View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Waseda University Repository

語彙の印象属性にもとづく対話韻律生成

Communicative prosody generation using impression attributes of lexicons

2010 年 9 月

早稲田大学大学院　国際情報通信研究科
国際情報通信学専攻　音声言語情報処理研究Ⅱ

グリーンバーグ陽子

目次

図目次

表目次

第１章　序論

1.1 音声合成のための韻律制御

書き言葉を音声出力する音声合成では、テキストからの音声合成（text-to-speech synthesis, TTS）と呼ばれるように、多くの場合、任意のテキストを入力としてそれを読み上げる音声の出力を実現する。TTS は、テキスト解析、韻律生成、スペクトル生成といった技術を必要とする。入力されたテキストは、単語辞書などを用いた言語解析がなされ、読み、アクセント、ポーズ位置など読み上げに必要な音声言語情報が抽出される。これらは、リズムやイントネーションなどの声の高さ、速さ、強さといったいわゆる韻律情報を導くのに使用される。これらの音声言語情報により、声の高さを担う声帯の振動数（基本周波数(fundamental frequency, F0)）、各音素に対する音響区分の時間長、それらの強度（振幅）といった韻律を担う音響特徴量が生成される。合成には、韻律情報とあわせて、読みから得られる音色を表すスペクトル情報を必要とする。このスペクトル情報の生成にあたっては、音声データベースから種々の方法で生成された音素、音節といった音声単位に対応する音響的情報が用いられ、音声合成器により最終的に抑揚のある滑らかな音声波形の生成が可能となる。

自然な音声を合成するためには、音声合成単位に対応して音素環境を考慮した適切なスペクトルと、音声をどのような高さ、長さ、強さで出力するかを決定する韻律の制御が非常に重要となる。これらの韻律のうち、音声の高さに関連する基本周波数 F0 は、日本語では、その局所的なパタンによってアクセントの違いを具現し、大局的なパタンは文の構造や区切りを明確にして内容の理解を助けるイントネーションを与える。F0 はいわゆる言語的なアクセントや文構造を如実に反映するが、それだけにとどまらず、書き言葉にない種々の情報を伝達し、場面に応じて自然な音声を作り出す上で鍵となる本質的な音声特徴量である。

F0 の制御には、アクセントや文構造と F0 時間変化パタン(以下、F0 パタンと略称する)の間に存在する規則ならびに、それら制御要因から F0 時間変化パタンを生成するモデルが用いられる。F0 生成過程を機能的に記述するモデルとしては、藤崎らによって提案された指令応答型の生成過程モデルがよく知られている[1]。モデルに関する詳細は本論文 4.3 節の説明に譲るが、指令応答型モデルでは、F0 パタンは、句頭から句末に向かって緩やかに下降するフレーズ成分と、局所的な起伏に対応するアクセント成分との重畳で、F0 パタンが表現されている。この F0 生成モデル関連の研究に限っても数多くの研究がなされている。日本語に対しては特に研究が進んでおり、高品質な読み上げ音声の出力を目的として、アクセント成分の強さ、フレーズ成分の強さのパラメータの量子化に関する研究[2]、これらの生成パラメータ生成の自動化をはじめ規則化を目指した種々の検討[3][4]がなされてきている。F0 パタン制御には、この生成機構のモデル化に拘泥せず、表層的なパタン生成を目指したモデルも数多く提案されている。例えば、HMM や、数量化 I 類などの最適化手法を用いて制御規則をコーパスから自動作成し、F0 パタンの概形を直接生成する方法[5][6]や、母音の重心点の F0 を設定し、その間を折れ線近似で接続することによってその F0 パタンを表現する点ピッチモデル[7]などである。基本周波数 F0 と共に、韻律を担う各音素に対応する音響区分の時間長（以下、音韻継続時間長と略称する）の制御は、音声のリズム、テンポといった韻律の自然性を与え、さらに個々の音韻の了解性にも影響する重要な課題である。音韻継続時間長は、全体の発声速度、音素の種類、前後の音素、音素が含まれる呼気段落や句のモーラ数、その文内の位置、統語的属性など種々の要因によって影響を受けることが知られている[8]。それらの要因を考慮した、音声合成における音韻継続時間長の制御については、音声データベースから統計的手法を用いた制御規則の推測が行われている。計算モデルとしては、数量化 I 類を用いるもの[9,10]、積和型の重回帰モデル[11，12]や拘束条件付き回帰木モデル[13]等が提案されている。

以上のように、これまでに、自然な音声合成を生成するために不可欠な、適切な韻律制御を目的とした、読み上げ音声の韻律規則に関するモデルの提案は盛んに行われてき

た。さらにコンピュータの能力向上と共に、コーパスの有効利用を図るコーパスベース音声合成技術の展開により、自然性の高い合成音声の出力が可能となってきた[14][15]。この合成音声の高品質化は、TTS として開発されてきた合成方法の枠を越え、より広い分野への応用を加速させており、その過程で、韻律制御の重要性が再認識されてきている。

1.2 多様な韻律生成に関する従来の研究

合成音声の高品質化に伴い、本来対象とされてきた読み上げ音声の出力に留まらない、対話音声が要求されるような場面での使用が期待されている。対話音声の合成を可能にするにあたっては、書き言葉を対象とした読み上げに必要な従来の言語情報に加え、話し言葉が持つ新たな情報の利用が必要である。これまでにすでに明らかにされているように、対話音声が具備すべき韻律特性は読み上げ音声と大きく異なっている。読み上げ音声との比較からの、生成過程モデルに[1]よる対話音声の制御特性の分析[16]や、韻律的特徴分析[17]、時間制御特性分析[18]などの従来研究により、対話音声の取り得る韻律のバリエーションの多様さが示されている。

近年、このような多様な対話韻律と関連して、感情音声、いわゆるExpressive speechに関する研究は精力的に行われている。それらの研究においては、対話韻律特性は、いわゆる言語情報と関連しない「パラ言語情報（言語外情報）」として、感情などを含む感性情報と同定し、感情や、発話場面などで指定される韻律の特性の解明や、感情音声の合成などの試みがなされている。喜び、悲しみ、怒りを表現する話し言葉に表出される韻律特徴の分析[19]、怒りの度合いに対応した韻律的特徴分析[20]、感情表現の内容によって用いられる音響特徴への重みづけの違いに関する研究[21]、また実際の対話場面での自然発話を分析対象とした、表出される感情とF0特徴との分析[22]や、話者感情とそれらを表現する際の対話韻律特徴の関係分析[23]など、感情や対話場面などで指定される対話韻律特性の解明は盛んに行われている。さらに、実際の合成においても、同様の手法で、怒り、喜び、悲しみの感情が指定する対話韻律音声の合成が行われている。

用いられる方式としては、コーパスベース[24][25][26][27][28]や、HMM　[29][30][31][32]などがある。HMMと素片接続型を組み合わせたハイブリッド音声合成を用いて、顔のアニメーションに合わせて、喜び、驚き、怒り、悲しみの感情音声をリアルタイムで出力する視覚的音声合成法の提案[33]もされている。その他にも、Gaussian Mixture Model (GMM)を用いて、平静な音声を、怒り、喜び、悲しみの3種類の感情音声へと変換する方法[34]や、感情音声の韻律パラメータに対し主成分分析を行うことで韻律の部分空間を算出し、それらと主観評価実験から抽出した感情を対応付けることで、感情から韻律を合成する手法の提案なども行われている[35]。以上のように、感情や発話場面、発話様式などのカテゴリによって出力韻律を規定することにより、読み上げ音声と大きく異なる対話韻律を反映させた、表現力のある音声の合成が試みられてきた。

1.3 対話音声合成のための韻律生成

多くの感情音声合成研究にみられるような、典型的な感情表現などのカテゴリにもとづく韻律制御によって、表情豊かな感情音声を出力することが可能となった。しかしながら、それらは感情音声の出力には有効であっても、対話システムに求められる対話音声の出力に対応するには十分でない。すなわち、対話システムに有効な音声合成の実現のためには、これまで取り組まれてこなかった、次に挙げる主に 2 つの課題を解決することが不可欠である。

1. 対話場面に出現する多様な韻律の規定

   対話韻律の制御を考えるにあたって、まず読み上げ韻律と大きく異なることが判明している対話韻律の規定を行う必要がある。対話韻律に影響を与えていると考えられる、対話者が表出する意図や発話状況の規定や表現などを定量的に取り扱うことは難しいが、対話韻律の規定は避けて通れない問題である。さらに対話音声合成のためには、それら規定内容が対話韻律制御特性と対応付けられることが必要である。

2. 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

読み上げ音声のような一方通行の発話に比べて、対話音声では、双方向から発話のやり取りという状況下での音声出力となる。従って、対話システムに求められる音声の合成には、対話場面で用いられる多様な韻律を、発話内容に即して、語句のレベルで動的に制御する仕組みが必要となる。

本論文では、以上の課題を解決し、対話システムに有効な、対話韻律の規定方法および、対話韻律制御方法の提案を行った。まず(1)の対話韻律規定の問題を、対話韻律の違いが聞き手に与える印象として検討した。さらに、その印象を入力情報、対応する韻律特徴を出力情報として、(2)の動的に対話韻律を制御する仕組みを解決する方法を提案した。

提案する対話韻律生成方法では、入力となる語彙の印象という属性に着目した。対話韻律に影響を及ぼすと考えられる要因は数多く考えられるが、それらをその対話場面の種々の状況から抽出し、音声合成に用いることは現実的に難しい。このため、そのような難しい対話発話に関する種々の詳細な対話の情報規定ではなく、対話の言語内容自体が、対話が伝える発話行為(speech act)と関連して、取りうる韻律を制約する点を取り上げた。例えば、「きれい」という対話発話は、多くの場合、好印象を与え、語彙によって取り得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」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、醸し出す雰囲気といった情報に関連した韻律が用いられることが多いと推察される。無論、現実の対話では、発話語彙が与える韻律だけでは不十分で、対話状況による種々の考慮が必要であると思われるが、工学的には、発話語彙の印象情報が与える対話韻律の追加だけでも有用であると考えられる。このように、対話韻律の制御問題を、入力される語彙が規定する対話韻律の推定問題として解くことを考えた。

1.4　本論文の構成と各章の概要

本論文では、対話システムに有効な対話韻律生成の実現を目指して、まず前節で述べた課題(1)に対応する対話韻律の規定方法を提案することから始めた（第 2，3 章）。対話韻律を規定するものとして印象という情報を用い、対話韻律の違いが与える聴覚印象の分析によって、対話韻律の取り得る自由度と、それらを規定するための印象の記述を行った。続いて、それら印象と印象が規定する韻律特徴の関係を入出力情報として、課題(2)に対応する動的な韻律制御を可能にする、入力語彙の印象属性にもとづいた対話韻律生成方法の提案を行った（第 4 章）。さらには、提案した韻律生成方法を拡張できる可能性の検証を行った（第 5，6 章）。以上、対話韻律生成の実現への課題と対応する本論文の章番号を図 1.1 に示す。各章の概要は次の通りである。

対話音声合成を行うためには、まず合成出力として多様な韻律の中から所望の対象韻律を規定するための、何らかの方法が必要である。またこの出力規定にあたっては、規定内容とそれに対応した対話韻律制御特性と対応付けられることが、音声合成を可能とするために必要である。第 2 章では、この出力規定の課題を、種々の韻律を持つ一語発

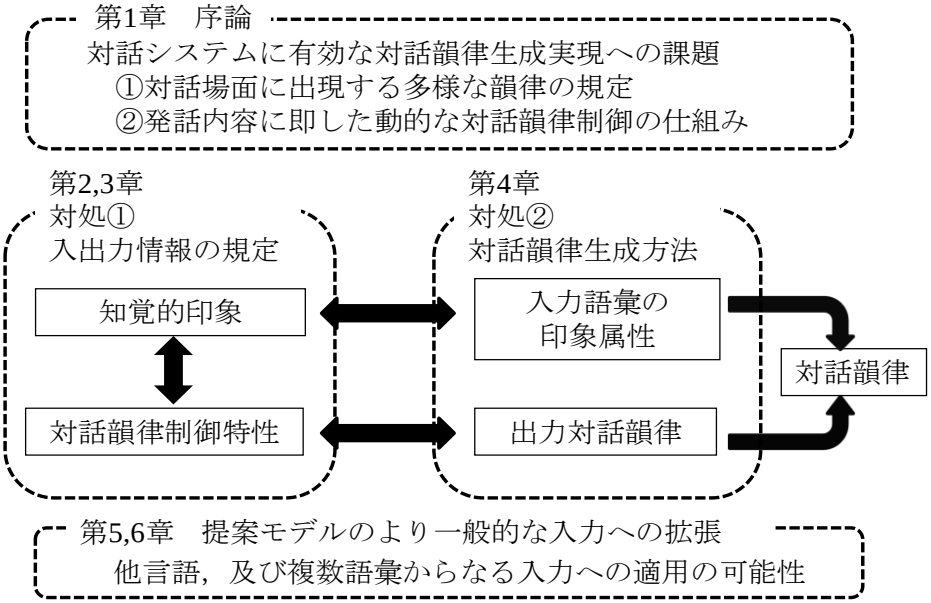


図1. 1 本論文の課題と章番号との対応

話「ん」を例に検討した。出力音声が与える聴覚印象にもとづいた規定方法を考え、多次元尺度構成法を用いた対話韻律の分析を行った。その結果、3 次元による対話韻律が与える印象の記述、また、それら印象表現と対応する対話韻律制御特性の関係が明らかとなった。続く第 3 章では、それら印象と対話韻律の対応関係が、入出力情報として対話韻律生成に用いることができる可能性の検証を行った。一語発話「ん」は語彙としての印象属性を規定しないため、種々の対話韻律をとれるが、一般的な語彙は「ん」に比べて限定され使用となり、その語彙自体が有する印象属性によって対話韻律がある程度規定できることが期待される。このため、一語発話「ん」の対話音声に見られた印象 - 韻律の対応関係が一般の語彙でも成立するかどうかを、印象表現に直接関連する対話表現を用いて調べ、その有効性を示した。

第 4 章では、前章までで明らかにした、印象と対話韻律の対応関係を参考に、語彙が与える印象にもとづいた対話韻律生成方法を提案した。提案する方法では、従来の言語情報にもとづく読み上げ調の韻律制御に加え、新たに、入力される語彙自体が有する印象属性にもとづいた韻律特徴を加えた対話韻律生成を行う。提案方法にもとづいて生成した音声サンプル用いた自然性聴取実験により、提案方法の妥当性を確認した。

第5章と第6章においては、提案した対話韻律生成方法の拡張の可能性について検証した。第5章においては、提案方法で扱う語彙の印象属性のレベルでは、言語依存性が低く、他言語の対話音声合成への適用の可能が考えられるため、英語を対象にした対話音声の合成を試みた。日本語語彙を対象とした韻律生成実験と同様の手順で、対話韻律生成を行い、自然性聴取実験によって、その妥当性の確認を行った。

第6章においては、単独語彙を用いて検証してきた提案方法を、より一般的な、複数語彙の組み合わせによって構成される入力文に展開するための検討を行った。異なる印象属性を有する複数語彙から構成される文発話の対話韻律分析を行った結果、各語彙の印象属性に対応する対話韻律制御特性を加え合わせることで、文全体の出力対話韻律が説明できることが判明した。すなわち、提案した対話韻律生成方法が、単独語彙のみでなく、より一般的な入力に対しても適用できる可能性が示唆された。

最終章では、本論文を総括し、提案した語彙が有する印象属性にもとづいた対話韻律生成方法の意義、判明した実験事実の整理と、一般的な対話音声合成への展開に向けた将来への課題を示した。

第 2 章　印象にもとづく対話韻律の規定

2.1　対話韻律生成に用いる情報の規定と制御方法

対話韻律生成を行うためには、会話場面に出現する豊富な韻律の中から、出力となる所望の対象韻律を規定するための何らかの方法が必要である。また音声合成を可能とするためには、出力される対話韻律を規定する内容は、韻律を制御する入力情報として扱わわれる必要がある。このためこの出力対話韻律を制御する入力情報の規定の課題を、種々の韻律を持つ一語発話「ん」を例に、発話される対話韻律特徴と、それらの違いが聞き手に与える聴覚印象の対応関係として検討した。

会話中に出現する一語発話「ん」の韻律特徴は、聞き手への伝達情報を表現していると考えた。日本語で、一語発話「ん」は、発する内容や状況において、感嘆詞や返答、つなぎ言葉などになり得る。一番重要なことは、「ん」自体は、特定の意味、またデフォルトのイントネーションやアクセントなどを持たない。従って、「ん」の韻律バリエーションによる聴覚印象を検討することで、出力対話韻律と、それらを制御する印象との関係を明らかにすることを期待した。以下、2.2節では、日常会話に出現した一語発話「ん」が伝達する内容の観察結果をもとに、F0特徴の違いによって与えられる聴覚印象の記述を行った。さらに、2.3節では、限られたデータによって示したF0特徴の分類の妥当性検証のために行った、大規模な日常会話データを用いた、F0特徴分類に関して述べる。2.4節では、観察されたF0制御特徴の違いが聞き手に与える聴覚印象の定量的表現を得るために、多次元尺度構成法（Multidimensiona Scaling, MDS）を用いた分析を行い、2.5節において、出力対話韻律特徴とそれらに対応する聴覚印象の関係を明らかにした。最後に2.6節でまとめる。

2.2 一語発話「ん」に対する聴覚印象にもとづく対話韻律の規定

2.2.1 聴覚印象にもとづく F0 特徴の分類

対話場面に出現する対話韻律特徴と、それらの違いが与える聴覚印象の規定を行うため、日常会話で多用される一語発話「ん」の韻律の多様性を、それら発話が伝達する表現内容との関係から観察した。会話中に出現した一語発話「ん」が、どのように聞き手によって理解されたかを分析することによって、言語情報を取り除き、純粋に韻律情報のみによって伝達される情報の規定、およびそれらに対応する対話韻律制御の自由度の観察を試みた。

分析に際しては、まず、30～35 歳の、友人同士の親しい関係にある成人女性 4 名が行った日常会話から得られた、42 の一語発話「ん」を用いた。話者らは、東京方言/標準語を話す日本語母語話者であった。会話は、静かでリラックスしたムードの中で行われ録音した。まず、WaveSurfer[36]のインターフェースを用いて、手動で「ん」のサンプルを取り出した。また例えば、「ん！美味しいね！」「ん～、嫌だなぁ」のように、「ん」の後続発話の言語内容から、「ん」が包含する伝達情報を予測できると考え、そ

表2.1　F0平均値・パタンによって分類された一語発話「ん」の後続発話

パタン / 高さ	上昇	平坦	下降	上昇＋下降
高	本当？♪ 何？何？♪	それで、それでっ♪ そしたら？♪	そうだったんだっ！♪♪ もちろんっ！♪	気にしないで!♪
↕	うそー？ そうなの？	そうだねっ いいよ	いいんじゃない♪ 構わないよ	そんな事ないよ
↕	そうなんだ↓	そうかなぁ でも、大丈夫？	まぁ、そうだね‥↓	だって、嫌だよ
低	本当？↓	それはどうかなぁ 違うんじゃないかな	そうだったんだ‥↓ いやっ、でもね･･･↓	

れぞれの後続発話の抽出も合わせて行った。「ん」の後続発話が示す伝達内容と、対応するサンプルの F0 を観察した結果、F0 の平均値・パタンによって、相手の発言に対しての返答、心的状況、問いかけなどの情報が担われていることが判明した。表 2.1 に、分類したサンプルの後続発話を示す。

2.2.2 聴覚印象を示す印象表現の規定

後続発話によって表現される伝達内容をより精確に表現するため、一語発話「ん」の対話韻律の違いが与える聴覚印象を印象表現として記述することにした。先の分析でみられた F0 の平均値とパタンを制御対象として考えるため、平均的高さとパタンが異なる一語発話「ん」を 12 種類(平均的高さ 3 種類(高･中･低)×パタン 4 種類(上昇･平坦･下降・上昇＋下降))用意した。音声発話は著者自身が行い、F0 の平均値とパタンが 12 種類の各カテゴリの典型例になるよう注意し、また意図的な感情表出を避けるため、特定の発話状況を意識しない発話を心掛けた。なお、作成したサンプルの F0 のパタンを図 2.1 に、最高と最低の F0 の高さを表 2.2 に示す。

これらの F0 平均値とパタンの異なる 12 種類の異なる音声を用いて、それら韻律バリエーションが聞き手に与える聴覚印象の違いを記述するための評定実験を行った。評定は東京方言／標準語を話す日本語母語話者の成人 5 名(男性 2 名、女性 3 名)が行った。評定実験では、表 2.1 作成時の経験を参考に、次に続くことが予想される句表現、またそれらから想定される発話者の表現しようとする内容を、極力、形容詞または副詞で直感的に表現してもらうように指示した。なお各サンプルに対する複数回答をその結果、67 表現が得られ、その中から複数被験者によって回答された、26 表現を印象表現語として選択した。具体的には、“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、怒っている、ふてぶてしい、うざい”であった。

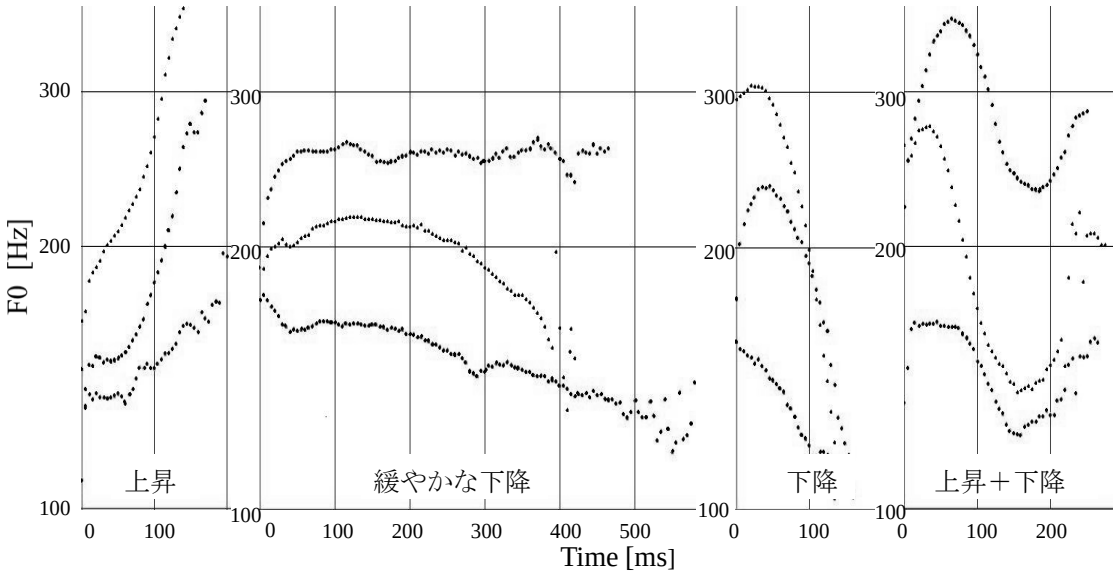


図2.1 聴覚印象を記述する印象表現語の選出に用いた
一語発話「ん」の音声サンプルのF0パタン

表2.2　聴覚印象を記述する印象表現語の選出に用いた
一語発話「ん」の音声サンプルのF0の最高・最低値

高さ / パタン	高		中		低	
	最高	最低	最高	最低	最高	最低
上昇 ↗	354.55.	182.20	282.58	142.17	194.21	98.24
平坦 →	264.55	232.39	213.23	178.90	162.77	124.97
下降 ↘	305.50	119.81	234.22	90.87	155.94	65.98
上昇と下降 ∫	363.46	222.15	273.08	153.48	163.17	111.44

2.3　実発話データに見られる一語発話「ん」の韻律分析

日常会話に出現した一語発話「ん」を用いた対話韻律分析により、F0 の平均値とパタンの違いが与える聴覚印象を、印象表現として記述することができた。しかしながら、分析に用いた日常会話データは、非常に限られたデータであった。このため韻律制御特性として挙げた F0 パタンの分類が、実際の日常会話に出現するパタンを十分に反映できているかを確認する必要がある。そのため大規模な実発話データ中から抽出した一語発話「ん」を用いた、F0 パタンの分類を行った。

2.3.1　音声資料の概要

日常会話場面で出現する一語発話「ん」の F0 のパタンの分類を行うため、JST(Japan Science & Technology Agency) CREST (Core Research for Evolutional Science and Technology) ESP (Expressive Speech Processing) プロジェクト[37][38][39]によって収録された膨大な量の日常会話発話データから抽出した一語発話「ん」の分析を行った。

このデータは、日本人母語話者の成人女性の 150 時間に渡る音声を収録したものであり、その中には、23,648 サンプルの一語発話「ん」が含まれていた。その中から 6,271 サンプルをランダムに抽出し、分析に用いた。

2.3.2　ベクトル量子化

実際の生活環境化で出現する一語発話「ん」の F0 のパタンにどのようなパタンが存在するかを把握するために、ベクトル量子化(VQ)を行うこととした。発話時間のばらつきが 50 ms～400 ms と大きかったため、量子化に先んじて、50 ms 毎に 7 個のカテゴリに分け、そのカテゴリ毎にベクトル量子化を行った。

量子化に際しては F0 パタンの値そのものを単に線形伸縮せず、生成モデル[1]に従った制御指令時間の変更によるパタンの再合成を行った。すなわち、手動抽出したアクセント指令位置を時間長に合わせて伸縮し、再合成した F0 パタンを用いた。VQ には K 平均法を用いた。クラスタの分割数は、図 2.2 に示すように、クラスタ数の増加に伴う総歪みの減少に適当な閾値を用いて、それぞれ発話時間の 7 個のカテゴリに 20 のクラスタに決定した。従って、140 のクラスタが得られた。

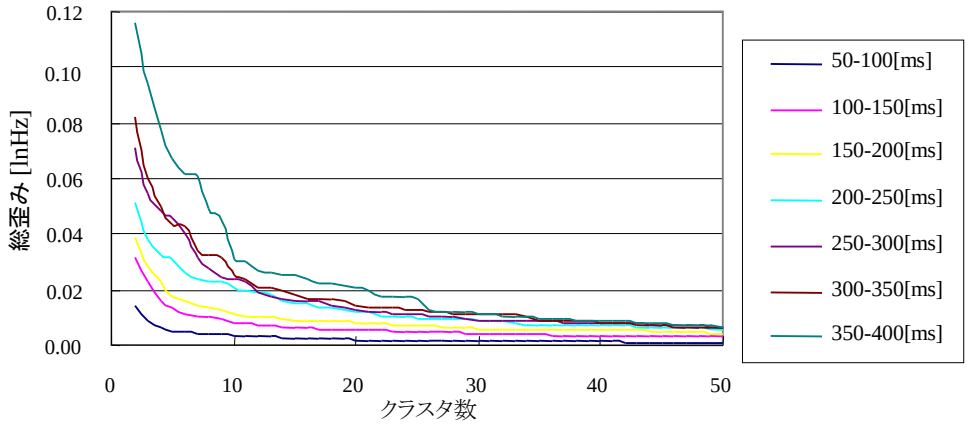


図2.2 クラスタ数の増加に伴う総歪みの減少

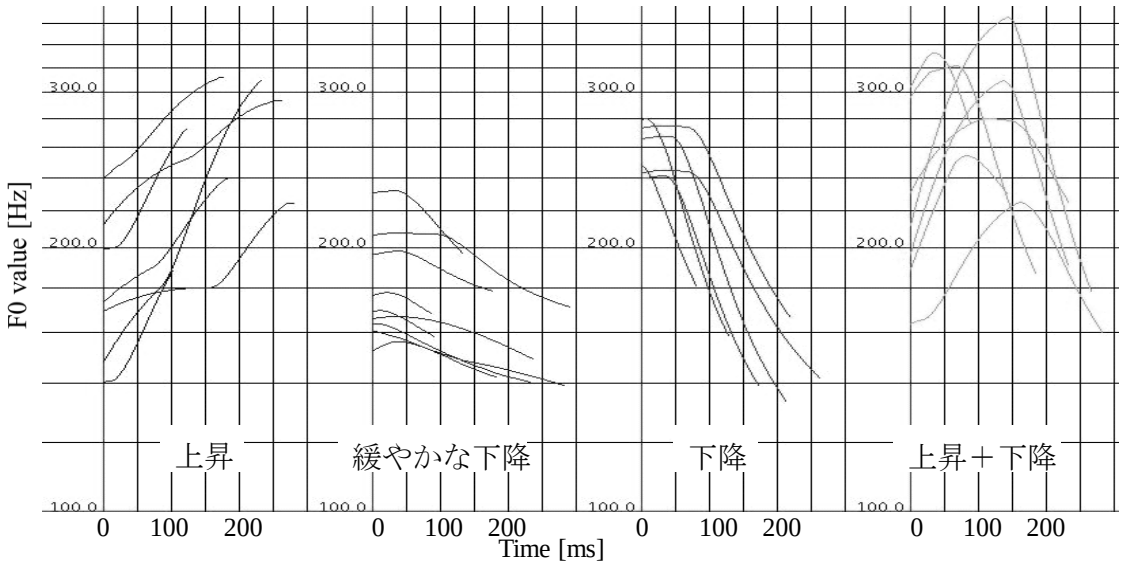


図2.3 日常会話で観測されたF0の典型パタン

2.3.3　日常会話における一語発話「ん」を用いたF0パタン分類

得られたクラスタを観測した結果、F0のパタンは、2.2節で示したものと同様の4つのパタン（上昇、緩やかな下降、下降、上昇＋下降）に大別できることが判明した。図2.3に、各クラスタの中心に最も近いF0 パタンの例を示す。

このデータには、好ましくない印象、例えば、不機嫌であったり、相手に不快感を与えるような表現などは含まれていない。このため、この4つのパタンで、一語発話「ん」の存在しうる全ての韻律パタンを網羅できたとは限らないと考えられるが、限られた「ん」のモデル音声で示したパタンが、最も一般的に出現するものであった。この事実から、録音対照としているような環境下にあっては、ここで取り扱うパタン形状が一般的であることが裏付けられた。

2.4 多次元尺度構成法を用いた聴覚印象の次元低下

2.4.1 印象表現による聴覚印象ベクトル表示

対話韻律によって与えられる印象を規定するために、聴覚印象のバリエーションや違いを言葉で記述する必要がある。韻律によって与えられる、漠然としたイメージを細かに記述し、また個人の感覚に依存する所が多い微妙なニュアンスを表現するために、印象を多次元で定量化するべきである。そのため、2.2 節で得られた 26 の印象表現語を用いて、対話韻律を制御する聴覚印象を表現することにした。まず印象表現語をもとに、F0 の平均値とパタンの違いが与える聴覚印象の違いを確認するため、2.2 節で用いたものと同一の 12 の一語発話「ん」の音声サンプルを用いて聴取実験を行った。

評定者らに、それぞれの一語発話「ん」のサンプルに対し、26 印象表現に 0(全く当てはまらない)～7(とても当てはまっている)の 8 段階評定を求めた。評定者としては先の評定者とは異なる、25 歳～33 歳の東京方言/標準語を話す、日本語母語話者の成人 5 名(男性 1 名、女性 4 名)を用いた。評定者は、必要なだけサンプルを聞きなおすことが許され、実験の平均所要時間は 30 分～40 分程度であった。

2.4.2 印象表現による対話韻律制御特性の分析

複数の印象表現語をもとに、対話韻律が伝達する情報を定量的に表現するため、多次元尺度構成法 MDS[40]を用いた分析を行った。MDS を用いることにより、距離を表す

データをもとに独立な次元を求め、各サンプルが従う構造や制約の多次元表現・理解を期待した。分析には、評定者が複数の場合の個人差を考慮に入れた INDSCAL アルゴリズムによる MDS を用いた。INDSCAL では、通常の MDS で得られる空間が個人によらず共通でありことを仮定しており、個人による対象相互の類似度の差異は、刺激空間に対する個人別の重みによって異なることによるとするモデルとなっている。

入力データは、26 の印象表現語をもとにした、各刺激間の評定値差によって得られる距離行列を入力データとした。適切な次元は、経験値から得られることが多いが、本分析では、表 2.3 に示す、比較的低次元で説明できる分散の割合(VAF)を参考に、3 次元を採用した。また、表 2.4 に示すように、3 次元に対するそれぞれの被験者の重みは適切なようであったので、全被験者の距離行列データを用いることにした。各軸の解釈を行うために、重回帰分析を用いて、それぞれの印象表現語に対する平均評定値を、3 次元空間に射影させた。図 2.4 に示すように、3 次元と、印象表現によって特徴付けられる軸との関係が観察された。

表2.3　各次元に対する分散の割合 (VAF)

	次元			
	1	2	3	4
VAF	0.7398	0.8036	0.816	0.5952

表2.4　3次元に対する各被験者の重み

被験者	次元		
	1	2	3
SY	0.9459	0.7651	0.8470
CA	0.4933	0.7668	0.7051
YY	0.6428	0.7821	0.8761
FY	0.7063	0.4332	0.6148
KK	0.5833	0.7868	0.6851

2.5　聴覚印象と韻律制御の関係分析

26 印象表現語を用いた、対話韻律特徴と聴覚印象の関係を分析した MDS 分析の結果を図 2.4 に示す。それぞれの次元に布置された音声サンプルの韻律特徴と、各軸に関連していた印象表現との対応を次元毎に以下にまとめる。また、射影された 26 の印象表現語をもとに、3 次元軸をそれぞれ、「好印象 - 悪印象」、「確信 - 疑念」、「肯定 - 否定」と呼ぶことにする。具体的には、“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”が「好印象 - 悪印象」の軸と、“納得、了承、疑い、迷い、疑問”が、「確信 - 疑念」の軸、“同意、否定、反論”と「肯定 - 否定」の軸が、それぞれ相互に関連している。

図 2.4(1)、(2)

第 1 次元・第 2 次元、第 2 次元・第 3 次元の平面上

　韻律特徴：　F0 平均値（高・中・低）

　印象軸　：「好印象 - 悪印象」

図 2.4(1)、(3)

第 1 次元・2 次元、第 1 次元・第 3 次元の平面上

　韻律特徴：　F0 パタン（下降、上昇＋下降、平坦、上昇）

　印象軸　：「確信 - 疑念」

図 2.4 (3)

第 1 次元・2 次元、第 1 次元・第 3 次元の平面上

　韻律特徴：　F0 パタン（下降、平坦、上昇、上昇＋下降）

　印象軸　：「肯定 - 否定」

以上のように、F0 特徴量と関連付けられる印象を 3 次元で近似的に記述することができた。またこれらの結果を、F0 制御の観点から見直すと、「好印象 - 悪印象」といった聴覚印象により F0 平均値を、「確信 - 疑念」、「肯定 - 否定」は F0 のパタンを制御することが考えられる。

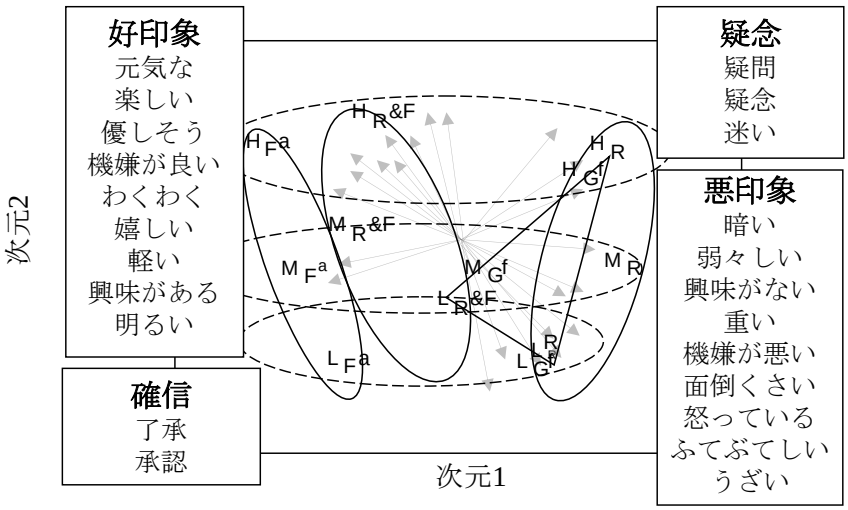


図2.4　3次元空間における印象表現語の投影
(1) 次元1 – 次元2

凡例：同一のF0平均値（H：高,M：中，L：低）を破線,F0のパタン（R：上昇，Gf:緩やかな下降，Fa：下降,R&F：上昇+下降）を実線でそれぞれ囲んである.

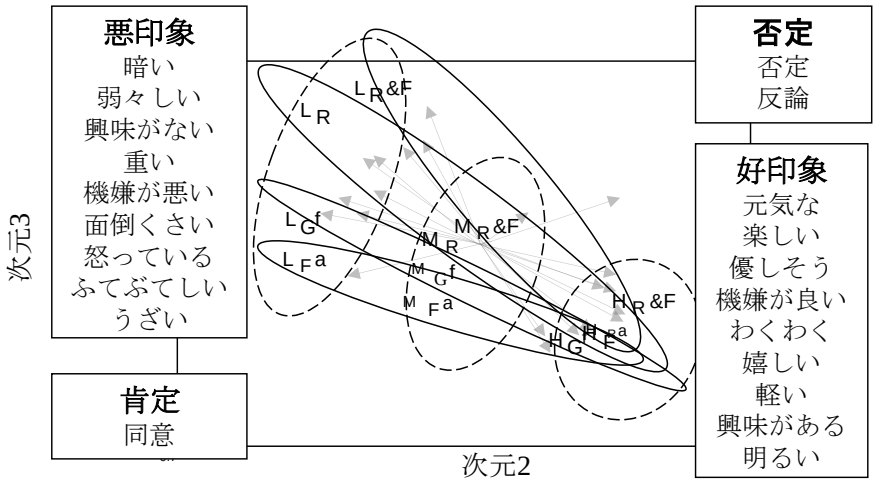


図2.4 (2) 次元2– 次元3

（図2.4(1)のタイトルおよび凡例を参照）

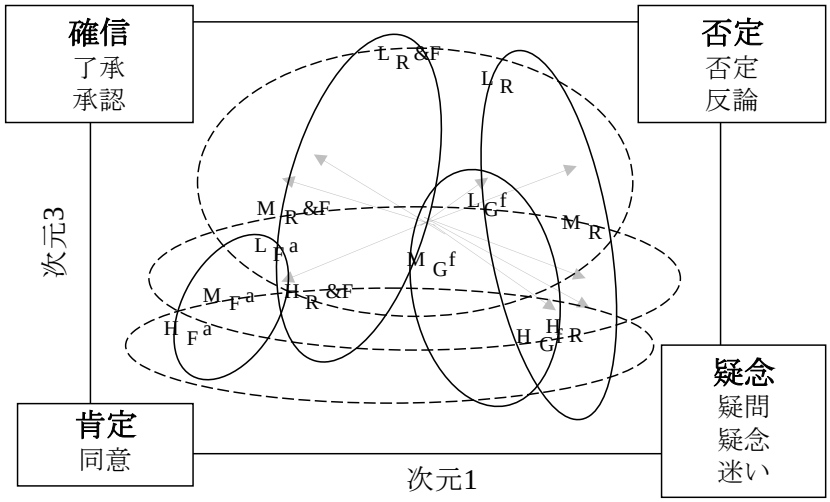


図2.4 (3) 次元1– 次元3

（（図2.4(1)のタイトルおよび凡例を参照）

2.6　2 章のおわりに

本章では、対話韻律生成に用いる入出力情報の規定を目的として、出力韻律を規定する方法の提案を行った。提案方法では、韻律を規定するものとして印象という情報に着目し、それら印象情報と対話韻律制御特性との対応付けを行った。

一語発話「ん」を参考に、対話韻律のバリエーションとそれらの違いが与える聴覚印象の対応関係を、MDSを用いて分析した結果、対話韻律の取り得る自由度と、それらを制御する印象の記述ができた。3次元（確信 - 疑念，肯定 - 否定，好印象 - 悪印象）による対話韻律が与える印象の記述ができ、さらには、それら聴覚印象と、F0特徴（F0の平均値とパタン）とに明確な対応関係が存在することが明らかとなった。すなわち、3次元聴覚印象によるF0制御の可能性を示した。

第 3 章　語彙の印象属性と対話韻律の聴覚印象

3.1　語彙が持つ印象属性の対話韻律生成への利用

一語発話「ん」を用いたMDS分析により、対話韻律特徴（F0の平均値とパタン）と、それらを制御する3次元の聴覚印象空間（好印象 - 悪印象、確信 - 疑念、肯定 - 否定）との関係が明らかとなった。これらの対応関係を用いて、「きれい」、「汚い」、「怪しい」といった入力される語句が規定する対話韻律の生成を考えることができる。すなわち、言語形式として意味を限定しない「ん」の韻律を規定するには、外部からの指定が必要であるが、「きれい」、「汚い」、「怪しい」では、それらの語彙が有する、「好印象」、「悪印象」、「疑念」といった印象属性によって、対話韻律を制御する入力情報の規定ができるのではないかと考えた。このためこれら各語彙の一語発話としての対話韻律と、各語彙が与える「好印象」、「悪印象」、「疑念」と同様な印象を与える一語発話「ん」の韻律特徴との共通性の確認を行うために、印象表現に直接対応する語彙発話の対話韻律分析を行った。以下、3.2節では、まず分析対象となる3次元の印象空間を代表する語彙群の選定、及び、それら選定した語彙の発話収録に関して述べる。 3.3節では、発話語彙の印象属性にもとづいた対話韻律の制御特性という観点から行った、対話韻律分析の結果を述べ、一語発話「ん」が示した分析結果との共通性の確認を行った。最後に3.4節でまとめる。

3.2　一語発話に見られる語彙の印象属性と対話韻律の聴覚印象

3.2.1　聴覚印象空間に対応する印象属性を有する語彙の選定

一語発話「ん」の分析が示した、好印象 - 悪印象、確信 - 疑念、肯定 - 否定からなる 3 次元の印象空間と韻律制御特性の関係にもとづいた対話韻律生成の可能性を検証するため、それらの印象表現に直接対応する語彙についても同様に、印象と韻律の対応関係が成立するかを調べる。このため、3 次元の印象空間を代表する語彙を選び、それら語彙発話の韻律分析によって、語彙の印象属性と対話韻律との対応関係の検討を行った。分析には、表 3.1 の語彙欄に示した 3 軸 6 方向（好印象、悪印象、確信、疑念、肯定、否定）の典型印象表現に対応する、日常よく使用される日本語の語彙を用いた。

実験に先立ち、選定した各語彙が、6 方向の印象表現を確実に反映しているかどうかを確認するために、各語彙に対する主観評価実験を行った。選定した 25 の語彙に対して、6 方向に対応する 16 の印象表現語が該当するか否かを、それぞれ、0（全く当てはまらない）～6（非常に良く当てはまっている）の 7 段階で評定させた。16 の印象表現語は、2.2.2 節で、3 次元の印象空間を導き出すのに用いた 26 個の印象表現語から、評定者への負担や好印象 - 悪印象に対応する表現語の多さによるバランスの悪さを考慮して選択した印象表現語である。具体的には、「好印象」 - 「悪印象」に対応する 8 表現語「わくわく、明るい、嬉しい、軽い」 - 「ふてぶてしい、暗い、悲しい、重い」、「確信」 - 「疑念」に対応する 4 表現語「納得、了承」 - 「疑い、迷い」、「肯定」 - 「否定」に対応する 4 表現語「同意、賛成」 - 「否定、反論」であった。評定実験には、東京方言／標準語を話す、28 歳～38 歳の日本語母語話者の 4 名（男性 1 名、女性 3 名）が参加した。

表 3.1 に各印象表現に該当する印象表現語の平均評定値、評定者間相関値の分布範囲を示す。平均評定値のうち「当該印象表現」は、選定した語彙が意図した印象表現に対応する印象表現語群に対して得た評定値の平均である。具体的には、例えば、“確信”の印象表現を意図して選択した“絶対”、“間違いない”、“確かに”、“納得”の 4

表3.1　対話音声収録で用いた語彙と意図したそれらの印象の一致

印象表現	語彙	平均評定値		評定者間相関の範囲
		当該印象表現	その他の印象表現	
確信	絶対、間違いない、確かに、納得	5.1	0.31	0.60－0.98
疑念	怪しい、疑わしい、迷う、何故	4.5	0.09	0.62－0.97
肯定	いいよ、そうだね、賛成、当り	5.4	0.25	0.87－0.98
否定	嫌だ、無理、外れ、反対、違う、	5.3	0.12	0.81－0.96
好印象	嬉しい、きれい、面白そう、面白い	5.1	0.04	0.89－0.99
悪印象	可哀想、汚い、難しい、退屈	4.5	0.06	0.92－0.98

つの語彙がそれぞれ、確信の印象表現に対応する印象表現語の“納得”、“確信”に対して得た評定値の平均値を指す。また、意図しないそれ以外の印象表現語に対する平均評定値を、比較として「その他の印象表現」の欄に示す。さらに、それぞれの語彙に対する評定が、評定者間で、大きなばらつきがなかったかを調べるために、語彙毎に、評定者間相関を求めた。表には、印象表現毎に、それぞれの語彙の評定における評定者間相関値の範囲を示す。表 3.1 に示すように、選定した語彙の、意図した印象表現語に対する平均評定値はどれも 4.5 以上となり、意図しない他の印象表現語における平均評定値 0.31 以下と比べて高い値となっている。また評定者間相関値は、いずれも 0.6 以上であった。このように、選定した語彙は意図した印象表現を反映できており、さらに、それらの評定は、妥当なものであることが確認できた。

3.2.2　対話音声収録

選定した語彙が対話場面において、どのように発話されているかを調べるために、音声収録を行った。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、それぞれの

印象表現に即した発話状況を語彙毎に設定した。発話者には、提示した状況を自由に想像してもらい、極力自発的な対話を促した。例えば、「確信」の印象表現の代表として選択した「絶対」という語彙に対しては、印象としての「確信」を意識付けるため「相手の不安を払拭するための発話」というように、相手を考えた状況を想像した発話を依頼し、発話を収録した。発話者は、25 歳～31 歳の東京方言／標準語を話す日本語を母語とする成人 4 名（男性 2 名、女性 2 名）であった。話者 1 名につき、表 3.1 に示す 25 の語彙に対応する合計 25 発話を静かな環境で録音した。また比較のため、対話音声の録音後に、同一発話内容の読み上げ調発話も録音した。

今回の音声収録における発話者はいずれも声優ではなかったため、相手を想定した自然発話の発声は難しかったと思われる。そのため韻律分析に先立ち、収録された対話音声が、対話に出現する発話として十分自然な音声であったかの確認のため、自然性評価実験を行った。各発話に対し、どのような印象がどの程度知覚されるかを、0（全く当てはまらない）～6（とても良く当てはまっている）の 7 段階で評定させた。評定項目としては、前節の主観評価実験と同じ 16 の印象表現語を用いた。評定者は、先の発話者とは異なる、24 歳～37 歳の日本語母語話者 5 名（男性 3 名、女性 2 名）であった。また音声刺激は、反復聴取可能な形で提示した。収録した発話が自然なものであれば、発話から得られる印象は、より多くの聞き手によって共有されると考えられる。従って、得られた 100 発話（＝25 発話×4 話者）のうちから、当該印印象表現に対して比較的高い得点（4 以上）と、全体評定において、比較的高い評定者間相関（0.70 以上）を示した 24 発話を選出し、それらを韻律特性の分析対象とした。内訳は、6 つの印象表現、確信、疑念、肯定、否定、好印象、悪印象に対して、それぞれ 4、6、7、2、5、0 発話であった。今回の収録においては、悪印象の印象属性を備えた語彙の発話を得ることができなかった。今回のような人工的な対話設定下では、特に、話者にとって悪印象を表現するのは難しかったと思われる。

3.3　語彙の印象属性による対話韻律の説明可能性

発話語彙の印象属性が制御する対話韻律特徴という観点から、収録した対話発話韻律の分析を行った。一語発話「ん」が示した、3 次元印象空間に対応する F0 特徴（F0 の平均値・パタン）との共通性を確認するため、3 軸 6 方向の印象表現に直接対応する語彙群毎の発話の韻律制御特性の分析を行った。まず、図 3.1 に、読み上げ調音声と比較した、全ての対話音声の F0 平均値を示す。なお、同一印象表現内では F0 平均値が高いものから低いものの順に並べた。F0 平均値については、好印象（ - 悪印象）の印象表現に対応する語彙の発話で大きな影響が観測された。すなわち好印象の印象属性を有する語彙では、明らかに対話音声の方が F0 平均値は高くなっていた。また、確信、肯定に該当する語彙では、1 サンプルずつの例外はあるが、全体としてそれぞれ疑念、否定に対応する語彙よりも、対話音声の F0 平均値は高い傾向にあった。今回は、悪印象に該当する語彙の発話を得ることはできなかったが、上記の傾向により、悪印象のものは、好印象のものとは逆の低い F0 平均値を示すことが予想される。

次に、対話音声と読み上げ調音声との F0 パタンの違いを、図 3.2 に例示する。まず、

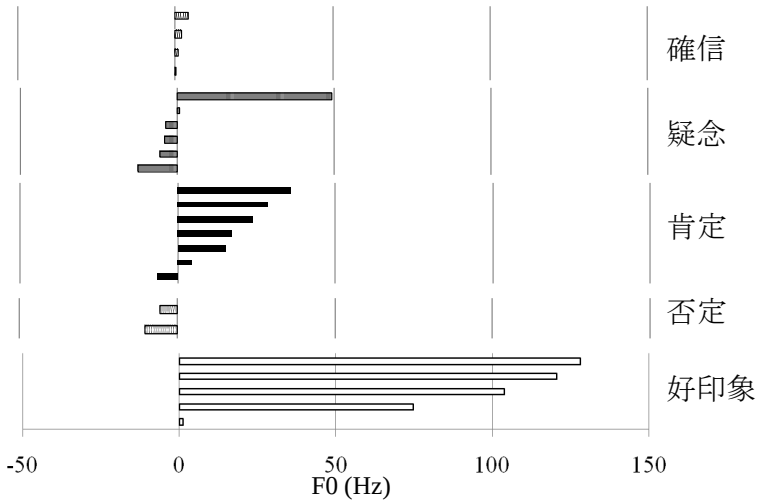


図3.1 読み上げ音声とのF0平均値の違い

パタンの比較を容易にするため、比較する 2 サンプルの発話時間長の、より短い方を伸長することで同じ長さに合わせた。図 3.2 にも見られるように、全体的に、確信の印象属性を有する語彙では、対話音声は下降の形状を示していた。また、疑念のうち、純粋な“疑問”に対応する語彙では上昇を、“疑念”を表現する語彙では緩やかな下降の形状を示すことが判明した。同様に、肯定に対応する語彙では下降、否定に対応する語彙では上昇＋下降の形状を示していた。なお、好印象の印象表現に対応する語彙では、形状の変化について一定の傾向は認められなかった。以上の結果は、2.5 節で一語発話「ん」の対話韻律の F0 パタンに観測された傾向と一致する。2.5 節で示した 3 次元の聴覚印象空間においては、確信 - 疑念の軸は、下降、上昇＋下降、緩やかな下降、上昇の順で、肯定 - 否定の軸は、下降、緩やかな下降、上昇、上昇＋下降の順で、F0 パタンと相関していた。つまり、確信は下降、疑念は上昇、または緩やかな下降、肯定は下降、否定は上昇+下降とそれぞれ対応していた。また、好印象 - 悪印象の軸は F0 パタンと明確な相関を持たなかった。

また、今回の対話発話分析によって、F0 制御特徴に加えて新たに発話時間長も、入力印象表現に対応する韻律制御特徴として挙げられることが判明した。図 3.3 に、読み上げ調音声と比較した、全ての対話音声の発話時間長を示す。各対話音声の時間長から対応する読み上げ調音声の時間長を差し引いたものである。なお、同一印象表現内では発話時間長が長いものから短いものの順に並べた。全体的に、確信と肯定の印象表現に該当する語彙の発話では、対話音声の発話時間長が読み上げ調音声のものより短くなる傾向があった。反対に、疑念と否定の印象表現に該当する語彙では、対話音声の発話の方が長かった。なお、好印象の印象表現に該当する語彙では、一定の傾向は見られなかった。5 発話の内、3 つは対話音声の方が長く、2 つは短かった。

以上のように、2.5 節で述べた一語発話「ん」の分析によって示された、確信 - 疑念、肯定 - 否定、好印象 - 悪印象の 3 次元の印象空間と F0 制御特性（平均値とパタン）との関係が、それら印象表現に直接対応する語彙についても成立していた。さらに、発話時間長も対話韻律制御に重要な役割を担っていることが判明した。

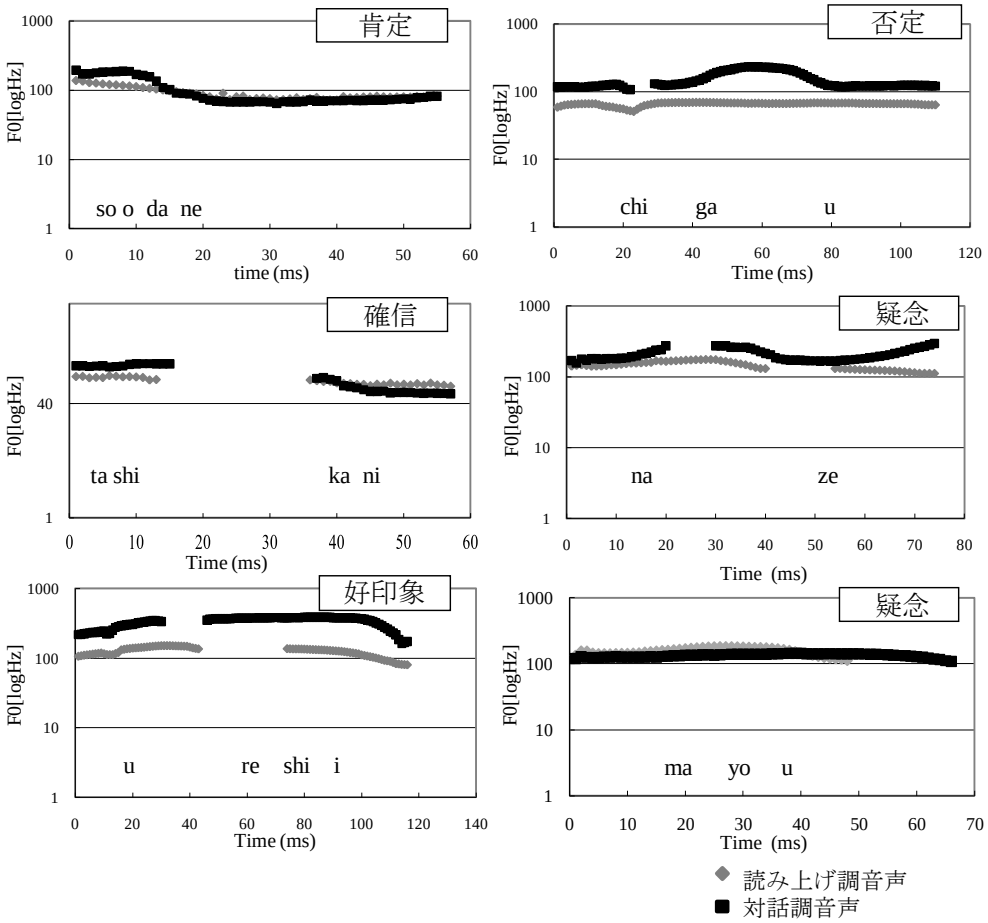


図3.2　対話音声と対応する読み上げ音声のF0パタンの比較

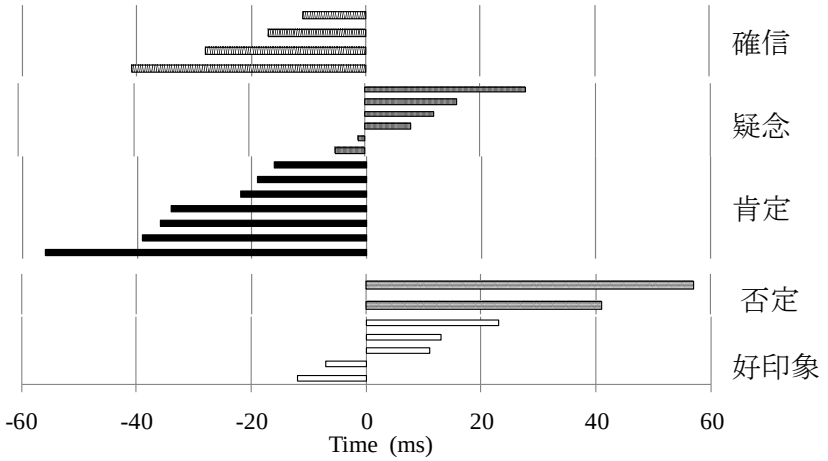


図3.3 読み上げ音声との発話時間長の違い

3.4　3 章のおわりに

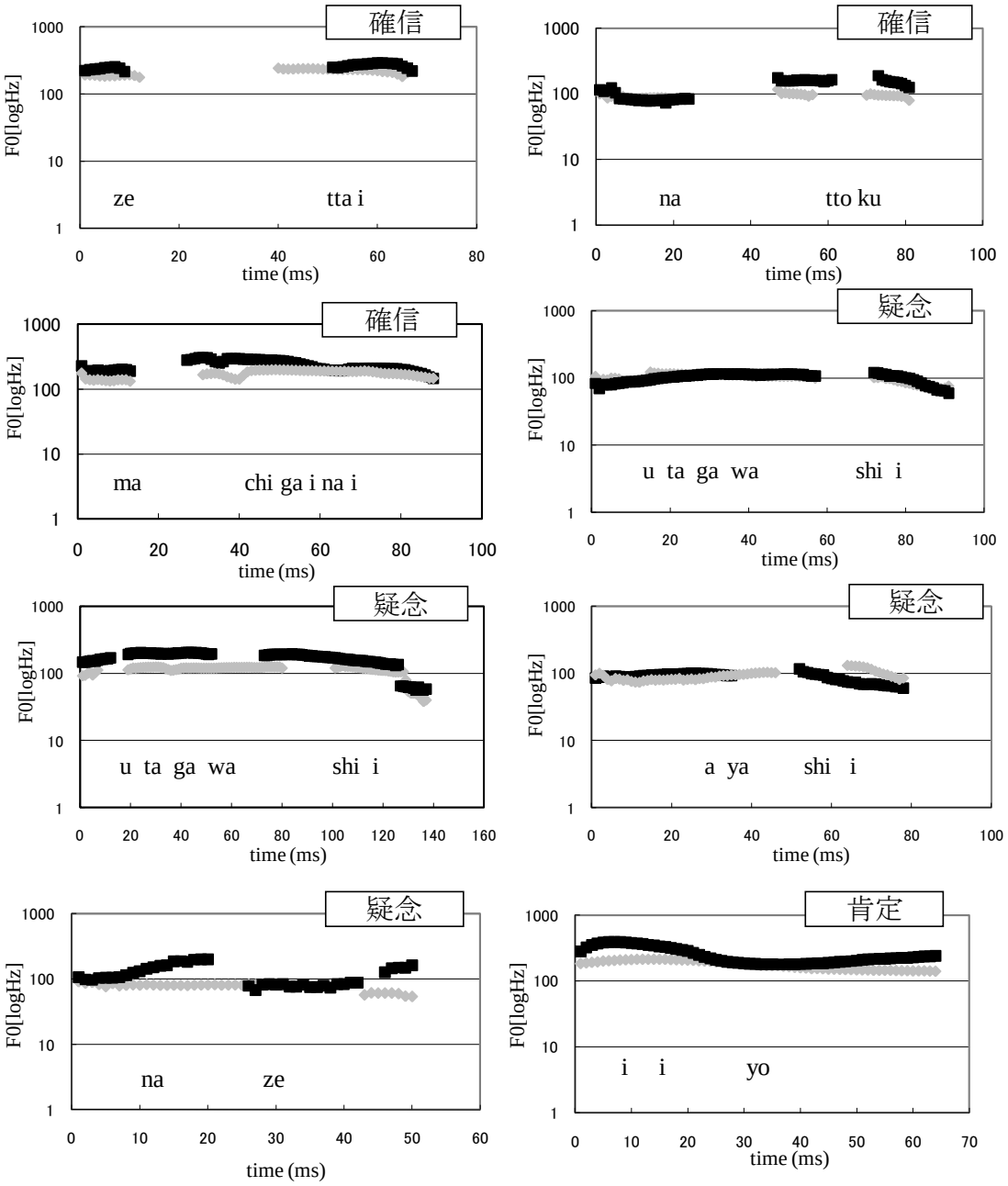
本章では、一語発話「ん」の対話韻律分析の結果が示した、印象 - 韻律の対応関係が、対話韻律生成の入出力情報として用いることができる可能性の検証を行った。このため、それら印象表現に直接対応する一般語彙の対話韻律の分析を行い、「ん」が示した印象 - 韻律の関係が、一般の語彙においても成立するかを調べた。

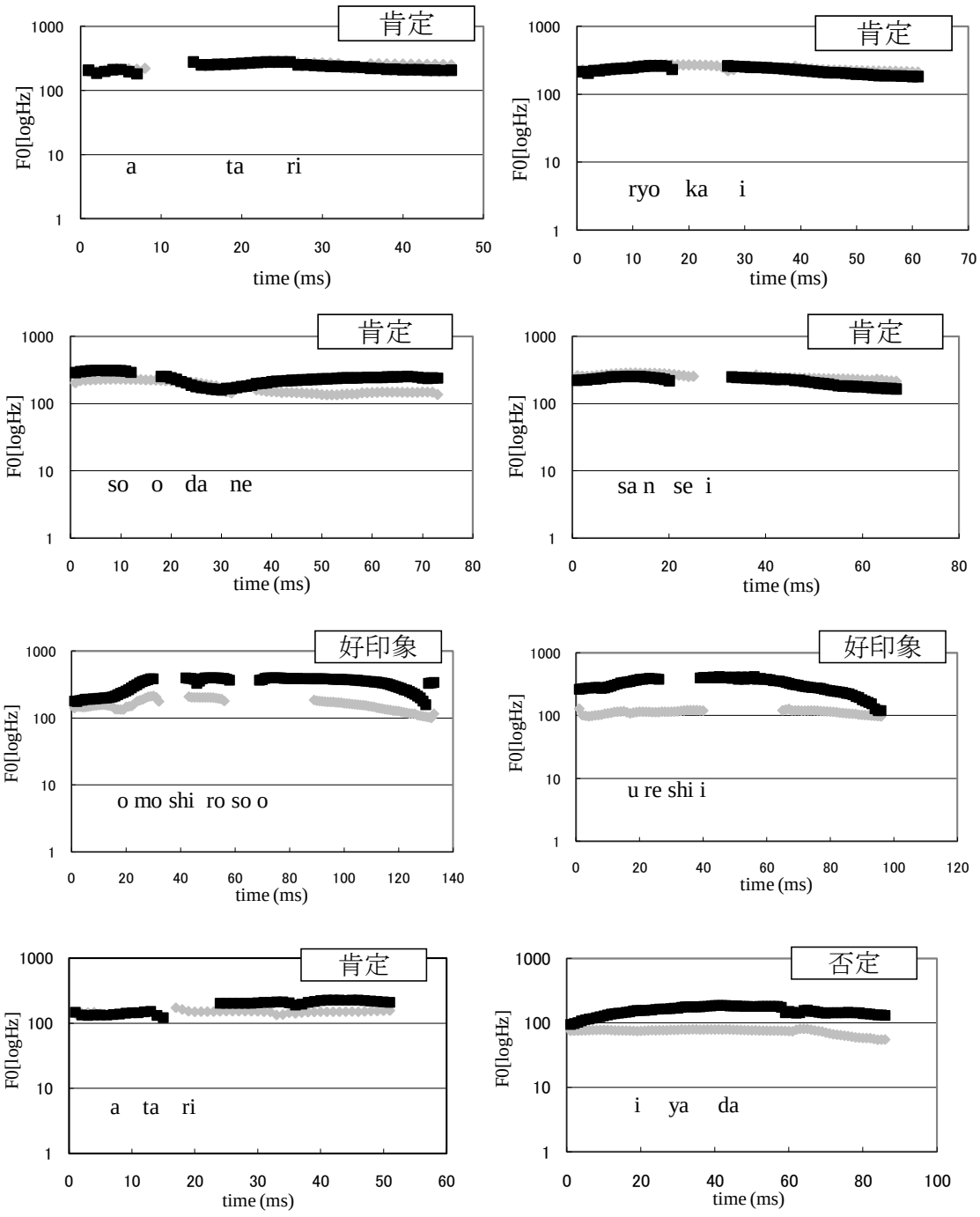
一語発話「ん」は印象属性を有する言語形式を持たないため、種々の韻律を取り得るが、一般の語においては、語彙の印象属性により、出力される対話韻律をある程度限定できることが考えられる。このため 3 軸 6 方向の典型印象表現（確信，疑念，肯定，否定，好印象，悪印象）を印象属性として有する語彙を選定し、それらの語彙発話の対話韻律分析を行った。その結果、一語発話「ん」が示した聴覚印象と対話韻律の対応関係との共通性が定性的に確認でき、さらには、F0 特徴に加えて、発話時間長も語彙の印象属性と関連することが判明した。すなわち、対話音声合成における出力韻律を規定する情報として、入力語彙の印象属性を利用出来る可能性が示唆された。

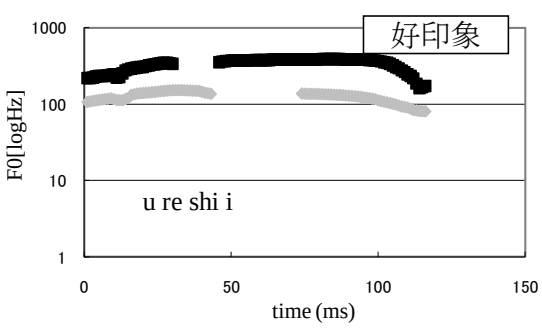
3 章の付録

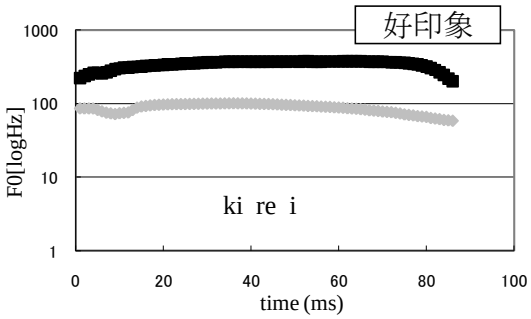
以下に、24 発話の F0 パタンのうち、代表パタンとして示した 6 つのパタン以外の、18 発話全ての F0 パタンを示す。

なお、それぞれ、■は対話調音声、◆は読み上げ調音声を示す。




肯定
F0[logHz]
1000
100
10
1
a ta ri
0 10 20 30 40 50
time (ms)
肯定
ryo ka i
0 10 20 30 40 50 60 70
time (ms)
肯定
so o da ne
0 10 20 30 40 50 60 70 80
time (ms)
肯定
sa n se i
0 20 40 60 80
time (ms)
好印象
o mo shi ro so o
0 20 40 60 80 100 120 140
time (ms)
好印象
u re shi i
0 20 40 60 80 100 120
time (ms)
肯定
a ta ri
0 10 20 30 40 50 60
time (ms)
否定
i ya da
0 20 40 60 80 100
time (ms)


好印象
1000
100
10
1
F0[logHz]
u re shi i
0
50
100
150
time (ms)


好印象
1000
100
10
1
F0[logHz]
ki re i
0
20
40
60
80
100
time (ms)

第 4 章　語彙の印象属性にもとづく対話韻律生成モデル

4.1　語彙の印象属性を用いた対話韻律の制御

対話場面に出現した発話の韻律分析により、一語発話「ん」の分析が示した、確信 - 疑念、肯定 - 否定、好印象 - 悪印象からなる 3 次元の印象空間と韻律制御特性（F0 平均値とパタン、発話時間長）の関係が、印象表現に直接対応する一般の語彙でも同様に成立することが定性的に確認できた。このことは、入力となる語彙の印象属性に応じた韻律制御の可能性を示唆する。すなわち、例えば、“好印象”の印象属性を備えた「きれい」という語彙に対しては、“好印象”が規定する制御特性（高い F0 平均値）を、その読み上げ音声に追加することによって、対話韻律を生成できることが考えられる。このため本章では、入力語彙の印象属性にもとづいた対話韻律生成方法の提案を行い、その妥当性の検証を行った。以下、4.2 節において、まず、提案する対話韻律生成方法の概要を述べる。提案方法では、従来の読み上げ韻律に、入力語彙の印象属性に応じた対話韻律特徴を追加することで対話韻律を生成する。韻律特徴の追加に際しては、そのままの特徴量に対してではなく、韻律を生成する過程でのパラメータのレベルでの生成を考えた。このため、4.3 節では、指令応答型の F0 制御モデルに関して述べ、対話韻律の生成における、その使用に関して述べる。4.4 節では、一語発話「ん」が示した聴覚印象による対話韻律制御特徴を参考に、入力語彙の印象属性に応じた対話韻律生成のための具体的な手順に関して述べる。さらに 4.5 節において、それら生成した対話韻律音声サンプルを用いた自然性評価実験について述べる。最後に 4.6 節でまとめる。

4.2　対話韻律生成モデル

前章までの分析で、一語発話「ん」が示した 3 次元の印象空間と対話韻律制御特性の関係が、一般の語彙においても成立することが定性的に確認できた。これにより、それら印象 - 韻律の対応関係を入出力情報として用いた、図 4.1 に示すような、入力語彙自体の印象に対応した対話韻律生成を考えることができる。図 4.1 に示すように、入力語彙は、従来の TTS で行われている読み上げ調の韻律だけでなく、対話韻律の特徴量の推定にも用いられる。あらかじめ各語彙に対して印象属性を付与した語彙辞書を用いて、入力された語彙（列）に対する印象情報を得る。この印象情報としては、入力語彙の、確信 - 疑念、肯定 - 否定、好印象 - 悪印象からなる印象表現空間における 3 次元座標のベクトルといったものを想定しているが、本検討では、各語彙に対応した印象表現そのものとなる。対話韻律付与モデルでは、この印象情報を用いて、対話韻律成分を作成する。この作成にあたっては、次節で述べるように、指令応答モデルのような韻律生成を念頭においた、生成のための制御指令レベルのパラメータ生成を考えている。これらは、従来の読み上げ韻律成分で生成された読み上げ韻律に加えられ、最終的な対話韻律が生成される。

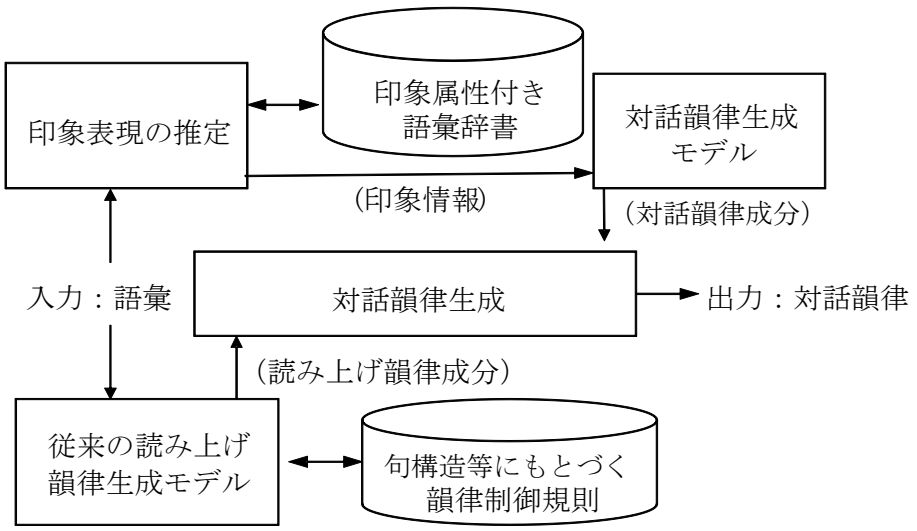


図4.1 語彙の印象属性を用いた対話韻律生成

本検討では、入力語彙の印象属性を表現するための空間と、対応する印象情報は、確信 - 疑念、肯定 - 否定、好印象 - 悪印象からなる 3 軸 6 方向そのものを、韻律変形としては F0 の平均値とパタン、及び、発話時間長の簡単な操作だけに限定している。以下では、これら各軸に対応する対話韻律生成の妥当性の確認を試みる。この確認ができれば、それらの重みつき組み合わせに対応する 3 次元印象空間内での任意のベクトルに対する対話韻律の生成の可能性も明らかとなる。また同様の方法論を用い、本検討で用いている印象表現以外の語彙属性への展開も期待できる。

4.3　指令応答型基本周波数制御モデルを用いた対話韻律生成

読み上げ韻律成分に対話韻律成分を追加するために、単純な加算処理が可能で、少数のパラメータで F0 パタン生成を記述できる、指令応答型の生成過程モデル[1]を採用した。生成過程モデルでは、F0 パタンは 2 種類の成分、すなわち、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の、二つの成分から構成される。フレーズ成分が I 個、アクセント指令が J 個ある場合の F0 パタンは以下のように表される。

$$\ln F_0(t) = \ln F_{min} + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj} \{G_a(t - T_{1j}) - G_a(t - T_{2j})\} \quad (1)$$

第 1 項の F_{min} は基本周波数の基底値、第 2 項の A_{pi} と T_{0i} は、それぞれ i 番目のフレーズ指令（インパルス）の大きさと生起時刻、第 3 項の A_{aj} と T_{1j}、T_{2j} は、それぞれ j 番目のアクセント指令（ステップ）の振幅と立ち上り時刻、立ち下がり時刻である。また、$G_p(t)$はフレーズ制御機構のインパルス応答関数、$G_a(t)$はアクセント制御機構のステップ応答関数であり、それぞれ次のように表される。

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (2)$$

$$G_a(t) = \begin{cases} \min[1 - (1 + \beta t) \exp(-\beta t), \gamma], & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (3)$$

ここで、α、β は、それぞれフレーズ制御機構とアクセント制御機構の固有角周波数、γ はアクセント成分が有限時間内に一定値に達することを保証する相対飽和値である。

4.4　対話韻律生成実験

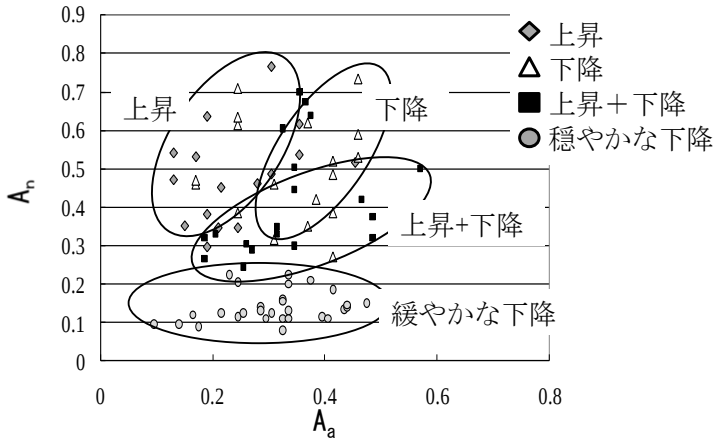
4.4.1　対話韻律生成実験のセットアップ

本論文で提案する対話韻律生成方法の妥当性を検証するために、読み上げ韻律成分で生成された読み上げ韻律をもとに、入力される語彙に応じた対話韻律の生成を行った。入力語彙が有する印象属性として 3 次元印象空間を用い、各軸の両端（確信、疑念、肯定、否定、好印象、悪印象）に位置すると近似的にみなせる単一のアクセント指令を持つ語彙を採用した。用いた語彙は 3.2 節で選定した、各印象表現に対応していると確認された語彙群の中から印象表現毎に 2 語彙を無作為に選んだ合計 12 語彙であり、それらを表 4.1 に示す。これらの代表的な語彙に対する対話韻律成分の追加が妥当であれば、より一般的なこれらの重みつき組み合わせとして規定される印象空間内のベクトルへの妥当性も推察できると考えた。

音声合成への適用を意図した場合、読み上げ韻律を作成する必要があるが、ここでは、合成システムが与える韻律ではなく、実際の読み上げ発話を分析した値そのものを使用している。これは、一般に音声合成では原音のパラメータ値に近いものを用いた場合に、変更するパラメータが与える主観的差異が最も明確に見られることが多いことによる。

表4.1 対話韻律生成に用いた語彙

印象表現	語彙
確信	絶対、確かに
疑念	疑わしい、迷う
肯定	賛成、当たり
否定	嫌だ、違う
好印象	きれい、面白そう
悪印象	汚い、退屈



注記：楕円は各分布の重心を中心とし分布の大部分を含む

図4.2　各F0パタンを持つ「ん」のA_aとA_p値

F0 パタンの生成にあたっては、 実際の読み上げ音声に対し、 種々の韻律事象記述に対する指令応答モデルによるこれまでの使用例を参考にした修正を行うこととした。図 4.2 に示すように、「ん」の印象表現の違いは F0 生成パラメータでは、フレーズ指令の大きさ A_p と、アクセント指令の大きさ A_a にみられることから、これら 2 つのパラメータの変形による可能性を考えた。「疑念」についても、これら 2 つのパラメータの変形だけとした。純粋な「疑問」では、語末の上昇が必須となるが、上昇を伴わない「疑念」を表現する語彙を対象とした。これにより、単純な上昇成分の追加による印象評定への直接的な関与は起こらず、他の印象と同様に、2 つのパラメータのバランスが与える影響だけを確認できると考えた。

提案した対話韻律生成方法における、表 4.1 に示した語彙が与える印象情報は、3 軸 6 方向（確信、疑念、肯定、否定、好印象、悪印象）のいずれか 1 つに対応する。それぞれの入力語彙に対応する読み上げ音声を収録し、それらの読み上げ韻律をもとに、対応する印象表現に応じて F0 平均値とパタン、発話時間長の生成を行なう。対話韻律成分の追加に際しては、まず、生成過程モデル[1]を用いて、読み上げ発話の韻律分析を行った。次に、音声分析変換合成法 STRAIGHT[41]を用いて、読み上げ発話の韻律制御パラメータに韻律生成パラメータ値を追加した。次に、その韻律生成パラメータ値の設定に関して述べる。

4.4.2　一語発話「ん」の韻律特性を用いた対話韻律制御

実際の韻律生成に際して、読み上げ韻律成分の読み上げ調の韻律に対して追加する、対話韻律成分で作成される、入力語彙に応じた対話韻律生成パラメータの値を設定した。値は、実際の対話音声に見られる韻律の変動幅を反映するため、2.3 節で述べた JST/CREST ESP プロジェクトによって収録された膨大な量の日常会話発話データから抽出した一語発話「ん」の分析をもとに決定した。2.3 節で用いた一語発話「ん」6,271 サンプルのうち、各クラスタの中心に近かった F0 のパタンに対応する音声サンプルを、それぞれの F0 パタンカテゴリの代表として選び、生成パラメータの設定に用いた。サンプル数は、上昇は 16、緩やかな下降は 30、下降は 18、上昇＋下降は 21 であった。さらに、各パタンカテゴリの特徴量を抽出するために、読み上げ調の「ん」を参照サンプルとして用意した。読み上げ調の「ん」は第一著者によって発話された。F0 のパタンが平坦になるよう、また F0 の高さと発話時間長ができるだけ中立的になるよう注意して発話された。

各生成パラメータ値の算出方法を以下に示す。A_a と A_p は、F0 パタンに反映されるので、4 つの F0 パタンカテゴリ毎に韻律生成パラメータ値を算出した。まず、生成過程モデル[1]を用いて、全てのサンプルから A_a と A_p の値を取り出した。次に、4 つのカテゴリ（上昇、緩やかな下降、下降、上昇＋下降）毎に平均値を算出した。それらの値を読み上げ調参照サンプル発話から取り出した値によって除算し、4 つずつの A_a、A_p の韻律生成パラメータ値を得た。F_{min} 値は F0 平均値に反映される。単純化のため高めと低めの 2 つの値を算出した。上記発話サンプルから F_{min} 値の最も高いあるいは低いもの各 10 サンプルを取り出し、それぞれの平均値から読み上げ調参照サンプル発話の F_{min} 値を減算することで、2 つの値を得た。発話時間長についても同様に、最も長いあるいは短いもの各 10 サンプルを取り出し、それぞれの平均値を読み上げ調参照サンプル発話の時間長で除算することで、長めと短めの 2 つの値を得た。

表4.2　対話韻律生成に用いた韻律生成パラメータ値

	確信	疑念	肯定	否定	好印象	悪印象
F_{min}	30	-35	30	-35	30	-35
A_p	1.99	1.81	1.99	2.08	1	1
A_a	2.26	0.60	2.26	1.86	1	1
発話時間長	0.75	1.3	0.75	1.3	1	1

注記 :表示した値を それぞれの対応する読み上げ音声のF_{min}(Hz)に加算, $A_{a,}$, A_p, 発話時間長に乗算することで 対話音声に変換する

対話韻律生成を行うのに、韻律生成パラメータ値を 3 軸 6 方向の印象表現のそれぞれに対して配分した。そのために、3.3 節で述べた各印象表現に対応する語彙と、その対話韻律制御特徴との関係を用いた。まず、F0 パタンに関しては、確信の印象表現に該当する語彙の対話音声は「下降」、疑念は「上昇」または「緩やかな下降」、肯定は「下降」、否定は「上昇＋下降」を、それぞれ示していた。従って、「下降」の A_a と A_p の値を確信と肯定に、「緩やかな下降」の値を疑念に、「上昇＋下降」の値を否定に、それぞれ適用した。次に、F0 平均値に関しては、好印象、確信、肯定の印象表現に該当する語彙の対話音声は高く、反対に、悪印象、疑念、否定の印象表現に該当する語彙のものは低かった。この傾向に従い、好印象、確信、肯定には高い F_{min} 値を、悪印象、疑念、否定には低い値を適用した。最後に、発話時間長は、疑念と否定の印象表現に該当する語彙の対話音声では長く、確信と肯定の印象表現に該当する語彙のものでは短かった。従って、疑念と否定には長めの発話時間長を、確信と肯定には短めのものを適用した。以上の手続きで設定された、読み上げ韻律に追加するための、対話韻律生成パラメータ値を表 4.2 に示す。

4.5　対話合成音声の自然性評価実験

提案した方法の妥当性を検証するために、前節で得た韻律制御パラメータをもとに、対話韻律の生成を行い、自然性評価実験を行った。評価実験は、入力となる印象表現にもとづいて設定された韻律制御パラメータの妥当性と、入力語彙が与える印象表現が規定する対話韻律の妥当性を検証する、2 種類の実験であった。

4.5.1　一語発話「ん」の対話韻律制御特性の効果

一語発話「ん」の韻律特徴をもとに設定した、入力語彙が対応する印象表現にもとづく韻律生成パラメータ値が、自然性の向上に効果的であるかどうかを検証するため、自然性評価実験を実施した。このため、4.4.1 節で用意した表 4.1 の、3 軸 6 方向の各印象表現に対応する語彙を用いて、それぞれの印象表現毎に、関係する全てのパラメータ値を追加したサンプルを用意した。具体的には、確信、疑念、肯定、否定の 4 印象表現に該当する 8 語彙に関しては、読み上げ調韻律の他に、及び、F_{min} 値、A_a と A_p の値、発話時間長、の全てのパラメータ値を追加したサンプルを用意した。一方、好印象、悪印の印象表現に該当する 4 語彙に関しては、A_a、A_p、発話時間長の追加による変化は生じないので、読み上げ調音声の他に、F_{min} 値のみを修正した。韻律生成パラメータ値の追加に際しては、表 4.2 に示した値を、それぞれの読み上げ調韻律の F_{min} 値には加算、A_a と A_p の値には乗算した。また、発話時間長は、表 4.2 の値を乗算することにより、韻律制御パラメータ値を追加した。作成された対話韻律の F0 パタンの例を、もととなる読み上げ音声の韻律と合わせて図 4.3 に示す。また各図の下に、それぞれの読み上げ音声と修正を施した音声の A_a と A_p の値も合わせて示す。合計で 24 音声（＝12 語彙×2 韻律）を用意した。

自然評価実験では、24 の合成音声サンプルを 1 つずつランダムな順番で提示し、誰かの発言への返答としてどれくらい自然であるかを、0（不自然である）～7（自然であ

る）の 8 段階で評定させた。評定者は、東京方言／標準語を話す 24 歳～38 歳までの日本語母語話者 5 名（男性 2 名、女性 3 名）であった。

図 4.4 に、印象表現グループ毎に、読み上げ調音声サンプルと対話韻律生成パラメータ値を施した音声サンプルそれぞれに対する、全評定者の平均得点を示す。図 4.4 に示すように、全印象表現において、対話韻律音声が、同一発話内容の読み上げ調音声と比較して、対話音声としての自然性が向上していると評価されていた。また、韻律生成パラメータ値の追加による音声サンプルの自然性向上の効果を検定するため、1 元配置の分散分析を行った。その結果、全ての印象表現グループにおいて、それぞれ 5%の水準で有意差がみられた。この結果は、前節で入力語彙が与える印象をもとに設定した韻律生成パラメータ値が有効であったことを示す。これにより、提案方法における、一語発話「ん」の韻律制御特性を参考に設定した、韻律生成パラメータ値の選択の妥当性が確認できた。

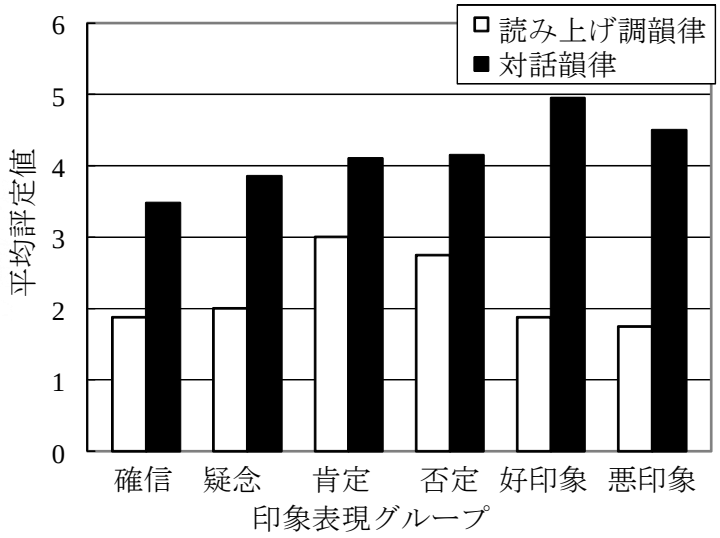


図4.4　韻律制御パラメータの追加修正による
音声サンプルの自然性の向上

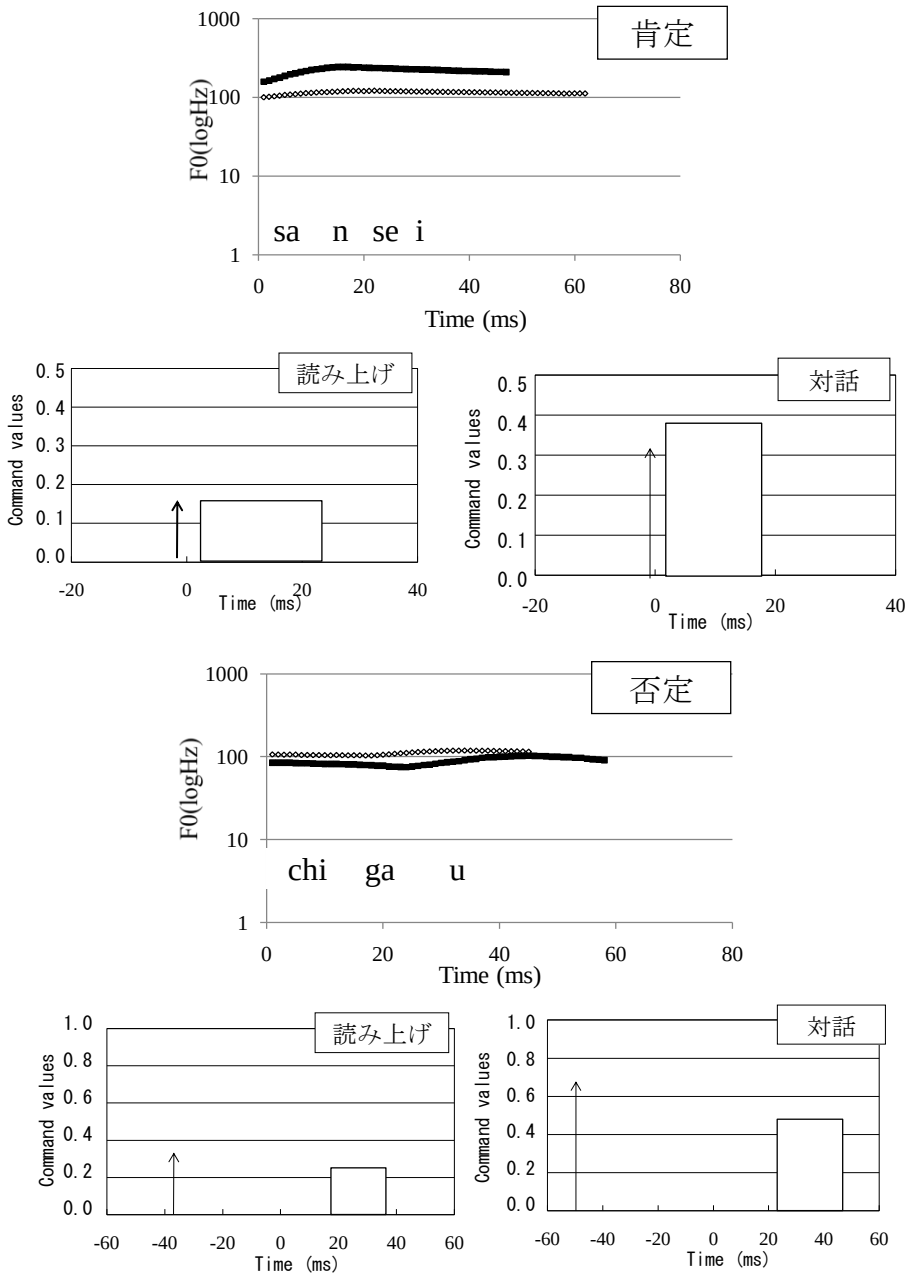


図4.3（1/2）　韻律生成パラメータの追加によるF0パタンの比較

（F0パタンの下に生成過程モデルのパラメータを図示する。上向き矢印はフレーズ指令の大きさA_pと生起時刻T_0を，矩形はアクセント指令の大きさA_aと立ち上がり時刻T_1、立ち下がり時刻T_2をそれぞれ示す。）

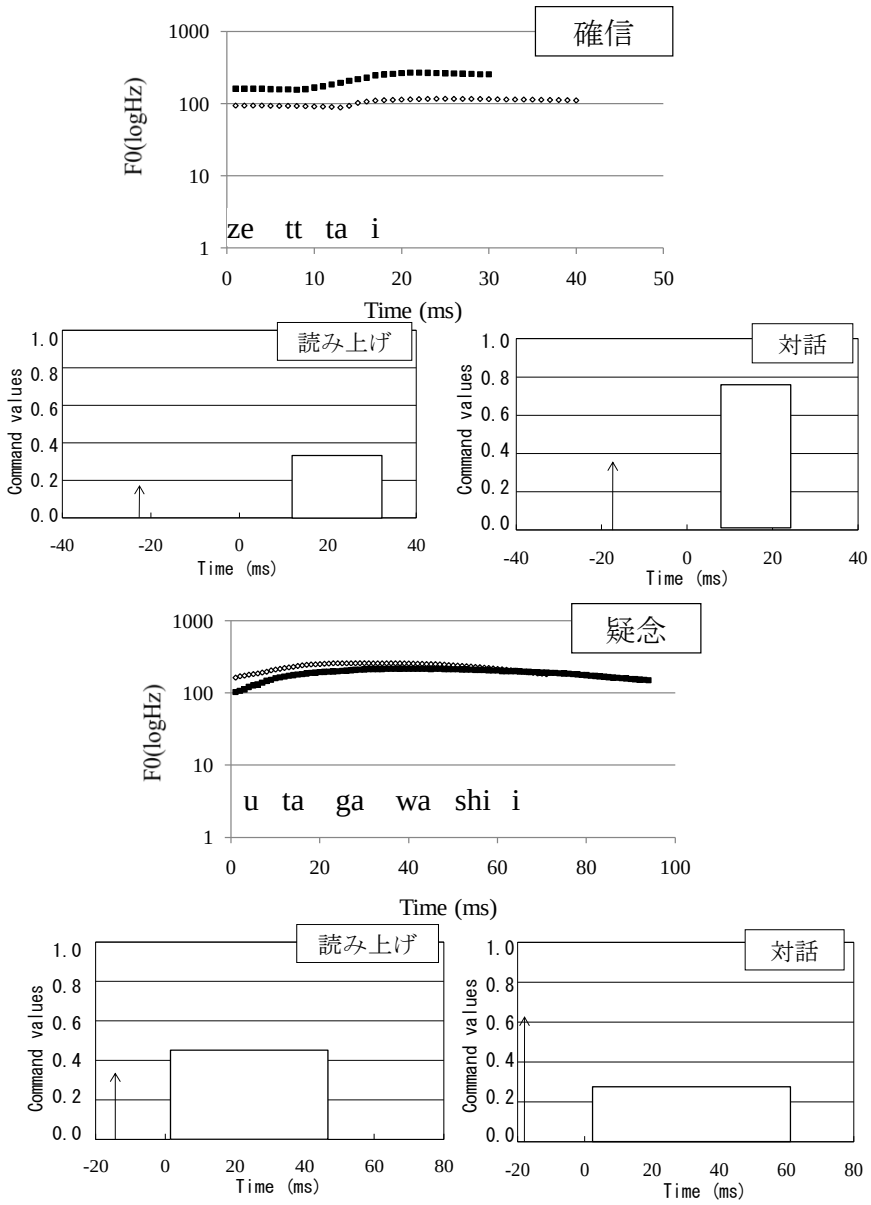


図4.3（2/2）

（タイトルは図4.3(1/2)を参照）

4.5.2　入力語彙の印象属性によって規定される対話韻律の妥当性

入力語彙の印象属性にもとづいた対話韻律音声が妥当であるか否かを検証する自然性評価実験を行った。表 4.1 の 3 軸 6 方向の各印象表現に対応する語彙に対して、入力語彙の印象属性に一致した韻律生成パラメータ、または一致しない韻律生成パラメータを施し、その自然性の評価を求めた。実験には、前節と同様の手順で、12 個の全部の語彙に対して、表 4.2 に示す 6 種全部のパラメータ値を施した音声、合計 72 音声（＝12 語彙×6 韻律）を用意した。

入力語彙の印象属性にもとづく対話韻律生成の妥当性を検証するため、それら合成音声サンプルを用いた自然性の検証を行った。前節の自然性評価実験と同様に、誰かの発言への返答としてどれくらい自然であるかを、0（不自然である）～7（自然である）の 8 段階で評定させた。評定者は、前節とは異なる、東京方言／標準語を話す 25 歳～36 歳までの日本語母語話者 5 名（男性 1 名、女性 4 名）であった。

なお、合成に用いた対話韻律の入力語彙への一致度の違いに着目した比較を行うため、用いた韻律の種類を次の 4 つのグループに分類した。各グループはそれぞれ、“読み上げ調韻律”、“一致グループ”、つまり、入力語彙が対応する印象表現と一致する韻律（例：入力語彙の印象表現が肯定で、追加対話韻律パラメータも肯定）、“同一次元逆グループ”、つまり、入力語彙が対応する印象表現と一致しない韻律のうち同一の次元で逆向き（印象空間内で 180 度）の印象表現の韻律（例：入力語彙の印象表現が肯定で、追加対話韻律パラメータが否定）、“中立次元グループ”、つまり、入力語彙が対応する印象表現と一致しない韻律のうち中立の次元（印象空間内で 90 度）の印象表現の韻律（例：入力語彙の印象表現が肯定で、追加対話韻律パラメータが、確信、疑念、好印象、悪印象）であった。前半の 2 グループは、前節の、読み上げ調、及び対話韻律サンプルに等しい。

図 4.5 に、韻律グループ毎の全評定者の平均得点を示す。韻律グループによる平均得点の違いを検定するため、1 元配置の分散分析を行った。その結果、1 %の水準で韻律グループの効果が有意であった。さらに、4 つのグループの評定値に対して、Tukey 法

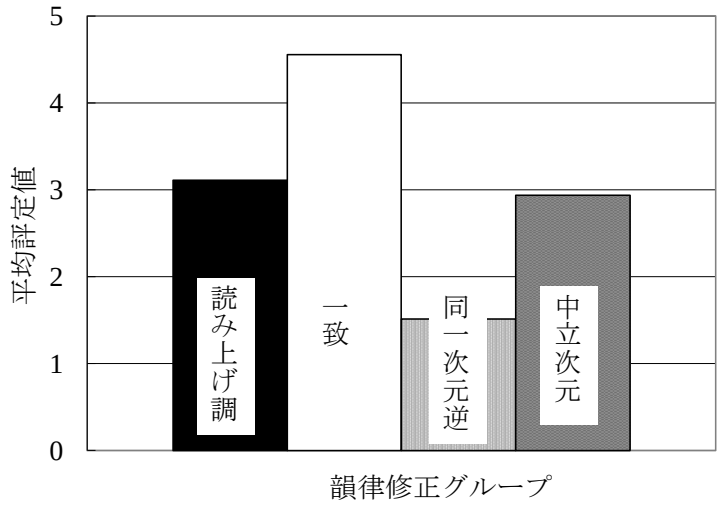


図4.5　入力語彙の印象属性に対応した韻律特性の
追加による音声サンプルの自然性の比較

にて多重比較を行った結果、読み上げ調と中立次元グループの間を除く全てのグループ間において、それぞれ 5%の水準で有意差が見られた。上記の結果は、入力語彙と整合した韻律追加のみが対話音声としての自然性向上に貢献すること、入力語彙と逆方向の印象に対応する韻律追加は自然性を低下させ逆効果であることを示す。これにより、語彙の与える印象表現にもとづく対話韻律の生成という、提案方法の妥当性を確認できた。

4.6　4 章のおわりに

本章では、前章までの分析結果にもとづき、対話場面に有効な TTS システムのための、対話韻律生成方法の提案を行った。発話内容に即して、語句のレベルで動的に対話韻律を制御する仕組みとして、入力語彙自体が有する印象属性を利用した韻律制御方法を考えた。

提案方法では、従来の読み上げ韻律に加え、新たに、語彙が有する印象属性にもとづいた制御を加えた対話韻律生成を行う。3 軸 6 方向の典型印象（確信，疑念，肯定，否定，好印象，悪印象）を印象属性として有する語彙の読み上げ音声に対して、各々が対応する聴覚印象が制御する一語発話「ん」の F0 平均値、F0 パタン、発話時間長の対話韻律特徴を追加することにより合成音声を作成した。F0 生成に際しては、韻律生成過

程を反映する生成過程モデルを用い、生成パラメータのレベルで対話韻律成分の追加を行った。これら合成した対話韻律音声に対して自然性評価実験を行い、提案した対話韻律生成方法の妥当性を確認することができた。

第 5 章　対話韻律生成モデルの他言語への適用

5.1　言語に共通する対話韻律制御情報としての語彙印象属性

提案した入力語彙の印象属性にもとづく対話韻律生成方法が、より幅広い入力に対応出来る可能性を検証した。提案方法で用いる、入力語彙が有する印象属性のレベルでは、言語依存性が低いことが考えられる。また特定の意味を有さない一語発話「ん」を用いた出力対話韻律の規定などとも考え合わせると、印象を用いた対話韻律生成のフレームワークの、他言語への適用の可能性が考えられる。

このため本研究では、提案した対話韻律生成方法が、日本語以外の言語への適用の可能性の検証のため、音声合成で広く用いられる英語を対象に、前章における、日本語の語彙を対象に行った対話韻律生成実験と同様の手続きで、対話音声の合成を試みた。以下、5.2 節では、3 軸 6 方向（確信、疑念、肯定、否定、好印象、悪印象）の典型印象表現に対応する語彙の選定、5.3 節では、それら選定した英語の語彙を入力とする、対話韻律生成実験の詳細を述べる。さらに 5.4 節において、生成した音声サンプルを用いて行った自然性評価実験に関して述べ、最後に、5.5 節でまとめる。

5.2　英語対話音声合成の試み

日本語の語彙を入力として検証した、印象 - 韻律の対応関係を用いた対話韻律生成方法の妥当性を、さらに確認し、また言語に依存しない生成方法であることを示すために、英語を入力とした対話韻律の生成を試みた。入力には、前章で行った日本語のサンプルと同様、3 軸 6 方向（確信、疑念、肯定、否定、好印象、悪印象）の典型印象表現に対応する、英語の語彙を選定した。分析には、表 5.1 の英語語彙欄に示した 18 の英語の語彙を用いた。

表5.1 対話音声収録で用いた英語語彙と意図したそれらの印象の一致

印象表現	英語語彙（和訳）	平均評定値		被験者間相関の範囲
		当該印象表現	その他の印象表現	
確信 (Confident)	Sure（確かに), Certainly（その通り), Of course（もちろん）	4.67	0.17	0.58 – 0.87
疑念 (Doubtful)	Shady（疑わしい）, Why（なぜ）, Suspicious（怪しい）	5.82	0.55	0.61 – 0.92
肯定 (Allowable)	Agree（賛成）, Okay（了解）, Bingo（当たり）	6.25	0.27	0.71 – 0.98
否定 (Unacceptable)	No（ううん）, Incorrect（間違い）, Wrong（違う）	6.5	0.90	0.74 – 0.91
好印象 (Positive)	Beautiful（きれい）, Glad（嬉しい）, Interesting（おもしろい）	4	0.06	0.72 – 0.89
悪印象 (Negative)	Dirty（汚い）, Sad（悲しい）, Bored（退屈）	3.4	0.14	0.88 – 0.96

実験に先立ち、選出した各語彙が、3 軸 6 方向の印象表現を確実に表現できているかどうかを確認するために、各語彙に対する主観評価実験を行った。選定した 18 の英語の語彙に対して、3.2 節で述べた 16 の印象表現語に準じたものを用い、各表現語に 0（全く当てはまらない）～6（非常に良く当てはまっている）の 7 段階評定を求めた。実験には、31 歳～39 歳のアメリカ英語を母語とする、成人 3 名（男性 2 名、女性 1 名）及び、日本語を母語としているが、英語圏に 4 年以上住んだ経験があり、英語が流暢である 39 歳および 42 歳の女性 2 名が参加した。なお、16 の印象表現とは、それぞれ確信－疑念(understanding、 approve、 doubt、 ambivalence) 肯定－否定(agreement、 sympathy、 deny、 objection) 好印象－悪印象(bright、 happy、 interested、 light、 dark、 sad、 not interested、 heavy)であった。

表 5.1 に各印象表現に該当する印象表現語の平均評定値、評定者間相関値の分布範囲を示す。平均評定値のうち「当該印象表現」は、選定した語彙が意図した印象表現に対応する印象表現語群に対して得た評定値の平均である。具体的には、例えば、“確信”の印象表現を意図して選択した、”Of course”、 “Certainly”、 “Sure”の 3 つの語彙がそれ

ぞれ、“確信”の印象表現に対応する印象表現語の“understanding”、“approve”に対して得た評定値の平均値を指す。また意図しないそれ以外の印象表現語に対する平均評定値を、比較として、「その他の印象表現」の欄に示す。表 5.1 に示すように、選定した語彙の、意図した印象表現語に対する平均評定値は 4 以上となり、意図しない他の基本印象語の平均評定値 0.9 以下と比べて高い値となっている。また評定者間相関値は、いずれも 0.58 以上を示していた。これにより、選定した語彙は意図した印象表現を反映した妥当なものであることが確認できた。

5.3　英語対話音声合成実験

英語の語彙が有する印象属性を用いた対話韻律生成の妥当性を検証するために、選定した語彙を用いて、対話韻律生成実験を行った。選定した語彙は、まず 28～39 歳のアメリカ英語母語話者 4 名（男性 2 名、女性 2 名）によって、読み上げ調で発話され、静かな環境下で録音された。提案した対話韻律生成方法における、表 5.1 に示した語彙が与える印象情報は、3 軸 6 方向（確信、疑念、肯定、否定、好印象、悪印象）のいずれか 1 つに対応する。それぞれの入力語彙に対応する読み上げ音声を収録し、それらに入力語彙の印象属性に応じた、F0 平均値とパタン、発話時間長の対話韻律成分の追加を行った。実際の韻律生成パラメータ値の追加に際しては、前章での日本語を対象とした対話韻律生成実験での手順と同様、まず生成過程モデル[1]を用いて、読み上げ発話の韻律分析を行った。次に音声分析変換合成法 STRAIGHT[41]を用いて、読み上げ発話の韻律パラメータに、韻律生成パラメータ値を追加した。

韻律生成パラメータ値に関しては、4.5 節で一語発話「ん」をもとに設定した表 4.2 のパラメータ値を用いた。2 軸 4 方向（疑念、確信、否定、肯定）の印象表現に対応する 12 の語彙に対しては、F_{min}、A_a / A_p、発話時間長の全てのパラメータ値の追加を行ったものを用意した。一方、1 軸 2 方向（悪印象/好印象）に対応する 6 つの語彙に関しては、F_{min} のみの追加を行った。最終的に読み上げ発話を含めた、合計 144 の発話サンプルを用意した。

5.4　英語対話合成音声の自然性評価実験

提案した入力語彙の印象属性を用いた対話韻律生成のフレームワークが、他言語に適用できる可能性を調べるために、提案方法にもとづいて生成した英語の対話音声を用いて、2 種類の自然性評価実験を行った。5.4.1 節では、入力語彙の印象属性に対応した対話韻律特徴を読み上げ韻律に追加することにより生成した、対話音声の自然性の向上に関して、5.4.2 節では入力語彙が対応する印象表現と出力される対話韻律特徴との直接的対応関係を調べるための実験についての詳細をそれぞれ述べる。

5.4.1　入力語彙の印象属性によって規定される対話韻律の妥当性

入力語彙が有する印象属性に対応した対話韻律特徴を追加することによって生成した英語の対話韻律サンプルの、対話音声としての自然性が向上しているかどうかを調べるための、自然性評価実験を行った。前節で用意した音声サンプルを、話者毎に 1 つずつランダムな順番で提示し、日常会話に出現する発話として、どれくらい自然であるかを、0（不自然である）～7（自然である）の 8 段階評定で求めた。実験には、31～37 歳のアメリカ英語を母語とする成人男性 3 名と、37～42 歳の英語圏に 4 年以上住んだ経験のある、英語に流暢な成人女性日本語母語話者 3 名が参加した。評定者には、1 話者毎に休憩を取るなど、無理のないリラックスした状態で評定を行うようお願いした。一人当たりの実験実施時間は、20 分～30 分程度であった。

3 軸 6 方向（確信、疑念、肯定、否定、好印象、悪印象）の典型印象表現に対応する語彙に対する全被験者の平均評定値を、図 5.1 に示す。図に示すように、全ての印象評定値に対応する語彙において、読み上げの発話よりも、対話韻律の追加を行った発話サンプルの方が自然であると評価された。また 1 元配置の分散分析の結果、全体でその差は 5%で有意であった。すなわち、語彙が与える印象を用いた対話韻律生成方法の妥当性は、日本語のみでなく、英語においても有効であることが示された。

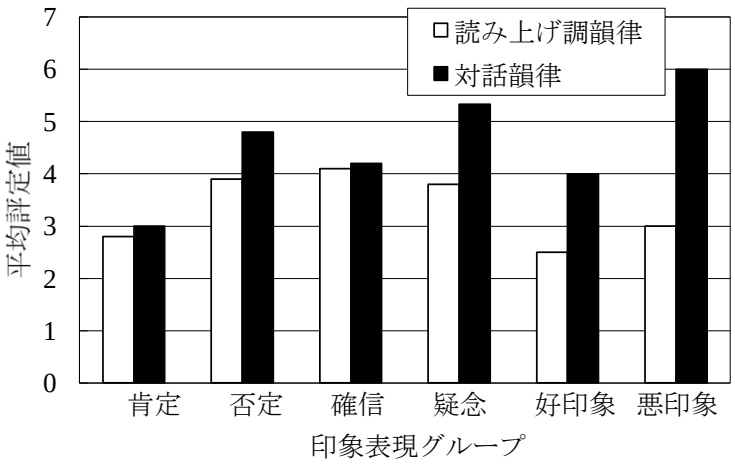


図5.1　韻律生成パラメータ値の追加による
英語音声サンプルの自然性の向上

5.4.2　語彙とその語彙発話の対話韻律が与える印象の直接的関係

入力語彙の印象属性によって規定される対話韻律の妥当性を確認するために、入力語彙が対応する印象表現と、対応する対話韻律が与える印象との直接的関係を調べた。前節で用意した、語彙印象属性を用いて生成した対話音声サンプルがそれぞれ、5.2 で用意した 16 の印象表現語に、0（全く当てはまらない）～7（非常に良く当てはまっている）の 8 段階評定を求めた。被験者は、5.4.1 節と同様の、アメリカ英語母語話者の成人男性 3 名と、英語の流暢な日本語を母語とする成人女性 3 名とした。音声サンプルと、評定項目の多さから、被験者への負担を考慮して、1 話者分の 36 音声サンプルを 1 セットとして提示し、無理のない範囲内で、実験を実施するようお願いした。その結果、一人当たりの評定音声サンプルは、2～3 セットであり、合計 504 サンプルに対しての評定を行なった。

入力語彙印象と、出力される対話韻律特徴の直接的な関係をみるために、5.2 節で得た入力語彙そのものに対する主観印象評定値と、対話音声サンプルの印象評定値を比較した。図 5.2 に、語彙が対応する印象表現カテゴリ（確信、疑念、肯定、否定、好印象、悪印象）毎に得られた、入力語彙と対話韻律音声サンプルの印象評定値間の平均相関値を示す。具体的には、例えば、“確信”のカテゴリ内には、“Of course”、“Certainly”、“Sure”の語彙そのものに対して、またそれらの対話発話について、5.2 節で用いたのと同様の 16 の印象表現語にもとづいて得た平均印象評定値の相関が含まれている。図 5.2 に示すように、全サンプルにおいて、非常に高い相関値（0.86 以上）が得られた。すなわち、入力語彙自体が与える印象と、対話音声が与える聴覚印象に共通性が認められた。従って、入力語彙が有する印象属性にもとづいた対話韻律生成方法の妥当性が、英語のサンプルについても確認することができた。

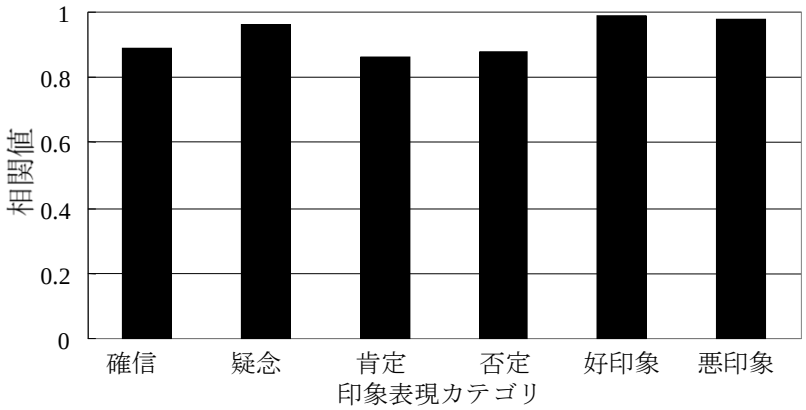


図5.2　語彙印象と対話韻律が与える印象の
直接的対応関係

5.5　5 章のおわりに

提案した対話韻律生成方法で扱う印象属性のレベルでは、言語依存性が低く、言語を問わず用いることができる可能性が考えられる。このため日本語で検証した提案方法が他言語にも適用できる可能性の検証を目的として、英語を対象とした対話音声の合成を行った。

合成には、日本語の韻律生成実験に用いた 3 次元 6 方向の典型印象表現（好印象、悪印象、疑念、確信、否定、肯定）を印象属性として有する英語語彙を用い、各々が対応する聴覚印象を持つ一語発話「ん」の対話韻律の追加を行った。聴取実験の結果、それら英語の合成音声サンプルの、対話韻律としての自然性の向上を確認した。すなわち、入力語彙の印象属性を用いた対話韻律生成方法が、他言語へも適応できる可能性を示すことができた。

第 6 章　複数語彙の印象属性と対話韻律の分析

6.1　一般の語彙列に対する対話韻律生成に向けた検討事項

日本語と英語の複数言語サンプルを用いた検証により、入力語彙の印象属性にもとづいた対話韻律制御の可能性を、単独語彙の入力に対して明らかにした。さらに提案した韻律生成方法の、より一般的な複数の語彙の組み合わせによって構成される入力への展開の可能性を検証した。

提案した語彙の印象属性にもとづく対話韻律制御が可能であるかどうかは、複数の語彙が関与した場合の制御特性が、単独語彙の制御の組み合わせで対処できるかどうかがキーポイントとなる。このため、種々の印象属性を持つ複数の語彙からなる対話音声の韻律を分析し、各語彙が全体の対話韻律へ与える影響を調べた。以下、6.1 節において、まず対話韻律の分析に用いる、3 軸 6 方向（好印象、悪印象、疑念、確信、否定、肯定）の典型印象に対応する語彙の組み合わせにより構成される入力のデザイン、及び、それら対話発話の収録、対話韻律の分析方法に関して述べる。続く 6.2 節では、入力を構成する各語彙の印象属性の効果が、それぞれどのように出力対話韻律に関与しているかという観点からの対話韻律分析結果を述べる。最後に 6.3 節でまとめる。

6.2　異なる印象属性を有する複数の語からなる文の対話韻律分析

提案した韻律生成方法をより一般的な入力への適用の可能性を検証するために、複数の語彙の異なる印象属性によって規定される韻律制御特徴に関する検討を行った。このため、異なる印象属性を有する複数の語彙から構成される文の発話の対話韻律の分析を行い、各語彙の印象属性が、その発話の出力対話韻律に与える影響を検証した。分析対象となる文は、3 軸 6 方向（好印象、悪印象、疑念、確信、否定、肯定）の典型印象表現のうち、異なる印象属性を備えた 2 つの語彙によって構成される、日常よく使用される日本語の文を用いた。表 6.1 に示すように、好印象/悪印象を表す 4 セットの形容詞（表 6.1(a)）と、疑念/確信、否定/肯定を表すそれぞれ 4 セットの終助詞（表 6.1(b)）から構

成した。日本語で終助詞は、様態を表現するのに用いられ、日常会話においても多用される。様々な終助詞が存在し、表に示すように、それらの印象を段階的に表現する終助詞が存在する。

語彙の印象属性による対話韻律の変化を、定量的に分析するために、表 6.1(c)に示す程度副詞を採用した。好印象/悪印象を表す形容詞は、それらの程度副詞を用いることにより、その度合いを表現した。確信 - 疑念、肯定 - 否定を表す終助詞を含む全ての文に対して同一の副詞を用いることが理想的ではあるが、否定の終助詞を伴う文に際しては、不自然な日本語を避けるため、ふさわしい相当の終助詞を用意した。これらの語の組み合わせにより、結果 256 の異なる文を用意した。

デザインした文の発話の対話韻律を分析するために、まず自然発話サンプルの収録を

表6.1　各語彙の全体の出力対話韻律への関与
を測定するために用いた入力を構成する語彙

(a)好印象 - 悪印象の印象表現に対応する形容詞

好印象	悪印象
きれい	汚い
うまい	まずい
優しい	厳しい
おもしろい	つまらない

(b)肯定 - 否定／確信 - 疑念の印象表現に対応する終助詞

度合い	
確信	だ
やや確信	よね
やや疑念	かな
疑念	なの(か)

度合い	
肯定	よ
やや肯定	かも
やや否定	ないかも
否定	ない

(c) 好印象 - 悪印象の度合いを表現する程度副詞

度合い	程度副詞	
	肯定文	否定文
強い	すごく	ちっとも
強め	相当	さほど
弱め	割合	たいして
弱い	そこそこ	あまり

行った。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、それぞれの文が与える印象に即した発話状況を文毎に設定した。発話者に発話内容に沿った状況を自由に想像してもらい、極力自発的な対話を促した。例えば、“すごく優しいよ”という発話に対しては、発話印象としての「好印象」及び「確信」、を意識付けるため、“ある人のことに対して、「どんなの人なのかなぁ・・・」と、少し不安げに尋ねる友人に対しての発話”というような発話状況を規定した上で、相手を考えた状況を想像してもらい、発話を収録した。デザインした 256 文のうち、64 文を 1 セットとして、発話者に提示し、心地良く発話できる範囲内で、無理のない発話をお願いした。その結果、一人当たりの発話数は、平均 2 セットで、合計 896 発話を得た。発話者は、東京方言／標準語を話す日本語を母語とする 24 歳～35 歳の成人 5 名（男性 2 名、女性 3 名）で、静かな環境で録音された。また比較検討のため、対話音声の録音後に、同一発話内容で読み上げ調発話の録音も行った。

入力文を構成する各語彙の印象属性が関与する韻律制御特性を分析するために、収録した対話音声と、読み上げ音声の比較を行った。各印象表現が与える出力対話韻律への効果を調べるために、2.3 の一語発話「ん」の分析結果を参考に、文を構成する各語彙有する印象属性と、それらの印象表現が制御する対話韻律の関係を分析した。具体的には次に述べる、印象表現と対話韻律の対応関係である。

(1) 好印象 - 悪印象の印象表現と、F0 平均値

好印象／悪印象が、高い／低い F0 平均値に影響を及ぼす。

(2) 確信 - 疑念、肯定 - 否定の印象表現と F0 パタン

確信 - 疑念は、下降、上昇＋下降、緩やかな下降、上昇の順に、

肯定 - 否定は、下降、緩やかな下降、上昇、上昇＋下降の順に影響する。

(3) 確信 - 疑念、肯定 - 否定の印象表現と F0 の発話時間長

確信・肯定／疑念・否定が、短い／長い発話時間長に影響を及ぼす。

分析にあたっては、まず対話音声と読み上げ音声の韻律特徴量の違いを比較した。F0 平均値と発話時間長に関しては、対話音声の特徴量から、単純に、読み上げ音声の特徴量を減算した。F0 パタンは、まず、音声分析プログラム Praat[42]を用いて、それぞれの対話音声サンプルの F0 の高さの 10 ポイントを抽出し、パタンの単純化を行い、それぞれのサンプルが F0 パタンの 4 パタン（上昇、平坦、上昇+下降、下降）あるいは、それ以外かを手動で分類した。

次に、抽出された韻律特徴に、各語彙の印象属性がどのように関与しているかを調べるために、好印象/悪印象の度合いを示す程度副詞の強さのレベル毎に、F0 平均値の特徴量を計算した。疑念/確信、及び否定/肯定の印象表現が、対話韻律に及ぼす影響に関しては、それぞれ疑念/確信及び、否定/肯定の終助詞の強さのレベル毎に、発話時間長と、4 つのパタン（上昇、平坦、上昇+下降、下降）に分類された F0 パタンの数を計算した。

6.3 語彙の印象属性にもとづいた対話韻律の理解

前節で述べた印象表現と対応する対話韻律特徴の関係を検証するため、対話韻律と読み上げ韻律との違い求めた。まず図 6.1 に、前節の(1)で述べた程度副詞によって表現される好印象/悪印象の印象の強さの度合いによる、読み上げ韻律に対しての対話韻律の F0 平均値の違いを示す。確信 - 疑念(a)、肯定 - 否定(b)の終助詞を伴う文においても、全体として、好印象 - 悪印象の印象表現に対応して右上がり／下がりの曲線を示していた。否定(b)の終助詞を含む文では、文意が逆転するため、肯定(b)とは逆の傾向を示していた。これは、好印象 - 悪印象の印象表現が、複数の語彙の印象属性の中の一つであったとしても、単独語彙の場合と同様に、F0 平均値に影響を与えていることを示唆する。すなわち、形容詞の好印象／悪印象の印象属性が、その対話韻律の高い／低い F0

平均値と直接的に関係していた。さらには、確信 - 疑念、肯定 - 否定の終助詞も、F0 の平均値に影響を及ぼしていることが判明した。

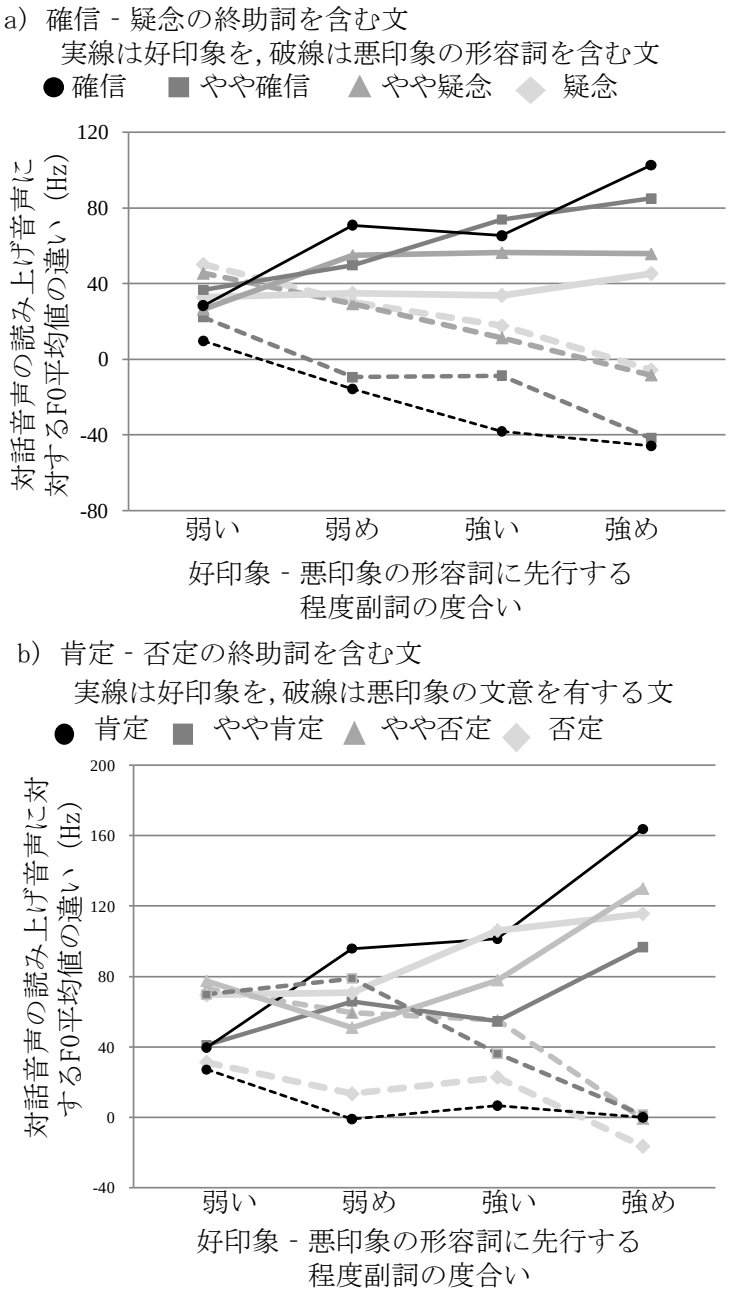


図6.1　好印象 - 悪印象の印象の強さの度合いに応じたF0平均値の変化

次に、前節の(2)で述べた、確信‐疑念、肯定‐否定の印象表現が F0 のパタンに与える影響を調べるため、それらの印象表現に対応する終助詞を含む文毎に、4 つのパタンに該当するサンプル数を数えた。図 6.2 に示すように、それぞれの F0 のパタンの最も多いサンプルの数が、下降、上昇+下降、緩やかな下降、上昇の順で、確信から疑念に、下降、緩やかな下降、上昇、上昇+下降の順で、肯定から否定に対応していた。この結果は、単独語彙での結果と一致する。

さらに、前節の(3)にあたる、確信‐疑念、肯定‐否定の印象表現の違いと発話時間長との対応関係を、図 6.3 に示す。図に示すように、確信、または肯定であればあるほど、その対話韻律の発話時間長は短く、疑念、否定であればあるほど長くなる傾向にあった。すなわち、異なる印象属性を有する複数語彙によって構成される文においても、確信‐疑念、肯定‐否定の印象表現と、発話時間長の直接的対応関係を確認することができた。極端な“疑念”および“否定”の印象表現に関しては、その発話時間長が短くなっていた。これは、本実験で用いた終助詞が、例えば、“怪しい”や“違う”など、“疑念”、“否定”の印象表現を意図した語彙とは違う、断定的な意味合いを持つ語彙を選択したことに原因が考えられる。従って、制御傾向には一貫性が認められたが、語彙の印象属性を選択する際の見直しの必要性があると考えられる。

これら単独語彙で確認した、語彙の印象属性と対応する対話韻律特徴の関係が、異なる印象属性を有する複数語彙からなる文においても成立することが確認できた。すなわち、この印象表現と対話韻律特徴の不変性により、入力を構成する各語彙の印象属性が制御する対話韻律特徴を加え合わせることによる、対話韻律制御の可能性が示唆された。

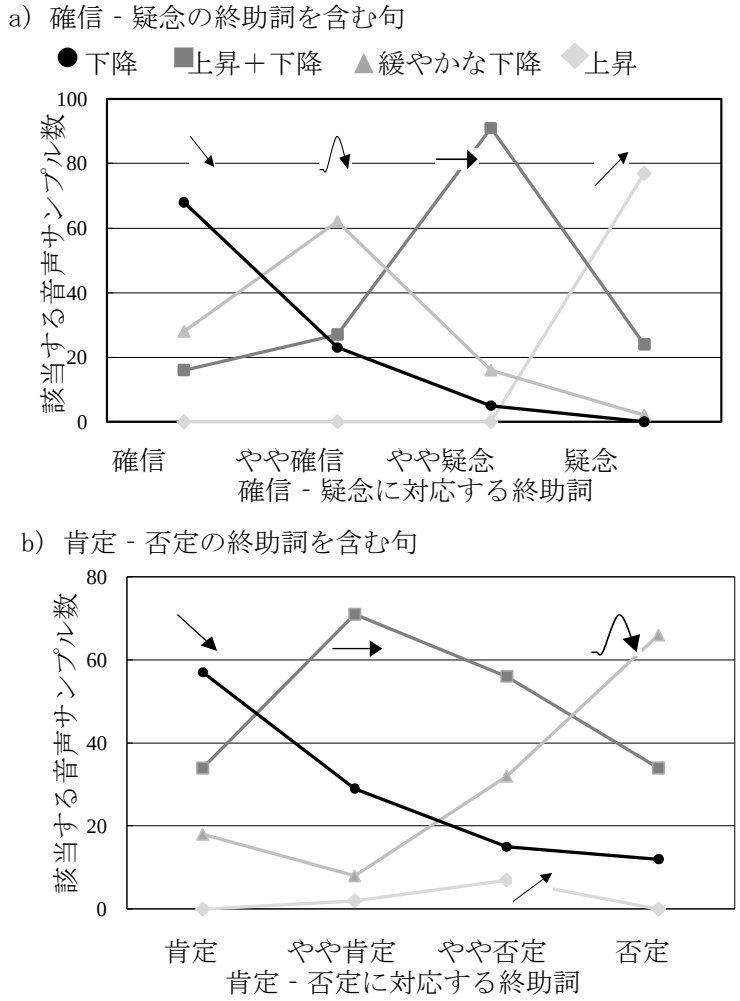


図6.2　確信 - 疑念，肯定 - 否定の印象の強さの
度合いに応じたF0パタンの分類

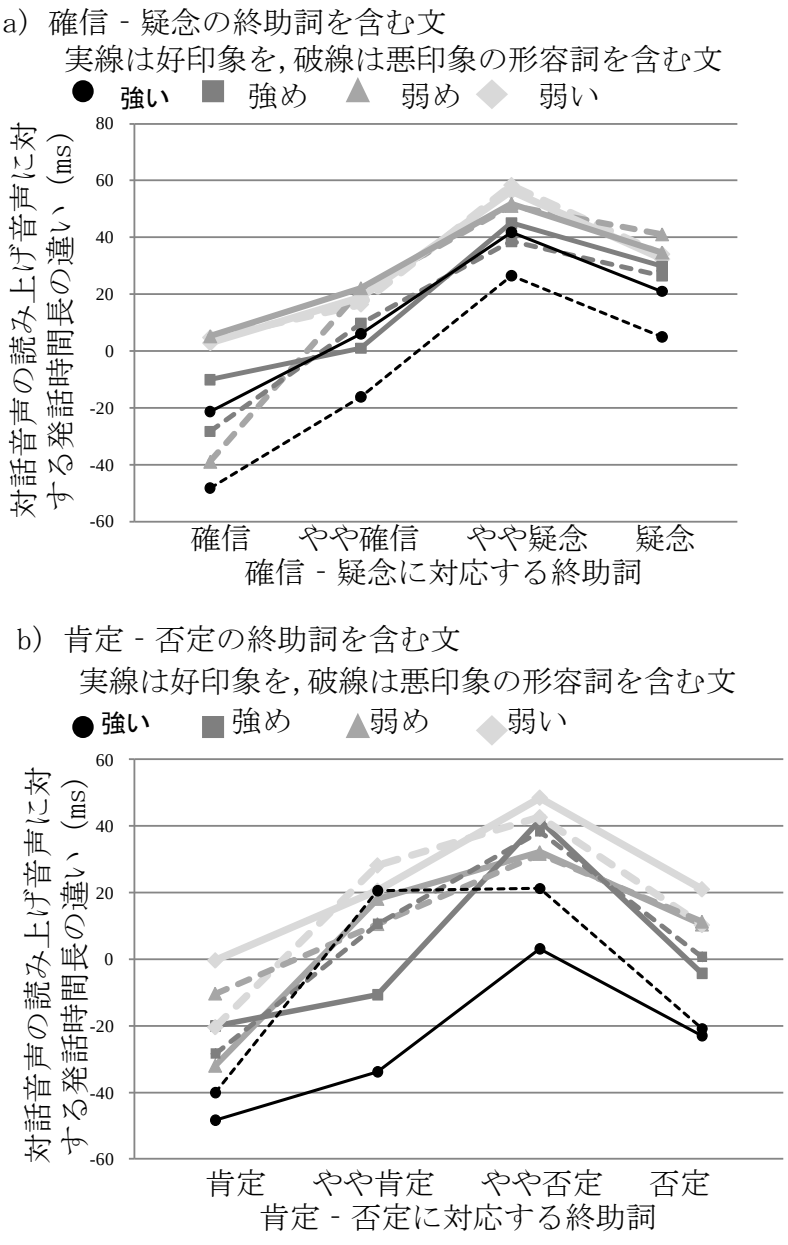


図6.3　確信 - 疑念，肯定 - 否定の印象の
強さの度合いに応じたF0平均値の変化

6.4　6 章のおわりに

本検討では、単独語彙を用いて検証した、語彙の印象属性を用いた対話韻律生成方法が、より一般的な入力へ適用できる可能性の検証をおこなった。このため、異なる印象属性を有する複数語彙からなる文の対話韻律の分析を行い、文を構成する各語彙の印象属性の、全体の出力対話韻律への影響を調べた。

分析対象とした文は、3 軸 6 方向（確信、疑念、肯定、否定、好印象、悪印象）の典型印象表現を印象属性として有する、形容詞、終助詞の組み合わせから構成した。対話韻律分析の結果、単独語彙において成立していた印象属性と韻律制御特性の対応関係、入力が複数語彙の場合でも有効であった。すなわち、発話を構成する語彙の印象属性に応じた韻律制御特性を加え合わせることで、全体の出力対話韻律が説明できるようであることが判明した。これにより、提案した対話韻律生成方法が、より一般的な入力に対しても柔軟に対応できる可能性が示唆された。

第 7 章　結論

7.1　本研究のまとめ

音声合成の分野においては、コーパスベース音声合成や波形合成方法の導入等により、出力される音声の品質は大幅に向上し、その適応範囲も広がった。テキスト入力の読み上げ等の音声出力を対象としていた音声合成の、対話型アプリケーションなどでも有効な対話音声の出力が望まれるが、現在まで、その対話韻律の制御は実現されていない。対話場面で有効な TTS システムの実現のためには、解決すべき課題が、主に 2 つある考えられる。まず、(1)対話場面に出現する多様な韻律を、どのように規定するかという問題。対話韻律に影響を与えていると考えられる種々の要因のうち、出力対話韻律の規定方法を解明し、またそれら規定内容が対話韻律制御特性と対応付けることが、対話音声合成を可能とするのに必要である。次に、(2)対話韻律は、発話内容に即して語句レベルで、動的に制御されなければいけないという問題。対話という双方向からの発話が行われる場面で有効な TTS システムのためには、入力される発話内容毎に、適切な対話韻律制御が行われる仕組みが必要である。

本論文では、これらの課題を踏まえて、対話システに有効な対話韻律の規定、および対話韻律生成方法の提案を行った。まず(1)を解決する、対話韻律を規定する情報として、対話韻律の違いが与える印象に着目した。さらに、その印象を入力情報、対応する韻律を出力情報として、(2)に対処するための、対話韻律生成方法の提案を行った。提案方法では、対話韻律を制御する情報として、入力として与えられる語彙の印象属性を用いることを考えた。対話発話で用いる言語内容は、その対話が取り得る韻律を制約すると考えられる。例えば、「きれい」という対話発話は、多くの場合、好印象を表現しており、従って取り得る対話韻律も限定される。このように、言語内容自体が有する印象という情報に着目し、語彙自体が与える印象を、対話韻律制御の要因として考えた。

まず第1章では、従来の音声合成技術が提供する読み上げ調の音声に代え、多様な韻

律生成を目的とした音声合成研究を紹介した。これらの研究はそれぞれの韻律の変化を与える方法として有効であるが、対話韻律を考える上で韻律制御の問題解決に本質的な問題が未解決であることを指摘した。すなわち、先に挙げた2つの問題の解明が不可欠であることを示し、本論文で扱う研究課題を明確にした。さらに、課題解決のために行った研究内容に対応する本論文の構成と概要について述べた。

第2章、第3章では、課題(1)に対処するための、対話韻律の規定方法を示し、対話韻律生成に用いる入出力情報の規定を行った。まず第2章においては、種々の韻律を持つ一語発話「ん」を例に、対話場面に出現する韻律の自由度とそれらを制御する印象の記述を行った。MDSを用いた「ん」の対話韻律の違いが与える聴覚印象の分析により、聴覚印象は3次元（確信 - 疑念、肯定 - 否定、好印象 - 悪印象）による近似的な記述が可能であり、さらには、それら聴覚印象とF0制御特徴（F0の平均値とパタン）とに明確な対応関係が存在することを示した。第3章では、それらの対応関係が、対話韻律生成において入出力情報として利用できる可能性を検証するため、一語発話「ん」が示したそれらの関係が、そのまま一般の語彙でも成立するかどうかを調べた。3次元6方向の典型印象表現（確信、疑念、肯定、否定、好印象、悪印象）に直接対応する語彙からなる対話発話の韻律分析を行った結果、一語発話「ん」と同様の関係が成立することが確認できた。さらに、F0特徴に加え発話時間長も、語彙が与える印象と関連することが判明した。

第4章においては、前章までの分析結果にもとづき、課題(2)を解決するための、語彙が与える印象にもとづいた対話韻律生成方法を提案した。提案する方法では、従来の言語情報にもとづく韻律制御に加え、新たに、語彙が有する印象属性にもとづいた制御を加えた対話韻律生成を行う。提案した対話韻律生成方法の妥当性を示すため、語彙の印象属性を用いた韻律生成実験を行った。前述の6種類の典型印象表現に対応する語彙からなる読み上げ音声を用い、各々が対応する聴覚印象を持つ一語発話「ん」のF0平均値とパタン、および発話時間長を用いて対話韻律を追加した合成音声を作成した。F0生成には、韻律生成過程を反映する生成過程モデルを用い、生成パラメータのレベルで対話

韻律成分の追加を行った。生成した対話韻律音声サンプルを用いた聴取実験により、提案した対話韻律生成方法の妥当性を確認した。

第5章、第6章では、提案方法のより一般的な入力への適用の可能性を検証した。第5章においては、提案方法で扱う語彙の印象属性のレベルでは、言語依存性が低く、他言語への適用の可能が考えられるため、英語を対象にした対話音声の合成を試みた。日本語語彙を対象とした韻律生成実験と同様の手順で、対話韻律生成を行い、自然性聴取実験によって、その妥当性の確認を行った。第6章においては、単独の語彙を用いて検証してきた提案方法を、より一般的な、複数語彙の組み合わせによって構成される入力文に展開するための検討を行った。異なる印象属性を有する複数の語彙から構成される文の対話発話の韻律を分析した結果、発話を構成する各語彙が有する印象属性に対応する韻律制御特性を加え合わせることで、全体の出力対話韻律が説明できることが判明した。これにより、提案方法がより一般的な入力に対しても適用可能である見通しを得ることができた。

7.2　今後の課題

本論文では、対話場面で用いる対話音声の生成を目的として、新たな韻律生成方法の提案を行った。検証に用いたサンプルは、典型的な一部の語彙に対してのみに限られている。しかしながら、提案方法の原理はより一般的な表現への拡張も可能であると考える。今回扱った、確信 - 疑念、肯定 - 否定、好印象 - 悪印象からなる3次元印象空間の3軸6方向の典型印象表現に対応する代表的語彙のみでなく、印象空間内に存在する多数の語彙にも拡張できると期待している。また、同様の方法論を用いて、本検討で用いた印象表現以外の印象属性への拡張も可能であると考えられ、今後より一般的な表現への適用に関する検討が必要である。

またより複雑な語彙の組み合わせから構成される入力に対しての、提案方法の有効性の確認や、対話の焦点の影響などの取り扱いに関する検討、会話の相手の発言内容などの情報が加味された時の対話音声としての自然性の問題に関する検討、提案方法を実際

の対話音声合成に用いるための制御モデルの構築など、提案方法の適用領域の拡大のために、さらなる検討が必要である。

さらに、実際の対話場面における対話韻律の制御に際しては、語彙の印象属性から得られる情報の限界の問題に加えて、発話者意図や発話状況などの種々の考慮が必要であると思われる。しかしながら、対話場面に出現する多様な対話韻律を、発話内容に即して自律生成する提案方法は、ロボットやアニメーションのような、リアルタイムでの音声出力が要求されるようなアプリケーションにも対応できると考えられ、対話場面に有効なTTSシステム実現への道を開くことができたと言える。

謝辞

本研究を進めるにあたり、多大なるご指導とご助言を賜りました早稲田大学大学院国際情報通信研究科　匂坂芳典教授に深く感謝致します。

本論文をまとめるにあたり、貴重なご意見を賜りました早稲田大学大学院国際情報通信研究科　山崎芳男教授、早稲田大学大学院国際情報通信研究科　河合隆史教授、早稲田大学スポーツ科学学術院　誉田雅彰教授、　早稲田大学理工学術院　小林哲則教授に感謝致します。

本論文中の研究の共同研究者として、懇切丁寧に御意見、御討論、御助言を頂きました、加藤宏明博士（現在、独立行政法人情報通信研究機構／ATR メディア情報科学研究所）、津崎実博士（現在、京都市立芸術大学）、渋谷渚様（現在、ソフトウエア情報開発株式会社）、李克様（現在、トランスコスモス株式会社）に深く感謝致します。

本論文中の研究に用いたデータやツールを提供して下さった、ニックキャンベル博士（現在、奈良先端科学技術大学院大学）、河原英紀博士（現在、和歌山大学）に深く感謝致します。

在学中に様々な面でお世話になりました、早稲田大学大学院国際情報通信研究科　匂坂研究室に所属する学生、また卒業生の皆様に深く感謝致します。

最後に、様々な面で辛抱強く支えてくれた、夫のグリーンバーグエリック、在学中に無事に産まれ、笑顔と元気を与え続けてくれた、息子のグリーンバーグ尚憧（ノア）、成澄（セス）に深く感謝致します。

参考文献

[1] H. Fujisaki and K. Hirose: “Analysis of voice fundamental frequency contours for declarative sentences of Japanese”, J. Acoust. Soc. Jpn. (E), 5, 233–242, 1984.

[2] K. Hirose, H. Fujisaki and M. Yamaguchi: “Synthesis by rule of voice fundamental frequency contours of spoken Japanese from linguistic information”, Proc. 1984 IEEE ICASSP, 1984.

[3] 平井俊男, 岩橋直人,匂坂芳典: “統計的手法を用いた基本周波数制御規則の自動抽出”,電子情報通信学会論文誌, D-II, Vol. J78-D-II, No. 11, pp. 1572-1580, 1995.

[4] 小川博正, 匂坂芳典: “発話情報を用いた F0 制御パラメータの自動抽出”,日本音響学会 2004 年春季研究発表会講演論文集, pp. 265-266, 2004.

[5] 匂坂芳典: “F0 パターン概形制御の定量的検討”,電子情報通信学会音声研究会資料, SP89-111,1990.

[6] 阿部匡伸, 佐藤大和: “音節区分化モデルに基づく基本周波数の 2 階層制御方式”, 日本音響学会, Vol. 49, No. 10, pp. 682-690, 1992.

[7] 箱田和雄, 佐藤大和: “文音声合成における音調規則”, 電子通信学会論文誌 J63-D, 9, pp.715-722, 1980.

[8] 海木延佳, 武田一哉, 匂坂芳典: “言語情報を利用した母音継続時間長の制御”, 電子通信学会論文誌, Vol. J75-A, No.3, pp.467-473, 1992.

[9] 酒寄哲也, 佐々木昭一, 北川博雄, “規則合成のための数量化 I 類を用いた韻律制御”, 日本音響学会講演論文集 3-4-17, pp. 245-246, 1986.

[10] 海木延佳, 匂坂芳典: “文音声における子音継続時間長の設定”, 日本音響学会秋季研究発表会講演論文集, Vol.1, pp. 259-260, 1990.

[11] J. Venditti and J. P. H. Santen J.: "Modeling segmental durations for Japanese text-to-speech synthesis", .Proc. 3rd ESCA Workshop on Speech Synthesis, pp. 31-36, 1998.

[12] J. P. H. Santen: “Contextual effects on vowel duration”, Speech Communication, vol.11, pp. 513-546- 128, 1992.

[13] M.D.Riley: “Tree-based modeling of segmental durations”, in Talking machines, pp. 265-273, Elsevier, 1992.

[14] 河井恒, 戸田智基, 山岸順一, 平井俊男, 倪晋富, 西澤信行, 津崎実, 徳田恵一: “大規模コーパスを用いた音声合成システム XIMERA”, 電子情報通信学会論文誌, J89-D(12), pp. 2688-2698, 2006.

[15] 全炳河, 大浦圭一郎, 能勢隆, 山岸順一, 酒向慎司, 戸田智基, 益子貴史, ブラックアラン, 徳田恵一: “HMM 音声合成システム(HTS)の開発”, IPSJ SIG Notes 2007(129), pp.301-306, 2007.

[16] N. Higuchi, T. Hirai and Y. Sagisaka: “Effect of speaking style on parameters of fundamental frequency contour,” in Progress in Speech Synthesis, J. P. H. van Santen, R. W. Sport, J. P. Olive, J. Hirschberg, Eds. (Springer- Verlag, New York), pp. 417–428, 1996.

[17] 坂田真弓, 広瀬啓吉: “対話音声の韻律的特徴の分析と合成”, 電子情報通信学会技術研究報告, SP 95(42), pp.55−62, 1999.

[18] N. Kaiki and Y. Sagisaka: “Prosodic characteristics of Japanese conversational speech,” IEICE Trans., E76-A, 1927–1933, 1993.

[19] Y. Hashizawa, S. Takeda, M. D. Hamzah and G. Ohyama: “On the differences in prosodic features of emotional expressions in Japanese speech according to the degree of emotion”, Proc. Speech Prosody 2004 Nara, pp. 655−658, 2004.

[20] S. Takeda, G. Ohyama, A. Tochitani and Y. Nishizawa: “Analysis of prosodic features of ‘anger’ expressions in Japanese speech”, J. Acoust. Soc. Jpn. (J), 58, pp. 561−568, 2002.

[21] N. Audibert, D. Vincent, V. Auberge and O. Rosec: “Expressive speech synthesis: Evaluation of a voice quality centered coder on the different acoustic dimensions”, Proc. Speech Prosody 2006, pp. 525−528, 2006.

[22] T. Johnstone and K. R. Scherer: “The effects of emotions on voice quality”, Proc. Int. Congr. Phonetic Sciences, San Francisco, pp. 2029−2031, 1999.

[23] R. Kehrein: “The prosody of authentic emotions”, Proc. Speech Prosody 2002 Aix-en-Provence, pp. 423−426, 2002.

[24] 桂 聡哉, 広瀬幸吉, 峯松 信明: “感情音声合成のための生成過程モデルに基づくコーパスベース韻律生成とその評価”, 電子情報通信学会技術研究報告, SP, 102(749) , pp. 31−36, 2003.

[25] N. Campbell, A. Iida, F. Higuchi and M. Yasumura: “A corpus-based speech synthesis system with emotion”, Speech Communication, 40, pp. 161−187, 2003.

[26] J. M. Montero, J. Gutiérrez-Arriola, S. Palazuelos, E. Enríquez, S. Aguilera and J. M. Pardo: “Emotional speech synthesis: From speech database to TT”, Proc. ICSLP 1998, pp. 923−926, 1998.

[27] M. Bulut, S. Narayanan, and A. Syrdal: “Expressive speech synthesis using a concatenate synthesizer”, Proc. ICSLP 2002, pp. 1265−1268, 2002.

[28] E. Zovato, A. Pacchiotti, S. Quazza, S. Sandri: “Towards emotional speech synthesis: a rule based approach”, Carnegie Mellon University Pit sburgh the City: 5th ISCA Speech Synthesis Workshop, 2004.

[29] 赤川達也, 岩野公司, 古井貞熙: “HMM を用いた話し言葉音声合成のためのモデルの検討”, 電子情報通信学会 技術研究報告, No. SP2007-3, pp. 13-18, 2007.

[30] M. Tachibana, J. Yamagishi, K. Onishi, T. Masuko and T. Kobayashi: “HMM-based speech synthesis with various speaking styles using model interpolation”, Proc. Speech Prosody 2004 Nara, pp. 413–416, 2004.

[31] 都築亮介, 全炳河, 徳田恵一, 北村正: “HMM 音声合成における感情表現のモデル化”, 電子情報通信学会技術研究報告, vol.103, no.264, pp. 25-30, 2003.

[32] K. Tokuda, T. Masuko, N. Miyazaki, T. Kobayashi: “Hidden Markov models based on multispace probability distribution for pitch pattern modeling”, Proc. ICASSP, Phoenix, pp. 229-232, 1999.

[33] J. Tao, L. Xin and P. Yin: “Realistic visual speech synthesis based on hybrid concatenation method”, IEEE Trans, 17, 3, pp. 469-477, 2009.

[34] H. Kawanami, Y. Iwami, T. Toda, H. Saruwatari, K. Shikano: "GMM-based Voice Conversion Applied to Emotional Speech Synthesis," Proc. 8th European Conference on Speech Communication and Technology, pp.IV-2401-2404, 2003.

[35] 森山剛, 森真也, 小沢慎治: “韻律の部分空間を用いた感情音声合成”, 情報処理学会論文誌, Vol.50, No.3, pp.1181-1191, 2009.

[36] WaveSurfer: Centre for Speech Technology (CTT) at KTH in Stockholm, Sweden, http://www.speech.kth.se/wavesurfer/, 2005.

[37] The JST/CREST Expressive Speech Processing Project, introductory web pages at: www.isd.atr.co.jp/esp

[38] N. Campbell: “The Recording of Emotional Speech; JST/CREST database research”, Proc. LREC2002, Vol. 6, pp. 2029-2032, 2002.

[39] N. Campbell: “Towards synthesizing expressive speech; designing and collecting expressive speech data ”, Proc. Eurospeech 2003, pp.1637–1640, 2003.

[40] W. Torgerson: “Multidimensional scaling: I. Theory and method”, Psychometrika 17, pp.401-419, 1952.

[41] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigné: “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds”, Speech Communication, 27, 187–207, 1999.

[42] Boersma, P. and Weenink, D : (http://www.fon.hum.uva.nl/praat)

研究業績一覧

査読付学術論文

1. ○ グリーンバーグ陽子, 加藤宏明, 津崎実, 匂坂芳典: “語彙が与える印象に基づく対話韻律生成”, 日本音響学会誌, 66 巻, 12 号, 2010.

2. ○Yoko Greenberg, Nagisa Shibuya, Minoru Tsuzaki, Hiroaki Kato, Yoshinori Sagisaka: “Analysis on paralinguistic prosody control in perceptual impression space using Multiple Dimensional Scaling”, Speech Communication, Vol.51 No.7 pp. 585-593, 2009. 7.

3. Yoshinori Sagisaka, Takumi Yamashita, Yoko Kokenawa: "Generation and perception of F0 markedness for communicative speech synthesis", Speech Communication 2005, Vol.46, pp.376-384, 2005.3.

査読付国際会議

4. ○Yoko Greenberg, Minoru Tsuzaki, Hiroaki Kato, Yoshinori Sagisaka: “Communicative prosody generation using language common features provided by input lexicons", Proc. SNLP2009, pp.101-104, 2009.10.

5. Mingzhao Zhu, Ke Li, Yoko Greenberg and Yoshinori Sagisaka: “Automatic extraction of paralinguistic information from communicative speech”, Proc. The 7th International Symposium on Natural Language Processing, pp. 207-212, 2007.12.

6. Ke Li, Yoko Greenberg and Yoshinori Sagisaka: “Inter-language prosodic style modification experiment using word impression vector for communicative speech generation” , Proc. INTERSPEECH 2007, pp. 1294-1297, 2007.8.

7. Ke Li, Yoko Greenberg, Nagisa Shibuya, Nick Campbell and Yoshinori Sagisaka: “On the analysis of F0 control characteristics of nonverbal utterances and its application to communicative prosody generation”, NATO SCHOOL on The Fundamentals of Verbal and Non-verbal Communication and the Biometrical, pp. 179-183, 2006.9.

8. ○Yoko Greenberg, Nagisa Shibuya, Minoru Tsuzaki, Hiroaki Kato, Yoshinori Sagisaka: "A trial of communicative prosody generation based on control characteristic of one word utterance observed in real conversational speech", Proc. Speech prosody 2006 pp. 37-40, 2006.3.

9. ○Yoko Greenberg, Minoru Tsuzaki, Hiroaki Kato, Yoshinori Sagisaka: "Communicative speech synthesis using constituent word attributes", Proc. INTERSPEECH 2005, pp.517-520, 2005.9.

10. ○Yoko Kokenawa, Minoru Tsuzaki, Hiroaki Kato, Yoshinori Sagisaka: “F0 control characterization by perceptual impressions on speaking attitude using Multiple Dimensional Scaling”, Proc, IEEE ICASSP 2005 Philadelphia, SP-P1. 3. (I-273), 2005.3.

国内研究会

11. ○グリーンバーグ陽子, 津崎実, 加藤宏明, 匂坂芳典: “発話印象表現に基づく対話韻律制御の分析(Analysis of communicative speech prosody control based on perceptual impression)" , 言語・音声理解と対話処理研究会（人工知能学会）SIG-SLUD-A501, pp.33-38, 2005. 6.

12. ○苔縄陽子, 津崎実, 加藤 宏明, 匂坂芳典: “基本周波数パタンに見られる発話態度の分析”, 情報処理学会研究報告 73, Vol. 2004, No. 74,pp. 87-92, 2004.7.

国内大会

13. 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, 2007.9.

14. 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, 2007.9.

15. 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”, 日本音響学会2006年秋季研究発表会講演論文集, 3-6-9, pp. 233-234, 2006.9.

16. ○グリーンバーグ陽子, 津崎実, 加藤宏明, 匂坂芳典: “ノンバーバル発話の韻律制御に基づく会話韻律の生成” , 日本音響学会2006年春季研究発表会講演論文集, 2-4-9, pp. 325-326, 2006.3.

17. 渋谷渚, グリーンバーグ陽子, 匂坂芳典: “基本周波数特性に基づく一語発話「ん」の分類について”, 日本音響学会 2005 年秋季研究発表会講演論文集, 2-6-7, pp. 271-272, 2005. 9.

18. ○グリーンバーグ陽子, 津崎実, 加藤宏明, 匂坂芳典: “入力語彙情報に基づく対話韻律制御” , 日本音響学会 2005 年秋季研究発表会講演論文集 2-6-8, pp. 273-274, 2005. 9.

19. ○苔縄陽子, 匂坂芳典, 津崎実, 加藤宏明: “多次元尺度構成法を用いた対話音声の

基本周波数パタン分析”, 日本音響学会 2004 年秋季研究発表会講演論文集, Vol.1, 3-2-21, pp. 357-358, 2004. 9.

著書

20. 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, 2006.6.

その他（特許）

21. 匂坂芳典, グリーンバーグ陽子, 津崎実, 加藤宏明: “音声合成装置,音声処理装置,およびプログラム”, 特開 2006-330060, 公開日 2006.12.7.