

音声認識による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システムの開発

藤井 亮^{1,a)} 渡辺 富夫^{2,b)} 石井 裕^{2,c)}

概要：著者らは、発話音声からコミュニケーション動作を自動生成する音声駆動型身体的引き込みキャラクタ InterActor を開発し、コミュニケーション支援の有効性を示している。従来の InterActor では、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、キャラクタで表現している。さらにこの身体動作に加えて、音声認識により言葉の意味から身体動作を生成することができれば、使用者の思いをキャラクタに表現させるといった、より表現性の向上したコミュニケーションが実現できると考えられる。本研究では、音声認識を併用することで、InterActor に言葉の意味に対応する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ている。

Development of a Speech-Driven Embodied Entrainment Character System with an Emotional Expression Function by Speech Recognition

RYO FUJII^{1,a)} TOMIO WATANABE^{2,b)} YUTAKA ISHII^{2,c)}

Abstract: We have developed a speech-driven embodied entrainment CG character called InterActor that automatically generates communicative motions and actions such as nods for entrained interaction from voice rhythm based on only speech input. In this paper, we develop an advanced embodied entrainment character system with the expressive function of affective motions and signs from the meaning of utterance by speech recognition as well as the entrained function of InterActor for affective avatar-mediated communication.

1. はじめに

今日、音声認識は多くのスマートフォンに導入されるなど、日常的に使用できる身近な技術になりつつある。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おり、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ている^{[1][2]}。また音声合成システムとの統合により、受付ロボットなどの案内エージェントが提案されている^{[3][4]}。一方、対面コ

ミュニケーションに代表される音声を介したコミュニケーションにおいては、音声に対するうなずき反応などによって、対話者同士が互いに引き込み合うことで円滑にコミュニケーションしている。著者らは、ロボットやCGキャラクタに引き込み動作を行わせることで、コミュニケーション支援につながると考え、発話音声から身体的引き込みを伴う豊かなコミュニケーション動作を自動生成するインタラクション技術 iRT を開発してきた。さらに iRT をCGキャラクタに組み込んだ音声駆動型身体的引き込みキャラクタ InterActor を開発し、コミュニケーション支援の有効性を示している^[5]。また、iRT を応用して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し、コミュニケーション支援を行うチャットシステム InterChat において、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も開発している^[6]。本研究では、従来の発話音声から身体動作を自動生成する InterActor に加え、音声認識を併用す

¹ 岡山県立大学大学院 情報系工学研究科
Graduate School of Systems Engineering, Okayama Prefectural University

² 岡山県立大学情報工学部
Faculty of Computer Science and Systems Engineering, Okayama Prefectural University

a) ryo_fujii@hint.cse.oka-pu.ac.jp

b) watanabe@cse.oka-pu.ac.jp

c) ishii@cse.oka-pu.ac.jp

ることで、InterActor に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のプロトタイプを開発している。また、画面に自分と相手のキャラクタを表示し、通信を行うコミュニケーションシステムの開発を行っている。

2. InterActor

音声駆動型身体的引き込みキャラクタ InterActor はディスプレイ上に表示される CG キャラクタであり、話し手と聞き手の両機能を備えている。また、各関節部位の曲げ動作や回転動作を組み合わせることで、多様なコミュニケーション動作を表現することができる。InterActor は対話者の語りかけに聞き手としてうなずきや身体動作などの身体全体で引き込むように反応する。一方で相手の音声が入った時には話し手としての身体的引き込み動作をすることで、インタラクションを円滑にし、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。

2.1 インタラクションモデル

InterActor のインタラクションモデルは、音声の ON-OFF パターンに基づくうなずき反応モデルを導入している(図 1)。このモデルでは、マクロ層とマイクロ層からなる階層モデルを用いてうなずきの予測を行っている。予測には $[i - 1]$ ユニット以前のユニット時間率 $R(i)$ (ユニット区間での ON 区間の占める割合、(2) 式) の線形結合で表される (3) 式の MA(Moving-Average) モデルを用いる。雑音は身体のゆらぎを示すノイズ成分である。

$$M_u(i) = \sum_{j=1}^J a(j) R(i-j) + u(i) \quad (1)$$

$$R(i) = \frac{T(i)}{T(i) + S(i)} \quad (2)$$

$a(j)$: 予測係数, $T(i)$: i 番目ユニットでの ON 区間,
 $S(i)$: i 番目ユニットでの OFF 区間, $u(i)$: 雑音

$$y(i) = \sum_{j=1}^K b(j) V(i-j) + w(i) \quad (3)$$

$b(j)$: 予測係数, $V(i)$: 音声データ, $w(i)$: 雑音

予測値 $M_u(i)$ がある閾値を越えて、うなずきが存在すると予測された場合には、処理はマイクロ層に移る。マイクロ層では音声の ON-OFF データ (30Hz, 60 個) を入力とし、MA モデルでうなずきの開始時点 $y(i)$ を推定する。予測値が閾値を越えた場合には InterActor をうなずかせる。瞬きについては、対面コミュニケーション時における瞬き特性に基づいて、うなずきと同時に瞬きさせ、それを基点として指数分布させている。その他の身体反応の推定にはうなずきの予測値を用い、うなずきよりも低い閾値で InterActor の各部位 (頭部, 胸部, 右肘, 左肘) のうちいくつかを選

択して動作させることでうなずきと関係付けている。話し手のモデルについても同様に、対面コミュニケーション時の音声と身体動作の特性から、音声の ON-OFF パターンに基づく身体全体の動作を予測するモデルと音声の振幅に基づく腕部動作モデルを導入している。身体動作モデルとしては全ての動きの ON-OFF の総和データから体の動くタイミングを予測し、閾値を越えたときに InterActor の各部位 (頭部, 胸部, 右肘, 左肘) のうちいくつかを選択し、動作させることで発話音声と関係付けている。

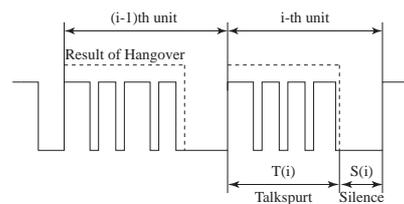
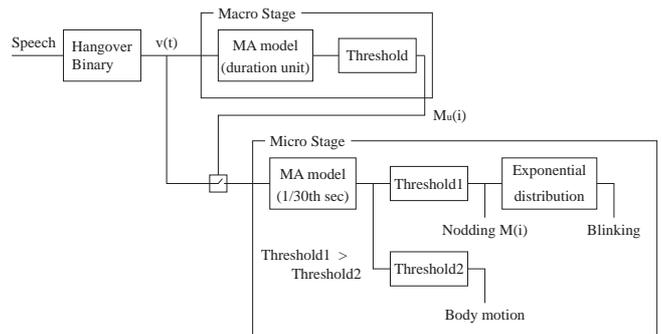


図 1 インタラクションモデル

Fig. 1 Interaction model.

3. コンセプト

従来の InterActor では、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、キャラクタで表現している。さらに音声認識により言葉の意味から身体動作を生成することができれば、使用者の感情をキャラクタに表現させるといった、より表現性の向上したコミュニケーションが実現できると考えられる。本システムのコンセプトを図 2 に示す。

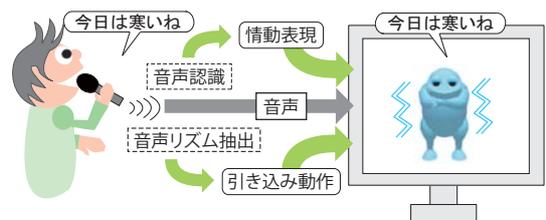


図 2 コンセプト

Fig. 2 Concept.

使用者の語りかけに対して、音声認識によって音声を文字列に変換し、その文字列から単語を抽出、単語の意

味にあった動作を生成することによって、InterActor に新たな動作を加え、感情や肯定・否定を表現する。従来の InterActor のうなずきなどの身体的引き込み動作に加え、キャラクタに肯定や否定、感情などの情動表現を追加することによって、キャラクタに自分の気持ちを表現させることができ、よりコミュニケーションを支援する。

4. 情動表現

InterActor に情動表現機能を付与するために、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する。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文字列を認識し、対応した意味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」のハートマークや「ひらめいた」の電球などの記号やシンボリックな表現と組み合わせた提示を行う。文字列に対応した動作はデータベース化しており、データベースには、352 種類の文字列とそれに対応した 105 種類の動作が登録されている。情動表現の例を図 3 に示す。

対応文字列352種類、動作パターン105種類			
文字列	やあ バイバイ	ごめん すみません	よし！ 頑張るぞ
動作			
文字列	好き かわいい	寒い 怖い	腹が立つ イライラする
動作			
文字列	ひらめいた 思いついた	やったー うれしい	悩む 参ったな
動作			

図 3 情動表現の例

Fig. 3 Example of affective expression.

5. システム開発

5.1 プロトタイプシステム

音声認識を併用することで、InterActor に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のプロトタイプシステムを開発した。音声認識には音声認識エンジン Julius を使用した^[7]。システムの使用画面を図 4 に示す。画面には、動作を行う InterActor と音声認識によって文字列に変換された音声

がウィンドウに表示される。マイクに話しかけている最中は、InterActor による話し手としての身体的引き込み動作を行い、言葉の切れ目に文中の単語によって情動表現を行う。

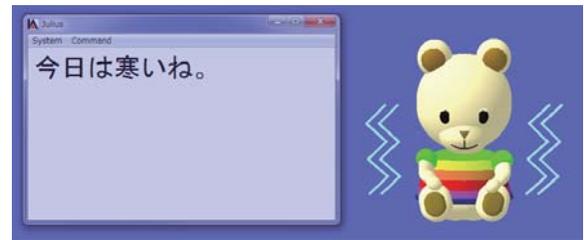


図 4 プロトタイプシステム使用画面

Fig. 4 Screen shot of the prototype.

5.2 コミュニケーションシステム

画面に自分と相手の情動表現機能を有する InterActor を表示し、通信するコミュニケーションシステムの開発を行った。システムの使用画面を図 5 に示す。画面には自分と相手のキャラクタが向かい合って配置されている。キャラクタはそれぞれ自身の話に対して、身振り手振りなどの話し手としての身体的引き込み動作と情動表現の動作を行い、相手の話に対して、うなずきなどの聞き手動作を行う。音声認識によって得られた文字列中に、対応する動作のある文字列が含まれている場合、図 5 のように、動作に加え、文字列を吹き出しによって表示する。情動表現を有する InterActor を用いることによって、音声のみの通話に比べ自分の思いをより伝えることができ、コミュニケーション支援に役立つと期待される。



図 5 通信システム使用画面

Fig. 5 Screen shot of the communication system.

6. おわりに

本研究では、従来の発話音声から身体動作を自動生成する InterActor に加え、音声認識を併用することで、InterActor に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提案し、プロトタイプを開発した。さらに、画面に自分と相手のキャラクタを表示し、通信を行うコミュニケーションシステムを開発した。これらのシステムの評価については今後の課題である。

謝辞 本研究の一部は科学研究費補助金 No.22300045 ,
No.24118707 の支援による .

参考文献

- [1] 尾形 正泰 , 大澤 博隆 , 篠沢 一彦 , 今井 倫太: Voicy:口
ポ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).
- [2] 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).
- [3] 鹿野 清宏 , Cincarek Tobias , 川波 弘道 , 西村 竜一 ,
李 晃伸: 音声情報案内システム「たけまるくん」および
「キタちゃん」の開発 (特別企画「音声認識デベロッパ
ズフォーラム」), 情報処理学会研究報告 , 2006-SLP-63,
pp.33-38, (2006).
- [4] 小林 宏: 超リアルな受付システムの開発 (ヒューマン・
エージェントインタラクション: 擬人化/顔インタフェ
ース再考), 情報処理学会研究報告 , 2004-SLP-52, pp.63,
(2004).
- [5] Tomio Watanabe: Human-entrained Embodied Interac-
tion and Communication Technology, Emotional Engi-
neering , Springer , pp.161-177, (2011).
- [6] 畔地耕太、渡辺富夫、山本倫也: 入力情報を同期表示
する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, ヒューマンインタフェースシンポジウム
2009DVD-ROM 論文集, pp.793-796, (2009).
- [7] 李 晃伸, 河原 達也: Julius を用いた音声認識インタフェ
ースの作成, ヒューマンインタフェース学会誌, Vol. 11, No.1,
pp.31-38, (2009).