

表現豊かな音声合成のための テキストからの句末音調ラベルの予測

中嶋 秀治^{†1} 水野 秀之^{†1}
吉岡 理^{†1} 高橋 敏^{†1}

表現豊かな音声において多様性を示す句末音調ラベルをテキストから予測する方法について述べる。本方法では、これまでの読み上げ口調の音声合成の言語解析の出力結果である単語の情報と、アクセント句およびイントネーション句の境界情報を用いる。そして、表現豊かな音声が発せられる場面、および、話者に依存したモデルを構築する。商品宣伝、電話対応の各場面のデータを用いて、句末音調ラベル予測評価実験を行なったところ、数個の特徴量を用いる提案法が、多量の特徴量に基づく従来法に比べて同等以上の一致率 (*Cohen's kappa*) を得ることを確認した。

Predicting Phrase Boundary Tone Labels from Text for Expressive Text-to-Speech Synthesis

HIDEHARU NAKAJIMA,^{†1} HIDEYUKI MIZUNO,^{†1}
OSAMU YOSHIOKA^{†1} and SATOSHI TAKAHASHI^{†1}

This paper describes the statistical method for predicting, from the input text, phrase boundary tone labels which shows more varieties in expressive speech than in reading speech. This method uses both lexical information and phrase boundary locations obtained by the language analyzer in conventional reading style text-to-speech synthesizers. Accordingly, obtained models reflect both domain and speaker specific speaking styles. Through the experiments, we confirmed that our model uses simple and a few features to achieve the same or higher correlation coefficients with correct answer labels (*Cohen's kappa*) than the conventional classification tree based method.

1. はじめに

ニュースや情報案内文を対象とした読み上げ口調の音声合成は自然性が改善され、実際に利用されるようになった。そして、音声合成の適用領域の拡大が求められ、e-Commerce や電話自動対応やエンターテインメントのような人に向かって語りかけるさまざまな場面での自然な口調を備えた音声の合成を目的とした研究が行なわれている¹⁾。本稿では、このような人に向かって語りかける場面での使用を念頭に置いた口調の音声を「表現豊かな音声 (expressive speech)」と呼び、議論を進める。

表現豊かな音声における特徴の一つとして、アクセント句の末尾での基本周波数 (fundamental frequency, F_0) の動きの多様性がある。従来の読み上げ口調の音声合成対象は情報案内目的の平叙文が多く、アクセント句ごとの F_0 の変化形状は、日本語の東京方言の読み上げ口調では、概ね「へ」の字型になる場合が多かった。しかし、表現豊かな音声での F_0 は、アクセント核の位置で降下した後も句末まで下がったままか、句末付近で再上昇するかといった多様な変化を示し、このような「句末音調」の分析的な研究が行なわれてきた^{2),3)}。

このような F_0 の変化形状は Tone and Break Indices (ToBI)⁴⁾ の Tone label を用いて記述される。日本語の東京方言の場合、アクセント句内の F_0 の変動は概ね「へ」の字型であること、アクセント句の1モーラ目と2モーラ目との間で声の高さが異なること、起伏型のアクセント句ではアクセント核が存在することは多くの場合に当てはまる。そのため、このような標準的な F_0 の変化形状を表わす Tone label のような情報は音声合成器の言語解析の結果から容易に得ることができる。しかし、表現豊かな音声に見られる句末音調は上述のようにさまざまであった。さらに、表現豊かな音声のテキストからの音声合成を想定し、テキストから句末音調ラベルを予測する研究は、筆者の知る限り、これまでに無かった。そこで本研究では、合成対象の文の言語表現から、句末での F_0 変化形状を記述する句末音調ラベル (アクセント核以後 F_0 が下がったままか、句末付近で再上昇するか否かのラベル) を予測する方式の検討を行なった。

F_0 変化形状を記す ToBI ラベルからの F_0 生成処理については、これまでも研究が進んでいる。例えば、各 ToBI のシンボルの意味する F_0 変動が起こるタイミングの予測⁵⁾ や、ToBI ラベルからの F_0 時系列の生成 (線形回帰モデルを用いるもの⁶⁾、指令応答モデルを用いるもの⁷⁾、および、HMM 音声合成の枠組みを用いるもの^{8),9)}) については既に効果が確

^{†1} 日本電信電話株式会社, NTT サイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation

認されている．そのため，それらの処理の前段に位置する，テキストからの句末音調ラベルの予測方式が完成すれば，テキストからの表現豊かな音声合成の実現に繋がる．実際の応用場面においては，最終的には演出家やクリエイターのような人間が調整を行なうとしても，テキストから自動予測された句末音調ラベルの精度が上げれば，調整作業開始時の初期ラベルとして有用となり，人間の手作業の負担を削減できる見込みがあるので，方式の確立が必要である．

なお，表現豊かな音声における他の特徴の一つとして「強調」があり，強調のラベルを利用した音声合成方式の有効性が確認されている^{5),9)}が，強調は各時間区間の前後の対比に基づく程度を表わすラベルであり， F_0 形状そのものを記述する句末音調ラベルとはラベルとしての性質が異なる為，強調ラベルのテキストからの予測は本研究の対象外とした．

2. 従来研究との関係

テキストから句末音調ラベルの予測を行なった従来研究であるが，英語を対象言語とし，平叙文が大半を占めるラジオニュースを対象に行なわれた研究が報告されている¹⁰⁾．この方法では，句末の位置でのみ，前の句末での予測結果とは独立に，句末音調ラベルの予測が行なわれた．予測には分類木が用いられ，予測のための特徴量（分類木の質問に現われる変数）として，単語，minor phrase, major phrase, 文といった様々な範囲の長さや範囲間の距離，そして，辞書参照して得られる句末の2単語の品詞，次の句頭の1単語の品詞，辞書に定義されている強勢といった約50個の情報が利用された．そして，約50分の音声データを用いて，それらの特徴量と句末音調ラベルの生起との間の相互情報量が調査された¹¹⁾が，その結果に基づく分類木構築前の特徴量の選択は行なわれず，自動学習された分類木を用いて評価が行なわれた．従来法で用いられた特徴量の一つである一文や一発話の長さには，ラジオニュースと本研究が対象とする表現豊かな音声の間では大きな違いがある．そのため，ラジオニュースを対象に得られた特徴量を用いるのではなく，表現豊かな音声から学習によって構築される予測方式を用いることが望ましいと考えた．

2.1 予測に用いる特徴量

これまでに文末詞と句末音調と談話機能（本稿では，意図や発話行為なども含めて，それらを「高次言語情報」と呼ぶことにする）との対応関係の分析が行なわれた³⁾．同じ文末詞でも，高次言語情報と句末音調には複数の可能性があることが報告されている．従って，従来法のように句境界前後の3単語の品詞だけで，高次言語情報の代わりとすることには限界が予想される．

高次言語情報の導入が期待されるが，精度が高く実用の域にある単語や品詞の系列の自動付与に比べて，高次言語情報はその定義や必要最小限の数（基底）すら変わり得る研究段階であり，それらの情報を精度高く予測できる処理方式が存在しない．よって，高次言語情報と句末音調との対応関係についての知見をそのまま用いることはできない．

そのため本研究では，従来の音声合成器の言語解析の結果として得られる形態素解析や句境界情報を用い，それらの分布（特徴量のベクトル）として高次言語情報も表現されることを期待して，予測方式を構成することにした．

2.2 モデルの予測範囲

文末詞を含まない「～さん」という呼び掛けの句（発話）を考えると，聞き手が近くに居る場合には句末での F_0 を上げずに発話し，一方，距離が離れていて大きな声で話しても良い場面であれば F_0 を上げることが容易に想像される．このように，聞き手との距離や発話する場面によって，同じ高次言語情報を持つ発話でも句末音調に複数の可能性があり，高次言語情報が句末音調の決め手になるとは限らない．本研究では，場面の情報が句末音調の決定に有効であることを期待し，予測モデルを場面毎に持ち，上記の曖昧性を防ぐことにした．

また，予測対象は F_0 形状を表わすシンボルではあるが，シンボルの意味する句末音調の上げ下げ自体には話者の個性が関係するので，高い予測性能の実現に有効であると期待し，話者毎のモデルを用いることにした．

このように，モデルによる高次言語情報や場面や個人の予測は行なわず，句末音調のみを予測するモデルとした．

3. 表現豊かな音声データ

3.1 表現豊かな音声の発話場面とデータサイズ

本研究では，童話の語り聞かせ（FT），コールセンターオペレータの電話対応（OP），マスメディアを通じた商品宣伝（AP）といった3つの場面において自然な韻律で発話された音声¹²⁾を用いる．どれも人に向かって語りかけるように発話された音声であり，AP 場面に見られる明かるく元気のよい印象の音声，OP 場面に見られる楽しい印象の感謝や真剣な印象の謝罪などの音声を含む点から，このデータベースの音声は「表現豊かな音声」の範疇に属する．

本研究で対象とする句末音調は，アクセント句内のアクセント核の位置で一度 F_0 が下降した後に，アクセント句の末尾に向けて F_0 が再上昇する現象である．これを意味するラベルが本データベースには付与されている．例えば，FT での「六兵衛よ 入」と呼びかける

表 1 表現豊かな音声データベースの諸元
Table 1 Summary of expressive speech database

	FT	OP	AP
# of sentences	64	104	152
# of accent phrases	684	1,061	1,550
# of speakers	3	3	3
Total length [min.]	13	14	20

発話や、AP での「簡単でしょう ↗」と同意を求める発話等においてみられる。このデータベースでは、前記の 3 場面実際の文を、3 人の話者が、各場面において自然な表現豊かな韻律 (expressive) と、淡々とした読み上げ韻律 (reading) の 2 つの韻律で発話している。収録発話数は話者や場面によって異なった。比較検討を行なう為に、各場面において、2 つの韻律間で、かつ、3 話者間で共通する発話のみを用いる。このような条件下での本音声データベースの諸元を表 1 に記す。アクセント句数と音声データの合計の時間長は 3 話者での平均値である。

3.2 句末音調ラベルの発生率

表 1 の本音声データベースでの句末音調の発生率を表 2(a) に示す。FT, OP, AP といった場面間でアクセント句数が異なる状況下での比較のため、頻度ではなく、相対頻度である発生率 [%] を記した。‘expressive’ の列は表現豊かな音声、‘reading’ の列は読み上げ音声である。最下の “all” の行は全アクセント句での句末音調ラベルの発生率である。 “with pause” と “w/o pause” の各行は末尾にポーズを伴う、及び、伴わないアクセント句での発生率である。FT では口調に関わらずほとんど発生しなかった。さまざまな情景描写の文が発生率を下げたと考えられる。OP と AP では全体の数% のアクセント句で発生した。expressive 列と reading 列との比較から、表現豊かな音声では発生率が高いことが確認できた。また、末尾にポーズを伴うアクセント句での発生率は全体でみた場合の 2 から 3 倍に高まる。

一般に手に入る CSJ のコーパス¹³⁾ の対話音声 (学術講演インタビュー (Int. Aca.), 模擬講演インタビュー (Int. Sim.), タスク指向対話 (T.O.D.)) での句末音調ラベルの発生率を表 2(b) に記す。電話応対場面の音声データ (OP) と CSJ との間で、ポーズを句末に伴う場合の句末音調ラベルの発生率は近い値になっており、本音声データベースの電話応対対話 (OP) が特殊な音声データベースではないことが分かる。

句末音調の発生と Pause の有無との間の相関性のような分析を、句末音調の発生と言語表現との間についても行なえば、明快な知識が得られるかもしれない。しかし、表現間の組み合わせは、少なくとも前後付近の単語を見る必要があり、膨大な数となる。特徴量の組み合わ

表 2 句末音調ラベルの発生率の分布
Table 2 Occurrence distribution of phrase boundary tone labels

(a) our expressive speech database

[%]	FT		OP		AP	
	expressive	reading	expressive	reading	expressive	reading
w/o pause	0.3	0.2	0.7	0.1	0.1	0.1
with pause	1.1	0.8	16.6	4.2	6.4	2.7
all	0.6	0.4	6.8	1.8	2.7	1.2

(b) CSJ database

[%]	Int. Aca.	Int. Sim.	T.O.D.
w/o pause	0.07	0.05	0.03
with pause	21.6	18.5	18.1
all	15.8	13.4	12.9

せ数が膨大で、かつ、データが少ない場合には、データスパースネスがしばしば問題となる。しかし、近年研究の進んだマージンを最大化する機械学習器はそのような状況下でも高い性能を示している。そこで、言語表現の組み合わせについては機械学習で自動的に獲得する方針を採用した。

4. 句末音調ラベル予測方式

4.1 予測問題の設定

本研究では句末音調ラベルの予測という問題を、アクセント句内で句末音調が発生するか否か (F_0 再上昇が起きるか否か) という 2 値分類の問題と考える。これまでに検討された F_0 生成方式⁹⁾ では、2 値ラベルとして句末音調ラベルが用いられている。まずは、その方式での利用を念頭に置き、句末音調が発生するか否かという 2 値の分類器を構築することにした。

予測に有効な特徴量とその組み合わせは機械学習によって重みづけされ、分類器の構成に利用されるようにする。本研究では分類器として Support Vector Machine¹⁴⁾ を用いる。

4.2 分類に用いる特徴量

3.2 で述べたように、ポーズを末尾に伴うアクセント句での句末音調の発生率が高かった。そのため、末尾にポーズを伴うか否かという情報を分類のための特徴量として利用する。他の情報として、複数個の単語の表層形 (出現形)、品詞といった音声合成の言語解析結果から得られる情報を用いて、どの程度までの予測が可能であるのかについて検討を行なうこと

にした。まとめると、あるアクセント句での F_0 の再上昇の有無の分類の為に、そのアクセント句から得られる、

- 句末のポーズの有無
- 句末から句の頭側に向かって順に取った N 個の単語の出現形（表層形）と品詞

を分類器のための特徴量として用いる。単語数の N は実験で様々に振る。アクセント句内の単語数が少なく N に満たない場合にはその不足分は“NULL”のラベルとする。

5. 実験

句末音調のシンボルを用いたコンテキストの追加により、HMM 方式での音声合成において F_0 の生成が改善されることは既に確認されている⁹⁾。そのため、ここではテキスト平文から各句末において F_0 が再上昇するか否かのシンボルの予測の評価に焦点を当てる。まず本提案手法の評価を行ない、最後に従来法¹⁰⁾ との比較を行なう。

5.1 方法

本予測方式を 5 分割交差実験で評価する。表 2(a) のように FT では句末上昇の発生率が低いので、評価対象場面は OP, AP のみとした。単語数 N は 1 から 5 まで変化させた。カーネル関数として多項式関数を用い、その次元数を 2 から 4 まで変化させ、様々な特徴量の組合わせのもとで実験した。

5.2 評価尺度

評価尺度として、i) 精度 (Acc.), ii) 再上昇の再現率 (Rec.), iii) 再上昇の適合率 (Prec.), iv) Cohen's $kappa$ ¹⁵⁾ の 4 つを用いる。句末で F_0 の再上昇の起こらない句の数のほうが圧倒的に多い。そのため、iv) の $kappa$ を用いることにより、句末での F_0 再上昇が生じる場合と生じない場合との間での数量のバイアスの影響を排除した評価を行なう。この尺度は -1 から 1 までの値をとり、1 に近いほど正解との一致度が高いことを意味する。

5.3 結果と考察

評価結果を表 3 に示す。単語数 N 、カーネル関数の次元を様々に振って実験した結果の中で最高の性能を示した結果を記した。電話応対場面 (OP) では単語数が 4、カーネル関数の次元数 (特徴量の組合せ数) が 2 または 3 の場合が、商品宣伝場面 (AP) では単語数が 2 または 3、次元数が 2 の場合が最高性能を示した。このように、OP のほうが AP よりも多くの単語数を必要とする結果となった。

OP においては、例えば、推察であることを伝える「程度が \nearrow と / (思います)」の事例に見られるように、分類に重要そうな助詞が句末 (‘/’ が句境界) に限らず句末よりも文

表 3 Prediction Performance

	speaker id	OP		AP	
		Closed	Open	Closed	Open
$kappa$	1	0.95	0.70	0.98	0.80
	2	0.93	0.77	0.99	0.84
	3	0.98	0.75	0.94	0.71
Acc. [%]	1	99.1	95.0	99.8	99.1
	2	99.4	98.2	100	99.1
	3	99.8	97.9	99.7	98.5
Rec. [%]	1	96.1	66.7	95.3	75.7
	2	92.5	72.0	100	77.3
	3	97.1	63.3	91.5	68.3
Prec. [%]	1	95.6	81.8	100	84.9
	2	94.9	85.7	98.3	91.9
	3	100	95.0	96.2	75.7

頭側の位置に来る事例が見られた。また、「デジープラザ \nearrow という / 」の事例のように、やはり F_0 の再上昇の位置が句末よりも前側に位置している事例が見られた。OP における分類の為に入力情報として多くの単語を必要とする理由の 1 つと考えられる。OP で再上昇の湧き出し誤りが起こった箇所は、再上昇となっても意味を変えるものではないものが多かったため、合成上は問題にならないと考えられる。

このような誤りがあるものの、AP, OP ともに、学習に用いていない評価データを用いた Open テストにおいて、再現率 (Rec.) には改善の余地はあるが、予測された箇所の適合率 (Prec.) は高い結果が得られた。また、 $kappa$ 値も 0.7 以上の高い一致率を得た。

5.4 従来法との比較

2 節で挙げた分類木を用いた従来法¹⁰⁾ との比較を行なった。従来法¹⁰⁾ は対象言語が英語であったので、本質を損なわないよう配慮し、分類に用いる特徴量の一部について付録のような置き換えを行なった。分類木構築には CART を実装した Wagon¹⁶⁾ を利用した。一般に、過学習への対策として、学習データの 25% を用いた木の枝狩りが行なわれるので試したが、その場合の従来法の $kappa$ は 0 から最大でも 0.5 という低い値となった。また、entropy 基準で構築した分類木の一致度も同様に低かった。そこで、頻度基準で、枝狩りは行なわず、葉の最小の事例数を 1 として分類木を構築した。そのようにして構築した分類木と提案法との間での $kappa$ を尺度とした比較の結果を図 1 に示す。

この図 1 の棒グラフの高さが Cohen の $kappa$ の値であり、各棒グラフの上から上下に飛び出たバーが各 $kappa$ の 95% 信頼区間である。白い棒グラフが提案法、灰色の棒グラフが従来法に対応する。

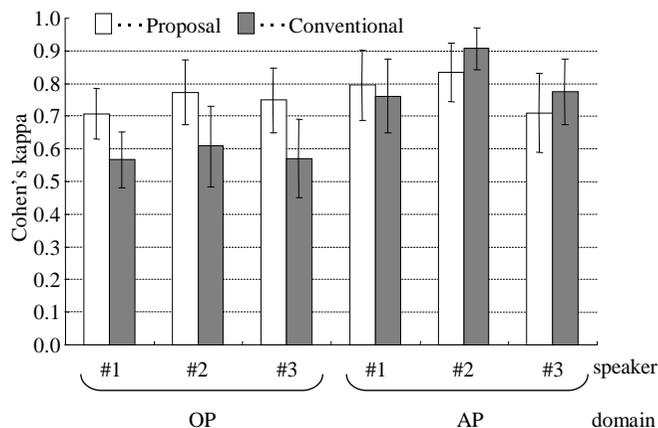


図1 Cohen の kappa での性能比較 (バーは 95%信頼区間)

Fig. 1 Performance comparison with Cohen's kappa coefficients (Small bars mean 95% confidence intervals)

OP では、従来法と提案法との間で信頼区間のバーの重なり部分が僅かであり、かつ、kappa の値が提案法において高くなっている。一方、AP では、信頼区間のバーの重なりは大きく、kappa はほぼ同じ値を示した。これらから、提案法は、従来法に比べて (AP のように) ほぼ同等、または (OP のように) 高い性能を有していると考えられる。

6. おわりに

テキストからの表現豊かな音声合成の実現に向けて、表現豊かな音声においてさまざまな動きを示す句末音調を、合成対象の文から予測する方式の検討を行なった。本手法では、従来の音声合成器の言語解析結果から得られる数個の特徴量と Support Vector Machine とを用いて、場面毎かつ話者毎にモデルを構築した。句末音調ラベル予測の評価実験を行なうことにより、本手法の有効性が確認できた。また、多量の特徴量を用いる分類木を用いる従来法との比較の結果、提案法によって同等またはそれ以上の一致率 (Cohen's kappa) が得られることを確認した。句末音調以外の F_0 形状を表現するシンボルはこれまでの東京方言の日本語の音声合成器の言語解析の結果 (句境界とアクセント型など) から予測できるので、 F_0 形状を記述するシンボルの主要なものはこれで予測が可能となった。

今後は、 F_0 形状の記述そのものとは異なる強調箇所を指定するラベルをテキストから予測する方式についての検討を行なう予定である。

参考文献

- 1) Tang, H., Zhou, X., Odisio, M., Hasegawa-Johnson, M., and Huang, T.S.: Two-stage prosody prediction for emotional text-to-speech synthesis, *Proc. Interspeech*, pp.2138–2141 (2008).
- 2) 前川喜久雄, 菊池英明, 五十嵐陽介: XJToBI:自発音声の韻律ラベリングスキーム, 情報処理学会研究報告音声言語情報処理,39-23, pp.135–140, (2001).
- 3) 石井カルロス寿恵, ニックキャンベル: 句末音調の機能的役割: 談話機能を中心に, 日本音響学会 2004 年春季研究発表会講演論文集 Vol.I, pp.235–236 (2004).
- 4) Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J.: ToBI: a standard for labeling English prosody, *Proc. ICSLP*, pp.867–870 (1992).
- 5) Yu, K., Mairesse, F., and Young, S.: Word-level emphasis modelling in HMM-based speech synthesis, *ICASSP*, pp.4238–4241 (2010).
- 6) Black, A.W., and Hunt, A.J.: Generating F_0 contours from ToBI labels using linear regression, *Proc. ICSLP*, pp.1385–1388 (1996).
- 7) 平井俊夫, 樋口宣男: 韻律ラベリングシステム J-ToBI のラベル情報を用いた重畳型基本周波数制御モデルパラメータの自動抽出, 電子情報通信学会論文誌, D-II, Vol.J81-D-II, No.6, pp.1058–1064 (1998).
- 8) 郡山知樹, 能勢隆, 小林隆夫: 対話音声合成のためのイントネーションラベルのタイミング予測, 日本音響学会 2011 年秋季研究発表会講演論文集, pp. 333–334 (2011).
- 9) 前野悠, 能勢隆, 小林隆夫, 井島悠介, 中嶋秀治, 水野秀之, 吉岡理: 多様な音声合成のための強調コンテキストの自動付与の検討, 日本音響学会 2011 年秋季研究発表会講演論文集, pp. 335–336 (2011).
- 10) Ross, K., and Ostendorf, M.: Prediction of abstract prosodic labels for speech synthesis, *Computer Speech and Language*, pp.155–185 (1996).
- 11) Ross, K.: Modeling of intonation for speech synthesis, Ph.D. dissertation, Boston University College of Engineering, (1995).
- 12) Nakajima, H., Miyazaki, N., Yoshida, A., Nakamura, T. and Mizuno, H.: Creation and Analysis of a Japanese Speaking Style Parallel Database for Expressive Speech Synthesis, *Oriental COCODSA*, paper-id=30, available from (http://desceco.org/OCOCOSDA2010/proceedings/paper_30.pdf) (accessed 2011-11-10) (2010).
- 13) 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明: 日本語話し言葉コーパスの設計, 音声研究, 4(2), pp.51–61, (2000).

- 14) Vapnik, V.: *The nature of statistical learning theory*, Springer (1995).
- 15) Carletta, J.: Assessing Agreement on Classification Tasks: The Kappa Statistic, *Computational Linguistics*, 22 (2), pp.249-254 (1996).
- 16) Edinburgh Speech Tools Library,
available from http://festvox.org/docs/speech_tools1.2.0/book1.htm
(accessed 2011-11-10).

付 録

本実験の従来法では、Ross ら¹⁰⁾の用いた特徴量を、本質を損なわないよう配慮しつつ、日本語向けに置き換えた。その詳細を文献 10) の末尾のリストに対応付けて以下に記す。一般的な点では、Ross らは句や単語の長さや位置を測る単位としてシラブルを用いたが、本研究ではモーラに置き換えた。また、minor phrase をアクセント句、major phrase をポーズにはさまれたアクセント句の系列に置き換えた。

A.1 Dictionary information

- Dictionary stress
英語では強勢の位置を利用していたが、対象が日本語であるためアクセント核の位置に置き換えた。ここでは平板型で核が無いが、起伏型で核が来る位置が長母音上か短母音上か、機能語か、の4つのカテゴリとした。

A.2 Part-of-speech

- 品詞の種類
Ross らは当時の分類木のシステム上の制限からカテゴリ数を8つに限定したが、今回用いたツールには制限が無いので、データベースに現われた全種の品詞を用いた。
- 単語クラスのカテゴリ
Ross らの用いた *less/more-accentable function word* は用いず、機能語、内容語、固有名詞の3つのカテゴリとした。

A.3 Prosodic phrase structure

- phrase break size
アクセント句末かポーズを有するアクセント句末かの2カテゴリとした。

A.4 New/Given status

- 本データベースは音素バランスを優先して収集された音声データベースであるので、*story* や *paragraph* が存在しないので、それらの範囲における情報の新旧は用いなかった。

A.5 Paragraph structure

- 上と同様の理由で、paragraph 内での句や文の位置や長さについての情報は用いなかった。

A.6 Label of other units

- pitch accent
Ross らの論文では、“none, high, downstepped, low”の4種であったが、アクセント句のアクセント型とした。
- boundary tone on word
本研究ではこれが予測対象であるので、用いていない。
- previous boundary tone
未使用との記載が文献 10) 内にあったので用いていない。
- number of syllables from the last prominence
句のアクセント型の核の位置から句末までの距離とした。平板型の場合は0とした。8以上は8とした。
- last preceding pitch accent type in the phrase
句のアクセント型か平板型であれば、‘H’、その他は‘L’とした。
- number of pitch accents since last minor break
句のアクセント型が平板型であれば0、起伏型であれば1とした。
- number of pitch accents since last major break
上記の累積数とした。