発話内単語の感情極性に基づき反応動作を行う
身体的引き込みキャラクタシステム西田 麻希子^{*1,*2}, 太田 靖宏^{*1}, 渡辺 富夫^{*3}, 石井 裕^{*4}Embodied entrainment character system with automatically generated response motions
based on the semantic orientations of words in utteranceMakiko NISHIDA^{*1,*2}, Yasuhiro OOTA^{*1}, Tomio WATANABE^{*3} and Yutaka ISHII^{*4}^{*1} Graduate School of Computer Science and Systems Engineering, Okayama Prefectural University

111 Kuboki, Soja-shi, Okayama 719-1197, Japan

^{*2} Faculty of Design, Okayama Prefectural University

111 Kuboki, Soja-shi, Okayama 719-1197, Japan

^{*3,*4} Faculty of Computer Science and Systems Engineering, Okayama Prefectural University

111 Kuboki, Soja-shi, Okayama 719-1197, Japan

Received: 31 March 2016; Revised: 25 August 2016; Accepted: 28 December 2016

Abstract

The prior research includes the development of a speech-driven embodied entrainment computer-generated character called InterActor, which automatically generates communicative motions and actions such as nods for entrained interaction from voice rhythm based on only speech input. Because the conventional InterActor character generates only positive actions from the verbal content but no negative actions, it is possible to aggravate a speaker's negative emotions by performing positive gestures in response to negative verbal content. In this paper, we develop an advanced speech-driven embodied entrainment character system that can respond to and improve the speaker's emotional state using speech recognition. In this system, the speaker's words are converted to text by speech recognition; their emotions are then estimated from character strings in the converted text. The system uses a database that quantifies each word and estimates the emotion associated with it. Then, the system automatically generates negative/positive motions based on the semantic orientations of words in utterance as well as entrained motions. Furthermore, we demonstrate the effectiveness of the system through three experiments: two role-play experiments for one user involving positive/negative scenarios, and a communication experiment for two remote users using the developed system.

Key words : Embodied interaction, Entrainment, Communicative agent, Speech recognition, Emotion

1. は じ め に

今日音声認識は、スマートフォンなどの携帯端末に導入され、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、様々な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が提案されている (Cassell, 2000, 川本他, 2002)。さらに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も提案されている (鹿野他, 2006, 橋本他, 2007)。これら音声認識を活用した自動応答システムは言語情報に基づくもので、音声リズムなど非言語情報に基づいた応答

は十分に検討されていない。一方人の対面コミュニケーションでは、言葉だけでなく、音声に対するうなずきや身振り・手振りが相互に同期することで、自然なコミュニケーションが行われている。母親の語りかけに乳児が手足を動かして応えるなど、本来別々の身体的リズムで行われている行動が相互に時系列的に関係が成立して同期する現象は引き込み現象と呼ばれている (Condon and Sander, 1974)。従来より対話者の身体性の共有に着目し、同期現象を創出する共有仮想空間を活用したシステム開発も行われており (Watanabe and Miwa, 2012)、人とのインタラクションを行う対話エージェントの研究開発においても身体的なかかわりはインタラクションに重要な役割を果たすと考えられる。また身体動作や情動表出による社会的相互作用に着目し、人やロボットとのインタラクションに応用する試みも研究が進められ (ペントランド, 2013, Woo, 2015)、使用者の感情推定に基づく応答動作を行うことで、より深くかかわりを実感できる対話エージェントの開発が期待されている。

著者らは、これまでに会話音声と身体動作が同期する身体的引き込みに着目し、インタロボット技術 iRT (InterRobot Technology) を開発してきた (Watanabe, 2011)。iRT は、音声と身体動作の関係をモデル化することで発話音声からコミュニケーション動作を自動生成し、身体的引き込みによりインタラクションを円滑にし、コミュニケーションを支援するものである。すでに、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 に実装することで、システムの有効性を示している (Watanabe et al., 2004)。iRT は発話内容に依存せず、発声の有無 (ON-OFF) に着目したもので、身体性の共有に特化して開発、展開されてきた。一方で、従来の InterActor は、iRT に基づいて音声リズムからうなずきなどの身体的引き込み動作を自動生成し、どのような発話内容に対しても引き込み動作を行うものの、使用者の感情によって動作を変化させる機能は実装されていなかった。そのため、“学校・仕事に行きたくない”や“何もしたくない”のような否定的な発話内容の場合においても、リズム同調的に自動生成されたうなずき動作を行うため、使用者がうなずきを肯定動作と解釈した場合、その否定的な発話感情を助長してしまう。

そこで本論文では、発声の有無のみに基づいた従来の InterActor の身体的引き込み動作に加えて、音声認識を導入して発話内容から使用者の感情を推定し、結果に基づいて反応動作を変化させることで、使用者の否定的な発話を抑制し肯定的な内容へ誘導、または肯定的な状態を保持することができ、会話意欲を促進する身体的引き込みキャラクタシステムを開発した。また開発したシステムを用いて、肯定的及び否定的なシナリオに基づくロールプレイによる自動応答エージェントシステムとしての評価実験とともに、2人1組での自由対話によるコミュニケーションインタフェースシステムとしての評価実験を行い、コミュニケーション支援への有効性を示した。

2. 発話感情表現に基づき反応動作を行う身体的引き込みキャラクタ

2.1 コンセプト

開発したシステムのコンセプトを図1に示す。使用者が対話エージェントとしてのCGキャラクタに話しかけると、キャラクタはうなずきなどの身体的引き込み動作を行う。うなずきは肯定的な意味を連想させるため、使用者が負の感情を含んだ発話を行った場合、使用者の負の感情が肯定されてしまい、さらなる負のイメージを使用者に与えてしまう可能性がある。そこで、従来の InterActor の身体的引き込み動作に加えて音声認識を併用して、発話文の単語が一般的に良い印象を持つか、それとも悪い印象を持つかを表した二値属性である感情極性から使用者の感情を推定し、それに対応した反応動作を生成する。使用者の否定的な発言に対しては、うなずくのではなく負の感情を緩和させる否定的な動作を行うことで、リズム同調を保ちつつ負の感情の抑制を促す。また肯定的な発言に対しては、正の感情を盛り上げる動作を行うことで正の感情がより促進される。これらの機能を用いることで、使用者の会話意欲のさらなる促進が可能となる。

2.2 音声駆動型身体的引き込みキャラクタ InterActor

従来の音声駆動型身体的引き込みキャラクタ InterActor は、基本性能として音声から自動生成される身体動作による話し手と聞き手の両機能を備えている。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自動的に反応し、話し手として口の開閉や身体動作など話者の発話音声に同期した動作をすることで、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。

Fig. 1 Concept of an advanced speech-driven embodied entrainment character system.

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 2, 檀原他, 2005). 音声データ $V(i)$ は 16bit 22.050kHz でサンプリングし、閾値で二値化するとともに、音節間の短時間の無音区間による発話の断片化を除去するために 133msec でハングオーバー処理を施している. このモデルでは、マクロ層とミクロ層からなる階層モデルを用いて、InterActor のうなずきの予測を行っている. 予測にはユニット以前のユニット時間率 (ユニット区間での ON 区間の占める割合、式 (2)) の線形結合で表せる式 (1) 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

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

$b(j)$: 予測係数

$V(i)$: 音声データ

$w(i)$: 雑音

Fig. 2 Listener's interaction model for speaker's voice.

予測値 $Mu(i)$ が閾値 0.20 を超えて、うなずきが存在すると予測された場合は、処理はミクロ層に移る. ミクロ層では音声の 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を推定する. 予測値が閾値 0.47 を超えた場合には 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をさせ、それを起点にして次の瞬きまでの時間間隔を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、胴部、右肘、左肘)のうちいくつかを選択して動作させることで話し手の発話音声と関連付けている。

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すべての動きの ON-OFF の総和データから体の動くタイミングを予測し、閾値を超えたときに InterActor の各部位(頭部、胴部、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連付けている。

2.3 発話感情推定方法

本研究では、使用者の発話感情推定に高村らの提案する単語感情極性対応表(高村他, 2006)を用い、発話文中の単語から正負の感情推定を行った。人間の感情区分としてはさらに細分可能であるが(Ekman et al., 1982, Parrott, 2001), 本研究では発話のみから推定を行うため感情極性に着目している。単語の感情極性とは、その単語が一般的に良い印象を持つか、それとも悪い印象を持つかを表した二値属性である(以下、感情極性対応表に基づく正の値を Positive, 負の値を Negative と表記)。例えば「良い」「美しい」などは Positive な極性、「悪い」「汚い」などは Negative な極性を持つ。感情極性値は、語彙ネットワークを利用して自動的に計算されたものであり、-1 から +1 の実数値を割り当てている。+1 に近いほど Positive であり、-1 に近いほど Negative である。対応表には約 5 万単語が登録されているが、本システムでは極性値が 0.7 以上の単語及び -0.7 以下の単語(8351 語)を採用し、登録された単語に紐づけられた極性値に基づいて、発話文の極性値を推定した。また、対応表には単語の基本形のみが登録されているため、例えば「悲しい」という表現に対し、「悲しかった」「悲しくなる」などといった活用の場合ではマッチングできない。そこで、文章の内容を考慮するためにオープンソースの形態素解析システム MeCab を用いて品詞や内容を判別する形態素解析を行った。これにより、「悲しかった」という活用の場合にも、「悲しい」という形容詞と「た」という助動詞に分割されるため、対応表に登録している「悲しい」という表現にもマッチングが可能となる。取得した単語情報と対応表にある単語を照合し、対応表に登録されている単語の場合には、その単語に紐づけられた値を参照する。発話文中の単語の極性値 x_i ($-1 < x_i < 1$)、およびその単語数 n を用い、以下の式(4)から発話文の感情極性値を算出した。「美しくない(美しい/ない)」といった否定語を含む場合は、「ない」の直前の単語の極性値は参照しない。

$$\text{感情極性値} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

x_i : 単語の極性値, n : 単語数

2.4 システム構築

音声認識エンジン Julius(李, 河原, 2009)により、発話音声をテキストに変換し、テキスト内の単語情報から感情極性を推定して、その推定結果に応じキャラクタの動作を変化させる。推定結果が Positive の閾値を上回った場合に、正の感情を盛り上げる動作、Negative の閾値を下回った場合に、負の感情を緩和させる動作をキャラクタの身体動作に関連付けた。Positive および Negative の閾値に関しては、予備検討として「すごく楽しかったエピソード」「自分のことが嫌になってしまったエピソード」というテーマのもと、キャラクタに対し話しかけるような口語口調で被験者に文章を記述させ(平仮名 400 文字程度)、そこから被験者毎の各感情値の平均を算出した。さらに、各個人の平均値から全体の平均値を算出することで、Positive/Negative の閾値を決定した(表 1)。音声情報を用いて全ての応答動作を自動生成するため、マイクと PC のみでシステムを構成できる。Windows 7 32bit PC (HP Core i5 CPU, 2.67GHz, 4GB メモリ), DirectX 及び Visual Studio で開発したプロトタイプ画面例を図 3 に示す。

Table 1 Results of Negative/Positive values of recorded past event by subjects.

Subject	1	2	3	4	5	6	7	8	9	10	Total: threshold
Negative	-0.58	-0.75	-0.92	-0.50	-0.92	-0.99	-0.73	-0.25	-0.99	-0.92	-0.76
Positive	0.56	0.86	0.85	0.65	0.97	0.58	0.58	0.99	0.65	0.52	0.72

Fig. 3 Example of communication scene using the system.

2.5 同調動作・緩和動作提示確率の決定

2.5.1 同調動作・緩和動作提示のための動作評価実験

発話文から推定した感情極性値に基づいて行う正の感情に対する強い同調及び負の感情に対する緩和動作として、多様な表現を行うために、動作をそれぞれ4パターンずつ用意した。強い同調動作として、A：連続うなずき＋笑顔，B：深いうなずき＋笑顔，C：うなずき＋拍手＋笑顔（笑い），D：うなずき＋笑顔，緩和動作としてE：手と頭を左右に振る（否定），F：腕を組む＋首をかしげる（疑問），G：口を開閉＋腕を上下（宥める），H：手を横に出す（ツッコミ）動作である（図4，図5）。本システムにおけるキャラクタが行う各動作の提示確率を検討するために、予備検討として同調動作，緩和動作それぞれについて一対比較実験を行った。各4パターンの動作をキャラクタシステムに付加したものをを用い、被験者に対し、各4パターンの動作の中から2つ選び、実際にシステムを使用させた上で、それぞれ強い同調動作・緩和動作として好ましいものを選択させた。発話内容として、強い同調動作の際は過去の肯定的な出来事，緩和動作の際は過去の否定的な出来事を思い出しながら1分間自由に発話させた。これを1人に対して各6通り（ $4C_2$ ）行った。モードの提示順序はカウンターバランスをとり、評価への影響を排除するために、総当たりにより提示順序を変更した。被験者は前節と同じ対象者で、同じ研究グループに属する22～24歳のボランティアの学生10名（男性8名，女性2名）である。否定的な発話内容が聴取あるいは記録されることの影響を排除するため、実験の様子は記録しなかった。

Pattern	Initial state	A : Smile + Multiple nod	B : Smile + Deep nod	C : Smile + Clap + Multiple nod	D : Smile + Nod
Motion					
Time [msec]	—	500	500	500	500
Range [rad]	—	Head : x 0.35	Head : x 0.56	Head : x 0.35 Upper arm : x 1.0, y 1.0, z 0.1 Forearm : z 1.0	Head : x 0.35

Fig. 4 Positive motions for a listener character.

Pattern	Initial state	E : Denial	F : Tilt head	G : Soothe	H : Stop joking
Motion
Time [msec]	—	500	500	1000	500
Range [rad]	—	Head : x 0.4 Upper arm : x 1.4, y 0.5 Forearm : 0.8	Head : x 0.1, y 0.1	Upper arm : x 1.0, y 0.5 Forearm : x 1.0, z 0.8	Head : y 0.2 Upper arm : x 1.0, y 1.0

Fig. 5 Negative motions for a listener character.

2.5.2 実験結果とシステム適用

各一対比較の結果を表 2, 3 に示す. 一対比較による評価を一義的に定めるために, Bradley-Terry モデル ($P_{ij} = \pi_i / (\pi_i + \pi_j)$, $\sum \pi_i = \text{const.} (= 100)$, $\pi_i : i$ の強さ, $P_{ij} : i$ が j に勝つ確率) を想定し, モードの強さ π を最尤推定した結果を図 6, 7 に示す. 強い同調動作は, A の連続うなずき+笑顔の動作, C のうなずき+拍手+笑顔(笑い)の動作, B の深いうなずき+笑顔の動作, D のうなずき+笑顔の動作の順で好まれる結果となった. また, 緩和動作は, G の口を開閉+腕を上下(宥める)動作, E の手と頭を左右に振る(否定)動作, H の手を横に出す(ツッコミ)動作, F の腕を組む+首をかしげる(疑問)動作の順で好まれる結果となった.

Table 2 Result of paired comparison for positive motions.

	A	B	C	D	Total
A		6	5	8	19
B	4		4	5	13
C	5	6		7	18
D	2	5	3		10

Fig. 6 Comparison of π for positive motions.

Table 3 Result of paired comparison for negative motions.

	E	F	G	H	Total
E		7	3	9	19
F	3		2	4	9
G	7	8		6	21
H	1	6	4		11

Fig. 7 Comparison of π for negative motions.

本研究ではこの実験結果から得られた各動作の好みの強さ π を確率分布とし, それぞれの強い同調動作・緩和動作の選択確率とした. この選択確率に従って, 身体的引き込み反応に加えて, 感情極性推定から算出した感情極性値が予め設定した Positive の閾値を上回った, あるいは Negative の閾値を下回った場合に, キャラクタが強い同調動作, あるいは緩和動作を行うキャラクタシステムを開発した. 開発したシステムにおいて, 音声入力終了してから音声認識された結果が表示されるまでの時間は, 動画による評価で約 167msec 程度であり, またその時点からキャラクタ動作が開始されるまでに約 99msec 程度で, 合わせて約 267msec 程度の時間遅れが生じている. 従来の InterActor は二値化した音声入力について約 133msec のハングオーバー処理を行っているため, 従来

のシステムとの時間遅れの差は約 133msec 程度となる。先行研究によって CG アバタを介したコミュニケーションにおいて、音声通話における遅延については 400 msec までが推奨され(伊藤, 北脇, 1987), また 500msec 程度の通信遅延では評価に影響を与えないことが明らかになっており(石井他, 2010), 本研究の時間遅れについて大きな影響はないと考えられる。

3. シナリオに基づくキャラクタへの語りかけによる評価実験

本研究では開発したシステムの評価として, Negative 及び Positive と判定されたシナリオに基づくロールプレイによる評価実験, 及び 2 人 1 組での自由対話による評価実験を行った。本章でのロールプレイによる評価実験では, 正負の感情に対して適切な動作表現ができているかの観点から, あらかじめ用意したシナリオに基づいて, キャラクタへの語りかけによって評価した。とくに Negative なシナリオについては, 本研究で課題とした使用者の負の感情を助長しない観点から, 負の感情を抑制する動作表現に加え, 比較対象としてあえてさらに同調を促す動作表現を用いてシステムを評価している。次章ではキャラクタを対話者それぞれのアバタとして扱い, 2 者間コミュニケーションによってシステムを評価した。音声認識による感情極性に対応した動作は, 相手キャラクタの聞き手動作として反映される。

3.1 Negative なシナリオに基づく評価実験

3.1.1 実験手順

提案手法を用いたキャラクタシステムの評価として, まず Negative な内容のシナリオ(図 8)に基づくキャラクタへの語りかけによる評価実験を行った(図 9)。本実験では, 表 4 に示すキャラクタの動作が異なる N1~N3 の 3 つのモードを用意した。N1 モードでは, 従来の InterActor が行ううなずきや身振りなどといった聞き手としての動作を行う。N2 モードでは, N1 モードの聞き手としての動作に加え, 本研究で開発した提案手法による感情推定を行い, 推定結果が Negative の閾値を下回った場合にキャラクタが負の感情に対し強い同調動作を行う。N3 モードでは, N1 モードの聞き手としての動作に加え, 本研究で開発した提案手法による感情推定を行い, 推定結果が Negative の閾値を下回った場合にキャラクタが負の感情を緩和させる動作を行う。

女性 21 歳 大学3年生
家族構成:父親 51 歳 母親 48 歳 兄 25 歳の 4 人暮らし

夏季休暇も終了し、前期の成績の結果が返却された。必修科目については無事取得出来ており問題なかったが、選択科目をいくつか落としてしまっていた。現状、進級に必要な単位のギリギリで履修登録を行っていたため、後期の履修可能分を全て取得出来なければ進級が危うく、留年の危機が迫っている。
バイトや部活などもあるため忙しく、かなりの過密スケジュールになると予想されるため、このままでは後期の単位取得も難しくなるのではないかと不安に駆られている。そのため、ここ最近は寝つきが悪く、良質な睡眠がとれていない。来年度には 4 年に進級し、卒業研究に加え就職活動を始めなければならないため、これから先のことについて考えていると不安が募り、気分が滅入っている。
現在は両親などと話し合いを行いつつ、今後のことについて考えている。昨晩は精神的不安などからあまり食欲もわかず、あまり眠れなかったため、気分が落ち込んだまま 1 日を過ごしている。

場面:日曜の午後 10 時、自宅の自分の部屋。来週の月曜から後期の講義が開始となる。食後、何をする様子もなく机に向かっている。

Fig. 8 Example of negative scenario for female subject.

Table 4 Communication modes in the negative scenario experiment.

Mode	Explanation
N1	Conventional InterActor
N2	InterActor with positive motions for negative utterance
N3	InterActor with negative motions for negative utterance

Fig. 9 Example of communication scene in the speaking experiment.

最初に、システムを十分理解した上で評価を行うために、各モードの違いと操作方法を説明した後、実際にシステムを試用させた。説明にあたっては、実験者が統制された説明を行うことができるように、実験に関する運用手順書を作成した。モードの切り替えや操作は、観察者がマイクとスピーカにより指示した。被験者にはアンケート用紙をあらかじめ配布し、評価項目ごとに記入させた。ただし、各モードを使って語りかけている途中は、実験に集中させるためアンケート用紙に触らないよう指示した。

実験は、最初に N1～N3 の 3 つのモードのうち 2 つを提示し、一対比較法により各モードの比較を行わせた。参加者には総合的に良い方を選択させた。提示場面を変更し、3 回 ($3C_2$) 繰り返した。モードの提示順序はカウンターバランスをとり、評価への影響を排除するために、総当たりにより提示順序を変更した。次に、各モードに対し身体的コミュニケーション支援の観点から定めた 6 項目について 7 段階官能評価 (中立 0) を行った。6 項目は、「楽しさ」「好み」「対話しやすさ (話しやすさ)」「安心感」「和み」「システムを使用したいか」とした。被験者に提示するシナリオは、年齢や家族構成を始め、背景や場面設定が記されているものを使用した。被験者は 19～24 歳の学生 24 名 (男性 12 名、女性 12 名) で、被験者には 500 円の図書カードを謝礼として渡した。

3.1.2 実験結果

一対比較結果を表 5 に示す。この結果を定量的に評価するために Bradley-Terry モデルを想定し、強さ π を最尤推定した結果を図 10 に示す。N3 の「聞き手としての動作、および負の感情を緩和させる動作」のモードが突出して高く評価されており、続いて N1 の「うなずきや身振り手振りなどの聞き手動作」のモード、N2 の「聞き手としての動作、および負の感情に対する強い同調動作」のモードの順に評価されている。

次に、7 段階官能評価の結果について、各項目の平均及び標準偏差を図 11 に示す。3 条件間の比較として Friedman 検定を実施した結果、すべての項目において優位水準 5% で有意差が認められた。さらに多重比較として Wilcoxon の符号順位検定を行った結果、N3 モードと N1 モードの間では、「安心感」の項目において優位水準 0.1% で有意差が認められ、「楽しさ」「好み」「和み」の項目において有意水準 1% で有意差が認められ、「対話しやすさ (話しやすさ)」「システムを使用したいか」の項目において有意水準 5% で有意差が認められた。また、N1 モードと N2 モードの間では、「好み」の項目において有意水準 0.1% で有意差が認められ、「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、「対話しやすさ (話しやすさ)」「安心感」の項目において有意水準 5% で有意差が認められた。さらに、N3 モードと N2 モードの間では、「好み」「対話しやすさ (話しやすさ)」「安心感」「システムを使用したいか」の項目において有意水準 0.1% で有意差が認められ、「和み」の項目において有意水準 1% で有意差が認められた。

3.1.3 考察

実験の結果、一対比較では N3 の「聞き手としての動作、および負の感情を緩和させる動作」のモードが突出して高く評価された。また、7 段階官能評価の結果において、N1 モードに対し、提案手法を用いた N3 モードで「楽しさ」「好み」「安心感」「和み」の項目で有意差が確認された。また N2 モードに対し、提案手法を用いた N3 モードで、「好み」「対話しやすさ (話しやすさ)」「安心感」「和み」「システムを使用したいか」の項目で有意差が確認された。また、「聞き手としての動作、および負の感情を緩和させる動作」の N3 モードに関する自由記述では「しんどいときにキャラクタが「そうじゃない」「大丈夫」と動作してくれたのがよかった」や「勇気づけてく

Table 5 Result of paired comparison in the negative scenario experiment.

	N1	N2	N3	Total
N1		14	1	15
N2	10		2	12
N3	23	22		45

Fig. 10 Preference π based on the Bradley-Terry model.

Fig. 11 Results of seven points bipolar rating in the negative scenario experiment.

れている感じがして元気になれるそうだった」といったコメントがあり、N3 モードが好意的に受け入れられていることが確認された。一方で N2 モードでは「N2 は本当につらいときに笑顔というか笑っていたらちょっと嫌かも」や「N2 はつらい気持ちなのに、それを打ち消すように笑ったりして、元気になる時もあるかもしれないけど、自分の気持ちを否定されているように感じた」などといったコメントが見られた。これらのことから、Negative な内容の発話を行った際に、それを否定、打ち消すことで発話者に対し楽しさ、和みなどを与えることができ、負の感情で発話された内容に対する反応として、好ましいと判断されたことがわかる。さらに、7 段階官能評価において N2 モードが N1 モードに対し低く評価されていることから、使用者の負の感情に対して適切な同調・緩和動作を用いることが必要であるといえる。

3.2 Positive なシナリオに基づく評価実験

次に Positive な内容のシナリオ (図 12) に基づきキャラクタへの語りかけによる評価実験を行った。実験環境および観察者からの指示等は前節と同様である。前節の実験結果より、負の感情に対しては適切な同調・緩和動作を用いることが必要であることが確認された。本実験では、正の感情に対する強い同調動作の効果を検討するため P1, P2 の 2 つのモードを用意した。P1 モードでは、従来の InterActor が行ううなずきや身振り手振りなどの聞き手としての動作を行う。P2 モードでは、P1 モードの聞き手としての動作に加え、本研究で開発した提案手法による感情推定を行い、推定結果が Positive の閾値を上回った場合に、キャラクタが正の感情に対し強い同調動作を行う。

実験はまず、P1, P2 の 2 つのモード用いて各モード 1 分間ずつ使用させた後、一対比較によって総合的に良かった方を選択させた。次に、各モードに対し身体的コミュニケーション支援の観点から定めた前節と同じ 6 項目について 7 段階官能評価 (中立 0) を行った。被験者は 20~24 歳の学生 24 名 (男性 12 名, 女性 12 名) で、被験者には 500 円の図書カードを謝礼として渡した。

3.2.1 実験結果

一対比較の結果、被験者 24 名全員が P2 の「聞き手としての動作、および正の感情に対する強い同調動作」のモードを選択した。また 7 段階官能評価の結果について各項目の平均及び標準偏差を図 13 に示す。Wilcoxon の符号順位検定を行った結果、P2 モードと P1 モードの間で全ての項目において有意水準 0.1% で有意差が認められた。

女性 21 歳 大学3年生
 家族構成:父親 51 歳 母親 48 歳 兄 25 歳の 4 人暮らし

私の誕生日パーティーが開かれた。友人たちがサプライズで企画してくれていたようだった。いつも通り一日の講義を受け終え自宅に帰ろうとすると、友人からある場所に来てほしいとの連絡があった。何も知らされていなかったため疑問に思いつつ一人で指定された部屋に入ると、綺麗に装飾された部屋に食事などがたくさん用意されており、友人たちが私を迎え入れてくれた。驚きとともに、嬉しさがこみ上げてきた。おいしい食事をたくさん食べることが出来たり、友人たちが余興を行ってくれたり、さらにはプレゼントまで用意してくれるなど、申し訳ないほどもてなしを受けた。非常に楽しい会であったため、これ以上ない幸せな気分である。最近では学業やバイト、部活などで忙しく、みんなで集まって遊ぶことなどが出来なかった。そのため、今回の誕生日パーティーではみんなが揃い有意義な時間を過ごすことで心身ともにリフレッシュすることが出来た。現在は片付けを終え、みんなからもらったプレゼントを開封しながら誕生日パーティーの余韻に浸っている。そのため、気分が高まった状態で過ごしている。

場面:誕生日パーティーの片づけを終え自宅に帰宅。自分の部屋でのんびりとしている。

Fig. 12 Example of positive scenario for female subject.

Fig. 13 Results of seven points bipolar rating in the positive scenario experiment. InterActor with positive motions (P2) was affirmatively evaluated in comparison with conventional InterActor (P1).

3.2.2 考察

P2 モードは一対比較、7段階評価ともに高く評価された。また P2 モードに関する自由記述では「羊がニコニコしてくれるとこちらも楽しくなって話したいなと思った」や「笑ってくれたりした方が反応が感じられてこちらも話しやすかった」といったコメントがあり、P2 モードが好意的に受け入れられていることが確認された。また、P1 モードに関する自由記述では「P1 モードは悩みや真剣に話を聞いてほしいときには話しやすいが楽しいときには話しづらかった」や「忙しい話をしているときは P1 モードもいいと思った」など状況に応じたキャラクタの動作提示が必要であるということが示唆された。これらの結果より、Positive な内容の発話を行った際に、それに同調し盛り上げることで、発話者に対して楽しさ、話しやすさなどを与えることができ、正の感情を盛り上げるための好ましい効果が得られたと考えられる。

4. 自由対話によるシステム評価実験

4.1 実験手順

前章の実験は、Negative/Positive それぞれの状況下にあることを限定した上でのキャラクタに対する語りかけによる実験であった。実際にシステムを使用した際の評価を行うためには、一般的な対話状態における発話者の感情推定の結果に応じて、同調・緩和動作を行うキャラクタシステムを用いたコミュニケーション実験を行う必要がある。そこでキャラクタシステムに通信機能を導入することにより、遠隔で 2 者対話を実現するコミュニケーションシステムを開発した (図 14, 15)。2.4 節で構築したシステム構成を用いて、2 台の PC を 1Gbps Ethernet で接続し、音声を送受信している。

システムの評価として、2 人 1 組での自由対話による評価実験を行った。実験風景を図 16 に示す。本実験では、次に示すキャラクタの動作が異なる α 、 β の 2 つのモードを用意した。 α モードでは、発話時の身振り手振りな

Fig. 14 Schematic diagram in the communication system configuration. Fig. 15 Screen shot of communication scene between remote users.

ど話し手動作に加え、うなずきなどの聞き手としての動作を行う。βモードでは、αモードの従来動作に加え、本研究で開発した提案手法による感情推定を行うもので、推定結果が Positive の閾値を上回った場合にキャラクターが正の感情に対し強い同調動作を行い、推定結果が Negative の閾値を下回った場合にキャラクターが負の感情を緩和させる動作を行う。

被験者は 19～22 歳の男女学生で、同性同士の友人関係である者を 2 人 1 組とした 12 組 24 名 (男性 6 組, 女性 6 組) である。被験者には 500 円の図書カードを謝礼として渡した。実験は 2 つの部屋を用い、それぞれ別の個室に分かれてシステムを用いて日常の談話をさせた。最初に、各モードの違いと操作方法を説明した後、実際にシステムを数分間試用させた上で実験を行った。

実験はまず、α, β の 2 つのモードを各 3 分間ずつ使用させた後、一対比較によって総合的に良かった方を選択させた。次に、各モードに対し身体的コミュニケーション支援の観点から定めた前章と同じ 6 項目について 7 段階官能評価 (中立 0) を行った。

Fig. 16 Example of the experimental scene under free conversation.

Fig. 17 Result of paired comparison in the communication experiment under free conversation.

4.2 実験結果

一対比較の結果を図 17 に示す。βモードの「聞き手としての動作、および正の感情に対する強い同調動作、および負の感情を緩和させる動作」のモードが高く評価されていることが分かる。次に、7 段階官能評価について各項目の平均及び標準偏差を図 18 に示す。Wilcoxon の符号順位検定を行った結果、βモードとαモードの間で「楽しさ」の項目において有意水準 0.1% で有意差が認められ、「好み」「和み」「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、また「対話しやすさ (話しやすさ)」「安心感」の項目では有意水準 5% で有意差が認められた。

Fig. 18 Results of seven points bipolar rating in the communication experiment under free conversation. InterActor with Negative/Positive motions (β) was affirmatively evaluated in comparison with conventional InterActor (α).

4.3 考察

実験の結果、 β モードが一対比較、7段階官能評価ともに高く評価されており、システムの有効性が示された。 β モードに対する自由記述では「うなづく動作や笑う動作がある方がリアルに話しているように見えて話がしやすくなった」や「苦労話をしたときに緩和動作をとってくれてとても和んだ」といったコメントがあり、 β モードが好意的に受け入れられていることが確認された。一方で「「やばい」のようなマイナスでもプラスでもとらえることのできる言葉にも対応できるとよりよくなと思った」「Negative なワードを言ってない時に盛り上げる動作をしてくれることが何回かあったのでそっちに目が行ってしまい、話の内容が頭に入っていないことがあった」などといったコメントもあり、発話文取得の際の認識精度や、動作の提示タイミングなどに対して改善すべき意見が得られた。

5. 結 言

本論文では、音声駆動型身体的引き込みキャラクタ InterActor に音声認識を導入し、発話感情推定を行うキャラクタシステムの開発を行った。認識した発話音声から感情極性値を算出し、Positive の閾値を上回った場合には、強く同調することで正の感情を盛り上げ、Negative の閾値を下回った場合には、それを打消し、緩和させる動作を行うことで負の感情を緩和させた。また、開発した感情推定手法を適用したキャラクタシステムを用いたロールプレイによる評価実験において、シナリオに基づくキャラクタへの語りかけによって、感情極性に対応して適切に反応動作を行うことが評価された。さらに2人1組によるコミュニケーション実験により、システムの有効性を示した。本論文では、否定的な発話を緩和する動作に焦点を当ててシステム開発を行ったが、音声認識によりさらに詳細に対応動作を生成することで、高次元インタラクションをデザインすることが可能であり、そのようなコミュニケーションシステムの開発が期待される。

文 献

- Cassell, J., Embodied conversational interface agents, Communications of the ACM, Vol. 43, No. 4 (2000), pp. 70–78.
- Condon, W. S. and Sander, L. W., Neonate movement is synchronized with adult speech, Science, Vol.183 (1974), pp. 99–101.
- 檀原龍正, 渡辺富夫, 大久保雅史, 音声駆動型身体引き込みキャラクタ InterActor が発話音声に与える効果, 日本機械学会論文集 C 編, Vol.71, No.712 (2005), pp. 152–159.
- Ekman, P., Friesen, W. V. and Ellsworth, P., What emotion categories or dimensions can observers judge from facial behavior emotion in the human face, New York: Cambridge University Press (1982), pp. 39–55.
- 橋本卓弥, 平松幸男, 辻俊明, 小林宏, ロボット受付嬢 SAYA を用いたリアルなうなづくに関する研究, 日本機械学会論文集 C 編, Vol.73, No.735 (2007), pp. 3046–3054.

- 石井裕, 瀬島吉裕, 渡辺富夫, 通信遅延環境における自己の身体的アバタ動作遅延提示の効果, ヒューマンインタフェース学会論文誌, Vol. 13, No. 1 (2010), pp. 23–30.
- 伊藤憲三, 北脇信彦, 会話音声の時間的特徴量に着目した遅延品質評価法, 日本音響学会誌, Vol. 43, No. 11 (1987), pp. 851–857.
- 川本真一, 下平博, 新田恒雄, 西本卓也, 中村哲, 伊藤克亘, 森島繁生, 四倉達夫, 甲斐充彦, 李晃伸, 山下洋一, 小林隆夫, 徳田恵一, 広瀬啓吉, 峯松信明, 山田篤, 伝康晴, 宇津呂武仁, 嵯峨山茂樹, カスタマイズ性を考慮した擬人化音声対話ソフトウェアツールの設計, 情報処理学会論文誌, Vol.43, No. 7 (2002), pp. 2249–2263.
- 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 (2009), pp. 31–38.
- Parrott, W., Emotions in social psychology, Psychology Press (2001).
- アレックス (サンディ) ペントランド著, 柴田裕之訳, 安西祐一郎監訳, 正直シグナル 非言語コミュニケーションの科学, みすず書房 (2013) (原著: Pentland, A., Honest signals: how they shape our world, The MIT Press (2010)).
- 鹿野清宏, Cincarek Tobias, 川波弘道, 西村竜一, 李晃伸, 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 (2006), pp. 33–38.
- 高村大也, 乾孝司, 奥村学, スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, Vol. 47, No. 2 (2006), pp. 627–637.
- Watanabe, T., Human-entrained embodied interaction and communication technology, Emotional Engineering, Springer (2011), pp. 161–177.
- Watanabe, T., Okubo, M., Nakashige, M. and Danbara, R., InterActor: speech-driven embodied interactive actor, International Journal of Human-Computer Interaction, Vol.17, No. 1 (2004), pp. 43–60.
- Watanabe, T. and Miwa, Y., Duality of embodiment and support for co-creation in hand contact improvisation, Journal of Advanced Mechanical Design, System, and Manufacturing, Vol. 6, No. 7 (2012), pp. 1307–1318.
- Woo, J., Botzheim, J. and Kubota, N., Verbal conversation system for a socially embedded robot partner using emotional model, Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (2015) pp. 37–42.

References

- Cassell, J., Embodied conversational interface agents, Communications of the ACM, Vol. 43, No. 4 (2000), pp. 70–78.
- Condon, W. S. and Sander, L. W., Neonate movement is synchronized with adult speech, Science, Vol.183 (1974), pp. 99–101.
- Danbara, R., Watanabe, T. and Okubo, T., Effects of speech-driven embodied entrainment character “InterActor” on talker’s speech, Transactions of the Japan Society of Mechanical Engineers, Series C, Vol.71, No.712 (2005), pp. 152–159 (in Japanese).
- Ekman, P., Friesen, W. V. and Ellsworth, P., What emotion categories or dimensions can observers judge from facial behavior emotion in the human face, New York: Cambridge University Press (1982), pp. 39–55.
- Hashimoto, T., Hiramatsu, S., Tsuji, T. and Kobayashi, H., Research on realistic nod with receptionist robot SAYA that has human-like appearance, Transactions of the Japan Society of Mechanical Engineers, Series C, Vol.73, No.735 (2007), pp. 3046–3054 (in Japanese).
- Ishii, Y., Sejima, Y. and Tomio Watanabe, T., Effects of delayed presentation of self-embodied avatar motion with network delay, Transactions of Human Interface Society, Vol. 13, No. 1 (2010), pp. 23–30 (in Japanese).
- Itoh, K. and Kitawaki, N., Evaluation method for delay effect on conversation based its time characteristics, Journal of Acoustical Society of Japan, Vol. 43, No. 11 (1987), pp. 851–857 (in Japanese).
- Kawamoto, S., Shimodaira, H., Nitta, T., Nishimoto, T., Nakamura, S., Itou, K., Morishima, S., Yotsukura, T., Kai, A., Lee, A., Yamashita, Y., Kobayashi, T., Tokuda, K., Hirose, K., Minematsu, N., Yamada, A., Den, Y., Utsuro, T. and Sagayama, S., Design of software toolkit for anthropomorphic spoken dialog agent software with customization-oriented features, IPSJ Journal, Vol.43, No.7 (2002), pp. 2249–2263 (in Japanese).

- Lee, A. and Kawahara, T., Developing a speech recognition interface using julius, Journal of Human Interface Society, Vol. 11, No. 1 (2009), pp. 31–38 (in Japanese).
- Parrott, W., Emotions in social psychology, Psychology Press (2001).
- Pentland, A., Honest signals: how they shape our world, The MIT Press (2010).
- Shikano, K., Cincarek, T., Kawanami, H., Nishimura, R. and Lee, A., Development and evaluation of Takemaru-kun spoken guidance system and portability to Kita-chan and Kita-robo systems, IEICE technical report, 2006-SLP-63 (2006), pp. 33–38 (in Japanese).
- Takamura, H., Inui, T. and Okumura, M., Extracting semantic orientations using spin model, IPSJ Journal, Vol. 47, No. 2 (2006), pp. 627–637 (in Japanese).
- Watanabe, T., Human-entrained embodied interaction and communication technology, Emotional Engineering, Springer (2011), pp. 161–177.
- Watanabe, T., Okubo, M., Nakashige, M. and Danbara, R., InterActor: speech-driven embodied interactive actor, International Journal of Human-Computer Interaction, Vol.17, No.1 (2004), pp. 43–60.
- Watanabe, T. and Miwa, Y., Duality of embodiment and support for co-creation in hand contact improvisation, Journal of Advanced Mechanical Design, System, and Manufacturing, Vol. 6, No. 7 (2015), pp. 1307–1318.
- Woo, J., Botzheim, J. and Kubota, N., Verbal conversation system for a socially embedded robot partner using emotional model, Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (2015), pp. 37–42.