

早稲田大学大学院 基幹理工学研究科

博士論文概要

論文題目

対話音声合成のための音声表現の多様化に関する研究

Studies on Diversification of Speech Expressions
for Conversational Speech Synthesis

申請者

岩田 和彦
Kazuhiko IWATA

2022年12月

音声対話は、人にとっては最も自然なコミュニケーションの手段である。そのため、古くから究極のヒューマン・マシン・インタフェースとして待望され、機械との間で声による情報の入出力を可能にするための技術として発展してきた。近年、音声対話技術は、基礎研究から実用化の段階へと急速な進展を遂げ、実用的なサービスにも利用されるようになってきた。そして今や、ロボットを話し相手にして、対話を楽しむことができるようにまでなっている。

人同士のコミュニケーションでは、言語情報だけでなく、話し手の意図や態度、感情などのパラ言語・非言語情報も含めた情報交換が行われる。声の表情は、感情の表出以外にも、話し手と聞き手との関係や、話題、場の雰囲気など、発話の背景にある様々な要因に応じて変化する。本論文では、パラ言語・非言語情報が音声を媒体として伝達される際に韻律や声質（ここでは、話者の個人差に起因する聞こえの違いは含めない）に現れる特徴を、「音声表現」と呼ぶことにする。

ところが、話し相手のロボットが、従来のようにどのような場面でも常に同じ口調でしゃべっていたのでは、話は一向に弾まないだろう。話の内容やその場の状況などに応じて表情豊かで多彩な口調を使い分けて話し掛けてきたら、ロボットに対する親しみが増し、対話がより楽しいものになると期待される。

また、日本語の話し言葉では、文末に現れる終助詞や助動詞などのいわゆる文末詞が話し手の態度や判断などのモダリティを表す役割を果たすが、文末詞が表す意味は文末詞のイントネーション（以下、「音調」と記す）とも密接に関係している。このため、言葉での表現は同じでも、文末音調の違いに応じて異なる意図を伝えることができる。文末詞の音調が発話全体の意味を支配することを示す実験結果も、報告されている。このことは、裏を返せば、合成音声にも適切な文末音調を付与しないと、誤ったメッセージが伝わってしまう危険性があることを意味している。実際、文末音調の違いで意味が変わる表現は、日本語を外国語として学ぶ学習者にとっては、コミュニケーション上の誤解を生む要因となっている。

音声対話を利用したサービスや製品が一般にも広く浸透し、合成音声を耳にする機会が増えてきている中で、音声合成技術に求められる性能や機能も、大きく変化してきている。対話音声らしい、生き生きとして表情豊かな合成音声を提供することが、急務となっている。こうした背景から、本研究では、音声対話システムの応答音声に豊かな表情を与えることができる音声合成の枠組みを構築することを目的として、合成音声における音声表現の多様化に向けた具体的な二つの目標を掲げた。これらを実現するための研究に取り組み、それぞれで成果を上げることができた。

【目標 1】対話状況に応じた音声表現の多様化（第 2 章～第 3 章）

対話を取り巻く状況や発話内容に応じて音声表現を使い分けることができる機能を提供する、音声合成の枠組みを構築した。本研究で選定した複数の音声表現を使い分けることで、音声対話システムの応答が表情豊かになるこ

とを示すことができた。

【目標 2】文末音調による意図表現の多様化（第 4 章～第 7 章）

文末詞とその音調の組合せと、それによって聞き手に伝わる話し手の意図との関係を明らかにし、文末音調によって意図を伝える表現手法を構築した。合成音声の聴取実験を通して、文末音調がもつ表現力の高さを裏付ける結果が得られたことは意義深い。

本論文では、これら二つの目標の実現に向けた議論を、以下のような構成により展開する。

第 1 章では、本研究が取り組む対話音声の合成を巡る近年の動向について概観するとともに、本研究の目的について詳述した。

第 2 章では、音声対話システムの応答を表情豊かにするためには、どのような音声表現を用意すべきかを論じた。人同士の対話では、そのときの話題や場の雰囲気などの要因に応じて様々な話し方や声質などの音声表現が用いられることから、対話を取り巻く状況の違いに応じた音声表現に着目した。感情の 2 次元平面モデルを手掛かりとして、四つの象限に対応するような状況 I～状況 IV と原点に対応する状況 0 の、5 種類の対話の状況を選定した。ロボットと人との模擬対話を用いた主観評価実験で、ロボットの全ての発話を状況 0 のモデルで合成した従来型の応答に比べ、発話文ごとに状況 I～状況 IV のモデルを適宜使い分けて合成した応答は、より自然で対話らしいとの評価を得た。選定した音声表現は様々な対話の場面で汎用的に利用可能であり、それらを対話の中で使い分けることで、音声対話システムの応答が表情豊かなものになることが確認できた。一方で、合成音声の品質に関する課題も見つかった。

第 3 章では、第 2 章の評価実験で浮かび上がった課題の改善に取り組んだ。特に、状況ごとの音声に、別人の声のようにも聞こえるほどの極端な声質の違いが生じてしまっていた。異なる音声表現を収集する際に、それぞれが適切な表現となるようにすることに重点を置き、異なる表現間の調和を保つことを考慮しなかったことに原因があると考えた。本論文では、複数の異なる音声表現を対話のコンテキストの中で使い分けたときの表現全体を通じた自然性を、個々の表現の自然性と区別するために、「調和性」と呼ぶことにする。表現全体の調和性を確保するために、対話の流れの中で異なる音声表現を順に表出させることを考え、話し手の心的状態が次々と変化するように進行するスキットを利用した音声収集手法を提案した。併せて明瞭性、自然性の改善に向け、音素連鎖などのバリエーションを確保することも考慮した。合成音声の評価の結果、新たに収集した音声データによる合成音声では、明瞭性、自然性が向上し、異なる音声表現を対話の中で使い分けたときの調和性も保たれていることが確認でき、提案した音声収集手法の有効性を示すことができた。

第 4 章では、対話音声の文末で、実際にどのような基本周波数（以下、「F0」と

記す)の形状が用いられているかを具体的に明らかにした。対話調の音声データから抽出した文末 F0 形状を、合成音声の文末音調の付与に用いるテンプレートとして利用できるようにするために、時間軸と周波数軸を正規化する方法を検討した。その上で、階層的クラスタリングによって形状を分類し、どのような文末音調が存在するかを見渡せる樹形図に示した。これらの中には、上昇調、上昇下降調などの、言語学の分野での分類に相当する特徴をもった形状も見られた。

第 5 章では、クラスタリングによって分類した多数の文末 F0 形状の中から、合成音声の文末音調の付与に利用するテンプレートを選び出すために、聴感上の判定を基準として導入した絞り込みの手法を考案した。一つのクラスタを分割した後の二つのクラスタのセントロイドが音調として差異があるかどうかを、相槌の音声「はあ」に付与した音調の対比較実験によって明らかにした。その結果、似通った形状でも聴感上ははっきり違うと感じられるものや、見掛け上は違いがあっても音調としての差異はほとんど感じられないものがあり、テンプレートの絞り込みに聴感上の判定を基準として導入することの必要性が示された。この結果を利用して、音調としての差異が顕著なものを選ぶような絞り込みを試みたところ、言語学の分野の先行研究における分類との対応が見られる形状が選ばれた。「上昇調」や「くだりのぼり音調」などのような名称で呼ばれている音調を合成音声に付与するための、具体的な F0 形状のテンプレートを獲得することができた。

第 6 章では、文末詞とその音調の組合せに着目し、それらと聞き手に伝わる話し手の意図との関係を明らかにした。聴感上も音調としての差異が顕著な 6 種類の F0 形状を文末音調として付与した合成音声の聴取実験により、文末音調には、文末詞と伝えたい意図の組合せに応じて適不適があることを示した。その結果を整理して、文末詞と伝えたい意図の組合せに適した文末 F0 形状をテンプレートの中から選択する、発話意図の表現手法を提案した。音調に応じて異なる意図が伝わる可能性がある文末詞を伴う発話文では、伝えたい意図に適した文末音調を付与しないとコミュニケーションに齟齬を来すことになる。本研究により、対話音声合成における文末音調の適切な制御の必要性を、改めて示すことができた。

第 7 章では、第 6 章で扱った主たる意図に付加される微妙なニュアンスに着目し、文末音調が違うだけでも聞き手には異なるニュアンスが伝わることを明らかにした。これまでほとんど議論されてこなかった言外の意図ともいべき付加的なニュアンスについても、文末音調による表現手法の実現可能性を示すことができた。従来の文末音調の分類では同じ音調として扱われる F0 形状の中には、F0 の動きが微妙に異なり、伝わるニュアンスに違いが見られる様々なバリエーションがあることも明らかになった。これらの結果は、文末音調を従来よりも細分化した上で、それぞれによって伝わる意図やニュアンスを具体的に明らかにしていくことで、合成音声による表現の幅が更に広がる可能性があることを示している。

最後に、第 8 章では、本研究の成果のまとめと今後の課題について述べた。

早稲田大学 博士（工学） 学位申請 研究業績書

氏名： 岩田 和彦

印

(2023年 1月 現在)

種類別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
論文	○岩田和彦, 小林哲則, “異なる音声表現の調和を保つことに力点を置いた対話音声合成のための音声収集手法の設計,” 電子情報通信学会論文誌D, Vol.J106-D, No.1, pp.57-65, Jan. 2023.
論文	○岩田和彦, 小林哲則, “対話音声合成の表現力向上に向けた文末音調の制御による付加的なニュアンスの表現に関する実験的検討,” 電子情報通信学会論文誌D, Vol.J102-D, No.6, pp.442-453, June 2019.
論文	○岩田和彦, 小林哲則, “対話音声合成の表現力向上に向けた文末詞と音調の組合せによる発話意図の表現に関する実験的検討,” 電子情報通信学会論文誌D, Vol.J100-D, No.11, pp.938-948, Nov. 2017.
論文	○白井克彦, 岩田和彦, “音声合成のための単語の強調表現の規則化,” 電子情報通信学会論文誌A, Vol.J70-A, No.5, pp.816-821, May 1987.
講演	小林哲則, 岩田和彦, “会話向け音声合成システム,” 信学技報, SP2014-93, Vol.114, No.303, pp.19-24, Nov. 2014.
講演	岩田和彦, 小林哲則, “発話意図に様々なニュアンスを付加して伝える文末音調の分析,” 日本音響学会2013年秋季研究発表会講演論文集, 3-P-24, pp.455-456, Sept. 2013.
講演	K. Iwata and T. Kobayashi, “Expression of speaker's intentions through sentence-final particle/intonation combinations in Japanese conversational speech synthesis,” Proc. 8th ISCA Workshop on Speech Synthesis (SSW), OS6-4, pp.235-240, Barcelona, Spain, Aug. 2013.
講演	K. Iwata and T. Kobayashi, “Speaker's intentions conveyed to listeners by sentence-final particles and their intonations in Japanese conversational speech,” Proc. IEEE ICASSP 2013, SP-L9.5, pp.6895-6899, Vancouver, Canada, May 2013.
講演	岩田和彦, 小林哲則, “終助詞とその音調とによって聞き手に伝わる発話意図の分析,” 信学技報, SP2012-77, Vol.112, No.281, pp.31-36, Nov. 2012.
講演	岩田和彦, 小林哲則, “対話音声における終助詞に応じた文末音調と聞き手に伝わる発話意図との関係,” 日本音響学会2012年秋季研究発表会講演論文集, 3-Q-15, pp.393-394, Sept. 2012.
講演	K. Iwata and T. Kobayashi, “Expressing speaker's intentions through sentence-final intonations for Japanese conversational speech synthesis,” Proc. INTERSPEECH 2012, Mon.P2b.03, pp.442-445, Portland, USA, Sept. 2012.
講演	岩田和彦, 小林哲則, “対話状況に応じたHMMを持つ音声合成システム,” 日本音響学会2011年秋季研究発表会講演論文集, 3-Q-3, pp.355-356, Sept. 2011.
講演	岩田和彦, 小林哲則, “対話音声における文末音調と聞き手に伝わる発話意図との関係の分析,” 日本音響学会2011年秋季研究発表会講演論文集, 2-8-5, pp.295-296, Sept. 2011.
講演	K. Iwata and T. Kobayashi, “Conversational speech synthesis system with communication situation dependent HMMs,” Proc. Workshop on Paralinguistic Information and its Integration in Spoken Dialogue Systems (IWSDS 2011), pp.109-119, Granada, Spain, Sept. 2011.
その他 (論文)	高津弘明, 福岡維新, 藤江真也, 岩田和彦, 小林哲則, “会話によるニュース記事伝達のための音声合成,” 人工知能学会論文誌, Vol.34, No.2, pp.B-I65_1-15, March 2019.
その他 (論文)	白井克彦, 小林哲則, 岩田和彦, 深沢克夫, “ロボットとの柔軟な対話を目的とした音声入出力システム - WABOT-2 における会話系,” 日本ロボット学会誌, Vol.3, No.4, pp.104-114, Aug. 1985.

早稲田大学 博士（工学） 学位申請 研究業績書

氏名： 岩田 和彦

印

(2023年 1月 現在)

種類別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
その他 (講演)	I. Fukuoka, K. Iwata, and T. Kobayashi, "Prosody control of utterance sequence for information delivering," Proc. INTERSPEECH 2017, pp.774-778, Stockholm, Sweden, Aug. 2017.
その他 (講演)	福岡維新, 岩田和彦, 小林哲則, "発話系列を扱う会話音声合成," 信学技報, SP2016-74, Vol.116, No.414, pp.59-64, Jan. 2017.
その他 (講演)	大町基, 岩田和彦, 小林哲則, "距離感を与える音声の特徴分析と合成," 信学技報, SP2009-89, Vol.109, No.356, pp.159-163, Dec. 2009.
その他 (講演)	岩田和彦, 三留幸夫, "パソコン向け音声合成ソフトウェアを利用したコミュニケーション支援装置の試作," 情報処理学会第51回全国大会講演論文集(分冊1), 4T-8, pp.373-374, Sept. 1995.
その他 (講演)	岩田和彦, 高橋一裕, 三留幸夫, 西浦充, "マルチメディア・パソコンにおけるテキスト音声合成の利用," 1995年電子情報通信学会総合大会講演論文集(情報・システム1), SD-9-8, pp.381-382, March 1995.
その他 (講演)	岩田和彦, 三留幸夫, "発話テンポに依存しない韻律構造のモデル化," 1994年電子情報通信学会春季大会講演論文集(第1分冊), SA-5-2, pp.486-487, March 1994.
その他 (講演)	岩田和彦, "文章朗読音声における韻律的特徴の分析," 日本音響学会平成6年度春季研究発表会講演論文集, 2-Q-11, pp.285-286, March 1994.
その他 (講演)	岩田和彦, 高橋一裕, 三留幸夫, 永野敬子, "パソコン向けソフトウェア日本語テキスト音声合成," 日本音響学会平成5年度秋季研究発表会講演論文集, 2-8-13, pp.245-246, Oct. 1993.
その他 (講演)	岩田和彦, 三留幸夫, "発話テンポの影響を考慮した継続時間長制御モデル," 日本音響学会平成5年度春季研究発表会講演論文集, 3-8-3, pp.241-242, March 1993.
その他 (講演)	K. Iwata and Y. Mitome, "Prosody generation models constructed by considering speech tempo influence on prosody," Proc. ICSLP 92, pp.1155-1158, Banff, Canada, Oct. 1992.
その他 (講演)	岩田和彦, 三留幸夫, "発話テンポと種々の音声区間長との関係," 日本音響学会平成4年度秋季研究発表会講演論文集, 1-5-12, pp.239-240, Oct. 1992.
その他 (講演)	岩田和彦, 三留幸夫, 宮本牧, 渡辺隆夫, "自動通訳システムINTERTALKERにおける日本語音声合成," 情報処理学会第44回全国大会講演論文集(分冊3), 6P-8, pp.225-226, March 1992.
その他 (講演)	岩田和彦, 三留幸夫, "発声速度が韻律に及ぼす影響," 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-21, pp.247-248, March 1992.
その他 (講演)	岩田和彦, 三留幸夫, 友竹世光, 赤松実, 亀谷潤, 小澤一範, 渡辺隆夫, "残差制御型音声合成方式を用いた日本語テキスト音声変換システム," 信学技報, SP90-56, Vol.90, No.335, pp.15-22, Nov. 1990.
その他 (講演)	K. Iwata, Y. Mitome, and T. Watanabe, "Pause rule for Japanese text-to-speech conversion using pause insertion probability," Proc. ICSLP 90, pp.837-840, Kobe, Japan, Nov. 1990.
その他 (講演)	K. Iwata, Y. Mitome, J. Kametani, M. Akamatsu, S. Tomotake, K. Ozawa, and T. Watanabe, "A rule-based speech synthesizer using pitch controlled residual wave excitation method," Proc. ICSLP 90, pp.185-188, Kobe, Japan, Nov. 1990.
その他 (講演)	K. Iwata, K. Ozawa, Y. Mitome, and T. Watanabe, "A Japanese text-to-speech conversion system using pitch-controlled residual wave excitation," 119th Meeting of the Acoustical Society of America, RR4, Pennsylvania, USA, May 1990.

早稲田大学 博士（工学） 学位申請 研究業績書

氏名： 岩田 和彦

印

(2023年 1月 現在)

種類別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
その他 (講演)	岩田和彦, 小澤一範, 三留幸夫, 渡辺隆夫, “残差制御型合成方式を用いた日本語テキスト音声合成システム,” 日本音響学会平成元年度秋季研究発表会講演論文集, 3-P-18, pp.311-312, Oct. 1989.
その他 (講演)	岩田和彦, 市川昌子, 小澤一範, 三留幸夫, 渡辺隆夫, “残差制御型音声合成方式を用いた規則音声合成システム,” 日本音響学会平成元年度春季研究発表会講演論文集, 2-7-19, pp.207-208, March 1989.
その他 (講演)	岩田和彦, 三留幸夫, 小澤一範, “残差制御型規則音声合成システムにおける韻律制御規則の検討,” 日本音響学会平成元年度春季研究発表会講演論文集, 2-7-17, pp.203-204, March 1989.
その他 (講演)	岩田和彦, 市川昌子, 小澤一範, 渡辺隆夫, “残差制御による音声合成システムの検討,” 日本音響学会昭和63年度秋季研究発表会講演論文集, 3-2-7, pp.183-184, Oct. 1988.
その他 (講演)	岩田和彦, 三留幸夫, 伏木田勝信, スティーブクレイトン, “英語テキスト音声変換システムの試作,” 昭和62年電子情報通信学会情報・システム部門全国大会講演論文集(分冊1), 169, p.170, Nov. 1987.
その他 (講演)	K. Shirai, K. Iwata, and T. Ohno, “Pitch contour control in Japanese conversational speech,” Proc. IEEE ICASSP 1986, pp.2043-2046, Tokyo, Japan, April 1986.
その他 (講演)	岩田和彦, 大野毅, 白井克彦, “対話文章合成のための韻律制御規則の検討,” 昭和61年度電子通信学会総合全国大会講演論文集(分冊6), S26-2, pp.367-368, March 1986.
その他 (講演)	岩田和彦, 大野毅, 白井克彦, “会話文章における基本周波数パタンの制御規則について,” 日本音響学会音声研究会資料, S85-42, pp.317-324, Oct. 1985.
その他 (講演)	岩田和彦, 大野毅, 白井克彦, “会話文章におけるピッチパターンについて,” 日本音響学会昭和60年度秋季研究発表会講演論文集, 2-3-21, pp.195-196, Sept. 1985.
その他 (講演)	岩田和彦, 大野毅, 白井克彦, “単語合成における音韻性を考慮したピッチパターン制御,” 日本音響学会昭和60年度春季研究発表会講演論文集, 1-5-18, pp.137-138, March 1985.
その他 (講演)	岩田和彦, 白井克彦, “規則合成システムにおける韻律パタンの生成,” 日本音響学会音声研究会資料, S84-45, pp.349-356, Oct. 1984.
その他 (講演)	岩田和彦, 阿部匡伸, 小熊幸雄, 白井克彦, “素片編集規則合成の実験用システム,” 昭和59年度電子通信学会総合全国大会講演論文集(分冊6), 1643, p.196, March 1984.
その他(講演)	その他25件
その他(著書)	白井克彦(編著), 音声言語処理の潮流, コロナ社, 東京, 2010. (3章分担執筆)
その他(著書)	岩田和彦, “音声認識技術とその応用,” NEC技報, Vol.56, No.9, pp.19-22, Nov. 2003.
その他(著書)	岩田和彦, 高橋一裕, 三留幸夫, “音声合成,” NEC技報, Vol.47, No.8, pp.65-71, Sept. 1994.
その他 (著書)	K. Shirai, T. Kobayashi, Y. Komori, N. Hashimoto, K. Iwata, Y. Fukazawa, and J. Yazawa, “Speech I/O system realizing flexible conversation for robot – The conversational system of WABOT-2,” Bulletin of Science and Engineering Research Laboratory, Waseda University, No.112, pp.53-79, Sept. 1985.
その他(著書)	その他4件
その他(特許)	岩田和彦, 音声応答装置, 日本電気株式会社, 特許第3601411号, 2004.10.01.
その他(特許)	岩田和彦, 音声認識型通信制御装置, 日本電気株式会社, 特許第3327213号, 2002.07.12.
その他(特許)	その他17件