[3] D. Hirst, A.D. Crist, and R. Espesser : Levels of representation and levels of analysis for the description of intonation systems, Merle Horne (ed.) *Prosody : Theory and Experiment*, Academic Press, (2000) pp. 51-87.
[4] 小林，板橋，速水，竹澤：日本音響学会研究用連続音声データベース，音響学会誌 Vol. 48, No. 12 (1992) pp. 888-893.
[5] 堂下，新見，白井，田中，溝口：音声による人間と機械の対話，第7編，第2章，PASD コーパス：重点領域研究模擬対話音声コーパス，オーム社 (1998) pp. 361-375.

3.3 話し言葉コーパスの韻律ラベリング

3.3.1 目的

話し言葉のデータベースないしコーパスを構築するにあたって，ぜひともほしい研究用付加情報の1つが韻律特徴の情報である．音声の韻律特徴は，パラ言語情報の伝達媒体となる点において，話し言葉による情報伝達の本質にかかわっているからである[1]．

本稿では，共通語の自発音声に対する韻律ラベリングの手法として筆者らが「日本語話し言葉コーパス」[2]のために開発した X-JToBI を紹介する．音声の韻律特徴は，持続時間長，イントネーション（F_0 パターン），声質（発声様式）など様々な姿で伝達されるが，X-JToBI はそのうち持続時間長とイントネーションに焦点を当てたラベリング体系となっている．以下ではまず，X-JToBI の出発点となった J_ToBI の概要を紹介し，次いで X-JToBI の概要を紹介する．

3.3.2 J_ToBI

ToBI は，音声の音韻論的分析に立脚して個別言語のイントネーション特徴をラベリングするための一手法である．当初英語を対象として提案された ToBI は[3]，すぐに多くの言語に拡張されていった．日本語を対象とする J_ToBI は J.Venditti らによって 1995 年に提案された[4]．J_ToBI では，表 3.5 に示した4層のラベルによって日本語（東京方言）の特徴をラベリングする．

Word 層には，語境界と語の音形がラベルとして表示される．たとえば「これは大変だ」という発話であれば，{kore} {wa} {taiheN} {da} のように語（ないし形態素）の情報がそれぞれ1個のラベルをなし，音声信号の時間軸上で各語

表3.5　J_ToBIラベル

層	表現する情報	ラベル
Word層	形態論情報	
Tone層	音韻論的tone	%L，L%，H%，HL%，H-，H*+L
BI層	韻律境界の強さ	0，1，2，3（補助記号としてm，…）
Misc層	備考情報	

（形態素）の終端位置に付与される．

Tone層ラベルには，東京語の語彙的アクセント（H*+L），アクセント句頭におけるF_0上昇の頂点を表現するフレーズトーン（H-），アクセント句の句頭におけるF_0の低さを表現する句頭境界トーン（%L），同じくアクセント句末のF_0特徴を表現する句末境界トーン（L%，H%，HL%）がある．

東京語のアクセント句は，それが有核句，すなわち語彙アクセントを含む句であれば，%L，H-，H*+L，L%というトーンの連鎖から構成され，無核句すなわち語彙アクセントを含まない句であれば，%L，H-，L%という連鎖から構成される．これらの記号列は，アクセント句の冒頭が相対的に低いピッチ（%L）で始まった後すぐに上昇し（H-），語彙アクセントがあればそこで下降し（H*+L），最後もまた低く終わる（L%）ことを表現している．

アクセント句の終端には，言語ないしパラ言語情報の伝達に重要な種々の局所的音調が生じる．J_ToBIでは，これを表現するためにH%とHL%が用意されている．これらのラベルはいずれもL%の直後に付与されて，L%H%で上昇調をL%HL%で上昇下降調を表現する．

BI層ラベルは韻律境界の強さに関する作業者の判断を離散化したもので，原則として1～3の整数値をとる．「1」は語境界，「2」はアクセント句境界，「3」は，そこで発話のピッチレンジがリセットされるアクセント句境界（イントネーション句もしくは中間句—intermediate phrase—境界）に該当する．J_ToBIではこのほかに「0」を用いるが，これは{kore}{wa}が{korya}となるように，形態素境界が縮退した場合に用いられる．最後に，Misc層は種々の備考情報を記入するための層として提供されている．

3.3.3 J_ToBI の問題点

すでに述べたように，J_ToBI は日本語の音韻論的分析をもとに構築された体系であるから，日本語を母語とする作業者であれば比較的容易に習熟することができる．しかし，J_ToBI が立脚している音韻分析はもっぱら朗読音声の分析から帰納されたものであるため，自発性の高い話し言葉では対処不可能な現象—たとえばフィラー（filled pause）や語断片—に遭遇することがある．

「エー」「アノー」「マ」などのいわゆるフィラーについては，それを1つのアクセント句と見るのか，あるいは先行／後続アクセント句の一部と見るのか，そのイントネーションをどう記述するのか，そもそもフィラーとフィラーでないもの（「アノ」や「マー」などの音形はフィラーとしても通常の語としても出現し得る）をどう区別するかなどの問題を解決しなければならない．「（ワ）私が」「（コ）これを」「日本人（ダ）ではない」のような発話に含まれる語断片についても，フィラーと同じ問題を指摘できる．

フィラーや語断片は文字に転記し得る現象だが，そうでないタイプの現象もある．代表は句末音調のインベントリ不足である．川上蓁が，東京方言の句末上昇に4ないし5種の区別を認めていることからわかるように[5]，現在の J_ToBI の句末音調（上昇調は2種のみ）インベントリは明らかに不足している．

また，アクセント句頭に生じるピッチ上昇のバリエーションも記述困難である．アクセント句頭句末に生じる局所的音調は，パラ言語情報の伝達に重要な役割を果たしていることが明らかにされているから（第2章第3節参照），これも適切にラベリングしておきたい．

最後に，分節音ラベルの必要性を指摘する．J_ToBI は，F_0 パターンの記述に焦点を絞ったラベリング体系であった．しかし，自発音声の情報を適切に解釈するためには分節音の情報が欠かせない（たとえば3.3.4 d.項で説明する PNLP）．また，自発音声には朗読音声に生じることの稀な分節音の変異も多量に生じるから，その特徴解明のためにも分節音ラベルが欠かせない．

3.3.4 X-JToBI

X-JToBI という名称は eXtended J_ToBI に由来しており，J_ToBI を自発

表3.6　X-JToBIラベル

層	表現する情報	ラベル
単語（Word）層	形態論情報	CSJでは短単位の音形
分節音層	分節音情報	p, t, k, a, i, u, e, o, 〈pz〉, 〈sv〉, etc.
トーン層	音韻論的tone	%L, L%, H%, HL%, LH%, H-, A,〉, pH, pL, FL, FH, etc.
BI層	韻律境界の強さ	1, 1＋p, 2, 2＋p, 2＋b, 2＋pb, 3, P, 〈D, 〈F, PB, etc.
プロミネンス層	Toneの変異	PNLP, FR, HR, EUAP
注釈（Misc）層	備考情報	AYOR etc.

音声向けに拡張したラベリング仕様を意味している．X-JToBIラベルの一部を表3.6に示す．表3.5と比較すると2つの層が新設され，各層ラベルのインベントリも拡張されている．以下では，X-JToBIの主要な拡張点についてのみ説明する．X-JToBIに関する詳細は文献［6］を参照されたい．

a．分節音層および分節音ラベル

X-JToBIでは子音，母音，ポーズなどの種類と境界位置を表す分節音ラベルを新しい層として導入した．これによって，たとえばToneとそれを担う分節音との時間関係を検討できるようになった．

表3.6の第3列のラベルのうち，＜pz＞はポーズ，＜sv＞は母音終端での非意図的な声帯振動の持続を示すラベルである．「日本語話し言葉コーパス」の分節音ラベリングでは，このようないわゆる音素の範疇に入らないラベルも用いている（詳細は文献［7］参照）．ユーザーは必要なラベルだけを適宜選択すればよい．分節音ラベルはもっとも数の多いラベルであるため，付与作業が楽ではない．しかし，Word層の情報（音素列情報）があれば，自動アラインメントツールを利用できる．自動アラインメントの精度は人手作業に劣るが，目的によってはそのまま利用することもできる．文献［8］は，「日本語話し言葉コーパス」における自動アラインメントについての解説を含んでいる．

b．BI層インベントリの拡張

J_ToBIを用いて「日本語話し言葉コーパス」のサンプルを試行的にラベリングしてみたところ，作業者間の不一致がもっとも目立ったのがBI層であった[9]．J_ToBIにはフィラーや語断片のBIに関する規定がないことがその原因である．X-JToBIにおけるBIの拡張は，1)中間値の許容と，2)非流暢性に関する拡張の2種類に大別できる．

(1) 中間値

X-JToBIでは，N＝1，2，3の整数にA＝b，p，bpのアルファベットを結合させたN＋A形式の中間値BIを許容する．N＋は何らかの理由でその境界がNよりも強いと判断されたことを示し，Aはその根拠と判断された現象を示す．たとえば，1＋pはアクセント句内部の語境界にポーズ（＜pz＞）が存在することによって，通常の1よりも強い境界となっていることを示す．たとえば，「フクオカ＜pz＞ニ（行きました）」のようなケースである．2＋pは，アクセント句直後に顕著なポーズが存在していることにより，通常のアクセント句境界（BI＝2）よりも強く知覚される境界である．2＋bは，アクセント句末に句末音調（BPM，後述）が生じているために通常の2よりも強いが，ピッチレンジはリセットされていないので3よりは弱いと判断された境界である．これら2つの特徴が共起している境界は2＋bpとする．

(2) 非流暢性

表3.6中のP，＜D，＜F，PBなどは，非流暢性（disfluency）に関するBIラベルである．Pは，「フク＜pz＞オカニ」のように語中にポーズが生じた場合に付与するラベルである．＜Dは，Dとともに語断片の前後に付与される．同様に，＜Fは，Fとともにフィラーの前後に付与される．PB(parasitic boundary）は稀に用いられる特殊なラベルで，アクセント句末に2個の句末音調が継起していると判断された場合（たとえばL%H%H%）に，最初の句末音調の終端に付与される．PBが用いられる代表的な発話は，いわゆる半疑問の直後にアクセント句境界があり，そこにも上昇音調（H%）が生じているケースである（詳しくは文献［6］p.12参照）．

c．トーン層インベントリの拡張

トーン層ラベルの拡張は句末音調のインベントリに関するものが多いが，そ

のほかにフィラー用ラベルも導入した．またJ_ToBIでは，H*+Lであったアクセントのラベルを Aに変更した．

(1) 句末音調ラベル

句末音調ラベルに追加されたのは，LH%と HLH%である．従来の H%などと同じく，これらもまたL%に後続する．L%LH%はいったん低いピッチが持続した後に上昇が始まる音調（文献［5］の「反問の上昇」），L%HLH%は上昇した後下降し，最後にまた上昇する音調である．これらの音調を，先行するL%も含めて，複合境界音調（boundary pitch movement，BPM）とよぶ．

BPMを研究するためには，F_0曲線の頂点や谷の位置の時間情報が重要である．しかし，J_ToBIではHL%を1つのラベルとして発話末に付与していたために頂点の時刻を知ることができなかった．この問題を解決するために，X-JToBIにはポインターという補助記号を導入した．

ポインターには pHと pLの2種類があり，それぞれBPM中のF_0の頂点と谷底に付与される．L%HL%を例にとると，まずF_0が上昇を開始する位置にL%が，上昇の頂点に pHが，そして発話末に HL%が付与される．X-JToBIのユーザーは，発話末のラベルを見ることによって，BPMのタイプが上昇下降調であることとBPM末のF_0値を知ることができ，また pHによって頂点の時刻と上昇頂点のF_0値を知ることができる．

pLは，LH%における低ピッチ区間の終端と HLH%における2つの頂点に挟まれた谷底の位置を示すために利用する．

ラベル＞はエクステンダーとよばれ，トーンの延長を表現するために用いる．たとえば「コノマエサー，カレガー」の下線部にL%H%を用い，さらに上昇を終えたピッチの高さをそのまま維持しながら母音を延長する発話がある（若い女性に多く見られる）．この場合，下線部のF_0は／￣の形状をなす．X-JToBIでは上昇の頂点（高ピッチ平坦部の始端）に H%を付与し，高ピッチ平坦部の終端に＞を付与して通常のL%H%と区別する．

エクステンダーはまた，アクセント句頭における低ピッチの延長現象をラベリングするためにも利用される（文献［6］参照）．

(2) フィラー用トーン

フィラーの多くは低いピッチで発音されるが，ときには明らかに前後よりも高く発音されることがある．ただし，ピッチの変動は認められず，フィラー全

体を通して一定のピッチが維持される．X-JToBIでは，BI層で<FとFに囲まれたフィラー区間のピッチを高いか低いかに二値的に分類し，FHないしFLで表現している（なお，韻律特徴の面からみてフィラーをどう認定するかについては文献［6］のp.17参照）．

d．プロミネンス層ラベル

トーン層のラベルは，イントネーションの外形を記述するものであった．それに対して，プロミネンス層のラベルはトーン層で規定された音調の変種を表現する．

PNLPは，L%HL%が句末の2モーラにまたがって実現されていることを表す．たとえば「データが」という発話末に上昇下降調が生じる場合，通常のL%HL%であれば上昇下降は最終モーラである「ガ」の内部で生じる．しかし，「デー ⌐タ¬ が」のように，次末（penult）モーラが頂点を担う発話もある．句末に上昇下降調が生じているという点で，外形はL%HL%と同じであるがトーンと分節音のアラインメントが異なっている．このような場合，X-JToBIではトーン層ラベルにはL%HL%を用い，アラインメントが通常と異なっていることを示すためにプロミネンス層にラベルPNLPを与える．

PNLPはpenult non-lexical prominenceの意であるが，この名称が示しているように語彙的に定まったプロミネンス，すなわちアクセントはPNLPの対象ではない．たとえば，「横浜まで」が「横浜 ⌐マ¬ デ」となってもPNLPの対象とはせず，助詞が単独でアクセント句を構成しているとみなす．

FRとHRは，L%H%の変種を表すプロミネンス層ラベルである．FRは，上昇の開始点が次末ないし次々末モーラにある変種に付与される．これは文献［5］が「浮き上がり調」とした変種であり，FRはfloating riseの略である．

HR（hooked rise）は，やはり文献［5］の「つりあげ調」である．これは句末の2モーラが1音節として発音され，その始端でピッチがステップ状に急上昇しているように聞こえる変種である．典型的には「〜です」「〜ます」で終わる発話末に生じる．

EUAPはemphasized unaccented accentual phraseの略である．無核句が連続している発話において，特定の句が強調されることによってそのピッチレンジが前後の句に比べて拡大されている場合，強調された句の末尾にこのラベ

ルを付与する．なお，X-JToBIでは有核句が強調されていることを示すラベルは規定していない．

3.3.5　問題点

前述した種々の改良によって，X-JToBIの記述力はJ_ToBIに比べて格段に向上している．しかしすべての問題が解消されたわけではない．X-JToBIのMisc層のラベルAYOR（At Your Own Risk）は，何らかの理由でラベルが一意に決定できなかった箇所を示すラベルである．AYORの大多数は，引用の発話（「～という」の類）に付与されている．

AYOR以外にも問題がある．ここでは，紙幅の関係で1つだけ重要な問題を紹介しておく．X-JToBIによる自発音声のラベリングでも，BIの判断には困難が少なくない．特にピッチレンジのリセットを判断するに当たって，どの程度の時間範囲を参照するかは理論上も重要な問題である．

図3.9は，4個の有核アクセント句からなる発話のF_0形状の模式図である．A1～4はアクセントによって形成されるF_0のピークを，B1～3はアクセント句境界を示している．B1は明らかに2であり，B3は明らかに3であるが，B2は判定が難しい．A2とA3を比較すれば，ピッチレンジがほぼ同じである（ダウンステップが生じていない）から，3と判定できる．しかし，A1と比較すれば，A2のピッチレンジは明らかに狭いので，2と判定できる．

「日本語話し言葉コーパス」のラベリングでは，作業効率を重視してピッチレンジの比較は直接隣接するアクセント句との間でだけ行った．しかし，理論上も実用上もこれでよいという保証はない．BI値に3+ないし4を新設する

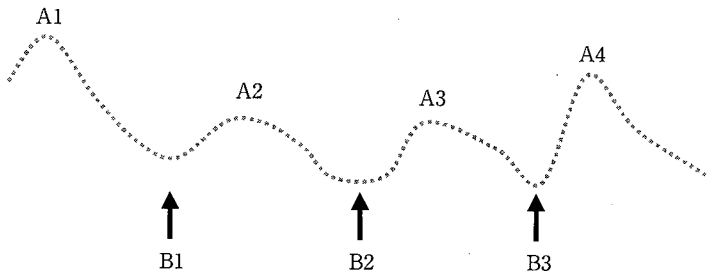

図3.9　BIの判定が困難なイントネーション

のが1つの解決策であるが，その境界の言語的特性をどう規定するかは今後の日本語イントネーション研究の課題である．

［参考文献］

［1］ 前川喜久雄：「パラ言語的情報」別冊国文学「現代日本語必携」53（2000）pp. 172-175.
［2］ 前川喜久雄：『日本語話し言葉コーパス』の概要，日本語科学，15（2004）pp. 111-133.
［3］ Silverman, K. et al : ToBI : A Standard for Labeling English Prosody, *Proceedings of the Second International Conference on Spoken Language Processing* (*ICSLP 92*), 2 (1992) pp. 867-870.
［4］ Venditti, J. : Japanese ToBI Labeling Guidelines, *Ohio State University Working Papers in Linguistics*, 50 (1997) pp. 127-162（当初 1995 年に Web 上で公開）.
［5］ 川上蓁：文末などの上昇調，国語研究，16（1963）pp. 25-46.
［6］ 前川喜久雄・五十嵐陽介・菊池英明・米山聖子：「日本語話し言葉コーパス」のイントネーションラベリング Version 1.0，日本語話し言葉コーパス付属電子文書（2004）.
［7］ 前川喜久雄・菊池英明・藤本雅子・米山聖子：『日本語話し言葉コーパス』の分節ラベリング Version 1.0，日本語話し言葉コーパス付属電子文書（2004）.
［8］ 菊池英明・前川喜久雄・五十嵐陽介・米山聖子・藤本雅子：日本語話し言葉コーパスの音声ラベリング，音声研究，7-3（2003）pp. 16-26.
［9］ 菊池英明・前川喜久雄：自発音声韻律ラベリングスキーム X-JToBI によるラベリング精度の検証，日本音響学会 2002 年秋季研究発表会講演論文集，1（2001）pp. 259-260.

＊［6］，［7］は http://www.kokken.go.jp/katsudo/kenkyu_jyo/corpus/よりダウンロード可能.

3.4 ジェスチャを伴う韻律コーパス

3.4.1 はじめに

音声対話において表出する韻律（プロソディ）として，一般には基本周波数，パワー，時間情報などの音声情報がおもに研究されているが，韻律の概念をもっと広くとらえた場合，視線やうなずきなどのジェスチャも韻律情報と考えることができる．音声対話におけるジェスチャなどを研究するためには，ジェスチャもアノテートされた音声コーパスの存在が望ましい．本節ではマルチモーダル対話コーパスの構築，およびジェスチャの分析例としてうなずきの機能に関する研究について紹介する．