

# 音声グループアシスタントを用いた対面協調 Web 検索環境の構築と評価

山本 卓嗣<sup>1,a)</sup> 高田 秀志<sup>2,b)</sup>

## 概要 :

複数人がスマートフォンなどのモバイル端末をそれぞれ持ち寄り、ある共通の目的を達成するために協調して Web 検索を行う「対面モバイル協調 Web 検索」という作業がある。モバイル協調 Web 検索では、各作業者が同時に Web 検索を行う場面が多く発生するため、作業員間で他者が何をしているか把握しづらいという問題が発生する。本研究では、この問題に対して、Web ページの検索、および、Web ページの共有を音声アシスタントを用いて行う協調 Web 検索作業環境を提案する。以前、音声アシスタントを用いるに当たっての検討事項を調査するために行った実験では、モバイル端末に搭載されている Siri を用いて、音声で検索クエリを入力することで、作業員同士が何をしているか把握なりやすくなる可能性があることがわかった。一方で、Siri などの音声パーソナルアシスタントを用いると、他者のモバイル端末に誤入力してしまう等の課題点のみうけられた。従って、本研究では、音声グループアシスタントとして、Web ページの検索機能、および、Web ページの共有機能を搭載したスマートスピーカを構築する。さらに、このスマートスピーカを用いた対面モバイル協調 Web 検索環境を構築し、評価実験を行う。本稿では、評価実験に向けて構築した音声グループアシスタントを用いた対面モバイル協調 Web 検索環境について、及び、構築したシステムの予備評価実験の結果を示す。

## Development and Evaluation of a Co-located Collaborative Web Search Environment Using Voice Group Assistant

TAKATSUGU YAMAMOTO<sup>1,a)</sup> HIDEYUKI TAKADA<sup>2,b)</sup>

**Abstract:** "Co-located mobile collaborative Web search" is the work that several people search and share information on the Web to achieve a common purpose, using mobile devices such as smart phones that each worker brings. In mobile collaborative Web search, there are many scenes where each worker performs Web search at the same time, leading to a problem that it is difficult to share what the others are doing. In this research, to solve the problem, we will consider using a voice assistant or voice assistants equipped with voice recognition function. We previously conducted an experiment to investigate in issues on using the voice assistant. In the experiment, it was found that each worker could easily understand what others are doing by voice inputting a search query using Siri installed in mobile devices. On the other hand, using personal assistants such as Siri, we found problems such that spoken voice reached and was recognized by other people's mobile devices. Therefore, in this research, we create a smart speaker equipped with Web search function and Web page sharing function as voice group assistant. Furthermore, we create co-located mobile collaborative Web search environment using this smart speaker and we conduct an experiment to get user's evaluation. In this paper, we show the result of preliminary experiment of co-located mobile collaborative Web search environment using voice group assistant.

<sup>1</sup> 立命館大学大学院情報理工学研究科  
Graduate School of Information Science and Engineering,  
Ritsumeikan University, Shiga 525-8577, Japan

<sup>2</sup> 立命館大学情報理工学部

College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

<sup>a)</sup> is0240hx@ed.ritsumeikan.ac.jp

<sup>b)</sup> htakada@cs.ritsumeikan.ac.jp

## 1. はじめに

2018年現在、日本におけるスマートフォンの普及率は、79.2%、タブレット型端末は40.1%、モバイル端末に至っては95.8%であり[1]、一人一台モバイル端末を持つようになっている。それに加え、世代別インターネットの利用動向では、50代以下の世代でモバイル端末でのインターネットの利用の伸びが顕著であり、60～64歳では、モバイル端末の利用が自宅のパソコンを上回った。また、インターネットを利用する目的としては、Web上の情報を検索、収集することが挙げられており、モバイル端末はWebを利用する主なデバイスに移り変わってきている。

このような背景から、複数人である共通の目的を達成するためにWeb上の情報を検索し、共有を行う「モバイル協調Web検索」[2]という作業が日常的に行われるようになってきている。モバイル協調Web検索は、複数人が同じ場所に集まる対面環境や、グループチャット等を用いたりリモート環境で行われる。グループチャット環境でのモバイル協調Web検索を検索Botによって支援することを目指した研究[3]では、グループチャットに検索Botを導入することで、「共有、コミュニケーション、合意形成に達する容易さ」等の有用性が示された。我々の研究では、このグループチャットで形成される「マルチユーザ+Bot」という環境を対面モバイル協調Web検索で再現することを目指す。「マルチユーザ+Bot」を対面環境で再現する方法として、作業者が持ち寄ったそれぞれのモバイル端末を、音声パーソナルアシスタント(Siri<sup>\*1</sup>やGoogle Assistant<sup>\*2</sup>等)を用いて再現する方法と、音声グループアシスタント(Google home<sup>\*3</sup>やAmazon Echo<sup>\*4</sup>等)を用いて再現する方法がある。本研究では、2つの方法のうち、どちらで再現すべきかを調査するために、調査実験を行った結果[4]に基づき、Botの役割を担うグループアシスタントとしてスマートスピーカを導入する。

モバイル協調Web検索でのスマートスピーカの利用を検討する場合、マルチユーザ向けの機能を考案する必要がある。本研究では、マルチユーザ向けの機能として、Web検索機能とマルチユーザがそれぞれ持つモバイル端末のWebページ共有をアシストする機能を搭載するモバイル協調Web検索での利用に向けたスマートスピーカを構築する。さらに、構築するスマートスピーカがモバイル協調Web検索のユーザに与える効果を検証・評価するための評価実験を行い、音声グループアシスタントを用いた対面協調Web検索環境の検証・評価を行う。

以下に本稿の構成を示す。2章では、対面モバイル協調Web検索作業の課題点、音声アシスタントを用いたマルチユーザの作業支援、関連研究について述べる。3章では、本研究で構築する音声グループアシスタントの機能、実装について述べる。4章では、本環境の予備的な評価実験につ

いての結果と考察、および、今後行っていく評価実験の方針を述べる。

## 2. 研究方針

本節では、対面モバイル協調Web検索作業についての課題を指摘する。その後、マルチユーザでの作業を支援すること、既存のスマートスピーカの現状、本研究の関連研究について述べる。

### 2.1 対面モバイル協調Web検索

モバイル協調Web検索作業は、複数人の作業員によって、図1に示すような流れで行われる[5]。単独検索フェーズとは、作業員がそれぞれ持ち寄ったモバイル端末を用いたWeb検索を行うフェーズであり、個人作業である。意見交換フェーズとは、Web検索によって集まった情報について作業員同士で意見を交換するフェーズであり、共同作業である。候補比較フェーズとは、作業員同士の話し合いによって生まれたいくつかの候補を作業員同士で選定していくフェーズであり、共同作業である。

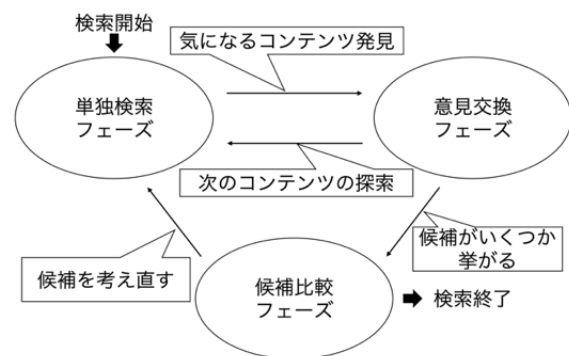


図1 モバイル協調Web検索作業の流れ

このように、モバイル協調Web検索作業は個人作業と共同作業を往復することによって行われる。しかし、個人作業と共同作業を往復することはユーザによっては困難であり、モバイル協調Web検索作業で重要とされている「作業の共有」「意識の共有」[6]が図りにくいという問題がある。従って、個人作業と共同作業の切り替えを支援するような対面モバイル協調Web検索環境を構築することで、より協調的な作業が実現できると考える。既存環境よりも協調的な作業を実現するために、本研究では、検索クエリの

\*1 Siri <https://www.apple.com/jp/ios/siri/>

\*2 Google Assistant [https://assistant.google.com/intl/ja/\\_jp/](https://assistant.google.com/intl/ja/_jp/)

\*3 Google Home [https://store.google.com/product/google\\_home?hl=ja](https://store.google.com/product/google_home?hl=ja)

\*4 Amazon Echo [https://www.amazon.co.jp/dp/B07V1KHDCS?tag=googhydr-22&ref=pd\\_sl\\_32ia8287yw\\_b](https://www.amazon.co.jp/dp/B07V1KHDCS?tag=googhydr-22&ref=pd_sl_32ia8287yw_b)

\*5 Google Speech to Text API <https://cloud.google.com/speech-to-text>

\*6 Open JTalk <http://open-jtalk.sp.nitech.ac.jp>

入力と作業者同士の Web ページの共有を音声操作で実現する環境を提案する。これにより、作業者同士で他者の状態が把握しやすくなり、「作業の共有」「意識の共有」[6]がよりしやすい環境が構築できると考える。

## 2.2 マルチユーザの作業支援

Google Home<sup>\*3</sup>等の製品化されているスマートスピーカは、ユーザがコマンド会話を話すことによって、ユーザのやりたいことを音声インタフェースを介して実現するデバイスである。音声認識の認識精度を向上させる研究は日々向上している。しかし、音声インターフェースをどのようにして既存環境に組み合わせるかについては十分な研究結果は得られていない[7]。スマートスピーカは、対面環境下におけるマルチユーザでの利用も十分に考えられる。従って、グループワーク等での使用を目的とした機能を検討することで、グループワークを支援できる可能性があると考えられる。

実際に、オフィス会議にスマートスピーカを用いることによる作業効率化を目的とした研究[8]の研究結果は、ユーザ同士の会話ログの解析にとどまっておりに実際にシステムの構築には至っていない。そこで、本研究では、日常生活で行われる対面協調 Web 検索でのスマートスピーカの利用方法を検討し、実際にその環境に必要な機能を搭載したスマートスピーカを構築する。また、構築したスマートスピーカを用いた協調 Web 検索作業がユーザにどのような影響を与えるかを検証・評価する。

## 2.3 関連研究

グループチャット環境へ検索 Bot を導入し、ユーザの作業を支援することを目的とした研究[3]では、「ユーザの互いの活動に対する意識が高まる」、「共有、コミュニケーション、合意形成に達する容易さ」、「より高いレベルでの楽しみ」といった有用性をユーザに与えることが示されている。しかし、この研究では Bot のアクションを人間が実行する Wizard of OZ[9] 実験を実施しており、実際に Bot の構築には至っていない。

マルチユーザの使用を想定したスマートスピーカについての研究[7][8][10]や、スマートスピーカの日常利用について検討した研究[7][8][11]は多く行われている。特に、マルチユーザの会議をスマートスピーカを用いて支援することを目的とした研究[8]では、ユーザ同士の会議での自然言語による会話からスマートスピーカの適切な起動トリガとなるワードを選定することを行っている。しかし、この研究は、マルチユーザの会話ログの解析にとどまっている。

グループワークでのスマートスピーカの利用を目的とした研究[7]は、スマートスピーカの既存製品を用いたワークショップを行っている。しかし、既存製品が搭載する機能は、シングルユーザ向けの機能であり、ユーザの機能要

求を満たせないと考える。そこで、本研究では、日常的に人々が行っている対面モバイル協調 Web 検索[2]で利用されるスマートスピーカを想定し、その場面でのユーザの機能要求を満たすスピーカの構築を行う。さらに、構築するスマートスピーカを用いた対面モバイル協調 Web 検索環境を構築し、スマートスピーカがユーザに与える効果を検証・評価する。

## 3. 音声グループアシスタントの構築

本節では、対面モバイル協調 Web 検索環境に音声グループアシスタントとして導入するスマートスピーカの機能と実装について述べる。

### 3.1 機能

構築する音声グループアシスタントは、検索機能と共有機能を有する。また、本音声グループアシスタントを「イザナギ」と名付け、この名前をスマートスピーカの起動トリガとした。



図 2 検索機能

検索機能の説明を図 2 に示す。検索機能は、マルチユーザの一人が音声グループアシスタントに「起動トリガ+検索機能の選定」、「検索クエリ」の 2 ステップで音声グループアシスタントに話しかけると、Web 検索結果がそれぞれのユーザのモバイル端末に表示される機能である。



図 3 共有機能

共有機能の説明を図3に示す。共有機能は、マルチユーザの一人が音声グループアシスタントに「起動トリガ+共有機能の選定」、「Web ページ共有の際の送信者名」の2ステップで音声グループアシスタントに話しかけると、送信者がそのとき開いている Web ページを他のユーザに共有する機能である。

### 3.2 音声認識の実装

音声グループアシスタントの構築には、Raspberry Pi にマイク付きスピーカを接続し、Google Speech to Text API<sup>\*5</sup>を用いる。音声入力を用いた、検索機能、共有機能の実現方法を図4に示す。

このAPIを用いることで、スピーカへの音声入力をテキスト変換し、テキストに応じて、検索、共有機能を実現する。

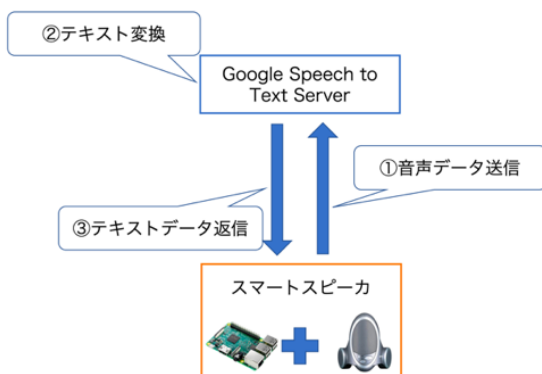


図4 実装概要

また、音声出力には、Open Jtalk ライブラリ<sup>\*6</sup>を用いて、テキストファイルから作成した音声合成の wav ファイルを用いる。

### 3.3 検索機能、共有機能の実装

音声グループアシスタントを用いた対面モバイル協調 Web 検索環境を構築するために、iPad, Raspberry Pi, Macbook Pro を使用する。各端末の通信は、Macbook Pro をローカルサーバとする CGI ライブラリを用いたサーバ・クライアント方式で行う。図5に音声グループアシスタントを用いた Co-located モバイル協調 Web 検索環境の実装の概要を示す。

初期設定として、モバイル端末の端末名は、ユーザの共有時に用いる名前に設定する。

モバイル端末には、Web 検索機能とクライアント機能を搭載するアプリケーションを構築し、Web ページが遷移する度に Flag, 端末名, URL をサーバに Post する。この Flag は、モバイル端末からの Post か、スピーカからの Post かを判定するためのものであり、モバイル端末からの Post の場合は 0、スピーカからの Post の場合は 1 の値を

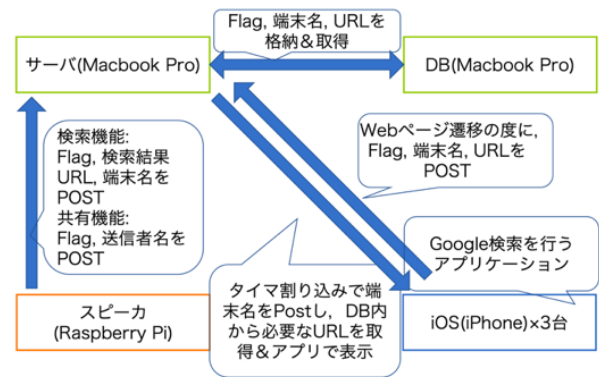


図5 実装概要

取る。サーバは、受け取った Flag, 端末名, URL を DB に格納する。さらに、モバイル端末は、タイマ割り込みを用いて、バックグラウンドでサーバにリクエストを送り、DB 内で Flag が 1 のデータが見つかった場合、その行の URL を取得し、アプリケーション上に表示する。

スマートスピーカは、常に音声認識を行い、作業者の「イザナギ 検索」または、「イザナギ 共有」という発話内容(起動トリガ+機能の選定)の入力待ち状態をとる。検索機能の実現の際、スピーカは「イザナギ 検索」という発話内容を認識すると、「検索クエリを教えてください」という応答を音声出力し、検索クエリ入力待ち状態に入る。検索クエリが音声入力されると、Google 検索を行い、Flag, 端末名, 検索結果 URL を Post する。サーバは Post されたデータを DB に格納する。

また、共有機能の実現の際、スピーカは「イザナギ 共有」という発話内容をスピーカが認識すると、「誰が共有しますか」という応答を音声出力し、送信者名の入力待ち状態に入る。送信者名が音声入力されると、Flag, 送信者名を Post する。サーバは、Post された送信者名が含まれる端末名を DB 内で検索し、最新の URL を取得する。その後、Flag, 送信者名, 取得した URL を DB に格納する。

## 4. 予備評価実験

本節では、評価実験に用いる音声グループアシスタントの使いやすさについて、ユーザ評価を得るために行った予備的な評価実験について述べる。

### 4.1 実験方法

本実験の目的は、3章で述べた実装方法で構築する音声グループアシスタントで評価実験を行う前に、システムの改善点等を得ることである。

本実験の被験者は、情報系の学生 2 名である。被験者 2 名は、図6のように隣り合わせて着席した。

被験者は、それぞれモバイル端末 (iPad) を 1 台ずつ所持し、構築した Web 検索アプリケーションと音声グループアシスタントを用いて、モバイル協調 Web 検索を行っ



図 6 作業者間の位置関係

た。実施した作業の内容は、大阪での晩御飯処の選定である。被験者は、Web 検索と、お互いのモバイル端末の Web ページを共有する際に、音声グループアシスタントを使用する。

また、実験の概要を説明する際に、音声グループアシスタントの検索機能、共有機能を被験者 2 名にそれぞれ練習として使用してもらった。

#### 4.2 評価方針

本実験では、20 段階評価と自由記述欄を設けた自作のアンケートへの回答と、作業風景を撮影した動画の観察をもとに、4 章で述べた音声グループアシスタントのユーザ評価を得る。アンケートの質問内容を表 1 に示す。それぞれの質問には自由記述欄を設け、なぜその値で評価したかを回答してもらった。

表 1 アンケート内容 (20 段階評価+自由記述)

設問	質問内容
Q.1	音声検索、共有後の iOS 端末での Web ページ表示はスムーズでしたか？
Q.2	音声検索、共有の精度は十分でしたか？
Q.3	各機能の音声入力に 2 ステップを要しましたが使いやすいかったですか？
Q.4	音声出力から、次に何を言えばいいかわかりやすかったですか？

#### 4.3 結果

まず、ユーザ評価アンケートの結果を図 7 に示す。

Q.1. の評価値に対する自由記述では、「Air Drop などと比較してもそんなに遅く感じなかったので、速い方なのだと思うから。」「検索、共有後に、実行されているか、少し不安を感じる時があった」という回答を得た。

Q.2. の評価値に対する自由記述では、「正しく音声認識されるには、しっかり滑舌良く話さなければならなかった」

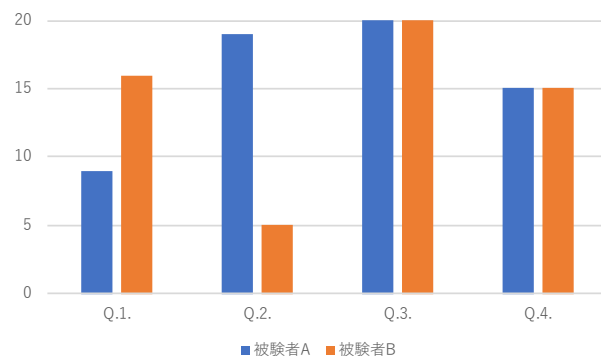


図 7 アンケート結果

という趣旨の回答を 2 名から得た。

Q.3. の評価値に対する自由記述では、「シンプルで使いやすいかった」「1 ステップの場合、音声入力を話す前に発話内容をしっかり決める必要があり、3 ステップの場合、手間が増えると思った」という趣旨の回答を得た。

Q.4. の評価値に対する自由記述では、「クエリという単語を一般的な人がどれだけ理解できるか気になった」という回答を 2 名から得た。

その他の自由記述には、「イザナギという名前が発音しにくい」という回答を得た。

さらに、本実験のタスクは 7 分 52 秒で達成され、検索機能は 2 回使用された。検索クエリは「大阪 晩御飯」と、「大阪 テレビで紹介された店」であった。共有機能は 1 回使用された。共有された Web ページは、お好み焼き屋についての Web ページであった。また、今回は、検索機能、共有機能の音声入力時の誤認識はなかった。

#### 4.4 考察

Q.1. の回答は、「Air Drop などと比較してもそんなに遅く感じなかったので、速い方なのだと思うから。」「検索、共有後に、実行されているか、少し不安を感じる時があった」というように、意見が別れた。既存の近距離間情報共有手法と比較すると、スムーズに感じる事が考えられる。また、低く評価した被験者については、音声入力为正しく認識されたかわからないため、このような不安が生じたと考えられる。

Q.2. の評価値に対する自由記述では、「正しく音声認識されるには、しっかり滑舌良く話さなければならなかった」という趣旨の回答を 2 名から得た。しかし、被験者 2 名の評価値は、5 と 19 となり、大きく別れた。これは、共有機能の際に、被験者は自身の名前を発話する必要があり、音声認識で認識しづらい固有名詞があったためと考えられる。また、滑舌や音声入力の発話内容について、日常会話よりも意識しながら使用していると考えられる。この問題の改善案として、作業を始める前に被験者がそれぞれ名前を言って登録できるようにする等が考えられる。

Q.3. についての回答は、2名共に高い評価を得た。従って、2ステップの音声入力の問題ないと考えられる。

Q.4. についての回答は、2名共に高い評価を得た。しかし、自由記述で「クエリという単語の使用を避けた方が良い」という同じ改善案が記述されていた。評価値が高くなった原因として、被験者が共に情報系の学生であったことが考えられる。そこで、評価実験の際は、現在の「検索クエリを教えてください」という音声出力を、「検索ワードを教えてください」等に改善する。

Q.5. の「イザナギという名前が発音しにくい」という回答から、評価実験の際は、起動トリガを音声認識しやすく、かつ日常会話であまり用いられない日本語の単語に変更する。

## 5. おわりに

本稿では、音声グループアシスタントを用いた対面モバイル協調 Web 検索環境の構築と評価について述べた。調査実験の結果、対面モバイル協調 Web 検索作業中の Web 検索のクエリ入力を音声で行うことにより、他者が何をしているか把握しやすくなる可能性があることがわかった。また、対面モバイル協調 Web 検索作業には、音声グループアシスタントを導入する方針に至った。さらに、評価実験のために構築した音声グループアシスタント及び対面モバイル協調 Web 検索環境の操作性についてのユーザ評価を得るために予備評価実験を行った。この結果、構築した音声グループアシスタントの改善すべき点がわかった。

今後は、予備評価実験の実験結果からわかったシステムの改善点を改善し、本実験を行い、音声グループアシスタントを用いた対面モバイル協調 Web 検索環境がユーザ与える効果を検証・評価する。

## 参考文献

- [1] 総務省：総務省データベース。
- [2] Morris, M. R., Lombardo, J. and Wigdor, D.: WeSearch: supporting collaborative search and sensemaking on a tabletop display, *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM, pp. 401–410 (2010).
- [3] Avula, S., Chadwick, G., Arguello, J. and Capra, R.: SearchBots: User Engagement with ChatBots During Collaborative Search, *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, ACM, pp. 52–61 (2018).
- [4] 山本卓嗣, 高田秀志ほか：音声エージェントを用いた対面協調検索環境の検討, 研究報告グループウェアとネットワークサービス (GN), Vol. 2018, No. 16, pp. 1–6 (2018).
- [5] 奥梓, 小牧大治郎, 荒瀬由紀, 原隆浩, 上向俊晃, 服部元, 西尾章治郎：携帯端末を用いた協調 Web 検索におけるコンテンツ比較支援インタフェース, *DEIM Forum 2010* (2010).
- [6] 岡田謙一：協調作業におけるコミュニケーション支援, 電子情報通信学会誌= THE JOURNAL OF THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS, Vol. 89, No. 3, pp. 213–217 (2006).
- [7] Porcheron, M., Fischer, J. E., McGregor, M., Brown, B., Luger, E., Candello, H. and O'Hara, K.: Talking with conversational agents in collaborative action, *companion of the 2017 ACM conference on computer supported cooperative work and social computing*, ACM, pp. 431–436 (2017).
- [8] McGregor, M. and Tang, J. C.: More to meetings: challenges in using speech-based technology to support meetings, *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ACM, pp. 2208–2220 (2017).
- [9] Dahlbäck, N., Jönsson, A. and Ahrenberg, L.: Wizard of Oz studies—why and how, *Knowledge-based systems*, Vol. 6, No. 4, pp. 258–266 (1993).
- [10] Luger, E. and Sellen, A.: Like having a really bad PA: the gulf between user expectation and experience of conversational agents, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, pp. 5286–5297 (2016).
- [11] Porcheron, M., Fischer, J. E. and Sharples, S.: Do animals have accents?: talking with agents in multi-party conversation, *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ACM, pp. 207–219 (2017).