., Brown, B., Kawaguchi, I., Jaber, R., Solsona Belenguer, J., & Kuzuoka, H.	4. 巻 Vol. 3, No. CSCW
2. 論文標題 Designing with Gaze: Tama--a Gaze Activated Smart-Speaker	5. 発行年 2019年
3. 雑誌名 Proceedings of the ACM on Human-Computer Interaction	6. 最初と最後の頁 Article No.176
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3359278	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 川口 一画, 葛岡 英明, マクミラン ドナルド	4. 巻 21(3)
2. 論文標題 スマートスピーカーにおける注視の入出力を用いたインタラクションの効果	5. 発行年 2019年
3. 雑誌名 ヒューマンインタフェース学会論文誌	6. 最初と最後の頁 269-278
掲載論文のDOI（デジタルオブジェクト識別子） 10.11184/his.21.3_269	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 Barry Brown
2. 発表標題 Designing with Gaze: Tama -- a Gaze Activated Smart-Speaker
3. 学会等名 CSCW2019（国際学会）
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------