

# シナリオ対話における感情音声合成を用いた対話システムの 評価と感情付与方法の検討

加瀬 嵩人<sup>†1,a)</sup> 能勢 隆<sup>†1</sup> 千葉 祐弥<sup>†1</sup> 伊藤 彰則<sup>†1</sup>

概要：近年、非タスク指向型の音声対話システムへの需要が拡大しており、様々な研究がされている。それらほとんどの研究は言語的な観点から適切な応答の生成を目指したものである。一方で人間同士の会話においては、感情表現や発話様式などのパラ言語情報を効果的に利用することにより、対話を円滑に進めることができると考えられる。そこで我々はシステムの応答の内容ではなく、応答の仕方に着目し、感情音声合成を対話システムに用いることを試みる。本研究ではまず、適切な感情付与を人手により与えた場合に実際に対話システムの質が向上するかを複数のシナリオを作成して主観基準により評価する。次に、感情付与を自動化するために、システム発話に応じた付与とユーザ発話に協調した付与の2つの手法について検討を行う。評価結果から、感情を自動付与することで対話におけるユーザの主観評価スコアが向上すること、またユーザ発話に協調した感情付与がより効果的であることを示す。

## 1. はじめに

近年、AppleのSiri[1]やNTTドコモのしゃべってコンシェル[2]など、ユーザが音声によって対話的に処理を行うサービスが登場し、普及が進んでいる[3]。これらのサービスでは情報検索などのタスク指向対話システムが中心的な機能として提供されている一方で、ユーザの何気ない発話に対して雑談的な応答を返す機能も提供されており、雑談を目的として利用するユーザが少なくないというデータが報告されている[4]。このような雑談的な対話の例としては、対話エージェントに対する好きなものについての質問・応答や、自分の趣味に関する話などが考えられる。このように、音声対話システムがより我々の身近になりつつあり、システムがユーザの入力に対してどのような発話内容を応答として返すかについては様々な研究が行われている[5]。しかし、システムが発話する音声そのものに目を向けると、その話し方は人間に比べて抑揚がない読み上げ調のものがほとんどである。このような読み上げ調の音声は検索結果などの情報をユーザに伝える目的としては十分であると考えられるが、先に述べた雑談など、より人間同士の普段の会話に近い音声対話を実現する上では未だ十分とは言えない。

これまでに、タスク指向型ではなく非タスク指向型である雑談のような対話についての研究がされ[6],[7]、ユー

ザを楽しませることを目的とした応答を行うことのできる対話システムについて検討が行われている。柴田らは雑談対話において自由な応答を実現させるためにWebから妥当な発話文を選択する手法を提案しており、約66%の発話において自然で意味のある応答が生成できることを示している[8]。また水野らはユーザを楽しませるために語彙の拡張を行った概念辞書を知識として用いることで、ユーモアを含んだ応答のできるシステムを構築した[9]。これら雑談対話システムについての研究は言語的な研究が中心である。好きなものについての話や、趣味に関する話では、発話内容が多様であってもそれを伝えるための応答の仕方が画一的であれば、場合によっては対話の楽しさがユーザに十分に伝わらないことが予想できる。以上のことから我々は、非タスク指向型の対話システムにおけるシステムの応答の仕方について検討を行う。

応答の仕方に関連した研究としては、応答タイミングや発話の韻律情報に着目した研究が行われている[10],[11]。西村ら是对話における人同士の発話の基本周波数(F0)の相関に着目し、ユーザの発話の基本周波数に合わせてシステムの発話のF0を変動させる手法を提案している[10]。また武藤らは応答タイミングや発話長を考慮することにより、ユーザの人工物に対する印象が変化することを明らかにしている[11]。一方で音声応答を生成する音声合成については、近年統計モデルの一つである隠れマルコフモデル(HMM)に基づく手法[12]についての研究が幅広く行われている[13]。HMM音声合成が従来の文献[14]などの波

<sup>†1</sup> 現在、東北大学  
Presently with Tohoku University  
<sup>a)</sup> taketo.kase.t6@dc.tohoku.ac.jp

形素片選択接続方式と比べて優れている点として、感情表現・発話様式などを柔軟にモデル化し [15], 制御できることが挙げられる [16], [17]. HMM 音声合成は最近では対話システムへも導入されており [18], 情報案内やデジタルサイネージなどへの利用の可能性が示されている。

感情音声を音声対話システムに応用する例として, Acostaらの研究がある [19]. Acostaらは, 人間同士(学生とカウンセラー)の対話の感情極性をラベル付けし, 機械学習によって感情応答パターンを推定した. また, それを音声対話システムに応用することによって, システムとの間のラポール(信頼関係)が増強することを示している. 彼らの研究では, 感情推定に直前のユーザ発話の音響的特徴を利用している. 彼らの研究での対話は非タスク指向対話であるが, 本研究で目標としている雑談対話とはやや異なり, カウンセリング対話である. そのため, カウンセリングにおけるラポールと雑談対話における「豊かさ」「楽しさ」が同じであるかどうかは明らかではない. また, システム発話に対する適切な感情付与がユーザ発話にどのように影響するのか, ユーザやシステム発話の意味的な内容が適切な感情付与にどう関係するかについても明らかになっていない. 感情音声合成を用いて雑談などのより人間同士に近い対話を行える対話システムを実現するためには, 実際にシステムとの対話を行うことによってユーザの印象がどのように変化するかを評価・分析することが必要不可欠である.

本研究では, 感情音声合成を対話システムに導入して主観評価実験を行い, 感情付与によってどのような効果が得られるかについて分析する. 実験では日常生活を想定した複数のシナリオを用意し, あらかじめ人手により適切に付与された感情を付与した場合とそうでない場合について, 被験者にディスプレイに表示されたエージェントとの対話を行ってもらい, 対話やエージェントに関する4つの項目に対して評価してもらおう. また, 感情付与を自動化するために, 感情表現語に基づくユーザ発話の状態推定を行い, システム自身の発話から感情を決定した場合, およびユーザの発話に協調して感情付与を行った場合について, 感情付与を行わない場合, ランダムに付与した場合との比較を行い, その有効性を検証する.

## 2. 音声対話システムにおける適切な感情による音声合成の有効性

従来の音声対話研究では感情音声合成を利用することがどのように有効であるかについて詳しい検討はされていない. そこで本節では, 用意したシナリオに対し, あらかじめ人手によりシナリオの各発話内容に応じて合成音声に適切な感情を付与した場合に, 対話の豊かさやエージェントへの印象が改善するかどうかについて調べる.

表 1: シナリオの例 (人手による適切な付与)

Table 1 Example of a scenario (manual emotion labels)

Speaker	Speech	感情
User	おはよう.	
System	おはようございます. 昨日はよく眠れましたか?	平常
User	はい, よく眠れました.	
System	それは良かったです.	喜び
User	今日の天気は?	
System	晴れです. 晴れの日は気分が良いですね.	喜び
User	そうですね.	
System	今日どこかへ行かれるのですか?	平常
User	ツーリングに行ってきます.	
System	天気の良い日のツーリングは最高ですね. 気をつけて行ってきてください.	喜び

### 2.1 感情音声合成を用いたシステムに対するユーザの満足度

まず, 適切に感情を付与した合成音声を用いた場合に, 用いない場合に比べて対話システムに対するユーザの満足度がどのように変化するかを調べるため, 主観評価実験を行った.

#### 2.1.1 主観評価実験

本実験では, 被験者間の対話内容の違いが評価結果に与える影響を排除するため, あらかじめ用意したシナリオを用いた. 12名の被験者にエージェントとシナリオに沿った対話を行ってもらい, 対話とエージェントについての満足度を評価させた. シナリオの内容はエージェントとの日常会話を想定したもので, 「晴れの日の会話」「試合に勝って帰宅した時の会話」の2つのポジティブなシナリオ, および「雨の日の会話」「試合に負けて帰宅した時の会話」の2つのネガティブなシナリオの計4つの異なるシナリオを用意した. 各シナリオの長さは10ターン程度である. 実験にはMMDAgentを使用しており, 合成音声もツールキットに付属の音声(メイちゃん)を使用した. 付与する感情は通常・喜び・悲しみの3種類とし, 以下の3つの条件でシステムの合成音声に対し各感情を発話単位で付与した.

- 感情付与なし(すべて通常)
- 人手による適切な感情付与
- 不適切な感情付与(適切とは逆の付与)

適切な感情付与を行うためにあらかじめ10名の作業員による人手での付与を行った. 作業員はシナリオ全体を読み, それぞれの発話文に対し, 前後の文脈を考慮してシステムの応答として適切であると思う感情を付与した. 10名のうち7名以上が一致した感情を採用し, 7名以上の一致

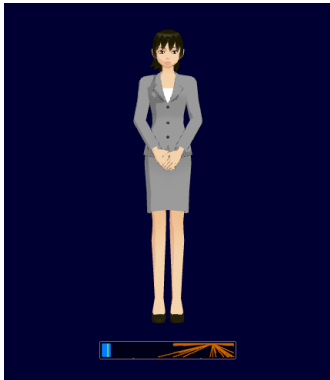


図 1: 本研究で用いた対話エージェント

Fig. 1 Agent used in the experiment

がなかった場合は通常スタイルとした。4つのシナリオに含まれる18のシステム発話のうち16発話において7名以上の評価が一致し、喜びが6発話、悲しみが4発話付与された。不適切な付与は適切な付与で付与された喜びと悲しみを逆に付与したものである。表1に適切な付与がされたシナリオの例を示す。

被験者は防音室内で椅子に座り、机に置かれたノートPCにおいてヘッドセットを用いてエージェントとの対話を行った。3つの感情の付与方法の提示順はランダムとした。対話を行うエージェントは図1に示す女性のエージェントであり、PCの画面に4cm × 10cm程度の大きさで表示される。被験者には4つすべてのシナリオにおいて先に示した3つの条件で対話を行い、対話が終わり次第、音声対話における満足度を以下の4つの指標により評価してもらった。

- 対話の豊かさ
- 対話の楽しさ
- エージェントの人らしさ
- エージェントの印象

評価スコアは「1: Bad」から「5: Excellent」の5段階とした。なお、実験後に各被験者にはシステム発話の感情の変化が認識できたかについてアンケートを行い、全ての被験者が何らかの感情の変化を認識できたことを確認した。

### 2.1.2 実験結果と考察

図2に評価項目ごとに、3種類の感情付与の方法に対する全被験者の平均スコアを示す。感情付与の方法を要因とする1元配置分散分析をそれぞれの評価項目について行ったところ、3種類のスコアの差は対話の豊かさ、楽しさ、エージェントの人らしさおよび印象において、それぞれ1%有意 ( $p < 0.001$ ,  $p < 0.001$ ,  $p = 0.009$ ,  $p < 0.001$ ) であり、感情付与によって評価に違いが生じたことが言える。

次に、各評価項目について Bonferroni 補正による多重比較検定を行った結果を表2に示す。

結果から全ての評価項目において、適切な付与を行った場合がその他2つと比べて有意に高いスコアを得ている。

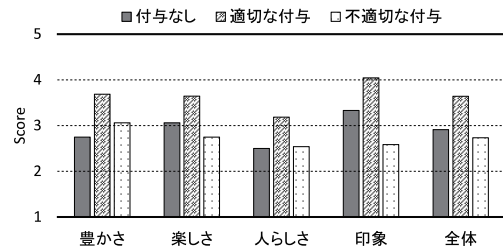


図 2: 主観評価結果

Fig. 2 Score of subjective evaluation

また、エージェントの印象に関しては付与なしより不適切な付与の方が低いスコアとなっている。各被験者はシステム発話の感情の変化を認識しており、エージェントの印象については“悲しみ”を付与すべきところで間違った“喜び”を付与してしまうと「馬鹿にされたような気分」となるという意見も得られた。以上のことより感情音声合成を適切に用いることはシステムとの対話やエージェントの評価に対して効果的であるといえる。一方で付与方法を間違えると評価が下がってしまい、さらに本来意図するものとは反対の印象をユーザに与えてしまうこともあったため、感情の付与方法の検討が重要であると考えられる。

### 3. 対話システムへの発話状態推定の導入

前節ではシステムの発話に対する感情の付与を手で行っていた。実験で用いたような限られたシナリオに対しては、このような人手による付与が可能であるが、人的コストが大きいという問題がある。このため、比較的規模の大きい対話システムにおいてはこのような人手による付与は現実的ではない。そこで、本節ではシステムやユーザの発話内容から発話状態を推定することで、感情の付与を自動的に行う手法を検討する。

システム発話については発話文からの推定、ユーザ発話については発話文や発話音声の韻律からの推定が考えられる。徳久らは Web を利用して作成したコーパスによってテキストから感情を推定する手法を提案している [20]。また、高野らは8つの感情について、感情表現辞典 [21] に含まれる感情関連語を用いて推定を行っている [22]。一方で、発話音声の韻律から感情の分析、推定を行う研究は数多くされており [23][24]、Acosta らの研究では実際の対話音声から感情推定を行っている [19]。しかし、機械に対して話す場合のユーザの思う「認識しやすい発話」に対して韻律から感情推定するのは難しいと考えられる。これらに基づき、本研究では発話文を利用したテキストからの状態推定のみについて検討する。

#### 3.1 感情表現語に基づくユーザ発話の状態推定

本研究では、あらかじめ感情表現語を用意し、これとのマッチングを行うことで感情推定を行う。感情表現語とし

表 2: Bonferroni 補正による多重比較検定の結果 ( $p$  値)

Table 2 Result of multiple comparison test with Bonferroni correction (p-value)

組み合わせ	豊かさ	楽しさ	人らしさ	印象
付与なしと適切な付与	< 0.001**	0.005**	0.019*	0.001**
付与なしと不適切な付与	0.525	0.259	> 1.000	< 0.001**
適切な付与と不適切な付与	0.022*	< 0.001**	0.031*	< 0.001**

\*\* : 1%の水準で有意差が認められたもの

\* : 5%の水準で有意差が認められたもの

表 3: 感情表現語の例

Table 3 Examples of emotion expression words

単語	フレーズ
好き 幸せ	興味津々
嫌い 弱気	気が滅入る
腹立つ 美しさ	心地良い
むかつく 失望	あきれてものが言えない
安い 合格	物足りない
軽い 病気	非の打ち所がない

て用いるものは小林らによって収集された用言約 5,000 件の評価極性辞書 [25] と東山らによって収集された名詞約 8,500 表現 (複合名詞含む) の評価極性辞書 [26] を乾らが人手で極性を付与・チェックしたものである [27]. 感情表現語の例を表 3 に示す. なお, 表に示す通り, 感情表現語には単語だけでなくフレーズも含まれている. この辞書は極性として positive と negative を付与されているが, これを喜びと悲しみに対応させて利用する. 以下の 3 つの規則に従って発話文の中でマッチした感情表現語の極性を発話文の状態の推定結果とした. マッチングを行うために発話文を茶筌 [28] によって形態素解析し, それぞれの形態素の原型と品詞が感情表現語と一致した場合にマッチとした. また感情表現語がフレーズである場合にはフレーズ内の単語が全て一致した場合にマッチとした. 感情表現の後ろに否定語 (「ない」「ぬ」等) があった場合は極性を反転させた. また, 発話文に複数の感情表現が存在した場合は文末に近いものを推定に使用した. これに関しては高野ら [22] によって文末に近いものの方がより意味を持つことが示されている.

### 3.2 推定精度の評価

3.1 節で述べた推定方法による感情推定の評価実験を行った. 入力文として 2.1.1 節で述べた主観評価実験用に作ったシナリオ文 100 文と ASJ-JIPDEC [29] Vol.7 に収録されている 4 つの自由対話からランダムに 100 文を用いた. 実験結果の評価方法として, シナリオ文に関しては人手による付与との一致率を調査した. また, ASJ-JIPDEC に関し

表 5: ASJ-JIPDEC コーパスにおける感情推定精度 (%)

Table 5 Emotion estimation accuracy for sentence in ASJ-JIPDEC corpus (%)

適切である	不適切とはいえない	不適切である
64.7	21.6	13.7

ては推定結果の正しさを主観的に評価した. 評価者は 10 名である. この際, 評価項目として以下に示す 4 項目で評価をした. この評価項目は高野ら [22] の評価方法を参考にしている.

- その感情を付与することは適切である
- その感情を付与することは不適切とはいえない
- その感情を付与することは不適切である

それぞれの結果を表 4 と表 5 に示す. 表 4 は推定結果と人手による適切な付与との一致率を示したものであり, 表 4 より全体として 73.0%の精度を示した. また表 5 は ASJ-JIPDEC の文 100 文に対する 10 名の評価を平均したものであり, 表 5 に示すように 86.3 %の文において「適切」または「不適切とはいえない」という評価を得た. 不適切とはいえないと評価されたものには「でもゴールデンウィークの頃の旅行っていうのは前々から予約しないとね (喜び)」といったようなものがあり, 悲しみが付与される可能性も考えられるが, 喜びを付与したとしても間違いとは言えない, というような評価である. 以上のことからある程度高い精度で推定できているため, この推定手法を用いる.

## 4. 合成音声への感情の付与方法の検討

3.1 節で述べた状態推定を用いる際はシステムの発話文自体から推定を行うことが考えられるが, 一方で, 直前のユーザの発話に協調した応答をシステムが返すことによって, ユーザの満足度が向上することも考えられる. そこで, このような感情の付与方法の違いによってユーザが得る満足度の変化に違いが見られるのか検証するために主観評価実験を行った.

表 4: シナリオ文における感情推定精度 (%)  
Table 4 Emotion estimation accuracy (%)

シナリオ番号										
1	2	3	4	5	6	7	8	9	10	Ave.
94.4	58.3	50.0	60.0	83.3	75.0	87.5	80.0	100.0	85.7	73.0

表 6: シナリオの例 (自律的付与)

Table 6 Example of a scenario (self-estimation of emotion)

Speaker	Speech	
User	好きな食べ物は何ですか?	
System	ラーメンが好きです.	喜び
User	なぜですか?	
System	たくさん種類があって飽きないからです.	喜び
User	何ラーメンが好きですか?	
System	とんこつです.	平常
User	良いですね.	
System	美味しくてつい替え玉してしまいます.	喜び
User	わかります.	
System	ラーメン食べなくなってきました.	平常

表 7: シナリオの例 (協調的付与)

Table 7 Example of a scenario (cooperative estimation of emotion)

Speaker	Speech	
User	ただいま.	
System	お帰りなさい. 今日の試合はいかがでしたか?	平常
User	負けてしまいました.	
System	そうでしたか.	悲しみ
User	接戦を逃してしまいました.	
System	惜しかったんですね. 次は勝利したいですね.	悲しみ
User	はい, 頑張ります.	
System	お疲れ様でした. ゆっくり休んでください.	喜び

#### 4.1 主観評価実験

シナリオとしてあらかじめ全 12 対話を作成した. 内容は日常会話が 6 対話 (「天気についての会話」「試合についての会話」「早起きについての会話」各 2 対話ずつ), 趣味の話が 6 対話 (「野球についての会話」「映画についての会話」「食べ物についての会話」各 2 対話ずつ) で, それぞれ 10 ターン程度のものであり, 趣味の話に関してはユーザと

エージェントの嗜好が一致している場合としていない場合をそれぞれ 3 対話ずつ用意した. 各シナリオについて 10 名の被験者により評価を行った. 2.1 節の実験と同様に音声対話における満足度を「対話の豊かさ」「対話の楽しさ」「エージェントの人らしさ」「エージェントの印象」の 4 つの指標により評価してもらった.

主観評価実験では以下の 4 つの条件でシステムの応答に感情付与を行った.

- 感情付与なし
- 自律的付与 (システム発話に応じた付与)
- 協調的付与 (ユーザ発話に協調した付与)
- ランダムに付与

自律的付与ではシステム発話文に対し状態推定を行い, 推定された感情を付与した. 協調的付与では対象となるシステム発話の直前のユーザ発話 (入力発話) に対し状態推定を行い, 推定された感情を付与した. システム発話 56 発話のうち自律的付与では喜びが 20 発話, 悲しみが 7 発話, 協調的付与では喜びが 20 発話, 悲しみが 5 発話, それぞれ付与された. 表 6 および表 7 に自律的付与, 協調的付与がされたシナリオの例をそれぞれ示す.

自律的付与ではユーザの質問に答える発話や同意をする発話などに感情が付与されていることがわかる. ただし, この手法では「はい」などの一言の発話に対して感情を付与することができないという問題がある. これに対し, 協調的付与では同意だけでなく相づちにも感情が付与されており, 感情表現語を含まない短い発話に対しても感情を付与できることがわかる.

#### 4.2 実験結果と考察

図 3 に各評価項目の結果を示す. 感情付与の方法を要因とする 1 元配置分散分析をそれぞれの評価項目について行ったところ, 付与方法の間のスコアの差は豊かさ, 楽しさ, 人らしさ, エージェントの印象においてそれぞれ 1% 有意 ( $p < 0.001$ ,  $p < 0.001$ ,  $p = 0.002$ ,  $p < 0.001$ ) であり, 感情付与の手法によって評価に違いが生じたことが言える. さらに各評価項目について Bonferroni 補正による多重比較検定を行った結果を表 8 に示す.

対話の豊かさに関しては付与なしとそれ以外との間で有意な差が見られる. これによって, 今回行った 3 つの付与方法においては, 感情の付与を行うことでユーザは対話が

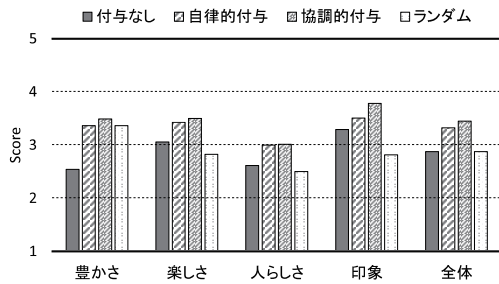


図 3: 主観評価結果

Fig. 3 Score of subjective evaluation

表 8: Bonferroni 補正による多重比較検定の結果 ( $p$  値)  
Table 8 Result of multiple comparison test with Bonferroni correction

(a) 豊かさ			
	自律的付与	協調的付与	ランダム
付与なし	< 0.001**	< 0.001**	< 0.001**
自律的付与		> 1.000	> 1.000
協調的付与			> 1.000
(b) 楽しさ			
	自律的付与	協調的付与	ランダム
付与なし	0.018*	0.002**	0.348
自律的付与		> 1.000	< 0.001**
協調的付与			< 0.001**
(c) 人らしさ			
	自律的付与	協調的付与	ランダム
付与なし	0.051	0.037*	> 1.000
自律的付与		> 1.000	0.004**
協調的付与			0.002**
(d) 印象			
	自律的付与	協調的付与	ランダム
付与なし	0.040	< 0.001**	< 0.001**
自律的付与		0.119	< 0.001**
協調的付与			< 0.001**

\*\* : 1%の水準で有意差が認められたもの

\* : 5%の水準で有意差が認められたもの

豊かになったと感じるということがいえる。

対話の楽しさに関してもまずランダムな付与とその他2つの付与との間に有意な差が見られる。これは、対話中において的外れな感情付与が行われることで、ユーザの気持ちさがめてしまったことが理由であると考えられる。また付与なしと2つの付与との間にも有意な差が見られる。これより今回検討を行った2つの付与方法を行うことで、ユーザがその対話を楽しいと感じるようになることがわ

かった。

エージェントの人らしさに関してはまず、ランダムな付与が他の2つの付与に比べて低い評価となっている。これは、一定の規則を持った付与がされていないことでユーザが対話の流れに違和感を感じた結果であると考えられる。他の2つに関しては的外れな付与が起きにくいために、より人間らしいと評価されたものと考えられる。また、付与なしと協調的付与において有意な差が見られる。これは相づちの部分でユーザ発話に協調した感情が付与されるため、ユーザがシステムに対して一定の賢さを感じた結果であると考察できる。

エージェントの印象に関してはランダムとその他2つの付与、付与なしとの間に有意な差が見られた。対話中に的外れな感情付与が起きることは、エージェントの印象も悪くしてしまっていることがいえる。例えば、楽しい対話中に悲しい話し方をされるとその対話相手の印象が悪くなることは明らかである。また、付与なしと協調的付与との間にも差が見られた。これはエージェントという立場がはっきりしている中での対話であったため、ユーザに協調した感情付与がよりエージェントの印象を良くしたのだと考えられる。これより、ユーザに協調的付与によって実際にユーザが協調されているという感覚を持ったともいえる。

以上のことから今回検討した2つの付与方法はランダムな付与を行うより対話の楽しさやエージェントの人らしさ、印象に関してより良い評価をもたらすことがわかった。さらに、自律的付与を行った場合に付与なしとの差がみられたのは4つの評価指標のうち、対話の豊かさおよび楽しさの2つのみについてであるのに対し、協調的付与を行った場合は、4つの指標全てにおいて付与を行わない場合と有意な差がみられた。また、対話シナリオの中にはユーザの勝ち負けに関する対話が含まれており、この対話では協調的付与の効果がより大きく現れた。これよりユーザの感情に大きく関わるような対話において協調的付与がより有効である可能性が考えられるため、より詳細な検証が必要である。以上のことから、対話エージェントに対する感情付与方法として協調的付与がユーザの持つ印象に対して有効であることがわかった。これは、Acostaらのシステム[19]において、直前のユーザ発話から感情推定を行っていることと整合する。

## 5. むすび

本論文では、非タスク指向型の音声対話システムにおいて人間とシステム間のより良いインタラクションを実現するために、感情音声合成の利用がもたらす効果について検証を行い感情付与方法の検討を行った。

感情音声合成がもたらす効果に関しては、適切な感情付与を行ったときユーザの持つ印象が向上することが明らかとなった。しかしながら不適切な感情付与を行うことで評

価が低くなってしまったため、感情付与方法の検討の必要性も示唆された。次に感情付与方法の検討として、本論文では自律的付与と協調的付与の2つについて主観評価実験によって効果を検証した。実験結果から、協調的付与においては、4つの指標全てにおいて付与なしの場合と有意な差がみられ、付与による効果を確認することができた。しかしながら、実際の対話では対話相手との人間関係や立場が印象等に大きく影響することが考えられる。また、人は対話をする際、自分の発話と対話相手の発話の両方を考慮しながら応答の仕方を判断しているはずである。よって感情付与方法に関してはこれら2つの問題について検討・検証が必要であると言える。

今後の課題としては先に挙げたものとして、エージェントとの人間関係や立場、対話の内容に応じた付与方法の検討、システム発話とユーザ発話の両方を考慮した付与方法の検討である。さらに今回は感情推定を発話文のみから行っているが、これに関しても本来人は対話相手の発話内容と韻律から推定を行っている。これを実現するためには人が機械と話している場合の感情推定として、より小さな韻律の変化を考慮できる推定方法の検討が必要である。

#### 参考文献

[1] Apple: Siri, Apple (online), available from <https://www.apple.com/jp/ios/siri/> (accessed 2015-04-14).

[2] 吉村 健: シャベってコンシェルと言語処理, 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2012, No. 4, pp. 1-6 (2012).

[3] 東中竜一郎, 貞光九月, 内田 渉: R&D ホットコーナーソリューションシャベってコンシェルにおける質問応答技術, NTT 技術ジャーナル, Vol. 25, No. 2, pp. 56-59 (2013).

[4] 大西加奈子, 吉村 健: コンピュータとの自然な会話を実現する雑談対話技術, NTT DoCoMo テクニカル・ジャーナル, Vol. 21, No. 4, pp. 17-21 (2006).

[5] McTear, M. F.: Spoken dialogue technology: enabling the conversational user interface, *ACM Computing Surveys (CSUR)*, Vol. 34, No. 1, pp. 90-169 (2002).

[6] 徳久良子, 寺島立太: 雑談における発話のやりとりと盛り上がりとの関連, 人工知能学会論文誌, Vol. 21, pp. 133-142 (2006).

[7] Meguro, T., Higashinaka, R., Dohsaka, K., Minami, Y. and Isozaki, H.: Analysis of listening-oriented dialogue for building listening agents, *Proc. SIGDIAL*, pp. 124-127 (2009).

[8] 柴田雅博, 富浦洋一, 西口友美: 雑談自由対話を実現するための WWW 上の文書からの妥当な候補文選択手法, 人工知能学会論文誌, Vol. 24, No. 6, pp. 507-519 (2009).

[9] 吉田裕介, 萩原将文: 複数の言語資源を用いたユーモアを含む対話システム, 知能と情報 (日本知能情報ファジィ学会誌), Vol. 26, No. 2, pp. 627-636 (2014).

[10] Nishimura, R. and Nakagawa, S.: Response timing generation and response type selection for a spontaneous spoken dialog system, *Proc. Automatic Speech Recognition & Understanding Workshop (ASRU)*, pp. 462-467 (2009).

[11] 武藤ゆみ子, 高野弘二, 大良宏樹, 小林洋平, 山本知仁,

三宅美博: 音声対話インタフェースにおける発話タイミング制御とその評価, ヒューマンインタフェースシンポジウム 2007 論文集, pp. 639-642 (2007).

[12] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T. and Kitamura, T.: Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis, *Proc. EUROSPEECH*, pp. 2347-2350 (1999).

[13] Zen, H., Tokuda, K. and Black, A.: Statistical parametric speech synthesis, *Speech Communication*, Vol. 51, No. 11, pp. 1039-1064 (2009).

[14] Black, A. and Campbell, N.: Optimising selection of units from speech database for concatenative synthesis, *Proc. EUROSPEECH*, pp. 581-584 (1995).

[15] Yamagishi, J., Onishi, K., Masuko, T. and Kobayashi, T.: Acoustic modeling of speaking styles and emotional expressions in HMM-based speech synthesis, *IEICE Trans. Inf. & Syst.*, Vol. E88-D, No. 3, pp. 503-509 (2005).

[16] Tachibana, M., Yamagishi, J., Masuko, T. and Kobayashi, T.: Speech synthesis with various emotional expressions and speaking styles by style interpolation and morphing, *IEICE Trans. Inf. & Syst.*, Vol. E88-D, No. 11, pp. 2484-2491 (2005).

[17] Nose, T., Yamagishi, J., Masuko, T. and Kobayashi, T.: A style control technique for HMM-based expressive speech synthesis, *IEICE Trans. Inf. & Syst.*, Vol. E90-D, No. 9, pp. 1406-1413 (2007).

[18] Lee, A., Oura, K. and Tokuda, K.: Mmdagent A fully open-source toolkit for voice interaction systems, *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8382-8385 (2013).

[19] Acosta, J. C. and Ward, N. G.: Achieving rapport with turn-by-turn, user-responsive emotional coloring, *Speech Communication*, Vol. 53, No. 9, pp. 1137-1148 (2011).

[20] 徳久良子, 乾健太郎, 松本裕治: Web から獲得した感情生起要因コーパスに基づく感情推定, 情報処理学会論文誌, Vol. 50, No. 4, pp. 1365-1374 (2009).

[21] 中村 明: 感情表現辞典, 六興出版 (1979).

[22] 高野憲悟, 萩原将文: 感情関連語を用いた感情推定法の提案とニュースサイトのアクセス解析への応用, 日本感性工学会論文誌, Vol. 11, No. 3, pp. 495-502 (2012).

[23] 多田和彦, 矢野良和, 道木慎二, 大熊 繁: 感情遷移における急激な韻律特徴変化の検出による感情遷移判別法, 知能と情報: 日本知能情報ファジィ学会誌, Vol. 22, No. 1, pp. 90-101 (2010).

[24] 重永 實: 感情の判別分析からみた感情音声の特性, 電子情報通信学会論文誌 A, Vol. 83, No. 6, pp. 726-735 (2000).

[25] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 2, pp. 203-222 (2005).

[26] 東山昌彦, 乾健太郎, 松本裕治: 述語の選択嗜好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会発表論文集, pp. 584-587 (2008).

[27] 乾健太郎: 日本語評価語極性辞書, 東北大学 (オンライン), 入手先 (<http://www.cl.ecei.tohoku.ac.jp/>) (参照 2015-04-14).

[28] 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座: ChaSen-形態素解析器, 奈良先端科学技術大学院大学 (オンライン), 入手先 (<http://chasen-legacy.sourceforge.jp/>) (参照 2015-04-14).

[29] 日本音響学会: 研究用連続音声データベース (ASJ-JIPDEC), 日本音響学会 (オンライン), 入手先 (<http://research.nii.ac.jp/src/ASJ-JIPDEC.html>) (参照 2015-04-14).