

韻律的特徴の総合的なモデル化と、感情の表現・伝達過程

Integrated Modeling of Prosodic Features and Processes of Emotional Expressions

東京工科大学コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology

大野澄雄
Sumio OHNO

< 研究協力者 >
東京工科大学大学院工学研究科
Graduate School of Systems Engineering, Tokyo University of Technology

長島大介
Daisuke NAGASHIMA

It is well known that fundamental frequency, speech rate, and source intensity are three major quantities that represent prosody, and their temporal variations (contours) carry not only linguistic information, but also para- and non-linguistic information. Except for the fundamental frequency, however, the processes of generating their contours have not been formulated by quantitative models. This report firstly presents models for the processes of generating contours of the relative speech rate and the relative source intensity from their underlying commands. Together with the model that has been already established for the fundamental frequency contour, they represent an integrated approach for the analysis and synthesis of prosodic features. The usefulness of the approach is then illustrated by an analysis of the effects of emphasis upon the three prosodic features in utterances of English by native and nonnative speakers. Finally, the results of a preliminary study are presented on the prosodic features in the expression of various emotional states of the speaker.

Keywords: Integrated modeling, Prosodic features, Speech rate, Source intensity, Generating process, Effects of emphasis, Expression of emotional states

1 はじめに

音声の韻律的特徴は、音声の言語情報のみならず、パラ言語情報および非言語情報を表現する重要な特徴である [1]。本研究では、音声によって表現されるこれらの情報が、韻律的特徴として現れる様相を客観的・定量的な分析手法を用いてモデル化することにより、これを高品質で自然性の高い音声の合成や、自然に発話された音声からそれらの情報の抽出に応用することを目標としている。

音声の韻律的特徴のうち、特にその基本周波数パターン (以下、 F_0 パターンと呼ぶ) に関しては、既に藤崎らが提案した生成過程に基づく指令 応答モデルを用いることにより、音声の言語情報に対応して生起する少数の指令群でその特徴を正確に表現することができ、また指令の大きさの連続的な変化により種々の音声表現をとらえることが可能である。ここでは、発話速度および音源強度について、 F_0 パターンのモデルと類似したアプローチをとり、それ

らの変化の定式化・モデル化を試みた。

また、非言語情報、特に発話に込められた感情表現に関する情報が韻律的特徴としてどのように現れるかについて、モデルへの入力指令の変化として明らかにした。

以下、本報告では、 F_0 パターン、発話速度、および音源強度の 3 つの韻律的特徴を統一的に扱うためのモデルに関する成果を概説し、これらの 3 つの特徴が特に音声の強調表現においてそのような相互関係を有するののかについて検討した結果を述べる。また、感情表現と韻律的特徴の関係に関して、感情の分類、音声による感情表現の伝達特性について検討し、さらに F_0 パターンについてモデルへの入力指令との関係について検討した結果を述べる。

2 各種韻律的特徴の定量化のためのモデル

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) の時間的变化パターンとして観測できる。これらの時間的变化は、複雑な様相を呈するが、種々の離散的あるいは連続的な心理的・生理的・物理的要因に基づいて生成されるものである。生成過程をモデル化することにより、それら要因の影響を定式化し、根源的な伝達情報に基づく制御指令からそれぞれのパターンが生成するものと考えることができれば、種々の表現の把握を容易にする特徴の定量化が可能となる。

特に音声の F_0 パターンに関しては、既に藤崎らにより生成過程に基づく指令 応答モデルが提案されている [2]。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、共通日本語をはじめとして諸言語に対しても多く実証されている [3]。

一方、音声の発話速度や音源強度については、その局所的な振舞いは、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がかりがあるものと考えられる。従って、これら2つの韻律的特徴については、相対的な変化量を定量化し、モデル化することを試みた。

以下は、そのためにまず、比較的明確に表現でき、また韻律的特徴が重要な役割を担っていると考えられる音声の強調表現を対象として、モデル化を試みた結果について述べる。

2.1 基本周波数パターン生成モデル [2]

上述の F_0 パターン生成過程のモデルは、図1に示すように対数 F_0 パターンが、句頭から句末に向

かって上昇とその後の緩やかな下降を示すフレーズ成分と、語のアクセントに対応する局所的な起伏を示すアクセント成分、および発話単位中で、ほぼ一定値をとるベースライン成分の総和として表現できるとするものである。さらに2つの変化成分であるフレーズ成分とアクセント成分はそれぞれ、インパルス状のフレーズ指令、ステップ状のアクセント指令に対する臨界制動2次線形系の応答として表現できる。これら2つの指令とその応答の関係は、喉頭制御の生理的・物理的根拠に基づくものであり、このモデルの妥当性は日本語をはじめ多くの諸言語の F_0 パターンの分析により実証されている [4]。このモデルは次式で表すことができる。

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^I A_{P_i} G_P(t - T_{0i}) + \sum_{j=1}^J A_{A_j} \{G_A(t - T_{1j}) - G_A(t - T_{2j})\}, \quad (1)$$

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

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

ここで $G_P(t)$ は、フレーズ制御機構のインパルス応答を、また $G_A(t)$ は、アクセント制御機構のステップ応答を表している。また、 Fb は基本周波数の基底値、 A_{P_i} と T_{0i} はそれぞれ i 番目のフレーズ指令の大きさと生起時点、 A_{A_j} 、 T_{1j} と T_{2j} はそれぞれ j 番目のアクセント指令の大きさ、開始時点と終了時点である。 α と 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であり、話者による変化は少なく、多くの対象についてそれぞれ3および20程度になることが分かっている。

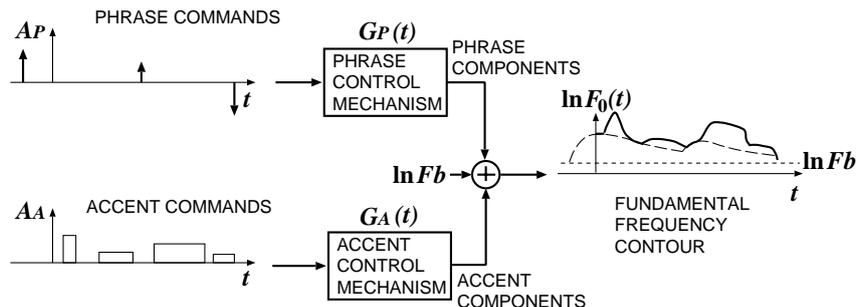


図1: F_0 パターンの生成過程モデル

2.2 発話速度変化パターン生成過程モデル

筆者は藤崎とともに、言語情報が同一の2つの異なる発話に対して、一方の発話速度の局所的な振舞いを、他方(通常、標準的・中立的な発話)のそれを基準とした相対的な局所発話速度として定量化する手法を提案した[4]。この手法は、2つの発話の分節的特徴の系列間の時間対応を非線形時間軸正規化により求め、時間軸正規化関数の導関数を局所的発話速度比の時間関数 $R(t)$ として表すものである。具体的には、DTWにより時間離散的に得られる時間軸正規化関数を平滑化することによりその導関数を近似的に求める。ここでは $R(t)$ の対数値を「発話速度変化パターン (R contour)」と呼ぶ。発話速度変化パターンが何らかの加速・減速制御に基づいて生成されるものとし、これらの発話速度変化指令を正負のステップ関数、また制御機構を臨界2次線形系であると仮定すると、 F_0 パターン生成過程モデル同様に次式のように発話速度変化パターンを定式化することができる(図2)。

$$\ln R(t) = \sum_{k=1}^K A_{Rk} \{G_R(t - T_{1k}) - G_R(t - T_{2k})\}, \quad (4)$$

$$G_R(t) = \begin{cases} 1 - (1 + \kappa t) \exp(-\kappa t), & t \geq 0, \\ 0, & t < 0. \end{cases} \quad (5)$$

ここで A_{Rk} は k 番目の発話速度変化指令の大きさ、 T_{1k} と T_{2k} は、それぞれ k 番目の発話速度変化指令の開始時点と終了時点を表す。

2.3 音源強度変化パターン生成過程モデル [5]

音源強度に関して、発話速度同様に相対的な意味付けを行うこととする。ここでは、2つの発話間の分

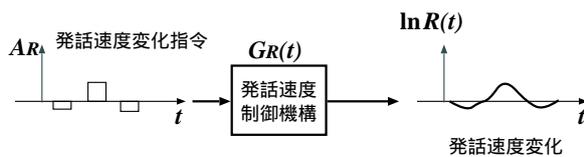


図 2: 発話速度変化パターンの生成過程モデル

節的に対応する時点間の一方の発話(通常、標準的・中立的な発話を基準とする)に対する音源強度の比の時間変化を「音源強度変化パターン (I contour)」と呼び、この時間関数 $I(t)$ を上述の発話速度変化パターンと同様の仮定に基づいて次のように定式化する(図3)。

$$\ln I(t) = \sum_{l=1}^L A_{Il} \{G_I(t - T_{1l}) - G_I(t - T_{2l})\}, \quad (6)$$

$$G_I(t) = \begin{cases} 1 - (1 + \lambda t) \exp(-\lambda t), & t \geq 0, \\ 0, & t < 0. \end{cases} \quad (7)$$

ここで、 A_{Il} は l 番目の音源強度変化指令、 T_{1l} と T_{2l} は、それぞれ l 番目の音源強度変化指令の開始時点と終了時点である。

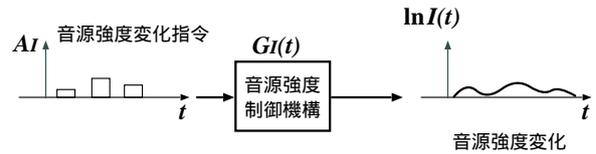


図 3: 音源強度変化パターンの生成過程モデル

2.4 生成過程モデルに基づく分析 [6]

2.4.1 音声資料・分析条件・分析例

音声資料として、英文 “We are nine very young men.” を種々の強調パターンで読み上げたものを用いる。強調の有無に関して、(1) 特に強調を行わない中立的な発声と、(2) 6単語のうち1つを強調した発声、を行った。また強調には、低・中・高の3段階の強調を行った。なお、話者は2名で、一方(話者A)は英語の母国語話者、他方(話者B)は日本語を母国語とする話者で、それぞれ1つの強調パターンについて5回ずつ発声した。

音声は 10 kHz, 16 bit でデジタル化し、10 ms 間隔で基本周波数を抽出し、 F_0 パターン生成過程モデルに基づいて、パラメータを A-b-S 法により求めた。また分節的特徴量として 12 次のケプストラム係数を 10 ms 間隔で求め、DTW による時間軸正規化を行い、それをもとに発話速度変化パターンと音源強度変化パターンを求めて、それぞれのモデルのパラメータを A-b-S 法により求めた。

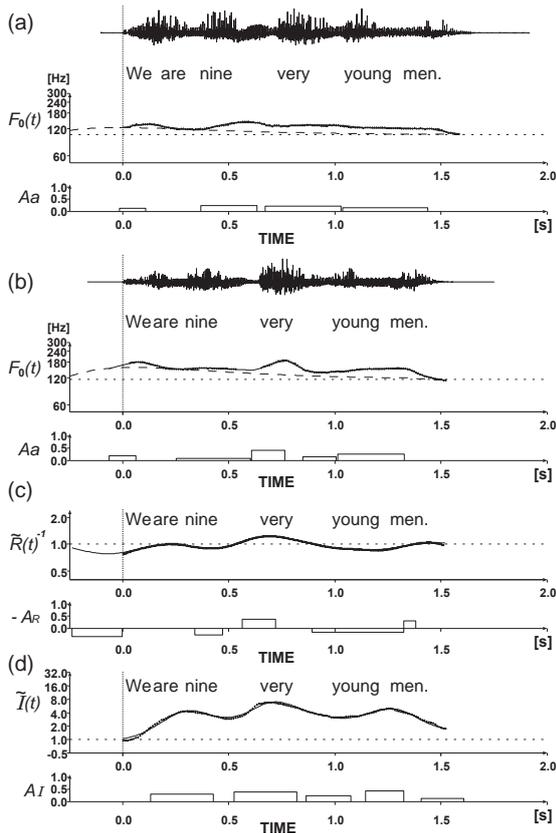


図 4: F_0 パターン、発話速度変化パターン、音源強度変化パターンの分析例

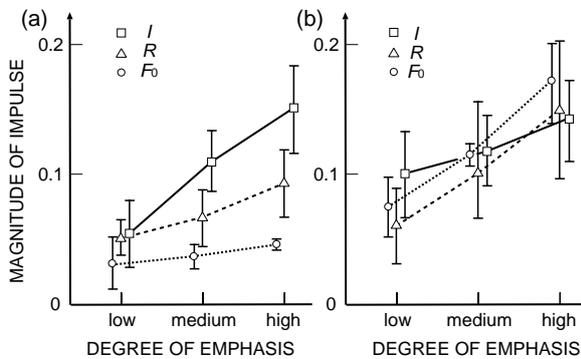


図 5: 3 段階の強調の程度に対する韻律指令の力積の大きさ

分析例を図 4 に示す。これは中立的な発話に対して、“very” を強調を中程度に強調した話者 A の発話を分析したものであり、上から順に (a) 中立的な発話の F_0 パターンとアクセント指令、(b) 強調発話の F_0 パターンとアクセント指令、(c) 中立発話に対する強調発話の発話速度変化パターンと発話速度変化指令、(d) 中立発話に対する強調発話の音源強度変化パターンと音源強度変化指令、を表している。ただし、発話速度変化については、強調した場合の多くに発話速度が小さくなる傾向があることを考慮

して、減速指令が正となるように、パターンは逆数を、指令に関しては符号を逆としている。

2.4.2 3 つの韻律的特徴に対する強調の影響

上述の各韻律指令が強調の程度によりどのように変化するかを調べるため、指令の大きさとその持続時間の積 (力積の大きさ) を指標とする。図 5 は話者 A と話者 B のそれぞれについて、3 段階の強調の程度に対する各韻律指令の力積の大きさを示したものである。母国語話者が音源強度 > 発話速度 > F_0 パターンの順で強調の程度を制御する特徴が大きく作用しているのに対して、非母国語の日本語話者は特に F_0 パターンと発話速度パターンによって強調の程度を制御している。

3 感情表現と韻律的特徴

3.1 感情の分類

“感情” とは、特定の事象に関する主観的評価を指す。感情には、基本的なもの、それらの複合的なものが存在するとする立場が一般的であるが、基本的なものが何であるかは必ずしも一致していない。本研究に取り組むにあたり、感情の起源が環境の変化に対する適応に由来するという立場から、具体的な対象を持つ感情として (1) 愛好 - 嫌悪、(2) 怒り - 怖れ、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。ただし、(4) は (1) ~ (3) と比べてやや異質であり、否定的・消極的な準備に対して、肯定的・積極的な対象が生起した驚きと、その逆の場合の驚きなどが混在するため、一般用語としての「驚き」、「期待」として捉えた場合には、状況によって (1) ~ (3) の他の次元の要因が複合的に関与するものと考えられる。また、これらの分類は、心理学的な見地に基づくものであり、日常生活での発話行為において、音声による表出の必要性、具体的な表出の程度や頻度の点では必ずしも一様ではない。

3.2 音声による感情の種類伝達能力

上述の 4 次元 8 種類に分類したそれぞれの感情が、音声によってどの程度伝達しうるかについて、自然音声を用いた聴取実験を行なった。

表 1: 音声資料に用いたテキストと対象とした感情

No.	テキスト	愛好	嫌悪	怒り	怖れ	喜び	悲しみ	驚き	期待
1.	駅まで迎えにいきますよ。								
2.	私が届けに行くよ。								
3.	おじいさんに会ってくる。								
4.	明日(あした)は雪だよ。								
5.	明日(あした)から夏休みだ。								
6.	明後日(あさって)はもうクリスマスだ。								

表 2: 9 種の表現を同定する聴取実験における刺激 反応の混同行列。下線は比較的大きな混同率 (> 0.2)

刺激	反応								
	中立	愛好	嫌悪	怒り	怖れ	喜び	悲しみ	驚き	期待
中立	.813	.021	.073	.031	.000	.000	.031	.010	.021
愛好	.198	.427	.021	.010	.052	.115	.031	.021	.125
嫌悪	.146	.000	.635	.125	.031	.000	.063	.000	.000
怒り	.010	.000	<u>.281</u>	.667	.031	.000	.000	.010	.000
怖れ	.094	.010	<u>.302</u>	.010	.177	.000	<u>.333</u>	.042	.031
喜び	.135	<u>.271</u>	.000	.000	.000	.219	.000	.156	.219
悲しみ	.031	.000	<u>.260</u>	.021	<u>.292</u>	.000	.354	.010	.031
驚き	.021	.188	.000	.000	.000	.115	.000	.521	.156
期待	.094	.198	.063	.010	.010	<u>.281</u>	.042	.135	.167

3.2.1 音声資料

上述の分類に基づいて、同次元の感情について同一のテキストを用意し、その両極を表現するための状況設定を行った。表 1 は、音声資料に用いたテキストと、それらをいずれの感情を込めて発話させたかについて示したものである。話者は、特別な訓練を受けた経験の無い成人男性 1 名(大学生)で、状況設定が書かれたカードを参照しながら、表 1 のテキストを、「中立」および「強」、「中」、「弱」3 段階の感情の度合いで発声した。

3.2.2 音声による感情の種類伝達能力

刺激音声は上記の音声資料のうち感情の度合いが「中」のものに「中立」を加え、9 表現 × 3 文 × 2 回分の 54 発話を用いた。これをランダムにヘッドホンを通して、成人男女 16 名(いずれも大学生)の被験者に呈示し、刺激に対して 8 種類の感情に「中立」を含めた 9 表現のうち、いずれに相当するか強制選択させ、回答用紙にマークさせた。なお、この実験は、被験者の識別能力を問うのではなく、発話により表現しうる感情の可能性について調べるものである。

表 2 は、刺激 反応の混同行列である。全般的には呈示した 1 文の発話に対して、概ね期待した感情が正しく受容されているが、刺激「怒り」に対する反応「嫌悪」、刺激「怖れ」に対する反応「嫌悪」

「悲しみ」、刺激「悲しみ」に対する反応「嫌悪」「怖れ」、刺激「期待」に対する反応「喜び」について混同する割合が比較的大きい。

また、「驚き 期待」の次元を除けば、同次元内で混同することはほとんどなく、対立する両極の基本感情については、音声の特徴上に明確な違いとして現れうることが示唆された。なお、「驚き 期待」については、内部的な準備の状況に対する対象の生起についての感情表現であり、準備している状況が積極的なもの、消極的なものなど種々の状況を想定することができるため、音声現象としてはやや複合的な状況を含むものとなり、この次元内での混同が多くなったものと考えられる。

3.3 感情の識別に関与する韻律的特徴

音声中に含まれる感情について、どのような音響的特徴がその表現に寄与するかを明らかにするため、「平静」の音声に対して、(1) F_0 パターン、(2) 局所発話速度、(3) 声帯音源強度、(4) 分節の特徴の 4 種の特徴について、感情を込めた音声のものと段階的に置換したものを刺激として、被験者に感情を判定させる実験を行った。ただし、(3) については音源強度の近似的な指標として、LPC 分析により得られるゲインを、(4) については LPC 係数を用いた。感情表現として「喜び」「怒り」「嫌悪」「悲しみ」の 4 種を取りあげ、男声話者 1 名に、これらに「平静」

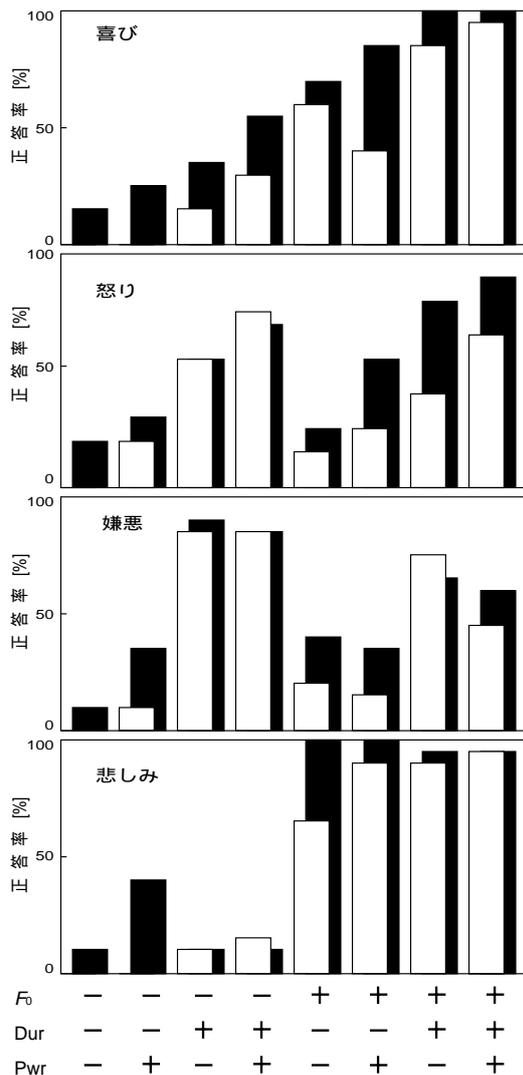


図 6: 特徴の置換合成音声に対する感情の判別結果 (図の詳細は本文参照)

を含む 5 種のカテゴリの発声をさせた。発話内容は『私は山へ行きます。』で、LPC による分析、パラメータの置換、再合成を行い刺激を作成した。

被験者は大学生 10 名で、刺激に対して 5 種のカテゴリから 1 つを強制選択させた。図 6 は段階的に特徴を変更した感情音声の判別結果であり、対象とした感情と判断した正答率を示したものである。韻律的特徴に対しては、- が「平静」の特徴量を、+ が感情を込めた音声の特徴量を採用したものであり、 F_0 が「平静」の LPC 係数を、 F_0 が感情を込めた音声の LPC 係数を用いた場合の正答率である。この結果から、「喜び」については F_0 パターンおよび発話速度が、「怒り」については発話速度および音源強度が、「嫌悪」については発話速度が、「悲しみ」については F_0 パターンが、感情表現の伝達に大きく寄与していることが分かる。

3.4 喜び 悲しみに関する韻律的特徴の分析

3.4.1 音声資料

音声資料として、同一のテキストに対して、「喜び」と「悲しみ」の 2 種の感情をそれぞれ強・中・弱の 3 段階の強さで表現した発話と、比較のために特別な感情を込めない中立な発話とを用意する。テキストは、表 3 に示す 5 文で、指定した感情を表現するために発話の状況を設定し話者に示した。例えば文 (1) 『見渡す限り海だ。』に対しては、「喜び」について『夏休み、友達と南の島へ来たあなたは、目の前の景色に心を打たれて一言』、また「悲しみ」について『船が難破し、漂着した所が小さな無人島だとわかって絶望のあまり一言』という状況を設定した。

話者は、いずれも演劇経験のある成人男女各 2 名であり、簡易防音室内において、ディスプレイに表示される状況設定の文章と発話テキストに従い、「中立」「中」「弱」「強」の順で感情の強さを変えて発声したものを録音した。

3.4.2 韻律特徴パラメータと分析手法

韻律的特徴を定量化するため、音声の F_0 パターンについては、その生成過程モデル^[2]に基づき分析を行ない、基底周波数 F_b 、フレーズ指令の大きさ A_p 、及びアクセント指令の大きさ A_a をそれぞれ求める。また発話速度については、大局的な特徴として平均発話速度 R を求める。

基本周波数の分析は、上述の音声資料に対して、変形自己相関関数法を用いて 10 ms 間隔で F_0 の値を抽出し、実測の F_0 パターンに対する最良近似を与えるモデルのパラメータの値を決定した。また、発話速度は波形及び短時間平均パワーの視察により発話の開始時点及び終了時点を決して求めた。

3.4.3 分析結果

中立発話に対して、感情を込めた発話の韻律的特徴量がどのように変化するかをとらえるため、分析

表 3: 収録に用いた発話テキスト

(1) 見渡す限り海だ。	(4) 今日はもう帰るよ。
(2) 雨が降ってきた。	(5) 数学が 70 点だ。
(3) もう夜だよ。	

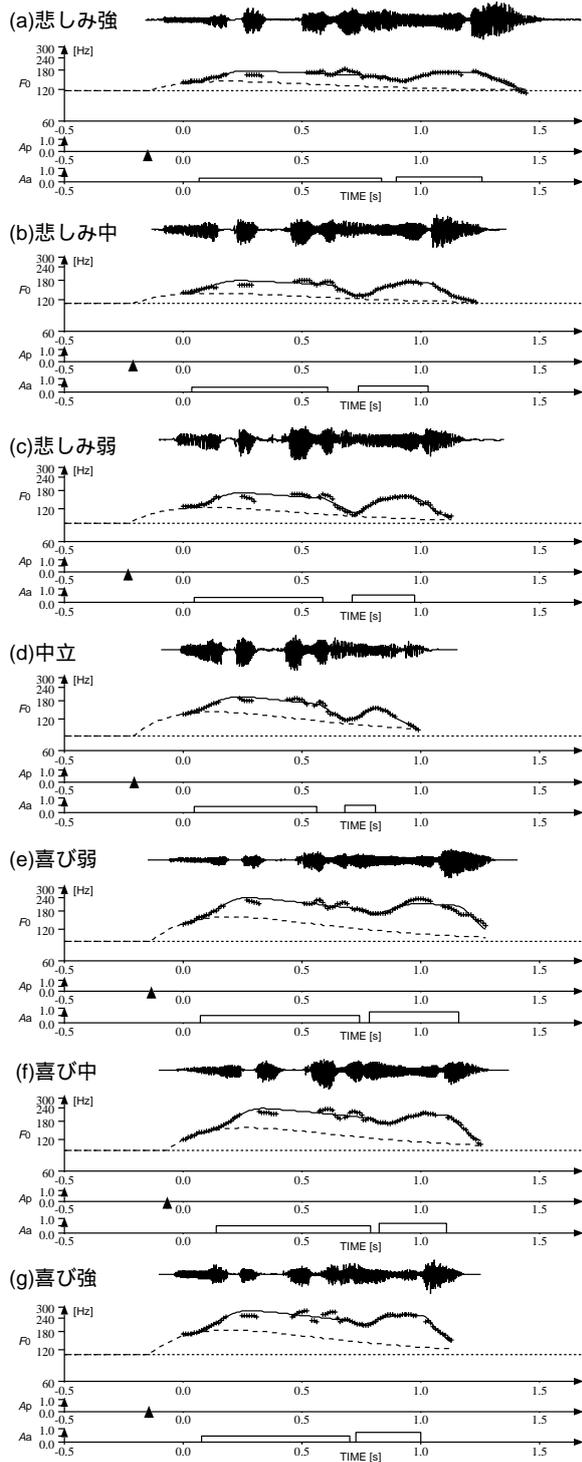


図 7: 分析例 (テキスト「見渡す限り海だ。」)

から得られたパラメータについて、中立発話の対応するパラメータの値からの変化分を比として求めた。

図 8 は、話者のうち男性 1 名分のデータについて、4 種類の韻律的特徴量に対してそれぞれ中立発話との比を求めた結果である。横軸は、中央から右に向かって「喜び」の強さが大きくなり、中央から左に向かって「悲しみ」の強さが大きくなるように配置

表 4: 中立発話に対する各種パラメータの平均値の有差検定結果

特徴量	悲しみ			喜び		
	強	中	弱	弱	中	強
<i>Fb</i>	+++	++	+	++	++	+++
<i>Ap</i>	--	---	-	++		+
<i>Aa</i>	---	---	--			++
<i>R</i>	---	---	---	---	---	---

+++ / ++ / + : 有意水準 1% / 5% / 10% で平均値大
 --- / -- / - : 有意水準 1% / 5% / 10% で平均値小

した。図中のボックスは平均値 ± 標準偏差の範囲を表わしており、最小値・最大値間を直線で結んで示したものである。

また表 4 は、図 8 と同一のデータに関して、感情を込めた発話のパラメータと、同一テキストの中立発話の対応するパラメータとを対にして 2 つの母平均の差の検定を行なった結果である。‘+’は中立発声と比べて平均値が大きいと言える場合であり、同様に‘-’は平均値が小さいと言える場合で、その数が多いほど有意水準が高いことを示している。

(a) 基底周波数 「喜び」、「悲しみ」双方において、感情の強さが大きくなるのに従って、基底周波数は単調に高くなる傾向がある。「悲しみ」の方がその傾向が顕著である。

(b) フレーズ指令の大きさ 「喜び」については、一定量の若干の増大が見られる。一方、「悲しみ」については、その強さが大きくなるのに従って、より大きく抑圧される。その結果、基底周波数の変化と併せると、「悲しみ」の抑揚は、平均的な基本周波数が極端に低くならず、起伏が少なく平坦になる。

(c) アクセント指令の大きさ 「喜び」、「悲しみ」双方において、感情の強さには大きく影響されないものの、「喜び」についてはアクセント指令が大きくなり、「悲しみ」については若干抑圧されている。

(d) 発話速度 「喜び」、「悲しみ」双方において、感情の強さが大きくなるのに従って、単調に平均発話速度が遅くなっている。「悲しみ」については、その傾向が顕著である。理由の一つとして、いずれの感情を込めた場合にも、文末伸長が生じており、その強さが大きいほど顕著に現れる点が挙げられる。「悲しみ」に関しては、比較的全体的に単調に遅く

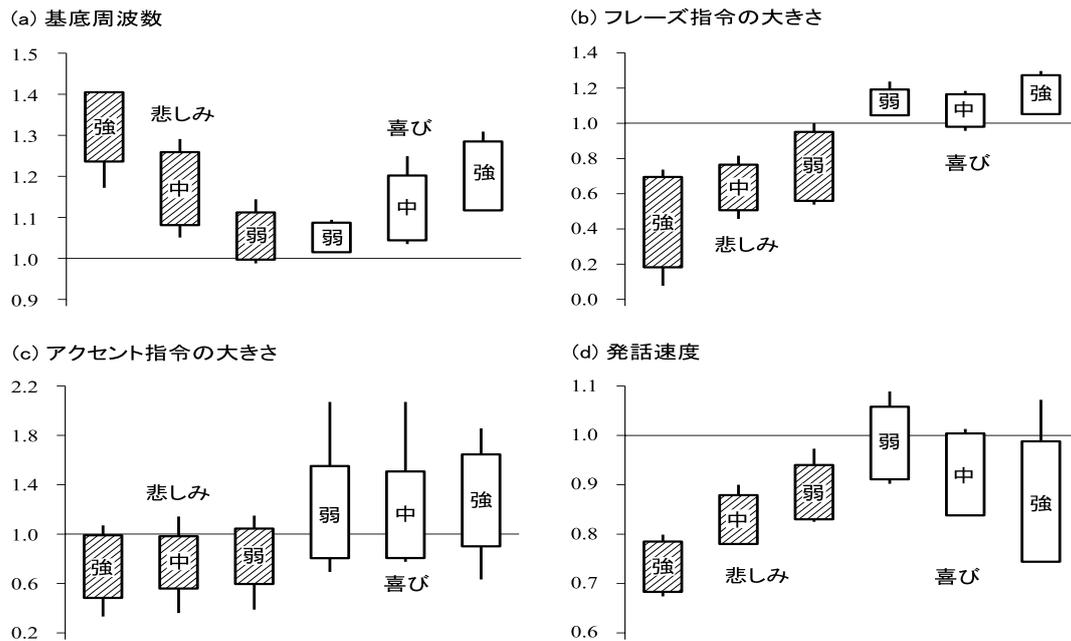


図 8: 中立発声に対する各種感情を込めた発話の特徴量の比

なっているのに対して、「喜び」では局所的な発話速度の変動が激しく、必ずしも発話速度全体が中立発声と比べて遅くなっているわけではない。

4 おわりに

音声の韻律的特徴を統一にとらえるため、 F_0 パターンに関する生成過程モデルを参考に、発話速度および音源強度に関して、それらの標準的な発話からの変化量の定式化・モデル化を行なった。それらのモデルに従って、強調表現における各種韻律的特徴の変化についてしらべた。

また、非言語情報として感情の表出に関して、受容特性、キーとなる韻律的特徴などを聴取実験により明らかにした。最後に、相対する感情の種類として、「喜び」と「悲しみ」を取り上げ、中立的な発声と比べた場合の韻律的特徴量の変化について特に F_0 パターンのモデルへの入力指令に関して詳細に調べた。

参考文献

[1] 藤崎博也, “日本語の音調の生成モデルによる分析,” 国際化する日本語 - 話し言葉の科学と音声教育, pp. 124–140, クバプロ (1993).

[2] H. Fujisaki and S. Nagashima, “A model for the synthesis of pitch contours of connected speech,” *Annual Report of Engineering Research Institute, University of Tokyo*, vol. 28, pp. 53–60 (1969).

[3] H. Fujisaki, S. Ohno and C.-F. Wang, “A command-response model for F_0 contour generation in multilingual speech synthesis,” *Proc. 3rd ESCA/COCOSDA International Workshop on Speech Synthesis*, pp. 299–304 (1998).

[4] S. Ohno and H. Fujisaki, “A method for quantitative analysis of the local speech rate,” *Proc. 4th European Conference on Speech Communication and Technology*, vol. 1, pp. 421–424 (1995).

[5] 杉山佳三, 藤崎博也, 酒巻宏, 大野澄雄, “韻律的特徴によって表現される強調の定量的分析,” 日本音響学会 2001 年春季研究発表会講演論文集, vol. 1, pp. 265–266 (2001).

[6] S. Ohno, Y. Sugiyama and H. Fujisaki, “Quantitative analysis of the effects of emphasis upon prosodic features of speech,” *Proc. 7th European Conference on Speech Communication and Technology*, pp. 661–664 (2001).