様式 C-19

科学研究費補助金研究成果報告書

平成２２年　６月２３日現在

研究種目：基盤研究（Ｂ）
研究期間：2006～2009
課題番号：18300063
研究課題名（和文）　出力単語の語彙属性を用いた対話韻律制御に基づく音声合成

研究課題名（英文）　Speech synthesis with communicative prosody driven by the impressions of output lexicons

研究代表者
匂坂　芳典（SAGISAKA YOSHINORI）
早稲田大学・理工学術院・教授
研究者番号：70339737

研究成果の概要(和文)：　対話場面での音声を合成するために､対話韻律生成方法を提案した。提案法では、対話における発話内容自体がその韻律を限定することに着目し、出力語彙が与える印象が規定する特徴量を用いて、対話韻律を生成する。これにより、従来の読み上げ調日本語合成音声の大幅な自然性向上を達成した。さらに、中国語と英語の対話音声合成への適用可能性、逆の機能となる対話音声からの印象推定可能性を実証し、提案法の汎用性を確認した。

研究成果の概要（英文）：A scheme for communicative prosody generation was proposed to synthesize speech needed for conversational purposes. Using the correlation between communicative prosody and impression attributes of lexicons constituting output, the proposed scheme enables prosody control for conversational speech output. Perceptual experiments showed the superiority of the speech synthesized with the proposed communicative prosody to the conventional one with reading style prosody. Further application to Chinese and English speech synthesis and the reverse technology of impression extraction from speech clarified the usefulness of the proposed approach.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００６年度	3,200,000	0	3,200,000
２００７年度	4,100,000	1,230,000	5,330,000
２００８年度	4,100,000	1,230,000	5,330,000
２００９年度	3,400,000	1,020,000	4,420,000
総　計	14,800,000	3,480,000	18,280,000

研究分野：総合領域
科研費の分科・細目：　情報学　知覚情報処理・知能ロボティクス
キーワード：　音声情報処理

１．研究開始当初の背景

言語情報を入力とした音声合成技術においては、コーパスベース音声合成の導入により、出力される音声の品質は大幅に向上し、その適応領域は広がった。従来、音声合成ではテキスト入力読み上げ等の音声出力を対象としており、得られる合成音声はテキストの読み上げに適したものとなっている。このため、カーナビゲーション、ロボットをはじめ、ゲームやコールセンターなど、利用者に対するコミュニケーションの音声出力として使用する上で、対話音声としての自然性の

不備が指摘されていた。開発時点では対話音声を生成するための方法論すらなく、自然性を大きく左右する対話韻律の制御特性の把握、制御要因の特定、韻律・声質生成のモデル化が求められていた。

２．研究の目的

本研究では、発話毎の多様性を実現した、自然性の高い対話音声の合成出力を目的とする。とりわけ自然性に大きな影響を与える対話韻律の制御方法について検討する。このため、以下の課題の検討を行う。

（１）対話音声における韻律制御特性の把握と出力単語が持つ情報との相関分析

（２）対話韻律制御の数理モデル、合成法

（３）提案対話韻律制御方法の有効性検証

（４）他言語への展開可能性等の汎用性追及

３．研究の方法

対話音声合成の必要性は早くから認識されたが、本格的な検討が進められて来なかった。この理由は、対話者が表出する意図や発話状況の規定や表現、定量的な取り扱いの難しさにある。これらの情報は，いわゆる「パラ言語(＝言語外)情報」として、伝統的な言語学でも直接扱ってこなかった情報である。パラ言語に関しては、感情音声合成研究に見られるように、あらかじめ規定された感情や、発話場面などで指定される典型的な韻律に限られた分析・合成の検討が進められてきた。これらの従来研究にみられる、先験的なカテゴリ分けに基づく一様な全体制御は，対話場面で用いられる多様な韻律を扱うには不十分である。対話音声の韻律生成には，発話内容に即して語句のレベルで動的に表出される韻律の適切な制御が不可欠である。これまでに、対話音声を特徴付け、出力を可能とする情報の規定とそれに基づいた韻律生成・音声合成の方法論は提案されていない。

単なる読み上げでない対話音声は、言語形式だけでなく、その対話韻律を規定するための入力情報を必要とする。ヒトの発話を考えた場合、図１に示すように、対話音声生成のための入力情報として発話意図等が考えられるが、具体的に情報内容を規定することは難しい。このような対話発話に関する情報の直接的な規定は難しいが、その対話の言語内容は，対話が伝える発話行為(speech act)と関連してその対話が取りうる韻律を制約する。例えば、「きれいだ」という対話発話は、多くの場合、好印象を与え、語彙によって取りうる対話韻律は限定される。すなわち、この対話発話は，悪印象を与える「汚い」，疑念を与える「奇妙だ」といった語句が想起させる対話韻律とは違った、発話語彙が与える印象、醸し出す雰囲気に即した韻律が用いられることが多いと推察される。

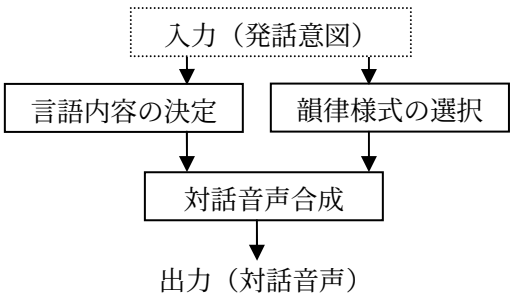

図１　対話音声の生成

本研究では、このような、対話韻律が合成すべき言語内容と相関を持つことに着目し、合成対象として与えられる出力言語自体から対話韻律を推定する問題として解くことを考えた。この考え方による対話韻律制御の可能性を調べた。出力単語が有する印象情報を用い、対話音声の韻律制御特性との相関関係を調べ、対話韻律の制御方法を考える方法を用いた。

４．研究成果

（１）出力単語の印象属性と対話韻律の分析

まず、実際の対話音声にみられる韻律制御要因の特定と、合成用入力情報の抽出を考えた。このため、一語発話「ん」の対話韻律に対して進めた解析方法と得られた結果[1]の一般化を考えた。「ん」は、最も短い言語形式からなる発話であるが、その多様な韻律によって種々のコミュニケーション情報を伝達する。「ん」の対話韻律の分析では、実験で得られた２６の印象表現（納得、了承、疑い、迷い、疑問、同意、否定、反論元気な、楽しい、優しそう、機嫌が良い、わくわく、嬉しい、軽い、興味がある、明るい、暗い、弱々しい、興味がない、機嫌が悪い、重い、面倒くさい、ふてぶてしい、怒っている、うざい）に対して多次元尺度構成法（MDS）により得られた３次元（「好印象‐悪印象」、「確信‐疑念」、「肯定‐否定」）の心理表現空間内の位置と、各発話の韻律形状の対応が付くことが分った。

声帯の基本周波数(F0)平均値の高低は「好印象」「悪印象」と対応して変わり、F0 時間変化形状は「確信‐疑念」、「肯定‐否定」に対応して変化する。すなわち、高い音声は好印象を表し、低い音声は悪印象を与える。また、上昇、平坦、上昇＋下降、下降と変わるにつれて「疑念」から「確信」へ、また、上昇＋下降、上昇、平坦、下降と変わるにつれて「否定」から「肯定」へと印象が移ってゆく。分析の過程で、このような韻律形状と印象との関連は極めて一般性を持つことに気がついた。一語発話「ん」に限らず、通常の単語発話でも似たような単語の印象属性と韻律制御の相関関

係が認められると考えられる。もし、そのような相関関係が示されれば、多義性を持つ「ん」の場合には外部からの指定が必要であるが、単語の場合はその単語が有する印象属性を規定でき、それを用いた対話韻律の制御を考えることができる。この着想が正しいことを確認するため、３次元６種（好印象‐悪印象、確信‐疑念、肯定‐否定）の印象表現に対応する、日常よく使用される日本語からなる単語発話を用いて対話韻律の分析を行なった[12]。分析対象とする音声は、日常会話場面になるべく近い状況下での発話を促すため、各々の発話内容に即した対話状況を設定して収集した。

この結果、会話場面において収録された２４音声サンプルの対話韻律を分析により、３次元の印象空間を代表する語彙と、それら語彙の対話発話の F0 特徴量に対応関係が存在することが定性的に確認できた。さらには発話テンポも同様に出力単語の印象属性と深く関連することが判明した。

さらに、複数の単語からなる出力内容の対話制御への展開可能性を探るため、このような印象属性が組み合わされた場合について対話韻律の分析を行った。分析には、一単語の場合と同じ３次元６種の印象を与える単語を対象とし、好印象/悪印象を表す４組８個の形容詞、疑念/確信、否定/肯定を表す４組８個の終助詞を用い、副詞により程度を変化させた計２５６個の表現を使用し、同様な音声の収集、分析を行った。この結果、出力を構成する複数の単語それぞれが有する各印象属性による単独の対話韻律制御特性を重ね合わせた制御傾向をとることが判明した。

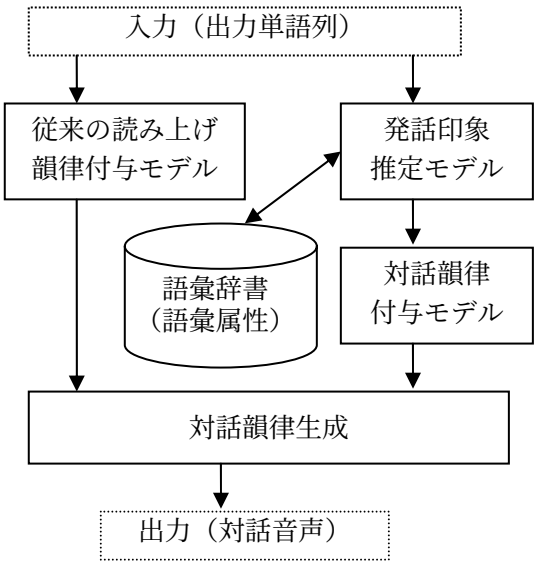

図２　出力語彙属性に基づく対話韻律制御

以上の観測結果から、同一の印象を持つ単語は、共通な対話韻律制御傾向を表わすことが確認できた。これにより、出力音声の構成単語が有する語彙印象属性を用いて、対話韻律を制御できる可能性が明らかとなった。

（２）対話韻律制御の数理モデル・合成法

前述の分析結果に基づき、出力単語の語彙印象属性に基づいた韻律制御を考えた[12]。図２に示すように、従来の書き言葉の読み上げに用いる韻律制御付与モデルにより、句のアクセントや韻律構造に基づくイントネーションを求め、これに対話韻律を加える。対話韻律の付与は、入力単語列を基にした発話印象の推定に基づいて行われる。この推定に用いる出力単語の語彙属性は、予め印象評定情報等を付与して語彙辞書に記載し、用いる。

従来の読み上げ韻律に対する対話韻律の追加は、F0 そのものの単純加算では、妥当な時間変化形状が期待できない。このため、対話韻律制御の F0 数理モデルとして、指令応答型の基本周波数生成モデル(通称、藤崎モデル)を用いた。、この生成過程モデルを用いることにより、生成パラメータのレベルで加算が可能となる。指令応答型生成モデルでは、F0 の時間変化特性は、句頭から句末に向かって緩やかな下降を示すフレーズ成分と、語の局所的な起状を示すアクセント成分によって次式のように表現できる。

$$\ln F_0(t) = \ln F_{\min} + A_p G_p(t - T_0) + [A_{a1}\{G_t(t - T_1) - G_t(t - T_2)\} + A_{a2}\{G_t(t - T_3) - G_t(t - T_4)\}]$$

式中､第１項の F_{min} は基本周波数の基底値、第２項の A_p と T_0 は、それぞれフレーズ成分の大きさと生起時刻、第３項の A_{aj} と T_j は、それぞれアクセント成分(２個の場合)の大きさと開始・終了時刻を示す。また、$G_p(t)$はフレーズ制御機構のインパルス応答関数、$G_t(t)$はアクセント制御機構のステップ応答関数であり、時間と共に指数的に減少する関数を用いて表される。

生成過程モデルのこれまでの研究にならい、本研究の対話韻律制御では、F0 については基底値 F_{min} ならびに、フレーズ成分の大きさ A_p とアクセント成分 A_{aj} および T_j を制御対象とした。また、時間長については、全体の発話テンポを考えた。単一単語だけからなる出力を対象とする場合、これらの値は、フィードフォワード型の３層のニューラルネット等により、印象表現を入力として推定できることが判明した[13] ①。この実験では、

入力として多次元尺度構成法で次元低下した３次元の印象ベクトルを用い、生成過程モデルパラメータを推定した。種々の実験条件の比較により、生成過程モデルパラメータ間の依存関係を考慮したニューラルネットの設定ならびに、元の２６次元の印象表現より、F0 形状に直接関連する３次元の印象ベクトルを用いた推定のほうが高い性能を示すことが判明した。

（３）提案対話韻律制御方法の有効性検証

提案した対話韻律制御の有効性を調べるため、対話韻律生成実験を行った。実験には、３次元６種（好印象‐悪印象、確信‐疑念、肯定‐否定）の印象毎に各２個、計１２の単語を用いた。対話韻律制御の効果を直接測ることを目的としているため、読み上げ韻律に起因する品質劣化を軽減する必要がある。このため、これらの単語を読み上げた音声を用い、それらを分析し、生成過程モデルパラメータの実測値を用いることとした。また、対話韻律の付与に関しては、単語自体による影響を除去するため、一語発話「ん」の韻律変化を用いた。すなわち、一語発話「ん」に対する各印象に対応する韻律を持つ発話、および読み上げ発話に対して、生成過程モデルパラメータの実測値を求め、それらの差異を用いて、１２単語の生成過程モデルパラメータ値に変更を加えた。参考のために、表１に変更に用いた値を示す。

有効性検証のため、これら１２単語の読み上げ韻律、対話韻律の合成音声を STRAIGHT 合成法を用いて作成した。また、妥当性を調べるため、異なった対話韻律を与えた合成音声も作成し、自然性受聴試験を行い、それらの平均オピニオン値（７段階）を求めた。この結果、本研究の提案である当該単語の印象特性に合わせた対話韻律が最もよい自然性を示し、他の韻律制御による合成音声と統計的な有意差があることが判明した。平均オピニオン値の結果を図３に示した。

（４）他言語への展開可能性等の汎用性追及

本研究で提案した対話韻律制御法は、日本語を中心として検討を進めてきたが、現時点で取り扱っている印象に関連する制御は言語に依存せずに多言語にも適用が可能であると考えられる。これを確認するため、中国語と英語の対話韻律生成を試みた[2][7]。実験には、３次元６種（好印象‐悪印象、確信‐疑念、肯定‐否定）の印象に対応する中国語と英語単語を選択し、同様の対話韻律制御を行い、合成音声を作成した。中国語と英語それぞれの母語話者による自然性受聴試験では、提案した韻律制御方法による合成音声の優位性が認められ、本制御法の有効性を確認することができた。また、本制御法の逆利用を考え、一語発話「ん」の印象認識法を考案し、実験によりその有効性を確認した[6]。

表１ 対話韻律生成のための各パラメータ変化量

	確信	疑念	肯定	否定	好印象	悪印象
Fmin	30	-25	30	-25	30	-25
Ap	1.99	1.42	1.99	2.08	1	1
Aa	2.26	2.19	2.26	1.86	1	1
発話長	0.75	1.3	0.75	1.3	1	1

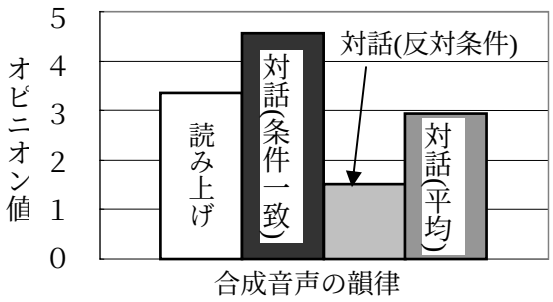

図３ 対話韻律を持つ合成音声の自然性評価

５．主な発表論文等

〔雑誌論文〕（計２件）

[1] Y. Greenberg, N. Shibuya, M. Tsuzaki, H. Kato, and Y. Sagisaka, "Analysis on paralinguistic prosody control in perceptual impression space using multiple dimensional scaling" Speech Communication Vol.51 No.7 pp. 585-593 2009 査読有

[2]Yoshinori Sagisaka“Towards Computing Phonetics”中国語音学報 Vol.1 pp. 23-37 2008 査読有

〔学会発表〕（計 14 件）

[1] Y. Sagisaka, H. Kato, M. Tsuzaki, Y. Greenberg S. Nakamura and C. Hansakunbuntheung "Computing prosody variations" Proc. IWSLPR (CDROM) 2009（招待講演）Kalkota(India) 査読無

[2] Y. Greenberg, M. Tsuzaki, H. Kato, and Y. Sagisaka "Communicative prosody generation using language common features provided by input lexicons" Proc. SNLP pp. 101-104 2009 Bangkok(Thailand) 査読有

[3] Y. Sagisaka "Corpus-based speech synthesis from reading speech to communicative speech" The first International Workshop on Spoken

Languages Technologies for Under-resourced languages 2008（招待講演）Hanoi(Vietnam) 査読無

[4] Y. Sagisaka, Y. Greenberg, K. Li, M. Zhu, M. Tsuzaki and H. Kato "Synthesis and Recognition of Communicative Prosody" The 8th Phonetic Conference of China and International Symposium on Phonetic Frontiers 2008 （招待講演）北京 査読無

[5] Y. Sagisaka, Y. Greenberg, K. Li, M. Zhu, M. Tsuzaki and H. Kato "Communicative prosody processing for synthesis and recognition of para-linguistic information" ICCA 2008 （招待講演）Yangon(Myanmar) 査読無

[6] M. Zhu, K. Li, Y. Greenberg and Y. Sagisaka "Automatic extraction of paralinguistic information from communicative speech" Proc. the 7th Symposium on Natural Language Processing 2007 pp.207-212 2007 Pattaya(Thailand) 査読有

[7] K. Li, Y. Greenberg and Y. Sagisaka "Inter-language prosodic style modification experiment using word impression vector for communicative speech generation" Proc. Interspeech 2007 pp.1294-1297 2007 Vietri sul Male(Italy) 査読有

[8] Y. Sagisaka "Prosody Generation for Communicative Speech Synthesis" Proc. Taiwan-Japan Joint Workshop on Speech Science and Technologies pp.83-90 2007 台北(台湾) 査読無

[9] 朱 明朝, 李 克, グリーンバーグ 陽子, 匂坂 芳典 「自然発話の韻律情報に基づく聴覚印象の自動抽出」 日本音響学会 2007 年秋季研究発表会講演論文集 pp.387-388 2007 山梨 査読無

[10]李 克, グリーンバーグ 陽子, 匂坂 芳典 「印象表現ベクトルに基づく言語間韻律変換」 日本音響学会 2007 年秋季研究発表会講演論文集 pp.295-296 2007 山梨 査読無

[11] Y.Sagisaka "Towards Computing Phonetics" Proc. The 7th Phonetic Conference of China and International forum on Phonetic Frontiers. (Plenary Talk) 2006 北京（中国）査読無

[12] Y.Greenberg, N.Shibuya, M.Tsuzaki, H.Kato, Y.Sagisaka "A trial of communicative prosody generation based on control characteristic of one word utterance observed in real conversational speech" Proc. Speech prosody pp. 37-40 2006 Dresden(Germany) 査読有

[13] 李 克, グリーンバーグ 陽子, 渋谷 渚, 匂坂 芳典「印象表現によるパラ言語情報を用いた韻律制御」２００６年秋季日本音響学会講演論文集 pp. 233-234 2006 金沢 査読無

[14] 匂坂 芳典「音声研究の輝かしい展開を求めて－数理モデルからの提案－」日本音声学会創立 80 周年記念フォーラム資料 p.5 Sep. 2006 （招待講演）東京 査読無

〔図書〕（計 2 件）

① K. Li, Y.Greenberg, N.Shibuya, N.Campbell, Y.Sagisaka "On the analysis of F0 control characteristics of nonverbal utterances and its application to communicative prosody generation" in NATO Security through Science Series E: Human and Societal Dynamics Vol.8 IOS Press. 179-183 2007 査読有

② 匂坂芳典,グリーンバーグ 陽子, 山下 琢美「語彙情報を用いた会話韻律生成について」音声文法研究会編 くろしお出版 文法と音声Ⅴ 第 9 章 pp. 135-145 Jun. 2006 査読無

６．研究組織

(1)研究代表者

匂坂 芳典（SAGISAKA YOSHINORI）

早稲田大学・理工学術院・教授

研究者番号：70339737

(2)連携研究者

小林 哲則（KOBAYASHI TETSUNORI）

早稲田大学・理工学術院・教授

研究者番号：30162001

誉田 雅彰（NONDA MASAAKI）

早稲田大学・スポーツ科学学術院・教授

研究者番号：90367095