3ZA-4

音声認識による動作・情動表現機能を有する音声駆動型身体的引き込みキャラクタの動画視聴による評価

四方　拓†　　藤井　亮‡　　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

音声認識は,「しゃべってコンシェル」や「Siri」など多くのスマートフォンに導入され，日常的に使用できる身近な技術になりつつある．音声インタラクションシステムやコミュニケーションシステムの提案がなされ，音声認識技術を用いた対話エージェント等の研究開発も積極的に進められている [1][2].

著者らは，発話音声から身体的引き込み動作を自動生成する InterActor に音声認識による動作・情動表現機能を付与し，円滑なコミュニケーションを促す身体的引き込みキャラクタシステムを開発し，評価実験により，コミュニケーション支援への有効性を示している [3]．しかし，前実験では，直接的な情動表現としてハートや感嘆符などの記号を用いたシンボル表現を中心に評価しており，身体表現のみによるロボットなどの物理メディアに導入した際の音声認識による身体表現効果を検討する必要がある．

そこで本研究では，従来の InterActor に音声認識を併用し，物理メディアに導入することを考慮し，シンボル表現を用いない動作・情動表現機能を付与した身体表現のみによる効果を検証する評価実験を行っている．

2　システム開発

2.1　コンセプト

本システムのコンセプトを図 1 に示す．InterActor はディスプレイ上に表示される CG キャラクタであり，音声リズムからうなずきや身振り手振りなどの身体的引き込み動作を自動生成し，表現している．聞き手動作として，対話者の語りかけに対して身体全体で引き込むように反応し，話し手動作としてリズム同調的な動作を行うことで，インタラクティブなコミュニケーションを実現している．従来の InterActor の身体的引き込み動作に加えて，音声認識による言葉の意味に対応した動作・情動表現および記号を用いたシンボル表現をキャラクタに追加することによって，使用者の思いが伝わりやすくなり，会話意欲の促進を働きかけるとともにより円滑なコミュニケーションが実現できる．

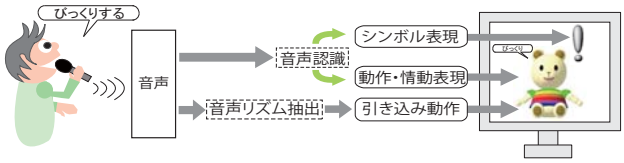

図 1: コンセプト

2.2　音声認識による動作・情動表現

音声認識には音声認識エンジン Julius を使用した．InterActor に動作・情動表現機能を付与するために，タイピングのリズムから身体的引き込み動作を自動生成するチャットシステム InterChat におけるテキスト情報から対応するキャラクタ動作を適用する技術を使用した [4]．音声認識によって音声を文字列に変換し，文字列内にある特定の単語を認識し，対応した動作をキャラクタの身体動作に関連付けた.「こんにちは」「バイバイ」のように，頭や手などの身体で反応する動作だけでなく,「好き」のハートや「びっくり」の感嘆符などの記号を用いたシンボル的な表現と組み合わせた提示を行う．単語に対応した動作はデータベース化しており，InterChat におけるチャット特有の表現を削除し，音声対話でよく使われるキーワードを追加するといった変更を加えた．現在，データベースには，383 種類の単語とそれに対応した 105 種類の動作が登録されている．動作・情動表現の例を図 2 に示す.

図 2: 動作・情動表現の例

Evaluation by video viewing of speech-driven embodied entrainment characters with emotional expressions and motions.
†Hiraku Shikata · Graduate School of Systems Engineering, Okayama Prefectural University
‡Ryo Fujii · Graduate School of Systems Engineering, Okayama Prefectural University
††Tomio Watanabe · Faculty of Computer Science and System Engineering, Okayama Prefectural University
‡‡Yutaka Ishii · Faculty of Computer Science and System Engineering, Okayama Prefectural University

3　評価実験

3.1　実験概要

身体表現のみによる効果の有効性を評価するために評価実験を行った．実験では出力動作や，発話量を一定にするために，同じ音声データから3つの動画を作成し評価した．Aモードの動画では，キャラクタがInterActorの話し手･聞き手動作を行う．Bモードの動画では，キャラクタがInterActorの話し手・聞き手動作に加え，音声認識による挨拶などの身体動作を行う．またCモードの動画では，ハートや感嘆符などの記号を用いたシンボル表現の効果を検討するために，InterActorの話し手･聞き手動作に加え，シンボル表現を含めた動作･情動表現を行う．被験者には聞き手を想定して，3つの動画視聴によるシステムの評価を行わせた．まず，AからCのモードからランダムに2つのモードを抽出し，同一画面上で同時視聴させて，どちらが総合的に良いか一対比較を行わせた．モードの提示順序はカウンターバランスをとり，3モードの比較で3(＝ ${}_3C_2$) 回行った．次に，AからCのモードを再度個別に視聴させ，5項目（楽しさ，好み，思いが伝わる，話の聞きやすさ，システムを使用したいか）について7段階で官能評価させた．被験者は19歳から24歳の男女学生24人である．

3.2　実験結果

一対比較の結果を表1に示す．Bradley-Terryモデルを想定し、各モードの強さπを最ゆう推定した結果を図3に示す．Cモードが最も高く，次いでB，Aの順で評価された．

表 1: 一対比較の結果

	A	B	C	total
A		3	3	6
B	21		4	25
C	21	20		41

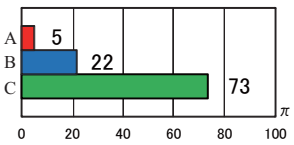

図 3: Bradley-Terry モデルによる π の値

7段階評価の結果を図4に示す．Wilcoxonの符号順位検定を行った結果，CとAの間では「楽しさ」「思いが伝わる」「システムを使用したいか」の項目において有意水準0.1 %で,「好み」「話の聞きやすさ」の項目において有意水準1 %で有意差が認められた．CとBの間では「楽しさ」「思いが伝わる」の項目において有意水準1 %の有意差が認められた．AとBの間では「楽しさ」「思いが伝わる」「話の聞きやすさ」の項目において有意水準0.1 %で,「好み」「システムを使用したいか」の項目において有意水準1 %で有意差が認められた．

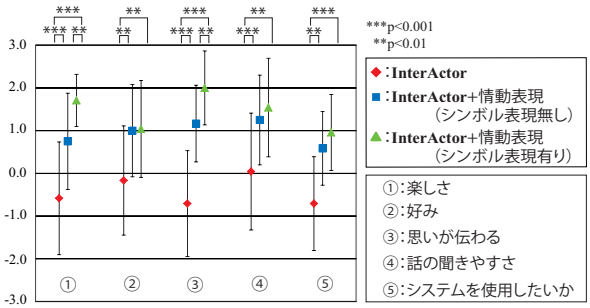

図 4: 7 段階評価の結果

3.3　考察

実験結果から，Cモードの有効性が再確認された[3]．一方，7段階評価の結果から，全ての項目でAモードに対してB，Cモードの間に有意差があり，効果が確認された．ここで，BモードとCモードの間に，2つの項目で有意差が確認できるが，それ以外の項目では同程度の評価である．このことから，シンボル表現を用いない身体動作のみの表現でも，コミュニケーション支援に有効であると考えられる．

4　まとめ

本研究では，身体表現のみによるロボットなどの物理メディアに導入した際の音声認識による身体表現の効果を検討するために，InterActorに音声認識を併用し，シンボル表現を用いない動作・情動表現機能を付与し，情報提示における聞き手を想定した評価実験を行った．その結果，シンボル表現を含む動作・情動表現機能の効果が再確認されたが，含まない場合でも十分に効果があることが示された．

参考文献

[1] 嵯峨山 茂樹, 西本 卓也, 中沢 正幸:擬人化音声対話エージェント (＜特集＞音声情報処理技術の最先端), 情報処理学会誌 vol.45, No.10, pp.1044-1049 (2004).

[2] 尾形 正泰, 大澤 博隆, 篠沢 一彦, 今井 倫太: Voicy: ロボットモーションを伴ったつぶやきシステムの提案, 情報処理学会研究報告.HCI, ヒューマンコンピュータインタラクション研究会 報告 2011-HCI-144 No.3, pp.1-4 (2011).

[3] 藤井 亮, 四方 拓, 服部 憲治, 渡辺 富夫, 石井 裕:音声駆動型身体的引き込みキャラクタシステムにおける音声認識による動作・情動表現提示の評価, 第14回計測自動制御学会システムインテグレーション部門講演会論文集, pp.1945-1948 (2013).

[4] 服部 憲治, 渡辺 富夫, 山本 倫也:タイピング駆動型身体引き込みキャラクタチャットシステム Inter-Chat, ヒューマンインタフェース学会論文誌 Vol.15, No.4, pp.53-62 (2013).