

(410) 情報 その他

発話単語に基づく身体引き込み画像を重畳合成した 実映像対話システムの開発

Development of a Video Communication System which Superimposes Embodied Entrainment Images based on Spoken Words

窪田 正勝[†]

Masakatsu Kubota[†]

渡辺 富夫^{††}

Tomio Watanabe^{††}

石井 裕^{††}

Yutaka Ishii^{††}

[†]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

Web カメラを用いた実映像対話は、私達にとって身近な技術の一つである。遠隔地でも互いの顔を見ながら通話を楽しむ機会は、今や決して珍しいことではない。

著者らはこれまでに相手顔画像を合成した身体的引き込み観客キャラクタを用いた実映像対話システムを開発し、実験において対話促進や合意形成対話での主張の促進に対して評価されるなど、システムの有効性を確認している[1]。

また、発話音声により有効に活用する手法として、オンライン検索画像を用いた身体引き込み発話促進システムを開発した[2]。発話単語から画像をオンライン検索し、検索画像を仮想空間に提示する。画像は仮想空間内の聞き手キャラクタと共にうなずき反応を示す。これにより、使用者の発話を促進する。但しこのシステムは、使用者が一人で発話を行うことを想定して開発している。

本研究では、Web カメラを用いて、話し手ユーザと聞き手ユーザを映像提示する。話し手ユーザの発話単語から画像を重畳合成し、うなずき反応を行う、実映像対話システムを開発した。話し手ユーザの発話が画像として可視化され、且つ画像がうなずき反応を示すことで、話し手の発話促進を行うことを目的とする。また、提示画像を二者間で共有することにより、発想支援や共創支援の効果が期待できる。

2 関連研究

発話内容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、画像の身体引き込み反応について、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]。

3 システム概要

3.1 コンセプト

本システムのコンセプトを図1に示す。本システムではまず、話し手ユーザと聞き手ユーザが同一画面上に PIP (Picture in Picture) で表示されている。ここで話し手ユーザが発話を行い、発話内容が可視化された画像を重畳合成する。画像はその後の話し手ユーザの発話に対しうなずきによる身体引き込み反応を示し、発話を促す。相手映像だけでなく自己映像の表示により、自己映像を相手映像と対比的に確認でき、話し手に安心感を与えることができる。また提示画像と自己映像との対比により、回想的に過去の出来事についての発話支援効果が得られる。

本研究では将来的に、人間とエージェントとの対話において画像提示の有効性を検証することを目指している。本システムでは、まずは対話相手をエージェントではなく人間として、自然な対話を行うことで、より文脈に即した画像提示を試みる。

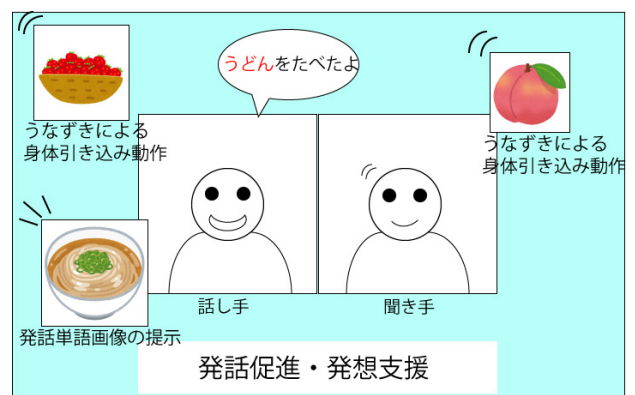


図1：コンセプト

3.2 システム概略

本システムの使用画面を図2に示す。システム開始時には、仮想空間内に話し手側と聞き手側の

映像が提示されている。話し手側が発話を行うと、音声認識により発話をテキスト化する。テキストの中から予め登録された単語が出現すれば、その単語に対応した画像を重畳合成して提示する。単語と画像はそれぞれ一単語につき一枚の画像が割り当てられている。そのため、同じ単語が複数回出てきた場合は同じ画像が複数枚提示される。画像の提示位置はランダムに設定されているが、コミュニケーションを阻害しない目的で、話者映像の中央付近には画像が提示されない。システムを終了するまでは画像は次々と提示されていく。提示された画像のうなずき反応については、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生成する技術 iRT (Inter Robot Technology)により、うなずき動作を自動生成する[5]。画像のうなずきの様子を図3に示す。複数の画像が提示された場合には、その全ての画像が発話に対しうなずき反応を示す。

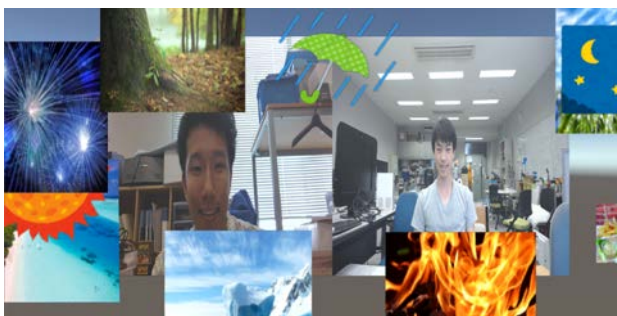


図2：システムの使用画面

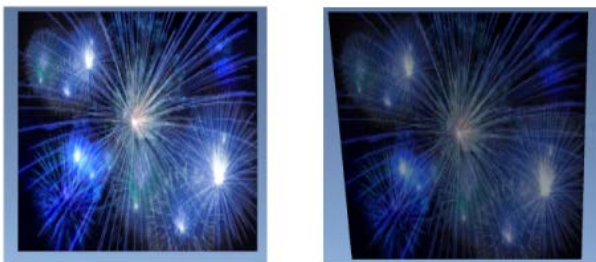


図4：画像のうなずき
(左：待機中 右：うなずき)

3.3 システム構成

本システムの構成を図4に示す。仮想空間の構築にはUnityを用いる。音声認識にはAmiVoiceSP2を用いる。また、WebカメラにはLogicool Webcam C930eを使用している。話し手側と聞き手側に提示する映像は、PCの画面を複製して提示している。つまり、使用者の二人は同じ映像をそれぞれ別々のディスプレイで見ながら対話を行う。提示する身体引き込み画像については予め400個の単語と

画像を登録しており、発話を基にこの中から画像を適宜提示する。ローカルな環境に画像を用意することで、話し手ユーザの個人的な思い出写真などを登録することができる。未登録の単語に対して、オンライン画像を用いて反応提示させることも可能ではあるが、現状のシステム構成では、発話から画像提示までに数秒の遅延が発生するなどの問題がある。学生2組4名に本システムを使用してもらったところ、「話した画像が出てくるのが面白い」「相手の映像よりも、登場する画像を見ていた」「相手映像はあまり見ていなかった」などのコメントが得られた。

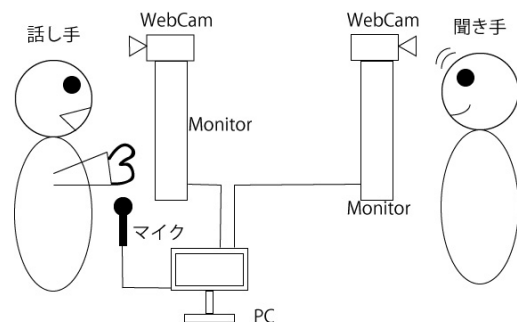


図4：システムの構成

4 おわりに

本研究では、話し手ユーザと聞き手ユーザの二者間における発話単語の身体引き込み画像を用いた実映像対話システムを開発した。提示画像のうなずきによる発話促進効果に加え、発想支援および共創支援の効果が期待できる。今後はシステムの有効性を、発話促進、発想支援、共創支援の観点から評価実験を行い検証していく。

参考文献

- [1] 石井, 渡辺: 相手顔画像を合成した身体的引き込み観客キャラクターを用いた実映像対話システム: ヒューマンインタフェース学会誌, 17 巻, 3 号, pp. 265-274(2015).
- [2] 窪田, 渡辺, 石井: 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: 情報処理学会第 81 回全国大会, pp. 4-155-4-156 (2019).
- [3] 吉野, 長野: 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. 523-524(2012).
- [5] 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).