修士論文

基本周波数パターン生成過程モデルに基づく
コーパスベース韻律生成における発話スタイル制御

平成23年2月9日

指導教員　　広瀬 啓吉　教授

情報理工学系研究科電子情報学専攻

48-096424　見原 隆介

概 要

音声認識，音声合成といった音声信号処理の技術の発達によって，音声の入出力を伴うインターフェースは近年益々身近なものとなっている．音声合成の分野では，高品質かつ任意の入力テキストに対応できる音声合成システムも実用化されている．

統計的に音響パラメータを生成する音声合成手法としてHMM音声合成手法がある．音声信号の挙動を定常信号間の確率的遷移として表現することで，任意のテキストに対応できる音声合成手法として注目されている．また，音声の韻律を担う基本周波数の挙動を表現するモデルとして，基本周波数パターン生成過程モデル (F_0 モデル) がある．このモデルは人間の声帯制御機構の仕組みに即したもので，発話全体にわたる韻律の動きを自然な形で記述しやすい韻律生成系として，HMM音声合成と併用する手法が検討されている．

しかし，任意のテキストに対応するためには，一名の話者から大量の音声データベースを構築する必要があるため，音声合成システムに多様な韻律表現を持たせるために，データ収集のコストが大きな問題となっている．このため，言語情報を明瞭な音声として表現する機能は確保しつつも，韻律の表現の幅は狭く，読み上げ調に限定されている音声合成システムがほとんどである

これに対して，我々は F_0 モデルを用いた韻律生成系に並行して，比較的少数のデータベースから抽出した差分情報を用いて韻律の表現性を拡張する手法を検討している．これは感情，態度，発話意図などによって変化する韻律的特徴について，異なる特徴を持つ音声の間から韻律パラメータの差分情報を抽出し，韻律制御に利用するものである．差分という形でパラメータを学習しているため，一名の話者から得た学習データを様々な話者の音声合成システムに応用できることが考えられ，任意のテキストに対応でき，低コストかつ拡張性の高い韻律制御手法として期待される．

本研究では丁寧，ぞんざいといった発話スタイルを表現の対象とし，入力されたテキストに応じて推定された韻律パラメータの差分値を読み上げ調の音声に適用して目的の発話スタイルの韻律を生成する手法を検討した．丁寧，ぞんざいの二つの発話スタイルをターゲットとして，話者クローズド，話者オープンの韻律制御を試みた結果，主に丁寧の発話スタイル表現について，有効な再現性が確認された．

目 次

図 目 次

表 目 次

第1章　序論

1.1 研究の背景

計算機技術の急速な発展に伴い，情報機器も人々にとって身近な存在になりつつある中，人々と情報機器との円滑かつ直感的なインタラクションを実現するインタフェース技術の重要性はますます高まっている．直感的なインタラクションの一形態として音声による入出力が想定され，どのユーザにも気軽に使用できるマン・マシン・インタフェースの開発が期待される．

人間同士のコミュニケーションにおいて，音声による情報伝達が韻律の変化によって非常に多様な表現力を持っていることから機械側から出力される韻律情報の制御の高度化もまた，上記のインターフェースの実現にあたって大きな課題の一つとなってくる．

また，音声合成技術の発展により，任意のテキストに対応できる高品質な音声合成系 (Text-to-Speech，TTS) が近年実用化されてきているが，多くは特徴のない読み上げ調の発話スタイルによる音声を合成するものであり，韻律の表現力の観点から見れば音声合成技術はまだ発展途上にある．

現在では任意のテキストに対して特定のスタイルの韻律を得るためは，個別に大規模な音声試料のデータベースを用意する必要があるが，今後更に広がるであろう多様な用途に効率的に対応してゆくためには，既存のデータベースから得られる読み上げ調の韻律を加工して表現の幅を広げる韻律制御手法の開発が必要である．

1.2 研究の目的

韻律は言語情報の伝達は勿論，意図・態度・感情といったパラ・非言語情報の伝達に大きな役割を果たしているため，合成音声の情報伝達の幅を広げる上では，韻律を的確に制御することが重要な課題となっている．

我々は生成過程モデルに基づいて基本周波数パターン (F_0 パターン) を生成し，音声合成に利用するコーパスベース手法を開発してきた [2]．この F_0 パターンを対象とした生成過程モデルの枠組みを，基本周波数パターン生成過程モデル (F_0 モデル)[1] と呼んでいる．この手法は入力されたテキストの言語情報を抽出し，F_0 モデルにおける韻律パラメータを統計的に決定し，文全体の F_0 パターンを生成するものである．様々な言語情報を網羅した数百文規模の音声試料に基づいてパラメータを学習することで，任意のテキストに対応して韻律を生成できるシステムとして提案されている．ちなみに，数百文というデータベースの規模は，他の多くの TTS においても同様である．

しかし，韻律による表現の幅に着目すると，その多くが読み上げ調の音声に限定されてしまっているのが現状である．音声合成技術を様々な用途により適切に対応してゆくために，柔軟かつ多様な韻律表現が可能な音声合成システムが実現する事が望ましい．これに関連する問題点として，音声データベースから抽出・学習した韻律パラメータを直接合成音声に利用している手法では，音声提供者の話者性や発話スタイルによる影響を多分に受けるという点がある．このため，一つ一つの韻律表現を実現するためには逐一大量のコーパスを必要とする点がある．

そこで小規模なコーパスを追加することで韻律表現を効果的に制御する手法を考える．この手法が実現すれば，既に比較的潤沢に存在する読み上げ調のコーパスを利用できるため，高品質な音声合成と多様な韻律表現とを両立する上でコストを大きく削減できる．

本研究では，入力項目をテキスト (漢字かな混じり文) とし，一貫した流れにより各 F_0 モデルパラメータの差分情報を推定し，それを読み上げ調の韻律に適用することで，目的の発話スタイルの韻律を持った合成音声を出力するシステムを実装する．差分情報を推定するためのコーパスを少量の音声試料によってこのシステムを構築することで，効率的かつ高品質な韻律制御の実現を目指す．

1.3 本論文の構成

本論文は全 6 章から構成される．第 2 章では，本研究の基礎となる音声合成技術について説明する．まずはじめに，現在の音声合成分野の主流である HMM 音声合成技術について，続いて F_0 モデルに基づく韻律生成手法について触れる．第 3 章では，発話スタイルや感情を対象とした F_0 モデルに基づく韻律制御手法および発話スタイル表現に関する先行研究を紹介する．第 4 章では，本研究で実際に用いた韻律生成系および韻律制御手法について説明する．第 5 章では，評価実験を説明し，結果について述べる．第 6 章では本論文のまとめを行い，今後の展望について述べる．

第2章　音声合成のための要素技術

2.1 はじめに

本章では，F_0 モデルに基づく音声合成の枠組みに関わる要素技術を説明する．まず最初に，任意のテキストに対応する音声合成システムの土台となる隠れマルコフモデル (HMM)[3] および HMM 音声合成技術について説明する．続いて，コーパスを作成する上で必要となる言語情報を抽出するための自然言語処理技術について説明する．最後に，人間の発声機構に関する説明を交えながら，基本周波数パターン生成過程モデル (F_0 モデル) の概要を紹介する．

2.2 隠れマルコフモデル

隠れマルコフモデル (Hidden Markov Model, HMM) は，対象をマルコフ過程と仮定し，連続的な信号列を複数の状態間の遷移によって表す確率モデルである．音声信号処理の分野では，音声のスペクトル変化特性を表す統計モデルとして，音声認識，分析，合成などにおいて広く用いられている．HMM は非定常な音声信号の時系列を複数の定常信号源間の遷移という形で表現する．

本研究では，音響パラメータの生成のために，HMM 学習ツールキットである Hidden Markov Model Toolkit(HTK)[4] と，HTK に基づく音声合成ツールである HMM-based Speech Synthesis System(HTS)[5] を使用する．

HMM 音声合成方式 [6] は，HMM をフレームごとの音声特徴量の系列に適用し，音素レベルの音響事象を一つの音響状態として学習する．まず，音声データに対してメルケプストラム分析を行い，各フレームのスペクトル情報を係数ベクトルとして HMM を学習するして音響モデルを得る．音声のケプストラムは，フレームごとに多次元ベクトルとして抽出することができる．この多次元ベクトルは人間が経験的に解釈・法則化することが困難であるため，次元数を設定して HMM で学習する事に大きな利点がある．

合成時には，入力されたテキストを解析した音素列に合わせて音響パラメータ時系列を音響モデルに基づいて出力し，音声を合成する．なお HTS では，MLSA フィルタなど一部の処理において，Speech Singal Processing Toolkit(SPTK)[7] を用いている．

図 2.1 は，音声合成システムの構成を表す．スペクトルモデルと F_0 モデルを独立に学習した場合，両モデル間で境界のずれが生じるため，F_0 の情報をスペクトル情報と合わせて特徴ベクトルとし，HMM の枠組みによって統一的にモデル化する．またモデル学習部では，音声データから得た各分析フレームごとの静的特徴量としてのスペクトルパラメータと F_0 パラメータから動的特徴量 (Δ，Δ^2) を算出し，静的特徴量・動的特徴量を結合したものを学習している．F_0 パターンは，有声区間では連続値をとる 1 次元となり，無声区間では値が定義されず 0 次元となる．この F_0 について，F_0 パラメータを多空間確率分布 HMM(MSD-HMM) でモデル化し，その HMM からパラメータ生成を行うことにより韻律を生成する手法が提案されている [8]．MSD-HMM は F_0 について連続分布 HMM と離散分布 HMM を特別な場合として含むように拡張した HMM である．

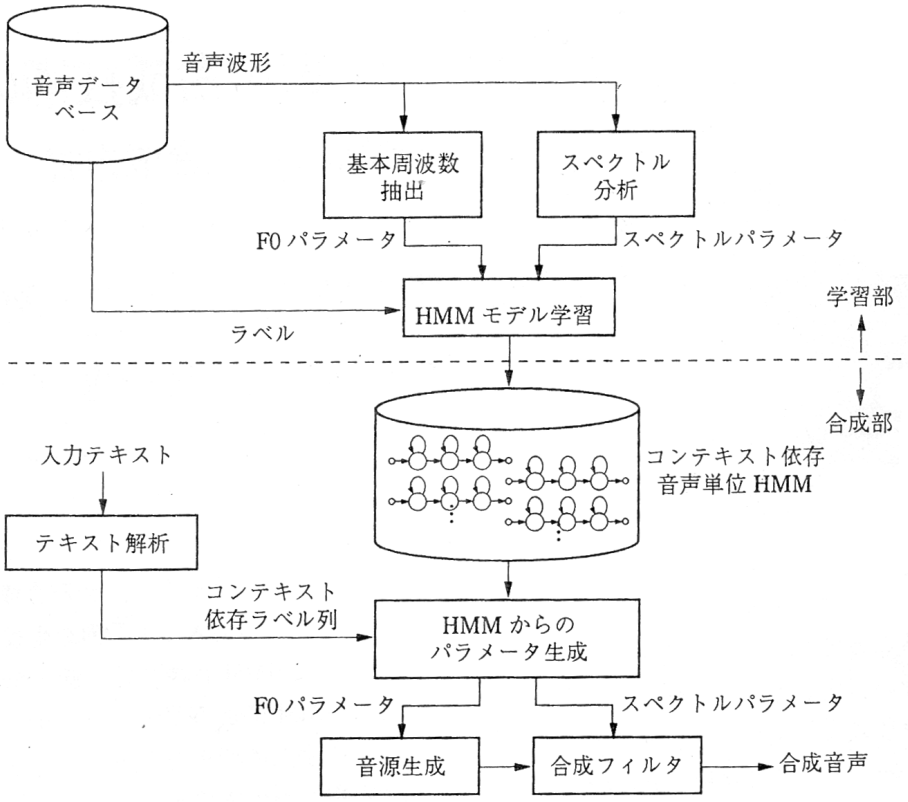


図 2.1: MSD-HMM に基づく合成音声システムの構成

各音響状態の音韻継続長は，HMM の状態継続長を多次元ガウス分布によりモデル化し，合成対象のスペクトルと F_0 に応じてパラメータを推定する．なお，上記のスペクトル情報，F_0，音韻継続長のパラメータは，アクセント型，構文情報，当該・先行・後続音素といった様々なコンテキストの影響を受けるため，コンテキストに依存したモデルを構築する．本研究は，F_0 モデルに基づいて F_0 パターンをモデル化して韻律生成を行うため，F_0 の取り扱いの点で図 2.1 の音声合成システムと異なる．

また，HMM 音声合成系において，小規模の音声資料による発話スタイル制御の研究例として，話者適応技術がある [9]．これは合成系の扱う音韻や韻律の特徴を，適応をかける小規模のデータから抽出した特徴に近づけるものであり，音声合成分野では合成音声の話者性や発話スタイルを変更する時に有効である．この手法は，スペクトル情報と基本周波数の両方について適応をかけることが可能であるが，話者性や発話スタイルといった情報がまとめて適応されてしまうため，同一話者で異なる発話スタイルの合成系を構築するためには適応前の話者本人や声の近い音声提供者から音声を収録する必要があるなど，有効な音声資料の条件が限定されてしまう．この点について，本研究では発話スタイル表現に現れる韻律の差分情報のみを扱うことで，とある話者から抽出した差分情報を任意の話者の音声合成系に適用できる，柔軟かつ拡張性の高い韻律制御系の実現を目指す．

2.3 自然言語処理

2.3.1 形態素解析

自然言語の文において，意味を担う最小の言語単位のことを形態素と呼ぶ [10]．単語は通常，一つまたは複数の形態素によって構成される．入力されたテキストに対して，それを構成する形態素を同定する処理は形態素解析とよばれる．形態素解析の広義的な役割には，次のようなものが挙げられる．

- 文の形態素列への分割 (分かち書き処理)
- 形態素への品詞の付与
- 形態素の語形変化の解析

本研究では，入力されたテキストに対する韻律パラメータを推定する上で，入力テキストを構成する形態素情報を用いる．形態素情報の抽出においては，日本語形態素解析システム茶筌 [11] を使用した．図 2.2 は，その解析結果例である．各品詞分類に属する各形態素について，その形態素の見出し語，読み，品詞，活用型などを記述した形態素辞書を作成する．出力形式は様々にカスタマイズすることができる．ここでは，左から漢字仮名交じり表記，カタカナ表記，読み，品詞，活用型，活用形，アクセント型，アクセント結合様式・アクセント価という順で出力している．

```
あらゆる  アラユル  アラユル  連体詞          無活用      基本形-一般  accent=3                          あらゆる
現実      ゲンジツ  ゲンジツ  名詞-一般       無活用      基本形-一般  accent=0:accent_con=C2            現実
を        ヲ        オ        助詞-格助詞-一般 無活用      基本形-一般  accent_con=名詞%F1                を
すべて    スベテ    スベテ    名詞-副詞可能   無活用      基本形-一般  accent=1:accent_con=C1            すべて
自分      ジブン    ジブン    名詞-一般       無活用      基本形-一般  accent=0:accent_con=C2            自分
の        ノ        ノ        助詞-格助詞-連体化 無活用   基本形-一般  accent_con=名詞%F1                の
ほう      ホウ      ホー      名詞-一般       無活用      基本形-一般  accent=1,0:accent_con=C3          ほう
へ        ヘ        エ        助詞-格助詞-一般 無活用      基本形-一般  accent_con=名詞%F1                へ
ねじ曲げ  ネジマゲ  ネジマゲ  動詞-自立       一段-一般   連用形-一般  accent=4                          ねじ曲げ
た        タ        タ        助動詞          助動詞タ    基本形-一般  accent_con=動詞%F1,形容詞%F2@-1   た
の        ノ        ノ        助詞-準体助詞   無活用      基本形-一般                                    の
だ        ダ        ダ        助動詞          助動詞ダ    基本形-一般  accent_con=名詞%F1,動詞%F1        だ
```

図 2.2: 茶筌の出力結果の例

2.3.2 構文解析

構文解析とは，文がどのような語，句，節によって成立しているかを調べる処理である[12]．語，句，節などは構成素と呼ばれ，これらが適切な配置で結合することで文の構文構造が形成される．構文解析は一般に，形態素解析により得られた語の列に対して適応される．語レベルの解析は形態素解析で行い，文レベルの解析は構文解析で行う．こうした処理の切り分けがなされるのは，形態素の構造は比較的複雑ではなく局所的に解析することができるため，より効率的な処理を行うほうがよいという考えからである．日本語構文解析システム KNP[13] は，日本語形態素解析システム JUMAN[14] によるテキストの形態素解析結果を元に，文節境界や文節間の係り受けの構造を推定する．図 2.3 はその出力結果の例である．図 2.4 は，構文解析から得た係り受けの距離と境界コードである．

```
# S-ID:1
* 1D <文頭><用言:弱><係:連体><区切:0-4>
あらゆる あらゆる あらゆる 連体詞 11 * 0 * 0 * 0 NIL <文頭><自立>
* 5D <ヲ><助詞><体言><係:ヲ格><区切:0-0>
現実 げんじつ 現実 名詞 6 普通名詞 1 * 0 * 0 NIL <漢字><自立>
を を を 助詞 9 格助詞 1 * 0 * 0 NIL <付属>
* 5D <数量修飾><数量><副詞><用言:弱><係:連用><区切:0-4>
すべて すべて すべて 副詞 8 * 0 * 0 * 0 NIL <数量修飾><自立>
* 4D <述並終点><助詞><体言><係:ノ格><区切:0-4>
自分 じぶん 自分 名詞 6 普通名詞 1 * 0 * 0 NIL <漢字><自立>
の の の 助詞 9 接続助詞 3 * 0 * 0 NIL <付属>
* 5D <外の関係><へ><助詞><体言><係:ヘ格><区切:0-0>
ほう ほう ほう 名詞 6 副詞的名詞 9 * 0 * 0 NIL <自立>
へ へ へ 助詞 9 格助詞 1 * 0 * 0 NIL <付属>
* -1D <文末><用言:強:動><レベル:C><区切:5-5><ID:(文末)><提題受:15>
ねじ曲げた ねじまげた ねじ曲げる 動詞 2 * 0 母音動詞 1 タ形 8 NIL <自立>
のだ のだ のだ 助動詞 5 * 0 ナ形容詞 21 基本形 2 NIL <文末><付属>
EOS
```

図 2.3: KNP による構文解析結果の例

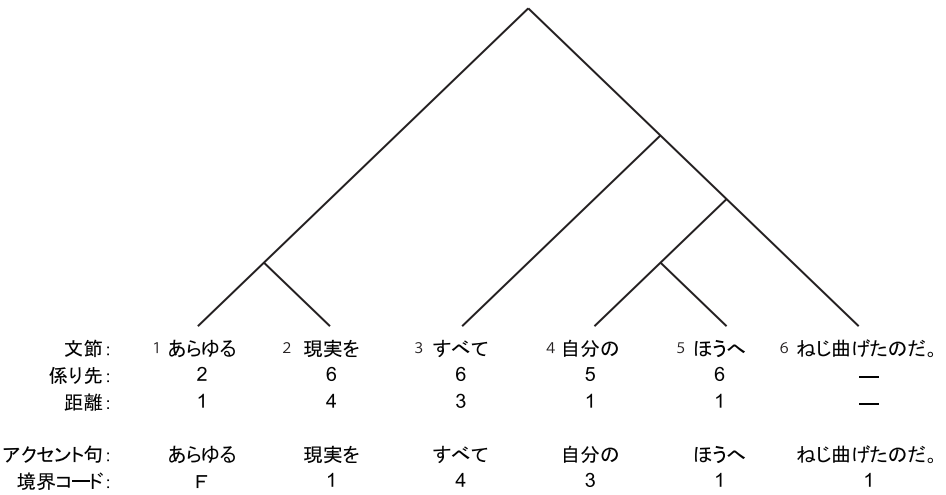


図 2.4: 境界コード

2.3.3 日本語アクセント結合規則

日本語のアクセントは，モーラ単位でのピッチ周波数の高低の変化によって表現される．このピッチ周波数の高い部分と低い部分との相対的な聞こえの差が安定して現れるのが日本語のアクセント表現の特徴と言える．単語のアクセントは，文字列から推定されるものではなく，単語に固有のものである．表 2.1 は，モーラ数の異なる単語に助詞 “が” を結合した場合の日本語のアクセント型である [15]．表 2.1 のそれぞれの項目において，ピッチ周波数が高い状態が続いている最後のモーラをアクセント核と呼ぶ [16]．アクセント結合に関する記述を分析・整理して，結合アクセント様式・結合アクセント価という形で規則化する方法が提案されている [17]．喜多らの研究では，複数の語句が結合した場合に起こるアクセント型変化について記述しており [18]，本研究でもこのアクセント結合規則を採用する．

型の種類		モーラ数				
		1	2	3	4	5
平板式		ハガ（葉）	ミズガ（水）	サクラガ（桜）	オハナミガ（お花見）	トナリムラガ（隣村）
起伏式	尾高型		ヤマガ（山）	ヤスミガ（休み）	イモートガ（妹）	モモノハナガ（桃の花）
	中高型			オカシガ（お菓子）	ミズウミガ（湖）	ニワカアメガ（にわか雨）
					ノミモノガ（飲み物）	ヤマザクラガ（山桜）
						オカーサンガ（お母さん）
	頭高型	キガ（木）	ハルガ（春）	ミドリガ（緑）	サンガツガ（3月）	オツキサマガ（お月様）

表 2.1: アクセント型

2.4 F_0 パターン生成過程モデル

2.4.1 人体における声帯制御の構造

人間は声帯の長さを変化させることで，声の高さを調節している．声帯の長さを変化させる仕組みは，質量・ばね系の運動を伴った力学系となる．すなわち，喉頭の制御は，質量・ばね系の運動を生じさせ，それが喉頭部分の動きから声帯の長さの変化に反映される．声帯のすぐ脇にある声帯筋に力をかけて引き伸ばすとき，そのばね定数は張力にほぼ比例することが実験的に知られている．さらに，甲状軟骨の平行移動と回転とが声帯の伸びを決定し，その動きがそれぞれが臨界制動二次系にモデル化されるので，声の高さの対数パターン (対数 F_0 パターン) にはそれらが二つの成分の和として表されるのである．

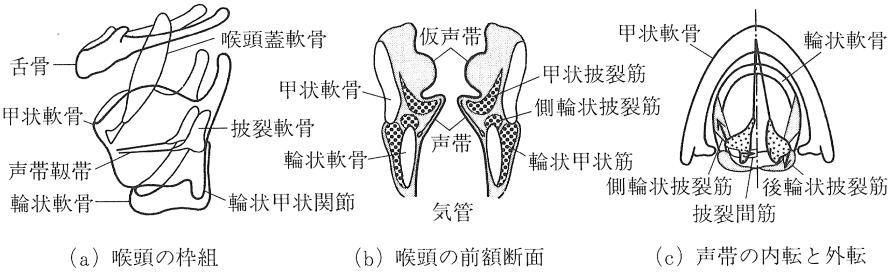


図 2.5: 喉頭の枠組みと内咽頭筋
(文献 [19] より引用)

喉頭 (larynx) の枠組みは，図 2.5 のように，舌骨と三種類の軟骨 (甲状軟骨 thyroid cartilage，輪状軟骨 cricoid cartilage，披裂軟骨 arytenoid cartilage) からなる．甲状軟骨は，その下にある輪状軟骨と左右 1 対ずつある関節を作る．その関節運動は，主に回転 (rotation) であるが，若干の並進運動 (translation) も生じる．図 2.6 は，声の高さ，すなわち声帯振動の基本周波数 (F_0) の調節機構を表す [19].

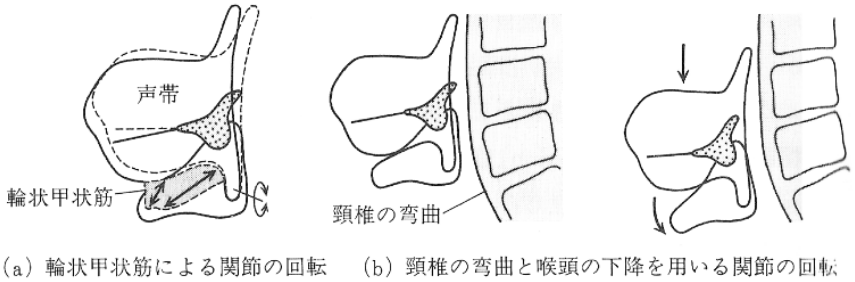


図 2.6: 声の高さの調節機構
(文献 [19] より引用)

このように声帯が引き伸ばされることよって F_0 が変わり，アクセントやイントネーションのような音声の韻律を構成する．輪状甲状関節の回転が起こり声帯が伸張すると，声帯の張力が増大し声が高くなる．輪状甲状筋は直接この関節の回転を引き起こす．咽頭が下降すると輪状軟骨は頚椎の湾曲に沿って移動し，関節の逆回転が生じ声帯が短くなる．

2.4.2 F_0 モデルの概要

F_0 モデルとは，対数 F_0 の時系列パターンをフレーズ成分とアクセント成分という二種類の成分の重畳によって表現するモデルである．図 2.7 に F_0 モデルによる $\log F_0$ の表現の様子を示す．フレーズ成分とアクセント成分は，それぞれインパルス信号，ステップ信号の指令信号によって生起する．フレーズ指令によって生起する $\log F_0$ パターンは，句頭から句末に向かって緩やかに下降する “へ” の字型の山となる．アクセント指令の場合は，個々の単語又は単語の連鎖に対応して，急峻な立ち上がり，立ち下がりをする起伏となる．この二つの成分を周波数の基底値 (F_b) に重畳することで，図 2.7 右側のような一連の F_0 パターンを得る．

第 2.4.1 項で説明した喉頭の動きにおいて，フレーズ成分は甲状軟骨の平行移動運動に対応し，アクセント成分は回転運動に対応している．つまり，F_0 モデルは喉頭の生理的・物理的特性に基づいて声帯振動数制御に関する部分をフレーズ制御機構・アクセント制御機構として分離して表現したものといえる [20, 21]．このため，各指令は発声機構の動き，ひいては言語情報に密接に関係していると考えられ，声帯の伸縮と振動数の関係から物理的にも妥当性が示されている．また，HMM 音声合成において F_0 の動きを文節的特徴とともにフレーム単位で推定しているのに対し，F_0 モデルでは少数の指令によって文全体の F_0 パターンを記述しており，広い時間領域に渡る F_0 の自然な動きを表現できる．

以上の点から，コーパスの規模が小さいものであっても，自然性の破綻が少ない統計的な韻律生成が比較的容易に実現できるモデルであると言える．

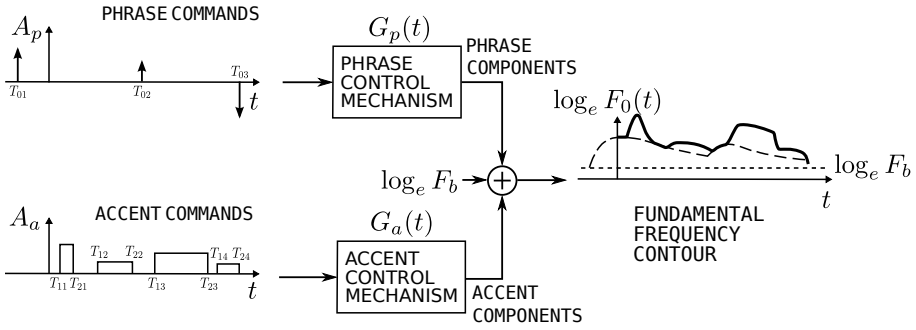


図 2.7: F_0 モデルによる $\log F_0$ の記述

2.4.3 フレーズ成分とアクセント成分

一般的な発話において，フレーズ成分は声帯振動の開始よりも 200〜400[ms] ほど先立って上昇が始まる．そして最大値に達してから，緩やかに下降して基底値 F_b に漸近する．なお，意識的な sp の挿入や発話の終端の場合は，急激な下降が起こる．

一方，アクセント成分は，個々の単語または連続した単語に付随して発生し，フレーズ指令によって生起した大きな起伏に上乗せする形で局所的な起伏として現れる．アクセントの頂点として相対的に高い拍となる発音の少し前から緩やかに上昇した後，途中で急激に上昇する．その後，そのアクセント指令が属するフレーズ指令の挙動に合わせて緩やかに下降する．この時，図 2.1 の平板式などのように高い拍が続く場合には起伏の高さを保つ．一つのアクセントの中で，高い拍から低い拍へ移る時は上昇と逆の挙動を示し，低くなる拍の発音に先行して緩やかな下降を始め，途中で急激な下降になり，フレーズ成分のみの状態に戻る．

これら二つの成分の挙動を定式化し，F_0 モデルによる F_0 パターン表現を数式化することを考える．フレーズ成分の挙動は，質量とバネ定数とを持つ 2 次の力学系が瞬間的な外力 (インパルス引力) を受けた場合の運動によく似ている．また，アクセント指令の挙動は，質量とバネ定数とを持つ 2 次の力学系がある時間中に持続的に一定の外力 (ステップ入力) を受けた場合の運動とよく似ている．このため，フレーズ成分およびアクセント成分を用いて，F_0 パターンを次式のように近似することができる．

$$log_e F_0(t) = log_e F_b + \sum_{i=0}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj}(G_a(t - T_{1j}) - G_a(t - T_{2j})) \qquad (2.1)$$

フレーズ成分を比較的時定数の長い線形 2 次系のインパルス応答，アクセント成分を比較的時定数の短い線形 2 次系のステップ応答に近似できるものとして，対数 F_0 パターンの変化をそれらの和として表している．ここで，式 2.1 におけるそれぞれの変数の内容は下記の通りである．

- F_b は F_0 パターンの基底値
- I，J はそれぞれ文中のフレーズ指令とアクセント指令の数
- A_{pi} は i 番目のフレーズ指令の大きさ
- A_{aj} は j 番目のアクセント指令の大きさ
- T_{0i} は i 番目のフレーズ指令が生起する時点
- T_{1j} は j 番目のアクセント指令の立上り時点
- T_{2j} は j 番目のアクセント指令の立下り時点

正のフレーズ指令のインパルスは文頭・文中のフレーズの先頭に生起して上昇させる．負のフレーズ指令のインパルスは文の終わりに生起してそれまでのフレーズ成分を下降させる役割を持つ．また，アクセント指令は正の方形波として個々の単語または単語連鎖ごとに生起してアクセント成分を生成する．

$G_p(t)$，$G_a(t)$ は，それぞれフレーズ成分に相当する系のインパルス応答，アクセント成分に相当する系のステップ応答であり，次式 2.2，2.3 によって表される．

$$G_p(t) = \begin{cases} \alpha^2 exp(-\alpha t) & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{2.2}$$

$$G_a(t) = \begin{cases} min[1 - (1 + \beta t)exp(-\beta t), \gamma] & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{2.3}$$

ここで α および β はそれぞれの制御機構の固有角周波数を表す．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．これらは話者ごと，発話ごとの変動は小さいため，表 2.2 のとおりに定数として定める．

α [rad/s]	β [rad/s]	γ
3	20	0.9

表 2.2: F_0 モデルにおける定数 α，β，γ の値

2.5 決定木

本研究では，F_0 モデルパラメータの推定のために決定木を作成し，入力されたテキストに応じて各パラメータの推定を行う．決定木の生成のアルゴリズムにはCART(Classification and Regression Tree)[22] を用いる．実装には，The Edinburgh Speech Tools Library[23] によるCARTに基づく決定木構築プログラムであるwagonを使用する．

決定木の例を2.8に示す．簡単のため，ここでは2変数の場合で説明する．決定木によるパターン識別は，パターンの分岐の条件となる属性の値について質問していくことで進行する [24]．ルートノードから始まり，各質問への答えに基づいて子孫のノードへと移動し，これを質問の無い葉ノードに至るまで繰り返す．葉ノードはカテゴリのラベルを持っており，テストパターンには，たどり着いた葉ノードのカテゴリが割り当てられる．

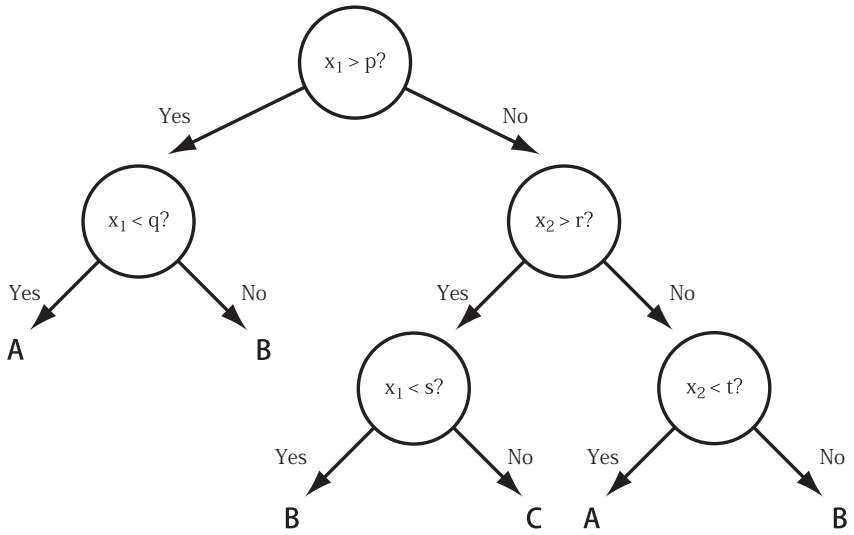


図 2.8: 決定木の例

数値データについて，各ノードの質問が “$x_i \leq x_{is}$ であるか？” という形である場合にデータがどのように分割されるかを視覚化した例を図 2.9 に示す．ノードでの各決定結果により，座標軸に垂直な超平面の決定境界と決定領域 R1, R2 に分割される．十分に大きい木を用いれば，どのような決定境界も任意の精度で近似することができる．決定木は，ニューラルネットワークなどの他の学習アルゴリズムとは異なり，構築されたルールを人が解釈するのが容易であるという特徴を持つ．

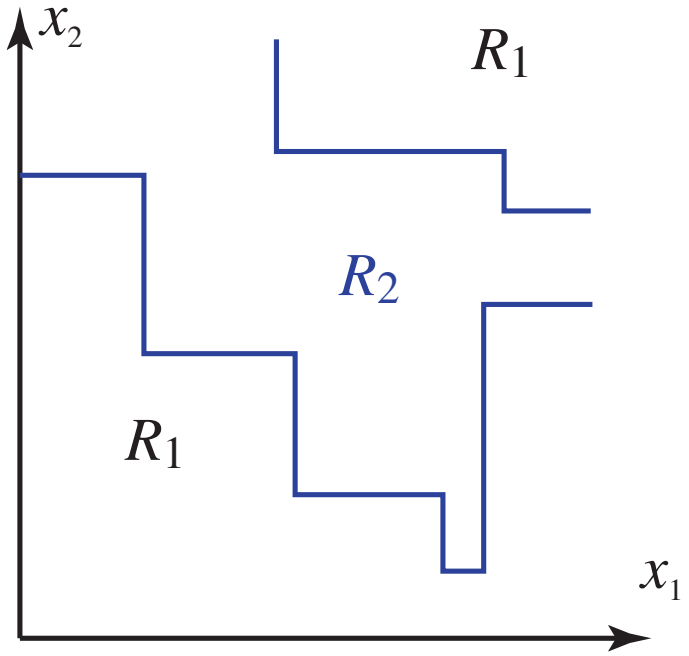


図 2.9: 決定木による決定領域の例 (2 カテゴリの場合)

第3章　先行研究

3.1 はじめに

本章では，F_0 モデルに基づく韻律制御に関連する先行研究を説明する．第一に，日本語の統語上の単位および韻律上の単位について触れ，F_0 モデルパラメータと言語情報との結びつきについて説明する．第二に，コーパスベース韻律生成の根幹となる韻律コーパスの構築について説明する．第三に，任意のテキストから韻律を生成するコーパスベース韻律生成の枠組みについて説明する．最後に，コーパスベース韻律生成を含め F_0 モデルに基づく F_0 パターンを対象とし，任意のテキストに対応するコーパスベース韻律制御手法について説明する．

3.2 日本語の統語上の単位，韻律上の単位

表 3.1: 日本語の統語上の単位

単位	意味
文節	「自立語＋付属語 (任意の個数)」の最長一致からなる単語のまとまり.
ICRLB	右枝分かれ境界で前後を区切られ，かつ左枝分かれ境界のみを含む単語連鎖. すなわち，修飾関係にある単語間の境界であり，一つの ICRLB は意味的に一つのまとまりをなしている. 例：「その上/3 月 20 日には/東京の地下鉄の車内で/ 毒ガスのサリンがまかれるという/前代未聞の怪事件も起き」
節	他の語句を修飾しない述語と，それを直接的・間接的に修飾する単語のまとまり. 例 1：「あなたはあなたの職場に行き/私は私の職場に行きます」(境界あり) 例 2：「あなたがやめた作業を私が引き継ぎます」(境界なし)
文	句点や疑問符で区切られた区間.

表 3.2: 日本語の韻律上の単位

単位	意味
韻律語	一定のアクセント型を示す音素連鎖として定義される．一つのアクセント成分に対応しており，韻律上の最小単位である．アクセント句と呼ばれることもある.
韻律句	一つのフレーズ成分の始点から次のフレーズ指令の始点，または音声の休止点までに含まれる韻律語の連鎖として定義される.
韻律節・韻律文	音声の休止点で区切られた韻律句の連鎖を表す．韻律節と韻律文は，最後にフレーズ成分のリセットが行われるか否かによって区別する．リセットが行われる場合が韻律文であり，行われない場合を韻律節であると定義する.

日本語の統語上の単位として，文節，ICRLB(Immediate Constituent with Recursively Left-Branching structure)，節，文があり，表 3.1 のように単位の規模が定められている．また，韻律上の単位としては，韻律語，韻律句，韻律節，韻律文の 4 つがあり，同様にして表 3.2 の通りに意味する単位の規模が異なる．この二種類の単位の間には，文節と韻律

語，ICRLB と韻律句，節と韻律節，文と韻律文，という形で 1 対 1 の対応関係を持つことが多い (図 3.1).

なお，表 3.2 では，フレーズ成分のリセットの有無によって韻律節と韻律文の区別をつけているが，実際の発話において文末が不明瞭に発声された場合やフレーズ成分のリセットが不明確な場合など，どちらの韻律単位であるかを判別できないこともある．

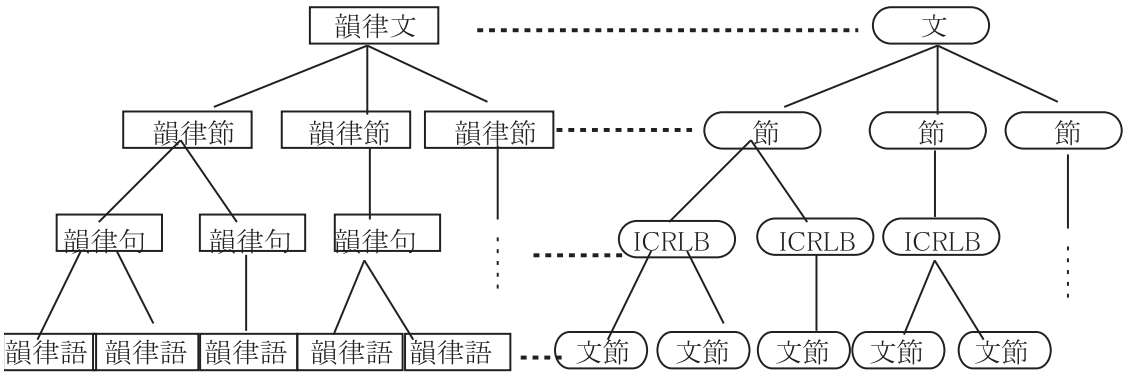


図 3.1: 韻律的特徴における区分と統語情報による区分の対応関係

3.3 韻律コーパス

コーパスベース手法におけるデータベース収集

F_0 モデルパラメータの統計的推定のために，言語情報の条件をバランスよく網羅した例文と，それを実際に読み上げた収録音声の組から抽出した，推定対象となるパラメータの情報を集めたデータベースをもとにして学習を行う．韻律パラメータを対象としたこのデータベースを韻律コーパスと呼ぶ．

推定の精度を十分なものにするためには，十分な量の韻律コーパスを準備する必要があるが，膨大な量の音声資料を一つ一つ手動でラベリング行うためには，膨大な労力を要する．

任意の人の音声データを少ない手順で学習用データとして利用できるようにすること，またデータベースの規模を削減し効率的な推定系を実現することは，今後の多様な音声合成システムの実現可能性や拡張性に大きく関わる．この点について，比較的小規模のデータベースから効果的な韻律制御系を検討する事を設定することが本研究の目的の一つとなる．

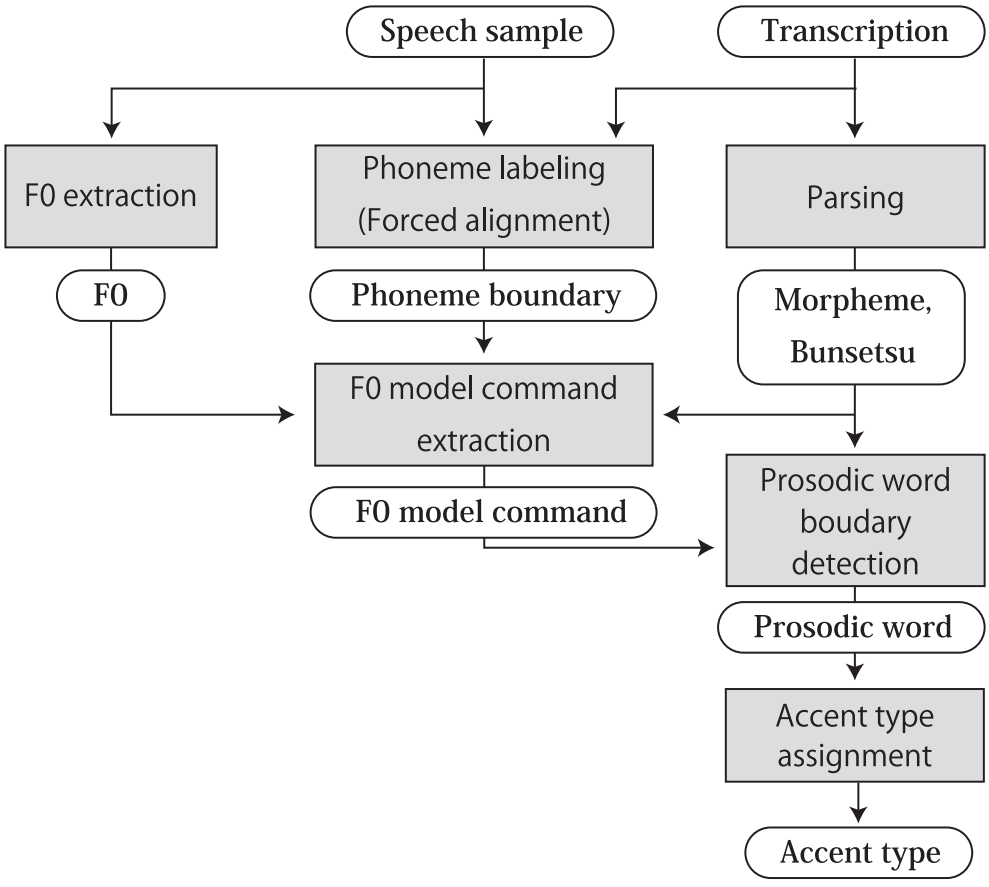


図 3.2: 韻律コーパス構築の流れ

韻律コーパス構築の流れ

韻律コーパスの作成のためには，F_0 モデルパラメータの学習・推定の単位である韻律語の境界を定める必要がある．そこで，入力した漢字仮名混じり文を韻律語単位へと分割する．以下に，具体的な処理の流れを記す．この処理を異なるテキストに基づく学習用データの全ファイルに対して行い，韻律コーパスを構築する．

1. 音声ファイルから基本周波数値を抽出し，F_0 モデルの枠組みに沿って分析して F_0 モデルパラメータファイルを得る．

2. 音声ファイルと，漢字仮名交じり文から得た発音ファイルから Julius を利用して，音素アライメントファイルを得る．本研究では，漢字仮名交じり文から読み仮名を得るには，茶筌の出力を利用する．

3. 漢字仮名交じり文から，茶筌を用いて形態素，品詞情報を得るとともに，JUMAN と KNP を用いて文節，統語情報を得る．この内容をまとめて言語情報ファイルとする．

4. 音素アライメントファイルと言語情報ファイルを参照して，実際の音声の時間情報と個々の形態素の対応関係を調べる．

5. 言語情報ファイルと F_0 モデルパラメータファイルを参照し，韻律語境界の抽出規則を用いて形態素系列を韻律語系列へと変換する．

6. アクセント結合規則に基づき，各韻律語のアクセント型を決定する．アクセント核をなす各モーラ母音開始時点とアクセント指令の立ち下がり位置との差分を T_2 として定義する．また，アクセント指令の生起タイミング T_1 を韻律語の先頭モーラの母音開始時点からの差分とする．

7. フレーズ指令はアクセント指令の間に 0 または 1 個存在すると仮定して検索を行う．存在した場合，時間的に後続する韻律語の先頭モーラの母音開始時点からの差分をフレーズ指令の時点情報 T_2 と定める．

3.4 F_0 モデルに基づくコーパスベース韻律生成

3.4.1 F_0 モデルパラメータの自動抽出

韻律コーパスを構築する上で，手作業でラベリングを行うことに膨大な労力が必要となることは第 3.3 項において述べた．この点を受けて，入力音声から F_0 モデルパラメータを抽出する手法がこれまでに検討されている．

F_0 モデルパラメータ自動抽出ツール AXPF[25] は，F_0 パターンファイルと音声ファイルから，F_0 モデルパラメータを自動抽出するプログラムである．F_0 パターンファイルには，各フレームごとの F_0 の値と，有声区間であるか無声区間であるかを記述している．音声ファイルは，サンプリング周波数 10kHz または 16kHz，量子化ビット数 16bit，Little Endian の RAW ファイルと，非圧縮，サンプリング周波数 48kHz の WAV 形式のファイルに対応している．

行番号	内容				意味
1	***/***.PAC				ファイルのパスとファイル名
2	100				フレーム長
3	03.12.24				最終更新日
⋮	-				-
7	241				総フレーム数
8	2				フレーズ指令総数
9	4				アクセント指令総数
10	60.909				F_0 の基底値 (F_b)
11	0.01				シフト長
⋮	-				-
14	0.001384 → 0.001123				AbS 前の F_0MSE → AbS 後の F_0MSE
⋮	-				-
21	0.2458	2.3939	0.5055	3	T_0 ，T_0'，A_p，α
22	0.5277	2.3939	0.5056	3	
23	0.0625	0.2222	0.6622	20	T_1，T_2，A_a，β
24	0.7712	1.4757	0.3689	20	
25	1.6553	1.9512	0.1576	20	
26	2.0765	2.1951	0.213	20	

表 3.3: PAC ファイルの例

表 3.3 に，AXPF が出力する F_0 モデルパラメータのファイル (PAC ファイル) の例を記す．音声ファイルの開始点を時刻 0 とし，フレーズ指令，アクセント指令について，指令の大きさや時点情報を記述している．なお，PAC ファイルは文全体の F_0 パターンを記述しているため，指令の時点情報として絶対的な値が記述されているが，テキストからの韻律生成においては，個別の指令の持つ時点情報として,「文節頭モーラから○○秒前」といったように相対的なものを扱う．入力されたテキストに応じて音韻継続長を決定し，指令の生起の基準点となるモーラの開始時刻が決定してから，その開始時刻と指令の持つ相対的な時点情報によって生起時刻を定めることになる．

本研究では，直接的には AXPF を対象とした研究を行ってはいないが，音声試料から F_0 モデルパラメータを抽出・編集する際の初期値を得るためにパラメータの自動抽出を適宜行う．F_0 モデルの編集には，FujiParaEditor[26] を利用する．FujiParaEditor は入力音声から F_0 モデルパラメータを抽出し，パラメータの編集および編集後の韻律を用いた音声の再合成ができる GUI である (図 3.3)．FujiParaEditor では，F_0 パターン抽出に Praat[27]，F_0 モデルパラメータ情報の入出力に PAC ファイルを利用している．

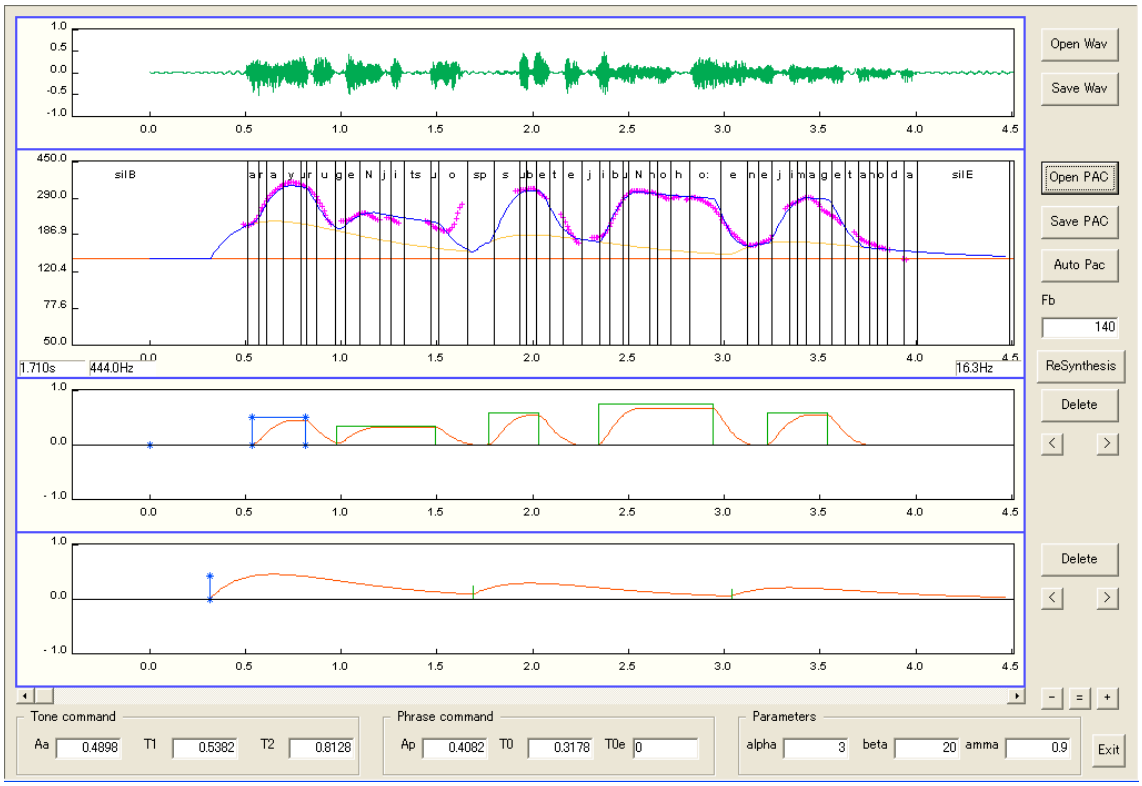


図 3.3: FujiParaEditor

3.4.2　テキストからの韻律生成システム

F_0 モデルパラメータの推定

表 3.4 は，推定の対象となるパラメータの一覧である．ここでの指令の時点情報は，各韻律語中で基準とする時刻からの相対距離 (タイミング) を用いて表し，フレーズ指令の生起タイミング T_0，アクセント指令の生起タイミング T_1 は韻律句の第一モーラの母音開始時刻 (図 3.4 の「あ」の開始時刻)，アクセント指令の終了タイミング T_2 はアクセント核モーラの終了時刻 (図 3.4 の「ゆ」の終了時刻) を基準としている．

表 3.4: F_0 モデルパラメータ推定項目

出力項目	カテゴリ数
先頭のフレーズ指令有無 PF	2 値
フレーズ指令の大きさ A_p	連続値
フレーズ指令のタイミング T_0	連続値
アクセント指令の大きさ A_a	連続値
アクセント指令の生起タイミング T_1	連続値
アクセント指令の終了タイミング T_2	連続値

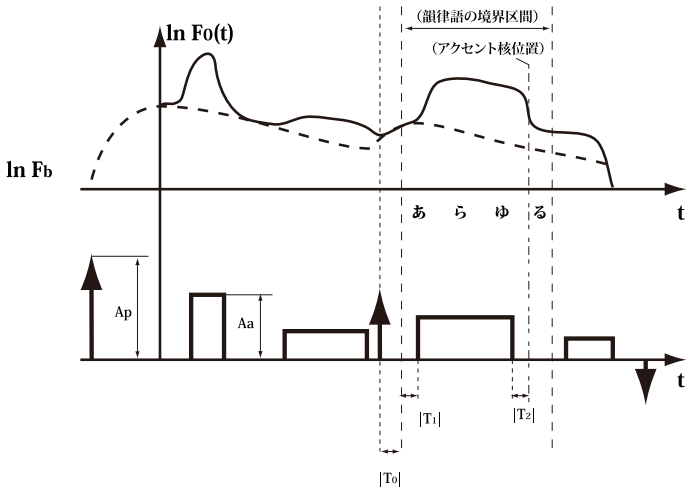


図 3.4: 推定の対象となるパラメータの概略

入力されたテキストに応じて，韻律コーパスを元に F_0 モデルパラメータを推定する流れを説明する (図 3.5).

第一に，入力された漢字仮名混じりテキストに対して形態素解析と構文解析を行い，言語情報を抽出するとともに，音韻継続長を決定する．第二に，フレーズ成分を推定する．文節境界ごとに PF を推定し，続いて T_0，A_p を推定する．ここで韻律句レベルまでの情報が決定する．第三に，韻律語境界を推定する．一つの韻律語は一つのアクセントと 1 対 1 対応することが多く，文節レベルまでの大まかな韻律構造が決定する．第四に，アクセント核の位置を推定する．アクセント核モーラの終了時刻が決定し，韻律語ごとのアクセントの詳細な型が決定する．最後に，T_1，T_2，A_a を推定し，アクセント成分を決定する．

なお，フレーズ指令の生起に関して次の前提条件を設定している．

- フレーズ指令は必ず文節境界の前に存在する
- 文頭，もしくは sp に続く文節境界の場合，文節境界から-300〜-100msec の区間に生起する
- 上記の条件に当てはまらない場合は，文節境界から-100〜0msec の区間に生起する

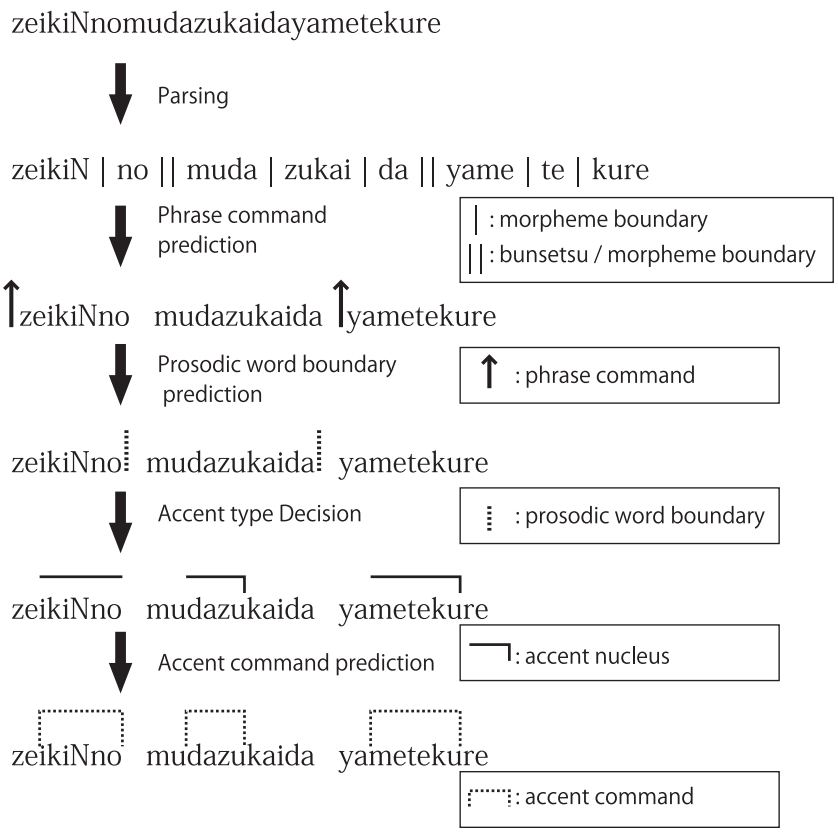


図 3.5: テキストからの F_0 モデルパラメータ推定の流れ

そのほかのパラメータの取り扱い

F_0 モデルパラメータの他に韻律生成に関わるパラメータとして，第 3.4.2 目のパラメータ推定における第一工程として挙げた音韻継続長およびショートポーズがある．音韻継続長は発話速度に直接的に影響するため，韻律生成において重要なパラメータである．また，ショートポーズ (休止，sp) とは，二つの韻律句の間など，大きなまとまりのある韻律単位の間に挟まれる無音区間であり，発話を適宜分割することで聞き手の発話理解を助け，可聴性を高める働きがある．標準的な話者の休止の位置と長さは，ともに局所的な文構造との間に高い相関がある [28]．HMM 音声合成システムにおいても音韻継続長の推定系が実装されているが，前述の F_0 パラメータ推定と同様の手法で継続長を推定する手法が越智らによって検討されている [29]．F_0 モデルパラメータと同じ情報に基づいて総合的に韻律を制御することが可能となる．

3.4.3 コーパスベース感情音声生成

コーパスベース韻律生成において，パラメータ推定の際の決定木への主な入力項目は関連の強い言語情報や韻律パラメータであるが，目的に応じてそれ以外の情報も入力項目として導入することが可能である．広瀬らの研究においては読み上げ調の音声に感情 (怒り，喜び，悲しみ) を込めた音声も併せて決定木を構築し，感情の種類なども入力項目に加えて一つのコーパスから目的の感情音声を合成する手法が提案されている [30]．この研究では，特に怒りの音声について効果的な韻律生成ができることが報告されている，

感情という発話スタイルを表現する点では本研究と同様の目的によるものであるが，状況ごとの指令を直接学習するか，差分値を学習するかという点で異なっている．またこの感情音声合成においては，それぞれの感情について 600 文前後の音声試料を収集しており，相応のコストがかかっている．この点に関しても改善案を提示することが本研究の目的の一つとなっている．

3.5　多様な韻律表現のためのコーパスベース韻律制御

3.5.1　差分推定による韻律制御

第 3.4.2 節の韻律生成系に加え，生成された F_0 モデルパラメータを修正する形で制御し，異なる非言語情報，パラ言語情報を持つ韻律を生成する枠組みについても研究を行っている．この韻律制御が比較的小規模の音声データに基づいて実現すれば，潤沢に存在する既存の読み上げ調の音声のデータベースを活用して効率的かつ高い拡張性をもって多様な韻律を表現することが可能になる．

韻律制御の大まかな流れとして，同一の話者による読み上げ調の音声および，目的の韻律的特徴を持つ音声を F_0 モデルに基づいて分析し，F_0 差分値を抽出する．続いて，コーパスベース韻律生成手法と同様にして構築された決定木を元に差分値を推定し，その出力結果を制御対象の F_0 パターンに適用する．

3.5.2　差分情報に基づくコーパスベース焦点制御

先行研究として，読み上げ調の音声の韻律を修正し，任意の文節に焦点を与える手法が越智らによって提唱されている [31]．この手法は文節境界におけるフレーズ指令の生起フラグ PF と大きさ A_p，アクセント指令の大きさ A_a を主な制御対象としており，パラメータの差分値は読み上げ調の音声と特定の文節に焦点を置いた音声から差分値を学習している (図 3.6)．

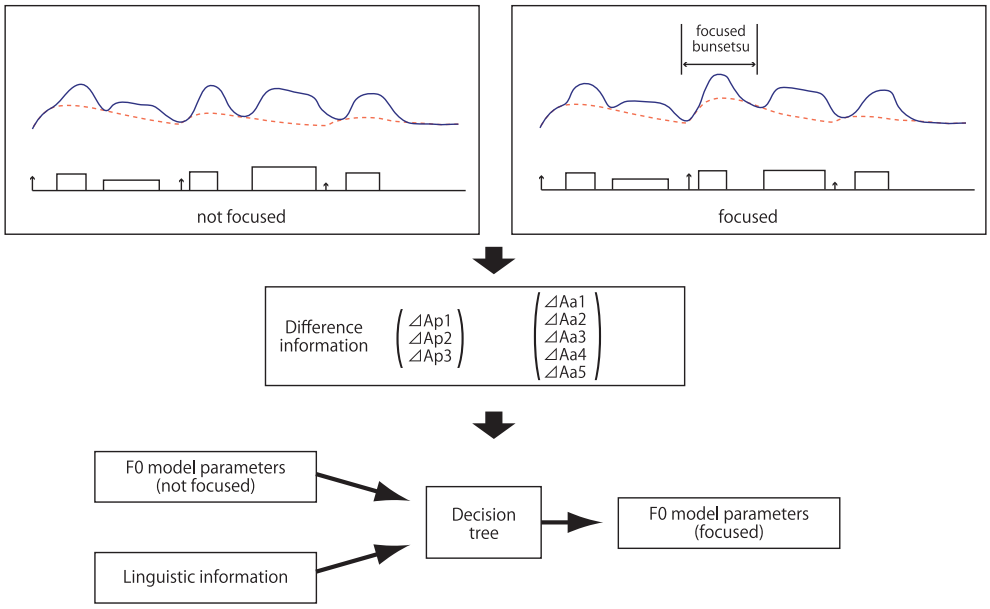


図 3.6: 焦点制御のための差分推定

図 3.7 の F_0 パターンの点線部分に示したように，フレーズ成分の形状は，300msec ほどかけて大きく山を作り，緩やかに下降している．アクセント指令を生起させる時，アクセント成分がフレーズ成分のどの部分に上乗せされるかによってその挙動が大きく異なる．焦点を置く文節の直前には sp を挿入し，フレーズ指令を立ち上げるとともに大きなアクセント指令を設定する．これにより文節頭の F_0 は急峻に上昇するため，焦点のおかれた韻律語として周囲の韻律語との差異化をはかることができる．

差分学習に用いている音声試料は，どの文節にも焦点を置いていない音声と，特定の文節に焦点を置いた音声である．任意の文節に焦点を与えられるように学習するため，テキスト内の文節数に応じて複数個の焦点のパターンを用意し，焦点のない音声との差分を抽出している．結果として，ATR503 文のうち 50 文程度と比較的少量のデータによって有効な焦点制御ができる事が報告された．

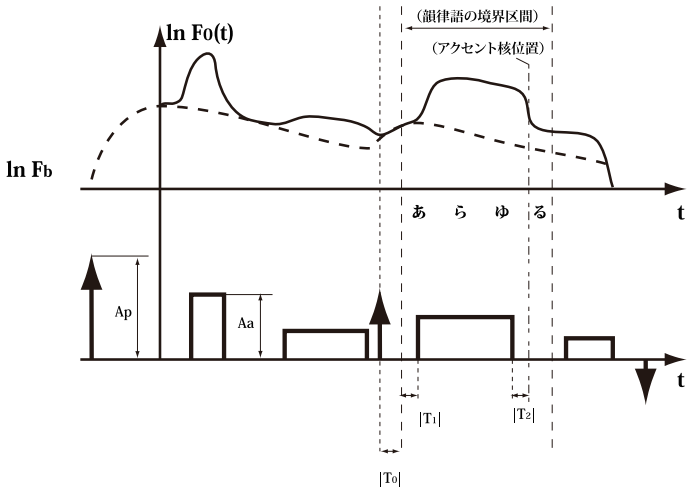


図 3.7: 推定の対象となるパラメータの概略 (再掲)

第4章　コーパスベース発話スタイル制御

4.1 はじめに

本章では，本研究における具体的な実験系の枠組みを説明する．

まず最初に，発話スタイル表現を目的とした差分情報に基づくコーパスベース韻律制御手法について述べる．続いて，任意のテキストに対応して差分を推定するパラメータの説明および決定木への入力項目を説明する．最後に，音声試料の分析の条件を示し，生成された決定木および F_0 パターンについて考察を行う．以後，読み上げ調の発話スタイルを平静と表記する．

4.2 発話スタイル制御を伴うコーパスベース韻律生成系

4.2.1 差分情報の韻律コーパス構築

本研究では，韻律パラメータの差分情報を推定するための決定木を構築し，推定結果の値を平静の発話スタイルの韻律パラメータに反映して目的の発話スタイルを表現することを目指す．決定木を構築する要素 (決定木における質問の対象となる入力項目) は，“推定対象に関連の深い言語情報 (品詞，活用形など)”，“当該パラメータ周辺の韻律パラメータ (指令の大きさ，生起の有無など)”，“平静の発話スタイルにおける当該パラメータ” の三種類に大別できる．

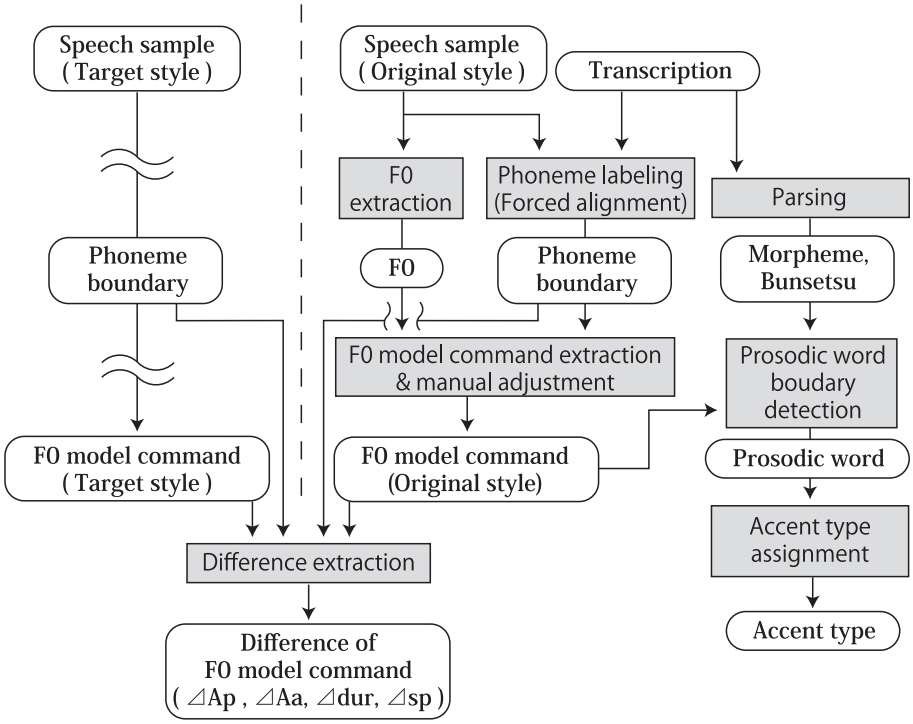


図 4.1: 韻律コーパス (差分) 構築の流れ

差分情報の韻律コーパスの構築の流れを図 4.1 に表す．第 3.3 目に記した韻律コーパス構築 (図 3.2) をベースとした流れでコーパスの構築を行う．図 3.2 の作業との主要な差異は下記のとおりである．

- F0 パターンおよび F0 モデルパラメータの抽出・編集に FujiParaEditor を用いる
- F_0 パラメータ自動抽出の後，FujiParaEditor 上で適宜手作業による調整を行う
- 平静スタイルと目的のスタイルとの間でパラメータの差分を抽出する

F_0 モデルパラメータ以外のパラメータについては，図 3.2 の韻律生成系のコーパス構築と同様にして平静の音声 (Original style) から抽出したものを用いる．

4.2.2 フレーズ指令，アクセント指令の差分情報抽出

FujiParaEditor を用いて，発話スタイルごとの F_0 モデルパラメータを抽出する．

まずはじめに，平静の発話スタイルの音声を分析し，平静時の韻律構造を得る．続いて目的の発話スタイルの音声を分析し，平静時の韻律構造を参考に F_0 モデルパラメータを調整する．

本研究における差分抽出・適用は，平静の韻律構造を変えないという条件の下で行っている．韻律構造の変化は，F_0 モデルにおいてフレーズ指令やアクセント指令の個数や配置，対応するテキストの範囲などが変化することに当たる．発話スタイルの変化が大きい場合には，アクセント型自体の変化も生じることが予想される．しかし，数十文程度の音声試料では，韻律構造が一致する場合の F_0 パラメータの差分の学習と，韻律構造自体の変化の学習を両立できない恐れがある．このため，小規模のデータに基づく効率的な韻律制御を検討するため，韻律構造は変えずにという制御する項目を絞っている．

上記の条件の下，平静の発話スタイルと韻律語の対応がうまくとれたものについて，発話スタイル間で指令の大きさ A_p，A_a の差分 $(\Delta A_p, \Delta A_a)$ をとる．

決定木を作成する際に質問項目の候補として導入するパラメータは表 4.1 である．

表 4.1: 差分推定のための入力項目 $(\Delta A_p, \Delta A_a)$

共通の入力項目
当該句の文内位置
当該 (先行) 句のモーラ数
当該 (先行) 句のアクセント型
当該文節開始点から先行フレーズ指令までのモーラ数
境界コード
当該文節境界直前の休止の有無
文節境界直前の休止長
先行句のフレーズ指令の有無
当該句の A_p

固有の入力項目 (ΔA_p)	固有の入力項目 (ΔA_a)
平静時の当該句の A_p	平静時の当該句の A_a
先行句の A_p	当該 (先行) 句の有する単語数
当該句の韻律句内位置	当該 (先行) 句の最初の単語の品詞
	(〃) の活用形
	当該 (先行) 句の最後の単語の品詞
	(〃) の活用形

4.2.3 音韻継続長の差分情報抽出

julius によって抽出された音韻セグメンテーションの結果から得た音韻継続長を比較し，発話スタイル間の差分をえる．発話速度の変化の大半は母音の継続長の変化によってもたらされるため，抽出される差分も母音において比較的大きいものが得られると予想される．なお音韻継続長の差分については，発話スタイル間の継続長の差ではなく，継続長の比を差分情報として用いている．決定木の構築に利用する入力項目は表 4.2 の通りである．

表 4.2: 差分推定のための入力項目 (duration)

当該 (前後) の音素の種類
当該 (前，当該句末の) 形態素の品詞
当該句頭 (句末) の活用形
当該 (先行) 句のアクセント型
当該句 (休止間の区分) の文内位置
当該 (先行) 句のモーラ数
先行形態素の活用形
当該句直後の文節境界の境界コード
先行 (後続) 休止長
先行 (後続) 休止までのモーラ数
当該呼気段落モーラ数
当該 (先行) 句の単語数
平静時の音素継続長

4.2.4　ショートポーズ継続長の差分情報抽出

ショートポーズ (sp) の場合も音韻継続長と同様，差ではなく比を差分情報として用いている．決定木の構築に利用する入力項目は表 4.3 の通りである．

sp の継続長は無音区間の長さとなるため，連続している音声の個々の音素継続長の変化よりも，その変化は顕著に聴取者に知覚されるものと予想される．

表 4.3: 差分推定のための入力項目 (sp)

当該文節境界での休止長
当該文節境界での休止の有無
先行休止からのモーラ数
文のモーラ数
文頭から当該文節境界までの文節数
先行文節のモーラ数
先行文節末の形態素の品詞
先行文節の助詞の格
先行文節における並列構造
先行文節末の形態素の種類
次文節のモーラ数
次文節頭の形態素の品詞
次文節頭の形態素の活用形
境界コード

4.2.5　小規模のコーパスに基づくルールベース韻律制御

上記のコーパスベース韻律制御系と並行して，簡単なルールベースによる制御について検討を行う．コーパスベース手法が決定木を用いて統計的にパラメータ推定系を構築するため，人間では把握しきれない複雑な制御についても近似的な制御規則を見出すことが可能である．一方，ルールベース手法はデータの分析の結果に基づいて，意図的なに制御規則を記述する手法であり，制御対象となるパラメータにある程度明確な法則が見られる場合に有効である．発話スタイルの変化は，話者の発話意図を反映した結果として現れるため，比較的容易に制御規則を設定できるものと予想される．

4.3 F_0 モデルパラメータ分析

4.3.1 合成条件

対象とした音声資料は，下記の表 4.4 の通りである．

表 4.4: 自然音声の分析条件

音声提供者	男性話者 MMI(差分情報のみ), 女性話者 FTY
発話内容	ATR 連続音声コーパス 503 文 [32]
サンプリング周波数	16kHz
量子化ビット数	16bit

ATR503 文は，音素列の出現パターンを十分に網羅するよう選定されたテキストのセットであり，HMM 音声合成を含め音声分野において広くコーパスのテキストとして用いられている．本研究では，F_0 モデルパラメータの差分を学習する上で，このうち 50 文を学習用データとして用いた．学習データとは異なる 10 文の音声を評価用データとし，各文の言語情報や平静時の F_0 モデルパラメータを決定木に入力して差分値を推定した．ちなみに，十分に多くのサンプルが得られることを想定している場合，40〜50 サンプル程度を一つの葉ノードに仕分けしている．本研究においては，これと同様に 40 を基本の最低サンプル数として扱う．

なお，平静のスタイルから目的のスタイルへの変化について評価を行うために，韻律制御の対象となる平静状態の韻律については十分に自然性の高いものを用いる必要がある．このため平静の F_0 モデルパラメータは FujiParaEditor によって手作業で編集されたものを直接用いた．以後，平静の合成音声はこの韻律を用いて再合成されたもの，丁寧，ぞんざいの合成音声は平静の韻律に差分を適用して再合成されたものを指すものとする．ケプストラムについては，HMM 音声合成システムのツールキットである HTS[5] によって女性話者 FTY の平静音声 503 文より学習されたものを共通して用い，生成された F_0 パターンと合わせて HTS の枠組みで音声試料を合成した．なお，第 2.2 節において HMM 音声合成における話者適応技術に触れたが，今回の実験においてはスペクトル情報については変更を加えず，平静の音声資料から学習したものを各発話スタイルの音声に利用する．

4.3.2 基底周波数 (F_b) の設定

F_0 パターンを作成する上で，フレーズ指令，アクセント指令を重畳する基底周波数 F_b を定める必要である．F_b の高さによって F_0 パターン全体の聞こえの高さが変わり，話者性の認識にも大きく影響する．

F_b もアクセント指令，フレーズ指令と同様，統計モデルによる推定の対象とすることも考えられる．しかし，Fb そのものを F_0 パターンから自動的に推定することは難しい．

また，F_b は感情などによって変動はあるが，同一の話者，同一のコンディションのうちでは変動が少ないと考えられるため，音声試料から得た F_0 パターンを参考に一定値を設定して F_0 モデルパラメータの分析を行っている．F_b 抽出された F0 パターンから F_b を検討する基準として，次のような求め方が候補として考えられる．

1. すべての音声試料中での F_0 の最低値

2. それぞれの音声試料中の F_0 の最低値の平均値

3. すべての音声試料中での F_0 の平均値から $n \times \sigma$ を引いたもの $(n : 1 \sim 3)$

また，F_0 モデルが声帯制御機構と深い対応関係があるという特徴を踏まえ，適切な大きさの F_0 モデルパラメータが設定できるように考慮する必要がある．以上を考慮し，平静時の女性話者 FTY の F_b として 140Hz を設定した．これは話者 FTY による平静の発話スタイル発声 50 文の平均値からおよそ 2σ を引いた値に相当し，韻律句頭の A_p の値は 0.4〜0.5 程度となっている．

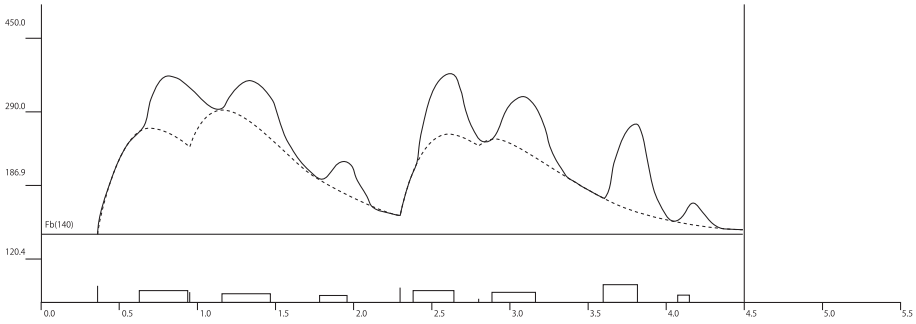
丁寧，ぞんざいの発話スタイルの自然音声の分析の際も，女性話者 FTY の F_b は平静のものと同様に 140Hz とした．様々な発話スタイルは平静の状態からの「ずれ」であると捉え，指令の大きさと時間情報の変更のみによって「ずれ」を表現する．また，同様にして男性話者 MMI の F_b を 85Hz と設定した．

4.3.3 各発話スタイルに見られる韻律上の特徴

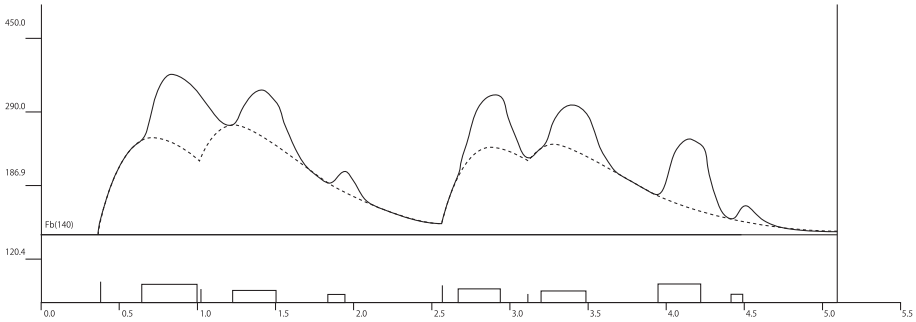
丁寧，ぞんざいの各発話スタイルについて，平静と比較した場合の特徴を記す．丁寧の場合，アクセントの無い部分 (フレーズ成分のみの部分) はやや低くなり，アクセントを置いた時の F_0 の高低差が大きくなる印象がある．また，発話速度がやや遅く，sp も十分に置かれることから，一つ一つの発話，アクセントが明瞭に発声されている．これは女性話者 FTY，男性話者 MMI のいずれにも共通して見られる．ぞんざいの場合，F_0 の動きが必ずしも平静や丁寧に見られる緩やかな山の形にならず，韻律語末や韻律句末が高いまま終わるなど，一般的なアクセント規則から逸脱した発声となることが多く見られる．また，sp が極端に短くなり (あるいは省略され)，矢継ぎ早に話をしている印象を受けるため，文中のアクセントの明瞭性もぼやけている．

4.4 合成結果

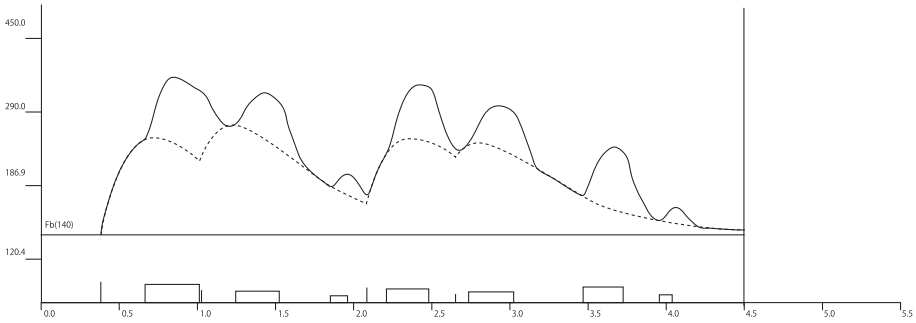
女性話者 FTY による差分情報を，女性話者 FTY の音声合成系に適用した合成音声の対数 F_0 パターンを図 4.2 に示す．図 4.2(a) の平静の F_0 パターンは制御を行う前の初期値に当たる．図 4.2(b) の丁寧の制御結果の大きな特徴として，主に文頭，韻律句頭のアクセント指令が強まっているほか，sp を含め発話全体の長さが伸びている．図 4.2(c) のぞんざいの F_0 パターンにおいては，丁寧と同様のアクセントの変化がある一方で，sp の継続長が非常に短く，2.0 秒付近において，3 つ目のアクセントの立下り直後にフレーズ成分による F_0 の隆起が見られる．



(a) 平静の場合



(b) 丁寧の場合



(c) ぞんざいの場合

図 4.2: 女性話者 FTY による差分情報に基づく合成音声の対数 F_0 パターン
(例：「学生はレポートを置くと，ちょっと頭を下げて出て行った.」)

続いて，各発話スタイルの音韻継続長を表 4.5 に示す．“order” の欄の “word” と “bunsetsu” はそれぞれ文中の単語と文節の番号でありそれぞれ先頭の音素に対応させている．図 4.5 の “from” および “to” は発声全体における各音素の開始時刻，終了時刻を示す (単位：sec)．文頭の “silB” と文末の “silE” はそれぞれ文の開始と終了をあらわす記号である．

表 4.5: 女性話者 FTY による差分情報に基づく音韻継続長の制御結果
(例：「学生はレポートを置くと，ちょっと頭を下げて出て行った.」)

		order		natural		polite		impolite	
id:	phoneme	word	bunsetsu	from	to	from	to	from	to
0	silB	0	0	0.000	0.480	0.000	0.490	0.000	0.490
1	g	1	1	0.490	0.520	0.500	0.550	0.500	0.550
2	a	−	−	0.530	0.590	0.560	0.620	0.560	0.610
3	k	−	−	0.600	0.660	0.630	0.690	0.620	0.690
4	u	−	−	0.670	0.690	0.700	0.760	0.700	0.760
5	s	−	−	0.700	0.780	0.770	0.870	0.770	0.870
6	e	−	−	0.790	0.880	0.880	0.940	0.880	0.940
7	i	−	−	0.890	0.910	0.950	1.020	0.950	1.010
8	w	2	−	0.920	0.990	1.030	1.090	1.020	1.090
9	a	−	−	1.000	1.070	1.100	1.140	1.100	1.140
10	r	3	2	1.080	1.110	1.150	1.190	1.150	1.190
11	e	−	−	1.120	1.200	1.200	1.280	1.200	1.270
12	p	−	−	1.210	1.300	1.290	1.360	1.280	1.360
13	o	−	−	1.310	1.380	1.370	1.430	1.370	1.430
14	o	−	−	1.380	1.450	1.440	1.500	1.440	1.500
15	t	−	−	1.460	1.560	1.510	1.580	1.510	1.580
16	o	−	−	1.570	1.630	1.590	1.660	1.590	1.650
17	o	4	−	1.640	1.670	1.670	1.720	1.660	1.730
18	o	5	3	1.680	1.830	1.730	1.820	1.740	1.820
19	k	−	−	1.840	1.920	1.830	1.900	1.830	1.910
20	u	−	−	1.930	1.960	1.910	1.950	1.920	1.960
21	t	6	−	1.970	2.060	1.960	2.030	1.970	2.040
22	o	−	−	2.070	2.160	2.040	2.150	2.050	2.150
23	sp	−	−	2.170	2.460	2.160	2.730	2.160	2.240
24	ch	7	4	2.470	2.540	2.740	2.830	2.250	2.340
25	o	−	−	2.550	2.610	2.840	2.910	2.350	2.420
26	q	−	−	2.620	2.710	2.920	3.010	2.430	2.510
27	t	−	−	2.720	2.780	3.020	3.070	2.520	2.580
28	o	−	−	2.790	2.810	3.080	3.130	2.590	2.650
29	a	8	5	2.820	2.890	3.140	3.200	2.660	2.710
30	t	−	−	2.900	2.970	3.210	3.270	2.720	2.790
31	a	−	−	2.980	3.010	3.280	3.340	2.800	2.860
32	m	−	−	3.020	3.090	3.350	3.420	2.870	2.940
33	a	−	−	3.100	3.120	3.430	3.480	2.950	3.000
34	o	9	−	3.130	3.200	3.490	3.550	3.010	3.070
35	s	10	6	3.210	3.320	3.560	3.670	3.080	3.180
36	a	−	−	3.330	3.360	3.680	3.710	3.190	3.230
37	g	−	−	3.370	3.420	3.720	3.810	3.240	3.320
38	e	−	−	3.430	3.490	3.820	3.860	3.330	3.370
39	t	11	−	3.500	3.550	3.870	3.940	3.380	3.450
40	e	−	−	3.560	3.630	3.950	4.020	3.460	3.530
41	d	12	7	3.640	3.680	4.030	4.100	3.540	3.600
42	e	−	−	3.690	3.770	4.110	4.160	3.610	3.660
43	t	13	−	3.780	3.830	4.170	4.230	3.670	3.730
44	e	−	−	3.840	3.890	4.240	4.290	3.740	3.790
45	i	14	−	3.900	4.030	4.300	4.400	3.800	3.890
46	q	−	−	4.040	4.180	4.410	4.520	3.900	4.050
47	t	15	−	4.190	4.230	4.530	4.580	4.060	4.110
48	a	−	−	4.240	4.310	4.590	4.660	4.120	4.190
49	silE	−	−	4.320	−	4.670	−	4.200	−

第5章　主観評価実験

5.1 はじめに

日本人の被験者を対象に，聴取実験を行った．本研究の韻律制御系において，十分な発話スタイルの再現性を有していること，自然性を損なうことなく制御ができていることの二点を確認する．

5.2 実験条件

実験 1，2 については，6 名の日本人を被験者とした．発話スタイル表現は話者ごと，また話者のコンディションによって異なる．このため，評価に先だって，目的の発話スタイルごとに評価基準を被験者に与える必要がある．そこで本実験に先立ち，女性話者 FTY の平静，丁寧，ぞんざいの各発話スタイルについて同一テキストの自然音声を 10 文ずつ提示し，発話スタイル間の違いについて被験者にイメージを持たせた．

また，実験 3 については，3 名の日本人を被験者とした．ここでの評価対象は，実験 1，2 において比較的良好な結果の得られた丁寧の発話スタイルとし，実験 1，2 と同様に，事前に男性話者 MMI の平静と丁寧および女性話者 FTY の平静の発話スタイルについて同一テキストの自然音声を 10 文ずつ提示した．

5.3 実験 1：発話スタイルの再現性の評価

平静の F_0 モデルパラメータを初期値として韻律制御を行った結果，十分に目的の発話スタイルが再現されているかを評価する．平静と丁寧，もしくは平静とぞんざいの二つの合成音声 10 文ずつを被験者に提示し，どちらの音声がよりよく目的の発話スタイルを表現しているかを 5 段階で評価させる．実験後，評価結果を表 5.1 の通りに解釈しなおし，スコアをつける．聴取の際，提示する音声の順番と発話スタイルの対応は固定ではなく，合成条件を隠して被験者に提示される．また，提示する順番を入れ替えたパターンも実験にかける．

表 5.1: 発話スタイルの再現性の評価基準
(例：音声 A：丁寧 or ぞんざい，音声 B：平静)

評価結果		スコアと評価基準
音声 A	→	5：有効である
やや音声 A	→	4：やや有効である
同じである	→	3：効果がない
やや音声 B	→	2：効果がなく，やや不適切な制御である
音声 B	→	1：効果がなく，不適切な制御である

5.4 実験2：合成音声の自然性の評価

平静の F_0 モデルパラメータを初期値として韻律制御を行った結果，韻律の自然性にどの程度影響が現れるかを評価する．丁寧，ぞんざいの各発話スタイルの合成音声 10 文ずつを被験者に提示し，韻律の自然性について表 5.2 のように五段階で評価させる．また，比較のために平静の合成音声についても同様に評価させる．聴取の際，合成音声の順番は 3 つの発話スタイルの音声の中からランダムに設定され，合成条件を隠して被験者に提示される．

表 5.2: 合成音声の自然性の評価基準

スコア	評価基準
5:	十分に自然である
4:	ほとんど自然である
3:	やや不自然な点もあるが，許容範囲である
2:	不自然さが気になる
1:	全く不自然である

5.5 実験3：話者オープンの韻律制御の評価

男性話者 MMI から抽出した差分情報を，女性話者 FTY の平静のパラメータに適用した場合 (話者オープン) の合成音声の評価を行う．平静と丁寧の合成音声 10 文ずつを用いて，発話スタイルの再現性および韻律の自然性の評価を行う．

5.6　実験結果

5.6.1　再現性の評価結果

再現性の評価結果を図 5.1 に示す.

丁寧の発話スタイルの場合，スコアの平均は 3.775 となり，顕著な変化は見られないものの，平静の合成音声と比べて概ね丁寧な印象があるという結果が得られた．一方，ぞんざいの発話スタイルの場合，スコアの平均は 2.6 となった．制御を加えていない平静の合成音声の方がぞんざいであると評価されるケースも多く，平静からぞんざいへの有効な変化が与えられなかったという結果となった.

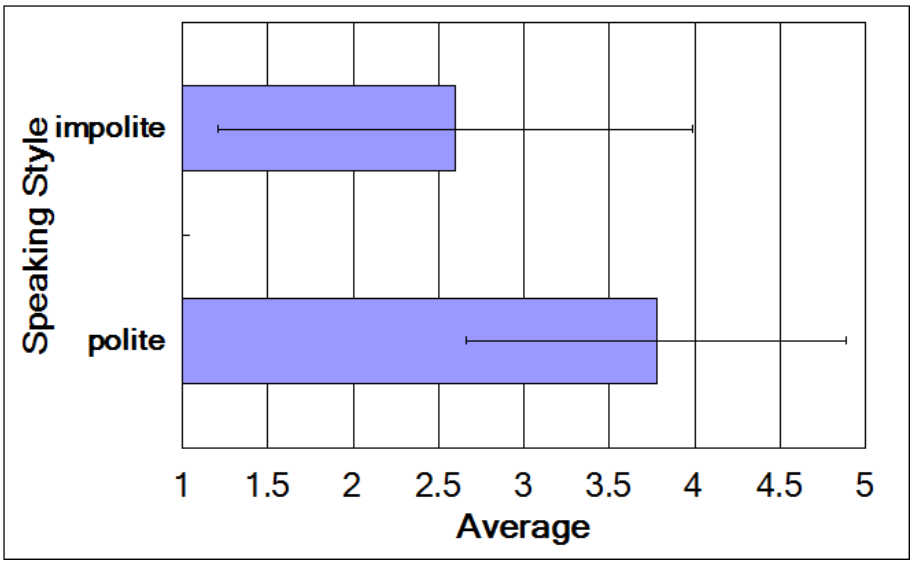


図 5.1: 再現性の評価結果

5.6.2 自然性の評価結果

自然性の評価結果を図 5.2 に示す．丁寧，ぞんざいのいずれの場合において，平静の合成音声のスコアと比較して大きな差が出ていないが，韻律の制御によって若干ながら自然性のスコアが落ちており，丁寧よりもぞんざいの方がその影響が大きかった．

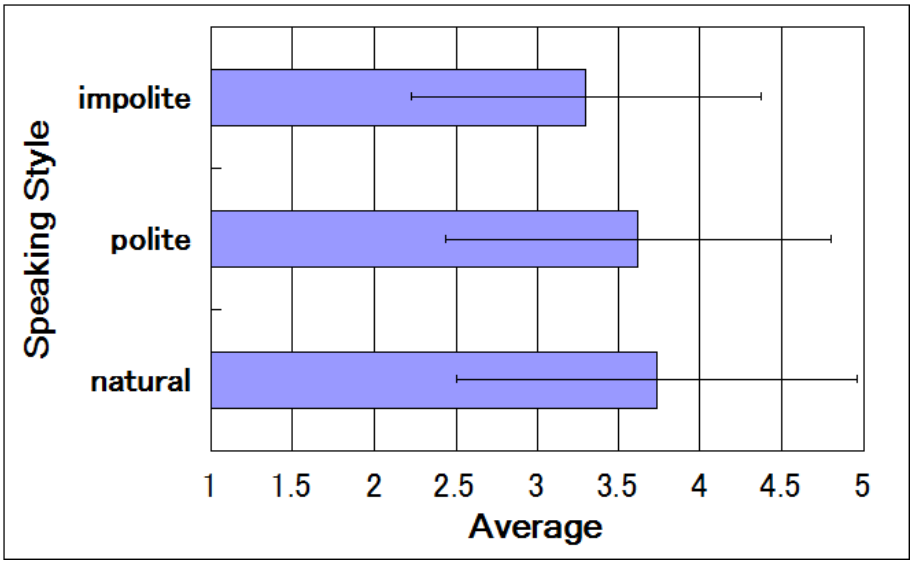


図 5.2: 自然性の評価結果

5.6.3 自然性評価における有意差検定

自然性への影響を考察するため，有意差検定を行った．平静と比較したときの指標 t_0 の値は，丁寧の場合に 0.822，ぞんざいの場合は指標 t_0 が 2.207 となった (表 5.3)．いずれの検定においても自由度は 118 だが，自由度 ∞，1% 有意水準の棄却域が 2.576 以上であることから，この実験において，制御前と制御後とで韻律の自然性には有意な差は見られなかった．

表 5.3: 自然性評価の有意差検定における条件および指標 t_0

スタイル	平均スコア	不偏分散	サンプル数	平静と比較した指標 t_0
平静	3.733	1.233	60	-
丁寧	3.617	1.180	60	0.581
ぞんざい	3.300	1.078	60	2.207

5.6.4 話者オープンの場合の評価結果

発話スタイルの再現性の評価におけるスコアは 4.233 となった．また，自然性評価においては，平静のスコアが 4.13 であったのに対し丁寧のスコアは 3.2 となり，女性話者 FTY による話者クローズドの韻律制御よりも顕著な差が現れた．自然性のスコアに対する有意差検定を行った結果，指標 t_0 は 3.332 となった．自由度 28 の 1% 有意水準の棄却域は 2.763 以上であるため，この検定の指標は棄却域に含まれ，制御前後でのスコアの間には有意な差があるという結果となった．

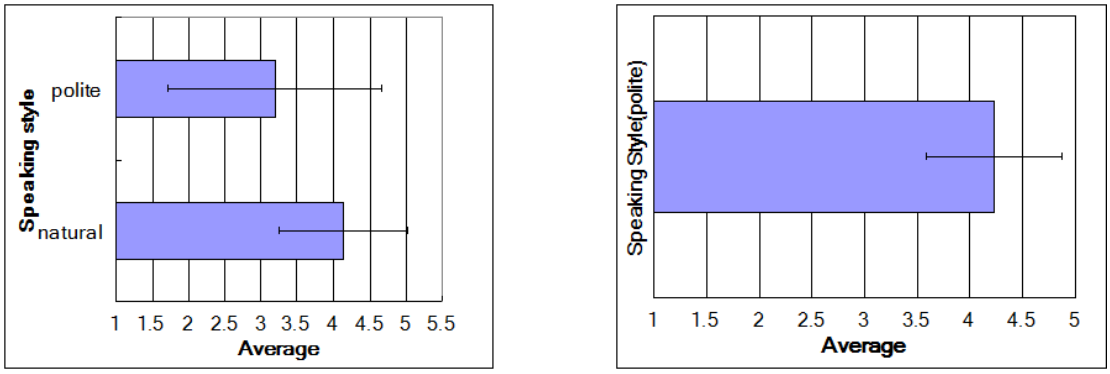


図 5.3: 自然性の評価結果

表 5.4: 自然性評価の有意差検定における条件および指標 t_0(話者オープンの場合)

スタイル	平均スコア	不偏分散	サンプル数	平静と比較した指標 t_0
平静	4.133	0.878	30	-
丁寧	3.200	1.475	30	3.332

5.7 考察

実験 1 の結果，40 文のデータを用いた場合でも，一般的な決定木構築と同様の条件で韻律制御を十分に実現しうることが確認された．アクセントの明瞭性の向上に大きく寄与しているものとして，韻律句頭のアクセントの動きが大きくなることが考えられる．韻律句頭であるかどうかという 2 値の分岐は，ほぼ全てのフレーズ指令に共通して重要な項目であるため，葉ノードの最低サンプル数を 40 とした場合でも十分に学習されている．

丁寧の発話スタイルについては，より明瞭なアクセントを実現したことで，有効な発話スタイル制御ができたものと思われる．平静の発話スタイルのアクセント表現にもある程度の明瞭さが含まれており，平静と丁寧との間に聴取の印象が似ているという評価者からの意見もあった．この点については，実験 2 の自然性評価において自然性の劣化に有意差が見られなかった一方で，スタイル制御の差分ももあまり顕著なものを得られなかったものと思われる．この問題の解決策として，各パラメータの差分情報を推定した後，推定値に掛け合わせる係数を設定することで，段階的にスタイル制御の度合いに強弱をつける事が挙げられる．

一方，ぞんざいの場合，平静と比較して sp が短いという点はよく反映されていたが，ΔA_p，ΔA_a によって指令が大きくなり，平静よりもアクセントが強調されてしまうケースもあり，効果的なぞんざいの発話スタイルの再現には至らなかった．今回の実験条件では，平静の韻律における指令の大きさのみを変えて表現をしていたが，ぞんざいの発話スタイルの場合，平静や丁寧とは韻律構造が異なり，その F_0 パターンを詳細に再現するためには，韻律構造の変更も含めた制御検討する必要がある．

実験 3 において，話者オープンの差分情報であっても，十分に有効性な韻律制御が可能であることが確認された．また，自然性の劣化において有意差が見られたものの，発話スタイルの再現性については比較的大きなスコアを得ることができた．発話スタイルの再現性と，韻律の自然性がトレードオフの関係にあることが考えられるが，F_0 モデルが人体の声帯制御機構と深い対応関係にあることからも，効率の高い韻律制御手法として有効性を確認できたものと思われる．

第6章　結論

6.1 結論

本研究では，F_0 モデルの枠組みに基づいて韻律パラメータの差分情報を学習し，平静の韻律パラメータを制御して目的の発話スタイルの韻律を得る手法を検討した．50 文程度の比較的小規模な音声資料のセットから抽出した差分情報を元に，任意のテキストに対応して F_0 および発話速度に関わる韻律パラメータを推定する枠組みを提案した．また，差分情報という相対的なものを用いることで，ある話者から得られた差分情報を任意の話者の韻律生成系に効果的に適用できることを確認した．

丁寧とぞんざいの発話スタイルをターゲットとして制御を行った結果，ぞんざいの発話スタイルについては，既存の F_0 モデルの枠組みで表現しづらい局所的な韻律表現を扱わなかったこともあり，発話スタイルの効果的な再現には至らなかった．その一方で，丁寧の発話スタイルについては，アクセントの明瞭性を向上させるという点で，有効性が確認できた．

6.2 展望

一方，ぞんざいの発話スタイルについては，韻律構造を変えないという条件が再現性に大きく制限を与えることとなった．音声合成システムに求められるであろう実用的な発話スタイルは，肯定的・否定的のいずれにしても十分に明瞭なアクセント表現を伴うものと予想されるため，大きく韻律構造を変える発話スタイルの実現の優先度はあまり高くはないと考えられる．しかし，ぞんざい音声に見られた韻律句末の F_0 が高くなるといった特徴は，その再現に検討の余地がある．

現在の F_0 モデルの枠組みでは，アクセント指令は最小でも文節レベルの範囲で生起するものとして扱っているため，単語レベル・モーラレベルでの詳細な韻律の変化には対応していない．文節末の助詞を強調するといった表現ができれば，先行研究で挙げた韻律句頭を強調する焦点制御とは異なる形で焦点・発話意図の表現が可能になる．このことから，F_0 モデルの枠組みを拡張し，アクセント表現の細分化して制御することは，本研究の枠組みの課題として挙げられる．ただし，F_0 モデルの大きな特徴として，声帯制御機構との深い対応関係を持っている点があるため，詳細なレベルでの指令の制御が声帯の挙動とどのような対応関係を持ちうるのか，という点には留意する必要がある．

また，発話スタイル表現の一つとして感情音声合成を考えた場合，F_0 だけでなく声の強さ (パワー) も大きな要素となる．これについては，F_0 モデルと同様，パワーの挙動についても生成過程モデルを用いて記述する手法が提案されている [33]．パワーの動きはアクセントの動きと強い関連があるため，生成過程モデルを用いた F_0，パワーの双方に渡る韻律制御系の実現の可能性が本研究の結果から伺える．

謝辞

本研究の指導教員であり度重なるご指導をいただきました広瀬啓吉先生ならびに峯松信明先生、日々の研究活動でお世話になりました高橋登技官、秘書の楠本由香里さん、磯部史子さん、池上恵さんに感謝の意を表します。

続きまして、本研究について多くのご助力をいただきました博士課程の越智さん、齋藤さんをはじめとする諸先輩方実験系の準備にご協力いただいた松田君、お忙しい中聴取実験にご参加いただきました学生の皆様に感謝いたします。

2011 年 2 月 9 日

見原 隆介

参考文献

[1] H. Fujisaki and S. Nagashima, “A model for synthesis of pitch contours of connected-speech,” Annual Report of Engineering Research Institute, University of Tokyo, Vol. 28, pp. 53-60, 1969.

[2] 広瀬啓吉, 峯松信明, 佐藤健太郎, “生成過程モデルに基づくコーパスベース基本周波数パターン生成”, 特定領域研究-韻律に着目した音声言語情報処理の高度化- 研究成果報告書, pp. 217-224, 2005.

[3] X. D. Huang, Y. Ariki and M. A. Jack, “Hidden Markov Models for Speech Recognition”, Edinburgh University Press, 1990.

[4] HTK Web site, http://htk.eng.cam.ac.uk/

[5] HMM-based Speech Synthesis System (HTS), http://hts.ics.nitech.ac.jp/

[6] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, T. Kitamura, “Speech parameter generation algorithms for HMM-based speech synthesis,” Proc. of ICASSP 2000, vol. 3, pp. 1315-1318, 2000.

[7] 音声信号処理ツールキット http://sp-tk.sourceforge.net/

[8] 吉村貴克, 徳田恵一, 益子貴史, 小林隆夫, 北村正, “HMM に基づく音声合成方式におけるスペクトル・ピッチ・継続長の同時モデル化”, 電子情報処理学会論文誌 (D-II), vol. J83-D-II, no. 11, pp. 2099-2107, 2000.

[9] 徳田恵一, “隠れマルコフモデルに基づいた韻律の統計モデル化手法と感情音声の生成”, 特定領域研究-韻律に着目した音声言語情報処理の高度化- 研究成果報告書, pp. 255-262, 2005.

[10] 宇津呂武仁, “テキスト処理”, 『文字と音の情報処理』, pp. 1-55, 2000.

[11] 形態素解析システム茶筌, http://chasen-legacy.sourceforge.jp/

[12] 島津明, “文章の解析と理解”, 『文字と音の情報処理』, pp. 1-55, 2000.

[13] 日本語構文解析システム KNP, http://www-nagao.kuee.kyoto-u.ac.jp/nl-resource/knp.html

[14] 日本語形態素解析システム JUMAN,
http://www-nagao.kuee.kyoto.u.ac.jp/mlresource/juman.html

[15] 秋永一枝 “共通語のアクセント”, 『日本語アクセント辞典』, 三省堂, 1981.

[16] 白井克彦, “音声の分析と合成”, 岩波書店, 『音声』, pp. 127-176, 2004.

[17] 匂坂芳典, 佐藤大和, ” 日本語単語連鎖のアクセント規則”, 電子情報通信学会論文誌, J66-D, No. 7, pp. 849-856, 1983.

[18] 喜多竜二, 峯松信明, 広瀬啓吉 “日本語テキスト音声合成を目的としたアクセント結合規則の構築と改良”, 電子情報通信学会技術研究報告, SP2002-26, pp. 13-18, 2002.

[19] 本多清志, “音声の生物学的基礎”, 岩波書店, 『音声』, pp. 93-125, 2004.

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

[21] 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』講演論文集, vol. 1, pp. 124-141, 1993.

[22] L. Breiman, J. Freidman, R. Olshen, and C. Stone, Classification and Regression Trees, Chapman and Hall/CRC 1998.

[23] The Edinburgh Speech Tools Library, http://www.cstr.ed.ac.uk/projects/speech_tools/

[24] R. O. Duda, P. E. Hart, D. G. Stork, 尾上守夫 (監訳), 『パターン識別』, 毎日コミュニケーションズ, 2001.

[25] 成澤修一, 峯松信明, 広瀬啓吉, 藤崎博也, “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.

[26] H. Mixdorff, FujiParaEditor:
http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html.

[27] Boersma, P., Weenink, D. Program Praat: doing phonetics by computer:
http://www.praat.org.

[28] 海木信佳, 匂坂芳典, “局所的な句構造によるポーズ挿入規則化の検討”, 電子情報通信学会論文誌, D-II Vol. J79-D-II, No. 9, pp.1455-1463, 1996.

[29] 越智景子, 広瀬啓吉, 峯松信明, “基本周波数パターン生成過程モデルを用いたテキストからのコーパスベース韻律生成とその評価”, 日本音響学会春季講演論文集, 3-8-4, pp. 231-232, 2007.

[30] K. Hirose, K. Sato, Y. Asano and N. Minematsu, "Synthesis of F_0 contours using generation process model parameters predicted from unlabeled corpora: Application to emotional speech synthesis," Speech Communication, Vol. 46, No. 3-4, pp. 385-404, 2005.

[31] K. Ochi, K. Hirose, and N. Minematsu, "Realization of prosodic focuses in corpus-based generation of fundamental frequency contours of Japanaese based on the generation process model," Proc. Int. Conf. Speech Prosody, CD-ROM, 2010.

[32] Y. Sagisaka, K. Takeda, M. Abel, S. Katagiri, T. Umeda, H. Kuwabara, "A Large-Scale Japanese Speech Database", Proc. ICSLP, pp. 1089-1092, 1990.

[33] 大野澄雄, "韻律的特徴の総合的なモデル化と、感情の表現・伝達過程", 特定領域研究-韻律に着目した音声言語情報処理の高度化- 研究成果報告書, pp. 119-126, 2005.

発表文献

1. 見原隆介, 齋藤大輔, 峯松信明, 広瀬啓吉, “二言語に渡る構造的表象に基づく音声・言語変換の実験的検討”, 日本音響学会春季講演論文集, 3-P-19, pp. 403-406, 2009.

2. 見原隆介, 齋藤大輔, 峯松信明, 広瀬啓吉, “二言語に渡る構造的表象に基づく音声・言語変換手法に関する考察”, 電子情報通信学会技術研究報告, SP2009-71, pp. 55-60, 2009.

3. 見原隆介, 越智景子, 広瀬啓吉, 峯松信明, “基本周波数パターン生成過程モデルに基づくコーパスベース韻律生成における発話スタイル制御”, 日本音響学会春季講演論文集, 1-Q-24, 2011 (to appear).