

修士論文

HMM音声合成のための
学習コーパスの韻律の観点からの
評価と選択 -基本周波数パターン
生成過程モデルの利用-



2015年2月5日

指導教員 広瀬 啓吉 教授

東京大学大学院 情報理工学系研究科
電子情報学専攻

48-136438 水上 智之

内容梗概

音声合成技術は、身の回りに増えてきている。駅の自動案内やカーナビの案内、スマートフォンなどの携帯端末を用いた日常を支えるシステム、医療支援やコミュニケーションの補助ツール、はたまた歌声やゲーム音声といったエンターテインメント利用など、至る所で聞く機会が生まれている。そういった音声合成システムは世界中で盛んに研究されてきている。初期の研究ではいかにも人工的な機械音を生成していたが、計算機の発達、大量のデータの利用、また統計処理や機械学習などの進歩によって音声合成技術は目覚ましく発展してきている。この技術によって、任意の話者、また任意のスタイルで、任意の文章を読み上げさせることが出来るようになりつつある。しかし、そのためには様々な障害があり、また生成される音声そのものの品質も十分であるとは言い難い。

そこで本研究では、統計的な音声合成手法の1つである HMM 音声合成の学習コーパスに着目して、合成の音声品質を向上させる手法を提案した。HMM 音声合成では HMM 学習に必要な音声コーパスを用いて音声を生成するためのモデルを構築していくが、このときに学習に用いると不都合なデータが度々出現してしまう。特に基本周波数の抽出ミスは顕著な問題で、今までに多くの抽出手法が編み出されてきたが、根本的な解決には至っていない。ここで、基本周波数を扱う基本周波数パターン生成過程モデルを用いることで、そういった学習に使うデータを評価することで学習に用いるサンプルを選択する手法を提案し、その有効性を示した。更に、学習データを分析することでどのような手法が採れるか検討していった。

目次

第 1 章	序論	1
1.1	背景	2
1.2	研究の目的	2
1.3	本論文の構成	3
第 2 章	関連研究	4
2.1	はじめに	5
2.2	音声の生成過程と音響特徴量	5
2.2.1	音声の生成過程	5
2.2.2	基本周波数	5
2.2.3	メル一般化ケプストラム	6
2.3	基本周波数パターン生成過程モデル	6
2.3.1	フレーズ成分	6
2.3.2	アクセント成分	7
2.3.3	基本周波数パターンの生成	7
2.3.4	基本周波数パターン生成過程モデルの生理学的、物理学的根拠	8
2.4	HMM 音声合成	10
第 3 章	基本周波数パターン生成過程モデルによる基本周波数差分の導入	13
3.1	はじめに	14
3.2	基本周波数差分	14
3.3	基本周波数差分が大きくなる要因	14
3.4	F0 差分を利用した HMM 学習の改善	16
第 4 章	基本周波数差分の大小を基にした学習コーパスの評価・選択	18
4.1	はじめに	19
4.2	実験条件	19
4.3	実験 1：音素単位での学習サンプル選択	19
4.3.1	実験 1 の条件	19
4.3.2	結果	22
4.3.3	考察	22
4.4	実験 2：F0 モデルによる学習サンプルの補間	25
4.4.1	実験 2 の条件	25

4.4.2	結果	25
4.4.3	考察	26
第 5 章	基本周波数差分の大小を基に付けたラベルによる HMM 学習	28
5.1	はじめに	29
5.2	フルコンテキストラベル	29
5.3	実験 1：基本周波数差分の大小を付与する手法	29
5.3.1	実験 1 の条件	29
5.3.2	結果	31
5.3.3	考察	31
5.4	実験 2：基本周波数差分が大きくなる要因を付与する手法	32
5.4.1	実験 2 の条件	32
5.4.2	結果	32
5.4.3	考察	32
第 6 章	結論	34
6.1	まとめ	35
6.2	今後の展望	35
	参考文献	37
	発表文献	39
付録 A	ATR コーパスの話者 MMI に付与したラベルデータ	40

目次

2.1	ソース-フィルタモデル(文献 [3] より引用)	5
2.2	F0 モデルによる F0 パターンの生成	8
2.3	喉頭の模式図	9
2.4	喉頭の構造	10
2.5	HMM 音声合成の概要(文献 [3] より引用)	11
3.1	抽出された F0 と F0 モデルパラメータによる F0	15
3.2	F0 差分が大きくなる要因	16
4.1	学習データの除外・分割手法	22
4.2	合成音声の F0 パターンの比較	24
4.3	学習データの補間手法	25
4.4	合成音声の F0 パターンの比較	26

表目次

4.1	実験条件	20
4.2	実験環境	20
4.3	除外音素・学習音素一覧	21
4.4	実験1(音素単位でのサンプル選択による実験)の結果	23
4.5	実験1の主観評価スコア	23
4.6	実験1の客観評価スコア	23
4.7	実験2の客観評価スコア	25
5.1	従来手法によるコンテキストラベル	30
5.2	基本周波数差分の大小を基にしたラベル	31
5.3	除外基準5%までの除外要因の個数	31
5.4	除外基準10%までの除外要因の個数	31
5.5	実験1の客観評価スコア	32
5.6	実験1の客観評価スコア	32

第1章

序論

1.1 背景

音声合成 (TTS:Text-To-Speech) は、任意の文章を与えたときにその文章を読み上げるシステムである。その利用法は様々で、鉄道用アナウンスや家電製品のアナウンス、電子書籍の読み上げや医療支援など多岐に渡り、音声合成技術の需要は近年増してきている。スマートフォンが広く普及して音声合成を利用したアプリケーションが登場したことで、音声合成は一般的なものと成りつつある面も見逃せない。そういった音声合成の手法としては、大量の音声コーパスから文章に合わせて波形を切り出してくる「波形接続法 [1]」や、音声コーパスを用いて統計モデルを構築して合成音声を生じる手法、特に統計モデルにHMM(Hidden Markov Model)を用いる「HMM 音声合成 [2]」など様々な手法がある。HMM 音声合成では、予めモデルを学習して構築しておくことで合成に用いるシステムに音声コーパスを保持しておく必要がない、HMM のモデルパラメータを適切に変換することで合成音声の声質や発話スタイルを変化させた柔軟な音声合成が可能である、といった特徴がある。その性能は進歩してきており、波形接続法と遜色ない音声を生じできると言われることもある。

HMM を用いた音声合成で一つ問題になるのは、HMM 学習に用いる学習コーパスに、分析の不安定性等から学習に悪影響を与えるサンプルが含まれることである。HMM 学習には学習コーパスを分析して得た特徴量を用いるが、この分析では必ずしも実際の音声から得られる正しい特徴量を抽出できず、不自然な乱れを持つことがあり、HMM の学習に悪影響を及ぼすことがある。例えば基本周波数パターンに着目すると、コーパス内の音声を持つ本来の音の高さに対して倍、あるいは半分の基本周波数が抽出される倍ピッチ・半ピッチエラーといった問題や、不規則な振動であるマイクロプロソディなどが問題となる。基本周波数の抽出に問題が起こった場合、同時にスペクトル包絡の特徴も乱れてしまうため、HMM の学習に相応しくないデータが出現してしまう。この不具合を解消する手法はあまり検討されておらず、学習コーパスの改善によって HMM 音声合成の品質が向上できると考えられている。

1.2 研究の目的

HMM 音声合成において、学習コーパスを分析して音響特徴量を得る段階で抽出の不具合が生じることで、HMM の学習に悪影響を与えるサンプルが含まれるという問題がある。このような学習サンプルを HMM の学習対象から除外して HMM を学習することで、合成音声の品質を向上させることを目的とする。また、そういった学習サンプルを除外するだけでなく別の特徴量によって補間したり、除外対象となったコーパスの選択基準を学習に反映させる手法を検討して、よりよい改善の方法を探っていく。

1.3 本論文の構成

本論文は全6章で構成される。第2章では本研究で用いているHMM音声合成システムと基本周波数パターン生成過程モデルとその周辺技術について述べる。第3章では、基本周波数パターン生成過程モデルを利用して基本周波数差分という考えを導入し、HMM音声合成を改善する手法を提案する。第4章では、第3章で提案した手法のうちHMM学習に用いるコーパスを評価・選択する手法について検討していく。第5章では、第4章で検討したコーパスの選択を学習時に用いるラベルに反映させることで、HMMの学習に対してどのような効果を与えるか検討していく。第6章では本論文の結論を述べ、今後の展望を記す。

第2章

関連研究

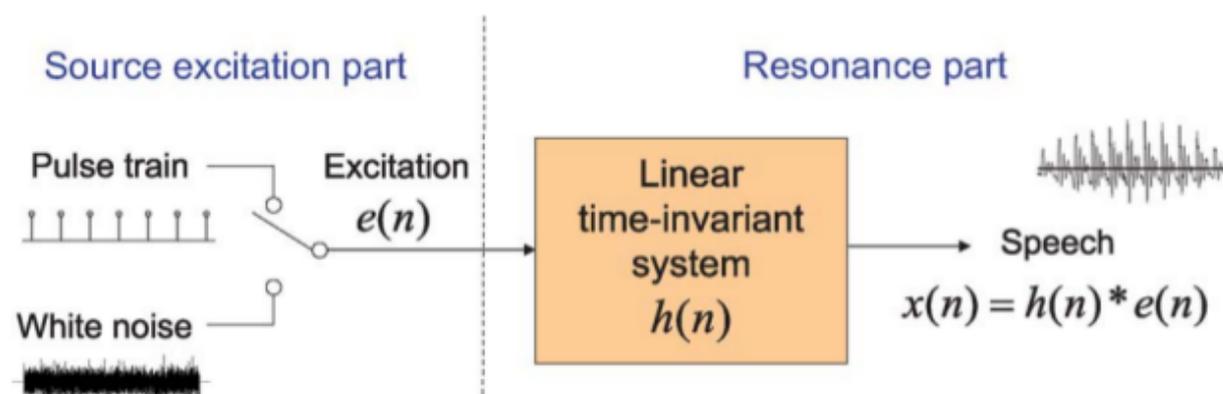


図 2.1: ソース-フィルタモデル (文献 [3] より引用)

2.1 はじめに

この章では、まず音声生成される過程を説明し、音声の特徴を表す音響特徴量を紹介する。次に、音響特徴量の1つである基本周波数を扱うモデルとして基本周波数パターン生成過程モデルについて紹介する。最後に音声合成の手法として盛んに研究されているHMM音声合成について述べていく。

2.2 音声の生成過程と音響特徴量

2.2.1 音声の生成過程

音声は、喉にある声帯の振動や声道を狭めることで生じる乱流を基に、声道の形状に従って周波数情報が変化して生成されるものである。声帯の振動は声の高さを決定し、声道の形状は個々の音の音韻性を決定するのに大きく関わっている。

ここで音声の生成過程はソース-フィルタモデルで表すことができる。図 2.1 にあるように、音声の音源として有声の場合はパルス波、無声の場合は白色雑音と置くことができ、声道特性は一種のフィルタとして考えることができる。

2.2.2 基本周波数

基本周波数とは、簡単に言えば人の声の高さに相当する特徴量である。これはイントネーションやストレス、アクセントなど発話スタイルや感情などに関わる重要な特徴量であり、基本周波数の扱いや制御が音声の研究では肝要となる。音声から基本周波数を抽出する方法としては、自己相関関数から周期性を探る方法 [4] や YIN らによる手法 [5]、ハーモニクモデルによる推定 [6] や HMM によるモデル化での推定 [7] など多種多様な手法が提案されている。その一方で、これといった確たる手法は提案されていないのが実状である。基本周波数の抽出が困難である理由として、人間の発声時々刻々として変化していくこと

や、発声は完全に制御されているわけではなく周期信号として見たときに非常に不安定であることが挙げられる。また、発声には声帯振動が伴う有声区間と、声帯振動を伴わない無声区間が存在しており、有声/無声の判別が困難であるという問題もある。

このように扱いが困難な基本周波数を上手に記述するモデルが求められ、その代表例として基本周波数パターン生成過程モデルが挙げられる。このモデルは少数のパラメータで基本周波数パターンを記述できるとともに、言語情報との対応が取れていることが知られている。

2.2.3 メル一般化ケプストラム

音声の周波数情報を表す特徴量として、メル一般化ケプストラム (MGC: Mel-generalized cepstral coefficients) という特徴量が考案されている。MGC は次式で表すことができる。

$$H(z) = \left(1 + \gamma \sum_{m=0}^M c_{\alpha, \gamma}(m) z_{\alpha}^{-m} \right)^{\frac{1}{\gamma}} \quad (2.1)$$

$$z_{\alpha}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (2.2)$$

これにより、安定的に解を求めつつスペクトル包絡の特徴をうまく再現することが可能になる。

2.3 基本周波数パターン生成過程モデル

基本周波数パターン生成過程モデル (以下 F0 モデル) とは、人間が発声する時の喉頭の動きをモデル化し近似したものである。このモデルでは基本周波数パターンを、句頭から句末に向けて緩やかな下降に対応するフレーズ成分、単語や単語の連鎖に対して局所的な起伏を示すアクセント成分と話者固有の基底周波数の和によって対数軸上で表現している。

2.3.1 フレーズ成分

フレーズ成分とは、発話内での F0 の大きな変動を示すものであり、声帯振動のはじまりより約 300 ~ 400ms 前から準備され始め、やや上昇しながら最大値に達した後、緩やかな下降を経てある一定の値に漸近し、発話の終端近くで急激に下降する成分である。単独発話の場合は 1 つ存在し、文の発話では句境界ごとに複数存在し得る。この成分は質量とばね定数を持つ 2 次の力学系が瞬間的な外力を受けた場合の運動に似ている。ここで、フレーズ成分を質量、ばね定数、摩擦抵抗を持った力学系のインパルス応答を用いて近似し、なおかつこの仮想的な力学系が線形性を持つ臨界制動系であると仮定すると、フレーズ成

分に相当するインパルス応答 $G_p(t)$ は次式で表すことができる。

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (2.3)$$

ここで α はフレーズ指令に対する系の速さを決定する係数であり、日本人話者の平均的な値として経験的に $\alpha = 3.0$ とすると良いことが知られている [8]。

2.3.2 アクセント成分

アクセント成分とは、発話内での F0 の局所的な変動を示すものであり、個々の単語や連続する単語に付随して表れる。アクセント成分は単語内で高いモーラの発音にやや先行して上昇を始めて一定の値に漸近し、高いモーラが続く間は高い値を保ち、低いモーラに移るのに合わせて先行して下降を始める。この成分は、質量とばね定数を持つ 2 次の力学系が一定時間続くステップ的な外力を受けたときの運動に似ている。フレーズ成分同様、アクセント成分に相当するステップ応答 $G_a(t)$ は次式で表すことができる。

$$G_a(t) = \begin{cases} \min[1 - (1 + \beta t) \exp(-\beta t), \gamma], & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (2.4)$$

ここで β はアクセント成分の立ち上がりの速さを決定する係数であり、こちらも平均的な値として経験的に $\beta = 20.0$ とすると良いことが知られている。また、上式の γ は実際の F0 パターンにおいて $G_a(t)$ が有限時間内に達する上限値を示しており、 $\gamma = 0.9$ として用いられる。

2.3.3 基本周波数パターンの生成

これらの 2 つの成分を用いることで、F0 パターンを非常に良く近似することができる。フレーズ成分を比較的時定数の大きい線形 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小さい線形 2 次系のステップ応答で近似できるとして、それぞれの成分が独立に表れるものと仮定すれば、F0 パターンはこれらの成分をそれぞれ定数倍したものと、話者固有の基底周波数を足し合わせることで表現できる。その模式図は図 2.2 のようになり、フレーズ成分はインパルス、アクセント成分は方形波の形で発話内に生じている。この 2 種類の成分と基底周波数が足し合わされることで、対数 F0 パターンを生成することができる。ここで、時刻 t における基本周波数を $F_0(t)$ とおくと、次式のように表現される。

$$\begin{aligned} \ln F_0(t) &= \ln F_b + \sum_{i=1}^I A_{p,i} G_p(t - T_{ps,i}) \\ &\quad + \sum_{j=1}^J A_{a,j} \{G_a(t - T_{as,j}) - G_a(t - T_{ae,j})\} \end{aligned} \quad (2.5)$$

ただし、式中の変数はそれぞれ以下のような意味を持つ。

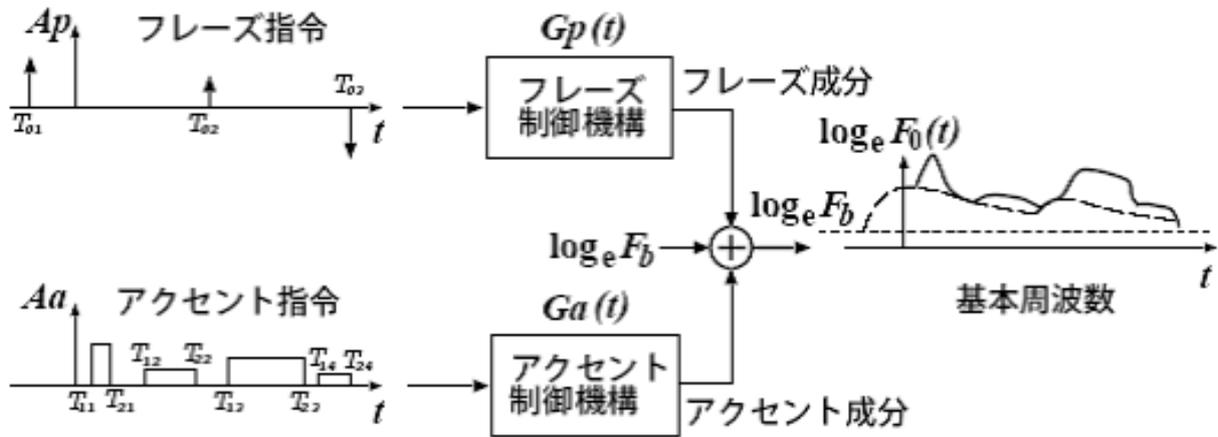


図 2.2: F0 モデルによる F0 パターンの生成

F_b : 話者固有の F0 パターンの基底値 (基底周波数)

I : 文中のフレーズ指令の数

J : 文中のアクセント指令の数

$A_{p,i}$: i 番目のフレーズ指令の大きさ

$T_{ps,i}$: i 番目のフレーズ指令の立ち上がり位置

$A_{a,j}$: j 番目のアクセント指令の大きさ

$T_{as,j}$: j 番目のアクセント指令の立ち上がり位置

$T_{ae,j}$: j 番目のアクセント指令の立ち下がり位置

2.3.4 基本周波数パターン生成過程モデルの生理学的、物理学的根拠

以上のように、F0 モデルではフレーズ成分、アクセント成分を分離して線形な力学系を仮定することで、基本周波数パターンを定量的に扱っているが、これは生理的、物理的に妥当であることが藤崎らによって示されている [9]。以下にその概要を示す。

i) 声帯の長さの変化と対数基本周波数との関係

声帯の長さは、喉頭の動きを制御することで変化している。図 2.3 に喉頭の概略図を示す。声帯は前後に引っ張られると薄い膜状になるが、その際左右の披裂軟骨を近づけることによって声門を狭め、さらに気流がある程度以上の強さになるとベルヌーイ効果で振動をはじめ、声門が開閉する。声帯のすぐ脇にある声帯筋に力を入れて引っ張るとき、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。ここで T を張力、 l を声帯の長さ、 a を $T = 0$ におけるばね定数、 b を比例定数とすると

$$\frac{dT}{dl} = a + bT \quad (2.6)$$

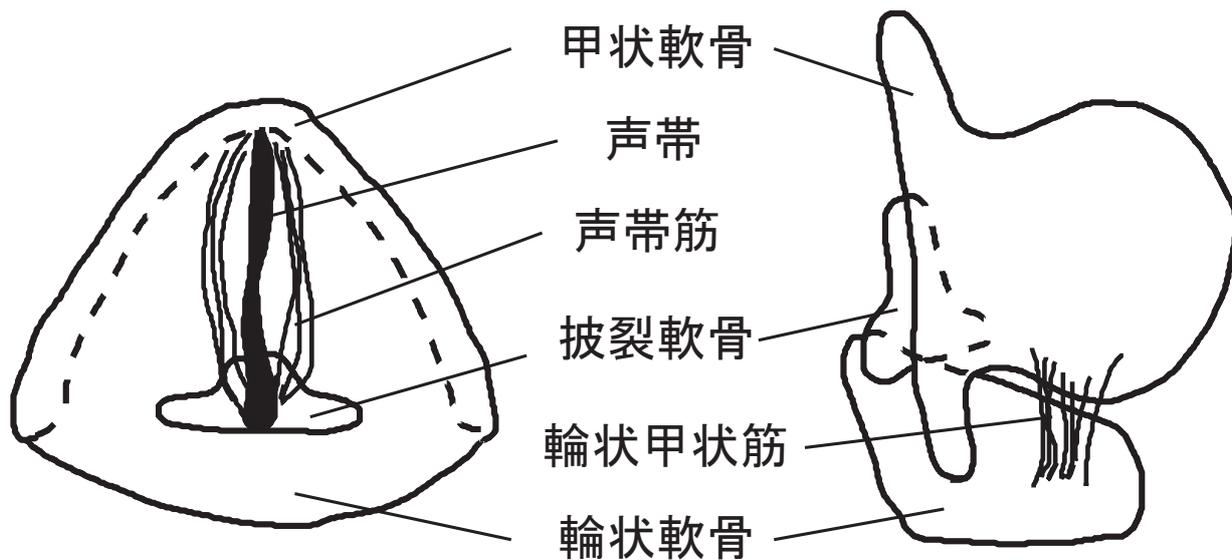


図 2.3: 喉頭の模式図

であるからこれを解いて

$$T = (T_0 + \frac{a}{b}) \exp(b(l - l_0)) - \frac{a}{b} \quad (2.7)$$

となる。ただし、 T_0 は声帯にあらかじめ加えた張力、 l_0 は $T = T_0$ のときの声帯の長さである。このとき $T_0 \gg a/b$ であるとすると

$$T \simeq T_0 \exp(bx) \quad (2.8)$$

となる。ただし x は l_0 を基準としたときの声帯の伸びである。一般的に弾性膜の固有振動数は張力の平方根に比例するので、弾性膜の振動の基本周波数は

$$F_0 = c_0 \sqrt{\frac{T}{\rho}} \quad (2.9)$$

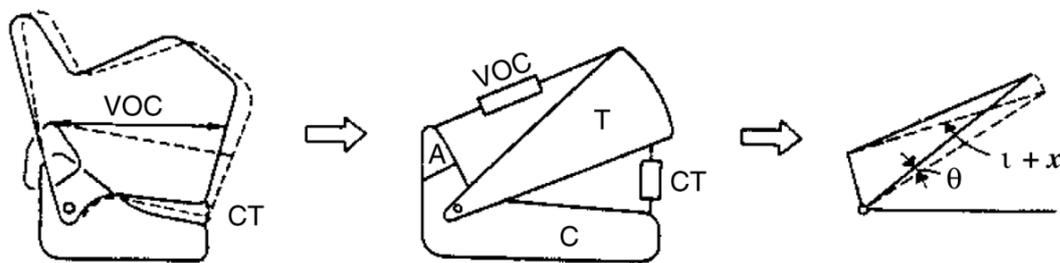
と与えられる。ただし ρ は弾性膜の単位面積あたりの密度、 c_0 は弾性膜の大きさに反比例する定数である。したがって (2.8) 式、(2.9) 式から

$$\ln F_0 = \ln F_b + \frac{b}{2}x \quad (2.10)$$

となる。ただし $F_b = \sqrt{\frac{T_0}{\rho}}$ である。また厳密には (2.10) 式の第 1 項は x によって多少変化するが、 $\ln F_0$ の値は主に第 2 項によって決まる。以上から x が時間的に変化する場合は考えると

$$\ln F_0(t) = \ln F_b + \frac{b}{2}x(t) \quad (2.11)$$

となり、これは基本周波数の対数が声帯の長さに比例して変化することを示している。



VOC : Vocalis M.
CT : Criothyroid M.

T: Thyroid
C: Cricoid
A: Arytenoid

l : Length of vocalis
 x : Elongation of vocalis
 θ : Angular displacement of thyroid

図 2.4: 喉頭の構造

ii) 喉頭の調節機構

声の高さの調節に關与する喉頭の構造を模式的に示すと、図 2.4 のようになる。輪状軟骨を基準に考えると、甲状軟骨は声帯筋と輪状甲状筋によって互いに逆の方向へ引っ張られて平衡状態を保っているが、輪状甲状筋の張力が少し強まると甲状軟骨を前へ倒そうとする。回転するとき、甲状軟骨は 1 つの質量と見なされ、声帯筋と輪状甲状筋はばねと見なすことができる。甲状軟骨と輪状軟骨をつなぐ輪状甲状関節には運動の自由度が 2 つあり、輪状甲状筋には、甲状軟骨を平行移動させる斜部と回転運動させる直部の 2 つの部分がある。この 2 つの運動を図から模式的に考えると、回転運動方程式が求まる。つまり甲状軟骨が平行移動する場合はインパルス関数入力に対する線形 2 次系の応答になり、甲状軟骨が回転する場合はステップ関数入力に対する線形 2 次系の応答になる。すなわち前者からフーズ成分が、後者からアクセント成分が生じていると見なすことができる。まとめると、甲状軟骨の平行移動と回転運動が声帯の伸びを決定し、それぞれが 2 次系の動きをするので、対数基本周波数パターンは、それらの 2 つの成分の和として反映される。平行移動による声帯の伸びを $x_1(t)$ 、回転運動による声帯の伸びを $x_2(t)$ とすると、(2.11) 式は

$$\ln F_0(t) = \ln F_b + \frac{b}{2}(x_1(t) + x_2(t)) \quad (2.12)$$

となる、以上の観点から基本周波数パターンは、フーズ成分とアクセント成分の和として表すことができ、従って F0 モデルが生理学的、物理学的根拠に基づくモデルであることが理解できる。

2.4 HMM 音声合成

HMM 音声合成システムは学習部と合成部の 2 つから構成されており、その概要は図 2.5 のようになる。

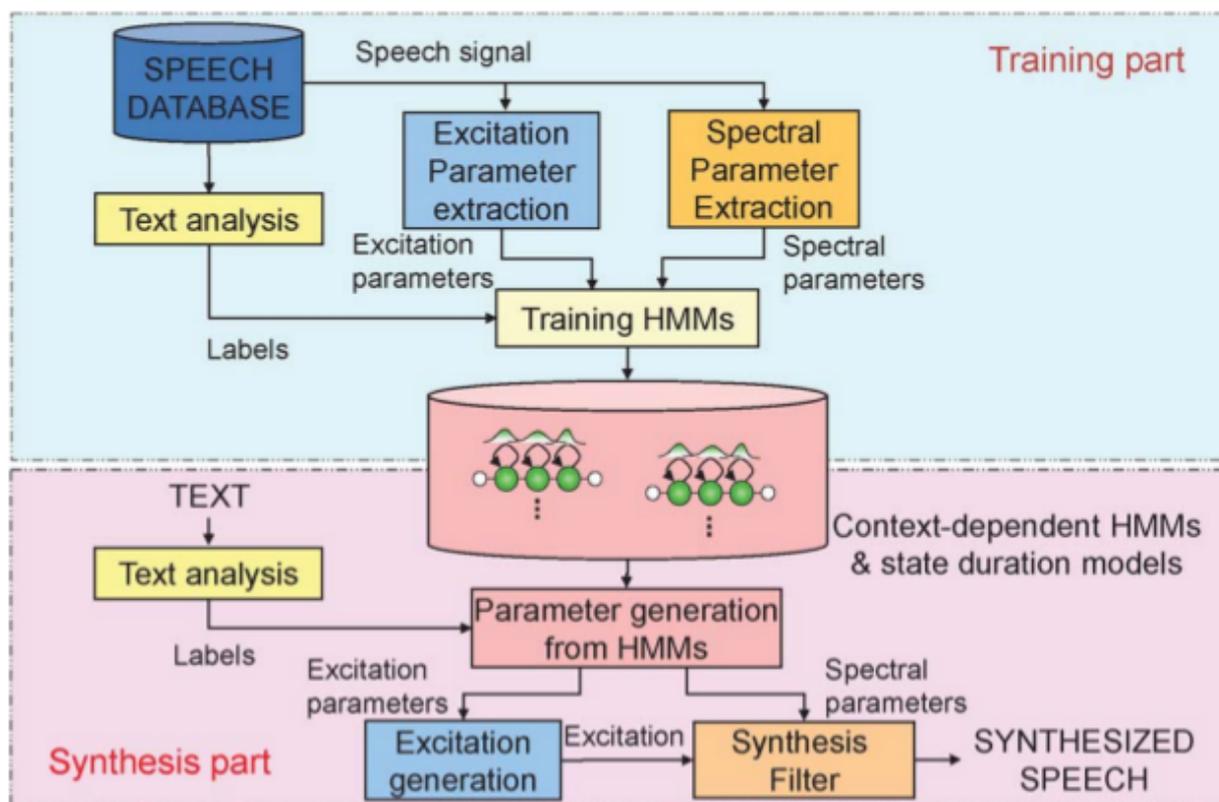


図 2.5: HMM 音声合成の概要 (文献 [3] より引用)

学習部では信号処理とテキスト処理を行っていく。信号処理の方では、音声コーパスから基本周波数やスペクトル包絡などの音響特徴量を得て HMM の学習に用いる特徴量ベクトルを形成する。テキスト処理の方では与えられた文章からその文章に対応したラベル情報を得る。これらの特徴量やラベルから HMM の学習をしてモデルを構築する。モデルの構築時には、当該音素と前後の音素情報やアクセント型、品詞、文長、文中での位置など、言語的な情報をコンテキストとして考慮するが、これは F0 や音素の継続長が言語的な情報に大きな影響を受けるからである。

合成部では、まず生成したいテキストに対応するラベルデータを入力データとして学習した HMM に与える。HMM はコンテキスト情報が一致する音素単位の HMM を連結した文 HMM を形成し、ここから尤度最大化基準の元で音声合成に必要な特徴量ベクトルを生成する。特徴量ベクトルを得たところでフィルタを通して目的とする音声を合成する。合成時には動的特徴量や発話内変動を考慮することで、自然な合成音声を実現している [10]。

HMM 音声合成では音声データからパラメータを取り出してモデルを構築していくため、学習時や合成時にパラメータを適切に扱うことで柔軟な音声合成を可能にしている。話者適応技術による HMM の変換や [11]、複数の話者 HMM の補間による音声モーフィングや [12]、感情や発話スタイルを補間した柔軟な音声合成などを行うことが出来る [13]。その一方で HMM では特徴量をフレーム単位で扱うため、アクセントや抑揚といった長時間に渡る韻律的特徴のモデル化は困難であるという問題もある [14]。

ここで音声の F0 を扱うモデルとして、次に述べる基本周波数パターン生成過程モデルがある [15]。このモデルは人間の発声の生理的な機構と対応しており、少数のパラメータで音声の基本周波数パターン (以下 F0 パターン) を表すことが可能である。HMM 音声合成と組み合わせて F0 パターンの制御を行ったものとして、松田 [16] や橋本 [17] の研究が上げられる。

第3章

基本周波数パターン生成過程モデル による基本周波数差分の導入

3.1 はじめに

この章では、まず第2章で記したF0モデルを基にした基本周波数差分という考え方を導入する。その後、基本周波数差分を用いることでHMM学習を改善する手法を提案していく。

3.2 基本周波数差分

HMM音声合成の枠組みでHMMの学習に用いる音声データは、時々刻々と変化するF0パターンを持っていると考えられる。ここで音声进行分析して得られたF0パターンの対数値と、F0モデルによって近似したF0パターンの対数値の、各フレームごとの差を基本周波数差分(以下F0差分)と呼ぶことにする。算出方法は以下の式の通りである。

$$F_{0diff}(t) = |\ln F_{0obs}(t) - \ln F_{0model}(t)| \quad (3.1)$$

$F_{0obs}(t)$ と $F_{0model}(t)$ はそれぞれ時刻 t における、抽出した F_0 と F_0 モデルによって近似された F_0 を表す。対数値を扱うのは、音声の周波数に対する人間の感覚は対数的であることが理由となる [18]。またF0差分は全て絶対値で扱うこととする。

分析によって得られたF0パターンが滑らかである場合は、そのパターンをF0モデルにより生成されるF0パターンで凡そ表現することができる。このとき、F0差分は小さくなる。しかし、分析によって乱れたF0パターンが得られた場合、乱れたフレームではF0モデルとの乖離が発生するためF0差分は大きくなってしまふ。F0モデルにより生成されるF0パターンはF0抽出誤りに頑健であることから、F0差分の大小から学習に問題のあるサンプルを指摘することが可能ではないかと考えられる。音声分析時のF0の乱れは同時にスペクトル包絡の推定にも影響を与えるため、F0差分が大きくなるようなサンプルをHMM学習に含めないことにより、合成音声の品質向上が期待できる。

図3.1では、音声分析で得たF0パターンと、推定したF0モデルパラメータによるF0パターンを示している。点線は各フレームの抽出されたF0を、実線はF0モデルパラメータから求められるF0パターンを示しており、実線と点線の差がF0差分になる。図3.1の「四畳半のアパートの我が家には、3人ものお客様が見えた」という文章の中で、「よ'じょ'うはん」に当たる部分では半ピッチの抽出ミスが見られており、F0差分が非常に大きくなっている。他にも、いくつかの部分で不自然に飛び飛びなF0パターンが得られてしまっており、そういった点ではF0差分が大きくなっている。

3.3 基本周波数差分が大きくなる要因

ここで、F0差分が大きくなる要因に着目してみる。F0差分が大きいということは、抽出されたF0とF0モデルで近似されたF0に大きなずれが生じていることになる。その原因としてはF0抽出側に問題があるか、F0モデルでは無視されているF0パターンが出現

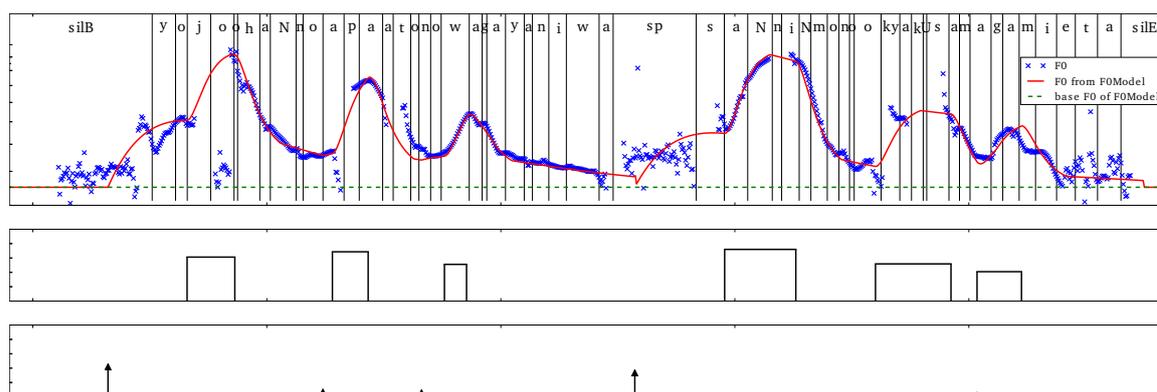


図 3.1: 抽出された F0 と F0 モデルパラメータによる F0(文:四畳半のアパートの我が家には、3 人ものお客様が見えた。)

しているなどが考えられる。先ほども挙げた通り倍ピッチや半ピッチは、F0 抽出に問題がある一例となるが、調べてみると大まかに以下のような場合が見られた。

1. 倍ピッチ・半ピッチエラー
2. 局所的なエラー
3. 子音の立ち上がり
4. 音素区間終端での下降・上昇
5. 音素区間内での乱れ
6. アクセント指令・フレーズ指令の推定ミス

倍ピッチ・半ピッチは、抽出時の処理の関係で本来の 2 倍もしくは半分の周波数が F0 として検出される場合である。この抽出ミスは F0 抽出ミスの代表的な例である。局所的なエラーとは、あるフレームにおいて局所的に抽出に失敗している場合である。これは、前述の倍ピッチ・半ピッチと被るケースもある。子音の立ち上がりとは、音素「k」や「t」のような破裂音などに多く、F0 を検出する際に子音特有の高周波部が F0 モデルでの近似より鋭くなっている場合である。音素区間終端での下降・上昇は、図に見られるように音素の末尾に向けて F0 が下降あるいは上昇している場合である。音素区間内での乱れとは、音素区間全体に渡って F0 が飛び飛びになっている場合である。また、今までの分類に該当しないものもここに含むものとする。アクセント指令・フレーズ指令の推定ミスとは、F0 モデルで近似する際にアクセント指令やフレーズ指令の指定が正しく行われていないため、本来正しいと思われる F0 を近似できていない場合である。図では、本来アクセントがある区間に対してアクセント指令を置いていないため、F0 モデルによる近似が失敗しており F0 差分が大きくなってしまっている。この場合は F0 の抽出ミスではなく、F0 モデルパラメータの推定ミスということになる。

一言に F0 差分が大きいと言っても様々な要因が存在することが分かる。これらの違いを上手く取り扱うことで、HMM 学習の質を高めていくことが可能になると考えられる。

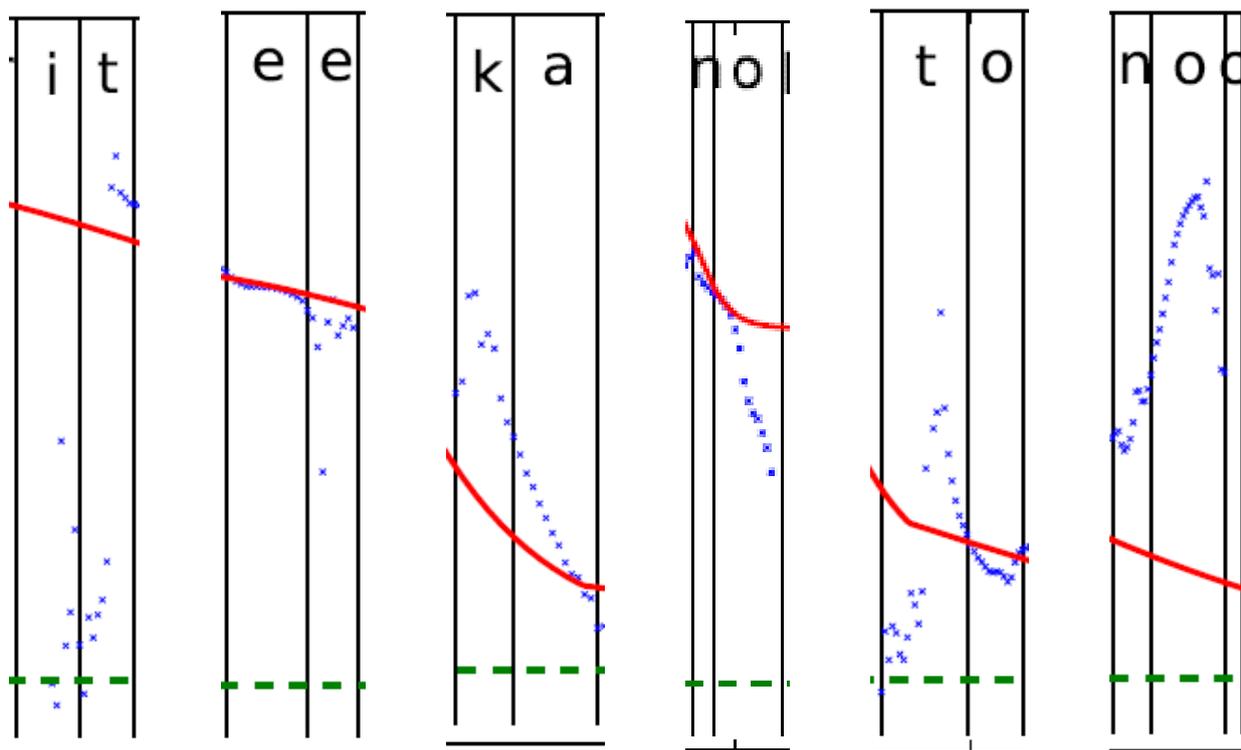


図 3.2: F0 差分が大きくなる要因の例。青い点が抽出された F0 で、赤い線が F0 モデルによる近似された F0 である。要因はそれぞれ左から順に上記の箇条書きに対応している。文章は ATR 日本語コーパスのものでありそれぞれ左から順に、A38 「普通、中距'離'トラックのドライバーは、中年の人が多い。」、B32 「リードが大きければ、牽'制'球を投げなければならない。」、A02 「一週間ば'か'り、ニューヨーク取材した。」、A11 「自然'の'研究者は、自然をねじ伏せようとしてはいけない。」、A45 「ワイン'と'日本酒とを問わず、原産地、成分表示を急ぐべきではないか。」、A29 「母は'脳'血栓の後遺症で、老人性痴呆症になり、一年前から入院中です。」、である。' 'で囲った文字の区間で、基本周波数差分が大きくなっている。

3.4 F0 差分を利用した HMM 学習の改善

ここで、F0 差分を用いて HMM 学習時のサンプル選択により HMM 音声合成を改善する手法を提案する。

1 つは、F0 差分の大小に着目し、HMM の学習対象からある音声区間を除外する手法である。F0 差分が大きい部分では F0 の抽出ミスや F0 モデルから極端に外れた F0 が存在しており、こういった学習サンプルを除外することは全体の品質の向上につながると考えられる。学習サンプルの扱いについては、単純な除外だけでなく、F0 モデルで近似した F0 パターンによる補間なども考えられる。これについては、第 4 章で詳細に検討していく。

もう 1 つは、同じく F0 差分の大小に着目し、HMM の学習時にフルコンテキストレベルに情報を乗せる手法である。先ほど紹介した通り F0 差分が大きくなる要因は様々あるの

で、それらが HMM 学習に対してどのように影響を及ぼすのか検討していくのが目的となる。これについては、第 5 章で詳細に検討していく。

第4章

基本周波数差分の大小を基にした 学習コーパスの評価・選択

4.1 はじめに

この章では、第3章で挙げた基本周波数差分を基に、HMM学習に用いる音声コーパスの評価と選択する手法をまとめる。実験は2つの手法で行った。1つめは基本周波数差分の大小を基に、学習コーパスを音素単位で選択する手法で、2つめは前述した手法で除外された区間に対し、基本周波数パターン生成過程モデルによるデータ補間を行う手法である。

4.2 実験条件

2つの実験に共通する実験条件を記す。音声データはATR日本語音声データベース [19] から男性話者 MMI を選び、全 503 文のうち、サブセット A から I までの 450 文で HMM を学習し、サブセット J の 53 文を合成して評価した。音声の分析は STRAIGHT [20] を用いて、F0、スペクトル包絡、非周期性指標を抽出した。分析条件は、フレーム周期 5 [msec] であり、F0 の探索範囲は話者 MMI では最小値 80 [Hz]、最大値 250 [Hz]、基底周波数は F0 の探索範囲の最小値と同じ値にした。HMM の学習に用いた特徴量は、0 から 39 次元までのメルケプストラムと 0-1、1-2、2-4、4-6、6-8 [kHz] の 5 帯域の平均非周期性指標、対数 F0、およびそれらの Δ 、 Δ^2 を含めた 138 次元のベクトルとした。メルケプストラムと平均非周期性指標は、スペクトル包絡特徴量と非周期性指標からそれぞれ SPTK¹ を用いて求めた。HMM は HTS-2.1² を用いて構築した。状態継続長分布を明示的に含んだ 5 状態 left-to-right HSMM を用い、各状態の出力は単一の対角共分散ガウス分布とし、決定木によるコンテキストクラスタリングを行い、木の停止基準には MDL 基準を用いた。

主観評価には、従来手法として学習コーパスを全て用いて HMM の学習を行う場合を想定し、その合成音声と、提案手法による合成音声との聴取実験による比較を行った。客観評価には、先ほどと同じ従来手法による合成音声と提案手法による合成音声との F0 歪みを計算して比較を行った。

4.3 実験 1：音素単位での学習サンプル選択

4.3.1 実験 1 の条件

まず、学習サンプルを音素単位で選択する手法を検討した。音素単位で見れば、学習に不要と思われる音声区間のみをよりの確に除外できると考えられるからである。

初めに音素アラインメントを求め、音響特徴量の抽出を行った。ここで、F0 抽出の問題で本来発話の無い無音区間にも F0 が抽出されてしまうことが判明した。HMM の学習対象を音素単位で検討する場合には、発声が無い無音区間と発声がある区間を同等に扱うのは不自然と考えられる。そのためこれらの区間について、つまり無音区間を意味する音素「silB、sp、silE」の中で F0 が抽出された場合は、その F0 を削除することにする。このよ

¹SPTK, <http://sp-tk.sourceforge.net/>

²HTS, <http://hts.sp.nitech.ac.jp/>

表 4.1: 実験条件

HMM 学習に用いる特徴量	対数 F0、0-1、1-2、2-4、4-6、6-8[kHz] の非周期性指標、0 ~ 39 次元のメル一般化ケプストラムとそれらの Δ 、 Δ^2
フレーム周期	5msec
話者	男性話者 MMI
F0 探索範囲	MMI:80 ~ 250[Hz]、FTY:150 ~ 500[Hz]
学習文章数	450 文
合成文章数	53 文

表 4.2: 実験環境

音素アラインメントの取得	Julius
特徴量抽出	STRAIGHT、SPTK
音声合成	STRAIGHT
HMM の構築	HTS-2.1
音声データ	ATR 日本語音声データベース

うな処理を施した F0 に対し F0 モデルパラメータを推定し、抽出された F0 パターンと、推定された F0 パラメータによる F0 パターンとの F0 差分を各フレームごと計算する。

ここで音素単位で学習サンプルを選択するために、図 4.1 のように音声を分割する手法を取ることにした。この処理によって HMM 学習から除外すべき音素が決定される度に、HMM 学習に用いる特徴量及びラベルデータを分割して学習に不要な音素区間を除外する。音素区間の除外基準を決定するために、F0 が抽出された各音素区間ごとその音素内の F0 差分の最大値を求め、学習コーパスの全音素区間から F0 差分最大値となる上位 5%、10%、30% に相当する音素区間までを除外の対象として、残った音素区間を用いて HMM を学習した。この基準により話者 MMI では F0 が抽出された 23609 音素中 1180、2360、7082 音素を除外対象とした。除外された音素は表 4.3 の通りになる。

従来手法による合成音声と提案手法による合成音声のどちらがより自然であるかを話者 MMI については 9 人、話者 FTY については 8 人の被験者が主観評価した。実験では J セットの 53 文からランダムで 20 文を選び、提案手法による音声と従来手法による音声の 2 つを評価してもらった。話者 1 人につき、20 文に対し 3 段階の除外手法による計 60 文の対比較を行った。評価は 5 段階とし、提案手法の方が優れているときを +2、提案手法の方がやや優れているときを +1、どちらともいえないときを 0、従来手法の方がやや優れているときを -1、従来手法の方が優れているときを -2 とスコアを付けた。

また、 F_0 歪みによる客観評価も行った。比較対象のうち録音音声とは予め録音された音声データとの比較であり、録音音声+除外とは録音音声进行分析して除外基準と同等以上の F0 差分が検出される音素区間は評価に使用しない比較である。このとき、学習からの除外音素が 5% の合成音声には同じく 5% の除外基準、10% には 10% の除外基準というようにそ

表 4.3: 除外音素・学習音素一覧

音素	除外 5%	除外 10%	除外 30%	総音素数
o	194	354	887	3438
a	203	348	958	3327
i	100	171	508	2248
u	57	115	433	2102
e	79	169	522	1843
k	88	210	624	1265
n	8	17	84	1178
t	109	216	659	1100
r	11	19	89	1085
sil	0	0	0	900
N	46	101	320	845
m	8	20	98	747
pau	0	1	2	690
g	19	28	105	652
s	29	107	305	644
d	9	32	141	569
sh	19	54	181	561
y	2	15	53	424
h	27	52	171	401
cl	61	83	123	388
w	5	7	35	378
b	3	13	78	333
j	25	50	149	333
I	2	4	13	291
U	0	3	9	261
ch	17	37	104	247
ts	13	28	99	233
z	7	15	86	204
f	20	45	88	143
p	7	16	53	123
ky	9	20	47	100
ry	1	2	10	37
gy	1	2	18	31
ny	0	1	5	27
hy	1	1	16	25
by	0	1	5	10
my	0	1	1	7
py	0	2	3	6
dy	0	0	0	1

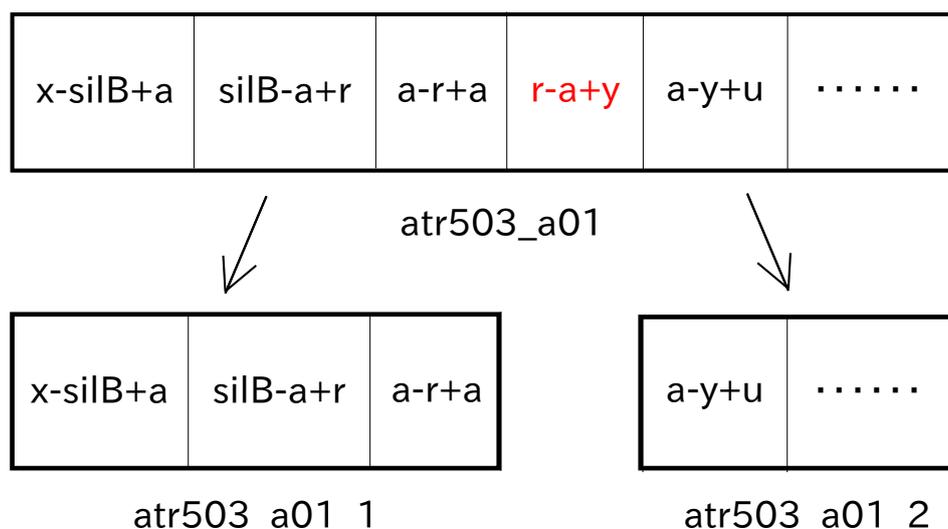


図 4.1: 学習データの除外・分割手法 (文「あらゆる現実を全て自分の方へねじ曲げたのだ」冒頭部分の音素「r-a+y」を除外する例)

れぞれ対応させて比較を行った。評価に用いる合成音声の合成時には録音音声の時間情報と同期するように合成を行った。評価は以下の式で行った。

$$RMSE = \sqrt{\frac{\sum (f(t) - \hat{f}(t))^2}{n}} \quad (4.1)$$

ここでは双方の音声が有声であるフレームだけを考慮して、時刻 t における F_0 の自然対数値を用いて算出する。また、 n は双方の音声が有声であるフレーム数である。

4.3.2 結果

結果は表 4.4 のようになった。話者 MMI で除外音素が 5% の場合には、主観評価のスコアが 5% の優位水準で 0.300 ± 0.148 となり、提案手法の優位性が示された。表 4.5 では主観評価のスコアをまとめている。これを見ると、0 点の評価が非常に多いことが分かる。表 4.6 では F_0 歪みをまとめているが、録音音声との比較では従来手法より歪みが大きくなったことが分かる。録音音声+除外の条件では、10% 除外による合成音声は F_0 歪みが従来手法を下回った。図 4.2 では、 F_0 パターンが改善された例を示す。「表現する能力を身につけることだ」という文章の中で、「表現」の部分など全体的に F_0 の乱れが収まっている点や、「能力」「身につける」の部分でわずかではあるが F_0 パターンがアクセントに従って自然になっていることが確認できる。

4.3.3 考察

この実験では主観評価により優位性を示すことができたが、これは学習サンプルの選択を音素単位で行う手法の有効性を示している。その一方でどの条件においても 0 点の評価

表 4.4: 実験 1 の結果

話者	除外音素	95%信頼区間
	5%	0.300±0.148
MMI	10%	0.139±0.147
	30%	-0.378±0.162

表 4.5: 実験 1 の主観評価スコア

話者	除外音素	+2	+1	0	-1	-2
	5%	23	51	69	31	6
MMI	10%	16	44	81	27	12
	30%	9	31	52	59	29

表 4.6: 実験 1 の客観評価スコア

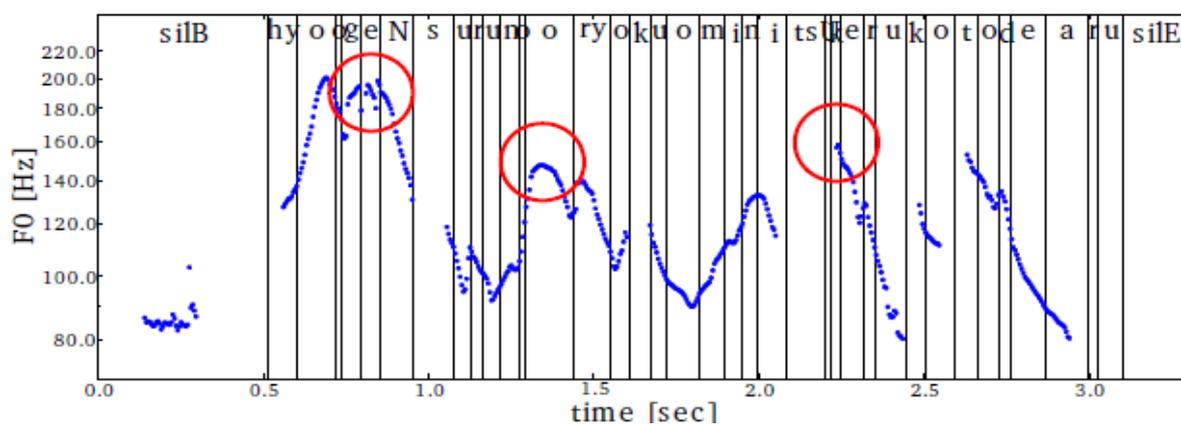
合成音声	録音音声	録音音声+除外
除外なし学習	$1.31 * 10^{-1}$	
5%除外学習	$1.34 * 10^{-1}$	$1.32 * 10^{-1}$
10%除外学習	$1.33 * 10^{-1}$	$1.29 * 10^{-1}$
30%除外学習	$1.40 * 10^{-1}$	$1.32 * 10^{-1}$

が多いが、これはHMMの変化に関係なく同じように合成される音声が多いことを意味しているのではないかと考えられる。どちらの話者を見ても除外音素の割合が増加するにつれて結果が悪くなる傾向が見られるが、これはHMMの学習量減少による音声の品質低下なので妥当な結果と言える。

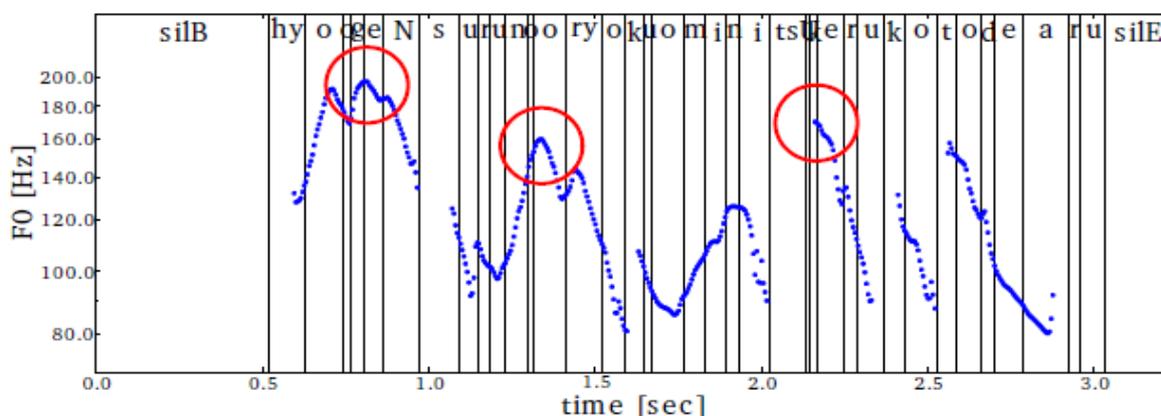
一方客観評価については、主観評価とはすぐわない結果が出た。録音音声との比較では、提案手法が除外しているようなF0差分が大きくなる区間を含めた状態での比較になるため、そのような傾向を再現できないことがF0歪みの増大につながっているのではないかと考えられる。その点を考慮した録音音声+除外の手法では、どの場合もF0歪みが小さくなっており、F0差分が大きい区間によるF0歪みの増大を示せたのではないかと考えられる。

今回の実験では音素区間ごとにF0差分の最大値を基準に上位何割かの音声区間をHMM学習から除外する手法を取ったが、あくまで話者MMIに適用した場合の結果なので、除外基準を割合で決めるのではなく、F0差分の最大値に対してある閾値を設けて除外する音素区間を決定する必要がある。F0差分が大きくなるような話者についてはより正しく問題となる区間を指摘するため、F0差分が小さくなるような話者については必要以上に学習データを削除しないように検討していくことで適切な閾値が見つかるものと思われる。

今回の実験で除外対象となった音素には偏りが生じている。そこで、全音素区間を同等



(a) 従来手法



(b) 提案手法

図 4.2: 合成音声の F0 パターンの比較 (文：表現する能力を身につけることである。)

に扱うのではなく音素ごとにそれぞれ除外する基準を設定する手法も考えられる。特に、コーパス中の出現数が少ない音素については、HMM 学習から除外したときの学習効果低減が激しくなるので慎重に扱う必要がある。

実験では音素区間ごとに各フレームの F0 差分の最大値を基準に除外する音素を決定したが、音素区間ごとに各フレームの F0 差分の平均値を基準に除外する方法も検討した。この手法による音声も聞いてみたが、改善があまり目立たなかった。最大値を用いる基準では倍ピッチの抽出ミスや局所的な不自然な F0 の変動に対して敏感に反応する一方で、平均値を算出すると局所的な抽出ミスの影響が小さくなって、除外したい音素区間をあまり的確に発見できず、HMM の改善が少なくなるからではないかと考えられる。

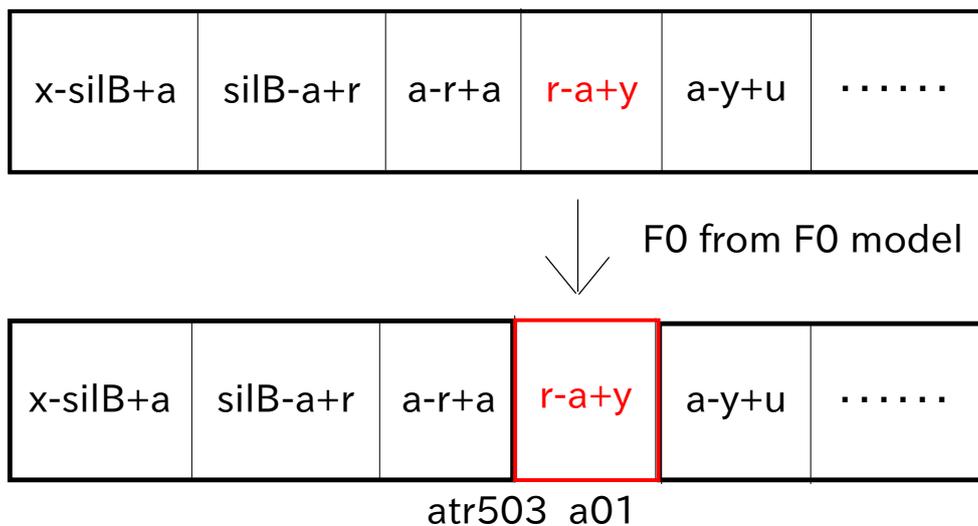


図 4.3: 学習データの補間手法 (文「あらゆる現実を全て自分の方へねじ曲げたのだ」冒頭部分の音素「r-a+y」を除外した後、補間している)

表 4.7: 実験 2 の客観評価スコア

合成音声	録音音声	録音音声+除外
除外なし学習	$1.31 * 10^{-1}$	
5%除外+F0	$1.38 * 10^{-1}$	$1.35 * 10^{-1}$
10%除外+F0	$1.41 * 10^{-1}$	$1.36 * 10^{-1}$
30%除外+F0	$1.29 * 10^{-1}$	$1.11 * 10^{-1}$

4.4 実験 2 : F0 モデルによる学習サンプルの補間

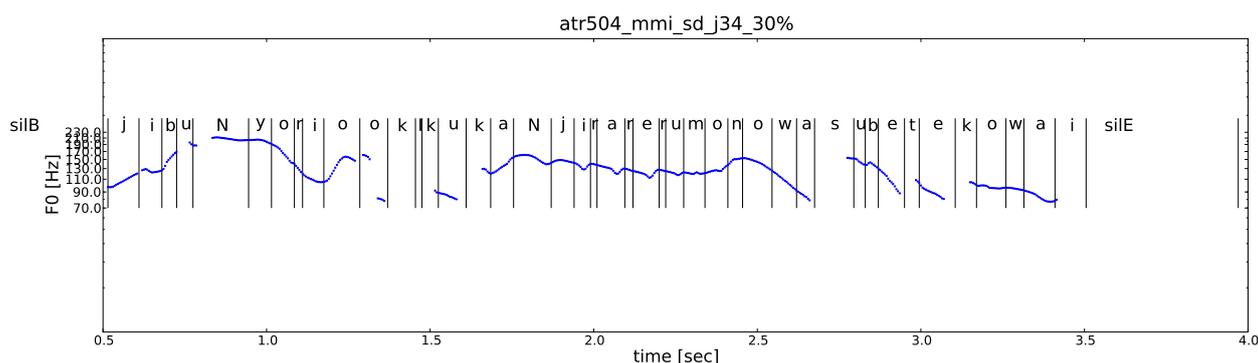
4.4.1 実験 2 の条件

実験 2 では、実験 1 で HMM 学習から除外された音声区間に対し、F0 モデルの F0 によってデータを補間することでデータの除外・削減を行わない手法を取ることにした。類似するの手法として橋本の手法が挙げられるが、こちらでは全ての学習データの F0 を置き換えている。

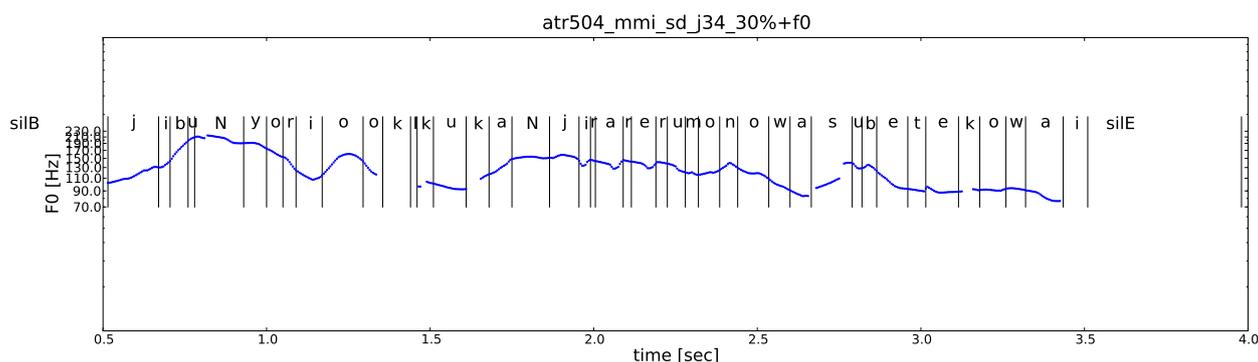
4.4.2 結果

表 4.7 を見ると、実験 1 と客観評価のスコアを比較したときに 5%除外と 10%除外では F0 歪みが増大しているが、30%除外では F0 歪みが減少している。特に録音音声+除外との比較では大きく歪みが減少している。

図 4.4 は除外 30%の学習による合成音声と、除外 30%+F0 の学習による合成音声の F0 パ



(a)30%除外による手法



(b)30%除外+F0による手法

図 4.4: 合成音声の F0 パターンの比較 (文：自分より大きく感じられるものは全て怖い。)

ターンを比較している。除外 30%による音声では前半の「じ'ぶ'んより」や「'お'おきく」の区間で F0 が途切れて乱れているが、除外 30%+F0 による音声では滑らかな F0 パターンが得られていることが分かる。

4.4.3 考察

この手法では、除外による学習量の減少を防ぐこと、また F_0 モデルによる滑らかな F_0 を学習に取り入れることで合成される音声の F_0 も滑らかになり自然な音声になることが期待され、実際にそのような効果が得られることが分かった。学習量の減少を防ぐ点についても、どの除外割合においても前述の除外手法のような学習量減少による劣化した音声はあまり生成されておらず、効果があったものと考えられる。客観評価のスコアについては、学習に用いる F_0 を F_0 モデルで近似した F_0 で置き換えたことで境界でのずれが生じるなど、少し F_0 歪みが大きくなるのは妥当な結果だと思われる。しかし、除外 30%においては F_0 歪みが小さくなっている。これは、多くの区間を平滑化された F_0 で置き換えたことで、結果的に歪みが小さくなったのではないかと考えられる。また、除外される音素区間が連続することで、除外対象となる音素区間が少ない場合に比べ不連続な点が少なくなったことが原因だと考えられる。

F_0 の置き換えを行うことで学習データの削減を防ぐことが狙いであったが、特に除外

30%のときはその効果が出ていたと思われる。図を見ると、除外30%の学習による音声では文章の前半でF0パターンが途切れがちになっているが、提案手法の学習による音声では滑らかなF0パターンが出力されている。この手法では、実験1で効果があった除外の程度が少ない場合よりも、学習に問題があると思われる区間を多く指摘した場合にこそ有用であると考えられる。

その一方でこの手法では、逆に学習に問題であると指摘した区間だけF0を置き換えるだけでは、除外対象となった音素区間の境界でF0の不連続が発生する部分もあることや、スペクトル情報との齟齬が発生する面もある。そのため、より適切なデータの扱いをすることで更に高い効果が得られるのではないかと考えられる。

第5章

基本周波数差分の大小を基に 付けたラベルによるHMM学習

5.1 はじめに

この章では、第4章で行った実験から派生して、学習データの選択ではなく学習時に用いるラベルデータに基本周波数差分の情報を付与することでどのような効果が得られるか検討してみた。ここでは、基本周波数差分の大小を付与した実験と、更に大小の要因を付与した実験の2つを行った。

5.2 フルコンテキストラベル

まず、HMM 音声合成の学習時にどのようなラベルが用いられているか述べる。HTS¹では、HMM 音声合成に以下の表 5.1 のようなフルコンテキストラベルを定義している。これによって、学習時に当該音素だけでなく先行・後続音素を考慮したり、文章の品詞情報やアクセント情報など韻律に関わる情報を考慮した学習を実行できる。

続いて、今回の実験で提案するラベルを掲載する。

除外基準のラベルでは、それぞれ除外 5%、10%、30%の順に 1,2,3 と指定している。それ以外の音素には 0 を指定した。除外要因のラベルでは、第3章で紹介した除外要因に対してラベル付けを行い、倍ピッチ・半ピッチエラー、局所的なエラー、子音の立ち上がり、音素区間終端での下降・上昇、音素区間内での乱れ、アクセント指令・フレーズ指令の推定ミス順に 1~6 と指定している。それ以外の音素には 0 を指定した。ラベリングでは話者 MMI の A~I セット 450 文のうち、F0 差分が大きい上位 10%の音素区間までを対象として、作業は筆者 1 人が行った。ラベリングの結果、対象となった 2360 音素の要因ごとの出現数は以下の表の通りになっている。更に細かいラベル情報は、末尾の付録を参照することにして、ここでは割愛する。

除外の要因として多くなったのは子音の立ち上がりと音素区間終端での下降・上昇である。これは、多くの子音や母音に傾向のため個数が多くなっている。倍ピッチ・半ピッチエラーに着目すると 148 個出現しており、これは音素区間で見ると 6%以上の音素でこれらのエラーが出現したことになる。アクセント指令・フレーズ指令の推定ミスも一定数見られたが、この推定精度は橋本の論文の結果とある程度の一致を見せている。

5.3 実験 1：基本周波数差分の大小を付与する手法

5.3.1 実験 1 の条件

実験 1 では、先ほど提案したラベルのうち除外基準についてのみ付与したフルコンテキストラベルを用いて HMM 学習を行った。ここでは 3 つの実験を、それぞれ 5%除外まで、10%除外まで、30%除外までを指定した状態のラベルを用意して行った。合成時に指定するラベルでは除外基準のラベルを 0 とし、除外基準に引っかけられない音声を生成するように指定した。それ以外の実験条件は第4章で述べた通りである。

¹HTS, <http://hts.sp.nitech.ac.jp/>

表 5.1: 従来手法によるコンテキストラベル

先行音素
当該音素
後続音素
アクセント句内モーラ位置 (単位: モーラ)
アクセント型とモーラ位置との差 (単位: モーラ)
先行品詞 ID
先行品詞の活用形 ID
先行品詞の活用型 ID
当該品詞 ID
当該品詞の活用形 ID
当該品詞の活用型 ID
後続品詞 ID
後続品詞の活用形 ID
後続品詞の活用型 ID
先行アクセント句の長さ (単位: モーラ)
先行アクセント句のアクセント型
先行アクセント句と当該アクセント句の接続強度
先行アクセント句と当該アクセント句間のポーズの有無
当該アクセント句の長さ (単位: モーラ)
当該アクセント句のアクセント型
先行アクセント句と後続アクセント句の接続強度
当該呼気段落でのアクセント句の位置
疑問文かそうでないか
後続アクセント句の長さ (単位: モーラ)
後続アクセント句のアクセント型
後続アクセント句と当該アクセント句の接続強度
後続アクセント句と当該アクセント句間のポーズの有無
先行呼気段落の長さ (単位: モーラ)
当該呼気段落の長さ (単位: モーラ)
文中での当該呼気段落の位置
後続呼気段落の長さ (単位: モーラ)
文の長さ (単位: モーラ)

表 5.2: 基本周波数差分の大小を基にしたラベル
基本周波数差分の大小を基にした除外基準
基本周波数差分が大きくなる要因

表 5.3: 除外基準 5% までの除外要因の個数

除外要因	出現数	出現数が多い音素
倍ピッチ・半ピッチエラー	147	a(00)
局所的なエラー	192	a(00)
子音の立ち上がり	225	a(00)
音素区間終端での下降・上昇	355	a(00)
音素区間内での乱れ	77	a(00)
アクセント指令・フレーズ指令の推定ミス	184	a(00)

表 5.4: 除外基準 10% までの除外要因の個数

除外要因	出現数	出現数が多い音素
倍ピッチ・半ピッチエラー	148	a(00)
局所的なエラー	353	a(00)
子音の立ち上がり	641	a(00)
音素区間終端での下降・上昇	776	a(00)
音素区間内での乱れ	151	a(00)
アクセント指令・フレーズ指令の推定ミス	291	a(00)

5.3.2 結果

客観評価のスコアでは、第 4 章の実験 1 と同程度かやや小さくなっている。また、録音音声+除外の比較では、従来手法である除外なし学習に対しても小さい値を示している。

5.3.3 考察

ラベル情報をつけることで、第 4 章の枠組みでは音素区間まるごと除外された区間でも学習に用いることが出来るものがあると F0 歪みに着目しても、第 4 章の実験 1 と比較して同程度かやや小さくなっている。これは生成された音声が一方向、それぞれの手法で比較すると、これも第 4 章の実験 1 の主観評価のスコアが高いものほど音声の品質も上がっている。これは、除外ラベル以外の音素区間によって作られる集合の大きさが違うことが原因で、5%除外ラベルの場合は残りの大きな集合が 95%の音素で作られるのに対し、10%除外ラベルでは残りの 90%、30%除外ラベルでは残りの 70%によって主な集合が作られることになる。すると、合成でメインで使う大きな集合が小さくなれば小さくなるほど学習効果も下がっていくので、このような結果に繋がるのが妥当だと言える。この除外基準につい

表 5.5: 実験 1 の客観評価スコア

合成音声	録音音声	録音音声+除外
除外なし学習	$1.31 * 10^{-1}$	
5%除外ラベル	$1.33 * 10^{-1}$	$1.31 * 10^{-1}$
10%除外ラベル	$1.34 * 10^{-1}$	$1.30 * 10^{-1}$
30%除外ラベル	$1.33 * 10^{-1}$	$1.24 * 10^{-1}$

表 5.6: 実験 1 の客観評価スコア

合成音声	録音音声	録音音声+除外
除外なし学習	$1.31 * 10^{-1}$	
5%除外+要因ラベル	$1.32 * 10^{-1}$	$1.30 * 10^{-1}$
10%除外+要因ラベル	$1.33 * 10^{-1}$	$1.30 * 10^{-1}$

でも、単純に割合で組むのではなくある絶対的な閾値を設定することで他の話者にも適用できるものと考えられる。

5.4 実験 2: 基本周波数差分が大きくなる要因を付与する手法

5.4.1 実験 2 の条件

続いて実験 2 では、先ほど提案したラベル全てを付与したフルコンテキストラベルを用いて HMM 学習を行った。ここでは 2 つの実験を、それぞれ 5%除外+その要因、10%除外+その要因を指定した状態のラベルを用意して行った。合成時のラベルには、除外基準、除外要因共に 0 を指定して、除外対象と関係ない音声を生成するように指定した。それ以外の実験条件は先ほど同様である。

5.4.2 結果

客観評価の値では実験 1 と同じような結果が得られており、録音音声+除外との比較ではわずかに従来手法より歪みが小さくなった。

5.4.3 考察

客観評価の結果を見ると、こちらの実験も大まかには先ほどの実験 1 と同じような傾向が得られていると考えられる。

ここで、生成された 2 分木を見ていくと、除外要因 6 (アクセント指令・フレーズ指令の推定ミス) と要因 0 (指定なし) や、除外要因 3 (子音の立ち上がり) と要因 0 (指定なし)、除外要因 4 (音素区間終端での下降・上昇) と要因 0 (指定なし) がまとめられた葉

が出現している。これは、アクセント指令・フレーズ指令の推定ミスは通常の音声区間と同等に扱うことができる、子音の立ち上がりや音素区間終端での下降・上昇は F_0 差分が大きくなるものの F_0 差分が小さい区間と同等に扱っても問題ないということが示唆される。

第6章

結論

6.1 まとめ

本論文では、基本周波数パターン生成過程モデルを用いて基本周波数差分という考え方を導入し、HMM 音声合成に用いる学習コーパスの評価・選択ならびに、ラベルデータの付与による学習への効果を検討した。第2章では、音声の生成過程と音響特徴量、基本周波数パターン生成過程モデルと HMM 音声合成システムについて述べた。第3章では、HMM 学習時に用いる音響特徴量の F_0 について、分析で得られた F_0 と、推定した F_0 モデルパラメータによる F_0 の差分を取って基本周波数差分を求め、基本周波数差分を基準にして HMM 学習を改善する手法を提案した。第4章では、基本周波数差分の大小によって HMM の学習から一定の音素区間を除外する手法を提案し、合成音声の評価実験を行った。その結果、提案手法による HMM を用いた音声合成は主観評価実験の結果、優位性があることを示した。また、一定の音素区間を除外する代わりに F_0 の補間をすることで、特に除外割合が大きくなるときに学習効果低減を防ぐことが可能であると確認した。第5章では、学習時に使用するフルコンテキストラベルに基本周波数差分の大小及びその要因を付与することでどのような効果が出るか検討した。これによりラベル付けした場合でも、学習対象を大きく分けることで学習効果が低下することが確認できた。また、子音の立ち上がりや音素区間終端での下降・上昇といった要因は通常の音素区間と同等に扱うことが可能であり、逆に倍ピッチや半ピッチ、局所的にエラーが発生するデータは学習に相応しくないのではないかと考えられる結果が得られた。

6.2 今後の展望

今後の検討事項として、音素区間の除去基準の検討や話者ごとの違いを吸収する必要がある。第4章でも述べたとおり、学習コーパスに著しい問題が見られない場合、音声区間を学習から除外すればするほど HMM 学習量が減少し合成音声の品質が下がってしまう。そのため、適切な閾値を求めて除外対象を絞っていく必要がある。また、今回の実験では各音素区間で F_0 差分の最大値を元に除外したが、この手法では音素ごとに除外される割合が異なり、除外する音素のバランスが取れていないので、音素の種類ごとに F_0 差分が大きい区間を除外することで各音素の学習量を確保する手法も検討したい。

F_0 差分が大きくなる要因をラベル付けしたことで、それぞれの要因ごとに分けて扱うことが可能になる。そこで、意図的に F_0 差分が大きくなるようにラベルを指定した音声を生成したり、どの程度までなら影響が出ないかなど詳細に検討が求められる。また、他の話者でも F_0 差分が大きくなる要因が同様のケースばかりなのかは調査の余地がある。

その他に、 F_0 モデルパラメータの推定ではある程度推定ミスが起きており、それに起因する F_0 差分の変化によって学習サンプルの誤った評価が生じている可能性も考えられる。そこでより精度の高い F_0 モデルパラメータ推定法を用いることができれば、より正確に F_0 差分を扱うことができ学習サンプル選択も正確に行うことができるのではないかと考えられる。

謝辞

まず本研究を進めるにあたり、学士時代から三年間ご指導を下さった広瀬啓吉教授と峯松信明教授、また齋藤大輔助教授に深く感謝を致します。また、研究室での活動を支えて下さった高橋登技官、秘書の池上恵氏と折茂結実子氏にも感謝を致します。

広瀬・峯松研究室の皆様にも感謝致します。特に、研究の指導など様々な点でお世話になった博士課程2年の橋本浩弥氏、研究で貴重な助言を下さった当時博士課程3年の鈴木雅之氏には感謝しております。また、修士課程の苦楽を共にした橋本君、藤垣君、笠原君、毛利さんや、お世話になった他の皆様方にも感謝の言葉を述べさせていただきます。

最後になりましたが、長い学生生活の中で支えてくれた家族や友人にも感謝致します。皆様、ありがとうございました。

2015年2月5日
水上智之

参考文献

- [1] A. Hunt, A. W. Black, “Unit selection in a concatenative speech synthesis system using a large speech database,” Proc. Int. Conf. Acoust. Speech Signal Process., pp.373–376, 1996.
- [2] K. Tokuda, T. Masuko, N. Miyazaki, T. Kobayashi, “Hidden Markov models based on multispace probability distribution for pitch pattern modeling,” Proc. ICASSP, pp.229–232, 1999.
- [3] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, K. Oura, “Speech Synthesis Based on Hidden Markov Models” Proc. IEEE, Vol.101, No.5, pp.1234–1252, 2013.
- [4] A.M. Noll, “Short-time spectrum and “cepstrum” techniques for vocal pitch detection,” J. Acoust. Soc. Am., vol. 36, no. 2, pp.269–302, 1964.
- [5] A. d. Cheveigne, *et al*, “YIN, a fundamental frequency estimator for speech and music” Journal of the Acoustical Society of America, vol. 111, no. 4, pp.1917–1930, 2002.
- [6] M.Asgari, *et al*, “Improving the Accuracy and the Robustness of Harmonic Model for Pitch Estimation” Proc. INTERSPEECH, pp.1936–1940, 2013.
- [7] Taylor. J, *et al*, “Modeling and estimation of the fundamental frequency of speech using a hidden Markov model” Proc. INTERSPEECH, pp.1926–1930, 2013.
- [8] H. Fujisaki, “From information to intonation,” Proceedings of 1993 International Symposium on Spoken Dialogue, pp.7–18, 1993.
- [9] H. Fujisaki, “Modelling the process of fundamental frequency contour generation”, Speech perception, production and linguistic structure, pp. 313–328, 1992.
- [10] T. Toda, K. Tokuda, “Speech Parameter Generation Algorithm Considering Global Variance for HMM-Based Speech Synthesis,” Proc. INTERSPEECH, pp.2801–2804, 2005.
- [11] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, T. Kitamura, “Speaker Interpolation in HMM-Based Speech Synthesis System,” Proc. of EUROSPEECH, vol.5, pp.2523–2526, 1997-9.

- [12] T. Masuko, K. Tokuda, T. Kobayashi, S. Imai, “Voice characteristics conversion for HMM-based speech synthesis system,” Proc. of ICASSP, vol.3, pp.1611–1614, 1997-4.
- [13] J.Yamagishi, K.Onishi, T.Masuko, T.Kobayashi, “A study on adaptation technique for speech synthesis with arbitrary speaker’s voice and various speaking styles,” Proc. the 2003 Spring Meeting of the Acoustical Society of Japan Acoustical Science and Technology, Acoustical Society of Japan, vol.1, no.1-6-25, pp.271-272, 2003.
- [14] K. Hirose, “Speech prosody in spoken language technologies,” Journal of Signal Processing, vol.12, no.1, pp.7–16, 2008.
- [15] H. Fujisaki, K. Hirose, “Analysis of voice fundamental frequency contours for declarative sentences of Japanese,” J. Acoust. Soc. Japan (E), vol.5, no.4, pp.233–242, 1984.
- [16] T. Matsuda, K. Hirose, N. Minematsu, “HMM-based F0 contour synthesis using the generation process model,” Acoustical Science and Technology, Acoustical Society of Japan, 2012.
- [17] H. Hashimoto, K. Hirose, N. Minematsu, “Improved Automatic Extraction of Generation Process Model Commands and Its use for Generating Fundamental Frequency for Contours for Training HMM-based Speech Synthesis,” Proc. INTERSPEECH, 2012.
- [18] 三浦種敏, “新版 聴覚と音声,” 電子通信学会, 1980.
- [19] A. Kurematsu, Kazuya Takeda, Yoshinori Sagisaka, Shigeru Katagiri, Hisao Kuwabara, and Kiyohiro Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis,” Speech Communication, vol.9, pp.357–363, 1990.
- [20] H. Kawahara, I. Masuda-Katsuse, A. de Cheveigne, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,” Speech Communication, vol.27, no.3–4, pp.187–207, 1999.

発表文献

国際会議論文

- [1] T.Mizukami, H.Hashimoto, K.Hirose, D.Saito, and N.Minematsu, “Selection of Training Data for HMM-based Speech Synthesis from Prosodic Features - Use of Generation Process Model of Fundamental Frequency Contours -” Proc. International Conf. on Speech Prosody, 2014

国内全国大会論文

- [2] 水上智之, 橋本浩弥, 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よるHMM 音声合成学習コーパスの評価と選択,” 日本音響学会春季講演論文集, 3-P-32d, 2013-3.

学位論文

- [3] 水上智之, “HMM 音声合成における基本周波数パターン生成過程モデルによる学習サンプル選択,” 東京大学工学部電子情報工学科卒業論文, 2013.

付録 A

ATR コーパスの話者MMIに 付与したラベルデータ

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

ここでは、第5章で用いたラベルデータを掲載する。ラベリングの対象となったのは、ATR コーパスの話者 MMI の中で HMM 学習に用いた A セット~I セットまでの 450 文であり、その中で基本周波数差分が大きい音素区間の上位 10% にラベリングをした。ラベルは第5章の表の上から、1~6 としている。また、音素は triphone 表記である。

文章	音素	ラベル
A01	i-ts+u	3
	e-t+e	3
	t-e+j	2
A02	u-u+k	2
	a-k+a	3
	o-k+u	3
	u-o+sh	2
A03	o-N+d	4
	a-s+o	3
	b-u+sil	1
A04	ry-o+sh	2
	I-t+e	3
	N-o+k	4
A05	o-k+e	3
	a-o+k	2
A06	cl-p+o	3
	cl-p+o	3
	p-o+m	5
	N-p+o	3
	p-o+m	5
A07	sh-u+u	2
	a-j+u	5
	pau-y+o	3
	i-e+k	2
	k-e+b	1
	N-e+k	2
A08	e-N+k	4
	k-a+k	4
A09	ch-o+cl	4
	o-cl+t	5
	u-o+t	4
	U-t+o	3
	i-cl+t	2
	cl-t+a	3
A10	g-a+pau	1
	u-cl+k	5
	n-e+t	4
A11	i-z+e	3
	n-o+k	4
	j-i+f	2
A12	U-t+e	3
	pau-k+e	1
	N-ky+o	3
	n-a+sh	1
	e-o+t	4
	o-s+a	3
	b-a+pau	5
A13	pau-f+u	3
	a-s+u	3
A13	d-a+cl	4
	a-cl+t	5
	pau-g+o	3
	k-u+t	4
	I-k+a	3
A14	o-cl+p	5
	cl-p+a	3
	n-o+p	4
A15	e-e+ky	4
	sil-o+ky	4
	r-e+b	4
	n-i+d	6
	i-d+e	6
	o-d+e	3
	N-j+a	3
	i-k+a	3
	a-cl+t	5
	t-e+k	4
A16	a-f+u	2
	t-a+k	5
	a-k+u	3
	o-h+a	5
	h-a+k	4
A17	sil-z+e	3
	N-sh+a	3
	f-u+h	4
	u-h+e	3
	N-k+a	3
	b-i+pau	4
	pau-k+o	3
	a-o+k	4
	u-N+k	4
	N-k+a	3
A18	a-t+o	3
	b-u+k	2
	k-o+t	4
	o-i+cl	4
	i-cl+k	5
A19	cl-k+e	3
	i-t+e	3
A20	i-i+s	4
	o-f+i	3
A21	a-cl+t	5
	h-o+t	2
	U-t+o	3
A22	i-j+o	3
	I-k+o	5
A23	w-a+t	4
	g-i+pau	1
	u-ts+u	2
	u-s+e	2
	cl-t+a	2
A24	sil-g+a	3
	p-e+k	4
	e-k+I	2
A25	e-N+t	2
	e-f+U	3
A26	k-o+k	4
	h-o+t	4
	o-N+d	1
	o-f+u	3
	n-o+t	4
	t-e+ts	4
	u-k+i	3
	e-pau+m	2
	a-h+a	3
	A27	sil-g+o
e-N+h		4
N-h+a		3
pau-k+e		2
a-h+o		3
w-a+pau		2
o-e+t		2
r-a+cl		1
a-cl+k	5	
A28	U-cl+p	5
	u-e+pau	4
	pau-m+a	3
	m-a+t	4
	t-a+pau	4
	h-a+ch	4
A29	m-o+k	4
	d-e+s	4
	a-w+a	6
	w-a+pau	6
	pau-n+o	6
	n-o+o	6
	cl-s+e	3
	s-e+N	5
	sh-o+o	4
	d-e+pau	4
A29	i-h+o	5
	pau-i+ch	1
	a-e+k	4
	r-a+ny	4

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

A30	i-N+ch	4	
	s-U+sil	2	
	j-a+m	6	
	m-a+t	2	
	a-t+o	3	
	t-o+t	4	
	sh-a+ts	4	
	r-e+t	4	
	e-t+e	5	
	r-o+cl	4	
	o-cl+k	5	
	n-i+pau	4	
	p-e+ch	4	
	a-N+k	4	
i-e+t	4		
A31	sil-m+a	2	
	a-s+o	3	
	o-d+e	2	
	g-u+ch	4	
	u-ch+i	3	
	p-o+k	4	
	g-a+pau	4	
	d-e+h	2	
	U-k+o	3	
A32	ch-a+k	4	
	o-ch+u	2	
	n-i+pau	4	
	a-f+e	2	
	a-t+o	5	
A33	sil-i+N	6	
	i-N+t	4	
	N-t+a	3	
	w-a+pau	4	
	pau-g+o	3	
	o-j+u	6	
	j-i+k	2	
	r-a+h	1	
	t-o+ch	4	
	o-ch+u	3	
A34	a-s+a	3	
	s-a+k	4	
A35	d-e+o	6	
	e-o+s	6	
	d-e+h	6	
	k-i+ts	4	
	z-u+k	2	
A36	m-a+cl	4	
A37	sil-d+a	3	
	a-i+p	4	
	cl-d+o	3	
	d-o+ch	4	
	I-k+a	3	
	a-cl+t	5	
A38	u-u+ky	4	
A39	r-i+t	1	
	i-t+o	1	
	r-a+cl	4	
	a-cl+k	5	
	cl-k+u	3	
	t-o+k	4	
	pau-k+a	3	
	A40	sil-y+u	6
		u-u+z	6
		u-z+a	6
		m-o+s	2
		s-e+k	4
		e-k+i	3
		u-t+o	3
		o-N+t	4
		N-t+o	2
		a-s+e	3
		A41	sil-h+o
w-a+pau			2
k-o+t	4		
e-e+j	4		
U-t+o	3		
i-cl+t	5		
A43	r-a+pau	2	
	i-hy+a	3	
	hy-a+k	4	
	d-o+r	2	
	s-u+s	1	
A44	sil-j+u	3	
	pau-ry+o	6	
	o-j+u	6	
	r-u+k	4	
	k-o+cl	4	
	o-cl+t	5	
A45	sil-w+a	6	
	i-N+t	2	
	N-t+o	5	
	n-i+h	2	
	z-u+pau	4	
	a-N+ch	4	
	e-k+i	4	
	a-i+k	6	
	A46	a-j+u	3
		o-N+d	4
g-a+pau		1	
o-e+k		4	
d-e+pau		4	
n-i+cl		4	
A47	i-cl+p	2	
	n-i+cl	4	
	i-cl+p	5	
	N-t+o	1	
	t-o+sh	4	
A48	I-t+e	1	
	r-i+hy	4	
	w-a+h	4	
	y-o+k	1	
	o-k+u	3	
	o-o+t	4	
	d-a+cl	4	
	U-t+o	3	
	A49	m-o+k	1
		a-t+a	3
ch-i+k		4	
U-s+a		3	
r-e+pau		3	
A50	s-a+k	4	
	my-a+k	4	
B01	i-t+e	1	
	y-a+k	4	
	o-o+k	4	
B02	e-sh+o	3	
	o-o+pau	2	
	g-a+h	4	
B03	o-ts+u	3	
	sil-d+e	3	
	u-ch+i	3	
	o-o+ch	1	
	n-o+ky	2	
	o-ky+o	3	
	ky-o+k	4	
B04	e-N+j	4	
	o-ky+o	3	
	o-t+o	3	
	i-k+u	3	
	I-t+e	3	
B05	d-e+s	2	
	i-sh+o	3	
	o-h+o	3	
	o-o+k	5	
	o-k+a	3	
	o-s+e	3	
	N-k+a	1	
B06	g-a+ts	4	
	r-e+t	4	
	o-s+o	3	
	b-u+r	2	
B07	cl-k+u	2	
	e-N+sh	1	
	N-m+o	3	
	o-N+k	4	
	N-k+a	3	
B08	o-a+ts	4	
	e-cl+t	5	
	o-o+y	5	
	k-a+k	4	
	r-u+h	5	

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

B09	a-i+sil	4	
	n-o+t	2	
	a-s+a	4	
	hy-a+k	1	
	o-o+k	4	
	o-k+l	3	
B10	i-i+p	4	
	r-o+p	5	
	p-e+r	2	
	r-i+f	4	
	a-t+o	4	
B11	o-d+o	3	
	a-k+e	3	
	y-o+k	4	
	a-N+t	2	
	m-a+t	4	
B12	sil-ky+o	3	
	o-ts+u	3	
	u-N+d	2	
B13	sil-b+u	6	
	r-a+t	4	
	t-o+ky	4	
	i-ch+i	2	
	u-k+u	5	
B14	i-t+e	5	
	k-a+s	4	
	a-s+u	3	
B15	a-N+pau	4	
	e-k+i	3	
	n-a+k	4	
	U-t+e	3	
	pau-e+w	6	
	w-a+k	4	
	k-a+k	2	
	k-e+r	4	
	r-u+sil	4	
B16	o-o+t	4	
	o-t+o	3	
	t-e+pau	4	
	r-e+t	4	
B17	sil-g+e	3	
	n-i+ts	4	
	u-i+t	2	
	i-t+e	3	
	u-u+t	1	
	t-o+k	4	
	u-cl+t	5	
	t-e+k	4	
	o-t+a	3	
	B18	r-e+t	2
e-o+ch		4	
B19	a-h+o	3	
	u-ch+u	3	
B20	pau-e+k	1	
	o-h+l	3	
	sil-k+o	3	
	e-e+k	1	
	i-t+e	3	
	t-a+sil	4	
B21	sil-d+o	3	
	pau-a+s	4	
	a-a+f	1	
	e-t+e	4	
	t-e+t	4	
B22	r-u+k	4	
	pau-j+i	3	
	a-cl+t	5	
B23	i-b+o	3	
	cl-ch+a	3	
	o-s+a	3	
	s-a+t	4	
	gy-a+cl	4	
B24	m-e+t	4	
	a-i+cl	4	
	o-k+o	2	
B25	i-ts+u	2	
	a-i+ch	2	
	i-ch+i	3	
B26	pau-ny+u	6	
	n-a+k	4	
B27	sil-i+m	6	
	n-o+pau	2	
	N-f+u	5	
	o-k+u	5	
	pau-ch+o	3	
	e-t+e	3	
	i-n+a	6	
	n-a+t	6	
	a-i+k	4	
B28	k-o+t	1	
B29	I-t+a	3	
	k-a+ts	1	
B30	sil-k+a	5	
	g-a+k	2	
	a-k+u	3	
	b-a+ts	4	
	k-u+pau	1	
	w-a+pau	4	
	g-e+cl	4	
	e-cl+ky	5	
	cl-ky+u	2	
	u-gy+o	6	
	cl-sh+i	1	
	r-u+sil	5	
	B31	U-t+e	3
		o-sh+o	3
		sh-o+k	4
B32	g-u+ch	4	
	i-o+k	4	
	o-i+sil	4	
	sil-r+i	3	
	i-d+o	3	
	l-k+e	3	
	pau-k+e	3	
	e-e+ky	2	
	a-k+e	3	
	e-b+a	6	
b-a+n	6		
B33	g-a+k	4	
	a-k+a	3	
	b-a+pau	4	
	m-o+k	4	
	o-s+u	3	
B34	a-cl+p	5	
	pau-z+o	6	
	z-o+r	6	
	r-o+z	6	
	r-o+t	4	
	o-o+k	4	
	o-k+i	3	
B35	pau-j+i	3	
	a-e+t	4	
	e-t+e	5	
	t-e+pau	4	
	i-d+o	2	
B36	t-a+cl	3	
	r-e+ts	4	
	e-ts+U	2	
	z-u+cl	2	
	u-cl+t	5	
B37	e-N+y	6	
	N-y+a	6	
	a-ky+u	3	
	a-s+a	3	
	a-s+a	3	
	i-cl+ch	5	
	m-u+s	2	
B38	a-N+t	4	
	t-a+sil	5	
	i-r+o	6	
	r-o+N	6	
	o-N+ch	4	
	w-a+pau	5	
	m-o+t	4	
B39	i-cl+p	5	
	r-u+sil	2	
	n-a+r	2	
B40	i-f+o	3	
	o-N+j	4	
	ts-u+i	6	

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

B41	u-i+t	6	
	i-t+o	3	
	e-s+o	3	
	s-o+d	4	
	i-t+a	3	
B42	a-i+k	4	
	n-o+f	4	
	pau-h+a	3	
	a-cl+py	5	
	cl-py+o	3	
B43	sil-j+i	3	
	j-i+k	2	
	i-k+o	3	
	n-o+ch	2	
	o-ch+o	3	
	ch-o+k	4	
	n-a+cl	1	
	pau-a+ts	4	
	e-k+i	3	
	pau-s+o	3	
	i-o+t	4	
	o-sh+I	2	
	sh-I+t	2	
	I-t+a	3	
	a-ch+e	3	
	g-a+s	6	
	a-s+o	2	
	s-o+o	4	
B44	sil-j+o	3	
	gy-a+k	1	
	a-j+o	3	
	m-e+d	2	
	a-cl+t	5	
	o-o+t	5	
	o-t+o	3	
	o-i+k	4	
B45	sil-m+o	6	
	t-o+h	4	
	o-h+i	3	
	o-i+g	4	
	o-t+o	3	
	o-e+m	2	
	o-a+cl	2	
	g-a+pau	4	
	a-o+s	2	
	o-s+a	3	
B46	e-s+o	3	
	o-h+a	2	
	e-N+f	4	
	N-f+u	3	
	B47	r-e+t	6
		e-t+o	3
		n-a+k	5
		pau-j+i	6
		n-i+t	4
		i-sh+a	3
sh-a+ts		4	
a-a+t		4	
k-a+cl		4	
a-cl+k		5	
B48	e-k+i	2	
	i-t+e	4	
	s-e+k	4	
	u-N+g	6	
B49	N-g+a	6	
	g-a+a	6	
	U-sh+a	6	
	sh-a+w	6	
	a-w+a	6	
	w-a+pau	6	
	i-N+j	6	
	N-j+a	6	
	o-j+o	2	
	a-t+o	3	
B50	o-t+o	3	
	y-u+u	6	
	u-u+j	6	
	u-j+o	6	
C01	e-o+ch	2	
	n-o+k	4	
	g-o+pau	4	
C02	a-i+g	1	
	o-o+ky	4	
	o-i+k	4	
C03	U-t+e	3	
	b-u+cl	4	
	w-a+t	4	
	n-o+t	4	
	pau-k+o	3	
	k-o+k	4	
	o-k+u	3	
	e-N+k	4	
	o-h+a	3	
	p-o+d	2	
C04	o-i+t	4	
	i-t+e	5	
	a-N+j	4	
	j-i+cl	1	
	pau-s+a	3	
	m-o+cl	4	
	C05	u-sh+a	3
pau-u+m		6	
C06	o-s+e	3	
	e-ts+u	3	
C07	e-k+a	2	
	u-f+u	2	
	o-o+f	4	
	o-f+u	2	
	pau-j+u	3	
	y-o+cl	4	
	I-t+e	3	
C08	pau-s+o	6	
	s-o+n	6	
	o-n+o	6	
	o-t+o	6	
	n-i+cl	4	
C09	a-ky+u	2	
	pau-i+i	6	
C10	sil-n+o	6	
	r-i+pau	4	
	n-o+f	4	
C11	o-f+U	2	
	U-k+o	3	
	k-e+e	6	
	e-e+j	6	
	e-j+i	6	
C12	N-t+a	5	
	a-ts+u	2	
	j-i+t	4	
C13	t-e+pau	4	
	k-e+t	4	
	o-o+ch	4	
C14	e-t+e	2	
	i-i+p	4	
	u-i+t	4	
C15	t-a+sil	2	
	pau-t+a	5	
	i-o+k	2	
C16	a-i+k	2	
	z-u+k	4	
	u-k+a	2	
	e-t+e	3	
	i-t+a	3	
	r-a+pau	1	
	pau-s+a	6	
	s-a+N	6	
a-N+j	6		
C17	N-j+u	6	
	cl-p+u	3	
	a-k+a	3	
	a-cl+t	5	
C18	sil-i+ch	4	
	a-e+k	2	

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

	e-k+a	3
	I-s+o	6
	s-o+k	6
	o-k+a	6
	n-i+ch	4
	e-t+e	1
C18	I-t+a	3
	u-r+i	3
	r-i+t	5
	i-t+o	3
	pau-j+o	3
C19	sil-h+i	1
	o-i+k	4
	U-t+o	3
	pau-s+a	3
	s-a+cl	4
C20	i-t+a	3
	a-k+u	3
	pau-j+o	3
	pau-sh+a	3
	pau-sh+u	3
C21	u-k+o	5
	k-o+t	4
	n-i+cl	4
	o-N+j	4
	I-k+i	3
C22	o-k+e	3
	o-j+u	2
	o-j+o	3
	o-o+h	4
	r-i+t	4
C23	m-o+a	6
	t-o+k	4
	e-z+u	3
	t-o+t	4
	o-t+a	6
C24	t-a+e	6
	u-t+a	2
	m-u+s	2
	a-f+u	3
	e-t+e	2
C25	sil-i+ts	4
	o-e+t	4
	e-t+a	3
	t-a+k	4
	h-a+cl	4
C26	i-sh+i	3
	pau-w+a	6
	t-e+k	4
C27	o-o+s	2
	N-sh+o	3
	o-sh+a	3
C28	pau-t+a	3
	a-s+e	3
	g-o+ch	4
C29	o-ch+a	3
	z-a+ts	4
	a-ts+u	2
	i-N+ky	4
	n-o+t	4
	m-e+t	4
	e-t+e	5
	pau-e+p	4
	g-u+cl	2
	cl-t+a	6
	t-a+n	6
	d-a+k	4
C30	a-k+e	2
	w-a+t	4
	i-k+a	3
	r-u+t	4
	u-t+o	3
	y-u+u	2
	r-u+i	6
	u-i+k	6
	U-s+e	2
	a-a+r	4
	a-r+u	3
	r-u+sil	5
C31	a-i+ky	4
	i-ky+a	3
	ky-a+k	4
	n-a+cl	2
C32	a-k+o	3
	t-o+ch	1
	U-t+e	3
	t-e+ts	4
C33	a-k+e	3
	pau-s+o	3
	d-a+k	4
	o-o+k	4
	n-o+k	4
C34	pau-i+z	6
	w-a+p	4
	r-i+g	4
	o-N+d	3
	e-o+k	4
C35	sil-k+o	1
	w-a+t	4
	sh-o+o	2
C36	sil-h+o	3
	i-s+a	3
	d-e+pau	4
	h-a+g	6
	a-g+e	6
	k-u+k	4
	t-a+t	4
C37	t-a+k	4
	a-sh+i	3
	U-sh+o	3
C38	o-k+o	1
	u-N+pau	4
	o-N+d	6
	N-d+o	6
	e-t+e	3
C39	y-a+cl	4
	a-cl+t	5
	sil-s+e	3
	e-N+t	1
	N-t+e	1
C40	hy-o+o	4
	o-o+k	1
	m-o+k	4
C41	u-d+e	6
	d-e+a	6
	pau-j+i	3
C42	n-o+s	1
	r-e+t	4
	e-k+u	3
C43	m-u+k	1
	a-N+k	1
	N-k+e	3
	sh-I+t	2
	sil-g+o	3
C44	d-e+k	4
	s-U+k	2
	U-k+a	3
	o-a+t	4
	t-a+t	4
C45	m-u+k	1
	w-a+k	4
	pau-s+a	3
	a-gy+o	6
	o-o+f	6
C46	o-f+U	6
	sil-k+o	3
	h-a+h	1
	r-u+t	4
	t-o+k	4
C47	o-k+e	3
	g-a+h	2
	o-u+k	4
C48	n-o+ts	4
	o-s+a	3
	o-o+t	4
C49	pau-f+U	3
	i-e+h	2
	sil-h+a	3
C39	pau-i+cl	4
	r-u+k	4
C40	w-a+pau	5

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

C50	pau-t+e	3	
	sil-s+e	3	
	n-a+k	1	
	o-o+ky	4	
	o-ky+o	3	
	g-a+f	4	
	a-f+U	3	
	sh-a+b	2	
	o-o+t	2	
D01	sil-j+u	3	
	N-k+a	3	
	n-o+ky	4	
	o-a+t	4	
	a-t+a	3	
	r-e+t	4	
	t-a+k	4	
	g-u+cl	4	
	u-cl+t	2	
o-ky+u	3		
D02	a-i+pau	2	
	n-o+k	4	
	o-e+t	4	
	e-t+e	3	
	i-t+a	5	
D03	i-ch+i	3	
	t-a+h	2	
	i-i+sil	4	
D05	ry-o+k	4	
	t-o+ky	4	
	e-a+r	4	
	r-u+sil	2	
D06	i-cl+p	5	
	o-N+g	6	
	N-g+a	3	
	r-u+pau	4	
	s-e+k	4	
	o-k+a	5	
D07	o-f+U	3	
	U-t+a	3	
D08	I-t+e	3	
	i-ts+u	3	
	n-o+k	4	
	o-k+a	6	
	k-a+t	6	
	a-t+o	3	
	u-f+u	3	
	u-z+u	3	
	z-u+k	4	
D09	r-o+t	4	
	o-t+o	3	
	r-a+b	2	
	e-N+r	4	
	r-o+cl	4	
	o-cl+py	5	
	r-u+sil	4	
	D10	u-ts+u	3
		d-a+k	4
		pau-y+u	6
a-k+a		3	
D11	d-o+r	1	
	pau-f+U	1	
	e-t+o	2	
	y-u+k	2	
D12	y-a+cl	4	
	a-cl+p	5	
	b-a+pau	1	
	k-e+cl	4	
	w-a+ts	4	
	e-k+u	3	
	D13	j-o+o	1
o-a+p		4	
a-t+o		3	
o-o+ky		2	
U-s+a		3	
i-e+t		2	
e-t+a		5	
D14	d-a+ch	4	
	a-ch+i	3	
	m-o+cl	2	
	o-cl+p	5	
	j-u+k	4	
	u-k+e	3	
	w-a+r	1	
	m-u+k	4	
	u-k+e	3	
	D15	sil-t+a	3
i-o+ts		1	
o-ts+U		2	
ch-a+d		4	
D16	sil-g+o	3	
	sh-U+cl	5	
	U-cl+k	5	
	cl-k+i	2	
	g-o+t	1	
	r-i+k	4	
D17	sil-h+a	3	
	pau-j+u	6	
	i-k+a	3	
	i-N+t	4	
D18	U-t+a	3	
	i-t+a	3	
	n-i+j	4	
D19	o-r+e	6	
	r-e+b	6	
	e-b+a	3	
	a-o+k	4	
	u-h+o	1	
D20	o-s+o	3	
	n-o+d	2	
	sil-t+a	1	
	pau-o+o	6	
	h-e+t	4	
	e-sh+o	3	
	sh-o+t	4	
	o-t+e	3	
	pau-t+a	3	
	t-a+ch	4	
	a-ch+i	3	
	a-t+a	3	
	n-a+k	2	
n-a+i	2		
D21	b-i+t	4	
	a-h+i	3	
	y-o+cl	4	
	cl-t+e	3	
	i-k+u	3	
D22	sil-j+u	3	
	r-i+s	2	
	r-u+s	4	
	i-N+sh	4	
	N-sh+a	3	
	sh-a+n	5	
	I-sh+a	6	
	sh-a+w	6	
w-a+k	4		
D23	m-o+k	4	
D24	u-N+d	4	
	a-i+t	2	
D25	I-s+a	2	
	o-o+r	6	
	o-r+i	6	
	i-e+k	4	
	a-e+cl	4	
D26	y-a+k	2	
	o-sh+I	2	
	r-u+sil	5	
D27	sil-i+i	6	
	a-a+t	4	
	a-t+o	3	
	pau-k+i	3	
D28	i-u+k	4	
	a-h+a	3	
	i-o+pau	2	
	m-e+t	4	
D29	g-i+pau	4	
	pau-f+u	1	
	u-j+o	6	
	o-o+r	6	
	o-r+i	6	
	r-i+o	6	

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

	a-i+t	4		N-f+u	2		pau-a+s	2
	i-t+o	3		u-z+a	3		a-k+u	2
	u-k+i	3				D46	w-a+pau	4
	i-k+o	4	D38	o-b+e	3		e-f+U	5
	k-o+k	3		N-ky+o	2		U-k+u	3
	o-k+o	2		b-e+k	4		k-u+t	4
	i-s+u	3		e-k+i	3		u-n+a	6
	u-k+o	5		i-k+a	3		n-a+o	6
	d-e+k	4		w-a+pau	4		o-n+i	6
D30	i-k+a	3		N-t+o	3		n-i+i	6
	a-cl+t	5		t-o+k	4		i-i+s	6
D31	m-a+r	2		e-N+s	4		i-s+o	3
	o-h+i	3		a-r+u	2		s-o+i	5
	a-N+z	4	D39	o-k+i	3		i-s+o	3
	pau-d+o	3		d-o+k	4		o-t+o	3
	n-i+pau	4		pau-s+e	3	D47	g-a+cl	4
	r-o+o	6		i-ch+i	2		r-i+k	4
D32	i-k+o	2		t-a+ch	4		U-t+a	3
	o-g+i	1		a-ch+o	3		g-a+k	4
	j-u+ts	2	D40	t-e+k	4	D48	i-s+a	6
	pau-s+a	2		n-o+h	4		s-a+i	6
	e-o+ts	4		g-a+pau	1		g-a+f	4
	u-i+ky	4	D41	U-k+o	3		a-f+u	3
	i-ky+u	3		U-t+o	3		k-o+t	4
	t-e+pau	5		t-o+sh	4		o-t+o	3
	pau-a+k	4		i-N+g	6		m-o+ts	4
	I-t+e	3		g-a+pau	2	D49	sil-h+a	2
	t-a+sil	2		N-h+o	3		o-t+o	3
D33	i-s+a	3	D42	o-o+g	6		k-a+j	4
	e-t+e	1		o-g+a	6		a-j+i	3
	i-s+o	3		j-i+t	2		N-t+o	3
D34	sil-i+t	4		o-ch+o	3		i-ts+U	2
	u-i+t	2		cl-p+i	2		pau-j+u	3
	i-t+e	5		r-i+k	1	D50	e-i+t	4
	a-f+u	3	D43	e-z+a	3		p-o+cl	4
	e-N+d	2		U-s+a	3		a-r+i	6
D35	y-o+k	4		I-t+a	3		r-i+t	6
	n-o+t	4		u-cl+t	5		i-u+k	2
	o-t+a	3		n-o+h	4		N-d+e	2
	t-a+ch	4		o-h+a	3		sh-I+t	2
D36	I-t+a	1		N-sh+a	6	E01	I-t+a	3
	t-a+ch	5		sh-a+k	6		d-e+pau	4
	ch-i+f	4		i-r+u	2		pau-ch+o	3
	i-f+u	3		d-e+u	6		t-a+b	2
	w-a+pau	1	D44	o-f+U	3	E02	e-g+i	3
	a-k+u	3		m-o+t	4		e-i+k	4
	u-o+sh	2		o-t+o	3	E03	i-cl+p	5
	a-N+h	2		a-k+a	3		o-f+u	3
	pau-t+o	6		w-a+pau	1		b-a+pau	2
	t-o+o	6		m-o+t	4		pau-g+a	3
	k-o+o	6		o-t+o	3		N-s+e	3
	o-o+n	6	D45	sil-g+a	2		o-r+o	3
D37	a-k+u	3		r-u+t	4		t-a+pau	2
	o-N+f	4		pau-i+cl	2		i-i+s	4

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

	a-s+U	3		N-p+a	3											
E04	n-i+ky	4	E16	sil-j+a	3	E25	i-t+o	6	E26	sil-d+a	3					
	i-ky+u	2		u-o+t	4		pau-f+u	3		a-a+ts	4					
	I-t+e	2		t-a+k	1		a-k+a	2		r-u+k	4					
	o-s+o	2		U-s+a	2		o-ts+u	2		t-o+k	4					
E05	a-t+a	3	E17	sil-g+a	3	E26	i-t+a	5	E27	o-t+o	3					
	r-u+sil	4		a-N+sh	4		sil-j+i	6		e-n+o	2					
E06	m-e+pau	4		I-k+i	2		e-N+r	2		o-s+e	3	E28	U-k+e	3		
	j-o+t	4		n-a+r	2		m-a+s	4		m-a+s	4		E29	w-a+t	4	
	t-a+ch	4		o-s+a	3		sil-j+i	6		N-o+pau	2			E30	o-N+pau	4
	i-a+ts	4		s-a+ts	4		e-N+r	2		cl-t+a	3		g-a+pau		4	
	r-e+t	4	w-a+ch	4	E18	sil-g+o	6	o-o+i	6	i-n+i	6					
E07	j-i+k	4	sil-g+o	6	E19	g-o+g	1	i-n+i	6	n-i+a	6					
	k-o+g	4	g-o+g	1		w-a+m	2	i-a+r	6	i-a+r	6					
	n-o+ky	4	w-a+pau	4		m-a+t	4	r-o+o	2	sil-z+u	3					
	a-i+t	2	pau-t+a	5		i-N+p	4	u-g+i	3	u-g+i	3					
	i-t+a	3	pau-m+e	6		n-o+ry	4	o-cl+t	5	o-cl+t	5					
	t-a+i	2	e-o+t	4		o-k+o	3	k-o+t	2	e-t+a	5					
E08	n-i+cl	2	t-e+k	4	E20	o-k+e	3	sil-s+e	1	E31	U-ts+u	3				
	o-N+j	4	s-u+sil	4		k-e+cl	4	sil-s+o	3		E32	sil-s+o	3			
	g-o+k	4	E19	sil-b+o		6	e-cl+t	5	o-o+t	4		o-o+t	4			
E09	ch-u+u	4	w-a+m	2		u-k+o	5	i-t+a	3	i-t+a		3	E33	t-a+k	4	
	g-o+k	1	m-a+t	4		k-o+t	2	t-a+k	4	t-a+k		4		E34	k-o+t	4
	i-sh+i	3	i-N+p	4		o-t+o	3	k-o+t	4	k-o+t		4			pau-m+o	6
	d-o+cl	4	E20	n-o+ry		4	a-cl+t	5	m-o+cl	6		m-o+cl	6		o-cl+t	5
	o-j+i	1	E21	o-k+o		3	t-a+sil	4	a-k+e	3	a-k+e	3	n-o+by	2		
	i-t+e	5		pau-h+a		3	o-t+o	3	n-o+by	2	n-o+by	2	a-cl+t	5		
E10	r-e+h	4		h-a+a		2	a-cl+t	5	a-cl+t	5	a-cl+t	5	cl-t+a	2		
	e-h+o	2		a-a+h	2	t-a+sil	4	cl-t+a	2	cl-t+a	2	o-ts+u	3			
	o-o+k	1		s-e+t	2	E22	t-o+k	4	o-ts+u	3	o-ts+u	3	w-a+ky	4		
	o-k+a	3		k-o+ky	1	r-e+t	4	r-e+t	4	w-a+ky	4	s-a+t	4			
	g-e+t	4		t-o+k	4	i-a+t	4	i-a+t	4	s-a+t	4	t-o+pau	4			
	r-a+sh	4		r-e+t	4	g-a+k	4	g-a+k	4	t-o+pau	4	o-ts+u	3			
E11	U-ky+a	3		i-a+t	4	a-k+u	3	a-k+u	3	E35	o-ts+u	3	o-ts+u	3		
	a-N+d	4		g-a+k	4	u-cl+k	5	u-cl+k	5	E36	o-ts+u	3	o-ts+u	3		
E12	m-a+s	4	a-k+u	3	i-ts+u	2	i-ts+u	2	E35		o-ts+u	3	o-ts+u	3		
	s-a+n	6	u-cl+k	5	E23	sil-k+o	3	sil-k+o			3	E35	o-ts+u	3	o-ts+u	3
	n-i+k	4	i-ts+u	2	g-o+h	4	g-o+h	4		E35	o-ts+u		3	o-ts+u	3	
	i-k+a	2	E23	sil-k+o	3	n-i+pau	4	n-i+pau	4		E35		o-ts+u	3	o-ts+u	3
	e-j+i	2	sil-k+o	3	a-t+a	5	a-t+a	5	E35			o-ts+u	3	o-ts+u	3	
	u-N+py	4	g-o+h	4	o-s+o	3	o-s+o	3		E35		o-ts+u	3	o-ts+u	3	
	N-py+o	3	n-i+pau	4	E24	r-a+pau	4	r-a+pau			4	E35	o-ts+u	3	o-ts+u	3
E13	o-t+o	6	a-t+a	5					E35		o-ts+u		3	o-ts+u	3	
	t-o+n	6	o-s+o	3						E35	o-ts+u		3	o-ts+u	3	
	e-N+s	4	E24	r-a+pau	4						E35	o-ts+u	3	o-ts+u	3	
	E14	I-ch+i	6						E35			o-ts+u	3	o-ts+u	3	
		ch-i+n	6							E35		o-ts+u	3	o-ts+u	3	
		i-n+o	6								E35	o-ts+u	3	o-ts+u	3	
		i-k+u	2						E35			o-ts+u	3	o-ts+u	3	
d-a+k		4						E35		o-ts+u		3	o-ts+u	3		
a-k+e		3								E35	o-ts+u	3	o-ts+u	3		
d-a+cl		4							E35		o-ts+u	3	o-ts+u	3		
E15	e-k+a	2						E35			o-ts+u	3	o-ts+u	3		

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

	u-ts+u	3
	U-k+e	2
	i-N+t	4
	a-i+t	4
	o-d+o	3
	o-ry+o	3
	o-k+u	3
	k-u+g	5
	u-g+a	6
	g-a+h	6
E37	t-o+pau	2
	sh-i+o	6
	i-o+k	6
	o-k+e	3
	n-a+ts	4
	a-s+u	3
	o-u+k	4
	k-e+pau	5
	u-cl+t	5
	cl-t+o	2
	cl-t+a	3
E38	o-o+ky	2
	i-o+pau	4
	m-a+ch	2
	I-t+e	3
E39	y-u+k	1
E40	i-p+i	3
	n-o+ch	4
	o-ch+i	3
	g-a+cl	4
	a-cl+k	5
	o-sh+o	3
	o-t+o	3
	o-o+k	4
	a-cl+t	5
E41	sil-u+m	6
	u-m+i	6
	n-o+pau	4
	o-f+U	2
	a-s+u	6
	s-u+m	6
	u-m+i	6
	k-e+t	4
	a-s+a	3
E42	I-k+a	2
	u-h+i	2
	n-i+k	4
	k-a+k	4
	e-t+a	5
E43	i-i+t	4
	w-a+t	4
	e-e+k	4
	e-k+i	2
E44	sil-h+a	3

	t-e+pau	2
	t-a+sil	2
E45	sil-i+m	6
	o-k+o	3
	sh-a+k	1
	t-a+t	4
	a-t+e	3
	t-a+k	1
	d-a+sil	1
E46	k-u+s	4
E47	o-N+f	1
	N-f+u	3
	u-e+t	4
	e-t+e	3
	I-cl+t	5
	cl-t+o	3
E48	i-s+a	3
	a-i+k	4
	o-N+j	6
	N-j+u	3
	r-o+k	4
	I-s+a	3
	s-a+pau	4
	pau-k+a	3
	b-u+j	6
	u-j+i	6
	i-ch+o	6
	ch-o+o	6
	m-a+ch	4
	o-o+k	4
E50	a-s+e	2
	N-k+a	3
	i-sh+i	1
	sh-i+pau	1
	pau-j+u	3
	o-k+i	6
	k-i+n	6
	r-i+t	4
	i-t+o	3
F01	i-N+d	4
	a-f+i	3
	n-i+ky	4
	w-a+k	1
	a-ch+i	3
	e-t+e	5
	cl-t+e	5
F02	a-i+cl	4
	o-cl+t	2
	cl-t+e	3
	t-o+gy	4
	n-o+t	1
F03	g-a+t	4
	z-o+k	4

	t-o+t	4
F04	r-e+k	4
	k-i+sh	2
	u-N+j	4
	o-m+u	3
	u-k+u	3
	n-a+ch	4
	o-h+a	3
	m-e+r	2
F05	i-N+sh	4
	i-N+sh	4
	g-o+t	4
	n-o+t	2
	o-t+a	3
	g-o+sh	4
	s-e+ts	3
	e-ts+u	4
	o-s+e	3
	s-e+ts	4
F06	gy-a+k	4
	a-k+u	2
	o-o+k	4
	I-t+e	3
	m-o+t	4
F07	o-k+o	3
	i-sh+o	3
	o-o+k	2
	e-t+e	6
	t-e+y	6
F08	m-e+k	2
	a-cl+t	4
	cl-t+a	3
	n-a+k	4
	i-e+t	4
	t-e+pau	4
	o-a+cl	2
F09	n-o+ky	4
	o-ky+u	3
	u-k+U	2
	g-i+t	2
F10	y-u+k	5
	y-a+k	2
	a-k+u	3
	y-u+k	4
	a-f+u	3
	w-a+t	2
	a-t+o	3
	o-o+ch	4
	o-ch+I	5
	I-t+e	3
F11	sil-d+a	3
	d-e+p	4
	e-p+a	2
	t-a+ch	4

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

F12	sil-a+i	2
	a-t+o	2
	sh-i+pau	4
	m-o+ky	4
	n-o+f	4
	b-a+ts	4
F13	n-o+h	4
	u-u+s	2
	cl-k+u	3
	k-u+ts	4
	U-s+a	3
	r-e+t	4
	e-t+a	5
F14	o-ch+o	6
	g-a+k	4
	sh-I+t	2
	g-o+pau	4
	pau-y+o	3
	m-a+r	2
	d-a+cl	4
	a-cl+t	5
F15	w-a+pau	1
	o-f+U	1
	d-a+k	4
	d-a+sil	1
F16	y-a+h	4
	a-h+o	3
	k-a+cl	4
	a-cl+t	5
F17	m-a+cl	6
	a-cl+p	5
	r-u+m	6
	u-m+a	6
	e-N+ch	4
	n-i+pau	4
	m-o+cl	2
	cl-t+e	3
F18	i-u+ch	4
	u-ch+u	3
	t-o+s	1
	n-o+t	4
F19	a-s+a	3
	a-i+k	4
	pau-d+e	3
	d-e+pau	4
	i-o+sh	4
	m-e+t	4
F20	e-e+k	4
	e-k+a	2
	r-e+t	4
F21	pau-a+k	1
	o-o+s	4
	I-ch+i	3
F22	o-k+o	3
	sh-i+pau	1
	h-e+y	5
	a-i+cl	4
	i-cl+t	1
F23	e-ts+U	2
	a-cl+t	5
F24	sil-w+a	6
	u-k+a	3
	g-a+pau	2
	pau-z+u	6
	u-i+ch	4
F25	a-k+i	3
	o-f+u	2
	i-N+t	4
	a-cl+t	5
F26	g-e+ts	4
	a-i+h	4
	i-h+e	2
F27	sil-y+o	6
	j-i+t	4
	a-s+a	3
	z-a+k	4
	a-cl+t	5
F28	cl-t+e	3
	u-N+h	4
	N-h+o	3
	n-i+k	4
	e-N+ry	4
F29	k-u+s	2
	o-a+k	4
	a-k+e	3
	r-u+t	4
	u-t+o	3
	t-o+pau	1
	I-s+o	3
e-cl+p	5	
F30	e-a+cl	1
	a-cl+t	5
F31	y-a+ky	4
F32	o-N+pau	4
	pau-s+o	1
	e-o+h	4
	o-h+a	3
F33	u-i+t	4
	n-i+i	2
F34	sil-d+o	3
	e-i+cl	4
	i-cl+t	5
	r-e+t	4
F35	sil-i+m	6
	a-i+i	4
	pau-s+e	3
F36	a-i+k	4
	i-k+o	3
F37	w-a+s	4
F38	m-o+pau	2
	a-k+e	3
	o-m+a	2
F39	N-t+o	3
	e-i+k	4
	pau-g+o	3
	o-p+a	3
	o-k+a	3
F40	d-e+ch	4
	o-o+k	4
	a-d+o	3
	o-N+t	4
	sil-k+o	3
	e-k+U	2
	u-s+a	3
F41	u-t+o	3
	t-o+pau	4
	I-t+e	6
	t-e+o	6
	e-o+r	6
	m-a+s	4
	a-s+U	2
	o-z+o	6
F42	o-N+j	6
	N-j+i	6
	n-o+t	4
	n-i+k	4
	w-a+pau	1
	e-t+e	3
	k-o+t	4
F43	w-a+k	4
	r-e+t	4
	o-o+s	2
	o-g+a	6
	g-a+a	6
	cl-k+u	3
F44	o-sh+i	6
	sh-i+d	6
	i-d+e	6
	N-k+i	3
	sil-i+cl	4
	i-cl+p	5
	i-cl+p	5
	o-N+t	2
t-o+b	2	
k-o+k	4	
n-o+pau	4	
k-i+N	6	
i-N+y	6	
I-k+a	3	

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

	k-a+N	5
	I-t+i	3
	t-i+i	5
	i-i+sh	4
F45	k-o+o	6
	o-o+k	1
	o-k+u	1
	n-i+ch	2
	w-a+pau	4
	r-u+sil	4
F46	u-i+f	6
	i-f+U	3
	U-t+o	3
	w-a+t	4
	i-k+a	1
	pau-t+e	6
	e-o+f	6
F47	sil-o+b	2
	a-k+e	6
	k-e+w	6
	e-w+a	3
	e-a+r	4
F48	u-h+a	2
	pau-a+ch	6
	I-k+o	6
	k-o+ch	6
	n-i+ts	1
	u-ch+i	3
	t-a+h	1
	h-o+k	4
	N-k+a	3
F49	w-a+t	4
	pau-i+ch	4
	o-s+e	3
	i-N+sh	4
	t-a+sil	4
F50	n-o+h	4
	u-h+o	5
G01	n-o+k	4
	a-cl+k	5
	cl-k+o	2
	u-k+e	3
	k-e+t	4
	e-t+e	3
	a-u+ts	4
G02	w-a+t	4
	g-a+f	4
	a-f+U	2
	U-t+o	3
	t-o+ts	4
	pau-ch+I	3
	I-ch+i	6
	ch-i+w	6
	a-o+o	6

	k-u+k	4
	U-cl+t	5
G03	sil-o+k	4
	o-k+a	2
	b-o+t	4
G04	a-N+r	6
	N-r+o	6
	r-o+N	6
	pau-e+N	6
	n-o+k	2
	a-sh+i	2
	i-o+m	1
G05	t-e+ch	4
	g-a+k	1
	o-k+u	3
	n-o+k	4
	k-o+k	4
	r-o+o	2
	g-o+k	4
G06	b-a+pau	2
	pau-a+k	4
	k-u+h	4
G07	cl-sh+a	3
	n-o+ch	4
G08	a-h+o	2
	a-d+e	2
G09	i-d+e	2
	u-t+o	3
	pau-z+u	6
	a-k+e	3
	s-a+k	4
G10	sil-t+o	5
G11	d-o+k	4
G12	e-N+sh	2
	cl-k+i	3
	k-i+n	5
	b-u+k	1
	k-o+t	4
G13	o-s+u	3
	e-N+k	4
	g-a+ch	4
	n-e+t	4
G14	m-o+k	4
	k-a+cl	2
	a-cl+t	5
	z-o+t	1
	o-j+o	2
	i-cl+t	5
G15	i-i+ch	4
	z-u+k	4
	pau-by+o	3
	u-k+o	3
	u-t+o	3

G16	pau-s+o	3
G17	r-u+f	1
	y-o+ky	4
	y-a+k	4
G18	n-i+cl	4
	w-a+k	4
	a-k+e	3
	k-e+cl	4
	pau-y+u	3
	o-a+t	4
G19	sil-o+y	6
	a-N+k	4
	d-a+k	4
	m-o+pau	2
	e-k+o	3
	a-s+e	3
	t-o+k	4
	d-a+sil	2
G20	a-sh+a	3
	u-o+sh	3
	m-e+t	2
	e-t+e	2
	m-a+cl	4
	t-a+sil	1
G21	d-e+k	4
	o-sh+i	3
	a-e+t	4
G22	pau-h+o	3
	m-o+o	6
	o-o+h	6
	m-a+s	4
G23	z-a+f	4
	a-f+u	3
G24	a-o+p	4
	o-p+o	2
	o-p+o	3
	I-t+e	3
	o-o+k	4
	u-s+a	3
	r-e+t	4
G25	e-cl+k	1
	m-u+k	4
	t-a+k	4
	a-cl+t	1
G26	r-e+t	4
	N-p+e	2
	i-s+u	3
	t-e+k	1
G27	g-a+ts	1
	u-t+o	1
	a-k+a	3
G28	r-a+p	4
	a-p+u	2
	p-u+cl	4

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

G29	a-m+e	2	G37	sil-y+u	6	G44	j-i+o	1		
	w-a+b	4		j-o+o	4		i-o+h	1		
	a-b+e	3		y-a+r	4		o-s+a	3		
	cl-d+o	3		a-r+i	3		N-sh+o	3		
	d-o+n	5		ry-o+k	4		o-j+u	2		
	N-t+o	3		n-o+k	4		e-N+k	4		
	I-r+a	3		k-o+k	1		m-i+t	4		
	o-s+o	3		o-o+r	6		i-t+o	3		
	j-i+cl	1		o-r+a	6		G45	u-ch+I	2	
	a-ts+u	3		r-a+i	6			e-e+t	4	
	G30	o-sh+a		3	sh-a+k			4	e-t+o	3
		pau-h+o		3	o-i+e			4	o-s+a	3
		h-o+n		6	i-e+e			5	g-a+pau	2
o-n+e		6	G38	r-a+N	6	o-g+a	2			
e-m+a		6		a-N+d	6	G46	cl-ky+o	3		
m-a+d		6		o-s+e	6		ky-o+o	5		
a-d+e		3		s-e+r	6		t-o+cl	4		
d-e+k		4		e-N+k	4		u-ch+u	2		
I-s+a		3		r-i+k	4		n-o+ch	4		
cl-t+e		3		i-k+o	3		i-N+pau	4		
pau-y+a		6		s-o+t	4		i-N+s	4		
G31	sh-o+cl	4		d-e+t	4		e-o+k	4		
	G32	pau-m+a		6	k-u+sil	2	b-e+k	4		
u-o+k		2		G39	sil-g+e	3	G47	i-j+o	3	
e-h+a		2	e-N+z		4	G48		o-k+u	3	
cl-t+a		3	N-z+a		3			y-o+cl	4	
G33	sil-n+i	6	r-u+k		4	o-cl+t		5		
	u-s+a	3	b-a+ts	4	a-i+s	4				
	U-k+o	1	G40	o-N+g	4	e-N+t		4		
	u-u+k	2		u-s+o	3	N-t+a		3		
	u-k+o	3		n-o+ch	4	a-k+a		3		
	r-u+k	4		n-a+h	4	u-a+k	4			
G34	sil-j+u	6	G41	sil-y+a	6	G49	sil-j+i	6		
	cl-s+a	3		o-a+k	4		U-t+o	3		
	a-i+t	4		o-j+i	3		t-o+s	2		
	u-e+n	2		u-f+u	3		o-a+t	4		
	d-e+pau	4		a-d+o	3		a-i+t	4		
	cl-s+a	3		d-o+N	6		i-h+I	2		
	n-a+s	4		o-N+ky	6		I-t+o	3		
	a-cl+t	5		G42	sil-b+a		3	G50	i-j+o	6
G35	u-m+a	3	o-s+a		3	w-a+k	4			
	m-a+d	5	w-a+pau		2	d-a+k	4			
	m-o+t	4	o-o+t		6	a-k+a	3			
	a-i+k	4	o-t+e		6	e-a+r	4			
	i-k+a	3	i-r+u		2	H01	g-a+k	1		
	a-cl+t	5	pau-m+e		6		r-u+t	4		
G36	a-r+e	6	w-a+ts		1		u-t+o	3		
	r-e+g	6	i-t+o	3	d-a+cl		4			
	g-a+k	4	m-a+cl	4	H02	i-cl+p	5			
	n-o+sh	4	G43	g-a+k		4	N-i+ch	4		
	o-sh+U	2		i-N+ch	4	i-b+a	3			
	u-sh+o	3		d-e+k	2	H03	o-sh+a	3		
	i-k+a	3		k-u+j	6		H04	gy-a+k	4	
				u-j+i	6	N-n+i		6		

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

H05	n-i+ch	6	
	m-o+k	4	
	n-a+k	1	
	i-e+tt	4	
	o-s+a	3	
	r-e+tt	4	
	e-t+e	1	
H06	I-t+a	6	
	t-a+ch	6	
	n-o+sh	4	
	u-k+o	3	
	e-sh+o	3	
	o-cl+tt	5	
	e-i+r	4	
H07	sil-n+a	6	
	e-k+a	3	
	k-a+pau	4	
	o-sh+o	3	
	b-e+tt	4	
H08	sil-a+r	6	
	sh-a+ch	2	
	a-ch+o	3	
	e-t+o	3	
H09	I-t+e	3	
	t-e+i	5	
	i-k+u	3	
	k-u+s	2	
H10	N-o+k	4	
	o-k+o	2	
	i-N+tt	4	
H11	sil-k+o	2	
	k-o+k	4	
	a-N+h	4	
	i-e+tt	4	
	e-t+e	3	
	u-sh+o	3	
	N-sh+I	4	
	H12	sil-k+e	3
k-e+N		6	
e-N+s		6	
U-sh+o		3	
y-a+k		4	
a-k+o		3	
pau-t+o		3	
r-o+g		6	
o-g+a		6	
a-sh+i		3	
U-s+e		1	
u-t+a		3	
n-o+k		4	
o-o+tt		4	
o-t+o		3	
H13		t-o+sh	5
		i-s+a	3
	s-a+k	4	
	o-k+i	2	
	n-o+hy	2	
	i-t+o	6	
	t-o+y	6	
H14	sil-y+u	6	
	o-a+tt	4	
H15	cl-k+i	3	
	u-N+tt	4	
	t-a+k	4	
	a-k+a	3	
	k-a+cl	4	
	t-e+pau	4	
	o-n+i	6	
	cl-t+a	3	
	H16	n-o+ch	4
k-i+s		1	
n-i+pau		2	
H17	cl-f+u	3	
	a-k+a	5	
H18	sil-j+u	3	
	n-o+k	4	
	a-i+s	2	
	u-sh+I	2	
	u-h+i	2	
	g-o+sh	2	
H19	sil-i+ch	4	
	i-n+e	2	
	j-u+u	4	
	e-t+a	3	
	a-k+o	2	
	k-o+k	4	
	i-f+U	5	
	cl-ch+a	3	
	H20	t-o+s	4
r-o+cl		1	
o-cl+p		1	
a-n+i		6	
n-i+i		6	
i-i+cl		4	
H21		sil-y+o	6
	a-s+a	3	
	w-a+tt	4	
	H22	r-a+g	2
r-a+tt		4	
o-h+i		2	
u-o+tt		4	
n-o+h		1	
r-e+k		4	
w-a+cl		4	
H23		sil-h+I	2
		i-t+o	3
H24	o-sh+o	3	
	o-o+ky	2	
	ky-a+k	4	
	a-k+U	3	
	t-o+cl	4	
	w-a+pau	2	
H25	pau-h+o	2	
	h-o+h	4	
	u-f+U	3	
H26	a-s+u	2	
	r-u+tt	4	
	t-o+ts	4	
	o-ts+u	1	
	m-u+cl	4	
	cl-t+e	3	
H27	o-o+r	2	
	n-a+k	2	
	e-t+e	3	
	e-o+cl	4	
H28	r-u+sil	5	
H29	i-p+u	3	
H30	n-i+cl	6	
	o-N+e	6	
	o-cl+tt	1	
	cl-t+e	2	
	t-e+k	4	
	e-k+a	3	
	o-r+e	2	
	H31	g-a+f	4
cl-t+e		3	
u-i+k		4	
g-a+cl		2	
u-k+o		3	
k-o+tt		4	
H32	o-N+k	6	
	N-k+a	2	
	j-i+cl	2	
	I-k+a	6	
	a-o+ts	1	
	o-ts+U	2	
	n-a+k	2	
H33	o-b+e	2	
	o-o+tt	4	
	o-k+a	2	
	m-o+cl	4	
	o-cl+p	5	
H34	y-a+ky	4	
	sil-d+o	3	
	m-o+d	2	
	pau-ky+o	3	
	ky-o+k	4	
H35	e-a+cl	1	
	i-k+o	2	

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

H36	pau-j+i	3
	k-e+cl	4
	i-s+a	3
	y-a+ky	4
H37	i-s+a	3
	o-ky+o	3
	r-o+t	4
	o-t+o	3
	w-a+k	4
H38	a-k+o	3
	n-a+cl	4
	sil-t+a	3
H40	y-a+k	4
	pau-a+k	1
H41	a-o+k	4
	i-sh+u	2
	pau-g+u	6
	N-t+a	3
	n-o+k	4
	o-k+a	3
H42	u-d+a	2
	t-o+pau	5
	pau-ky+u	1
	k-a+k	4
	g-a+k	4
	a-k+u	3
	k-u+cl	2
	u-cl+k	2
	cl-k+i	3
	i-t+o	6
	t-o+e	6
	g-a+k	4
H43	a-N+sh	4
	N-sh+i	3
	N-sh+I	2
	t-e+pau	4
	e-o+k	4
	m-o+d	2
	d-e+k	2
	e-k+i	3
	r-e+b	2
	pau-y+a	6
	y-a+j	2
	m-o+t	4
	t-o+b	4
	s-e+r	4
	H44	sil-h+e
h-e+e		6
a-i+sh		6
i-sh+I		6
n-a+k		4
a-g+e		6
H45	i-k+a	3
	e-N+b	6

H47	a-s+u	3
	g-o+sh	2
	o-o+k	4
	z-u+k	4
	d-a+k	4
	a-k+e	3
	n-o+k	4
	s-o+k	4
	r-e+t	4
	H48	b-a+pau
e-N+k		1
m-o+k		4
o-k+o		3
o-k+o		3
o-sh+I		3
d-e+pau		4
e-e+k		2
a-ts+u		3
u-o+ts		4
H49	U-t+o	3
	t-e+pau	4
	a-t+o	3
H50	u-u+ch	4
	u-ch+u	2
	a-y+u	6
	U-cl+t	5
	e-i+t	4
	t-a+k	4
	d-o+pau	2
	i-t+a	3
	a-j+i	3
	u-t+o	1
I01	e-N+h	4
	N-h+e	3
	d-a+cl	4
	o-s+a	3
I02	d-a+cl	4
	i-t+o	3
I03	u-r+e	2
	e-i+r	4
	k-o+k	1
	t-o+k	4
	o-k+o	3
	o-k+o	3
	y-u+t	4
	a-k+a	3
	i-ch+i	2
	I04	t-e+ch
g-u+cl		4
u-cl+t		5
I05	b-u+ts	4
	n-o+k	4
	b-o+t	4

I06	u-g+a	3
	cl-t+a	1
	b-i+t	1
	t-a+k	4
I07	m-a+cl	2
	pau-h+o	3
	h-o+k	4
I08	o-o+k	4
	sil-a+k	4
	N-s+e	3
	e-N+t	2
	N-t+o	3
	U-k+i	6
	k-i+a	6
	w-a+pau	1
	g-a+k	4
	a-k+o	3
I09	a-cl+t	5
	cl-t+e	3
	u-u+k	4
I10	o-s+o	2
	i-N+r	4
I11	r-e+k	4
	i-t+e	2
I12	i-sh+i	3
	sh-i+h	4
	h-a+k	4
I13	d-a+k	4
	a-k+e	2
	I-k+a	3
I14	o-o+t	4
	o-s+e	3
	r-u+sh	6
	z-e+k	4
	n-a+p	1
	cl-ch+a	3
	a-s+e	2
I16	o-o+k	2
	o-k+a	1
	u-k+a	3
	r-o+t	4
I17	r-e+t	4
	e-t+a	5
	a-k+e	3
I18	I-k+u	3
	a-cl+t	5
	sil-i+m	6
I19	r-a+N	2
	r-u+t	4
	o-h+a	2
I20	sil-k+o	3
	e-e+k	4
I21	o-N+t	4
	N-t+e	3

付録 A ATR コーパスの話者 MMI に付与したラベルデータ

	o-o+ts	4
	r-e+k	4
I22	w-a+t	4
I23	sil-k+e	3
	I-t+a	3
	t-a+sh	1
	a-sh+o	3
	u-f+u	3
I24	o-n+o	2
	pau-k+e	3
	ky-o+k	4
	U-k+u	3
I25	a-i+t	4
	m-e+ts	4
I26	k-a+t	4
	n-i+ch	4
	b-e+k	2
I27	sh-o+k	1
	u-o+t	4
	o-t+o	2
	e-m+o	6
	m-o+sh	6
	o-sh+a	3
I28	cl-p+o	3
	o-u+t	4
	u-t+a	3
	a-ky+a	3
	ky-a+sh	1
I29	w-a+k	4
	a-s+o	3
	m-a+cl	4
	a-cl+t	5
	t-a+sil	1
I30	pau-d+o	6
	o-f+U	3
	o-o+i	5
I31	g-a+pau	4
I32	n-a+ts	2
	a-s+u	6
	s-u+m	6
	u-m+i	3
	i-d+a	3
	d-a+t	4
	N-b+a	3
I33	sil-s+a	2
	pau-w+a	3
I34	sil-g+o	3
	o-f+u	2
	d-o+s	4
	pau-b+e	3
	e-a+cl	4
	t-a+k	4
	I35	m-o+cl
u-k+a		2
n-o+h		1
I36	k-a+k	1
	e-a+r	2
I37	i-o+t	4
	t-a+t	4
	u-k+o	2
I38	n-a+k	4
	e-e+t	4
	U-k+a	2
	r-u+t	4
	t-o+k	4
	I39	i-z+u
u-ts+u		2
ts-u+t		4
u-t+e		3
e-o+h		4
o-h+i		2
I40	k-a+r	2
	i-N+ch	4
	u-k+i	3
	r-u+k	4
	pau-d+o	3
	I41	o-ts+u
a-y+o		6
y-o+r		6
i-k+e		6
k-e+r		6
sh-o+k		4
i-t+o		3
I42	ky-a+k	2
	o-f+a	3
	o-t+e	2
	i-k+a	3
	i-k+i	3
I43	sil-o+o	6
	o-b+e	3
	pau-sh+i	2
	g-a+cl	4
I44	a-i+i	1
I45	sil-h+e	3
	U-sh+u	3
	n-a+k	4
	o-ts+u	3
I46	i-sh+o	3
	r-u+h	3
	u-h+o	4
	h-o+d	3
	o-d+o	4
	z-u+k	1
	u-t+o	6
	t-o+i	6
I47	r-o+k	4
	i-s+a	3
	t-e+k	4
	e-t+a	1
I48	sil-my+o	6
	s-o+r	2
	o-i+t	4
	r-u+r	2
	pau-j+o	3
	r-i+k	4
	e-e+d	2
	e-a+ch	4
I49	h-a+j	6
	i-m+e	6
	m-e+t	6
	e-t+e	3
	t-e+ch	4
	i-t+a	2
I50	sil-z+a	3
	o-o+cl	4
	r-u+k	4
	k-o+t	4
	o-s+o	3