

## エージェントによる接客トレーニングシステムの研究

## A Training System with a Character Agent for Communicating the Mind of Service

守田 航大<sup>\*1</sup>  
Kodai Morita

角 薫<sup>\*2</sup>  
Kaoru Sumi

<sup>\*1</sup> 公立はこだて未来大学大学院システム情報科学研究科知能情報科学領域  
Graduate School of Information Science, Future University Hakodate

This paper describes a dialogue system for communicating Japanese-style of service-mind as typified by paying attention to the little things and feeling compassion for a customer's feelings. Users are able to learn taking communication with character agent using the system. We think this system is possible to spread Japanese-style of service-mind around the world. Our system uses a speech recognition system, a speech synthesis system and a dialogue system, and it also uses Microsoft Kinect for facial expression and gesture recognition. This system is evaluated for the effect of a system using subjects by comparing with previous system using EEG machine.

## 1. はじめに

日本の接客サービスは世界でも定評があり、例えばエアラインの客室乗務員のサービスランキングでは上位に位置している。接客サービスではお客様とのコミュニケーションを取ることが最も大切なことで笑顔が重要であると言われている<sup>[1]</sup>。本研究では、エージェントとの対話により接客サービスを学習するシステムを開発した。日本のサービスマインドをシステム化することで、近年の増加傾向にある海外からの労働者にもわかりやすく伝えられると期待できる。

## 2. 関連研究

ELIZA<sup>[2]</sup>による非指示的カウンセリングは抜き出したキーワードを決まり文句に埋め込むといった単純なチャット形式のプログラムで、目的を持った対話ができるわけではなく対話を長引かせるための設計である。近年は音声対話が身近になってきており、例えば携帯電話では iOS 向けの音声認識を利用した操作補助機能の Siri, Android には Google 音声認識が搭載されている。SimCoach<sup>1</sup>はアメリカの軍人用のメンタルケアシステムである。主に退役軍人を対象にしたもので、戦場で精神的な傷、トラウマを抱えた兵士の心を癒す目的で開発され、エージェントを用いた音声対話によるアプリケーションへの期待は高まっている。日本においてもポジティブシンキングを促す引きこもり克服支援のためのシステムがあり、エージェントを使ってメンタルケアを目的としている<sup>[3]</sup>。

## 3. エージェント・トレーニングシステム

本研究では 3D モデルのアニメーションを用いるために C# (Visual Studio 2010) と XNA を使用した。本研究は、脳波測定器 (EEG) を用いたサービスマインドを伝えるキャラクターエージェント対話システム<sup>[4][5][6][7]</sup>に Kinect による表情認識とジェスチャー認識の機能を拡張したものである<sup>[8][9]</sup>。システム概念図は図 1 の通りである。

本システムでは Google 音声認識と AITalk<sup>®2</sup>を用いた。本システムの対話の相手を作成するために 3D モデルのキャラクタエ

ージェントを使用する。このエージェントには、本研究では「微笑み」、「悲しみ」を使用する。またエージェントの口の形をそれぞれ「あ」、「い」、「う」、「え」、「お」と用意し、モーフィングを利用した。さらに、エージェントは「浅いお辞儀」、「深いお辞儀」等のジェスチャーを行い、エージェントは一定間隔で瞬きをするようになっている。システムのスナップショットは図 2 である。対話システムとして Artificial Intelligence Markup Language (以下 AIML) を使用する。AIML は自然言語ソフトウェア構築のための XML を応用したマークアップ言語である。AIML のトピック機能により、話の内容ごとに会話を分けた。

### 3.1 表情認識システム

システムの Kinect を用いた表情認識では Kinect for Windows SDK の Face Tracking という機能、Face Tracking Basics-WPF サンプルをベースに、インターフェースに合うように組み込んだ。サンプルでは顔を検出するのみであるが、Kinect for Windows の Face Tracking の Animation Units/AU という表情を特徴づけるデータを用い、眉の角度、口角の角度などを基準として判別した。本研究ではこの AU の特徴量をそのまま使用している。口を横に引き伸ばしている状態 (AU2 が 0.4 以上) で笑顔と判定し、それ以外で眉をしかめている状態 (AU3 が 0.4 以上) で悲しい顔と判定している。また、口が開いている (AU1 が 0.5 以上) 場合は除外している。また、Kinect for Windows SDK のスケルトン検出は複数人のユーザ検出に対応していたが、本研究では、一人を特定する必要があるため、複数人検出された場合は顔の面積が最も大きい人 (最も Kinect に近い人) になるように実装した。

<sup>1</sup> SimCoach, <http://www.simcoach.org/>

<sup>2</sup> AITalk, <http://www.ai-j.jp>

連絡先: 守田 航大, 公立はこだて未来大学, 〒048-8655 北海道函館市亀田中野町 116-2 角薫研究室,  
odayakajanai.blue@gmail.com

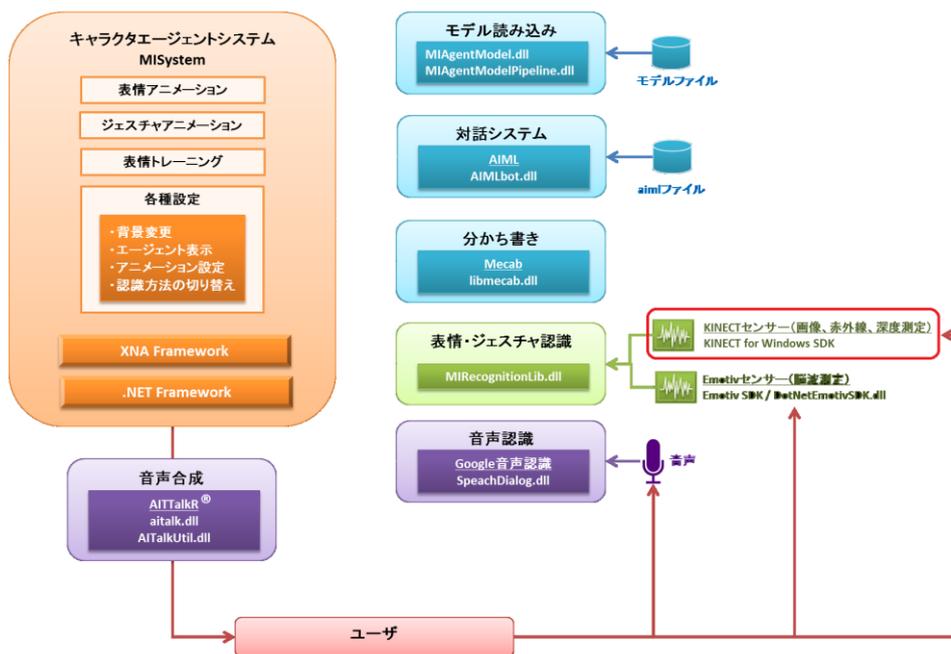


図 1 システム概念図



図 2 システムのスナップショット

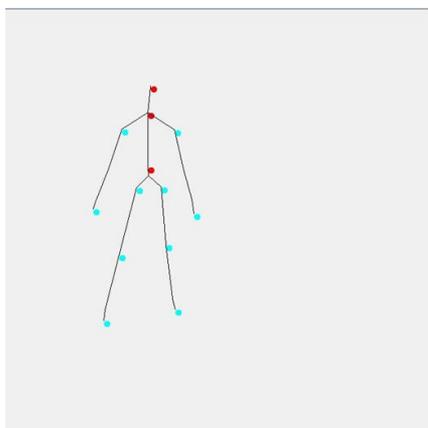


図 3 システムのスナップショット

Kinect for Windows SDK の骨格検出をベースにし、取得している骨格のポイントを骨格と同時に取得している RGB 映像での座標に照らし合わせ、どの程度骨格の座標に変化があるかを読み取ってジェスチャ認識を行った。Kinect の骨格認識では頭、肩、右手、左手、腰、右足、左足など計 20 箇所の骨格ポイントを検出することができるが、本研究のジェスチャ認識では頭、背骨、腰の三か所の骨格のポイントを使用して認識している。図 3 は骨格認識プログラムのスクリーンショットである。3 つの点は今回使用した頭、背骨、腰のポイントである。それぞれ頭のポイントを hpoint、肩の中央のポイントを scpoint、腰のポイントを spoint と定義している。骨格の座標は一番左上を(0,0)とし、右に行くほど X 座標の値が大きくなり、下に行くほど Y 座標の値が大きくなる。spoint の Y 座標と hpoint の Y 座標の差が一定以上あった場合に立っていると、立っている状態からの hpoint の遷移状態と、scpoint と spoint の差から浅いお辞儀なのか、深いお辞儀なのかを判断している。認識できるジェスチャは「浅いお辞儀」と「深いお辞儀」を検出できるようにしている。「浅いお辞儀」では約 15 度の角度でお辞儀と検出され、「深いお辞儀」では約 30 度以上の角度で深いお辞儀と検出されるようにした。

### 3.3 接客トレーニング対話システム

接客トレーニング対話システムでは、「褒め方について」、「クレームについて」、「挨拶について」の 3 つのトレーニングコンテンツからなる。それぞれのトレーニングで音声認識、表情認識、ジェスチャ認識を行っている。会話パートでは接客に関する本<sup>[1]</sup>を参考にして内容を作成されている。褒め方には褒めるテクニックと笑顔のトレーニングを、クレームについては対処方法と悲しい顔と深いお辞儀のジェスチャを、挨拶ではどのお店でも使う基本的な挨拶についてと笑顔と浅いお辞儀のトレーニングをすることができる。初めはエージェントの自己紹介と、接客業について何を聞きたいかとユーザーに促すパートがある。その後トレーニングに入り、トレーニングではまずエージェントがお手本となる言葉・ジェスチャ・表情を見せ、次にユーザーがトレーニングするという流れ

## 3.2 ジェスチャ認識

である。トレーニングでは、指定された表情を作り、ジェスチャをしながら指定された台詞を読むといったものである。この際、読み上げる文に関しては音声認識と AIML を使用して話しているのか判断し、表情とジェスチャは Kinect によって判断する。トレーニングでは音声認識、表情認識、ジェスチャ認識を同時に行い、全てのトレーニングがうまくいけば、次のトレーニングに進める。しかし、一つでも認識が失敗すればもう一度同じトレーニングを行う。

#### 4. 実験

実験では被験者 12 名にシステムのインターフェース、使いやすさかどうか、昨年度に比べて改善されたかなどの評価をもらった。被験者は大学 3 年生から 4 年生で男性 9 名、女性 3 名、平均年齢 21.8 歳であった。被験者には褒め方について、クレームについて、挨拶についての 3 つのトレーニングを受けてもらい、トレーニング後に、アンケートを答えてもらった。アンケートは表 1 に記載するもので、質問 2 から質問 8 までは 5 段階評価とともにその理由を答えてもらった。

表 1 アンケートの質問内容

問 1	なにか接客業をやっていた経験はありますか？ ありと答えた方は経験年数と、どのようなお仕事ですか？	
問 2	このシステムは楽しかったですか？	
問 3	また使ってみたいと思いましたか？	
問 4	このシステムは接客のトレーニングとして役に立ちそうですか？	
問 5	自分の表情は見えていたほうが良いですか？	
問 6	エージェントの印象についてどう思いましたか？	
	質問 6.1	印象はどうでしたか？
	質問 6.2	親切でしたか？
	質問 6.3	丁寧でしたか？
問 7	システムへ話す事はやりやすかったですか？	
問 8	システムからの音声は分かりやすかったですか？	

アンケートの結果、各質問の 5 段階評価の平均値は表 2 のような結果になった。良い評価を得られたものは 5 点、悪い評価を 1 点としている。さらに、表 2 には既存のシステムの結果の平均値も記載する。「本システム」とは Kinect を使用したものであり、「既存のシステム」とは脳波測定器を使用したものである。各平均値は小数点第二位までとし、小数点第三位を四捨五入している。この結果から、相関係数を求め、脳波測定器を利用した既存のシステムの結果と比較し t 検定を行った。

質問 1 より、接客業の経験のある者と、接客業の経験のない者では、接客業の経験のある者ほど本システムをまた使ってみたい、役に立ちそうと好意的な評価を得られた。また、各質問に対して相関関係を調べた結果、質問 2 と質問 3 において、0.77 と強い相関係数が見られた。よって、このシステムを楽しく感じた場合、このシステムをまた使ってみたいと感じる傾向にあることがわかる。

Kinect を用いた本システムと、脳波測定器を用いた既存のシステムの実験の各質問に対して t 検定を行った結果、質問 2「このシステムは楽しかったか？」では有意傾向にあった ( $t=0.36$ ,  $df=11$ ,  $p < 0.05$ )。質問 3「またこのシステムを使ってみたいか？」では有意であるという結果が出た ( $t=-3.22$ ,  $df=11$ ,  $p < 0.05$ )。どちらも脳波測定器を用いた既存のシステムの平均値の方が高か

った。よって Kinect を用いた本システムより、脳波測定器を用いた既存のシステムの方が全体的に評価が高かった。

表 2 システム 5 段階評価の平均値

	平均値(本システム)	平均値(既存システム)
質問 2	2.92	4.05
質問 3	2.83	4.00
質問 4	3.42	4.15
質問 5	4.25	2.45
質問 6.1	3.92	3.60
質問 6.2	3.42	4.15
質問 6.3	4.25	4.45
質問 6.4	4.08	3.75
質問 7	2.50	3.15
質問 8	4.00	4.20

#### 5. 考察

実験で得られた結果より、既存のシステムと比較するとシステム面において全体的に本年度は昨年度よりも平均値が低くなっている。また、実験時のアンケートでは、本年度は音声認識についての否定的な意見が目立った。認識率の低下が目立った理由として既存のシステムに加えて Kinect での処理を加えているためプログラム全体の処理が重くなったためだと考えられる。一方で「また使ってみたい」などの良い意見も挙げられていた。本研究では使用した音声認識に対する意見が否定的であり、エージェントやシステムの内容に対しては好意的な意見が多かった。音声認識を改善し、認識精度を上げることができればシステム全体に好意的な意見を持ってもらえると考えられる。t 検定では、質問 2 に対して有意傾向が見られ、質問 3 に対して有意であるという結果が出た。この 2 つはシステムに関する質問である。加えて、質問 2 と質問 3 は高い相関が見られ、「システムの楽しさ」が「また使ってみたいか」に影響を与えた。本システムではこの 2 つの評価平均値は低い数値となっており、音声認識を改善することで評価平均値が向上すると考えられる。また、「どの程度ダメなのか具体的な改善案が出ない」といった意見から、動作や表情が単純にできているかできていないかだけでなく、どの程度表情ができていないか、どの程度動作ができていないか等をユーザに示すことでさらに発展し、より良いシステムとして利用できるのではないかと考えられる。

#### 6. まとめ

本研究では接客トレーニングのためのシステムを開発し、実験によりシステムを評価した。

#### 参考文献

- [1] 森下 裕道: 心理接客術 お客様の心を一瞬でギョッとつかむ接し方, ソシム (2009)
- [2] Weizenbaum, Joseph: ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine, Communications of the ACM 9 (1): pp.36–45 (1966)
- [3] 高橋 識行, 角 薫: ポジティブシンキングを促す引きこもり克服支援のためのシリアスゲームの研究, 日本デジタルゲーム学会 2013 年度年次大会, pp.99-101, 日本デジタルゲーム学会 (2014)

- [4] 江端 竜次, 角 薫: サービスマインドを伝えるキャラクターエージェント対話システム, インタラクション 2013, (1EXB-45), 情報処理学会(2013)
- [5] Kaoru Sumi: Communication with a Virtual Agent via Facial Expressions, IJCAI Workshop on Empathic Computing, 4th International Workshop on Empathetic Computing (IWEC'13), Beijing, China (招待講演)(2013)
- [6] Kaoru Sumi, Ryuji Ebata: Human Agent Interaction for Learning Service-Minded Communication, iHAI2013, 1st international conference on Human-Agent Interaction(2013)
- [7] Kaoru Sumi, Ryuji Ebata: A Character Agent System for Promoting Service-Minded Communication, Intelligent Virtual Agents, Lecture Notes in Computer Science, LNAI8108, pp.438, Springer, (2013)
- [8] 守田 航大: サービスマインドを伝えるエージェント・トレーニングシステム, 公立はこだて未来大学卒業論(2013)
- [9] 守田 航大, 角 薫: 接客トレーニングのためのシリアスゲームの研究, 日本デジタルゲーム学会 2013 年度年次大会, pp.9-13, 日本デジタルゲーム学会(2014)