韻律コーパスの構築・分析と韻律モデルパラメータの自動抽出

Creation and Analysis of Prosodic Corpus and
Automatic Extraction of Prosody Model Parameters

筑波大学大学院 システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba

板橋 秀一
Shuichi Itahashi

＜研究協力者＞

筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering,
University of Tsukuba

布 社輝
Shehui Bu

筑波大学情報学類
College of Information Sciences,
University of Tsukuba

山本 幹雄
Mikio Yamamoto

筑波大学大学院 理工学研究科
Graduate School of Science and Engineering, University of Tsukuba

山本 佑
Tasuku Yamamoto

嶋田 幹貴
Motoki Shimada

We have been creating a Japanese prosodic corpus based on the existing speech corpora. We have chosen ASJ guide-task sentences, ASJ played dialogues and the Priority Areas Project corpus of played dialogues. We extracted fundamental frequency (Fo) and attached utterance labels to each sentence. We investigated occurrence frequencies of utterance units and Fo values. The result shows that the utterance unit length of the dialogue tends to be much shorter than that of the read speech. According to the clustering with 7 parameters derived from Fo patterns, sentence-initial phrases have similar characteristics in both read and dialogue speech. These results will be useful when users of the prosodic corpus select one which suits their objectives.

In the latter half of this paper, an automatic method to extract the discrete parameters of an F_0 model is proposed which is an extended version of what is called Fujisaki model. This method is based on the dynamic programming (DP) and the least mean square error (LMSE) methods. It is divided into two main steps in order to decrease the calculation time exhausted by the DP method. Furthermore, in order to detect the optimal number of phrase commands automatically, decrease of LMSE is used. From the results of the experiment on a set of 11 sentences spoken by four Japanese speakers, we obtained about 84% correct rate of phrase component detection by the proposed method.

Key words: Dialogue, Utterance unit, Fo pattern, Fo model, Fujisaki model, Dynamic programming (DP), Least mean square error, Two-step algorithm.

1. はじめに

音声認識の研究において近年統計的手法が用いられるようになった。統計的手法による音声認識モデルの学習のためには大量の音声データが必要となる。さらにコーパスに基づいた音声合成方式においても大量の音声データが要求される。このようなことから音声データベース／コーパスが音声研究を促進するために不可欠であることは広く認識されるようになった。音声コーパスは、種々の年代、性別、および方言の話者から得られたデータを含む必要がある。

これまでの音声研究においては、どちらかといえば音韻的特徴に注意を払い、韻律的特徴はあまり考慮されなかった。しかし、韻律的特徴は言語の構文や意味を伝達する際に重要な役割を果たしているので、それらを系統的・集中的に調査することが必要である[1]。

これまで、様々な種類の音声データベースが開発されているが、音声の韻律情報を含んでいるものは少ない。我々は最初から韻律音声コーパスを作成するのではなく既存の音声コーパスに韻律情報を付与することを進めて来た。以下ではまず対象とした音声コーパスおよび韻律コーパスの仕様を述べる。次に発話単位と

基本周波数についての分析結果を述べる[2]-[4]。これらの結果は、本韻律コーパスを利用する際のコーパス選択の指針となることを意図している。

本論文の後半では、韻律モデルパラメータの自動抽出について述べる。

2. 韻律コーパスの構築

2.1 対象とする音声コーパス

対象音声コーパスとしては次の３種を選んだ。

1) 日本音響学会「研究用連続音声データベース」 Vol. 4-6　各種案内読み上げ文
2) 日本音響学会「研究用連続音声データベース」 Vol. 7　模擬対話
3) 重点領域研究「音声対話」の対話音声コーパス Vol. 1-4

2.1.1 音響学会各種案内読み上げ文

ASJ コーパスは1992年に国内の大学等17機関の協力により音響学会の音声データベース委員会によって設計・作成された[5]。ASJ コーパスは全部で12474文から成り、男女各18人計36人の話者によって発話された案内タスク1027文のテキストおよび対応する音声波形を含んでいる。音声波形は16kHz、16ビットでディジタル化されている。内容は、地理ガイド5種、観光ガイド7種、コンサートガイド2種、パスポートに関する問い合わせ、スキーツアーガイドを含む対話を基に構成された。対話のタスクを予め2名の話者に与えて自由に対話をしたものをテキストに書き起こし、そのテキストから間投詞や言い誤り等の表現を除いて作成し直したテキストを、1名の話者が読み上げたものである。従って ASJ 案内タスク文は対話音声ではないが、対話の様相を反映した音声資料と言える。

2.1.2 音響学会模擬対話音声コーパス

模擬対話音声コーパスは1992年に国内の大学等17機関の協力で設計・構築された日本音響学会「研究用連続音声データベース」 Vol. 7 である[5]。道案内、観光案内、音楽会案内、パスポート取得問い合わせ、スキーツアー案内等の各種の案内をタスクとした模擬対話（37対話）を各々2名の話者が発声した音声波形と、その書き起こしテキストが収録されている。これは予め与えられたタスクに応じて2人の話者が模擬対話を行ったものを収録したものである。音声波形は16kHz, 16ビットでディジタル化されている。各対話は2名の話者が行い、全部で37対話あるので延べ74名が発声しているが、何人かの話者は複数の対話に参加しているので、全話者数は37名（男29名、女8名）となっている。

2.1.3 重点領域研究「音声対話」コーパス

「音声対話」コーパスは1994年に国内の12大学の協力により、文部省科学研究費補助金重点領域研究「音声・言語・概念の総合的処理による対話の理解と生成に関する研究」（略称「音声対話」）の音声コーパスワーキンググループによって設計・構築されたもので Vol. 1-4から成る[6]。Vol. 1-3には64対話6400文（約1GB）の男性40名、女性20名、計60名の話者による音声波形とその書き起こしテキストが収録されている。音声波形は16kHz, 16ビットでディジタル化されている。秘書システム、スケジュール管理、クロスワードパズル、旅行案内、テレフォンショッピング、地理案内、スケジュール調整、学生の雑談、留学生の雑談、間違い探しの計11種の模擬対話から構成されている。この中の10対話は、人間対機械の対話において、音声認識部を人間が代行する Wizard of Oz 方式で収録されている。この方式の場合は合成音声も収録されているが、合成音声については分析・ラベル付けはしていない。Vol. 4（31対話、23名、約415MB）には、地図課題が含まれている。これは、1人が道の記された地図を与えられ、もう1人に手もとの地図上にどのように道を引くかを教える対話である。

2.2 韻律コーパスの仕様

2.2.1 基本周波数（Fo）分析

分析時刻、短時間（分析フレーム内）音声パワー、基本周波数（Hz）を記録する。

(1) F_0抽出には原則として waves+/ESPS の中の Fo 抽出プログラムを利用する。
(2) 分析窓：ESPS の指定に従う。
(3) 分析フレーム間隔：5ms とする。
(4) Fo の誤り訂正：Fo の分析値が前後のフレームの値と著しく異なる場合は誤りとみなして修正する。修正には当該分析フレームの、前後フレームの Fo 値および音声波形情報等を参照する。

(5) 記録フォーマット：分析フレーム時刻（ms 単位）、Fo（Hz 単位）で以下の通り。

音声ファイル名

1フレームの時刻　Fo　Fo の訂正値　パワー

2フレームの時刻　Fo　Fo の訂正値　パワー

3フレームの時刻　Fo　Fo の訂正値　パワー

2.2.2 発話ラベル

(1) 韻律ラベル

D：平叙文　　Q：疑問文　　I：間投詞

E：感嘆詞　　N：音声・雑音　　S：合成音

（韻律ラベルは1つのファイルに１つとは限らず、ファイル名の後に記述してある）

(2) 対話文ラベル

対話文は全てローマ字で記録してある。

を：o　　ん：nn　　っ：q

(3) 対話文以外のラベル

pause：無音、　noise：雑音　　lgh：笑い声、

swt：スイッチ、　uci：意味不明、

cough：咳・くしゃみ、

noise-0：対話相手の声が入っている場合

(4) 同時に会話している場合

-a：回答者

-q：質問者

-a2：二人目の回答者（話者が3人の場合使用）。

(5) ファイル名：音声波形ファイルと同じ名称とし、拡張子として lb をつける。

(6) フォーマット

ファイル名　　韻律ラベル

開始時刻　　終了時刻　　ラベル

：　　：　　：

例：pasd4_siz0001_011

0　1352　kyouhayoroshikuonegaishimasu-a

（大学に訪問しているアナウンサー）

856 3498　yoroshikuonegaishimasukitazawadesuaqyoroshikuonegaishimasu-a2　（大学教授）

1404 2873　sennseeyoroshikuonegaisimasu-q

（スタジオにいるアナウンサー）

3498 3882　pause

3. 韻律コーパスの分析

3.1 発話単位

以下では「200ms 以上の無音区間によって区切られた音声区間」を発話単位とする。図1と図2に読み上げ音声と模擬対話音声の発話単位長の出現頻度を示す。図2から模擬対話では100ms から200ms といった短い発話が多く集中して出現しているのに対し、図1から読み上げには1.3sec にもピークが見られ、その付近にも多くの発話が見受けられる。模擬対話では間投詞を非常に多用したり、また相手に発話を遮られたりすることによって発話単位長が非常に短くなることが分る。また発話単位長は発話速度によっても変動する。今後読み上げ音声と模擬対話音声の発話速度を調査する必要がある。

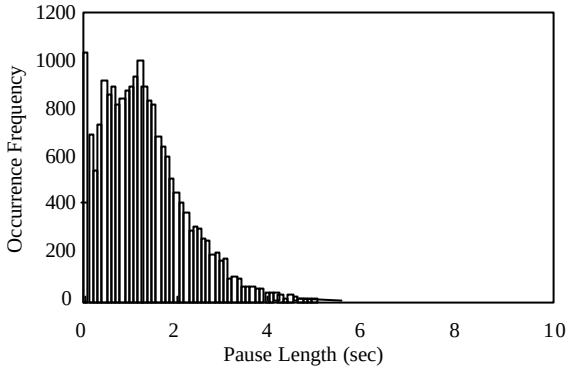

図1 :発話単位長の出現頻度　読み上げ）

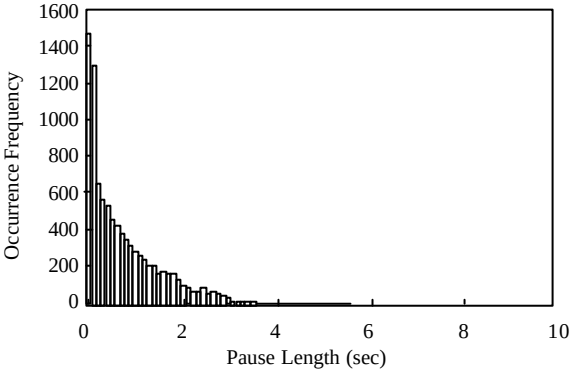

図2 :発話単位長の出現頻度　模擬対話）

3.2 発話単位長の主成分分析

発話単位長の出現頻度については3.1に示したが、その特徴をさらに明確にするために、主成分分析を行なった。発話単位長は0から9秒までを450ms ごとにクラス分けして、21個の説明変数とした。第1・第2主成分で累積頻度は80%を超えている。第1・第2主成分平面での分布を図3に示す。図3で□印は対話を、×印は読み上げを表している。この図から、実際上は第1主成分だけで対話と読み上げがかなり分離できることがわかる。第１・第2主成分の固有ベクトルを図4と図5に示す。これらの図から、発話単位長の短いものが対話に多く出現していることが分かる。

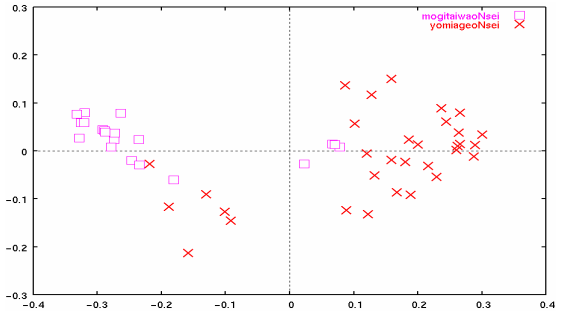

図3　第1-第2主成分平面上の発話単位長

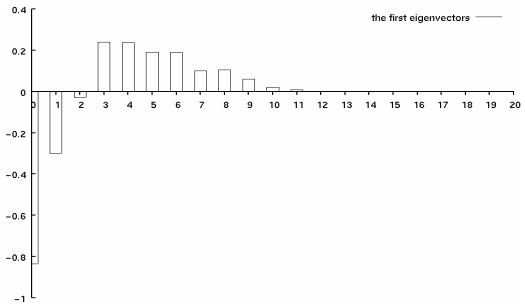

図4　第1固有ベクトル（横軸は発話単位長のクラス）

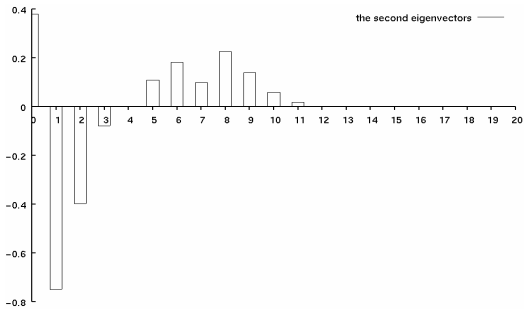

図5　第2固有ベクトル（横軸は発話単位長のクラス）

3.3　基本周波数に関する分析

図6に示すように基本周波数の時間的変化に関するパラメータとして次の7種を用いた。すなわち、F_0の開始・終了・最大・最小値、Fo の開始時点を基準とした最大値・最小値・終了点の時刻である。これらを用いて k-means 法によるクラスタリングを行った。各クラスタの初期中心にはサンプルの中からランダムに選び、最大クラスタ数は7とした。また、男女差と発話速度による影響を避けるために、時間および Fo の正規化を行った。図7と図8はそれぞれ読み上げと模擬対話の発話単位をクラスタリングした結果で、横軸は発話単位長、縦軸は Fo 開始点から最大値までの時間を示している。図7の文頭名詞と図8の文頭の句が2次元平面でほぼ同じ位置を占めることから、読み上げ・模擬対話に関わらず、文頭の句に関する韻律情報は類似の特徴を持つと考えられる。

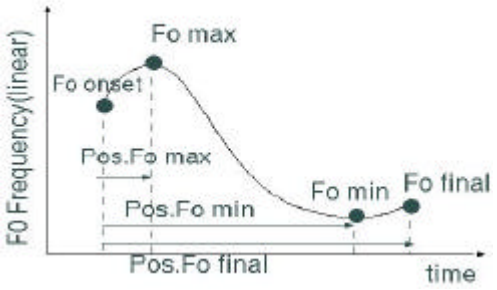

図6　韻律パラメータ

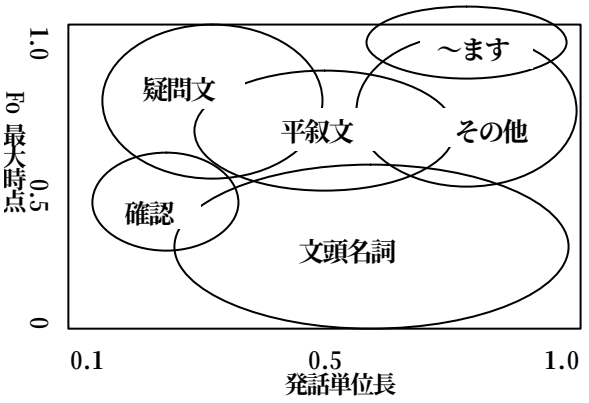

図7　発話単位のクラスタリング（読み上げ）

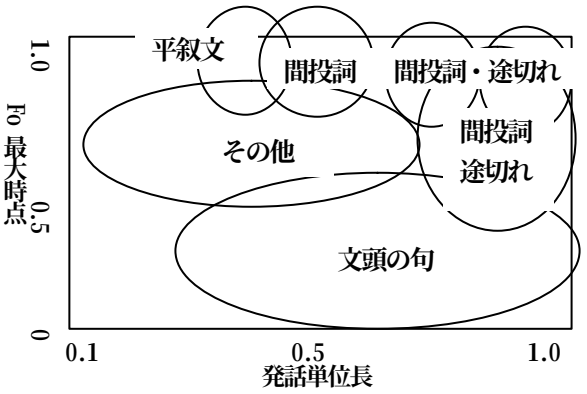

図8　発話単位のクラスタリング（模擬対話）

4.　Fo モデル・パラメータの自動決定

音声の基本周波数パターン（以下 Fo パターン）は音声コミュニケーションに重要な役割を果す。また、このパラメータは主に韻律の情報によって決定される。人間の耳は、他の音声信号パラメータの変化よりも Fo の変化に、より敏感である。近年、コンピューターと

マルチメディア技術の開発によって、実際の音声 Fo パターンを自動的に分析し近似することが必要とされるようになった。

Fo の生成のプロセスを数学的に説明する適切なモデルがあれば、Fo パターンと韻律学の関係を、定量的に分析することができる。藤崎モデル、Downstep モデルなどのようないくつかのモデルがこの問題に対処するために提案された[14]。多くの研究により、藤崎モデルが日本語等の言語についてイントネーションの変化を非常によく記述することができることが示された。一方、Fo モデルのパラメータを自動的に抽出する方法が必要である。幾つかの方法がこの問題を解決するために提案された。しかしながら、これらの方法では、良い初期値が必要とされることから、複雑な前処理を要求される[15, 16]。

本稿では、実測の Fo パターンから、Fo モデルのパラメータを自動的に抽出する方法について述べる。提案されたアルゴリズムは動的計画法および最小二乗法に基づいている。計算時間を減少させるために、このアルゴリズムは二ステップに分割する。最初に、フレーズ成分の境界を決定し、次に、フレーズおよびアクセント・コマンドの最適なパラメータを決定する。さらに、フレーズ・コマンドの最適数を自動的に決定するために、最小二乗誤差の減少量 $c(l)$を提案する。最後に、日本語の音声例を対象として、本提案手法を用いて分析を行い、本手法の性能を実験的に検討する[8, 9]。

5. Fo モデル

このモデルは1970年代に藤崎らによって提案された。このモデルは、フレーズ成分とアクセント成分、二つの成分から構成される。文献[10, 15, 16]によれば、藤崎モデルは以下のように表現される:

$$\ln\left(\hat{F}_0(t)\right) = \ln\left(F_{\min}\right) + \sum_{i=1}^{I} A_{pi} G_{pi}\left(t - T_{0i}\right) + \sum_{j=1}^{J} A_{aj}\left\{G_{aj}\left(t - T_{1j}\right) - G_{aj}\left(t - T_{2j}\right)\right\} \tag{1}$$

ここで、

$$G_{pi}(t) = \begin{cases} a_i^2 t e^{-a_i t} & : t \ge 0 \\ 0 & : t < 0 \end{cases} \tag{2}$$

$$G_{aj}(t) = \begin{cases} \min\left[1 - \left(1 + b_j t\right)e^{-b_j t}, g\right] & : t \ge 0 \\ 0 & : t < 0 \end{cases} \tag{3}$$

方程式(2)は、フレーズ制御機構の関数、また、方程式(3)は、アクセント制御機構の関数を示している。

発話された音声の Fo パターンは、個々の音韻による影響も加わって、非常に複雑な動きを示している。一般に、Fo パターンの基本となる成分は、吸気を伴うポーズの後、肺からの呼気圧の自然な減少に伴い、一定の減少率で降下するものとみなされるが、一定の高さから出発して一定の高さに向かって制御されつつ降下する特性を有する[12]。Fo パターンをよく近似するために、傾きを考慮し、式(1)の中の項$\ln(F_{\min})$を、直線 $b_i(t-T_{0i})+c_i$に置き替える[8, 9]。

$$\ln\left(\hat{F}_0(t)\right) = b_i\left(t - T_{0i}\right) + c_i + \sum_{i=1}^{I} A_{pi} G_{pi}\left(t - T_{0i}\right) + \sum_{j=1}^{J} A_{aj}\left\{G_{aj}\left(t - T_{1j}\right) - G_{aj}\left(t - T_{2j}\right)\right\} \tag{4}$$

6. 分析方法

6.1 パラメータ値の決定

文献[8, 9, 13]によって、方程式(4)で示されるような関数$\hat{F}_0(t)$によって、$F_0(t)$を近似することを仮定すると、我々は以下の関数によって平均二乗誤差を得ることができる:

$$j(t) = \frac{1}{T}\sum_{t=1}^{T}\left\{\hat{F}_0(t) - F_0(t)\right\}^2 w(t) \tag{5}$$

誤差は有声と判断された区間に対してのみ考慮するので、有声ならば1、無声ならば0をとる関数 $w(t)$をかける。

各モデルの持つ変数ごとについての偏微分を取り、これらを0とする連立方程式を解くことで、$j(t)$を最小にする近似関数を得ることができる。ここで、b_iは負の値を取り、一方A_{pi}，A_{aj}およびc_iは負の値を取らないとする。

6.2 パラメータ入力時刻の決定

モデルのパラメータ入力時刻を決定するために、動的計画法（DP 法）[8, 11, 13]を導入する。動的計画法の原理は多段問題を2段問題にすることと言える。その代わり、前の段階の最適の値を保持するための記憶テーブルが総計算量を減らすために必要となる。

最初に、長さ Tの発話を N 個に分割し、フレーム番号 n = (1, 2, ..., N)を仮定する。

$0 \equiv n_0 < n_1 < \cdots < n_K \equiv N$に対して、$j_k\left(n : n_k, a^{(k)}\right)$が k 区間目の二乗誤差を表すとすると、

$$j_k(n:n_k,a^{(k)}) = \frac{1}{N_k} \sum_{n=n_{k-1}+1}^{n_k} \left\{ \left(\hat{F}_{0k}(n-n_{k-1},a^{(k)}) - F_0(n) \right)^2 w(n) \right\} \quad (6)$$

ただし$n_0 = 0$，$n_K = N$，$N_k = n_k - n_{k-1}$また、$a^{(k)}$がパラメータA_p，A_a，b及びcを表す。ここで、最小化の手続を二つに分けることができる。１つはk番目の区間を最小化すること、また、もう一つは、１番目、２番目... k-1番目を最小化することである。ここで$g_k(n_k)$はk番目の段階における0とn_kの間の区間の最適解、pはパラメータの数を表す。

$$\begin{aligned} g_k(n_k) &= \min\left(a_1^{(k)}, \cdots, a_p^{(k)}; n_1, \cdots, n_k\right) \sum_{k=1}^{K} j_k(n:n_k,a^{(k)}) \\ &= \min\left(a^{(k)}; n_1, \cdots, n_k\right) \left\{ j_k(n:n_k,a^{(k)}) + g_{k-1}(n_{k-1}) \right\} \end{aligned} \quad (7)$$

6.3 二段アルゴリズム

DP 法によると、４つのパラメータA_{pi}，A_{aj}，b_iおよびc_iの計算は相当な計算時間を要する[8, 9]。計算時間を減小するために、二段アルゴリズムを提案した[8, 9]。このアルゴリズムは、最初 に DP 法を用いてフレーズ・コマンドの時刻を自動的に決め、次に１フレーズ成分内で最適な４パラメータを決定する。

第一段：フレーズ成分の式(2)および直線$b_i(t-T_0)+c_i$の結合を最小二乗法と DP 法を単純に使用して、フレーズ・コマンド時刻を安定かつ適切に抽出することができる。

第二段：第一段の結果によって、発話文をフレーズ・セグメントに分割する。その次に、各フレーズ・セグメントの処理は二つの手続きから構成される。一つは対応するフレーズ・セグメント中のアクセント・コマンドの最良の境界を探索することで、もう一つは最小二乗法によりフレーズ・セグメント内の４つのパラメータを算出することである。

6.4 フレーズ個数の自動判定

実際の分析では、提案された自動アルゴリズム中のフレーズ・セグメントの最適の数を決定することが必要である。ここでは、最小二乗誤差の減少量 $c(l)$と減少率 $d(l)$をとり上げる。

$$c(l) = \left| E(l) - E(l-1) \right| \quad (8)$$

$$d(l) = c(l) / E(l) \quad (9)$$

ここで、

$$E(l) = \frac{1}{N} \sum_{n=1}^{N} \left\{ \hat{F}_0^{(l)}(n) - F_0(n) \right\}^2 w(n) \quad (10)$$

$E(l)$はフレーズ・セグメント l個の場合、実測F_0と近似$\hat{F}_0^{(l)}(n)$の 平均最小二乗誤差, l はフレーズ・セグメントの個数である。$c(l)$と $d(l)$の最大値をフレーズ・セグメントの最適数の最良の候補と考える。次の実験はこの仮定を検討するために実行された。

7. 分析試験

7.1 実験データ

実験データとして、日本語11文を一セットとして選択した[10]。話者は20歳代の日本人男女各二名（男性：M1, M2；女性：F1, F2）であり、防音室で２回（一回目は練習）録音された。音声データは16kHz にサンプリングされ、16ビットで量子化された。AMDF 方法によって Fo パターンを抽出した。フレーム間隔は10ms である。実際に、発話者の個人性と言語的な曖昧性があるために、このセット中のある文のフレーズ個数を唯一に決定することができない場合もある。

7.2 実験結果

7.2.1 分析の例

提案された方法を評価するために、実験を行った。分析例は表１に示す男性 M1によって発話された「あの青い葵の絵はある」/*anoaoiaoinoewaaru*/である。図1は最小二乗誤差(LMSE)$E(n)$、最小二乗誤差の減少量$c(l)$および最小二乗誤差の減少率 $d(l)$をそれぞれ表示している。図1の中で示された $c(l)$および $d(l)$によって、フレーズ個数が3である場合 $c(l)$および $d(l)$が最大値であることが分かる。これから、フレーズの最適な個数が3であると考えることができる。

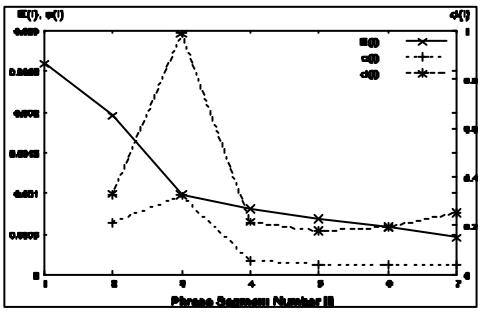

図1 日本語音声「あの青い葵の絵はある」の最小二乗誤差(LMSE)$E(l)$、減少量 $c(l)$、および減少率 $d(l)$

フレーズ個数が3である場合の実験結果を図2に示す。この図から、モデルの Fo パターンが実際の Fo パターンを十分良く近似していることが分る。

7.2.2 比較結果

c(*l*)の有効性を評価するために、収録音声データに対して実験を行った。表1で、項目 *c*(*l*)が判定個数を、項目 C が判定結果を表す。
c(*l*)は、式(8)が最大値となるときに判断された最適なフレーズ個数である。話者 M1、M2、F1および F2に対応する *c*(*l*)の正答率は81.8%と90.9%の間にある。対照的に、最小二乗誤差の減小率 *d*(*l*)の正答率は45.5%から72.7%までに及んでいる。*c*(*l*)の平均正答率は84.1%で、また、*d*(*l*)のそれは59.1%である。これらの結果によって、我々は、*c*(*l*)の正答率が *d*(*l*)より高いと言うことができる。

表1　The number of Phrases decided by the decrease of LMSE *c*(l).　*C*:(o : correct;×: incorrect)

No.	M1		M2		F1		F2	
	c(*l*)	*C*	*c*(*l*)	*C*	*c*(*l*)	*C*	*c*(*l*)	*C*
1	2	o	2	o	2	o	2	o
2	3	o	2	o	2	o	2	o
3	3	o	3	o	3	o	3	o
4	3	×	3	×	2	o	2	o
5	2	o	3	o	2	o	2	o
6	2	o	2	o	2	o	2	o
7	3	o	3	o	3	o	3	o
8	3	o	4	o	2	×	4	o
9	4	o	2	×	2	×	2	×
10	2	o	2	o	3	o	2	o
11	2	×	3	o	3	o	3	o
	81.8%		81.8%		81.8%		90.9%	

8. まとめ

既存の音声コーパス3種に対応する韻律コーパスを作成した。また、これらのコーパスについて、読み上げ音声と模擬対話音声から韻律情報を抽出し、その比較を行った。まず模擬対話では、間投詞や対話の割り込みによって発話単位長が短くなる傾向のあることが分った。さらに発話単位に関してクラスタリングを行った結果、読み上げ・模擬対話に関わらず、文頭の句は似通った特徴を持つことが分かった。これらの性質は、本韻律コーパスを利用する上で、利用者がコーパスを選ぶ際の参考となるものと考えられる[7]。

韻律の時間的変化を記述するモデルとしてよく知られている「藤崎モデル」に直線下降成分を加えたモデルについて、そのパラメータを動的計画法と最小二乗法により自動抽出する方法を提案し、その有効性を示した。また、フレーズ個数の自動判定方法の評価を 行った。これについては端点自由 DP 法の導入等、アルゴリズムの改善と高速化、より一般的な音声デー タによる検証、話者の増加などの課題が残されている。

韻律コーパスの一部については静岡大学の協力により J-ToBI による分析を行ったが、これについては省略する。

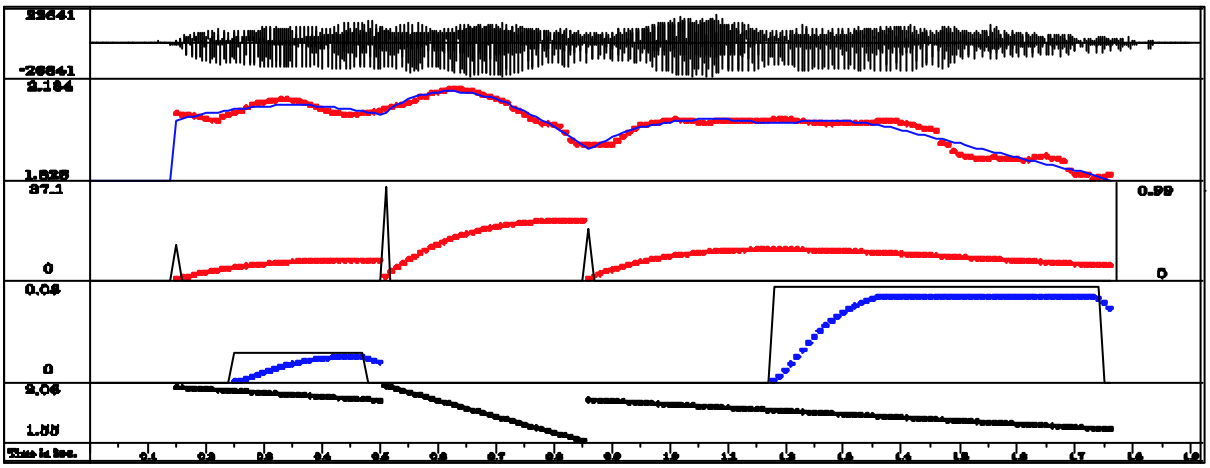

図2　フレーズ個数が3のとき、日本語音声「**あの青い葵の絵はある**」の分析結果。上から順に、音声波形、Fo パターン(点線)と近似パターン (実線)、フレーズ指令(実線)とフレーズ成分(点線)、アクセント指令(実線)とアクセント成分(点線)、直線成分(点線)をそれぞれ示す。

9. 参考文献

[1] 広瀬啓吉, “韻律情報の処理,” Journal of Signal Processing, Vol.2, No. 6, pp. 415-421 (1998)

[2] 嶋田, 山本, 板橋, “案内文タスク読み上げ音声の韻律分析,” 日本音響学会秋季研究発表会講演論文集, 3-10-6, pp. 339-340 (2002)

[3] 嶋田, 山本, 板橋, “案内文タスク読み上げ音声韻律コーパスの設計・構築と分析,” 日本音響学会春季研究発表会講演論文集, 1-p-31, pp. 413-414 (2003)

[4] 嶋田, 山本, 板橋, “読み上げ音声と模擬対話音声の韻律の比較,” 特定領域研究「韻律と音声処理」平成14年度第2回全体会議報告, pp. 47-50 (2003)

[5] 小林, 板橋, 速水, 竹澤, “日本音響学会研究用連続音声データベース,” 日本音響学会誌 Vol. 48, No. 12, pp. 888-893 (1992)

[6] 堂下, 新見, 白井, 田中, 溝口 :音声による人間と機械の対話, 第7編, 第2章, “PASD コーパス:重点領域研究模擬対話音声コーパス,” pp. 361-375オーム社 (1998)

[7] S. Itahashi, N. Ueda, M. Yamamoto, “Several Measures for Selecting Suitable Speech Corpora,” Proc. *Eurospeech'97*, pp. 1751-1754 (1997)

[8] S. Bu, M. Yamamoto, S. Itahashi, “A method of automatic extraction of Fo Model parameters,” Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition 2003, Tokyo, Japan, pp. 227-230 (Apr. 2003)

[9] S. Bu, M. Yamamoto, S. Itahashi, “Considerations on automatic parameters estimation of Fo Model,” Prep. Autumn Meeting of the Acoust. Soc. of Jpn, Paper 1-8-23, pp. 227-228 (Sep. 2003) (in Japanese)

[10] H. Fujisaki, K. Hirose, “Analysis of voice fundamental frequency contours for declarative Sentences of Japanese,” Jour. Acoust. Soc. Jpn. (E) Vol.5, No.4, pp. 233-242 (1984)

[11] S. Furui, “Digital speech processing, synthesis and recognition, second edition, revised and expanded,” Marcel Dekker Inc. (2001)

[12] K. Hakoda, H. Sato, “Prosodic Rules in Connected Speech Synthesis,” IEICE Trans. , Vol. J63-D, No. 9, pp. 715-722 (1980) (in Japanese)

[13] S. Itahashi, “Description of speech data patterns by several functions with applications to formant and fundamental frequency trajectories,” STL-QPSR 2-3, pp. 1-22 (Oct. 1978)

[14] H. Kubozono, “The organization of Japanese prosody,” Kurosio Publishers, Tokyo Japan (1993)

[15] H. Mixdorff, “A novel approach to the full automatic extraction of Fujisaki model parameters,” ICASSP2000, Vol. 3, pp. 1281-1284 (2000)

[16] S. Narusawa, N. Minematsu, K. Hirose, H. Fujisaki, “A method for automatic extraction of the fundamental frequency contours generation model,” IPSJ Jour. , Vol. 43, No. 7, pp. 2155-2168 (Jul. 2002) (in Japanese)