韻律制御に主体をおいた対話システム

Spoken Dialogue System utilizing Prosody Control

早稲田大学理工学部
School of Science and Engineering, Waseda University

小 林 哲 則
Tetsunori Kobayashi

.

Several issues concerning the use of prosodic information in conversation are investigated. The themes treated here are as follows: a) utterance timing control using prosodic information, b) mental state estimation using prosodic information, c) use of prosodic information for conversational speech recognition, d) use of prosodic information for conversational speech synthesis, and e) conversational system utilizing prosodic information. With the results of theses studies, a natural conversation system performing rhythmical conversation is realized.

Key Words: Pitch control, Dialogue system, Rhythm of conversation, Para-linguistic information

1. 研究の目的

本研究では，対話システムにおける韻律の利用の問題に焦点を当て検討を行った。

高度情報化社会を迎えようすとする今日，誰でも抵抗なく情報機器を利用できる環境を整えることは非常に重要である。この観点から音声対話インタフェースに寄せられる期待は大きく，これまでに多くの対話システムが開発されてきた。しかし、それらは読み上げに近い発話を認識し、単純な抑揚・タイミングで応答を返すだけのものであった。対話特有の言いよどみや言い直しを扱うものは少なく、また自然なリズム感を与えるものは皆無であった。結果として、音声対話としての魅力を感じさせるものは少なく，望まれる自然なインタフェースとしての役割を果たせなかった。この原因は、従来、発話個々の意味把握に集中するあまり、対話のリズムに重要な影響を与えるとともに、対話構造の把握に重要な役割を持つ韻律情報を無視してきたことにある。

そこで，本研究においては，韻律情報の利用によって対話調音声の理解を可能にした，高度な音声対話システムを実現することを試み，これによって次世代のヒューマンインタフェースに望まれる、気の利いた、より豊かなコミュニケーション機能を実現することを試みた。

具体的なテーマとその担当者は以下の通りである。

a) 韻律・表層的言語情報による発話タイミング制御(中川聖一：豊橋技術科学大学)：対話のリズムに大きく影響を与える相槌と割り込みについて検討を行い、特に雑談のような対話について，その最適な挿入タイミングをリアルタイムに検出する方式を明らかにし，実装する。

b) 心的状態推定における韻律の利用(白井克彦・菊池英明:早稲田大学)：対話者の心的状態を韻律情報から推定する枠組みについて検討する。

c) 対話理解における韻律利用（甲斐充彦：静岡大学）：言いよどみ・言い直しなどの対話特有な現象と韻律の関係や、発話意図と韻律の関係を実データの分析に基づいて整理し、それを用いて対話音声の高度な認識手法を確立する。

d) 対話音声合成における韻律制御（匂坂一典：早稲田大学）:入力言語情報による韻律制御の高度化に関する検討を行う。特に、豊富な韻律変化を実現するため、語彙の韻律的有標性の定量化、韻律的強勢の実現、話速制御モデル作成を試みる。また、韻律制御パラメータが合成音声の自然性品質に及ぼす影響を調べるため心理評価試験を行う。

e) 対話システムの構築（小林哲則：早稲田大学：a)－d)の成果を導入するためには，どのようなシステムアーキテクチャが望まれるのかについて検討し、これに基づいて韻律の利用に基づく高度な音声対話システムを実現する。また，韻律によって伝わるパラ言語の情報を認識するシステムを実現し，対話システムに組み込む。

以下，これらの成果の概要について述べる。

2.韻律・表層的言語情報による発話タイミング制御

本テーマにおいては，リズムある円滑な対話の実現に必要となる発話タイミング制御モデルを構築する観点から，韻律の利用法について検討した。

人間と機械が対話を行うことを考えるとき，機械が人間同士の会話と同様にあいづちや割り込みなどの応答を返すことができれば，より円滑な対話を行うことが期待できる．特に雑談のような対話に着目すると，その中ではたわいもない内容でありながら，あいづちや話者交替をタイミングよく繰り返すことによって継続されていく．すなわち，親しみやすい対話インタフェースを構築するには，このようなタイミングの考慮が不可欠である．本研究では，特に雑談のような対話に着目し，自然な雑談対話をする上で最も重要であるタイミング生成，すなわち，あいづちと割り込みのタイミングの判定を，韻律情報および表層的言語情報からリアルタイムに行うシステムの構築を行った．

まず，実際の人間同士の対話を分析した．小磯らの分析[1]によると，発話句音声末1 モーラ分のピッチ，パワーが特定の変動パターンに従った場合，対話相手に発話継続，終了，あいづちが出現する傾向が異なるとしている．また，言語情報としては，句末の助詞「ね」や「か」，もしくはタスクのキーワードの出現などがあいづちを誘発する．さらに，発話長が長い場合にもあいづちが打たれる場合が多い．

これらの分析に基づいて，この種々の要因を表現する特徴を素性とした決定木を用いてあいづちや話者交替のタイミング生成を行う方法を提案した．システムはまずユーザ発話のポーズを検出すると，100ms ごとに決定木によってあいづち/話者交替/何もしないというアクションを決定する．そのための素性として句音声末の100ms（およそ1モーラ分）のピッチおよびパワーの時間的変化の回帰係数，終端単語の品詞や発話の最後に現れた自立語の品詞情報，ユーザの発話長，および発話終端からの時間(ポーズ長)などを用いた．

人間同士の対話の片方の話者を，この決定木に置き換えることにより生成したあいづち・話者交替タイミングを，元の人間のものと比較した結果，オープンテストで，あいづちの分類において再現率57.4%，適合率24.8%，話者交替の分類において再現率44.8%，適合率71.6%の結果を得た．また，この決定木を天気予報を話題としたELIZA型応答生成と組み合わせた対話システムを構築し，被験者が実際に対話を行うことにより主観的に評価した結果，システムの返答内容に関しては改良が必要であるものの，あいづちタイミング自体はよいという評価が得られた．（中川）

3.対話者の心的状態推定における韻律の利用

本テーマにおいては，対話時に必要となる対話者の心的状態推定を行う観点から韻律の利用について検討した。

まず初年度において、システムと人間のタスク指向対話を収録し、特に心理的変化が顕著に現れると考えられるシステムの誤解発話に対するユーザ応答の分析を行ったところ、ユーザの戸惑いがポーズ長、パワー、基本周波数増加に影響を与えていることが分かった。こうした戸惑いや慣れなどのユーザの心的な状態をユーザ発話の韻律情報から推測するモデルを構築することの見通しを立てた。

次年度においては，韻律を利用した心的状態推定モデルをより精密に構築するために、韻律特有の単位(アクセント句）を人手で解析してラベル情報として利用することを試みた。これにより、喜びや怒りなどの基本感情の識別、発話における丁寧さの判別において精度が向上することを確認した。

最終年度においては，これらの成果に基づいて、話題境界の判別を題材に、韻律情報におけるアクセント句単位でのパラメータを用いて統計的なモデルを学習し、オープンデータに対しても人間と同程度の判別精度が得られることを確認した。

また、こうして得た様々なモデルを実装するための音声対話システムの汎用的なプラットフォームを開発した。（白井・菊池）

4．対話理解における韻律利用

本テーマにおいては，対話音声理解の観点から，韻律の利用について検討した。

自然な発話や対話音声の音声認識の研究において、特に十分な認識精度が得られない状況下では、人間のように聞き誤りの認識や聞き直しのような柔軟な振舞いを可能にすることが一層重要となる。そこで第一に、対話音声における重要語や、音声対話システムとの対話によくみられる繰り返しの訂正発話の韻律的特徴に焦点を当て、その分析や訂正発話の検出を検討した。また、これらと併せた頑健な対話音声理解のため、対話音声において頻出するフィラーの分析とモデル化の検討を行った。

対話音声における重要語の韻律特徴分析においては、ATR 対話音声データベースから55名分の110発話を無作為に選び、重要と思われるフレーズに人手でタグをつけ、当該フレーズ区間と発話全体、中央や述部の音声区間との韻律特徴の統計的な違いを分析した。結果

として、重要語部分の正規化されたピッチ、音素の持続時間およびパワーの統計量に関して、フレーズ全体の一般的傾向との違いを特徴付けることが示された。また、訂正発話の特徴分析および検出法の検討および評価のため、人間対機械および人間同士の２つの状況下での音声対話システムとの対話音声をそれぞれ収録した。この対話音声資料に関して、繰り返しの訂正発話においてその連続する２発話間の韻律特徴の変化に注目した分析の結果、１）対機械と対人間の両者で一貫した特徴は少ない、２）対機械の発話においては基本周波数の最大値や標準偏差の変化に有意な変化がみられる、３）車の運転（ゲーム）による認知的負荷を持った並行タスクを与えたときには有意な特徴が一部みられなくなる、などの知見が得られた。また、決定木 C4.5による機械学習のアプローチでこれらの特徴及び音響的特徴の DP マッチングによる検出法との併用を検討した。韻律特徴単独での検出率は低く、前述の統計的分析結果を裏付ける結果となった。そこで、音響的特徴のマッチングによる訂正発話検出法において、フレーズ単位の韻律的特徴のモデル化によって、検出候補の絞り込み・検証を試みた。結果は、一部の典型的な訂正発話における有効性を示唆する結果が得られている。また、フィラーに関する統計的分析の結果、同様に発話内の韻律的特徴変化を考慮することによって、フィラーの同定に寄与することが示唆された。（甲斐）

5.対話音声合成における韻律制御

本テーマにおいては，対話における合成の側面からから韻律の問題を考えた。

自然なマンマシンインタフェースを実現する上で、会話音声の韻律の制御は不可欠であるが，その制御特性についてはこれまで十分に調べられてこなかった．制御特性の定量的な分析を図るためには、制御モデル化と共に生成面、知覚面からの知識の拡充が急務である．我々は、これら多面的な理解を進めるため、①制御モデルによる自動分析を進めるための F0制御パラメータの自動抽出、②会話音声の発話語彙情報を用いた F_0 制御可能性に関する生成・知覚両面からの分析、③時間制御の自然性に関する聴知覚特性の測定を進め、会話音声制御機構の解明をねらった．

F0制御パラメータの自動抽出については，基本周波数制御の本質を理解し、音声合成に用いる高性能な F_0 制御規則を作成する観点から、生成モデルに基づく F_0 制御パラメータの分析が有効と考えられる．本研究では、入力された発話の情報を可能な限り利用し、F_0制御パラメータの抽出を行うことにした．抽出実験結果から、発話情報を最適パラメータ探索の初期値を予測する過程で用いた場合、抽出精度が向上することが確認できた．

発話語彙情報に基づく対話音声韻律制御については，語彙情報に基づいた F_0制御を目的とし、対話音声における程度副詞にみられる F_0変化について、生成、知覚、制御の観点から分析を行った．生成、知覚の観点から、対話音声において語彙情報が F_0自然性に与える影響を確認した．これらの知見に基づき、程度副詞が持つ語彙としての主観的強さによる F_0制御を提案した．語彙および被験者について、オープンなデータに対してアクセント句成分の大きさの予測実験を行い、この予測が効果的に行えることが判明した．これらの結果より、程度副詞の強勢の強さに基づいて、対話音声の為の F_0 制御パラメータの制御が可能である事を確認し、語彙情報に基づく F_0制御の可能性を確かめた．

文音声における音韻時間長伸縮に関する検討については，その知覚上の許容度を文音声を用いて調査した。従来，合成音声の音韻長制御の自然性向上を目的として、音韻長知覚特性が研究されてきたが、単語音声のみが対象とされてきた．音韻長知覚特性の音韻長制御への応用を考えると、文音声での調査が不可欠である．文音声を対象とした調査の結果、文節頭、文節中、文節末の順に許容度低下が大きいことがわかった．また、発話速度が速くなるにつれて、音韻長伸縮に対する許容度低下が大きくなることがわかった．さらに、音韻長伸縮に対する位置の効果が、発話時の音韻長制御特性と関連があることが示唆された．これらの知見は例えば、音韻長制御の際に重み付けをすることで合成音声の品質を向上させるなどの応用が期待できる。（匂坂）

6.対話システムの構築

本テーマでは，本班あるいは他班が与える知見を利用しながら，韻律を様々な形で利用することでリズムある自然な対話が可能なシステムを構築することを試みた。

まず，各グループで開発される様々な要素技術を組み込みやすいことを目的とした対話システムの枠組みについて検討した。提案するフレームワークはパブリッシュ－サブスクライブ機能のついた黒板型の情報共有方式を基本とする[3]。黒板型の情報共有によって，多数のモジュールが容易に情報を共有しあうことがで

き，モジュール追加等の拡張が容易になる。一方，パブリッシュ－サブスクライブ機能によって，必要な情報を予め登録（サブスクライブ）しておくと，情報が変わった瞬間にその情報が通知（パブリッシュ）される。このことによって単純な黒板では扱えない迅速な情報共有が実現できた。

次に，このフレームワーク上に，白井・菊池らの心的状態推定の手法を組み込み，パラ言語の理解能力を有する対話システムを構成した[4]。ここでは，提案に対する復唱発話の韻律から，提案を受けた人が提案を肯定的に受け取ったか否かを推定する機構を実装した。これを対話システムに組み込むことによって，言語情報でYES/NOを明示しない場合においても，発話者の意向を汲み取って，対話を円滑に進めることができるシステムが実現できた。

次に，中川の成果である発話タイミングの決定器を若干拡張する形で対話システムに組み込み，相槌と復唱が可能なシステムを構築した[5]。従来の相槌システムとは，発話内容と韻律の双方をタイミングの決定に用いていることが大きく異なる。発話内容の認識にはFSTを用いることで早期に内容を推定することが可能になっている。これらのことによって，相手の発話内容に応じた応答を，タイミングよく返すことが可能になった。

尚，これらの対話システムにおいては，聞きなおし機能の実装において甲斐の研究成果が，また合成において匂坂の研究成果が一部組み込まれている。

7.むすび

以上の研究の成果として，韻律の適切な利用に基づいて、リズム感のある対話が実現されるとともに、言いよどみなどの対話特有の現象が扱われることになり、音声対話の本質を捉えたインタフェースが実現された。このことは、より豊かなコミュニケーションという新たなマン･マシンインターフェイスな形成に道を拓く。また，今日，情報機器の利用に精通するものとしない者との間に深刻な情報収集機会の格差が生じつつあり，これが重大な社会問題になる可能性を持つが，自然な音声対話インタフェースの実現は，この問題の解決につながるであろう。

本研究では，対話における韻律の役割を，自然な対話のリズムの実現と，高度な対話理解・生成の両面に求め，実データの分析に基づいて，これらを実現する有用な対話制御モデルと対話理解モデルとを構築し，それらを基礎として自然な対話システムを構築したものであるが，韻律に着目して，自然性の高い対話を実現しようとする試みは国内外に例が無く，極めて独創的なものとして評価されよう。

参考文献

[1] 小磯 花絵，堀内 靖雄，土屋 俊，市川 熹，“先行発話断片の終端部分に存在する次発話者に関する言語的･韻律的要素について，”電子情報通信学会技術報告，NLC95-72，pp.25-30（1996）.

[2] J. Weizenbaum，“ELIZA -A computer program for the study of natural language communication between man and machine，”communications of the ACM，vol.9，no.1，pp.36-45（1965）.

[3] 松坂要佐，於久健太郎，小林哲則，“多機能ロボット開発のための情報共有アーキテクチャの設計と実装，電子情報通信学会論文誌，”D-I，Vol. J86-D-I，No.5，pp.318－329，May 2003.

[4] Shinya FUJIE，Yasushi EJIRI，Yosuke MATSUSAKA，Hideaki KIKUCHI and Tetsunori KOBAYASHI，“Recognition of Para-Linguistic Information and Its Application to Spoken Dialogue System，”IEEE ASRU2003，pp.231-236，Dec.2003.

[5] 藤江真也，福島健太，柴田大輔，小林哲則，“FSTと韻律情報を用いた相槌・復唱機能を有する対話ロボット，” 人工知能学会研究会資料，SIG-SLUD-A401-03，pp.15-20，Jun.2004.