日本語韻律コーパスの構築・分析と言語情報を利用した韻律ラベリング手法の評価

Development and Analysis of Japanese Prosodic Corpus and Evaluation of Annotation Method Using Linguistic Constraints

静岡大学情報学部

Faculty of Information, Shizuoka University

北澤　茂良, 桐山伸也，伊藤敏彦，ニック・キャンベル†

Shigeyoshi Kitazawa, Shinya Kiriyama, Toshihiko Itoh, Nick Campbell

＜研究協力者＞

静岡大学大学院情報学研究科

Graduate School of Information, Shizuoka University

三ッ田　佳史　　細川　雄太

Yoshifumi Mituta　　Yuta Hosokawa

This paper is about a prosodic corpus in Japanese: preparation of speech, recording, F0 extraction, phoneme segmentation, and prosodic labeling. We treated several different speeches in situation, speakers, styles, emotion, spontaneity. Extension of J-ToBI prosodic labeling is proposed. Another work is automatic generation of prosodic labeling. Although the generated prosodic labels are not sufficient, they are helpful as an initial value for the manual labeling work. We investigated V-V hiatus by listening to the whole phrases to estimate degree of discontinuity and the exact boundary of two phrases. The features of these V-V hiatus were phrase-initial glottalization and phrase-final nasalization, as well as phrase-final lengthening and phrase-initial shortening.

Key Words: Prosodic Corpus, MULTEXT, Hiatus, Prosody, EGG, J-ToBI, Phoneme Duration

1．はじめに

音声をデータベースとして蓄積する場合，その特徴を記述するラベル付けを行うことによって，データベースの利便性が大きく向上する．分節については，国際音声記号をはじめ確立した表記方法がある．しかしながら, 韻律については, そのような表記方法がなく，従って，韻律ラベル付けされたデータベースの開発も音素の場合と比較して遅れている．

本特定領域研究の中で「コーパス班」として韻律コーパスの作成とその自動化についての研究を行ってきた．この報告では，韻律コーパスの作成方法と作成した韻律コーパスについて述べる．次に，韻律コーパス作成自動化のための試みについて述べる．更に，コーパスに基づく韻律特徴の分析について述べる．

2．韻律コーパス

2.1．韻律コーパスの仕様

韻律コーパスの構成要素としては，音声信号，発話内容テキスト，信頼性の高いピッチ周波数，適切な韻律ラベル，等である．目的によって，またその詳細化の

程度によって，さらに付加的な情報を含めることがあるが，以上の4つの要素は，韻律コーパスとして不可欠である．

2.1.1. 音声収録

発話者の選抜，発話の状況，目的とする韻律情報が含まれているのかどうか，自発性発話の程度，録音品質（騒音，反響，スタジオ，無響室）など，韻律固有の特性に由来する問題がある．本コーパスは東京方言を基本として，新たに収録を行ったものと，既存の音声データベースに韻律情報を付与したものがある．

2.1.2. 音素ラベリング

分節音情報は韻律とは密接な関係があり，韻律事象の生起時点とそこでの対応音素の情報は重要であるので，高精度の音素ラベリングが必要である．HMM の音声認識アルゴリズムにより自動付加した後，スペクトログラム読み取りの手法によって分節区分情報を手動により付加した．

2.1.3. F0抽出

F0抽出はWavesurfer[3]組込みのESPSのアルゴリズムによって自動抽出し，狭帯域スペクトログラムと重ねて表示して，視察によって手修正した．

2.2. 韻律ラベリング

英語の韻律を記述する指標として単語境界情報，抑揚情報等を表記するTones and Break Indices（ToBI）システムが開発されている．ToBI には定量的な記述が欠けているが,F0は手修正した値を付与することにし，必要に応じて原音声信号に立返ることによって定量的な解析が可能になる．ピッチアクセント言語である日本語の F0値も音素レベルの分節化情報も信頼できるので,種種の韻律的仮説を検証することが出来る．

ToBI（Tone and Break Indices）の日本語版がJapanese-ToBI（J-ToBI）であり，東京方言の韻律的特徴を記述するために考案されたものである[1]．J-ToBI ラベルは，単語境界の情報を記す単語層，シンボルレベルで表し得る基本的な韻律事象の系列を表記するトーン層，各韻律境界における区切りの深さを表す BI 層，咳払い・言い淀み・強調など韻律に関係したその他の情報を記述する miscellaneous 層の4層構成になっている．

2.3. 韻律コーパスの作成

2.3.1. 日本語 MULTXT 韻律コーパス

EU 加盟国11言語を対象としたデータベース作成プロジェクトEUROM1[4]の仕様に従い，MULTEXT（多言語韻律コーパス）の日本語版を作成した．MULTEXT とは，EUROM1内の5ヶ国語（英・仏・独・伊・西）について音韻・韻律ラベリングを行った韻律コーパスである[5]．原稿のテキストは,1つが5〜6文で構成される40の小節で成り立っている．人名・地名などは各国独自のものを用いるが，全体の文意は保存されている．日本語版も，文意を保って翻訳されている[6]．

音声収録条件はEUROM1に準拠しているが，模擬自発発話と EGG 信号を追加した．話者は20代から40代の男女各3名の合計6名である．模擬自発発話の収録に際しては，小節ごとに想定した状況を指示し，役柄になりきるよう演じさせた．朗読との比較において，模擬自発発話スタイルの音声では，語・句単位の卓立がより明瞭となっている．音声サンプルは無響室で収録した．音声データおよび EGG（Electro Glotto Graph）信号の基本周波数（ピッチ周波数）を10 ms 毎に抽出した．EGG の波形分析から声帯波の開放率・閉鎖率の推定値を求めた．MOMEL アルゴリズムを適用して,スプライン曲線で整形したイントネーション曲線を抽出した[7]．

音素ラベルを付与し，音声学の専門家の手によるアクセント核の位置情報，無声化・鼻濁音化の有無の検聴を行っている．J-ToBI 韻律ラベリングスキームに基づく韻律ラベルを完了している．音素ラベル付与は，自動音素セグメンテーションを4人のラベラが手作業で修正した．

2.3.2. 重点領域研究対話音声　千葉大マップタスク

市川・堀内グループ（千葉大）で音素ラベルを付したもの（128対話中，8対話のみ ATR に委託して，音素ラベルの付加を行ないました.）について，F0抽出と修正を加え，韻律ラベルを付した．自然対話なので X-JToBI を援用してラベル付けを行った．

2.3.3. 読上げ音声天気予報

男女各１名のナレータが簡易防音室内で天気予報文２８９文を読上げ収録した音声(広瀬班で作成)に F0抽出修正，音素ラベルを手動で修正，J-ToBI ラベルを付与した．明瞭な音声で F0抽出，音素ラベル，韻律ラベルのいずれも確度が高い．

2.3.4. 案内模擬対話文

板橋(筑波大学)グループで韻律情報抽出した中から，研究用連続音声データベース（模擬対話文）から抜粋した J-ToBI ラベル付与のための音声データファイルおよび読上げテキストを選んで，さらに，話者は東京出身の男３名女２名でそれぞれ異なる文セットを読上げたものを選抜した（約２９分）に付いて，音素ラベルの付与，韻律ラベル（J-ToBI）を付与した．比較的明瞭な音声で F0抽出，音素ラベル，韻律ラベルのいず

れも確度が高い.

2.3.5. 模擬対話音声

板橋(筑波大学)グループで韻律情報抽出した中から,文部省科学研究費補助金重点領域研究「音声・言語・概念の統合的処理による対話の理解と生成に関する研究」(略称:音声対話)において編集したものVol.4の中の東京大学で収録した模擬対話(作成した原稿を読み上げている TOK2002)の中で話者が東京出身のもの４話者について(11'28")F0抽出,音素ラベル,韻律ラベルを付与した.

その中の東京方言の比較的自然な１対話(Vol.7 NTU1002D 9'24")についてF0抽出,音素ラベル,韻律ラベルを付与した.二人の対話をモノラルで収録しているため,発話が重なることがあり,その部分については正確な分析が行えなかった.F0抽出,音素ラベル,韻律ラベルは話者ごとに分離して付与した.

2.3.6. 模擬感情音声

これは,広瀬(合成)班で作成した発話スタイル音声データベースである.テキストはATR音韻バランス文の読上げ音声データである.次のスタイルでの発声を行わせたものである.1)丁寧に,2)ぞんざいに,3)悲嘆にくれて,4)明るく楽しく.発声は,2名のプロのナレータである.男性話者年齢:30代,女性話者年齢:30代,収録は,計算機室内に設置された簡易防音室で行った.

この一部の音声について,F0抽出,音素ラベル,韻律ラベルを付与した.

2.3.7. マルチモーダル対話音声

市川・堀内グループ(千葉大)で収録した,二人の対話者がそれぞれ独立した防音室に入って対話を行うことにより,相手の音声がマイクに入り込むことなくクリアな音声を収録するとともに,相手の正面からの顔および上半身の映像をプロンプタを通して提示して,表情やジェスチャーを見ながら対話を進めた.このとき収録した２チャンネルの音声についてF0抽出,音素ラベル,韻律ラベルを付与した.

2.4. 日本語の韻律表記の補強について

文字記述文の読上げの韻律現象でも,J-ToBIの枠組にない現象が多くある.アクセント核について,実際の揺れを詳細に記述する必要がある.聴覚印象に基づくものと,F0抽出による音響分析的なものとの食い違いについての詳細な検討が必要であるが,これまでの表記は曖昧であった.

東京語方言の,が行鼻濁音化の記述が必要である.同じく,無声化の程度(無声化傾向にあるものも含めて)と有無の記述が必要である.これらは,日本語の韻律として注目すべき現象である.

フレージングすなわち単語のチャンキングについての理解を深め,BIラベルの定義の厳密化が必要である.日本語において韻律的に重要なのは卓立とそれに伴うチャンキングおよびピッチレンジの大幅な変化であることから[8],発話の中心とそれに伴う句の範囲の判定が重要なラベリングとなる.自由発話における非流暢性の問題の取扱い,パラ言語的情報の記述は韻律の外だが重要な記述である.

3. 言語情報の利用による自動韻律ラベリング

韻律の持つ多様性により韻律ラベリングの自動化は困難であり,データベース作成は最終的に人手に頼らざるを得ない.我々の目的は,「韻律ラベリング支援システム」の開発である.すなわち,韻律ラベリングの完全自動化を目指すのではなく,適切なラベリング支援情報をラベラに提供することで,手動ラベリング作業の効率化を図ることを目標としている.今回,読上げテキストの言語情報に着目して,音素ラベル・J-ToBI[2][9]の初期ラベルを自動生成する手法を開発した.単語層,トーン層,BI層のそれぞれについて,例えば下記のような情報からラベルを推定することができる.

単語層:形態素解析結果から,ほぼ完全に自動付与が可能である.

トーン層: *F0* 曲線の山の頂きの形状はアクセント型の情報から推定可能である.裾の情報は音素セグメンテーション結果に含まれるポーズ位置の情報から得られる.

BI 層:構文解析結果を用いることでほぼ推定可能と考えられる.**BI 2** はほぼ文節境界に対応しており,それより細かい単語境界は一意に**BI 1** と定まる.**BI 4** は文末に対応する.

この観点から我々は,

(1) 音素ラベルとアクセント型の情報を用いたトーン層の自動ラベリング手法

(2) 構文解析を用いた**BI** ラベル値の自動推定手法

を提案する.

3.1. トーン層自動ラベリング

音素ラベルとアクセント型の情報を言語情報として利用した,トーン層ラベルの自動生成手法を提案する.本節の冒頭で述べた*F0*曲線の特徴に基き,**%L**,**%wL**,**L%**,**wL%**,**H-**,**H*+L** の6種類のラベルを自動生成することとした.各ラベルの生成規則を以下に示す.

(1) **%L/%wL** 音素ラベル中のポーズラベルの終端に**%L** を付与する．ただし，後続のアクセント句の先頭が「強い音節」である場合には，**%L** のかわりに**%wL** を付与する．

(2) **L%/wL%**アクセント句の終端に**%L** を付与する．ただし，後続のアクセント句の先頭が「強い音節」である場合には，**L%**のかわりに**wL%**を付与する．

(3) **H-**アクセントの立上り位置として，2 モーラ目の母音の中心位置に付与する．

(4) **H*+L** アクセント型からアクセント核があるとされるモーラの終端に付与する．

ここで「強い音節」とは，1)当該アクセント句が頭高型，2)1 モーラ目が長母音，のどちらかであることを意味する．

評価用テキストに対するアクセント型の情報は，アクセント辞書の見出しとアクセント結合規則を用いて手動で付与した．ラベラが手動で付与したラベルを正解とし，自動ラベリングの精度を検証した．時間軸上のずれは無視し，ラベルの記号にのみ着目して，全ラベル数（A)，正解ラベル数（C)，置換誤り数（S)，脱落誤り数（D)，挿入誤り数（I）を調査した結果を表1 に示す．ここで，置換誤り・脱落誤り・挿入誤りについては，トーン層のラベルを，1)**%L・%wL**，2)**L%・wL%**，3)**H-**，4)**H*+L ・*?**，5)**>** の五つのカテゴリに分類した上で，正解ラベルと生成ラベルの比較をそれぞれのカテゴリの中で行うことによって検出している．表1 から，全体の約80%は正しい種類のラベルが付与されていることが分かる．

表 **1**．トーン層自動ラベリング精度．各列は左から全ラベル数(A),正解ラベル数(C)，置換誤り数(S)，脱落誤り数（D)，挿入誤り数（I）をそれぞれ表す．

	A	C	S	D	I
Number	45172	36260	3655	5257	5383
Rate(%)	100.0	80.3	8.1	11.6	11.9

さらに，トーン層ラベルの各記号について，全ラベル数に対する完全正解ラベル数，準正解ラベル数の割合を算出した．ここで完全正解ラベルとは，正しい記号で，かつ，時間軸上のずれが50ms 以内の位置に付与されたラベルであり，準正解ラベルとは，記号は正しいものの，時間軸上のずれが50ms を超えるラベルを指す．また，全ラベル数に対する置換誤り数，脱落誤り数，挿入誤り数の割合も併せて算出した．これらの値を表2 に示す．各列は左から順に，全ラベル数に対する正解ラベル数(C)，準正解ラベル数(N)，置換誤り数(S)，脱落誤り数(D)，挿入誤り数(I)の割合であることを示す．

表2 から**%L**，**%wL** はほぼ完璧に付与可能であることが分かる．**L%**，**wL%**，**H-**の挿入誤りの割合がやや高いことから，実際の発話中では，今回辞書に基づいて検討した以上に，アクセント句結合が頻出していたことが窺える．今後の自動ラベリング精度の向上へ向けて，アクセント句境界の判定に基本周波数パターンの情報を取り入れることも必要になると考えられる．

表2．トーン層自動ラベリングにおける，記号別の正解／不正解ラベル数の割合．

Symbol	C(%)	N(%)	S(%)	D(%)	I(%)
%L	91.2	0.0	8.4	0.3	0.4
%wL	96.4	0.0	3.2	0.5	0.4
L%	75.2	2.4	16.2	6.3	10.8
wL%	81.8	2.1	6.4	9.6	13.3
H-	83.8	8.5	0.0	7.7	29.0
H*+L	67.2	24.2	0.0	8.6	7.9
_?	0.0	0.0	75.4	24.6	0.0
all	71.6	8.7	8.1	11.6	11.9

3.2. BI 層自動ラベリング

言語情報として構文解析結果を用いることで，**BI** ラベルの自動推定を行う手法を提案する．40 節からなる日本語MULTEXT 韻律コーパスの読上げテキストを評価用テキストとして構文解析を行った．解析には，日本語構文解析器KNP[10]を用いた．構文解析の単位となる文節に着目し，文節同士のつながりの深さ，及びポーズの有無を考慮した下記の規則によってBI 値1 〜4 を付与することとした．

1 文節内の単語境界に付与する．

2 文節の終端に付与する．

3 BI 2 のうち，読点のあるもの，及び直後にポーズがあるものに付与する．

4 文の終端に付与する．

表3． **BI** 層自動ラベリング結果NC，NA はそれぞれ各 **BI** 値における正解ラベル数，全ラベル数を表すCRはNCのNAに対する割合(正解率)を表す．

Index	NC	NA	CR(%)
1	13552	15014	90.3
2	6773	9322	72.7
3	2139	2834	75.5
4	1864	2220	84.0
all	24328	32377	75.1

日本語MULTEXT 韻律コーパス全480 節に対し，音素ラベルの時間情報を用いて，**BI** ラベルの自動付与を行った．手動ラベルを正解として，自動推定性能を評

価した．表3 にBI 値別のラベル値正解率を示す．

なお，all の行のNAの合計がBI 1 ～4 のNAの合計とは一致しないのは，この数が手動によって付与されたラベルの全数であるためである．すなわちこの数には，BI1 ～4 の数に加えて，2-・2m ・2p ・3-・3m の各ラベルの数も含まれる．当初の予想通り，BI 2 とBI 3 の識別に改善の余地があるが，BI 1 とBI 4 の推定性能は極めて高いことが分かった．

表4 のconfusion matrix にBI ラベルの置換誤りの分布を示す．今回，自動生成対象としたBI 1 ～4 の4種類のみに着目すると，BI 1 とBI 2 の間の置換誤り，及びBI 3とすべきところをBI 2 とする誤りが多いことがわかる．

表4.BI ラベルの置換誤りの分布を表すconfusion matrix．各行が正解ラベル，各列が自動生成したラベルの種類を表す．

	1	2	3	4
1	13552	1327	58	56
2	2414	6773	89	32
2-	708	908	8	1
2m	130	66	0	0
2p	0	2	0	0
3	166	452	2139	77
3-	153	950	31	0
3m	2	1	23	0
4	148	12	130	1864

例えば，「守って」「くれますから」という二つのアクセント句からなる動詞句についてKNP は一つの文節と解釈して解析を行うため，提案規則によると「守って」の末尾にはBI 値1 が付与される．これはBI 2 をBI 1 と誤る典型的な例であり，この問題の解決のためには，自動ラベリングによって生成されるアクセント句のモーラ数が長すぎる場合には，適切な単語のBI 1 をBI 2 に修正する可能性を提示するといった対処が必要になると考えられる．また，BI 1 をBI 2 と誤る例の多くは，一つのアクセント句を複数の短いアクセント句に分割してしまうことが原因であり，アクセント結合を考慮することによってある程度修正が可能と考えられる．

BI 3 をBI 2 と誤った例の多くは，次の句読点までのモーラ長が比較的長い部分に出現していた．人間は長い文章を発声する場合，適度に抑揚をつけて発声するが，今回自動生成したラベルによると，多数のアクセント句の連鎖を，抑揚をつけずに続けて発声したことになってしまったわけである．この問題に対しては，生成したBIラベルにおいて，BI 3(ないしBI 4)間のモーラ数が比較的長い場合，その区間を二分する位置にBI 3 を挿入するといった処理が有効と考えられる．

3.3．音素強制切り出しの性能評価

言語情報として読み上げテキストを利用し，HMM を用いた強制切り出しによって音素自動ラベリングを行

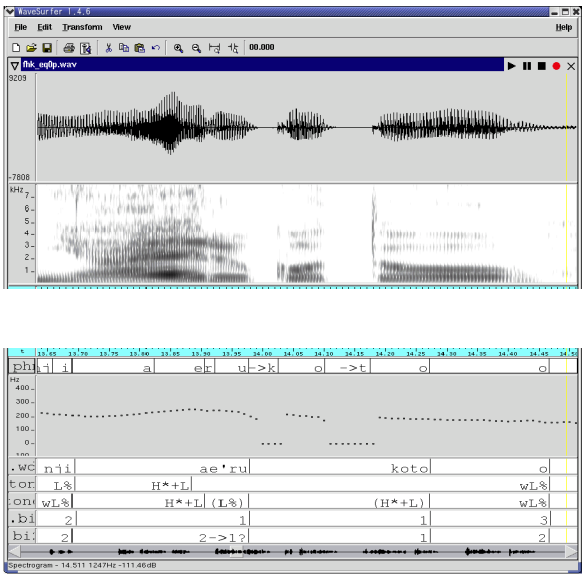

Figure 1. Prosodic labeling supporter.

った．ラベラが手作業で作成した音素ラベルを正解とし，自動セグメンテーション結果の誤差について，平均値と標準偏差を音素別に算出した結果，誤差が最も大きい音素でもその値は20ms であり，高い精度であることが分かった．音素別では，/r/や/y/の半母音・鼻音の/n/や/m/・有声破裂音（/b/,/d/,/g/）など，有声子音のラベリング精度が特に高いことが分かった．

生成した音素ラベルが正解ラベルの位置からプラス方向・マイナス方向のどちらにずれるかの割合を音素別に調査したところ，ほとんどの音素が正解よりも遅く誤りがちであるのに対し，音素/w/は正解よりも早く誤る傾向があることが分かった．ずれの時間軸上での方向をラベリング支援情報として付加することにより，手動ラベリングの効率化が期待できる．

3.4.韻律ラベリング支援システム

自動ラベリングの評価結果に基づき，図1 に示すようなラベリング支援システムが実現可能となる．

- **初期韻律ラベルの精度の改善**トーン層ラベルについて，アクセント句結合規則をより詳細に導入して，言語情報からのアクセント句推定精度を向上させ，初期ラベルの高品質化を図る．また，基

本周波数パターンの情報の利用を検討する.

- **「時間情報」の有効利用**モーラ数の多いアクセント句に対して，実際には複数のアクセント句からなる可能性があることを考慮し，トーン層ラベルに脱落誤りの可能性があることを教示する．BI 層についても，BI 1 ではなくBI 2 の可能性を示唆する．逆にモーラ数の少ないアクセント句に対しては，アクセント結合の可能性があるため，トーン層における挿入誤りの可能性と，BI 層における BI の置換誤り(BI 2 ではなくBI 1)の可能性を教示する.
- **時間軸上誤差の方向の教示**音素ラベルについて，実際の境界位置より早く誤る傾向にある音素は後ろへ，遅く誤る傾向にある音素は前方へと，修正の向きを併せて教示する.

図1 は，「会えることを」という発声内容に対する音素ラベル，及びJ-ToBI ラベルを表示したものである．音素ラベル層の‘→k’と‘→t’は，これらのラベルが矢印の方向へ修正されるべき可能性が高いことを表す．2 つあるトーン層のうち上段の“L% H*+L wL%”という記号列が，手動ラベリングによる正解ラベルである．下段のトーン層に現れる‘(L%)’と‘(H*+L)’は，この部分が「会える」と「ことを」という短い二つのアクセント句に対応しており，実際には一つのアクセント句として発声される可能性が高いことから，挿入誤りの可能性があるということを表している．BI 層上段の“2 1 1 3”は手動による正解ラベルである．下段の‘2→ 1?’はBI 1 の置換誤りである可能性を示唆している.

4．日本語における母音接続の音響的・韻律的分析

韻律は超分節的特徴(suprasegmental)といわれ，分節(segment)すなわち音素の特徴とは独立したものと思われがちであるが，聴取によって効率的に音声を理解する上で重要な手がかりとなっている連接や句読点，焦点，そして，強調などの韻律現象は，実際，分節的特徴と密接に関連している．一方，自然に話された流暢な音声では，句及び単語の境界は流暢さのために不明瞭になり，また，音素単位に分節することが困難になる．このことは，音声認識と音声合成で本質的に困難な問題となっている．本節では，単語や句の境界について，先行のアクセント句の終端モーラと後続アクセント句の先頭モーラが同一２母音からなる場合での母音接続(vowel-vowel hiatus)を分析する.

4.1 日本語の母音接続(hiatus)

日本語は開音節からなっているので，後続句が母音から始まるときには母音接続が生じる，すなわち，無音区間無しに同一母音が連続する．この母音接続は日本語では頻繁に生じる.

日本語 MULTEXT 韻律コーパス [6]から採取した例で最も多いのは助詞であった．先行アクセント句は形態素(例えば名詞)＋助詞であり，これに助詞又は頭位母音アクセント句が後続することによって母音接続が生じる．例としては，が+ある，は+あめ，しか+ありません，に+いって，て+エキゾ，と+おもう，を+おしえて，の+おたく，などがある.

次に多いのは副詞/頭位母音アクセント句である.例えば,まだ+あたらしい,いったい+いつ,もし+いきて,せっかく+うとうと,きちんと+おこなう,などである.

次に，少数の単語接続（複合語），例えば，こむぎ+いろ，タクシー+いちだい，などがある.

4.2母音接続分析法

句の韻律境界は音声・EGG 波形，広帯域･狭帯域スペクトログラム，を参照して決定したが，さらに，聴取によってアクセント句境界の分離度を評価した.

4.2.1句の聴取実験による分析[11]

ここでの調査対象は日本語のアクセント句境界にある母音接続の境界である．試料は女性話者 fhk の45音声である．もちろん，これらの母音接続には無音等の切れ目はない.

手で与えた境界を固定点として，先行句と後続句を分離して，聴取実験のために音声資料を切出した．切出し点は固定点を基準として一基本波周期を単位として声帯振動周期ごとに前後に５周期分移動させる．結果として１母音接続当り片側11個で合計22個の句単位の聴取実験用音声片を作成した.

句単位の音声は被験者にランダム順に提示した．被験者は句の音声の自然さを判断することが求められた.特に，各句の初めと終わりの部分に注意を払った．判定結果は0から５の段階で示した．５は“自然な”，０は“全く不自然な”である．各試料ごとに集計して平均し,+2,+1,0,-1,-2で示した.被験者は男子学生6名,女子学生2名であった.

4.2.2 電気声門図 EGG 波形の分析

電気声門図(EGG)の波形から解放率を求め,声門の開閉の周期より基本周波数を KAY の CSL ツール[12]を用いて求めた．声門解放率は声質に関連していて，50%以上は荒々しい声,50%は普通の声,20-30%は気息音の混ざった声である．解放率は時間軸に沿って滑らか

に変化しているが，急激な変化があったときは声門閉鎖音が生じたことを示す．EGG から抽出した基本周波数は平滑化した基本周波数と異なり，声門閉鎖と同期して瞬時基本周波数が一時的に低下する．

4.3 母音接続の分析結果

母音接続の分割は多くの場合に声門閉鎖として，あるものは鼻音化として解釈できる．結果としての句末母音の伸長と句頭母音の短縮は共通している．

4.3.1 句の音調で観測された母音接続

いくつかのアクセント句（3以上はまれ）が結びついてより大きな中間句・イントネーション句を形成する．イントネーション句間の境界は休止または疑似休止によって区切られる．F0が降下するのはイントネーション句の特徴であるが，ピッチリセットとも言う．

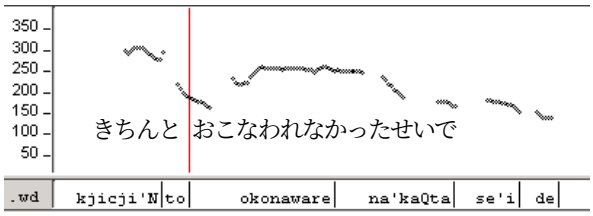

Figure 2. A pitch reset from fhko2r*.*

図1にはアクセント句 *kjicji'Nto* に焦点があって強調されてるが一方 *okonawarena'kaQta seede* は抑制される．後続句においてピッチリセットが観測され，ピッチ幅が圧縮される．聴取結果によれば-4ピッチ周期において先行句の聴取評点は1.0だが，後続句の評点は1.5であった．

4.3.2 句頭の声門閉鎖化

多くの場合，句頭母音は声門閉鎖によって強調される[13]．このことはまた，先行句末の母音が後続句頭の母音と同一である場合にも当てはまる．図2に示すのはこの例で，EGG の開放率は句境界付近で下降上昇する．ここが前後の句を聴取的にも分ける最適点でもある．

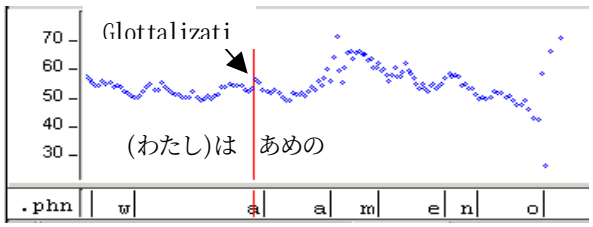

Figure 3. EGG open quotient in fhkr8r*.*

分析した45例の母音接続について，17例ではF0と開放率のくぼみとして声門閉鎖が明瞭に観測でき，23例では他の特徴が伴って弱い声門化が見られた．残りの5例では句末の母音の鼻音化が見られた．句頭の単語が強調された場合には句境界において明瞭な声門閉鎖が見られたが，先行句が助詞「が」で終わり，後続句頭に「あ」が続くときと，先行句が助詞「を」で終わり，後続句頭に「お」が続くときで，後続句が強調されないときには，句頭の声門閉鎖が不明瞭になる．

4.3.3 句末の鼻音化

鼻音化から非鼻音化への切替りが知覚的な母音接続を分解する手がかりとなっている．スペクトログラムの模様では高周波数部分のエネルギーが相対的に少ないことが鼻音化の特徴である．この対比は同一母音の連続でも観察できる．東京方言では句末の助詞「が」は鼻音化する．図3の例のでは助詞「が」が鼻音化し後続の「あ」は鼻音化してないことが高域のスペクトルに現れている．

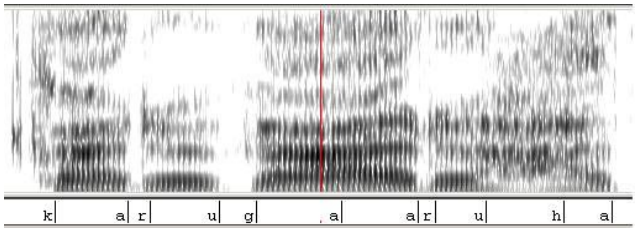

Figure 4. A contrast of nasalized+/- at the phrase boundary of /myuNzjkaruga-aruhazu/ in fhkr4r*.*

鼻音化は母音接続を分解するのに役立つ．我々に試料では+/-nasal の対立のみが4例，それに声門閉鎖が伴うのが12例あった．

4.4 BI と聴取評価値の関係

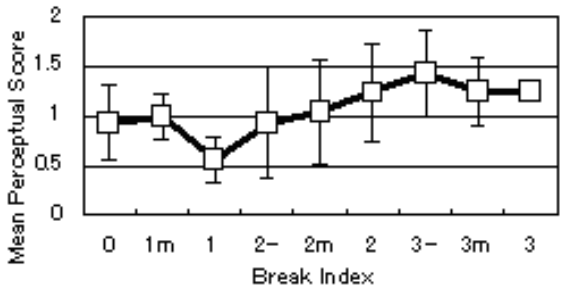

Figure 5. Break Index and mean perceptual scores at the best boundary point with standard deviation.

BI(Break Indices)は単語や句間の韻律的結合の度合いを示す[14]．知覚される句間の分離度の主観的評価値である．本研究での聴取評価値もまた分離度に関する．アクセント句境界で多くは BI 2と付けられるが，聴取によってまたイントネーション曲線によって 3-,2-,2m となる場合がある．BI はラベラの主観的判断によるので構文とは一致しない．

BI の値と聴取評価値の相関を図 4.に示した．聴取評価値が **BI** 値の増加と共に上昇するすなわちより分離度が強まる傾向がある．

4.5 分節音継続長の分析

句の終わりを示すために句末のモーラは伸長される，一方，後続の句頭のモーラはモーラ時間の遅れを取戻してモーラの等時性を維持するために短縮される．同一母音 VV の系列はおよそ２モーラ分の継続長があるが，同一母音間の境界は２母音の中間点より後方で見つかる．

調査した 45の統計では，先行句の句末母音の継続長は後続句頭の母音の継続長の平均1.7倍であった．後続句の単語が強調された場合には，先行母音は伸長されず，後続母音は通常の長さを維持するのでこの比は0.76まで低下することがあった．

5.まとめ

韻律コーパス班で開発したコーパスの一部について紹介した．新規に収録したのはEUROM1 に準拠しMUL-TEXT，マルチモーダルの対話音声，天気予報の読上げ，模擬感情音声である．また，既存の模擬対話音声データベースとマップタスク音声を使用した．いずれも，F0 抽出と手修正，音素ラベルの自動抽出と手修正を行い，更に，J-ToBI ラベルを付与した．これらのコーパスは日本語の韻律研究の基礎データとして広く利用できるようにする．

言語情報を用いた自動韻律ラベリング手法を提案した．音素境界とアクセント型の情報からJ-ToBI ラベリングスキームにおけるトーン層ラベルを生成させ，BI 値の自動推定を構文解析結果を利用して行った．さらに，HMM を用いた音素セグメンテーションの音素別の分別性能を検証した．実験を通して，提案手法によりtone 層の71.6%，**BI** 層の75.1%のラベルを正しく生成できたことを示した．この精度は手動ラベリングの初期ラベルとしては十分であると期待できる．実験で得られた知見を基に，音素毎のラベリング誤差の傾向，及びモーラ長に基づく複数のアクセント句分割方法の可能性を考慮したラベリング支援システムを考察した．今後は，ラベリング支援システムの試用実験を通して，システムの有効性を検証する．

J-ToBI ラベル付けした句境界を分離度（まとまり）を先行･後続句をそれぞれ聴取して評価した．多くの場合に最高聴取評価値（分離度）は単峰性の最大聴取評価値として得られた．よくある VV 母音接続のパターンは:句末助詞-母音 (1) +/-nasal の対比はスペクトログラムのパターンとして見える．先行の *ga* は鼻音化されるが後続の *a* は鼻音化されない．(2)後続句頭の母音は声門閉鎖になる．声門閉鎖はF0の低下と EGG の開放率のくぼみとして観察される．(3) *wo - o* の場合に声門閉鎖は顕著でなく，EGG 開放率は不安定であるが，スペクトルの変化が有効な特徴である．

句末の形容詞-*a* 母音あるいは母音で終わる単語-母音で始まる単語の場合，上記の場合より後続句頭の声門閉鎖がより強い．

EGG の開放率あるいはF0のくぼみ，また，基本周期の一時的伸長として観測される句頭の声門閉鎖は母音接続の特徴として重要である．

BI は聴取評価値と相関があることが示された．母音接続個所の母音の継続長は隣接母音の相互強調の関係に依存している．

参考文献

[1] M. Beckman, and G. Ayers, “The ToBI Handbook,” Technical Report, Ohio-State University, 1993.

[2] J. J. Venditti, “Japanese ToBI Labelling Guidelines,” Techni-cal Report, Ohio-State University, 1995.

[3] http://www.speech.kth.se/wavesurfer/

[4] Chan, D., Fourcin, A. et al., “EUROM - A Spoken Lan-guage Resource for the EU,” Eurospeech95, vol.1, pp.867-870, 1995.

[5] E. Campione, and J. Veronis, “A multilingual prosodic database,” ICSLP98, pp.3163-3166, 1998.

[6] S. Kitazawa, et al, “Preliminary Study of Japanese MULTEXT: a Prosodic Corpus,” ICSP2001, pp.825-828, 2001.

[7] Hirst, D. and Espesser, R., “Automatic Modelling of Funda-mental Frequency Using Quadratic Spine Function,” Travaux de l’Instiut de Phonetique d’Aix-en-Provence, No. 15, pp.75-85 (1993).

[8] Beckman, M. and J. Pierrehumbert, “Japanese Prosodic Phrasing and Intonation Synthesis,” Proceedings of the 24th Meeting of the Association for Computational Linguistics, pp.173-180, 1986.

[9]ニックキャンベル，“Tones and Break Indices(ToBI)システムと日本語への適用,”音響誌, vol.53, no.3, pp.223-229, 1997.

[10] S. Kurohashi and M. Nagao, “Building a Japanese Parsed Corpus while Improving the Parsing System,” ICLRE98, pp.719-724, 1998.

[11] Kitazawa Shigeyoshi, Kiriyama Shinya, Itoh Toshihiko, and Yukinori Toyama, 2004. Perceptual Inspection of V-V Juncture in Japanese, SP2004, 349-352.

[12] Instruction Manual Electroglottograph (EGG) Model 4338, Kay Elemetrics Corp., Lincoln Park, NJ 07035-1488 USA (April 1995).

[13] Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M., 1996. Glottalization of word-initial vowels as a function of prosodic structure, Journal of Phonetics, 24, 423-444.

[14] Venditti, Jennifer J., 2002. The J-ToBI model of Japanese intonation. In S. - A. Jun (ed.) *Prosodic Typology and Transcription: A Unified Approach*. Oxford: Oxford University Press.