早稲田大学大学院　基幹理工学研究科

博 士 論 文 審 査 報 告 書

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

2023年2月

文字では原則言語情報に限定した情報が伝わるのに対し，音声では話者の心的状態や話者の意図に係る微妙なニュアンス等多様な情報が言語情報と同時に伝わることとなる。音声対話が負担の少ないコミュニケーション手段として好まれるとされるのは，こうした多様な情報の同時伝達が，効率的あるいは効果的なコミュニケーションの基礎となるからと考えられる。

本研究は，これら多様な情報を運ぶ音声の表現（以下，音声表現）を明示的に制御しうる音声合成方式について，音声対話システムへの応用を指向しながら検討したものである。

音声表現の多様性を扱う研究は従来から数多く行われている。しかし，それらの成果は，対話の文脈で利用するには十分とはいえない。例えば，音声表現の多様性を扱う音声学的研究は，音調の型とそれが持つ機能とを結びつける有用な知見を与えるが，音調の型が具体的にどのような F0 形状によって表現されるべきかまでの情報は与えず，音声合成用途には利用できない。また，昨今進展するニューラルネットにおける End-to-End のアプローチは，多様な表現を極めて良好な音質をもって実現するものの，その多様化に係る変動要因は明示的には扱われず，音声表現の制御は困難である。

本研究では，対話状況に応じた話者の心的状態表現と，文末音調による意図表現に焦点をあて，その多様化に必要となる知見の整理を，音声対話システム用音声合成システムに直接利用しうる形で検討している。

本論文では，これらの議論を，全 8 章構成によって進めている。以下，各章の内容とその評価について述べる。

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ている。

第 2 章と 3 章は，話者の心的状態変化に応じた音声表現の制御について検討している。

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ている。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，それぞれの状況における発話を収集して複数の音声合成器を作成し，これを場面に応じて切替えて利用することを提案している。ロボットと人との模擬対話を用いた主観評価実験を通じて，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，対話の自然性を格段に改善することを確認している。

本章の成果は，対話で生じる話者の心的状態変化に追従した音声表現の生成に成功したものであるが，同時にそこで必要となる状態表現のパラメタは感情モデルにおける極少数のパラメタで良いこと示したものでもある。このことは，例えば近年主流をなす End-to-End 型のニューラルネットによる音声合成方式の学習データを補助情報つきで収集する際にも，その設計指針と

して利用されうるものであり，興味深い。

ついで第3章では,第2章で提案した方式における一連の発話の調和性(発話全体を通しで聞いたときの自然性）を改善することを試みている。第2章の方法は，状況毎に音声表現を使い分けることを可能にしたものの，各状況を独立に扱ったため，状況ごとの音声に極端な声質の違いが生じるなど，通しで聞いたときに違和感を生じることがあった。ここでは，対話の流れの中で話し手の心的状態が自然な形で次々と変化するよう設計したスキットを用意し，これに従って発話した音声を収集し，これを合成器の学習に用いることを提案している。学習データベース中の各音声表現が，周囲の音声表現との依存関係を保って発話されることで，自然な形で調和性が担保されるとしている。対話実験により，提案手法による合成音声では，異なる音声表現を対話の中で使い分けたときも調和性が保たれていることを確認している。

本章の内容は，対話における一連の発話の調和性という概念にはじめて注目したものであり，先見性の高い研究として評価できる。

第4章から第7章にかけては，文末音調の制御に基づく多様な発話意図の表出について検討している。

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について調査している。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した上で，階層的クラスタリングを適用し，どのような文末音調が存在するかを見渡せる樹形図として示している。

前章の結果得た F0 形状の樹形図は物理尺度に基づいて作成されたものであり，聴感上の特性を反映しない。第5章では，この樹形図に沿って順に一対比較による聴覚心理評価を行って，どの F0 形状が聴感上の差異を与えるミニマルなセットかを選ぶという方法を提案している。その結果，数ある F0 形状の中から，聴感上重要なパターンを，実施可能なコストで選ぶことに成功している。また，結果として得られた F0 形状の代表パターンは，音声学が与える音調分類と,よく整合する対応関係が見られることを報告している。

聴感上の差異を与えるミニマルな F0 形状の代表セットを選ぶという発想はユニークであり，また，それを物理尺度のクラスタリングと，聴覚心理実験の組合せによって選ぶという手法も興味深く，評価できる。

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ている。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ている。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ている。

扱った文末詞の種類が限定的という問題はあるものの，頻度の高い重要なものについて，F0 形状とそれが伝える意図との関係が整理されている。また，本章の成果は，従来音声学的に分類されていた音調の型に対し，具体的な F0 形状を結びつけたものとしても解釈でき，価値が高いものと評価できる。

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ている。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う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している。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している。

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の効果・可能性を示したものであり興味深い。

最後に，第 8 章では，本研究の成果のまとめと残された課題について述べている。

以上，これを要するに，本論文は，音声対話システムへの応用を指向して，対話場面に応じた多様な音声表現を出力可能とする音声合成方式を実現したものである。一部対象とする表現が限定的であるという問題はあるものの，重要な表現について，心的状態および意図／ニュアンスの明示的な指定によって，それに相応しい音声が出力できるしくみが実現されている。これらの技術は，対話システムに直接応用してその自然性を高めるばかりでなく，近年主流をなす End-to-End 型のニューラルネットに与える学習データの設計指針を与えるものとしても利用価値を持つ。よって，その工学的意義は高く，本論文は博士(工学)(早稲田大学)の学位論文として相応しいものと認める。

2023 年 2 月

審査員

主査	早稲田大学 教授	工学博士（早稲田大学）	小 林 哲 則
副査	早稲田大学 教授	博士（工学）（東京大学）	甲 藤 二 郎
	早稲田大学 教授	博士（工学）（早稲田大学）	小 川 哲 司
	名古屋工業大学教授	工学博士（東京工業大学）	徳 田 恵 一