

発話単語のオンライン検索による提示画像を用いた 身体引き込み発話促進システムの開発

窪田 正勝[†] 渡辺 富夫 石井 裕[‡]岡山県立大学大学院 情報系工学研究科[†] 岡山県立大学 情報工学部[‡]

1. はじめに

近年、音声認識技術は目覚ましい進歩を遂げている。Softbank 社の Pepper や Apple 社の Siri などに代表される様々なロボットやアプリケーションに搭載されており、我々にとって身近な存在になっている。

一方、著者らはこれまで、コミュニケーションにおいて重要な役割を果たしている、うなずきなどの身体動作に着目し、話者の発話音声からうなずきなどの身体引き込み動作を自動生成する技術 iRT (InterRobotTechnology) を開発してきた[1]。

先行研究として、VR を用いた身体引き込み発話促進システムを開発している[2]。このシステムは、使用者の発話単語の 3D オブジェクトを仮想空間内に生成する。その生成されたオブジェクトが発話音声に対するうなずきによる身体引き込み反応を示し、仮想空間内の人型聞き手キャラクターもうなずき反応を示す。また、ヘッドマウントディスプレイを使用して臨場感のある視覚提示を行うことで、使用者の発話促進や、視覚情報の提示による思い出想起促進を目的としている。しかしこのシステムでは、没入感を高めるために、空間に配置するための違和感のない 3D オブジェクトの制作に大幅な時間がかかるため、使用者の発話単語すべてに対応することができなかった。

そこで本研究では、より多様な情報に対応するために、発話単語から画像をオンライン検索し、検索結果の画像を平面ディスプレイに提示する。提示した検索画像がうなずきによる身体引き込み反応を示す発話促進システムを開発する。オンライン検索機能を実装することで従来の発話促進に加え、使用者の意図しない画像が

提示された場合でも、連想的に新たな発想を支援する効果も期待できる。

2. 関連研究

発話の可視化についての研究は、これまでも行われてきた。長野・吉野らの研究では、発話者の音声認識の結果から、文字や画像を提示することで発話内容の把握を支援している[3]。また瀬島らの研究では、身体引き込み反応する絵画オブジェクトを用いて、コミュニケーション支援への有効性を示している[4]。しかし、発話単語を元に提示された検索画像が、使用者の発話に対し何らかの聞き手反応を示す研究は行われていない。

3. コンセプト

本システムのコンセプトを図 1 に示す。発話があると、聞き手キャラクターがうなずきによる身体引き込み反応を示す。使用者の発話内に名詞が見つければ、その名詞で画像検索を行い、検索画像が見つければその画像を仮想空間内に提示する。提示された画像はその後、使用者の発話に対するうなずき反応を示す。うなずき反応による発話促進や、提示画像による発想支援が期待でき、とくに話す元気になる高齢者の発話促進に有効である。

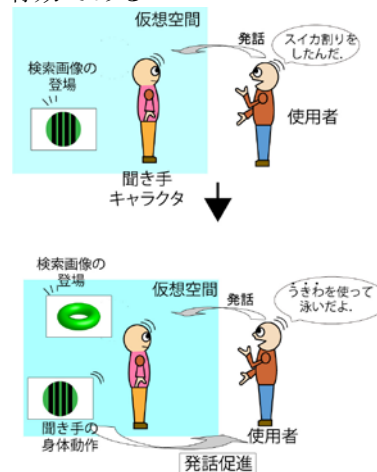


図 1 コンセプト

Development of an Embodied Entrainment Speech Promotion System Using Presentation Images by the Online Retrieval of Spoken Words

[†]Masakatsu-Kubota · Graduate School of Systems Engineering, Okayama Prefectural University

[‡]Tomio Watanabe · Faculty of Computer Science and Systems Engineering, Okayama Prefectural University

[‡]Yutaka Ishii · Faculty of Computer Science and Systems Engineering, Okayama Prefectural University

4. システム構築

4.1 システム構成

発話から画像提示までの流れを図 2 に示す。仮想空間の構築には Unity を用いる。発話音声声を音声認識した後、形態素解析を行い、その中から名詞を抽出する。音声認識には Google 社の WebSpeechAPI を用い、形態素解析には Mecab を用いた。次に抽出した名詞から、画像検索 API (pixabayAPI) を用いて画像をオンライン検索し、検索結果の中で 1 番目の画像を取得し、仮想空間内に表示する。発話から画像提示までは、5 秒程度かかるが、これは音声認識と画像検索で外部 API と通信を行っており、音声認識に約 2 秒、画像検索に約 3 秒時間を要しているためである。

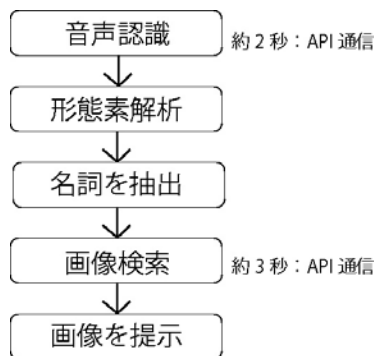


図 2 画像提示までの流れ

4.2 システム画面

システムの使用画面を図 3 に示す。システム開始時には、仮想空間内に聞き手キャラクターのみが配置されている。使用者が発話を行うと、音声認識により発話をテキスト化する。音声認識の結果は、画面左下に吹き出しと共に表示される。発話の中で名詞が見つければ、聞き手キャラクターが考え中であることを表現すると同時に、画像検索が実行される。画像検索中の画面を図 4 に示す。検索結果が得られれば、発話単語の画像が提示される。もし得られなければ、次の名詞を検索する。発話単語画像は聞き手キャラクターと共に iRT を用いて自動生成されるうなずき反応を示す[1]。提示画像は最大で 10 枚としており、11 枚目以降は最も古い画像から更新して提示を行う。画像の提示位置はランダムに設定しており、キャラクターの後ろに提示される。

5. おわりに

本研究では、発話単語からオンライン検索さ

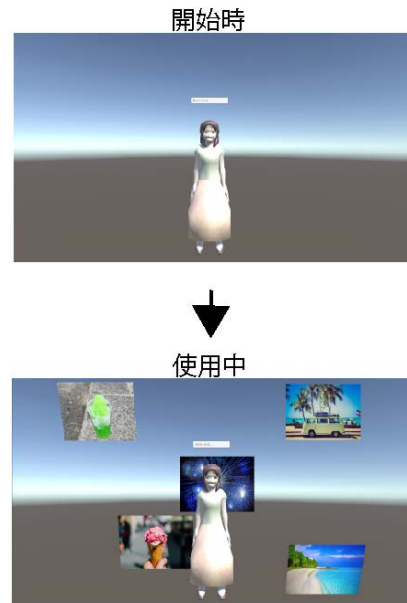


図 3 システムの使用画面

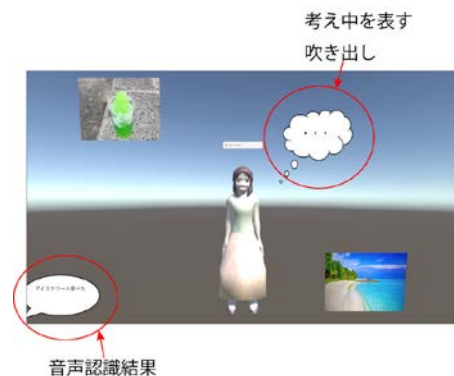


図 4 画像検索中の画面

れた提示画像がうなずきによる身体引き込み反応を示す、発話促進システムの開発を行った。本システムでは、聞き手キャラクターの役割はうなずきのみであったが、今後は使用者にとってインタビュアーとなる役割を実装し、発想支援システムとして展開していく予定である

参考文献

- [1] 渡辺：心が通う身体的コミュニケーション：日本機械学会誌, 121 巻, 1195 号, pp. 14—17 (2018).
- [2] 窪田, 渡辺, 石井：聞き手キャラクターと発話単語オブジェクトを用いた身体引き込み発話促進システムの開発：ヒューマンインタフェースシンポジウム 2018 論文集, pp. 795—198 (2018).
- [3] 長野, 吉野：拡張現実感技術を用いた発話可視化システム MIERUKEN の開発：情報処理学会研究報告, 2009-DPS-141, 20 号, pp. 1—8 (2009).
- [4] 瀬島, 石井, 渡辺：アバタコミュニケーション支援のための音声駆動形身体的引き込み絵画を用いた仮想観客システム：日本機械学会論文集, 78 巻, 786 号, pp. 523—534 (2012).