

# ユーモア発言を行う対話型エージェントに 適した韻律的特徴の調査

令和4年度 卒業論文

日本大学 文理学部 情報科学科 宮田研究室

得田 舜介

# 概要

我々は、ユーザの発言中の単語を、その単語から発音が近く、意味が遠く、認知度が高い単語に聞き間違えるボケをするエージェントを提案してきた。しかし、先行研究ではユーモア発言の言語的特徴にのみ焦点をあてており、非言語的特徴については扱ってこなかった。そこで、対話型エージェントが発言を音声で出力を行う際に、対話型エージェントの韻律的特徴によって、ユーザの印象が変わる可能性があると考えられる。実際に、人は話者と発話速度と声の高さによって話者への印象が変化することが分かっている。そこで本稿では、エージェントによるユーモア発言の非言語的特徴に焦点をあて、特にユーモア表現に適した韻律的特徴に注目する。ユーモアのある発言を行う対話型エージェントと音声で対話を行うシステムを構築し、対話型エージェントがユーモアのある発言を行う際の音声の速さ・高さを変化させる対照実験を行うことで、どのような韻律的特徴をもつ音声かユーモア表現に適しているかを明らかにする。具体的には、対話型エージェントの音声は速さと高さの異なる音声を9種類設定した。実験の参加者は1つの種類の音声で対話を行った後に、アンケートに回答し、その結果を比較、分析した。実験の結果から、対話型エージェントの音声の速さと高さによる影響は人に依存する可能性があるということと、速い音声はユーモアのある発言を行うエージェントに適していない可能性があると考えられることが明らかになった。

# 目次

<b>第1章</b>	<b>序論</b>	<b>1</b>
1.1	研究の背景	2
1.2	研究の目的	2
1.3	本論文の構成	2
<b>第2章</b>	<b>ユーモア生成と音声表現に関する研究事例</b>	<b>3</b>
2.1	ユーモア生成に関する研究事例	4
2.2	音声表現に関する研究事例	5
<b>第3章</b>	<b>研究課題</b>	<b>7</b>
3.1	問題の定義	8
3.2	研究課題の設定	8
<b>第4章</b>	<b>提案手法</b>	<b>10</b>
4.1	アプローチ	11
4.2	韻律的特徴を変化させる対照実験	11
<b>第5章</b>	<b>音声を用いたボケるエージェントの実装</b>	<b>13</b>
5.1	システムの全体像	14
5.2	音声を用いたボケるエージェントの実装	14
5.2.1	音声認識部	14
5.2.2	ユーモア生成部	15
5.2.3	音声の合成・出力部	16
<b>第6章</b>	<b>評価実験</b>	<b>18</b>
6.1	実験の目的	19
6.2	実験の概要	19
6.3	実験の手順	20
6.3.1	事前準備	20
6.3.2	実験時	20
6.4	実験の結果	21
6.4.1	対話後アンケートの結果	21
6.4.2	対話後アンケートの分析結果	22

---

6.4.3	実験後アンケートの結果 . . . . .	22
6.5	考察 . . . . .	23
<b>第7章</b>	<b>結論</b>	<b>24</b>
	<b>謝辞</b>	<b>26</b>
	<b>参考文献</b>	<b>28</b>
	<b>研究業績</b>	<b>31</b>

# 目 次

5.1	システム全体の構成図 . . . . .	14
6.1	ユーモア性 (N=6) . . . . .	21
6.2	親しみやすさ (N=6) . . . . .	21
6.3	対話継続意欲 (N=6) . . . . .	22
6.4	(N=6) . . . . .	23

# 表 目 次

4.1	評価指標 . . . . .	11
4.2	音声の速さ・高さの特徴 . . . . .	12
6.1	アンケート内容 . . . . .	19

# 第1章 序論

## 1.1 研究の背景

我々の生活空間には人とエージェントが対話を行うシーンが浸透し始めている [1]. Apple の Siri[2], Google Assistant[3], Amazon の Alexa[4] などが搭載された製品を家庭内で利用するユーザも増えている. 今後は音声によるエージェントとの対話機会はますます増えていくと思われる. さらに多くのシーンで対話型エージェントが活躍するためには, ユーザが対話型エージェントに対して親しみやすさを感じられるようになることが重要であると考えられる. ユーザが対話型エージェントに対して親しみやすさを感じやすくするためには, 対話型エージェントにユーモア性をもたせることが重要であると考えられている [5][6][7]. この考えに基づいて, 対話型エージェントにユーモアのある発言を行わせる研究を行ってきた [8][9][10].

## 1.2 研究の目的

1.1 より, 既存研究では対話型エージェントにユーモアのある発言を行わせる研究を行ってきた. しかし, ユーモアのある発言の言語的特徴にのみ焦点をあてており, 非言語的特徴について扱ってこなかった. 本研究では, エージェントによるユーモアのある発言の非言語的特徴に焦点をあて, 特にユーモア表現に適した韻律的特徴に注目する. 対話型エージェントの音声の速さ・高さを変化させる対照実験を行うことで, どのような音声の速さと高さがユーモアのある発言に適しているかを明らかにすることを目的とする.

## 1.3 本論文の構成

本論文の構成は次のとおりである.

2章では, ユーモア生成と音声表現に関する研究事例について述べる.

3章では, 本論文における問題の定義と研究課題について述べる.

4章では, 本論文における提案手法を述べる.

5章では, 音声を用いたボケるエージェントに関する実装について述べる.

6章では, 音声を用いたボケるエージェントに関する実験・考察について述べる.

最後に7章にて, 本論文の結論を述べる.



## 第2章 ユーモア生成と音声表現に関する 研究事例

本章では、ユーモア生成と音声表現に関する研究事例について述べる。これらは、どのようにユーモアを生成するか、さらに生成したユーモアをどのような音声表現で出力するのかという点で本研究と関係している。2.1節では、ユーモア生成に関する研究事例について紹介する。2.2節では、音声表現に関する研究事例について紹介する。

## 2.1 ユーモア生成に関する研究事例

ユーモア生成に関する研究事例として文献 [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] が挙げられる。Stock らは、文章の各単語の頭文字を連結している略称を、略称は置換せずに各単語を置換することでユーモアを生成する試みである。具体的には、もとの単語群の頭文字とその文脈の間の意味の不一致を検出するシステムを作成し、もとの単語群の文脈の反語となるような頭字語を作成することでユーモアを生成している。単語の置換時には、もとの単語群と新たに生成された単語群の関係が、意味は遠く、音は近くなるようにしている [11]。Tinholt らは、文中の代名詞関係に着目することでユーモア文を生成する試みである。具体的には、文章の意味が曖昧な代名詞関係を強調することで成立するユーモアに着目し、文章の意味が曖昧な代名詞関係が面白くないか、面白くなる可能性があるかを区別するシステムを作成する。このシステムを利用することで、面白くなる可能性がある入力文に対して、代名詞が指す正しい対象語とは異なる対象語を発見し、強調することで実現しようとしている [12]。Valitutti らは、人間とコンピュータの対話において、意図せず生じるユーモアに着目し、このユーモアを意図的に再現することでユーモアを生成する試みである。具体的には、テキスト予測システムを模倣し、ユーザが3, 4文字のテキスト入力を行うと、ユーザが入力していると予測される英単語の中で、接頭辞が一致し、単語が出現する確率が上位の単語を出力する [13]。Lee らは、画像にユーモアのあるタイトルを付ける試みである。具体的には、画像に対するユーモアのあるタイトル1つ以上をデータセットとして機械学習をすることで、画像に対して自動的にユーモアのあるタイトルを生成する。加えて、地域やトレンド等も考慮することで、ユーモア性の向上を狙っている [14]。真下ら、Umetani らと、Haraguchi らは、おかしみの構造図に基づき、Web ニュースから記事内の感情情報を用いて、つかみ、本ネタ、オチの3段構成の流れでの漫才台本の自動生成を行っている。つかみでは、挨拶を兼ねた最初の笑いとおチへの話題提供を行う。具体的には、挨拶として漫才台本生成時の月の行事に関する身近な話題を行い、その後最初の笑いとして表情ボケを行っている。表情ボケは、Web ニュースの記事内の感情表現に合わせて、あえて逆の表情にロボットがなることによってユーモアを生成するものである。本ネタでは、ニュース記事の内容を読み上げてユーザに説明しながら、同時に様々なボケを挟むことでユーモアを生成するものである。具体的には、感情ボケ、言葉遊びボケ、対立ボケ、過剰ボケの4種類をニュース記事の1文につき最大1度行う。オチでは、ニュース記事の内容を1つのキーワードで完結に表現し、最後にそのキーワードをお題に自動生成した謎かけでユーモアを生成している。さらに、漫才上演中の観賞の補助として、ロボットの台詞を画面にロボットのアイコンと吹き出しで表示するシステムを導入した。また、別のロボットプラットフォームを用いて実装を行うこと

で、この提案手法は柔軟性と拡張性が実証された [15][16][17]. 吉田らは、漫才形式の対話文の自動生成システムを提案している. 入力された文章から単語を選び、それを音の近い単語や、その単語が修飾している句につながる別の単語に置換し、ボケを生成している. 具体的には、まず入力文中の名詞となる単語を抽出する. その後、抽出された名詞単語それぞれについて、Google Ngram を用いて出現頻度を取得し、一定閾値以上となった単語からランダム選択された単語を置換元単語とする. 最後に、置換元単語から置換するボケ単語を選定する. ボケ単語の選定には、事前に準備した辞書内から音の近い単語を抽出する方法、置換元単語を構成する文字を入れ替えて単語を生成する方法、文解析で取得した係受け関係を用いる方法の3種類を用いている [18]. Dybala らは、ユーザの発言中の語彙を、異なる用法で使用した駄洒落文を生成するエージェントを提案している. 単に対話を行うエージェントと生成した駄洒落を用いるエージェントで、インタラクションに関する質問を行い、分析・比較する評価実験と、エージェントに対するユーザの感情の分析を行った. 評価実験の結果、駄洒落を用いるエージェントは単に対話を行うエージェントと比較すると、インタラクションを継続したいなどの多くの点で高いスコアになった. 感情分析の結果、駄洒落を用いるエージェントとの対話は単に対話を行うエージェントとの対話と比較すると、感情変化が大きくポジティブな変化をしていた. このことから駄洒落を用いるエージェントはユーザの気分を良くする対話型エージェントであることが分かった [19]. 関は、ユーザ発言文の一部をシステムが擬似的に聞き間違えて聞き返すようなユーモア発言を行うエージェントを提案している. 具体的には、日本語版 Wikipedia と分類語彙表、さらには分散表現を用いた語彙検索によりユーモア表現の変換候補辞書を作成し、聞き間違いの用例調査に基づく単語変換方法を設計する. そして、徳島大学の在学生向け対話システムに、親近感の向上を目的に提案手法に基づくユーモア表現を行える機能を実装した. 提案手法の有効性を検証するために、他手法との比較評価を行った結果、音韻の類似度や出現頻度などの変換条件を設計することで、ユーモアの喚起に有効であることが確認できた. また、単数変換と複数変換の併用がユーモア表現に多様性を与えることや、複数変換は変換元の文字数の影響を受けにくいため様々な発話文に対応できることを確認できた [20].

ユーモア生成に関する既存研究として、文献 [8], [9], [10] が挙げられる. これらの研究事例は、エージェントが不適合解決モデル [21], [22] に基づいてユーモアの生成を行っている. ユーザの入力中の特定の単語を、その単語から意味が離れており、音が近く、認知度が高い単語に聞き間違えて聞き返すボケをするエージェントを提案している. これは、ユーザの入力中の特定の単語とエージェントが出力する単語の意味を離すことで不適合を生成し、生成した不適合を音が近いという関連性をもたせ、解決することで、ユーモアを生み出すというものである [8][9][10].

## 2.2 音声表現に関する研究事例

音声表現に関する研究事例として文献 [23], [24], [25], [26], [27], [28] が挙げられる. Seaborn らは、エージェントの音声は人間とエージェントの対話において、重要な特徴に

なりつつあると考えている。調査により、人々がどのようにエージェントと対話し、反応し、エージェントの音声に影響されるかを理解することが、重要であることを明らかにした [23]。内田らは、話者の発話速度と高さによる聞き手への影響を調査している。具体的には、日本語話者4名と英語話者4名の音声を収集し、5段階の発話速度・高さの音声を作成した上で、発話速度と高さによる話者の性格印象にどのような影響を与えるか調査している。その結果、人は話者の発話速度と声の高さによって、話者への印象が変化することを明らかにした [24]。橋本らは、人工音声を使用して実験を行い、発話速度と高さが特性推論に及ぼす影響を検討している。具体的には、3段階の発話速度・高さの音声をテキスト読み上げソフトを利用して作成し、音声状態印象を、SD尺度12項目を用いて、7件法で測定した。また、性格特性に関する20項目に対して、1点から5点までの5件法で回答を求め、評価を行った。その結果、音声のパラ言語的特徴の違いによって、音声状態印象に影響を与え、異なる性格特性が推論されることが示された [25]。Tolmeijerらは、234人の参加者を対象とした探索的オンライン実験において、2段階の音声の高さと男性・女性の性別がステレオタイプのな特徴の帰属と信頼の形成に及ぼす影響を分析し、2つの異なるタスクにおけるジェンダーステレオタイプの調査をしている。その結果、音声アシスタントに対して暗黙のステレオタイプ化が起こることが明らかになった。また、性別が曖昧な音声と性別が設定された音声とでは、信頼度に有意な差がないことを明らかにした [26]。Iizukaらは、人間の自然発話で訓練された対話エージェントの合成音声、人間の聴者に与える影響について検討している。具体的には、自然発話を基にした合成音声の対話型エージェントとの対話した場合、一般的な音声コーパスを基にした合成音声の対話型エージェントと比べて、人間はより社会的な反応を示すと仮定をした。合成音声の作成方法による比較実験の結果、自然発話を基にした合成音声の対話型エージェントと対話した被験者ほど、応答時間が短く、バックチャンネル数が多い傾向が見られた。また、アンケート結果より、対話型エージェントとの対話を、より人間との対話に近いと評価する傾向が見られた [27]。Guoらは、sequence to sequenceの読み上げ音声合成システムを構築することを目的として研究している。具体的には、まず、音声エージェント用に設計された音声コーパスを、録音品質と話し方の両方を考慮する新しい録音方法で構築する。次に、会話文脈を考慮したend to endの読み上げ音声合成システムを提案する。提案手法は、会話文脈に応じた自然な韻律を生成し、発話レベルと会話レベルの両方において有意な嗜好性を獲得することが分かった [28]。

## 第3章 研究課題

本章では、本研究における問題の定義と研究課題について述べる。

### 3.1 問題の定義

ユーザが対話型エージェントに対して親しみやすさを感じやすくするために、対話型エージェントにユーモア性をもたせることが有効であると考えられている [5][6][7]。2.1 節で述べたように、ユーモアを生成する研究は多く行われている。既存研究ではエージェントのユーモア発言により、“笑い”を引き起こすことで親しみやすさを向上させるアプローチに取り組んでいる [8][9][10]。具体的には、ユーザの発言中の単語を用いて、その単語から意味が遠く、発音が近く、認知度が高い単語に言い換えるようなボケをする対話型エージェントを提案してきた。これらの既存研究では、対話型エージェントのユーモアのある発言の言語的特徴にのみ焦点を当てており、ユーザと対話型エージェントのインタラクションはテキストに限定していた。

一方、近年インタラクションを音声で行う対話型エージェントは多く普及しており [2][4][3]、対話型エージェントの非言語的特徴に関する研究の必要性が高まっている。この状況を踏まえると、言語的特徴量のみならずフォーカスしていた先行研究を拡張し、ユーモア発言を行う対話型エージェントの非言語的特徴に着目することには大きな価値があると考えられる。既存研究では、対話型エージェントが発言を音声出力で行う際に、2.2 節で述べたように音声の韻律的特徴により聞き手に影響を与えることまでは明らかにされている。しかし、対話型エージェントがユーモアのある発言を音声出力で行う際、どのような韻律的特徴があるとユーザがよりユーモアを感じられるのか明らかにされていない。この問題を明らかにすることで、対話型エージェントがユーモア表現を行う際に、ユーモア性がさらに向上し、親しみやすさ、対話継続意欲などの観点も同様に向上させることができると考えられる。

### 3.2 研究課題の設定

3.1 節で述べたように、対話型エージェントがユーモア発言を音声出力で行う際、どのような韻律的特徴がユーモア表現に適しているのか明らかにされていない。この問題を解決するために、対話型エージェントがどのような韻律的特徴でユーモア表現を音声出力で行うべきか明らかにする必要がある。2.2 節で述べたように、韻律的特徴により聞き手に様々な影響を与えることは明らかになっている [23][24][25][26][27][28]。しかしこれらの調査では、対話型エージェントがユーモアのある発言を音声出力で行う際の韻律的特徴に着目した取り組みは行われていない。そこで本研究では、対話型エージェントがユーモアのある発言を音声出力で行う際の韻律的特徴に着目し、適した韻律的特徴を明らかにすることを目的とする。

上記を踏まえ、本研究では対話型エージェントがユーモアのある発言を音声出力で行う際に、ユーモア表現に適した韻律的特徴を明らかにすることを研究課題として設定する。

具体的には、対話型エージェントの韻律的特徴を変化させる対照実験を行うことで、ユーモア発言を行う対話型エージェントに適した韻律的特徴を明らかにする取り組みを行う。対話型エージェントがユーモアのある発言を音声出力で行う際に、音声の速さ・高さを変化させる対照実験を行い、適した音声の速さ・高さを明らかにすることを旨とする。

## 第4章 提案手法



本章では、本論文における提案手法を述べる。

## 4.1 アプローチ

3.2節で定義した研究課題を達成するために、本研究では仮説を立てる。

本研究で提案するシステムは、対話型エージェントがユーモアのある発言を、特定の韻律的特徴で音声出力をする。それに対する表4.1の評価指標について検討する必要があると考えられる。

表 4.1: 評価指標

項目	評価指標の名称	内容
1	ユーモア性	ユーモア発言にどの程度ユーモアを感じたか
2	親しみやすさ	エージェントにどの程度親しみやすさを感じたか
3	対話継続意欲	エージェントと対話を続けたいと感じたか

本研究では、下記のように仮説を立てた。

- 対話型エージェントがユーモアのある発言を特定の韻律的特徴で行うことで、ユーモア性・親しみやすさ・対話継続意欲が向上する

韻律的特徴を変化させる対照実験を行い、韻律的特徴によって、ユーモア性・親しみやすさ・対話継続意欲がどのように変化するか調査する。

## 4.2 韻律的特徴を変化させる対照実験

4.1節で述べた仮説に基づき、対話型エージェントがユーモアのある発言を音声出力をする際の、韻律的特徴を変化させる対照実験を行う。韻律的特徴は表4.2にある特徴で、9手法の音声を作成する。本実験で変化させる韻律的特徴は、音声の速さと高さである。

表 4.2: 音声の速さ・高さの特徴

手法	音声の速さ・高さ
手法1	通常の時速・高さ
手法2	速さは通常の時速より速く、高さは通常の高さ
手法3	速さは通常の時速より遅く、高さは通常の高さ
手法4	速さは通常の時速、高さは通常の高さより高い
手法5	速さは通常の時速、高さは通常の高さより低い
手法6	速さは通常の時速より速く、高さは通常の高さより高い
手法7	速さは通常の時速より速く、高さは通常の高さより低い
手法8	速さは通常の時速より遅く、高さは通常の高さより高い
手法9	速さは通常の時速より遅く、高さは通常の高さより低い

手法それぞれの速さと高さのパラメータの詳細については後ほど記載する。

## 第5章 音声を用いたボケるエージェントの実装

本章では、音声を用いたボケるエージェントに関する実装について述べる。

## 5.1 システムの全体像

本研究で実装されたエージェントは、大きく分けると音声認識部、ユーモア生成部、音声合成・出力部の3つのシステムから構成されている。システム全体の構成を図5.1に示す。番号の順に処理が行われ、本研究では新たに、音声の入出力、音声認識、サーバ通信によるテキストの送受信、音声合成の実装した。

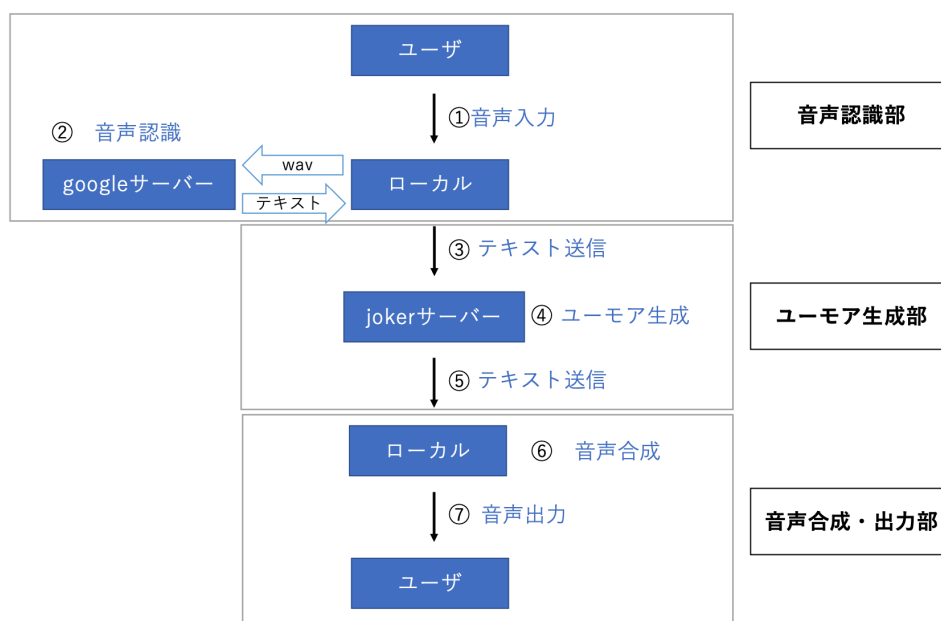


図 5.1: システム全体の構成図

## 5.2 音声を用いたボケるエージェントの実装

### 5.2.1 音声認識部

ユーザーからの音声入力を認識し、テキストに変換する実装について述べる。音声認識では、PythonのSpeechRecognitionライブラリを使用した。ユーザーからの音声入力の開始後、ユーザーが音声入力を終了し、沈黙になるまで音声入力を受け付ける。SpeechRecognitionを用いて、得られた音声をテキストに変換し保存する。

## 5.2.2 ユーモア生成部

本研究では、既存研究で提案してきたユーザの入力文中の名詞を聞き間違えて聞き返すというユーモア生成方法を利用する [8][9][10]. ユーモアを生成する実装は大きく分けると、ローカルとユーモアを生成するサーバ (joker サーバとする) で通信をする実装, ユーザの入力文中のどの名詞を聞き間違えるかを決める実装, その名詞をどのような単語に聞き間違えるかを決める実装の3つに分かれる.

### 5.2.2.1 サーバ通信

ローカルと joker サーバで通信をする実装について述べる. この実装では, 5.2.1 の実装によって得られたテキストを joker サーバに送信し, 生成されたユーモアのある発言のテキストをローカルに返す通信を行う. 具体的には, 通信には HTTP を用い, POST メソッドでテキストの受け渡しを行った. ローカルから joker サーバに送信されたテキストを用いて 5.2.2.2, 5.2.2.3 に記述の方法でユーモアのある文章を生成する. その後, 生成されたテキストをローカルに返すことで, ユーモアのある発言のテキストを利用できる.

### 5.2.2.2 置換元単語の決定

ユーザの入力文中の各名詞について, どの程度文のトピックに近いかを表すスコア  $s_t$  を算出し,  $s_t$  が最も高い名詞を置換元単語とする. 入力文中の全名詞からなる置換元候補群を作成する. 置換元候補群内の各名詞の  $s_t$  は下記のように算出する.  $i$  は  $s_t$  を算出する置換元候補群の名詞の位置,  $d_{s(k)}$  は算出する置換元候補群の名詞と  $k$  番目の名詞の概念距離,  $n$  は置換元候補群の名詞数となっている.

$$s_{t(i)} = \frac{1}{n-1} \sum_{k=1, k \neq i}^n d_{s(k)} \quad (5.1)$$

### 5.2.2.3 ボケ単語の決定

Wikipedia 日本語記事全文を MeCab [29] で形態素解析し, 不要品詞を除去したものをコーパスとする. このコーパス内の単語全てをボケ単語候補とする. 置換元単語に対して, 意味が遠く, 発音が近く, 認知度が高い単語をボケ単語候補の中から選定する.

意味が遠い単語を発見するために, 置換元単語とボケ単語候補に関する類似度  $u_{(k)}$  の逆数を用いて, 意味の遠さを表すスコア ( $s_s$ ) を算出する.  $s_s$  は下記のように算出する.

$$s_s = \frac{1}{n} \sum_{k=1}^n \frac{1}{1 + u_{(k)}} \quad (5.2)$$

発音が近い単語を発見するために、置換元単語とボケ単語候補それぞれの母音を文字列にし、その文字列間の編集距離  $d_e$  を用いて、発音の近さを表すスコア ( $s_p$ ) を算出する。 $s_p$  は下記のように算出する。

$$s_p = \frac{1}{1 + d_e} \quad (5.3)$$

認知度が高い単語を発見するために、日本語 Wikipedia 記事全文での出現回数  $f$  を用いて、認知度の高さを表すスコア ( $s_r$ ) を算出する。 $s_r$  は下記のように算出する。

$$s_r = \log f \quad (5.4)$$

算出した  $s_s$ ,  $s_p$ ,  $s_f$  を正規化・重み付けした合計値を  $s_h$  とする。すなわち、 $s_h$  は下記のように求める。 $w_s, w_p, w_f$  は重み係数である。

$$s_h = w_s s_s + w_p s_p + w_f s_f \quad (5.5)$$

$s_h$  が最も高い単語をユーモアのある発言に使用するボケ単語とする。ここまでの処理で決定したボケ単語を用いた、ユーザとエージェントの対話例を示す。

ユーザ      愛車がピカピカになって嬉しい

エージェント え?会社?

ユーザ      朝起きるのが苦手だ

エージェント え?かさ?

ユーザ      写真を撮りたいのに雨が降ってきて悲しい

エージェント え?家臣?

### 5.2.3 音声の合成・出力部

3.2より本研究の研究課題は、対話型エージェントがユーモアのある発言を音声出力する際に、ユーモア表現に適した韻律的特徴を明らかにすることである。音声の速さ・高さを変化させる対照実験を行うことで、韻律的特徴による効果を明らかにする。

#### 5.2.3.1 基準とする音声特徴

本研究では、基準とする音声特徴を設定する際に国内で高いシェアを誇っている Apple 製品に搭載されている音声アシスタントの Kyoko を参考に設定した。Kyoko は macOS・iOS に搭載されている日本語のテキスト読み上げ機能用の音声アシスタントである。Kyoko が“おはようございます”という文章を読み上げた際の音声の速さと高さを Praat を用いて計測したところ、音声の開始から終了までの時間が 1.04 秒、平均基本周波数が 224Hz で読み上げていた。この音声の開始から終了までの時間と平均基本周波数を基準にする。

### 5.2.3.2 音声合成

音声合成には, Open JTalk[30] を利用した. Open Jtalk は, 日本語テキストに基づいて音声を作成する HMM テキスト音声合成システムである. joker サーバから返されたユーモアのある発言のテキストを, Open JTalk を用いて音声合成することで, エージェントの音声を作成した. 音声合成する際に, パラメータを調整することで, 様々な韻律的特徴の音声を作成した. 作成する音声の速さ・高さの種類は, 表 4.2 の 9 種類である. 通常の速さの音声より速い音声は, 音声の開始から終了までの時間が通常の速さの音声より 0.5 倍の音声である. 通常の速さの音声より遅い音声は, 音声の開始から終了までの時間が通常の速さの音声より 2 倍の音声である. 通常の高さの音声の平均基本周波数を 100% とした際に, 通常の高さの音声より高い音声は, 平均基本周波数が 156.25% の音声とする. また, 通常の高さの音声より低い音声は, 平均基本周波数が 64% の音声とする. これらの値は, 内田らの研究 [24] を参考に設定した.

### 5.2.3.3 音声出力

5.2.3.2 で作成した音声を再生することで, 音声出力をする. 既存研究 [31] より, ロボットは 0 秒で反応するより, 1 秒反応を遅らせた方がユーザに好まれることが明らかにされている. この研究をもとに, ユーザの音声入力終了からの時間を計測し, 音声入力終了してから 1 秒後に音声出力をする.

## 第6章 評価実験



本章では、音声を用いたボケるエージェントに関する評価実験・考察について述べる。

## 6.1 実験の目的

本システムは、ユーザの音声入力に対して、聞き間違えるボケを音声出力する対話型エージェントである。しかし、対話型エージェントがユーモアのある発言を音声出力で行う際に、よりユーモアを感じる韻律的特徴がどのようなものであるか明らかになっていないという問題がある。この問題を明らかにするために、対話型エージェントがユーモアのある発言を音声出力する際の、音声の速さ・高さを変化させる対照実験を行う。本実験の目的は、対話型エージェントの音声の速さ・高さの違いによって、ユーザの対話型エージェントのユーモア表現への感じ方がどのように変化するか明らかにすることである。評価指標は、表 4.1 にある 3 項目で、対話型エージェントとの対話に関してアンケートを行い、結果から分析をする。

アンケートは、実験前アンケート、対話後アンケート、実験後アンケートの 3 種類のアンケートを行う。アンケートの詳細を表 6.1 に示す。Q1～Q3 が実験前アンケート、Q4～Q9 が対話後アンケート、Q10～Q16 が実験後アンケートとなっている。

表 6.1: アンケート内容

質問	質問内容	回答方法
Q1	対話型エージェントを普段から利用していますか？	はい：いいえ
Q2	どのくらいの頻度で利用していますか？	1日1回以上：週に2～6回：月1～2回
Q3	雑談できる対話型エージェントに対してどのような期待をもっていますか？	自由記述
Q4	音声は聞き取りやすいと感じましたか？	7:とても感じた～1:全く感じなかった
Q5	対話型エージェントの音声の速さをどのように感じましたか？	7:とても速いと感じた～1:とても遅いと感じた
Q6	対話型エージェントの音声の高さをどのように感じましたか？	7:とても高いと感じた～1:とても低いと感じた
Q7	対話型エージェントのユーモアのある発言にユーモアを感じましたか？	7:とても感じた～1:全く感じなかった
Q8	対話型エージェントに親しみやすさを感じましたか？	7:とても感じた～1:全く感じなかった
Q9	対話型エージェントと対話を続けたいと感じましたか？	7:とても感じた～1:全く感じなかった
Q10	もっともユーモアを感じた手法はどれですか？	手法1～9から1つ選択
Q11	対話型エージェントの音声の速さ、高さの違いについて、ユーモアの観点で感じたことを教えてください。	自由記述
Q12	もっとも親しみやすさを感じた手法はどれですか？	手法1～9から1つ選択
Q13	対話型エージェントの音声の速さ、高さの違いについて、親しみやすさの観点で感じたことを教えてください。	自由記述
Q14	もっとも対話を続けたいと感じた手法はどれですか？	手法1～9から1つ選択
Q15	対話型エージェントの音声の速さ、高さの違いについて、対話継続意欲の観点で感じたことを教えてください。	自由記述
Q16	何か感じたことがあればご記入をお願いします。	自由記述

## 6.2 実験の概要

本研究で行う対照実験では、対話型エージェントがユーモアのある発言を音声出力する際の、音声の速さ・高さを変化させる。実験時には、実験場所には実験参加者のみが行

き、本システムを起動して対話を行なってもらう。実験参加者は、20代の学生一名である。本実験で用いる韻律的特徴の手法は、表4.2の9手法である。

## 6.3 実験の手順

### 6.3.1 事前準備

事前準備は下記の手順で行う。

#### Step 1:

実験者は実験参加者に実験について説明を行う。

#### Step 2:

実験参加者は実験前アンケートに答える。

Step 1での説明の際に、実験時までには、実験参加者には対話型エージェントに話かける文章を9文用意してもらい、実験者は事前にどのような出力になるか確認をする。

### 6.3.2 実験時

実験時は下記の手順で行う。

#### Step 1:

実験者は本システムが実装されているpcを実験実施場所に設置し、サーバー起動する。

#### Step 2:

実験参加者は実験実施場所に着席し、マニュアルを確認する。

#### Step 3:

実験参加者はヘッドセットを着用し、本システムを起動する。

#### Step 4:

実験参加者は手法を選択する。

#### Step 5:

実験参加者は対話型エージェントと対話をする。

#### Step 6:

実験参加者は対話した手法に関する対話後アンケートに回答する。

#### Step 7:

実験参加者はStep 3～Step 6に関して、全ての手法で対話しアンケートを回答する。

Step 8:

実験参加者は実験後アンケートに回答する。

Step 2にあるマニュアルでは、本システムのエージェントについての説明、実験時に行ってもらった操作、全ての対話が終了した後に行う実験後アンケートの内容について記載する。

## 6.4 実験の結果

### 6.4.1 対話後アンケートの結果

表 4.1 の評価指標について、アンケートの Q7, Q8, Q9 で調査した。手法ごとのユーモア性、親しみやすさ、対話継続意欲の結果を図 6.1, 図 6.2, 図 6.3 に示す。

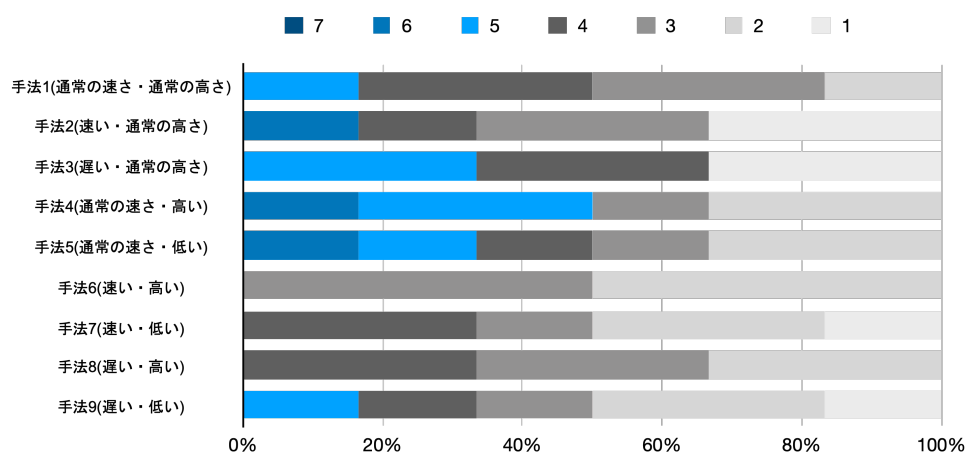


図 6.1: ユーモア性 (N=6)

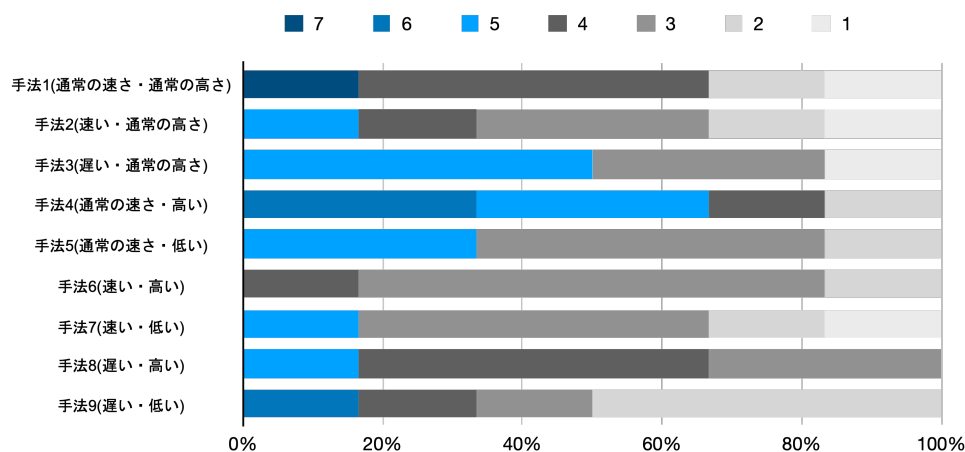


図 6.2: 親しみやすさ (N=6)

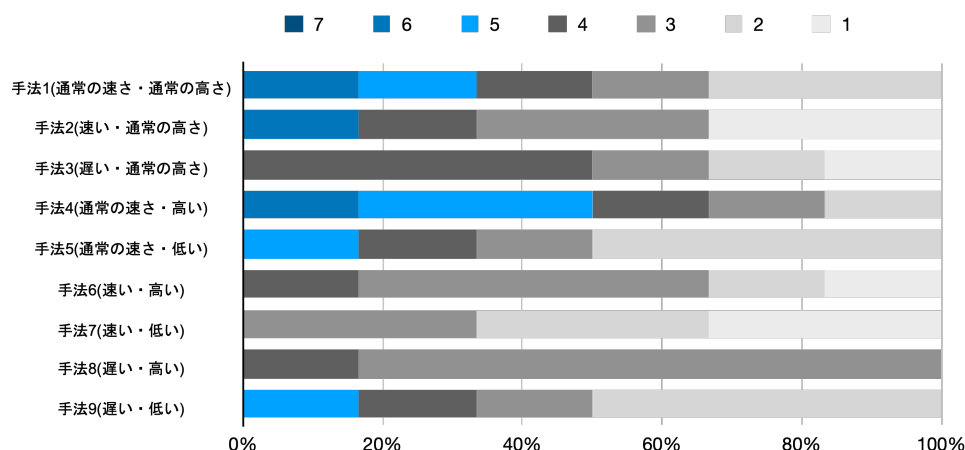


図 6.3: 対話継続意欲 (N=6)

### 6.4.2 対話後アンケートの分析結果

ユーモア性、親しみやすさ、対話継続意欲に関して、ウィルコクソンの順位和検定のホルム法、一元分散分析、二元分散分析を用いて分析を行った。

ユーモア性に関して、ウィルコクソンの順位和検定のホルム法で分析した結果、どの手法間においても有意差は確認されなかった。一元分散分析で分析した結果、手法によって差があると判断できず、人によって差がある可能性があることが明らかになった。二元分散分析で分析した結果、速さと高さには交互作用はなく、速さによる差がある可能性があり、高さによる差はないということが明らかになった。

親しみやすさに関して、ウィルコクソンの順位和検定のホルム法で分析した結果、どの手法間においても有意差は確認されなかった。一元分散分析で分析した結果、手法によって差があると判断できず、人によって差がある可能性があることが明らかになった。二元分散分析で分析した結果、速さと高さには交互作用はなく、速さによる差がある可能性があり、高さによる差はないということが明らかになった。

対話継続意欲に関して、ウィルコクソンの順位和検定のホルム法で分析した結果、どの手法間においても有意差は確認されなかった。一元分散分析で分析した結果、手法によって差があると判断できず、人によって差があると判断できないことが明らかになった。二元分散分析で分析した結果、速さと高さには交互作用はなく、速さによる差がある可能性があり、高さによる差はないということが明らかになった。

### 6.4.3 実験後アンケートの結果

もっともユーモア性、親しみやすさ、対話継続意欲を感じた手法について、実験後アンケートのQ10、Q12、Q14で調査した。結果を図6.4に示す。

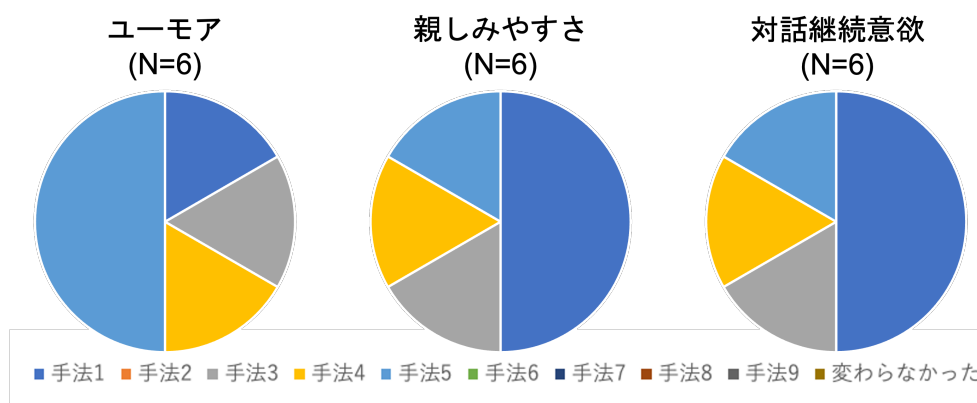


図 6.4: (N=6)

## 6.5 考察

対話後アンケートの結果より考察を行う。ユーモア性、親しみやすさ、対話継続意欲の観点において、ウィルコクソンの順位和検定のホルム法で分析した結果より、どの手法間においても有意差は確認されなかった。また、ユーモア性、親しみやすさにおいて、一元分散分析で分析した結果より、人によって差がある可能性があることが明らかになった。この2点より、音声の速さと高さによる影響は、人に依存する可能性があると考えられる。

二元分散分析で分析した結果より、ユーモア性、親しみやすさ、対話継続意欲の観点において、速さによって差がある可能性があるということが明らかになった。また、実験後アンケートの結果より、ユーモア性、親しみやすさ、対話継続意欲の観点において、速い音声は選ばれなかった。このことから、速い音声はユーモアのある発言を行うエージェントに適していない可能性があると考えられる。

## 第7章 結論

2.2節より、対話型エージェントが発言を音声出力で行う際の音声の韻律的特徴によって、聞き手に影響を与えることまでは明らかにされている。しかし、対話型エージェントがユーモアのある発言を音声出力で行う際、どのような韻律的特徴があるとユーザがよりユーモアを感じられるのか明らかにされていない。そこで、本研究では、ユーモアのある発言を行う対話型エージェントの、音声の速さ・高さを変化させる対照実験を行うことで、音声の速さと高さによる影響を調査した。具体的には、ユーザの音声入力に対して、音声認識をし、そのユーモアのある返答を音声合成し、出力するシステムを作成した。出力する音声の速さ・高さを変化させ、対照実験を行った。実験では、対話型エージェントの音声の速さ・高さに対する、ユーモア性、親しみやすさ、対話継続意欲の観点について、評価を行った。

実験時に行ったアンケートの結果より、音声の速さと高さによる影響は人に依存する可能性があるということ、そして、速い音声はユーモアのある発言を行うエージェントに適していない可能性があると考えられることが明らかになった。しかしながら、実験の参加者は少なかつたため、分析の精度については高くないと考えられる。上記を踏まえ、実験を引き続き行い、実験の参加者を増やして分析を再度行う必要があると考えられる。

アンケートの自由記述より、本稿で利用したユーモアのユーモア性についての意見がいくつかあった。このことから、本稿で行った実験で利用するユーモアの生成方法の変更や、うユーモア表現を増やすことで、ユーモア表現を行う多くの対話型エージェントに対して、有効である韻律的特徴を調査することができると考えられる。

# 謝辭



---

本研究と論文の作成にあたり、ご指導を頂いた指導教員の宮田章裕教授に感謝申し上げます。本研究を行うに当たって研究内容の取り組みにご協力いただいた、日本大学の大串旭さん、大西俊輝さん、呉健朗さんに感謝申し上げます。

## 参考文献

- [1] Daniel McDuff and Mary Czerwinski. Designing emotionally sentient agents. *In Communications of the ACM*, Vol. 61, No. 12, pp. 74–83, 2018.
- [2] Apple inc.: siri<https://www.apple.com/ios/siri/>(last visited on 2022/1/16).
- [3] Google llc.: Google assistant<https://assistant.google.com/>(last visited on 2022/5/20).
- [4] Amazon inc.: Alexa<https://www.amazon.co.jp/meet-alexa/>(last visited on 2022/5/20).
- [5] Peter Khooshabeh, Cade McCall, Sudeep Gandhe, Jonathan Gratch, and James Blascovich. Does it matter if a computer jokes. In *CHI EA '11: CHI '11 Extended Abstracts on Human Factors in Computing Systems*, pp. 77–86, 2011.
- [6] Ewa Luger and Abigail Sellen. Like having a really bad pa : the gulf between user expectation and experience of conversational agents. *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5286–5297, 2016.
- [7] Q. Vera Liao, Muhammed Mas-Ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N. Sadat Shami, and Werner Geyer. All work and no play? conversations with a question–and–answer chatbot in the wild. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, No. 3, pp. 1–13, 2018.
- [8] 呉健朗, 中原涼太, 長岡大二, 中辻真, 宮田章裕. ボケて返す対話型エージェント. 日本バーチャルリアリティ学会論文誌, Vol. 23, No. 4, pp. 231–238, 2018.
- [9] Kenro Go, Toshiki Onishi, Asahi Ogushi, and Akihiro Miyata. Conversational agents replying with a manzai-style joke. Proc. 33rd Australian Conference on Human-computer-interaction (OzCHI ' 21), pp. 221–231, 2021.
- [10] 呉健朗, 大西俊輝, 大串旭, 宮田章裕. ノリツッコミを行う対話型エージェント. 情報処理学会インタラクシオン 2022 論文集, pp. 39–47, 2022.

- 
- [11] Oliviero Stock and Carlo Strapparava. Getting serious about the development of computational humor. In *In International Joint Conference on Artificial Intelligence*, pp. 59–64, 2003.
- [12] Hans Wim Tinholt and Anton Nijholt. Computational humour: Utilizing crossreference ambiguity for conversational jokes. In *International Workshop on Fuzzy Logic and Applications*, Vol. 4578, pp. 477–483, 2007.
- [13] Alessandro Valitutti, Hannu Toivonen, Oskar Gross, and Jukka Mikael Toivanen. Decomposition and distribution of humorous effect in interactive systems. *Artificial Intelligence of Humor*, the AAAI Fall Symposium Series, pp. 96–100, 2012.
- [14] Dongkeon Lee, Seung-Ho Han, Kyo-Joong Oh, and Ho-Jin Choi. A temporal community contexts based funny joke generation. In *In Mobile Data Management*, 2017.
- [15] 真下遼, 梅谷智弘, 北村達也, 灘本明代. 文の感情を考慮した漫才ロボット台本自動生成手法の提案. 代7回データ工学と情報マネジメントに関するフォーラム, pp. F4–4, 2015.
- [16] Tomohiro Umetani, Satoshi Aoki, Kazuhiro Akiyama, Ryo Mashimo, Tatsuya Kitamura, and Akiyo Nadamoto. Scalable component-based manzai robots as automated funny content generators. *Robotics and Mechatronics*, Vol. 28, No. 6, pp. 862–869, 2016.
- [17] Kazuki Haraguchi, Kazuki Yane, Akira Sato, Eiji Aramaki, Isao Miyashiro, and Akiyo Nadamoto. Chat-type manzai application: mobile daily comedy presentations based on automatically generated manzai scenarios. In *Proceedings of the 18th International Conference on Advances in Mobile Computing Multimedia*, pp. 143–152, 2020.
- [18] 吉田裕介, 萩原将文. 漫才形式の対話文自動生成システム. 第13回日本感性工学大会, 2011.
- [19] Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka, and Kenji Araki. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Ambient Intelligence and Smart Environments*, Vol. 2, No. 1, pp. 31–48, 2010.
- [20] 関陽介. 用例に基づく疑似的な聞き間違いによりユーモア表現を生成する対話システム. 人工知能学会論文誌, Vol. 36, No. 2, pp. C–KA3\_1–10, 2021.
- [21] Seana Coulson and Robert F. Williams. Hemispheric asymmetries and joke comprehension. *Neuropsychologia*, Vol. 43, No. 1, pp. 128–141, 2005.

- 
- [22] Andrea C Samson, Christian F Hempelmann, Oswald Huber, and Stefan Zysset. Neural substrates of incongruity-resolution and nonsense humor. *Neuropsychologia*, Vol. 47, No. 4, pp. 1023–1033, 2009.
- [23] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. Voice in human-agent interaction: A survey. *ACM Computing Surveys*, Vol. 54, No. 81, pp. 1–54, 2021.
- [24] 内田照久, 中畝菜穂子. 声の高さと発話速度が話者の性格印象に与える影響. *The Japanese Journal of Psychology*, Vol. 75, No. 5, pp. 397–406, 2004.
- [25] 橋本和奈実, 古谷健. 発話速度と声の高さが特性推論に及ぼす影響—二段階推論仮説に基づいて—. *Japanese Journal of Applied Psychology*, Vol. 45, No. 1, pp. 15–25, 2019.
- [26] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. Female by default? – exploring the effect of voice assistant gender and pitch on trait and trust attribution. *CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, No. 455, pp. 1–7, 2021.
- [27] Takahisa Iizuka and Hiroki Mori. How does a spontaneously speaking conversational agent affect user behavior? *IEEE Access*, Vol. 10, pp. 111042–111051, 2022.
- [28] Haohan Guo, Shaofei Zhang, Frank K. Soong, Lei He, and Lei Xie. Conversational end-to-end tts for voice agent. In *Sound (cs.SD); Audio and Speech Processing (eess.AS)*, 2020.
- [29] Mecab: Yet another part-of-speech and morphological analyzer : <http://taku910.github.io/mecab/> (last visited on 2022/5/19).
- [30] Open jtalk : <https://open-jtalk.sp.nitech.ac.jp> (last visited on 2023/01/10).
- [31] 志和敏之, 神田崇行, 今井倫太, 石黒浩, 萩田紀博, 安西祐一郎. 対話ロボットの反応時間と反応遅延時における間投詞の効果. *日本ロボット学会誌*, Vol. 27, No. 1, pp. 87–95, 2009.

# 研究業績

## 査読付き国際会議

- (1) Kenro Go, Toshiki Onishi, Asahi Ogushi, Shunsuke Tokuda and Akihiro Miyata: An Evaluation of Motivation for Conversational Agents Replying to Manzai-Style Jokes. Transactions of the Virtual Reality Society of Japan, (2023 掲載予定).
- 

## 研究会・シンポジウム

- (1) 得田舜介, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 対話型エージェントのユーモア表現に適した韻律的特徴の調査, 情報処理学会シンポジウム論文集, マルチメディア, 分散, 協調とモバイル (DICOMO ' 22), Vol.2022, pp.570-575 (2022).
  - (2) 得田舜介, 呉健朗, 田中柊羽, 大西俊輝, 大串旭, 宮田章裕: 英語でボケる発言を行うエージェントの基礎検討, 情報処理学会インタラクション 2022 論文集, pp.291-294 (2022).
- 

## 受賞

- (1) 情報処理学会マルチメディア、分散、協調とモバイルシンポジウム (DICOMO2022) ヤングリサーチャー賞, 対話型エージェントのユーモア表現に適した韻律的特徴の調査, 受賞者: 得田舜介 (2022年7月).