音声の音調的特徴のモデル化とその応用
Modeling the Tonal Features of Speech and its Applications

東京大学　名誉教授

Professor Emeritus, University of Tokyo

藤 崎　博 也

Hiroya FUJISAKI

＜研究協力者＞ Cooperating Members

東京大学大学院 情報理工学系研究科

School of Information Science and Technology, University of Tokyo

ベルリン応用科学大学 計算機科学科

Faculty of Computer Science, Berlin University of Applied Sciences

成 澤　修 一　Schuichi NARUSAWA

顧　文 涛　Wentao GU

ハンスヨルグ ミックスドルフ　Hansjörg MIXDORFF

Abstract:

This report first gives the author's original definitions of three categories of information that are expressed and transmitted by speech, as well as the author's notions on the processes by which they are reflected on the acoustic characteristics of speech. It also emphasizes the need for modeling in a quantitative study of these characteristics. It then presents a model for the generation process of the contour of the voice fundamental frequency, which is one of the most important acoustic characteristics in many languages. The model is based on the physiological and physical mechanisms of controlling the vocal fold vibration. Experimental results are presented to show the applicability of the model to speech of many languages of the world, and classifications of various languages are shown on the basis of these results.

Keywords: Modeling, Tonal features, Fundamental frequency contour, Physiological and physical mechanisms, Classification of languages

1. はじめに

ヒトの音声は、話者がそれを意識するか否かに関わらず、種々の情報を表現している。筆者はこれらの情報を (1)言語的情報、(2)パラ言語的情報、(3)非言語的情報、の3つに大別している[1]。

ここで言語的情報とは、符号としての言語により規定される離散的な情報、すなわち辞書・統語・意味・談話等のレベルで、主として文字による表記が可能なもの、あるいはその前後の文脈から、一義的に、または高々有限個の選択の可能性を残して導出しうるものをさす。たとえば単語の読み (アクセントを含む)、文の統語構造や法（mood)、談話の焦点などに関する情報がそれである。これらは、原則的に離散的な状態（の一つ）を指定する情報であって、連続的・中間的な状態を対象としない。たとえば日本語の発話の中での単語 ' アメ ' のアクセント型は、その方言において許される有限個の選択枝の中の一つ (共通日本語の場合には頭高型の「雨」か平板型の「飴」のいずれか) であって、声の高さをどのように変えても、発話自体が不完全でない限り、その中間的なものは存在しない)。

一方、音声は上記のような離散的な情報ばかりでな

く、それ以外の情報も表現することができる。たとえば、文字による表記では同じ平叙文でも、断定／疑問／勧誘／反論など、さまざまな意図を込めて発音し、その意図をかなり明瞭に相手に伝えることができる。また、丁寧／ぞんざい、改まった／くだけた、などの話者の態度の区別を表すことができる。さらに、ゆっくり／早口、大声／小声、などの話し方（スタイル）を変えることにより、発話がどのような聞き手やその置かれた状況を対象としたものかを表すこともできる。なお、これらの例では、表現の対象とする情報が、あたかも範疇的であるかように記したが、同一の範疇の中でも量的な差があり、それも音声により表現することができる、という点で、さきにのべた言語的情報とは異なる。すなわち、疑問や断定の意図、あるいは丁寧な態度は、その表出に関わる特徴を調節することによって、意図や態度の程度までも表現することができる。筆者はこの種の情報を言語的情報と区別して、パラ言語的情報と定義する。ただし、言語的情報とパラ言語的情報に共通なのは、いずれも話者が音声によって表現するべく、意識的に選択するという点である。

なお、上では、「疑問」の意図に関する情報をパラ言語的情報に属するとしたが、「疑問」に関する範疇的な情報は、たとえば終助詞「か」を平叙文の末尾に加えることにより、言語的（すなわち符号的）にも表現することが可能である。ただし、それによって表現が可能なのは、あくまで範疇的・離散的な情報のみであり、連続的・定量的な情報は棄却される。この場合に表現される情報は、「疑問」に関する「意味」の情報であり、「意図」の情報とは区別すべきものである。換言すれば、意味は言語的、意図はパラ言語的である。

音声により表現される第３の種類の情報は、たとえば話者の個人的な特徴や、年齢・性別・健康などの身体的な状態に関するもの、あるいは気質・感情などの心理的な状態に関するもので、特定の発話の言語的な内容とは関係なく存在し、また、一般には、話者が意識的に制御していないものである。もちろんこれには例外もあり、個人的な特徴・年齢や感情も、話者が意識的に模擬することは可能であって、いわゆる声帯模写や、演劇における感情の表現はそのよい例である。なお、非言語的情報にも、離散的な側面と連続的な側面とがある。たとえば喜びや悲しみに関する情報は離散的であるが、音声に現れる感情の程度は、連続的に変化しうる。

2. 音声による情報の表出過程とそのモデル化[2]

図１は前記の各種の情報が、音声の生成過程にいかに関与するかを示す概念図である。段階①では、入力情報のうち、辞書・統語・意味・談話に関する離散的情報（言語的情報）にもとづいて、広義の文法規則の制約のもとに、通報（message）が計画され、形成される(message planning)。しかしながら、この段階の出力は、具体的な発話とは直結していない。通報を具体的な発話(utterance)として計画するのは、②の発話計画（utterance planning）の段階であり、ここでは言語情報には含まれない、話者の意図や態度に関するパラ言語情報を音声に付与するべく、韻律を含む発話の計画を立てる。ちなみに「韻律」とは、「一つの発話、または一連の複数の発話に一貫性を与えるための、個々の言語単位の発話の間の関連付け」と定義する。ある発話

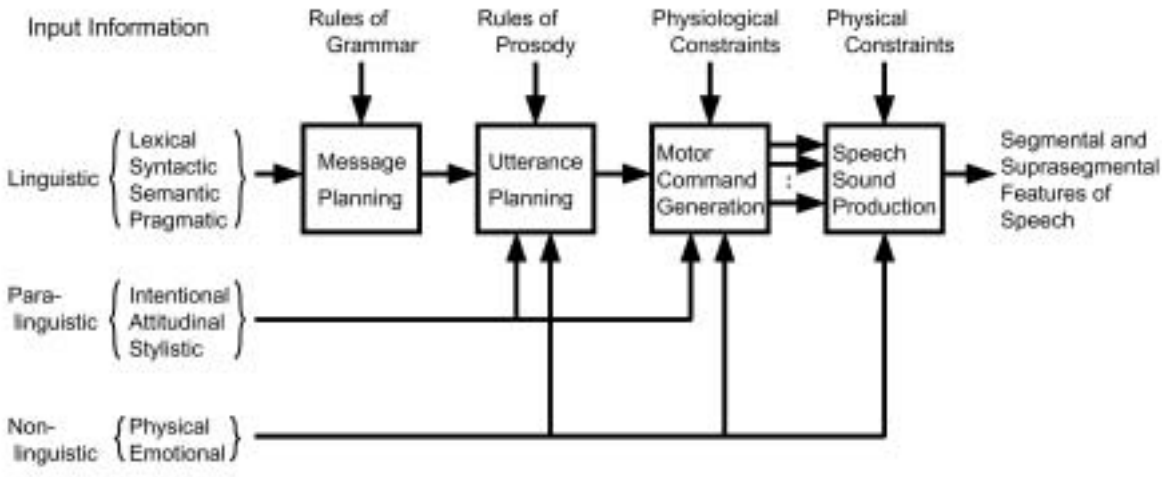

Figure 1. Processes by which various types of information are manifested in the segmental and suprasegmental features of speech

が特定の状況のもとでの発話として許容されるためには、文法規則とは別の、韻律規則とも言うべきものがある。具体的には、可能な韻律の単位と、その結合に関する制約の集合である。

発話の計画を実行に移すには、さらに各種の音声器官を制御することが必要であり、それを③の運動指令生成（motor command generation）の段階として明示した。ここから出力される運動指令には、調音器官の制御に関与し、主として分節的特徴の生成に寄与するものと、声門下の諸器官および喉頭の調節に関与し、主として超分節的特徴の生成に寄与するものとがあるが、両者の役割は完全に分離したものではなく、また、いずれも生理的な制約に従う。

これらの運動指令は、④の音声波生成（speech sound generation）の段階を制御し、その出力として各種の分節的・超分節的な特徴をもった音声が生成される。この段階が各種の物理的な制約に基づいて実行されるのはもちろんである。なお、パラ言語的情報は、②のみならず③の段階を通じて、また、非言語的情報は特に③と④の段階を通じて、生成される音声の特徴の上に反映される。

音声の音響的特徴と、それらが担う各種の情報との関係を明確に把握し、定量的に表現することは、音声言語の情報処理にきわめて重要である。しかしながら、現実に観測されるのは図１の諸過程の最終出力のみであり、そこから遡って種々の入力情報を推定することは複雑な逆問題である。韻律に関して言えば、それが複数の分節に関わる現象である以上、(1) まず、音響的特徴の時系列から、その原因となる運動指令を推定し、(2) つぎにそれらの指令から各種の情報を推定する、という２段階のアプローチが必要である。

本研究では、音声の韻律に関わる特徴のうち、特に声帯振動の基本周波数(F_0)の時間的変化のパターン(以下、F_0 パターン)を対象とし、喉頭のの制御に関する生理学的･物理学的な知見に基づいて F_0 パターンの生成過程を数学的にモデル化し、このモデルが多数の言語の音声の F_0 パターンに適用しうる一般性を持つことを実験的に確かめるとともに、それらの言語を喉頭制御に関する運動指令の見地から分類しうることを示した[3]。

3. 喉頭の制御機構の生理学的・物理学的特性と F_0 パターンの生成過程のモデル[4]

3.1 骨格筋の弾性特性

骨格筋(声帯筋を含む)の長さ方向の弾性特性は、すでに多くの研究者により実測されている[5, 6]。図２は筋肉に加えられた張力とそのスティフネスに関する、筆者の知る限りでは最初の実測結果である。この図から明らかなように、張力の広い範囲にわたって、両者の間には式(1)に示す線形の関係が成り立っている。

$$dT/dl = a + bT \tag{1}$$

ただし T は張力、l は筋肉の長さ、a は $T = 0$ における筋肉のスティフネスである。この式から、次式が導かれる。

$$T = (T_0 + a/b)\exp\{b(l - l_0)\} - a/b \tag{2}$$

ただし T_0 は張力の初期値、l_0 は筋肉の長さの初期値である。$T_0 \gg a/b$ の場合には、式 (2) は次式 (3)で近似できる。ただし x は筋肉の伸び $(l - l_0)$ である。

$$T = T_0 \exp(bx) \tag{3}$$

一方、任意の形状の弾性膜の振動の基本周波数は、次式で与えられる。

$$F_0 = c_0\sqrt{T_0/\sigma} \tag{4}$$

ただしσは膜の密度、c_0 は膜の大きさによって決まる定数である。

式(3)と(4)から、つぎの式(5)が得られる。

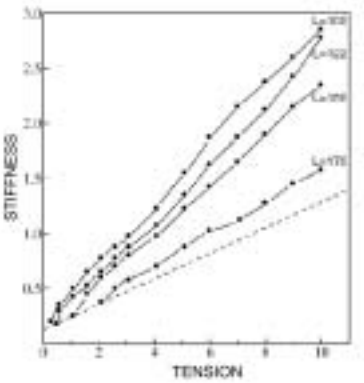

Figure 2. Stiffness of a skeletal muscle as function of tension at rest during isometric tetanic contraction initiated at different original length [5].

$$\log_e F_0 = \log_e (c_0 \sqrt{T_0 / \sigma}) + (b/2)x \qquad (5)$$

厳密には、式(5)の右辺の第１項も x に伴って僅かに変化するが、第２項による変化が圧倒的に大きい。式(5)に示すように、声帯の固有振動周波数の対数が、声帯の長さに比例して変化する成分をもつことは、種々の高さの音を持続的に発声した場合に声帯の長さの立体内視鏡を用いた観測によって確かめられているが、式(6)に示すように、声帯の長さ x が時間的に変化する場合にも成り立つ。

$$\log_e F_0(t) = \log_e F_b + (b/2)x(t) \qquad (6)$$

ここで F_b は式(4)における定数 $(c_0 \sqrt{T_0 / \sigma})$ である。

3.2 輪状甲状筋の役割

多くの言語において、声の高さの調節に最も重要な役割を果たすのは声帯の長さの制御であり、甲状軟骨と輪状軟骨の相対位置を変化させることによって行われる。いま、図３に示すように輪状軟骨を基準として考えると、甲状軟骨の運動には平行移動と回転の２つの自由度があり[7, 8]、前者は輪状甲状筋(cricothyroid muscle, 以下 CT)の斜部(pars obliqua)、後者は CT の直部(pars recta)の収縮により生ずる。これらの運動は、第１近似としては、図４に示すようにそれぞれ別の２次線形系として表現することができ、いずれも声帯の長さに微小な変化をもたらす。

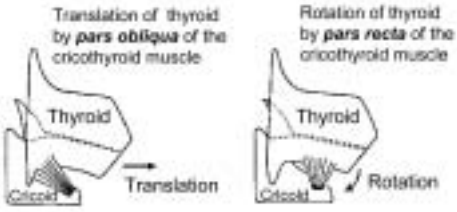

Figure 3. The roles of *pars obliqua* and pars recta of the cricothyroid muscle in translating and rotation the thyroid cartilages.

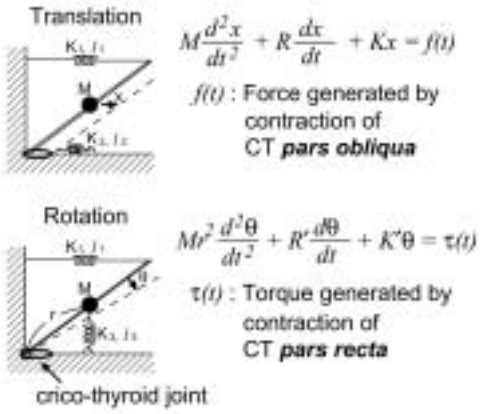

Figure 4. Equations of translation and rotation of the thyroid cartilage.

すなわち、CT の斜部が活動し短時間急激に収縮すると、甲状軟骨は一時的に微小な平行移動を起こし、その結果、声帯の長さに微小な変化 $x_1(t)$を生ずる。この場合の筋の収縮速度が系の応答速度と比較して十分に大きく、かつその持続時間が十分に短ければ、その運動は近似的にインパルス応答となり、$x_1(t)$も同じくインパルス応答として表現される。一方、CT の直部が活動し持続的に収縮すれば、甲状軟骨は輪状甲状関節(crico-thyroid joint)を軸として微小に回転し前屈する結果、その回転角に比例して声帯の長さに微小な変化 $x_2(t)$を生ずる。この場合の筋の収縮速度が系の応答速度と比較して十分に大きければ、その運動は近似的にステップ応答となり、$x_2(t)$も同じくステップ応答となる。これらの二つの運動が微小で互いに独立とみなしうる範囲では、声帯の長さの変化は $x_1(t)$と $x_2(t)$の和であり、その結果として観測される基本周波数の時間的変化は

$$\log_e F_0(t) = \log_e F_b + (b/2)[x_1(t) + x_2(t)] \qquad (7)$$

となる。すなわち、$\log_e F_0(t)$ は話者の声帯の物理的性質によってきまる固定項 $\log_e F_b$ と、2つの時間的変化成分 $x_1(t)$ と $x_2(t)$ の和として表される。ここで $x_1(t)$ 、$x_2(t)$ はいずれも正の値を持つ。なお、本稿では、以下、F_0 パターンとは $\log_e F_0(t)$ をさすものとする。なお、甲状軟骨の平行移動の時定数は、回転の時定数よりもはるかに大きいため、多くの言語に共通する特徴として $x_1(t)$ が句単位の比較的緩やかな音調の表現に、$x_2(t)$ が語または音節単位の比較的急激で局所的な音調の表現に用いられているのは興味深いことである。

3.3 正極性の局所成分をもつ F_0 パターンの生成過程のモデル[9, 10]

以上に述べた輪状甲状筋 CT による喉頭制御と、それに基づく F_0 パターン生成の生理的･物理的過程をやや簡略化し、理想化したのが図５のモデルである。ここでは CT の斜部の瞬間的な活動を理想化し、インパルス関数で表してフレーズ指令と名づけ、甲状軟骨の平行移動により F_0 パターンに変化を生ずる過程を臨

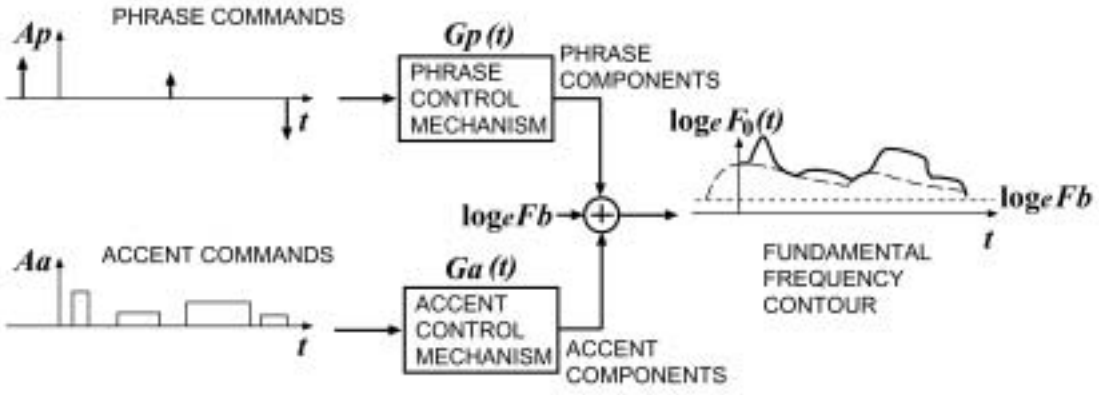

Figure 5. A functional model for the processes of generating the fundamental frequency contour of speech from phrase commands as impulses and accent commands as positive pedestal functions.

界制動の２次線形系で表現してフレーズ制御機構とよび、その出力をフレーズ成分とよぶ。また、CT の直部の持続的な活動を理想化し、ステップ関数で表してアクセント指令と名づけ、甲状軟骨の回転により F_0 パターンに変化を生ずる過程を臨界制動の２次線形系で表現してアクセント制御機構とよび、その出力をアクセント成分とよぶ。モデルの最終的な出力としての F_0 パターンはこれらの２種類の成分と固定項 $\log_e F_b$ との和として以下の式により表される。

$$\log_e F_0(t) = \log_e F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj} \{G_a(t - T_{1j}) - G_a(t - T_{2j})\} \quad (8)$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (9)$$

$$G_a(t) = \begin{cases} \min[1 - (1+\beta t)\exp(-\beta t), \gamma], & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (10)$$

ここで $G_p(t)$ はフレーズ制御機構のインパルス応答、$G_a(t)$ はアクセント制御機構のステップ応答であり、また

F_b：基本周波数の基底値
I ：発話中のフレーズ指令の数
J ：発話中のアクセント指令の数
A_{pi}: 第 i 番目のフレーズ指令(インパルス)の大きさ
A_{aj}：第 j 番目のアクセント指令(ステップ)の振幅
T_{0i}：第 i 番目のフレーズ指令の生起時点
T_{1j}：第 j 番目のアクセント指令の始端の時点
T_{2j}：第 j 番目のアクセント指令の終端の時点
α：フレーズ制御機構の固有角周波数
β：アクセント制御機構の固有角周波数
γ：アクセント成分の相対飽和レベル

なお、フレーズ制御機構とアクセント制御機構とは、厳密に臨界制動系であるという確証はないが、予備的な実験的検討の結果、臨界制動の仮定が近似的に成り立つことを確かめており、パラメータの数を少なくできることから、どちらも臨界制動系としている。また、αとβの値はそれぞれ図４の二つの力学系の定数によって決まるもので、話者によって異なり、また、一人の話者においても喉頭の制御の仕方により多少変化する可能性があるが、多数の言語の話者の音声を分析した結果、話者ごとに一定値を持つと仮定しても差支えないこと、さらに話者の個人差も、言語による差も比較的小さいことを確かめたため、α＝3/s、β＝20/ s の値を用いる。

さらに、臨界制動2次系のステップ応答は時間の単調増大関数で、原理的にはｔ＝∞ で１に漸近するが、実測の F_0 パターンでは、有限の時間内に一定値に到達するとみなせる場合が多い。これは、甲状軟骨の回転角に閾値が存在すると仮定することにより、近似的に表現することができる。この研究では、簡単のためにγ＝0.9 としたが、現実には、γの値は発話により、また話者によっても若干変化する。

α、β、γを定数として、フレーズ指令とアクセント指令を指定すれば、F_0 パターンは式(8)により一義的に算出される。このモデルが現実の F_0 パターンをどの程度まで近似しうるかを確かめるには、Analysis-by-Synthesis の手法を用いることができる。図

６は、共通日本語話者による「青い葵の絵は山の上の家にある」の F_0 パターンにこの手法を適用して、対数基本周波数尺度で自乗誤差を最小とするモデルによる最良近似を行った例であり、上から音声波形、F_0 パターンの実測値(+ 印)、式(8)による最良近似(実線)、その際に推定されたフレーズ成分(破線)とフレーズ指令、アクセント成分とアクセント指令、の順に示す。

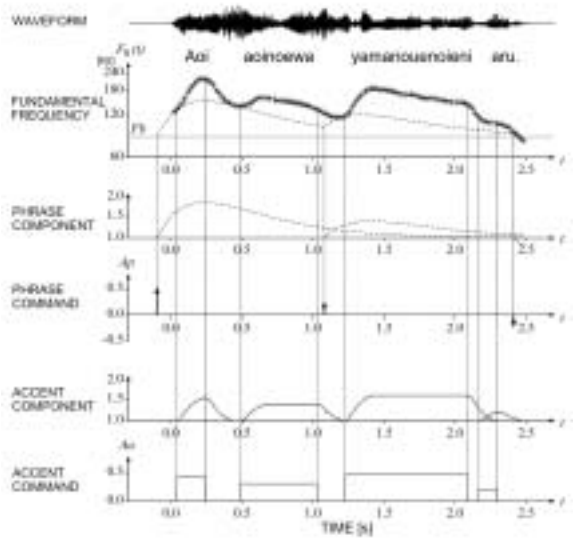

Figure 6. Analysis-by-Synthesis of an F_0 contour of the Japanese utterance: /aoi aoinoewa yamanouenoieniaru/.

この例のみならず、共通日本語の多数の音声資料の F_0 パターンに対して、上記のモデルは極めてよい近似を与えることが確かめられている。また、Analysis-by-Synthesis の手法を用いれば、所与の F_0 パターンをフレーズ成分とアクセント成分に一義的に分解することができ、それらのもとになるフレーズ指令とアクセント指令とは、発話の韻律に含まれる言語的およびパラ言語的情報と密接に対応することが示されている。

3.4 F_0 パターンの負の局所成分の生成における外喉頭筋の役割

一方、たとえばスウェーデン語のように、acute と grave の２種類のアクセントをもつ言語での grave accent は、F_0 の能動的な下げを伴うもので、acute accent とは反対に、負極性のアクセント指令によると考えられる。また、標準中国語などの声調言語においては、通常１つの音節が複数の声調を伴って発音される。それらの声調のあるものは正の局所成分をもつが、他のものは負の局所成分をもつと考えられる[11]。これらの言語の F_0 パターンにおける負の局所成分の生成には、ＣＴ以外の筋が関与するものと予想されていたが、その機構は明らかにされていなかった。以下はこの点に関する筆者の検討結果である。

ある種の言語の音声における F_0 の積極的な「下げ」と、胸骨舌骨筋 (sternohyoid muscle, 以下 SH) との間の関係に関しては、すでに若干の筋電図学的研究の報告がある[12, 13]。声の下げの機構に関しては、その他にも種々の仮説が提起されている。しかしながら、声の下げには、声帯の一端と直結している甲状軟骨の運動が必要であり、その運度にはやはり甲状軟骨に直結した別の筋肉の関与が必要であるにも関わらず、それらの仮説の中で、この点に着目したものはなかった。

一方、筆者はタイ語の声調に関する筋電図学的研究の報告[14, 15]を詳細に検討した結果、甲状舌骨筋 (throhyoid muscle, 以下 TH) の活動が常に ST と同時に記録されており、しかもその強さが SH をはるかに超えるものであることに気づいた。図７に示すように、TH は SH とは異なって甲状軟骨に直結した筋肉である。すでに指摘されているように、SH が声の能動的な下げに寄与することは事実であるが、それが声帯の長さ、ひいてはその張力を減少させることによって、最終的に声の下げを生ずるには、図８に示すように複雑な機構が必要である。すなわち、SH が緊張するこ

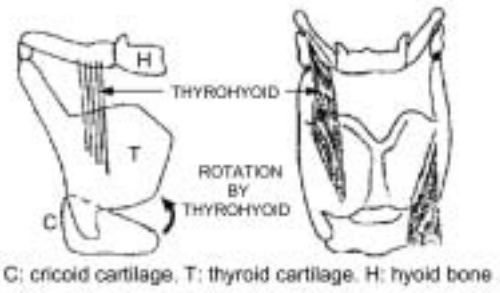

Figure 7. Role of the thyrohyoid muscle in laryngeal control.

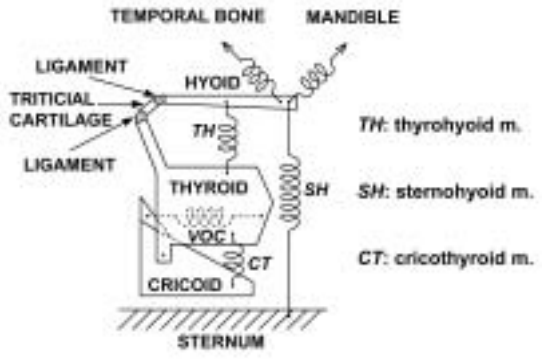

Figure 8. Mechanism of F_0 lowering by the activities of thyrohyoid and sternohyoid muscles.

とによって胸骨の位置が固定され、同時に TH の緊張が CT(の直部)の緊張とは逆方向のトルクを発生し、甲状張力を減少させ、最終的に声帯振動の基本周波数を下げる。なお、この際、甲状軟骨が逆方向に回転しうるのは、その上端と舌骨の間の結合が直接的ではなく、麦粒軟骨(triticial cartilage) を介する間接的なものであるためである[16, 17]。

3.5 正・負両極性の局所成分をもつ F_0 パターンの生成過程のモデル

上記のように、F_0 パターンの能動的な下降は、能動的な上昇とは別の機構を介して行われるのため、両者の時間的変化の速度は同一ではなく、従って厳密には両者を分離した定式化が必要である。すなわち、この場合の F_0 パターンの生成モデルは、式(11)、(12)、(13)により表される。

$$\log_e F_0(t) = \log_e F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{tj} \{G_t(t - T_{1j}) - G_t(t - T_{2j})\} \tag{11}$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{12}$$

$$G_t(t) = \begin{cases} \min[1 - (1+\beta_1 t)\exp(-\beta_1 t), \gamma_1], & t \geq 0 \\ 0, & t < 0 \end{cases}$$

（局所指令が正の場合）

$$G_t(t) = \begin{cases} \min[1 - (1+\beta_2 t)\exp(-\beta_2 t), \gamma_2], & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{13}$$

（局所指令が負の場合）

しかしながら、簡単のため、正・負の極性の局所指令に対する応答を同一の関数によって近似すれば、この場合のモデルは、局所指令が正の場合と同一となる。

3.6 種々の言語の音声の F_0 パターンに対するモデルの適用性の検証

(1) 正の局所成分のみを持つ言語

表１は報告者がこれまでにモデルの適用性を実験により確かめた言語の一覧表である。スペースの制約上、それぞれの言語に関する実験の詳細は省略するが、英語・ドイツ語・ギリシャ語・韓国語・ポーランド語・スペイン語（いずれも非声調言語）の各1文ずつの F_0 パターンのモデルによる近似の結果と、その際に得られたフレーズ指令およびアクセント指令を、これらの文とその英訳とともに図９(a)-(f) に示す。この図に示したのは僅かな例に過ぎないが、それぞれの言語についてはるかに多くの音声資料の分析を行っており、その結果は、上記のどの言語に関しても、共通日本語と同様、アクセント指令の極性を正のみに限定したモデルによって、極めてよい近似が得られ、従ってこれらの言語の通常の発話では、局所的な指令の極性を正に限定して差支えがないことが確かめられた。

Table 1. Languages for which the model has been tested by the author and his colleagues.

Language	Researchers	References
Japanese	Fujisaki et al.	[9, 10]
Cantonese	Gu, Fujisaki et al.	[30, 31]
English	Fujisaki et al.	[18]
Estonian	Fujisaki & Lehiste	[19]
German	Mixdorff & Fujisaki	[20]
Greek	Fujisaki , Ohno et al.	[21]
Hindi	Fujisaki & Ohno	[26]
Korean	Fujisaki & Ohno	[22]
Mandarin	Fujisaki et al.	[11, 27]
Polish	Fujisaki et al.	
Portuguese	Fujisaki & Narusawa	[25]
Shanghainese	Gu, Hirose & Fujisaki	[32]
Spanish	Fujisaki, Ohno et al.	[23]
Swedish	Fujisaki & Lijungqvist	[24]
Thai	Fujisaki, Ohno et al..	[28]
Vietnamese	Mixdorff & Fujisaki	[29]

(2) 正.・負両極性の局所成分を持つ言語

一方、スウェーデン語・ポルトガル語・ヒンディー語などの非声調言語、および標準中国語・タイ語・ベトナム語・広東語などの声調言語では、局所成分が正・負の両極性をもつことを確認した。図10 (a)-(c) は上記の非声調言語の、また図11 (a)-(d) は上記の声調言語の、各1文ずつの F_0 パターンのモデルによる近似の結果を示す。いずれの言語においても、モデルによる近似が実測の F0 パターンに極めて近く、両極性の局所的指令を用いるモデルが妥当なことが示されている。

なお、図10に示した３言語のうち、語のアクセントが本質的に負の極性を要求するものはスウェーデン語のみであり、ポルトガル語およびヒンディー語では、負のアクセント指令は辞書的情報ではなく、発話ないし句の冒頭に多く用いられ、強調などに関するパラ言語的情報の伝達に関与している。

一方、図11に示した声調言語では、いずれも１音節に１個または２個の声調指令があり、それぞれ正または負の極性をもつが、声調の中には、局所指令を欠く

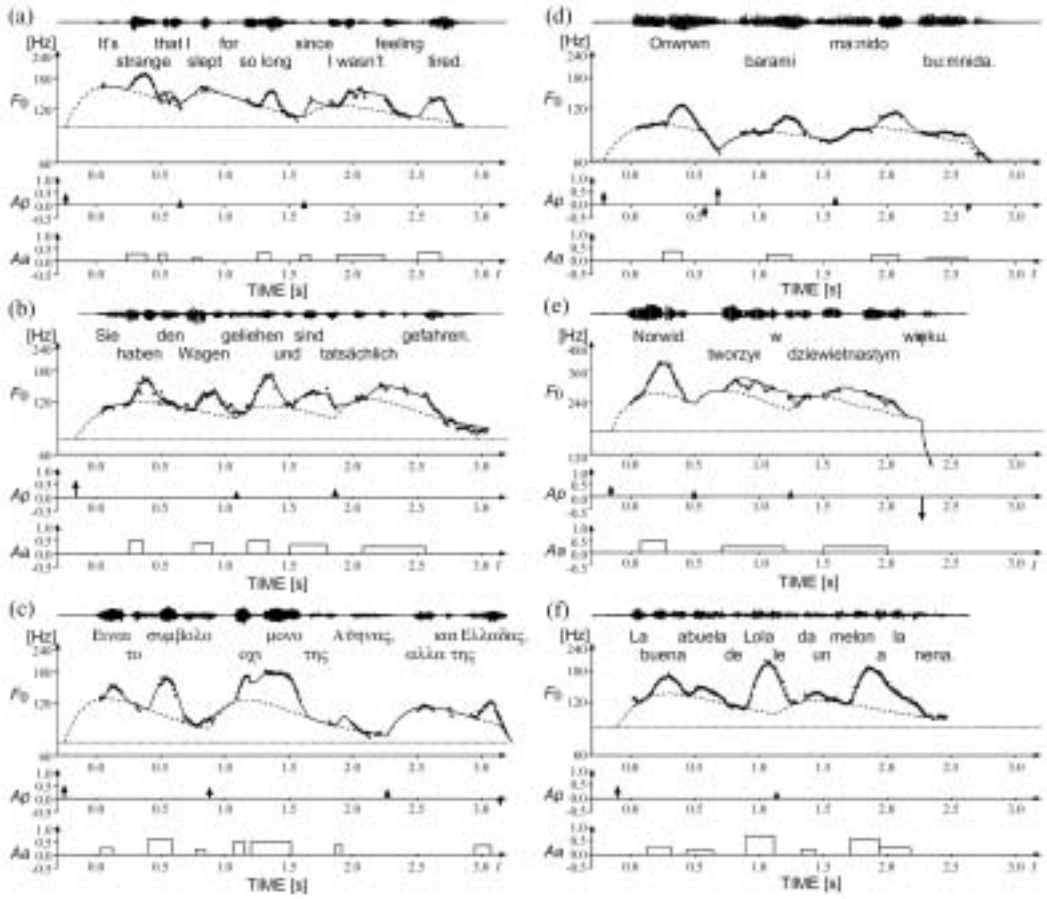

Figure 9. Examples of Analysis-by-Synthesis of F_0 contours of languages with positive accent commands. (a) English [18], (b) German [20], (c) Greek [21], (d) Korean [22], (e) Polish, and (f) Spanish [23].

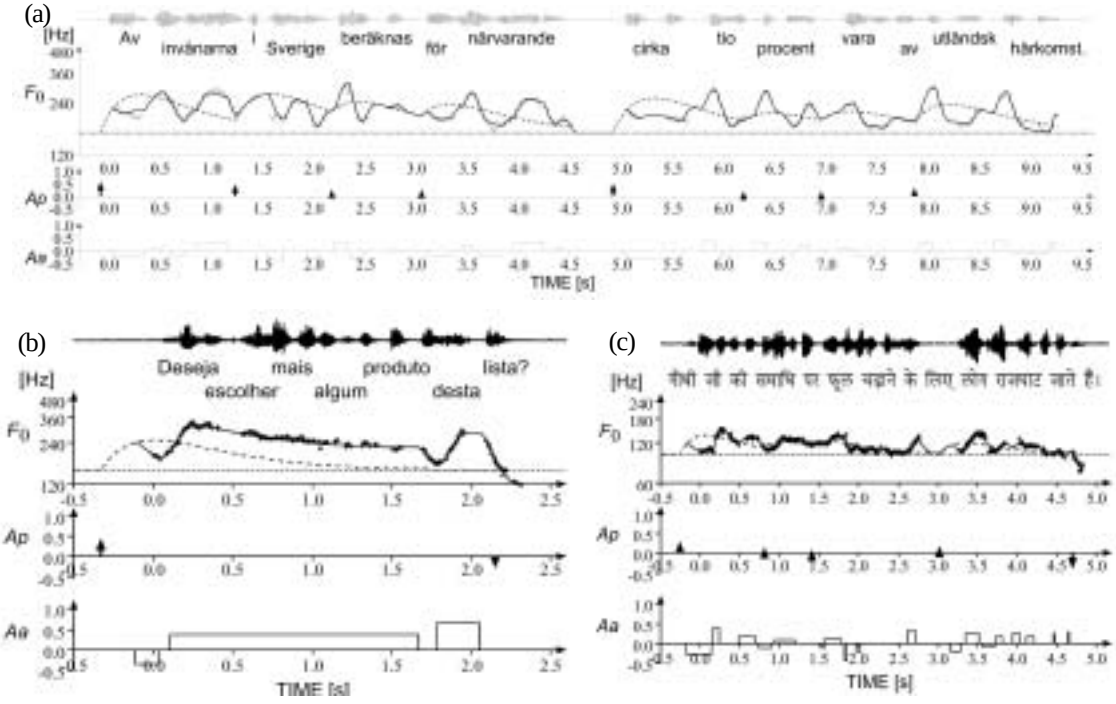

Figure 10. Examples of Analysis-by-Synthesis of F_0 contours of languages with positive and negative accent commands. (a) Swedish [24], (b) Portuguese [25], and (c) Hindi [26].

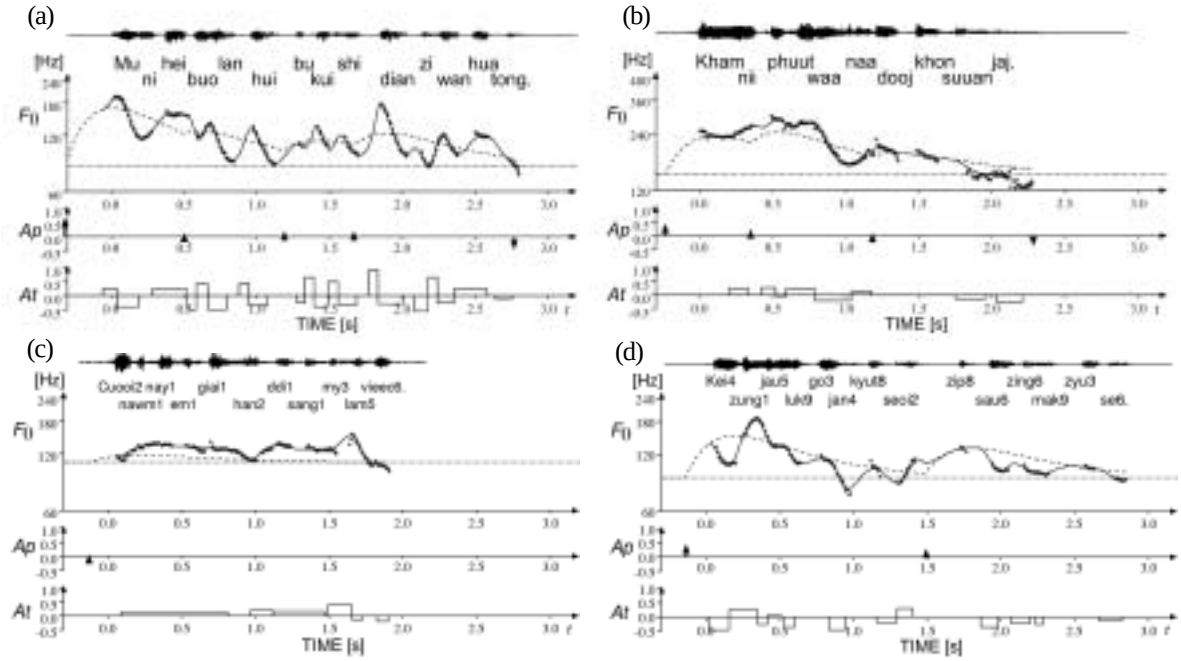

Figure 11. Examples of Analysis-by-Synthesis of F_0 contours of tone languages with positive and negative tone commands. (a) Mandarin [27], (b) Thai [28], (c) Vietnamese [29], and (d) Cantonese [30, 31].

もの、また、音節の前半または後半のみに局所指令をもつものなどがある。いま、この事態をやや簡略化し、音節の前半部分での局所指令の極性を横座標x、後半部分での局所指令の極性を縦座標yによって示し、声調指令のない場合をゼロに対応させると、上記の４つの声調言語における単音節語の声調型の布置は図12に示す通りであり、これらの諸言語の声調型の音韻論的な特徴と対比を明確に表現している。なお、高品質の音声合成のためにはこのような定性的記述だけでは不十分で、局所指令の大きさと音節内での生起時点に関するより詳細な情報が必要であることは勿論である。この点に関しては、標準中国語を対象として詳細な実験的検討を行い、成果を得ているがここでは省略する。

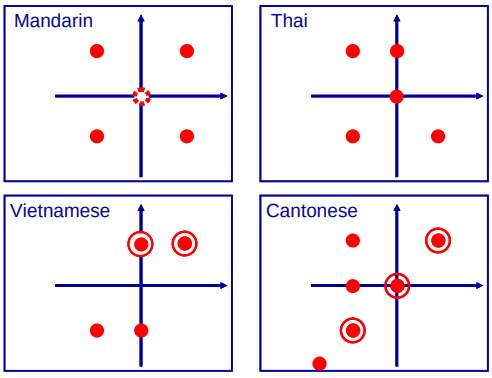

Figure 12. Patterns of tone command polarities for tone languages. Abscissa: polarity of commands in the early part of the syllable. Ordinate: polarities of commands in the latter part of the syllable.

表２は、報告者がこれまで分析の対象としたすべての言語を、局所指令の極性の見地から分類したものである。なお、通常の発話では正極性のみを用いる言語でも、状況により、また話者により、負の極性の局所指令を用いることがあり、特に British English の話者に比較的多く、皮肉・非難などのパラ言語的情報の表現に用いられること、また、話者によりかなりの差があることも確かめている [18]。

Table 2. Classification of languages according to the polarity of local commands.

Polarity	Languages
Positive only	English, Estonian, German, Greek, Korean, Polish, Spanish, …
Positive and negative	Hindi, Portuguese, Swedish Cantonese*, Mandarin*, Thai*, Shanghainese*, Vietnamese*

* Tone languages

4. おわりに

本報告では、まず音声によって表される情報を、言語的・パラ言語的・非言語的の３種類に分類し、かつそれらが音声の生成過程の主としてどの段階に関与するかを考察した。また、音声の音響的特徴からこれらの情報を抽出する上での、生成過程のモデル化の必要性を説き、多くの言語において韻律の表現に重要な役割を占める、音声の基本周波数パターンを取り上げ、その生成の生理学的・物理学的機構と、それに立脚し

た定量的なモデルとを示し、それが共通日本語をはじめ、多数の言語の音声に適用可能であること、また、このモデルに即して分析を行った結果、諸言語を音調的な特徴から２つの型に分けられること、特に声調言語に関しては、声調指令の極性に着目した音韻論的分類が可能であることを明らかにした。

文　献 (番号の**太字**は本研究の直接の成果を示す)

[1] 藤崎博也: “韻律研究の諸側面とその課題,” 日本音響学会平成6年秋季研究発表会講演論文集, 1, 287-288 (1994).

[2] Fujisaki, H.: “Prosody, models, and spontaneous speech,” In *Computing Prosody*, (Y. Sagisaka, N. Campbell and N. Higuchi, eds.) Springer-Verlag, New York, 27-42 (1996).

[3] Fujisaki, H.,: “Information, prosody, and modeling ― with emphasis on tonal features of speech,” *Proceedings of Speech Prosody 2004*, １-10 (2004).

[4] Fujisaki, H.: “A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour,” In *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O. Fujimura, ed.) Raven Press, 347-355 (1988).

[5] Buchthal, F. and Kaiser, E.: “Factors determining tension development in skeletal muscles,” *Acta Physiol. Scand.*, 8, 38-74 (1944).

[6] Sandow, W.: “A theory of active state mechanisms in isometric muscular contraction,” *Science*, 127, 760-762 (1958).

[7] Fink, B. R. and Demarest, R. J.: *Laryngeal Biomechanics*, Cambridge, Mass. (1978).

[8] Zemlin, W. R.: *Speech and Hearing Science, Anatomy and Physiology*, Prentice Hall, 1968.

[9] Fujisaki, H. and Nagashima, S.: “A model for the synthesis of pitch contours of connected speech,” *Annual Report, Engineering Research Inst., University of Tokyo*, 28, 53-60 (1994).

[10] Fujisaki, H, and Hirose, K.: “Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation,” *Preprints of Papers, Working Group on Intonation* (H. Fujisaki and E. Gårding, eds.) the XIIIth International Congress of Linguists, Tokyo, 57-70 (1982).

[11] Fujisaki, H., Hallé, P. and Lei, H.: “Application of F_0 contour command-response model to Chinese tones,” 日本音響学会昭和62年度秋季研究発表会講演論文集, 1, 197-198 (1987)

[12] Gårding, E.: “Word tones and larynx muscles,” *Working Papers, Dept. of Linguistics, Lund University*, 3, 20-46 (1970).

[13] Sagart, L., Hallé, P. et al.: “Tone production in modern Standard Chinese: an electromyographic investigation,” *Cahiers de Linguistique Asie Orientale*, 15, 205-211 (1986).

[14] Erickson, D. *A physiological analysis of the tones of Thai*, Ph.D. Dissertation, University of Connecticut (1976).

[15] Erickson, D. “Laryngeal muscle activity in connection with Thai tones,” *Annual Bulletin, RILP, University of Tokyo*, 135-149 (1993).

[16] Fujisaki, H., Tomana, R., Narusawa, S., Ohno, S. and Wang, C.: “Physiological mechanisms for fundamental frequency control in Standard Chinese,” *Proc. ICSLP 2000*, 1, 9-12 (2000).

[17] Fujisaki, H., Ohno, S. and Gu, W.: “Physiological and physical mechanisms for fundamental frequency control in some tone languages and a command-response model for generation of their F_0 contours,” *Proc. Int’l Symposium on Tonal Aspects of Languages*, 61-64 (2004).

[18] Fujisaki, H. and Ohno, S.: “Analysis and modeling of fundamental frequency contours of English utterances,” *Proc. EUROSPEECH ’95*, 2, 985-988 (1995).

[19] Fujisaki, H. and Lehiste, I.: “Some temporal and tonal characteristics of declarative sentences in Estonian,” *Preprints of Papers, Working Group on Intonation* (H. Fujisaki and E. Gårding, eds.) the XIIIth International Congress of Linguists, Tokyo, 121-130 (1982).

[20] Mixdorff, H. and Fujisaki, H.: “Analysis of voice fundamental frequency contours of German utterances using a quantitative model,” *Proc. ISCLP ’94*, 4, 2231-2234 (1994).

[21] Fujisaki, H., Ohno, S. and Yagi, T.: “Analysis and modeling of fundamental frequency contours of Greek utterances,” *Proc. EUROSPEECH ’97*, 1, 465-468 (1997).

[22] Fujisaki, H.: “Analysis and modeling of fundamental frequency contours of Korean utterances ― A preliminary study,” *Phonetics and Linguistics ― In Honour of Prof. H. B. Lee*, Seoul, 640-657 (1996).

[23] Fujisaki, H., Ohno, S., Nakamura, K., Guirao, M. and Gurlekian, J.; “Analysis of accent and intonation in Spanish based on the command-response model,” *Proc. ICSLP ’94*, 1, 355-358 (1994).

[24] Fujisaki, H. and Ljungqvist, M. and Murata, H.: “Analysis and modeling of word accent and sentence intonation in Swedish,” *Proc. ICASSP ’93*, 1, 211-214 (1993).

[25] Fujisaki, H., Narusawa, S., Ohno, S. and Freitas, D.; “Analysis and modeling of F_0 contours of Portuguese utterances based on the command-response model,” *Proc. EUROSPEECH ’03, 3,* 2317-2320 (2003).

[26] Fujisaki, H. and Ohno, S.: “Modeling the generation process of fundamental frequency contours of Hindi utterances,” 日本音響学会2003年秋季研究発表会講演論文集1, 217-218 (2003).

[27] Wang, C., Fujisaki, H., Tomana, R. and Ohno, S.: “Analysis of fundamental frequency contours of Standard Chinese in terms of a command-response model and its application to synthesis by rule of intonation,” *Proc. ICSLP 2000*, 3, 326-329 (2000).

[28] Fujisaki, H., Ohno, S. and Luksaneeyanawin, S., Analysis and synthesis of F_0 contours of Thai utterances based on the command-response model,” *Proc. 15th ICPhS* 1129-1132 (2003).

[29] Mixdorff, H., Nguyen, H. B., Fujisaki, H. and Luong, M. C.: “Quantitative analysis and synthesis of syllabic tones in Vietnamese,” *Proc. EUROSPEECH ’03*, 1, 177-180 (2003).

[30] Fujisaki, H., Gu, W. and Hirose, K.; “The command-response model for the generation of F_0 contours of Cantonese utterances,” *Proceedings of the 7th Int’l Conference on Signal Processing, Beijing*, 1, 655-658 (2004).

[31] 顧文涛, 広瀬啓吉, 藤崎博也, “広東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, 1, 395-396 (2004-9).

[32] 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, 1, 393-394 (2004-9).