

身体的引き込み動作と情動表現で思いを伝える メッセージ読み上げキャラクタシステムの開発

四方 拓^{1,a)} 小原 瑞希^{1,b)} 渡辺 富夫^{2,c)} 石井 裕^{2,d)}

概要：近年，テキスト情報を介したコミュニケーション技術が幅広く活用されている．しかし，これらのテキスト情報によるコミュニケーションでは，絵文字や顔文字といった表現の工夫がなされているものの，対話におけるうなずきや身振りなどの身体性が活かされていない．そこで，電子メールなどのリアルタイム性を求めないコミュニケーション場面を想定し，音声合成による合成音声から CG キャラクタに身振り等の引き込み動作，テキスト情報から言葉の意味に対応した情動表現を生成し，ユーザの思いを伝えるメッセージ読み上げキャラクタシステムを開発している．

Development of a Message Reading Character System for Conveying Feelings by Embodied Entrainment Motion and Emotional Expressions.

SHIKATA HIRAKU^{1,a)} KOHARA MIZUKI^{1,b)} WATANABE TOMIO^{2,c)} ISHII YUTAKA^{2,d)}

Abstract: In recent years, communication technology with text information is widely used. However, in communication with the text information, although contrivances such as pictograms and emoticons are used, body motions such as nod or gesture in dialogue are not used. In this paper, assuming the communication scene that does not ask for real-time such as e-mail, we develop a message reading character system for conveying user's feelings by embodied entrainment motion and emotional expressions by using text-to-speech synthesis.

1. はじめに

近年，遠隔コミュニケーションの手段の 1 つとして，Twitter[1] や LINE[2] といったテキスト情報を介したコミュニケーション技術が幅広く活用されている．一方，テキスト情報に音声合成を用いて，表現豊かにテキストの読み上げを行うエージェント等の研究開発も積極的に進められている [3][4][5]．

著者らは，音声リズムからロボットや 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ロポット技術 iRT を開発してきた．この 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 [6]．また，使用者の思いを表現するために，発話音声から身体的引き込み動作と音声認識を用いて情動表現を自動生成する音声駆動型身体的引き込みキャラクタを開発し，評価実験によりコミュニケーション支援の有効性を示している [7]．これらのシステムを電子メールなどのテキスト情報を介したリアルタイム性を必要としないシステムに導入することで，表現性が向上し思いが伝わるコミュニケーションが実現できると考えられる．

そこで本研究では，テキスト情報を介したリアルタイム性を必要としないコミュニケーション場面を想定し，音声

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) sikata0313@hint.cse.oka-pu.ac.jp

b) kohara_mizuki@hint.cse.oka-pu.ac.jp

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

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

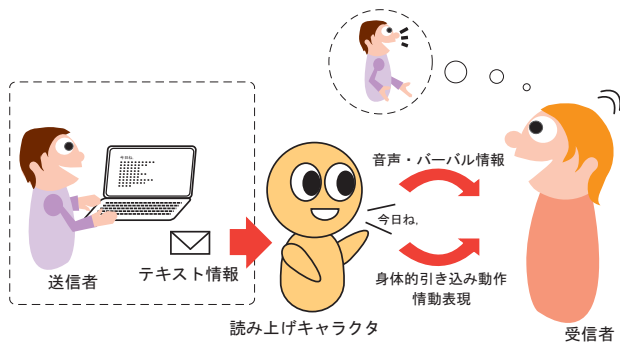


図 1 コンセプト
Fig. 1 Concept.

| | | | | |
|-----|----------------|-------------|---------------|------------|
| 文字列 | こんにちは すみません | やあ バイバイ | なるほど 思いついた | 寒い 凍える |
| 動作 | | | | |
| 文字列 | 悩む 照れる | いいえ 違います | 好き かわいい | 驚く びっくり |
| 動作 | | | | |

図 2 情動表現

Fig. 2 Examples of motions and emotional expressions.

合成を用いてテキストを音声に変換し、合成音声から CG キャラクタに身振り等の引き込み動作を生成するとともにテキスト情報から言葉の意味に対応した情動表現を自動生成し、思いを伝えるメッセージ読み上げキャラクタシステムを開発している。

2. コンセプト

本システムのコンセプトを図 1 に示す。テキスト情報を介したコミュニケーションでは、相手に文字情報のみで伝えるので、自己の思いや感情が伝わりにくい場合がある。そのため、絵文字や顔文字といった様々な表現の工夫がなされているものの、これらの表現はボタン選択やコマンド入力などの機能を用いたもので、表現の種類を多くすると操作性が低下する制約下での使用であった。そこで、CG キャラクタに対して、テキスト情報から言葉の意味に対応した情動表現、音声合成から得られた音声情報からは iRT を使用して発話音声から身体動作を自動生成して提示を行うことで、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。

3. システム開発

3.1 情動表現の提示方法

CG キャラクタに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する [8]。テキスト内にある単語を認識し、単語の意味に対応した情動表現をキャラクタに関連付けた。情動表現の例を図 2 に示す。「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やシンボリック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、挨拶などの日常でよく使われる単語と感情表現辞典 [9] を参考に単語を登録した。

3.2 情動表現の提示タイミング

先行研究の、あいさつインタラクションにおける動作に対する発声遅延の効果 [10][11] では、動作に対して発話が 300ms 程遅れることで、好ましいコミュニケーション効果が得られることが確認されている。本研究でも、合成音声の発話のタイミングと言葉の意味に対応した情動表現の提示タイミングを考慮することで、好ましい効果が得られることが期待される。そこで、情動表現の提示タイミングを自動調整するモデルを開発した。提示タイミングのモデルとして、音声の発話時間と発話音声の分節単位であるモーラ (mora) の関係から音声と情動表現のタイミングを同期させ提示を行うシステムを使用する。モーラとは、音韻論上、単一のリズムをなす発話音声の分節単位で、日本語の仮名 1 文字または長音「ー」、促音「っ」、撥音「ん」が 1 モーラとなる。ただし、「ゃ」「ゅ」「ょ」といった小書きの仮名は、その前の仮名と合わせて 1 モーラとなる (例えば、「チョコレート」では「チョ」「コ」「レ」「ー」「ト」で 5 モーラとなる)。本システムでは、初めにテキスト情報を合成音声に変換する。音声合成には Open JTalk を使用した [12]。次に、合成音声から音声の ON-OFF データを算出した。算出方法は、1/30 秒毎に音声の平均音圧が閾値を超えた場合を ON、そうでない場合を OFF とし、短い OFF 区間に対して、ON 区間に変換するハングオーバー処理を行った。また、あらかじめ合成音声による句読点や「?」「!」の記号などの無音時間を算出し、その値を参考に発話開始時から一定以上の無音時間を句読点等と判断した。さらに、句読点等の位置よりも前に存在している ON 区間を文章の発話時間とした。発話時間と文章のモーラ数の関係から、言葉の意味に対応した情動表現の提示タイミングを調整した。提示タイミングの自動調整モデルの概略図を図 3、提示タイミングの例を図 4 に示す。図 4 では、「すごく寒いです」の発話時間が 2400ms の場合、情動表現の提示タイミングは文章の開始時間の 900ms 後に動作を開始することを表している。

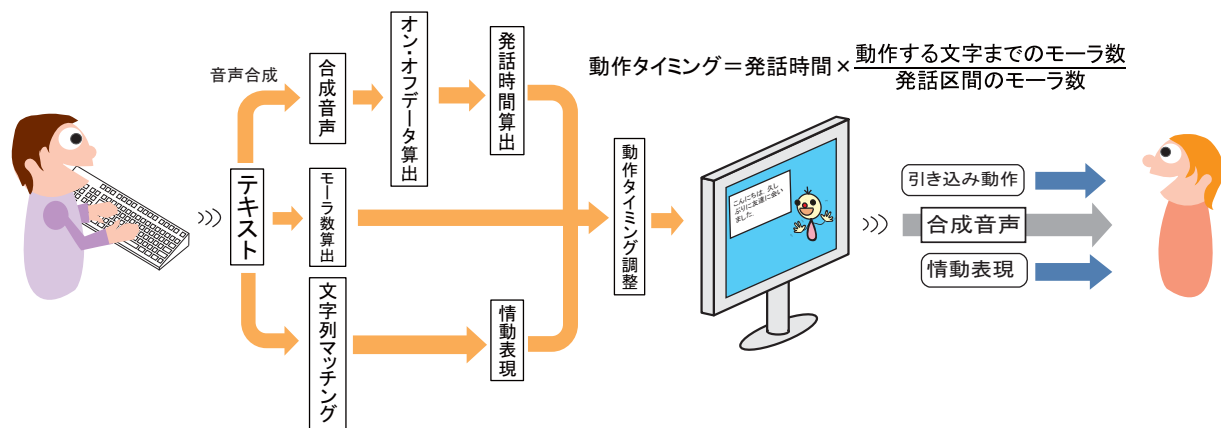


図 3 概略図

Fig. 3 Outline of the system.

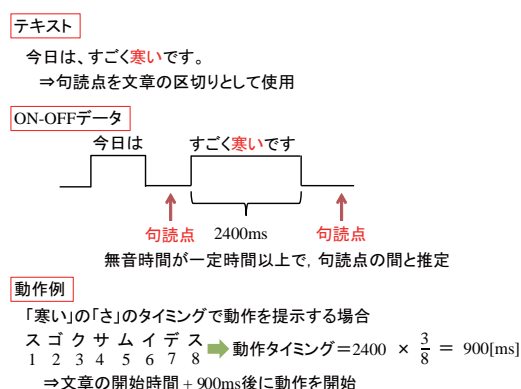


図 4 提示タイミングの例

Fig. 4 Examples of presentation timing.

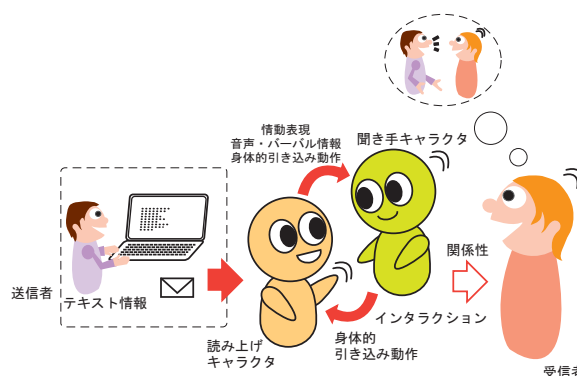


図 5 聞き手キャラクターを導入した提示

Fig. 5 Presentation in which a listener's character is introduced.

3.3 文の内容を考慮した情動表現の提示

従来の情動表現を行うシステムでは、テキストから言葉の意味に対応した情動表現を CG キャラクタに行わせており、例えば「寒い」という情動表現に対して、「寒い」「寒かった」「寒く」などデータベースに対応する情動表現の単語を複数用意する必要があった。また、「好き」という動作に対して、「好きではない」という言葉の場合も、情動表現を行ってしまい、文章によっては相手に間違った感情や思いを伝えてしまう可能性があった。そこで、文章の内容を考慮した新たに形態素解析を用いるモデルを提案する。形態素解析には MeCab を使用した [13]。形態素解析とは、文章を意味のある単語に区切り、データベースを利用して品詞や内容の判別を行うことである。例えば「寒かった」の場合「寒い」という形容詞と「た」という助動詞に分割され、動詞や形容詞の基本形の単語をデータベースに用意することで、情動表現が可能となる。また「好きではない」では、「好き」「で」「は」「ない」と分割され、「ない」という打消しの助動詞が使われている。そこで、情動表現を行う単語が存在し、その文章の区切り(句読点や「?」や「!」などの記号を区切りとした)までに「ない」「ません」といった打消しを意味する単語が存在すれば、否定文と判

断し動作をしないようにした。また、形態素解析を行うにあたって、従来のデータベースを改良した。現在、1000 種類以上の単語とそれに対応した 110 種類の情動表現が登録されている。このモデルを用いることで、より正確に使用者の思いを伝えることができる。

4. システム展開

今回開発したシステムでは、CG キャラクタを自己の代役として、キャラクタが受信者にメッセージを伝える手法としてシステム開発をしてきた。一方、対面コミュニケーションにおける音声を経たコミュニケーションにおいては、話し手に対する聞き手の反応などによって、対話者同士が互いにかかり合うことでコミュニケーションを行っている。そこで本システムにおいても、自己の代役だけでなく受信者の代役として聞き手の CG キャラクタを配置した提示を提案する。提案する提示方法を図 5 に示す。自己の代役となる CG キャラクタには話し手として今回開発したシステムを適用し、受信者の代役となる CG キャラクタには iRT を使用し、聞き手としてうなずきなどの身体動作を提示する。CG キャラクタ同士がかかり合うことで、受信者が一体感・共有感と送信者の思いを実感しながら

らメッセージを読む効果が期待できる。

5. おわりに

本研究では、テキスト情報から音声合成を用いて CG キャラクタに身振り等の引き込み動作とテキスト情報から言葉の意味に対応した情動表現を自動生成し、思いを伝えるメッセージ読み上げキャラクタシステムを開発した。また、提示方法として提示のタイミングや文の内容を考慮したモデルの開発も行った。さらに、システム展開として、聞き手キャラクタを導入した提示方法を提案した。これらのシステムの評価については今後の課題である。

参考文献

- [1] Twitter: <https://twitter.com/>, (2014/12/18 アクセス).
- [2] LINE: <http://line.me/ja/>, (2014/12/18 アクセス).
- [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ーズフォーラム」), 情報処理学会研究報告, 2006-SLP-63, pp.33-38, (2006).
- [4] 矢野美由紀, 吉田直人, 米澤朋子: ユーザのリアクションに応じたニュース読み上げアナウンサー, HAI シンポジウム 2013, pp.102-105 (2013).
- [5] 荒堀拓哉, 片上大輔, 角所考: 情報提供アナウンサーエージェントにおける表出モデル設計と印象評価 ~ スマートキャスト 24(TWENTY FOUR) の開発 ~, 人工知能学会論文誌, Vol.28, No.2, pp.179-186 (2013).
- [6] Tomio Watanabe: Human-entrained Embodied Interaction and Communication Technology, Emotional Engineering, Springer, pp.161-177, (2011).
- [7] 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 (SI 2013) 論文集, pp.1945-1948, (2003).
- [8] 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).
- [9] 中村明: 感情表現辞典, 東京堂出版, 1993.
- [10] 山本 倫也, 渡辺 富夫: ロボットとのあいさつインタラクションにおける動作に対する発声遅延の効果, ヒューマンインタフェース学会論文誌, Vol.6, No.3, pp.87-94, (2004).
- [11] Michiya Yamamoto and Tomio Watanabe: Time Lag Effects of Utterance to Communicative Actions on CG Character-Human Greeting Interaction, Proc. of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06), pp.629-634, (2006).
- [12] Open JTalk: <http://open-jtalk.sourceforge.net/>, (2014/12/18 アクセス).
- [13] MeCab: <http://mecab.sourceforge.net/>, (2014/12/18 アクセス).