

韻律の分析，定式化とモデル化*
Analysis, Formulation, and Modeling of Prosody

東京大学 名誉教授

Professor Emeritus, University of Tokyo

藤崎 博也

Hiroya FUJISAKI

Abstract: The aim of the Research Group on Prosody Modeling is to analyze, formulate and model prosody in an objective and quantitative manner and to elucidate the mechanisms and processes whereby linguistic, paralinguistic and nonlinguistic information is expressed by prosodic and segmental features of speech. As for the tonal aspects of prosody, the mechanisms for active lowering of fundamental frequency in tone languages are explained in the light of electromyographic data on extrinsic laryngeal muscles and the anatomy of human larynx. This explanation provides a basis for the model of fundamental frequency control already presented by Fujisaki and his co-workers. The applicability of the model with provisions for both positive and negative polarities for the local commands is experimentally confirmed for both non-tone languages including Hindi and Portuguese and tone languages including Mandarin, Cantonese, Shanghainese, Thai and Vietnamese. Together with our previous findings on eight other languages, these results lead to a binary classification of tonal features of languages in terms of local command polarities, and also to a finer classification of tone languages. A new method for automatic extraction of command parameters from observed contours of fundamental frequency is also developed. As for the temporal structure of Japanese, various dialects including the Tokyo dialect are analyzed and compared from the viewpoint of strength of the moraic constraint in three different domains: segmental durations, phonetic changes, and phonemic/phonotactic constraints. As for paralinguistic information, the process of its transmission is studied both from acoustic and physiological observations of speech production and from speech perception. Results of production studies show that both segmental and prosodic characteristics of speech are influenced by paralinguistic information. Results of perception studies show that paralinguistic information influences both phonological features such as boundary pitch movements and non-phonological features such as voice quality, and the former are found to be more language-specific than the latter. Furthermore, an integrated approach toward modeling the dynamic characteristics of speech rate and source intensity in line with the model of fundamental frequency control is presented, and is applied to a quantitative analysis of the effects of emphasis upon the three features. Finally, the information on the speaker's emotion is selected as an example of non-linguistic information. Based on a four-dimensional representation of emotion (love/hate, anger/fear, joy/sorrow, and expectation/surprise), natural utterances with three degrees of emotional strength along each dimension are used as stimuli in a perceptual test of emotion. The results indicate that the speaker's emotion is perceived correctly in most cases except those along the expectation/surprise dimension.

Keywords: Prosody modeling, Tonal aspects, Fundamental frequency control, Classification of languages, Automatic extraction of command parameters, Temporal structure of Japanese, Strength of moraic constraints, Production of paralinguistic information, Perception of paralinguistic information, Boundary pitch movements, Voice quality, Integrated approach to modeling prosodic features, Perception of emotion.

* 研究課題：韻律的特徴の分析、定式化およびモデル化 課題番号：

1. 研究の目的

音声の韻律的特徴は、文字言語によっては表現されない種々の情報を表現するものであるが、定量的な把握・記号化の困難さのために、従来十分な研究が行われず、それらの情報とその担い手としての基本周波数パターンや発話速度・テンポ・リズム等との間の関係が十分に定量的に把握されていなかった。この研究では、韻律によって表現される情報を (1)言語情報(語彙・統語構造・談話構造に関するもの)、(2)パラ言語情報(話者の意図・態度に関するもの)、(3)非言語情報(話者の個性・感情・個体差に関するもの)、に明確に分類し、まず韻律的特徴、特に基本周波数パターンおよび時間構造のパターンを精密に分析し、次にその結果とそれらの情報との間の関係を明確に定式化した上で、さらにそれらの情報から韻律的特徴を生成する過程を、音声言語情報処理に利用し得る形にモデル化すること、また、それらの情報を統一的に記述する枠組みを構築することを目的とした。なお、本研究班は特例として、表1の通りそれぞれ別の機関に所属し、専攻分野も異なる4名の研究者からなる学際的な研究グループとして組織され、韻律のモデル化を共通の目標としつつ、韻律の諸側面を分担し、研究を進めた。

表1．モデル班の研究組織

研究代表者	藤崎 博也 (東京大学名誉教授)
分担課題	音声の基本周波数パターンのモデル化
研究分担者	窪園 晴夫 (神戸大学文学部教授)
分担課題	日本語における時間制御の定式化
研究分担者	前川喜久雄 (国立国語研究所部門長)
分担課題	パラ言語情報の生成と知覚
研究分担者	大野 澄雄 (東京工科大学助教)
分担課題	韻律的特徴の統合的モデル化と非言語情報の表出過程の分析

2. 音声の基本周波数パターン(F_0 パターン)の生成に関する生理学的・物理学的機構の解明とモデル化、諸言語への適用性の検証と分類、および所与の F_0 パターンからの指令パラメータの自動抽出 (藤崎博也)

2.1 音声の基本周波数パターン(F_0 パターン)の生成に関する生理学的・物理学的機構の解明とモデル化

音声の F_0 パターンの生成過程に関しては、すでに代表者藤崎らにより定量的なモデル化が行われ、日本語のみならず多くの言語への適用性が示されていたが、声調言語など、 F_0 の局所的で積極的な下げを伴う言語の喉頭制御の機構は未解明であった。本研究では、まずこの点に関して検討を行い、従来知られていた胸

骨舌骨筋(sternohyoid muscle)のみではなく、甲状舌骨筋(thyrohyoid muscle)が同時に活動することにより、前者が舌骨の固定に、後者が甲状軟骨の回転に寄与し、その結果として F_0 の下げが生ずることを、外喉頭筋の筋電図学的所見と喉頭の解剖学的構造から、はじめて明らかにした[1, 2]。

2.2 モデルの適用性の実証と諸言語の分類

上記のモデルの妥当性を、標準中国語[3]・タイ語[4]・ベトナム語[5]・広東語[6]・上海語[7]を含む声調言語、およびポーランド語[1]・ポルトガル語[8]・ヒンディー語[9]を含む非声調言語の F_0 パターンに関して実証すると共に、すでに実証済みの日本語をはじめ多くの言語に関する知見[10]と総合して、それらの諸言語が、表2に示すように、局所指令(非声調言語ではアクセント指令、声調言語では声調指令)の極性が正のものと、正・負両極性のものとに大別されること、さらに各声調言語が、音節の前半部分と後半部分の声調指令の極性(正・零・負)から構成される2次元平面上での声調型の布置により区別されることを明らかにした[11, 12]。

表2．局所指令の極性に着目した諸言語の分類

極性	言語
正のみ	英語、エストニア語、ドイツ語、ギリシャ語、韓国語、ポーランド語、スペイン語
正と負	ヒンディー語、ポルトガル語、スウェーデン語、北京語*、広東語*、上海語*、タイ語*、ベトナム語*

* 声調言語

2.3 所与の F_0 パターンからの指令パラメータの抽出

一方、実測の F_0 パターンから指令のパラメータを推定することは、解析的には解けない逆問題であるが、実測の F_0 パターンに含まれる無音区間、子音調音の影響、測定誤差などの影響を軽減したのち、区分的3次曲線で近似すれば、局所指令の位置と振幅の第1近似値が解析的に求められることを示した[13]。また、それらの値を用いてフレーズ指令の位置と大きさの第1近似を求め、さらにそれらの第1近似値を出発点としてAnalysis-by-Synthesis法による逐次近似を行って、最適解を求める方法を提案し、従来手法よりも高い精度の推定が可能であることを示した[14]。さらに、入力音声のテキスト情報が既知の場合には、テキストに対する自然言語処理を行い、その結果得られる統語情報とアクセント情報とを利用して上記の方法による処理結果を修正することにより、より高い抽出精度が得られることを示した[15]。

3. 日本語のモーラリズムの実態に関する実験音韻論的研究(窪園晴夫)

日本語はモーラを基本とする言語(モーラ言語、モーラリズム言語)であることが知られている。その実態を、下記の4つの視点から考察した。

3.1 モーラリズムの揺れが生じる条件と範囲の解明 [16]

日本語のモーラリズムの言語であり、モーラの構造により「ビール」-「ビル」のような単語ペアが区別されている。その一方で、そのようなモーラ単位の制御が発話の一部で崩れることも知られている。本研究ではモーラリズムが崩れる条件と範囲を、音響実験と視覚実験の両面から分析した。音響実験の結果、(i)長音節(2モーラ音節)は語中より語末において短くなる傾向がある、(ii)短い音節(1モーラ音節)も語末において短くなる傾向があるため、長音節と短音節の差が語中より語末において小さくなる、以上2つの傾向を検出した。また視覚情報だけで長音節と短音節が区別できるかを見る視覚実験から、(iii)4種類の長音節の中でも長母音を含む音節だけが、特に語末という音声環境において、短音節と区別できなくなることを確認した。

3.2 モーラ単位の制御を示す諸現象の分析と他言語・方言との比較 [17]

日本語(東京方言)の言語現象を他の言語と比較分析し、(i)時間的現象、(ii)非時間的音声現象、(iii)音韻現象においてモーラ単位の制御が見られることを確認した。(i)について、母音と尾子音(音節末子音)との間に補償効果が観察されないという点において東京方言は英語や韓国語などの言語と異なっていることがわかった。(ii)の非時間的音声現象としては、たとえばサ行やタ行の子音が*/i/*や*/u/*の前で音価を変えてしまう音声的同化が観察される。ドイツ語をはじめとする諸言語では、母音は頭子音ではなく尾子音に強い同化作用を及ぼすことから、日本語に見られる同化作用はモーラを範囲(domain)とした制御を示唆するものである。同様のモーラ単位の制御は母音無声化という音声現象にも見られる。(iii)の音韻現象としては音素配列制約、(外来語における)母音挿入、言い間違い、吃音などの現象をあげることができる。たとえば音素配列制約では、ヤ行やワ行の母音生起に観察されるように、母音と頭子音の間に強い共起制限が見られ、母音と尾子音(撥音、促音)の間にはそのような共起制限は見られない。日本語以外の言語では逆のタイプの共起制限を示すものもあることから、日本語の共起制限はモー

ラ単位の制限であることがうかがえる。言い間違いや吃音といった発話の非流暢性現象においても、英語などではほとんど観察されないモーラ単位の音の交換や繰り返しを観察されることから、これらの現象でもモーラを単位とした制御がなされていることがわかった。

3.3 モーラ性の方言比較と時間制御のメカニズム [18]

日本語の方言の中でモーラ性が高いと言われる近畿方言、東京方言と比較して、音節方言とされる鹿児島方言における時間制御のメカニズムを考察した。音響実験の結果、鹿児島方言の特徴として、(i)重子音(促音)の部分(CC)は単子音部分(C)に比べるとはるかに長い。(ii)にもかかわらず、全体的な傾向として東京方言に観察されるようなモーラ単位での等時性は認めがたい。(iii)重子音の後ろの母音(閉音節に後続する母音)は単子音に後続する母音に比べ、一貫して短くなる傾向がある。(iv)等時性という点では、CVと(C)CVの部分が同じ長さになる傾向が認められる。以上の4つの特徴が検出された。

3.4 母音無声化をもとにした時間制御のメカニズム・単位の分析 [19]

母音の無声化は母音の持続時間と直接的な関係があると言われている。母音の無声化が実時間に制御された現象であるならば、母音の無声化が生じやすい音声・音韻環境ではすべての母音が短くなることが予想される。このことを実験的に確認するために、音響実験を行った。その結果、これまで無声化が起こりやすいと言われてきた音声環境において、*/a/*などの非無声母音も短くなること、また、促音に先行する母音は他の母音より長くなりやすいことを確認した。しかし、「摩擦音に挟まれた環境では母音が長くなりやすい」「同一モーラが連続する場合には最初の母音が短くなりやすい」という仮説については支持する結果が得られなかった。このことは、母音の無声化が母音の持続時間のみによって条件付けられているわけではないことを示唆している。

4. 音声に含まれるパラ言語情報の伝達機序に関する生成・知覚両面からの実験的検討(前川喜久雄)

4.1 生成機序に関する成果 [20, 21, 22]

- (1) 4種類のパラ言語情報 A 感心、D 落胆、I 無関心、S 疑い に対比のため N 中立と F 強調を加えた6種類のパラ言語情報を考察の対象とした。
- (2) 話者3名による436発話の同定実験を実施し、A,D,I,S は最低でも80%以上の精度で同定されることを確認した。

(3) 同定率50%以上の資料を音響分析して、音声基本周波数(F_0)、持続時間、母音のフォルマント周波数(F_1 , F_2)、有声音源のスペクトル傾斜など、検討対象としたすべての音響音声学的パラメータにおいて、パラ言語情報により組織的な影響を確認した。

(4) EMAによる調音運動計測により、「A」と「S」の間に観察される F_2 の組織的差異が舌の前後方向への移動と両唇距離の調整により生じていることを確認した。これによりパラ言語情報が音声の所謂分節的特徴にも影響を及ぼしていることを実証した。

(5) 声門の高速度ビデオ画像の解析によって、「N」「D」「S」の間に有声音源のスペクトル傾斜から予想されたとおりの振動様式の差が認められることを確認した。

(6) 音響分析の成果に依拠して自発音声の韻律ラベリング体系であるX-JToBIを考案し、本研究と並行して実施した『日本語話し言葉コーパス』の音声ラベリングに適用した。

4.2 知覚機序に関する成果 [23, 24]

(1) 上記4.1(2)で利用した同定実験データを多次元尺度法(MDS)で解析することにより、日本語母語話者におけるパラ言語情報の知覚空間を構成した。抽出された三つの次元は、DIM1が「顕著さ」、DIM2が「問合せ」、DIM3が「大きさ」の軸と解釈された。

(2) この3軸における各刺激音の布置と音響分析で得た種々のパラメータ類との回帰分析をおこない、DIM1には全体の持続時間、DIM2には発話末の F_0 形状、DIM3には全体の振幅などが有意に貢献していることを確認した。

(3) パラ言語情報の言語依存性の有無を検討するために、上記と同一の実験を英語を母語とする被験者群に実施した。中級程度の日本語学習歴を有する被験者群の知覚空間は日本語母語話者の空間と基本的に同一の構造を有していたが、学習歴皆無の被験者群の空間には母語話者のDIM2が欠落していた。これによりパラ言語情報には一定の言語依存性が存在することが示唆された。

4.3 音声生成の概念モデル

以上の成果に立脚してパラ言語情報と非言語情報を考慮した音声生成の概念モデルを提案した。このモデルの特徴のひとつはパラ言語情報の生成機序において、1)語ないし句レベルの音韻特徴を操作するものと、2)声質のみを操作するものとの2種類を区別する点である。パラ言語情報の言語依存性は1)に起因し、普遍性は2)に起因すると考えられる。

5. 韻律的特徴の統一的モデル化と、韻律的特徴による感情表現の検討 (大野澄雄)

5.1 各韻律的特徴の定量化のためのモデル [25, 26]

韻律で表現される高さ、長さ、強さの3つの要素は、物理的には音声信号の基本周波数(F_0)、各音の継続長(あるいは発話速度)短時間パワー(あるいは音源強度)の時間的変化パターンとして観測される。このうち、特に音声の F_0 パターンに関しては、既に藤崎らによる生成過程に基づく指令 応答モデルがあり、このモデルによれば、発話内容の言語的およびパラ言語的情報に対応する少数の入力指令によって、実測の F_0 パターンの特徴を正確に表現できることが実証されている。このモデルの指令 応答を参考にして、他の韻律的特徴の定量的な分析手法について検討を行なった。

音声の発話速度や音源強度について、その局所的な振る舞いは、例えば強調の無し・有りのような2つの表現間の相対的な変化として観測される特徴に重要な手掛かりがあるものと考え、まず2つの表現が異なる発話の特徴間の相対的な変化量を定量化し、モデル化することを行なった。発話速度に関しては、局所的な変化量を定量化する手法として、本研究開始以前に提案を行なった局所発話速度比の分析手法を用いて、変化量の時間関数を利用した。その時間変化パターンを F_0 パターンにおけるアクセント指令とその応答関数と類似した機構を導入することにより、発話速度変化指令とその応答関数の導入を行なった。

音源強度パターンについても、音源強度変化指令とその制御機構を仮定し、モデル化を行なった。その結果、 F_0 パターン、発話速度変化パターン、音源強度変化パターンの3種の韻律的特徴を統一的に取り扱うことのできる指令 応答モデルを構築し、強調表現を例とした分析を行なって検討を行なった。

5.2 感情表現と韻律的特徴 [27]

音声の伝達する情報のうち、非言語情報に着目して、非言語情報が韻律によりどのように表現されるかについて検討した。本研究では、非言語情報のうち、とくに感情の表現に関して調べた。

感情について分析を行なうのに先立って、まず感情の分類を行なった。感情の起源が環境の変化に対する適応に由来すると言う立場から、具体的な対象を持つ感情として、(1) 愛好 嫌悪、(2) 怒り 恐れ、の相対する2次元を、また抽象的な対象を持つ感情として、(3) 喜び 悲しみの1次元を、さらに対象の生起に対する準備の程度に関する感情として、(4) 驚き 期待、という1次元を考えることとした。

次に、これら4次元8種類に分類した感情が、言語表現を除外した音声のみによってその程度伝達しうるかについて聴取実験による検討を行なった。その結果、感情の種類を識別に関しては、驚き 期待の次元を除けば、意図した感情が聞き手にほぼ正確に伝達されることが明らかになった。また、感情の程度に関しても3段階程度の強弱をつけた発話であれば、聞き手にもそれらの段階が正確に伝わることを確認できた。

最後に喜び 悲しみに着目して、それらの感情を込めた音声を収集し、モデルに基づく分析を行ない、感情を込めない場合と比べた入力指令の変化について、詳細に分析を行なった。

発表論文

- [1] Hiroya Fujisaki, "The physiological and physical mechanisms for controlling the tonal features of speech in various languages," Proc. Workshop Prosody 2000, Krakow, pp. S1-S15 (2000-10).
- [2] Hiroya Fujisaki, Ryou Tomana, Shuichi Narusawa, Sumio Ohno and Changfu Wang, "Physiological mechanisms for fundamental frequency control in Standard Chinese," Proc. 2000 International Conference on Spoken Language Processing, Beijing, Vol. 1, pp. 9-12 (2000-10)..
- [3] Changfu Wang, Hiroya Fujisaki, Ryou Tomana and Sumio Ohno, "Analysis of fundamental frequency contours of Standard Chinese in terms of the command-response model," Proceedings of 2000 International Conference on Spoken Language Processing, Beijing, Vol.3, pp.326-329 (2000-10).
- [4] Hiroya Fujisaki, Sumio Ohno and Sudaporn Luksaneeyawin, "Analysis and synthesis of F_0 contours of Thai utterances based on the command-response model," Proceedings of 15th International Congress of Phonetic Sciences, Barcelona, Vol. 2, pp. 1129-1131 (2003-8).
- [5] Hansjoerg Mixdorff, H. Nguyen, Hiroya Fujisaki, Mai Luong, "Quantitative analysis and synthesis of syllable tones in Vietnamese," Proceedings of 8th European Conference on Speech Communication and Technology, Geneva, Vol. 1, pp. 177-180 (2003-9).
- [6] Hiroya Fujisaki, Wentao Gu and Keikichi Hirose, "The command-response model for the generation of F_0 contours of Cantonese utterances," Proceedings of 7th International Conference on Signal Processing, Beijing, Vol. 1, pp. 655-658 (2004-9).
- [7] 顧文涛, 広瀬啓吉, 藤崎博也, "上海語音声の基本周波数パターンの生成過程のモデル化," 日本音響学会

2004年秋季研究発表会講演論文集, vol. 1, pp. 393-394 (2004-9).

- [8] Hiroya Fujisaki, Shuichi Narusawa, Sumio Ohno and Diamantino Freitas, "Analysis and modeling of F_0 contours of Portuguese utterances based on the command-response model," Proceedings of 8th European Conference on Speech Communication and Technology, Geneva, Vol. 3, pp. 2317-2320 (2003-9).
- [9] 藤崎博也, 大野澄雄, "ヒンディー語音声の基本周波数パターンの生成過程のモデル化," 日本音響学会 2003 年秋季研究発表会講演論文集, Vol. 1, pp. 217-218 (2003-9).
- [10] Hiroya Fujisaki, "Prosody, information, and modeling with emphasis on tonal aspects of speech," Proceedings of the Workshop on Spoken Language Processing, Mumbai, pp. 3-12 (2003-1). (Invited plenary paper).
- [11] Hiroya Fujisaki, "Information, prosody, and modeling with emphasis on tonal features of speech," Proceedings of Speech Prosody 2004, Nara, pp. 1-10 (2004-3). (Invited plenary paper)
- [12] Hiroya Fujisaki, Sumio Ohno and Wentao Gu, "Physiological and physical mechanisms for fundamental frequency control In some tone languages and a command-response model for generation of the F_0 contour," Proceedings of International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages, Beijing, pp. 61-64 (2004-3). (Invited paper)
- [13] Shuichi Naruawa, Hiroya Fujisaki and Sumio Ohno, "A method for automatic extraction of parameters of the fundamental frequency contour," Proceedings of 2000 International Conference on Spoken Language Processing, Beijing, Vol.1, pp.649-652 (2000-10).
- [14] 成澤修一, 峯松信明, 広瀬啓吉, 藤崎博也, "音声の基本周波数パターン生成過程モデルのパラメータ自動抽出法," 情報処理学会論文誌, Vol.43, No.7, pp.2155-2168 (2002-7).
- [15] Keikichi Hirose, Yusuke Furuyama, Shuichi Narusawa, Nobuaki Minematsu and Hiroya Fujisaki, "Use of linguistic information of accent phrases for automatic extraction of F_0 contour generation model parameters," Proceedings of Oriental COCOSDA Workshop 2003, Singapore, Vol. 1, pp. 8-45 (2003-10).
- [16] Haruo Kubozono, "Temporal neutralization in Japanese," Papers in Laboratory Phonology 7, ed. by Carlos Gussenhoven and Natasha Warner, Mouton de Gruyter, pp.171-201 (2002-8).

- [17] Haruo Kubozono, "Speech timing and phonological structure," *Handbook of Japanese Phonology*, Cambridge University Press (Forthcoming).
- [18] 窪園晴夫, "鹿児島方言におけるアクセントの変化," *國文學*, Vol.48, No.4, pp. 46-52 (2003-3).
- [19] Haruo Kubozono, "Epenthetic vowels and accent in Japanese: Facts and paradoxes," *Issues in Japanese Phonology and Morphology*, ed. by Jeroen van de Weijer and T. Nishihara, Mouton de Gruyter, pp.13-142 (2001-11).
- [20] 前川喜久雄, "パラ言語的情報," 別冊国文学「現代日本語必携」, No.53, pp. 172-175 (2000-10).
- [21] Kikuo Maekawa and Takayuki Kagomiya, "Influence of paralinguistic information on segmental articulation Proceedings of ICSLP2000," Beijing, Vol. 2, pp. 349-352 (2000-10).
- [22] 前川喜久雄, 北川智利, "パラ言語情報の知覚," 日本行動計量学会第28回大会発表論文抄録集, Vol.1, pp.157-160, (2000-10).
- [23] 前川喜久雄, 北川智利, "音声はパラ言語情報をいかに伝えるか," *認知科学*, Vol. 9, No. 1, pp.46-66 (2001-3).
- [24] Masako Fujimoto and Kikuo Maekawa, "Variation of phonation types due to paralinguistic information: An analysis of high-speed video images," *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, Vol. 3, pp.2401-2404 (2003-8).
- [25] Sumio Ohno, Yoshizo Sugiyama and Hiroya Fujisaki, "Analysis and modeling of the effect of paralinguistic information upon the local speech rate," *Proceedings of 2000 International Conference on Spoken Language Processing*, Beijing, Vol.2, pp.353-356 (2000-10).
- [26] Sumio Ohno, and Hiroya Fujisaki, "Quantitative analysis of the effects of emphasis upon prosodic features of speech," *Proceedings of 7th European Conference on Speech Communication and Technology*, Aalborg, Vol. 1, pp.661-664 (2001-9).
- [27] 長島 大介, 大野 澄雄, "感情表現の韻律的特徴の分析 — 喜びと悲しみについて —," *日本音響学会2004年秋季研究発表会講演論文集*, vol. 1, pp. 273-274 (2004-9).