7

The J_ToBI Model of Japanese Intonation

Jennifer J. Venditti

7.1. INTRODUCTION

This chapter presents an overview of Japanese intonational structure and the transcription of this structure using J_ToBI, a variant of the general ToBI tagging scheme developed for Tokyo Japanese. Since the 'Japanese ToBI Labelling Guidelines' (Venditti 1995) were first distributed, J_ToBI has been used in numerous linguistic and computational contexts as a way to represent the intonation patterns of Japanese utterances. This chapter is intended not as a mere rehashing of the 1995 Guidelines, but rather as a comprehensive discussion of the fundamentals of Japanese intonation and the principles underlying the J_ToBI system.

In Section 7.2, we describe the prosodic organization of Japanese and its intonational patterns.[1] We discuss Japanese prosody from a cross-linguistic perspective, highlighting similarities between Japanese and other languages. Section 7.3 then provides an overview of the J_ToBI system. The discussion assumes the reader has some familiarity with intonation description, and with the general ToBI framework. Section 7.4 points out the differences between this new system and its predecessor, the Beckman–Pierrehumbert model presented in *Japanese Tone Structure* (Pierrehumbert and Beckman 1988). Section 7.5 gives an overview of the efforts toward automatization

The author would like to thank Mary Beckman, Sun-Ah Jun, and Kikuo Maekawa for insightful discussions and comments throughout the ongoing development of the Japanese ToBI system.

[1] This discussion and the J_ToBI system itself rely heavily on the model of Japanese tone structure put forth by Beckman and Pierrehumbert (see Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988, *inter alia*), which uses a tone-sequence approach to intonation modelling. However, a few important differences between J_ToBI and the Beckman–Pierrehumbert model will be discussed in Section 7.4. This approach is distinct from the superposition-based models of Japanese intonation (e.g. Fujisaki and Sudo 1971; Fujisaki and Hirose 1984; Venditti and van Santen 2000), which will not be discussed here.

of J_ToBI labelling, as well as the degree of labeller agreement using this system, and Section 7.6 lays out future directions for research on Japanese intonation.

7.2. JAPANESE PROSODIC ORGANIZATION AND INTONATION PATTERNS

7.2.1. *Pitch accents*

Japanese is considered a *pitch accent language*, in that the intonational system uses pitch to mark certain syllables in the speech stream. In this way it is similar to languages like English, which also uses pitch accents in its intonational system. However, there are several fundamental differences between the two. First, Japanese and English differ in the level (lexical vs. post-lexical) at which pitch accent comes into play. In Japanese, pitch accent is a lexical property of a word, and thus the presence or absence of an accent on a particular syllable in a Japanese utterance can be predicted simply by knowing what word is being uttered. Take for example the minimal pair shown in Figure 7.1.

Here, the verb /ueru/ in the phrase *uerumono* 'something to plant' is lexically-specified as unaccented, while that in *ue'rumono* 'the ones who are

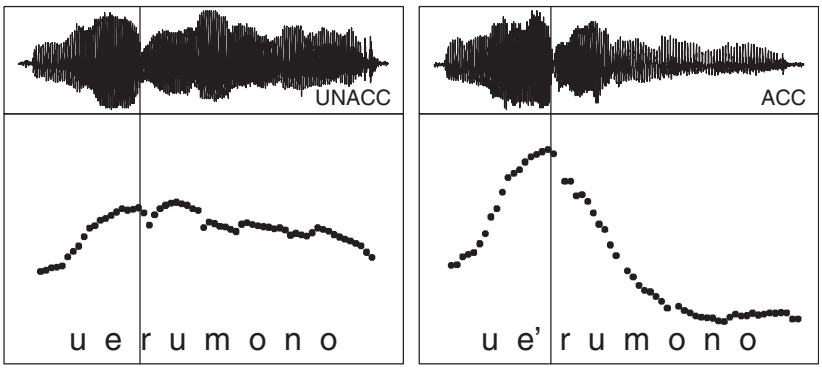

FIGURE 7.1 Waveforms and Fo contours of unaccented *uerumono* 'something to plant' (left) and accented *ue'rumono* 'the ones who are starved' (right) phrases, uttered by the same speaker. The x-axis represents the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contour. Both panels are plotted on the same frequency scale, and vertical lines mark the end of the second mora in each phrase.

starved' is specified as accented on the second mora /e/.[2] The accented phrase displays a precipitous fall in pitch starting near the end of this accented mora, while the unaccented phrase lacks such a fall.[3] This lexical distinction contrasts with languages such as English, in which pitch accents play a role at an entirely different level. In English, the location of metrically strong syllables in a word is determined at the lexical level, and it is these syllables (most often the strongest, or 'primary-stressed' syllable) which serve as docking sites to which pitch accents may be associated at a post-lexical level.

A second difference between the two languages is the function and distribution of pitch accents. In English, pitch accents serve to highlight (or make 'prominent') certain words or syllables in the discourse, and the distribution of pitch accents in an English utterance reflects this function. In a given utterance, there will be a number of metrically strong syllables that can potentially be made even more prominent by the association of a pitch accent. On which of these syllables pitch accents will fall is highly dependent on the linguistic structure of the utterance. That is, an interaction of various factors related to the syntax, semantics, pragmatics, discourse structure, attentional state, etc. will determine where the pitch accents are to be placed in English. In Japanese, in contrast, pitch accents are a lexical property of a given word, and thus they lack any such prominence-lending function. This leaves little room for variability in distribution of accents in a Japanese utterance.

A third difference between the languages is the shapes and meanings of the pitch accents themselves. In Japanese there is only one type of pitch accent: a sharp fall from a high occurring near the end of the accented mora to a low in the following mora. In English, the inventory of pitch accent shapes is far more diverse. There are a number of pitch accent shapes, in which the Fo can rise or fall to/from the accented syllable, or can maintain a local maximum/minimum on that syllable. Each shape has associated with it a specific pragmatic meaning which that accent lends to the overall meaning of the intoned utterance (see e.g. Pierrehumbert and Hirschberg 1990). The Japanese falling accent does not have any such meaning associated with it.

In summary, although both Japanese and English use pitch accent in their intonation system, the languages are in fact quite different with respect to the role that pitch accents play. The languages differ in the level at which pitch

[2] In the transcriptions, accented words contain an apostrophe after the vowel with which the accentual fall is associated; unaccented words lack such a marking.

[3] In the figure, the high to which the Fo rises in the accented case (right) is higher than that in the unaccented case (left). This systematic height difference has been reported in previous studies (e.g. Poser 1984; Pierrehumbert and Beckman 1988; and many others). However, while accented peaks do tend to be higher than unaccented peaks, there is a large amount of variability in both, and there are plenty of cases in read and spontaneous speech where this relative height relation is reversed. Future

accents come into play, in the function and distribution of accents, and in the shapes and meanings of the accents in the inventory.

7.2.2. *Prosodic groupings*

In addition to pitch accents, another important part of Japanese intonation is the grouping of words into prosodic phrases. Speakers can organize their speech into groups of intonational units, which are defined both tonally and by the degree of perceived disjuncture among words within/between groups. This grouping occurs at two levels in Japanese.

First, there is a lower-level grouping, such as that shown in each panel in Figure 7.1. The verb *ueru/ue'ru* is combined with the following unaccented noun *mono* 'thing or person', into a single prosodic phrase. This level of prosodic phrasing in Japanese is termed the *accentual phrase* (AP), and is typically characterized by a rise to a high around the second mora, and subsequent gradual fall to a low at the right edge of the phrase. This delimitative tonal pattern is a marking of the prosodic grouping itself, separate from the contribution of a pitch accent. Both panels in Figure 7.1 consist of a single accentual phrase with the delimitative tonal pattern, though the accented case (right panel) also shows the fall of the lexical accent.[4] The degree of perceived disjuncture between words within an accentual phrase is less than that between sequential words with an accentual phrase boundary intervening. In Tokyo Japanese it is most common for unaccented words to combine with adjacent words to form accentual phrases, though under some circumstances a sequence of accented words may combine, in which case the leftmost accent survives and subsequent accents in the phrase are deleted.

The second type of prosodic grouping in Japanese is the higher-level *intonation phrase* (IP), which consists of a string of one or more accentual phrases. Like accentual phrases, this level of phrasing is also defined both tonally and by the degree of perceived disjuncture within/between the groups. However, the tonal markings and the degree of disjuncture for the IP are different from those of the accentual phrase. The intonation phrase is the prosodic domain within which pitch range is specified, and thus at the start of each new phrase, the speaker chooses a new range which is independent of the former specification. Since there also is a process of *downstep* in Japanese, by which the local pitch height of each accentual phrase is reduced when following a lexically accented

investigations using large amounts of J_ToBI-tagged data are necessary in order to uncover the linguistic factors that are at work in determining this height relationship.

[4] Here, since the accent occurs early in the phrase, the delimitative initial rise is obscured.

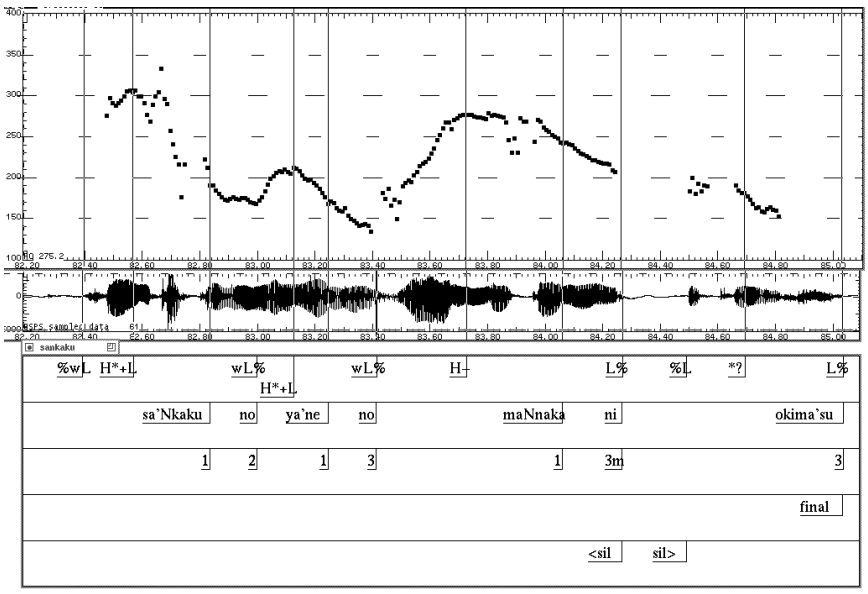

FIGURE 7.2 Fo contour, waveform, and J_ToBI transcription of the utterance ≪sankaku≫: triangle-GEN roof-GEN middle-LOC put 'I will place it right in the centre of the triangle roof'. The x-axis shows the time-course (in sec) of the utterance; the y-axis shows the frequency (in Hz) of the Fo. (Taken from Venditti 1995.)

phrase, one will often observe a staircase-like effect of accentual phrase heights, which is then 'reset' at an intonation phrase boundary. In addition to this behaviour of pitch range, the degree of perceived disjuncture between sequential words across intonation phrase boundaries is larger than that between words within or across accentual phrase boundaries.

Figure 7.2 contains a J_ToBI-transcribed example utterance showing words grouped into accentual phrases and higher-level intonation phrases.[5] The prosodic phrasing of this utterance was judged by a labeller as follows:

accentual phrasing	{ }	{ }	{ }	{ }
intonation phrasing	[ ]	[ ]	[ ]	[ ]
	sa'Nkaku no	*ya'ne no*	*maNnaka ni*	*okima'su*
	triangle-GEN	roof-GEN	middle-LOC	put

[5] At this point, the reader should focus his/her attention only on the Fo contour, the waveform, and the word tier (the 2nd from the top in the label window). A detailed discussion of the symbols in the other label tiers will be presented in following sections.

The accentual phrases *sa'Nkaku no* 'triangular' and *ya'ne no* 'roof-GEN' each are characterized by a rise then rapid fall in the Fo contour. These two APs combine to form the first intonation phrase, with *ya'ne no* being downstepped due to the pitch accent on *sa'Nkaku*, resulting in a staircase-like Fo trend. There is then an expansion of pitch range on the next phrase *maNnaka ni* 'middle-LOC'—this and the virtual pause between *ya'ne no* and *maNnaka* suggest an intonation phrase boundary.[6] The details of the labels in the tone (1st), break (3rd), and other label tiers will be discussed in the following sections.

In addition to the pitch range and disjuncture cues to intonation phrase boundaries, this prosodic unit is also characterized by optional rising or rise-fall tonal movements at its right edge. These movements serve to cue various linguistic and paralinguistic meanings of the utterance, such as questioning, incredulity, explanation, insistence, etc. (e.g. Kawakami 1963/1995; Venditti *et al.* 1998). Each intonation phrase in Figure 7.2 ends in a low tone without such movement, though examples of the various boundary pitch movements occurring in Tokyo Japanese will be discussed in Section 7.3.3.

This section has described the two levels of prosodic grouping in Japanese intonation: the accentual phrase and the intonation phrase. Each of these levels has analogues in other languages as well. Languages as diverse as French and Korean also have tonally-delimited groupings of words like the Japanese accentual phrase (Jun 1993; Jun and Fougeron 1995), and an even larger number of languages have boundary pitch movements which occur at the edge of larger prosodic units analogous to the intonation phrase. Of course, the specific tonal markings used in each language may differ.

English has intonation phrase boundary pitch rises that cue meanings such as questioning and continuation. However, unlike Japanese, English does not have a level of prosodic grouping analogous to the accentual phrase, though the pitch accents of English have a function similar to that of phrasing and pitch range variation in Japanese (see e.g. Venditti *et al.* 1996; Venditti 2000). As mentioned above, since the Japanese pitch accent is hard-coded into the lexical specification of a word, there is little room for variability in pitch accent distribution, as in English. However, the grouping of words into both accentual and intonation phrases (and the pitch range specification of those phrases) is dependent on an interaction of various factors such as the word accentuation, syntactic branching structure, focus, discourse structure, or attentional state, etc.—just those factors affecting English, albeit in a different way.

[6] The phrasing of the remainder of the utterance will be discussed below in Section 7.3.5 when we introduce phrasing/tonal mismatches.

This discussion of Japanese prosodic organization and intonation patterns in comparison with other languages is very important from a cross-linguistic perspective. It shows that the intonational systems of otherwise very diverse languages can be remarkably similar to one another, while maintaining their individual differences. These differences may in fact turn out to be the result of differing means to achieve similar goals. However, only more research on a variety of languages will show how far one can take these cross-linguistic comparisons. In this process, it is essential to be able to use a common framework like the ToBI system to facilitate comparison. With such a tool in hand, we will be much more prepared to start sorting out the similarities and systematic differences among various languages.

7.3. OVERVIEW OF THE JAPANESE ToBI TAGGING SCHEME

The J_ToBI intonation labelling scheme is consistent with the design principles of ToBI systems for English (see Silverman *et al.* 1992; Beckman and Hirschberg 1994; Beckman and Elam 1994) and other languages (this volume). As in other ToBI systems, the transcription consists of the speech and Fo records for the utterance, and a set of symbolic labels. The mandatory labels of a J_ToBI transcription are divided into five separate label tiers in which labels of the same type are marked: tones, words, break indices, finality and miscellaneous.[7] Other optional user-defined tiers can and should be added, as appropriate for the focus of research at each particular site.

The following sections describe the symbolic labels used in the various tiers of a Japanese ToBI transcription.[8] As mentioned in the introduction, J_ToBI for the most part closely follows the theory of Japanese tone structure put forth by Beckman and Pierrehumbert (see Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988, *inter alia*), though a few important differences between J_ToBI and the Beckman–Pierrehumbert model will be highlighted in Section 7.4.

[7] At present, some sites do not use the finality tier. This will be discussed further in Section 7.3.8.

[8] The system described here is identical to that outlined in the 'Japanese ToBI Labelling Guidelines' (Venditti 1995). The reader is referred to this work for more details of the transcription procedure (see also Campbell (1997) for an overview in Japanese). In addition, since the writing of this chapter, an extension of the J_ToBI tagging scheme, dubbed X-JToBI, has been developed by Maekawa and colleagues at the National Language Research Institute (NRLI) in Tokyo, for use in tagging their 'Corpus of Spontaneous Japanese' database (see e.g. Maekawa and Koiso 2000; Maekawa *et al.* 2002). This new scheme introduces additional labels that are necessary to transcribe the spontaneous speech phenomena that they have observed. The reader is referred to future work coming out of NRLI to track the development of this new X-JToBI scheme.

J_ToBI is intended as a tool: the entire purpose of the system is to provide a standard for prosodic labelling of diverse speech data, in order to promote continued research on Japanese intonation. The system is primarily qualitative, in that the symbols employed (and their positioning) reflect the phonological contrasts present in the language. As such, it is useful for those wishing simply to describe the intonational organization of Japanese utterances, for example a psycholinguist needing to describe the prosodic phrasing of his/her experimental stimuli. At the same time, J_ToBI can be a quantitative tool as well. A J_ToBI-labelled database can provide a valuable resource for those wishing to do quantitative modelling of Japanese intonation, for example a computational linguist needing to predict the Fo height relationship between the delimitative high and the accent high within an accentual phrase. Thus, Japanese ToBI is a general-purpose prosodic labelling tool that can be used in many different research contexts.

7.3.1. *Lexical accent tone*

The H*+L composite label placed within the accented mora is used to mark the lexical accent in accented accentual phrases. The H* portion indicates that the high part of the falling tone is associated with the accented mora itself, and the following +L indicates that a low occurs at some fixed point afterwards, usually within the following mora. This H*+L accent label is absent in unaccented words.

Figure 7.2 shows a full J_ToBI transcription of the example utterance ≪sankaku≫. In the tone tier (the 1st from the top in the label window), the H*+L labels on *sa'Nkaku* 'triangle' and *ya'ne* 'roof' mark the lexical accents. The downstep of *ya'ne no* is not explicitly marked (as downstep is in English ToBI), since it is entirely predictable from the lexical accent specification of the preceding phrase.

In many cases, the position of the H*+L label will coincide with the location of the actual Fo maximum (or in the case of a plateau, the start of the precipitous fall), as is the case in Figure 7.2. However, it is not uncommon for the peak to occur after the accented mora, but still be perceived as occurring on the accented mora (e.g. see Sugito 1981; Hata and Hasegawa 1988; Venditti and van Santen 2000). In such cases, two labels are placed: the H*+L is labelled within the accented mora, as usual, and an additional < label is used to mark the actual delayed Fo peak. That is, the H*+L label indicates that an accent is phonologically associated with that particular mora, regardless of whether the Fo peak occurs at that point or not. If necessary, the additional < label

pinpoints the actual location of this phonological event in the phonetic record. Careful labelling of the actual Fo event in J_ToBI transcribed databases is essential for research on Fo timing and peak alignment, and on systematic pitch range variation across phrases.

7.3.2. *Accentual phrase tones*

As described in Section 7.2.2, the accentual phrase in Japanese is tonally defined by an initial rise to a high around the second mora of the phrase, then subsequent gradual fall to a low at the right phrase edge. This tonal pattern is shown on the unaccented phrase *uerumono* in Figure 7.1 (left panel), and on the phrase *maNnaka ni* in Figure 7.2. The initial phrasal high tone is marked in J_ToBI by placing a H- label on the second mora of the phrase, while the final low boundary tone is indicated by L% placed at the phrase edge.[9] When the accentual phrase follows a pause (as it does in Figure 7.1), an additional delimitative %L tone is marked at the phrase onset, to provide an anchor from which the Fo rises. Thus, the complete tonal transcription of the APs shown in Figure 7.1 is:

unaccented AP	%L H-	L%
accented AP	%L (H-) H*+L	L%

Although delimitative tones such as these are found in a variety of intonational systems, the specific tones that each system employs (and which syllables the tones are associated to) will vary across languages. In Japanese there is an additional phenomenon that influences the tonal choice: accentual phrase-initial syllables which are either (i) heavy (i.e. two morae) and sonorant, or (ii) accented, display a rise starting from a higher Fo level than phrases starting with unaccented light (i.e. single mora) syllables. This complex difference in syllable weight affecting the Fo contour is encoded in a J_ToBI transcription by using %wL or wL% boundary tones. The %wL is marked at the beginning of post-pausal phrases, while the wL% is used at the right edge of phrases in cases where the next phrase begins with a heavy syllable or initial accent. Other languages have such language-specific phenomena as well, such as the influence of accentual phrase-initial consonant laryngeal features on the Fo contour in Korean (Jun 1993).

The tonal transcription in Figure 7.2 shows the delimitative accentual phrase tones. The utterance-initial phrase *sa'Nkaku no* is marked with a %wL

[9] Note that the H- phrase tone is labelled on all unaccented phrases, and on accented phrases only where the H- is distinguishable from the high of the lexical accent.

preceding and wL% following, due to the heavy accented initial syllable /sa'N/ and the following accented syllable /ya'/. The phrase *ya'ne no* is also followed by a wL%, due to the heavy syllable /maN/ following. Both phrases *maNnaka ni* and *okima'su* are labelled with L% at their right edge, since they are not followed by such a syllable. The final phrase *okima'su* begins with a %L, since it is post-pausal and starts with an (unaccented) light syllable.

The only phrase in this utterance that is marked with the H- phrase tone is the unaccented *maNnaka ni*, which shows a clear Fo peak around the second mora. Had the peak been delayed (as in the H*+L cases described above), the < would have been used to mark the late Fo event. It is often the case that the peak of the accentual phrase-initial H- rise is delayed to the third mora of the phrase, or even later. At present, it is unclear which factors influence this H- peak placement, though in some cases it appears that information status or speech rate may play a role: the peak is more likely to be delayed or undershot in old information, or in faster rates. This is still a very exciting open research question, which hopefully will be systematically investigated as J_ToBI-labelled databases become increasingly available.

7.3.3. *Intonation phrase tones*

The higher-level intonation phrase in Japanese displays tonal markings as well. As mentioned in Section 7.2.2, rising or rise-fall *boundary pitch movements* ('BPMs') often occur at the right edge of intonation phrases. The H% and HL% boundary tone labels are used to mark these BPMs, respectively.

The HL% is a boundary tone used to mark the rise-fall boundary pitch movement often found in the casual speech of younger speakers. Utterances containing this BPM type are often perceived as sounding 'explanatory' (Venditti *et al.* 1998). The H% boundary tone in the J_ToBI scheme described in the 1995 Guidelines is used for any rising BPM, regardless of Fo height, alignment, or meaning distinctions. However, the nature of H% rises in Japanese can be quite diverse (see e.g. Kawakami 1963/1995). For example, consider the two utterances in Figure 7.3: both are identical in segmental make-up (/*hontô ni na'ra no na no*/), and both consist of 2 APs grouped into one IP, with final BPM rise. As such, they have identical J_ToBI transcriptions (%wL H- wL% H*+L L% H%).

The utterances differ primarily in the height to which the Fo rises at the end of the phrase, and in the time-course of this rise. This difference results in a meaning distinction: the high-rising H% boundary tone (left) cues a question interpretation, while the mid-rising H% (right) cues an insisting

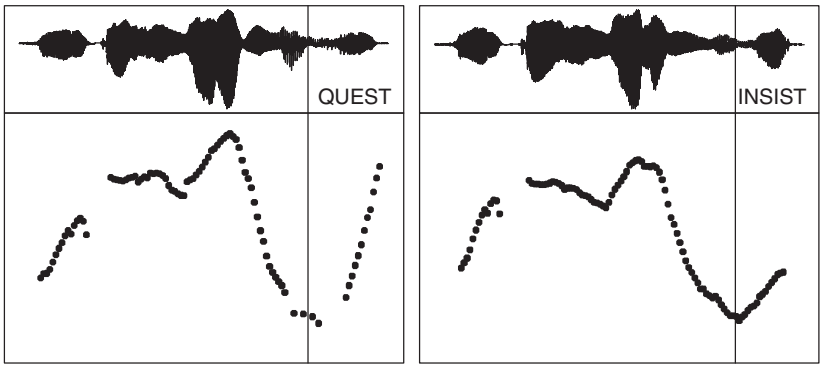

FIGURE 7.3 Waveforms and Fo contours of two productions of *hontô ni na'ra no na no*: really Nara-GEN-COP-QUEST, both uttered by the same speaker with the same tune. The left panel has a question interpretation 'Is it really the one from Nara?', while the right panel has an insisting interpretation 'It's really the one from Nara!'. The x-axis shows the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contours. Both contours are plotted on the same Fo scale, and the vertical bars mark the onset of the final mora *no* in each case.

interpretation. Venditti *et al.* (1998) have examined a number of rising BPM types in perception and production studies, and have concluded that the various BPMs in Tokyo Japanese not only cue statistically significant differences in meaning, but can be differentiated by Fo height, rise shape, and timing characteristics as well.

Figure 7.4 shows the shapes of the 5 different BPMs examined in Venditti *et al.* (1998). The figure plots multiple repetitions of raw Fo contours of the phrases *Na'oya ni* 'to Naoya' (left) and *Manami ni* 'to Manami' (right), uttered by a single speaker at a uniform speech rate. The rows show five different BPM types. The contoured lines trace the Fo values of each frame from the start of the phrase to the end of the rise (or the end of the fall in the explanatory type (row 5)). The solid vertical line marks the onset of the final mora *ni* (all contours are time-aligned by this point), and the dashed horizontal line marks a fixed arbitrary Fo reference height. Venditti *et al.* found that rises cueing a question interpretation (rows 1 and 2) are more 'scooped' (concave) and often rise to a higher Fo value than prominence-lending rises or insisting rises (rows 3 and 4). In addition, the timing of rises is different: the rise starts well within the vowel /i/ in *ni* in question BPMs (with the incredulity rise starting latest), while in other BPM types the rise starts at the onset of the final mora of the phrase (*ni*).

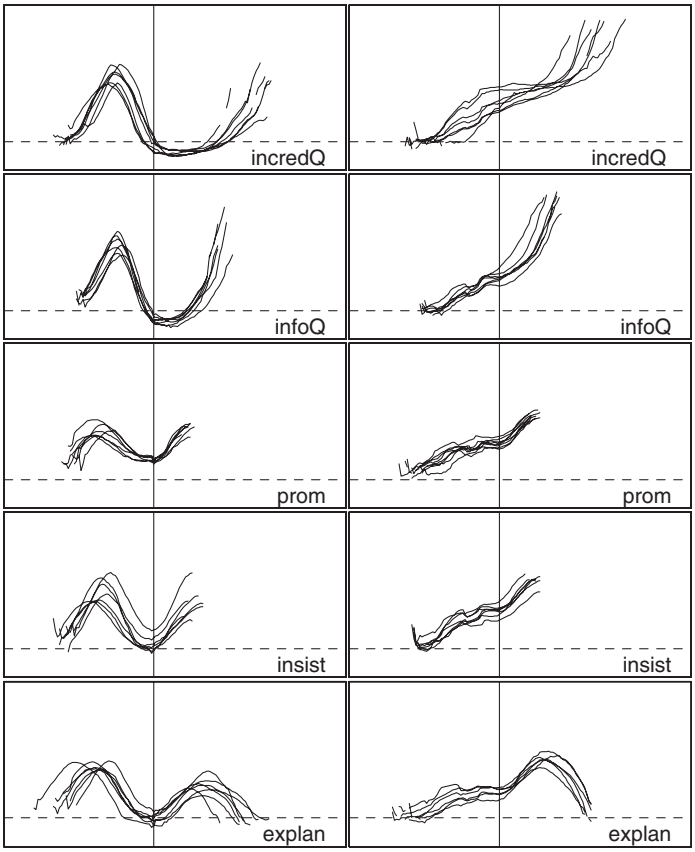

FIGURE 7.4 Fo contours of five boundary pitch movements: incredulity question (row 1), information question (row 2), prominence-lending rise (row 3), insisting rise (row 4), and the explanatory rise-fall movement (row 5). All phrases were uttered by a single speaker at a uniform speech rate on the phrases *Na'oya ni* 'to Naoya' (left) and *Manami ni* 'to Manami' (right). The x-axis shows the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contours. All panels are plotted with the same Fo and time scale. (Taken from Venditti *et al.* 1998.)

Under the J_ToBI system described in the 1995 Guidelines, all of the rising utterances (the first four rows) would be transcribed with an H% boundary tone at the right phrase edge. The accented phrase *Na'oya ni* would be transcribed as %wL H*+L L% H%, and the unaccented phrase *Manami ni* would be %L H- L% H%. However, each rise type has been shown to cue a categorically distinct meaning, and the question rises have a different Fo

shape than the other two rises; both of these facts suggest that the rises should somehow be distinctly represented in the transcription. Previous studies have shown that differences in pitch range can provide systematic cues to question interpretation in Korean (Jun and Oh 1996) and incredulity vs. uncertainty readings of the L*+H L- H% contour in English (Hirschberg and Ward 1992). In these cases, the phonological tonal transcription is identical in the two interpretations; the only difference is the overall range of the phrase. However, in the case of Japanese BPMs, not only is the pitch range different, but the timing (the alignment of the Fo rise with the segments) is distinct as well. This categorical difference in timing could be encoded in the tonal transcription by introducing an additional LH% boundary tone: in the left (accented) panel of Figure 7.4 both question types show a low region in the final mora preceding the rise (LH%), whereas the prominence-lending and insisting rise types start to rise right at the final mora onset (H%).[10] It is plausible that the low portion of the LH% boundary tone is present in the unaccented question BPMs (right panel) as well, albeit severely undershot. In such a revised system, the new inventory of boundary tones would be as follows:

H%	prominence-lending rise, insisting rise
LH%	incredulity and information question rises
HL%	explanatory rise-fall boundary movement

The difference between rises within each tonal category would then be attributed to differences in pitch range, voice quality, and the like, which do not come into play in a J_ToBI tonal transcription. Increasingly available spontaneous speech databases will be an invaluable resource in order to systematically investigate the acoustic properties of these BPMs, and also to determine their distribution function in connected discourse.

7.3.4. *Marking disjuncture*

Break indices ('BI') are one of the most important parts of a Japanese ToBI transcription, yet for some labellers these may be the most difficult to judge. Break indices are labels indicating the degree of prosodic association between adjacent words or phrases in an utterance. As such, they are primarily subjective values—measures of *perceived* disjuncture between adjacent

[10] The explanatory rise-fall BPM also starts its rise right at the onset of the final mora, which is consistent with the use of the HL% label. In this BPM, there is a marked lengthening of the final vowel (as in questions), which carries both high and low tones.

TABLE 7.1 Break index levels distinguished by the Japanese ToBI scheme

0 strong cohesion	Typical of fast speech or AP-medial lenition processes (e.g. lenition of a voiced velar stop to an approximant).
1 no higher-level juncture	Typical of the majority of AP-medial word boundaries.
2 medium degree of disjuncture	Typically corresponds to the tonally-defined *accentual phrase* (AP).
3 strong degree of disjuncture	Typically corresponds to the tonally-defined *intonation phrase* (IP).

words—and should therefore be labelled only after careful consideration of the sound record. There are various perceptual cues to disjuncture, including pausing, segmental lengthening, Fo lowering or resetting, creaky voice quality, etc. Listeners certainly can attend to all of these cues when parsing the stream of incoming speech.

The J_ToBI system currently distinguishes four degrees of disjuncture (on a scale from 0 (weak) to 3 (strong)) in the prosodic structure of Japanese.[11] All junctures between words in an utterance are assigned one of these break index values. The levels are summarized in Table 7.1, in order of increasing sense of disjuncture.

Figure 7.2 gives break index labels for the utterance ≪sankaku≫ (see the 3rd tier from the top in the label window). Break index levels 2 and 3 are arguably the most essential, since they show the higher-level prosodic phrasing of the utterance. A medium sense of disjuncture between adjacent words (BI 2) most often corresponds to the tonally-defined accentual phrase boundary. Likewise, a strong sense of disjuncture (BI 3) often corresponds to the tonally-defined intonation phrase boundary. However, there are a fair amount of mismatches between disjuncture and tonally-defined prosodic units, in both read and spontaneous speech. We will discuss these cases in Section 7.3.5.

For the most part, the break index levels and the tonally-defined phrases do match up. This is not a coincidence. As mentioned above, there are many

[11] J_ToBI labelling conducted at ATR in Japan also uses a level 4 break index, which represents an intonation phrase boundary occurring utterance-finally, which has a stronger sense of finality/completeness than do utterance-medial IP boundaries. However, the system described in the 1995 Guidelines and in this chapter does not include this additional level, but rather delegates this phenomenon to the finality tier (see Section 7.3.8).

perceptual cues to disjuncture, Fo movements being one of them. Unlike lexical accent, phrasing in Japanese allows for some degree of variability, and the prosodic structure that a speaker produces in a given utterance depends on an interaction of a number of linguistic factors, as outlined in Section 7.2.2. One way that speakers cue this prosodic parse (or 'chunking') of an utterance is by tonal movements: words are grouped into accentual phrases characterized by the delimitative tones, and APs are grouped into intonation phrases characterized by a certain pitch range and boundary tones. The initial rise of the accentual phrase cues the start of a new unit, and the pitch range reset at the start of an intonation phrase cues the beginning of an even larger unit. That is, it is the Fo rise itself that provides a major cue to the chunking of an utterance. Therefore, the close relationship between the perceived degree of disjuncture and the tonally-defined prosodic units is not considered circularity in the system, but rather it is a necessary result of Fo rising movements being one of the cues to disjuncture between words.

Another misconception is that labellers' judgements of BI 3 in Japanese is determined solely by the placement of pauses. Although it is true that pausing is often accompanied by the percept of a large degree of disjuncture, this is neither a necessary nor sufficient condition for marking BI 3. For example, there are numerous cases such as that shown in Figure 7.2, in which labellers judge a BI 3 between two words (here, *ya'ne no* and *maNnaka*) where no pause intervenes. As mentioned above, it is likely that the large Fo rise on *maNnaka* (or some other acoustic cues like pre-boundary segmental lengthening, etc.) results in the percept of large disjuncture between the two words. Likewise, there are many cases in spontaneous speech in which a pause is present, but no large disjuncture is perceived. These are cases of hesitations or disfluencies, and are discussed in detail in Section 7.3.6 and Figure 7.6 below.

7.3.5. *Mismatch between tones and perceived juncture*

The previous section described the levels of prosodic association between adjacent words currently recognized in the J_ToBI system. In most cases, break indices 2 and 3 correspond to accentual and intonation phrase boundaries, respectively. However, in some cases there is not such a clear mapping. There are cases in which the perceived degree of disjuncture is appropriate for an accentual phrase break, but there are clear tonal markings of an intonation phrase boundary. Likewise, the degree of disjuncture may seem large, yet the following AP appears to be in a downstepping pattern,

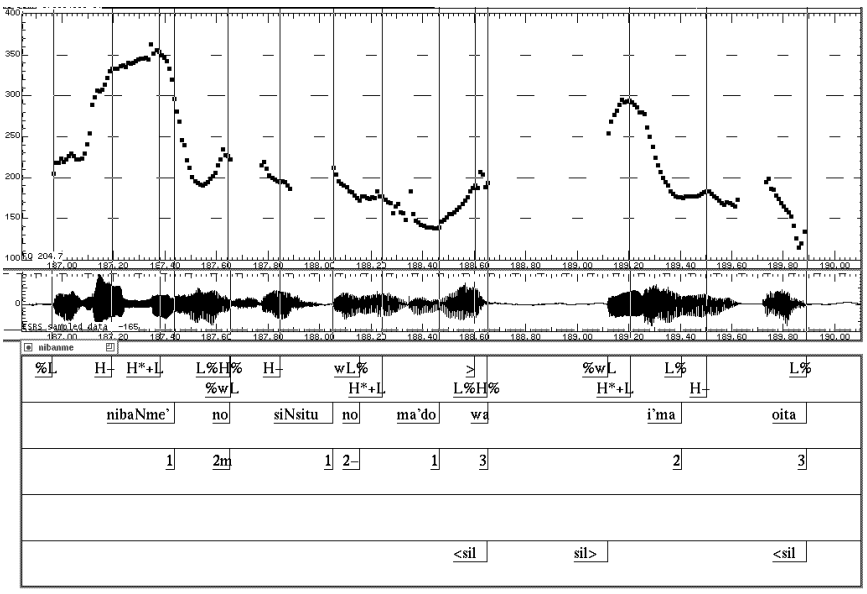

FIGURE 7.5 Sample J_ToBI transcription of the first part of the utterance ≪nibanme≫: second-GEN bedroom-GEN window-TOP now put 'I will put the second bedroom window below the first window which I just laid down'. (Taken from Venditti 1995.)

showing no signs of an intonation phrase break. Figures 7.5 and 7.2 show J_ToBI transcriptions of such cases, respectively.

In Figure 7.5, there is a boundary pitch movement (here, a H% prominence-lending rise) present on the final mora of the first phrase *nibaNme' no* 'second-GEN', suggesting an intonation phrase boundary, but there is no sense of a large disjuncture between this phrase and the following word *siNsitu* 'bedroom'. In fact, the downstepping of *siNsitu* due to the accent in *nibaNme'* suggests that there is no intonation phrase boundary intervening. Figure 7.2 shows another case of mismatch, in which there is a strong break (with pause) after the phrase *maNnaka ni* 'middle-LOC', though the pitch range on the final verb *okima'su* 'put' suggests that there is no intonation phrase break between the phrases. In such cases of mismatch, the break index value is labelled according to the perceived degree of disjuncture, and the accompanying diacritic 'm' is used. Thus, the BI labels in these two examples would be 2m and 3m, respectively.

At present, there are too few data available to conclusively determine what causes such mismatches. In the case of 2m, it is common to observe

utterance-medial BPMs in both read and spontaneous speech (e.g. Kawakami 1963/1995; Muranaka and Hara 1994; Nagahara and Iwasaki 1994), especially the prominence-lending rises, and these need not have a pause following or a strong disjuncture. Such a configuration would give rise to a 2m label. As for 3m, this type of contour is often observed in sentence-final position in Tokyo Japanese, in which the verbal predicate is set off from the rest of the sentence by a large juncture preceding, and is produced in a very narrow pitch range.[12] These casual observations about the distribution of mismatches cry out for a more detailed investigation using a large J_ToBI-labelled spontaneous speech database. With such a resource, it will be possible to make better generalizations about when tones and breaks coincide, and when they do not.

7.3.6. *Disfluent junctures*

It is common in spontaneous speech for the speaker to hesitate, stop abruptly and restart, or produce other types of disfluencies. Since the aim of J_ToBI is to describe the intonation of spontaneous as well as read speech, there must be a mechanism for marking such disfluent junctures. Following English ToBI, the diacritic 'p' following a break index value is used to mark these cases. The use of this diacritic on the break index tier is a cue that the corresponding tones on the tone tier may be incomplete or ill-formed.

Figure 7.6 shows three different productions of the fragment *ima no ma'do* 'the livingroom window', uttered by the same speaker in different contexts. The first panel shows a case where there is no disfluency. There are two accentual phrases in sequence: *ima no* 'livingroom-GEN' and *ma'do o* 'window-ACC', with a wL% boundary tone intervening. This internal juncture is label with BI 2. The second and third panels show cases of disfluencies. In both panels, the speaker stops abruptly after the words *ima no*, but then continues on with the following *ma'do* as if no disfluency had occurred (without restart). The difference between the two panels is the strength of the disfluent juncture. In the second panel, there is hardly any sense of disjuncture, and the whole fragment *ima no ma'do to* constitutes a single well-formed accentual phrase in terms of the tones. Thus, the BI value 1 at the disfluency reflects the fact that this juncture falls inside a larger unit (accentual phrase), and the 'p' diacritic flags the disfluency. There is no

[12] Such a contour is strikingly similar in function to the 'finality' contour described in Section 7.3.8, except that it lacks the H% prominence-lending rise. Without the rise, the break is labelled '3m', but with a rise it would be labelled '3'. However, further analyses of more data of this type may show that these are just two variants of the same animal.

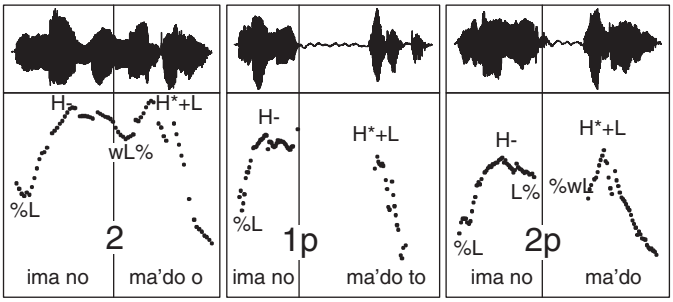

FIGURE 7.6 Waveforms and F0 contours of three productions of the fragment *ima no ma'do* 'the livingroom window', uttered by the same speaker in different contexts. The x-axis shows the time-course of the utterances; the y-axis shows the frequency (in Hz) of the F0 contours. Each contour is plotted on the same F0 scale, and the vertical lines mark the internal juncture. Break indices and tones are labelled for each phrase.

AP-final low tone after *ima no* here. In contrast, the sense of disjuncture in the third panel is stronger, with a clear L% boundary tone realized right before the disfluent region. In this case, the stronger juncture is marked by BI 2, and the 'p' flags the disfluency.

7.3.7. *Labeller uncertainty*

Japanese ToBI allows for marking of labeller uncertainty of both lexical accent realization and break index value. Accent uncertainty is most commonly found in regions of extremely reduced pitch range—for example, cases in which the pitch range of a phrase has been compressed due to the downstepping effect of a preceding accent, and/or by pragmatic or discourse factors. In these cases, the range is so compressed that the lexical accent (cued primarily by the sharp fall in F0) is hardly perceptible.[13] Such cases are often observed sentence-finally in Tokyo Japanese (see description of the 'finality' contour in Section 7.3.8). Figure 7.7 shows an example of such accent uncertainty. The sentence-final verb *okima'su* 'put' is lexically specified as

[13] But see the production study reported in Maekawa (1994) which shows that words containing 'degenerate accents' (those accents that are realized in a highly reduced pitch range and are often marked with the *? uncertainty label) differ systematically (albeit subtly) from unaccented words in their F0 slope. In addition, Maekawa (1997) presents data which show that such subtle differences in F0 slope can indeed bias listeners' accented vs. unaccented judgements in an identification (perception) task.

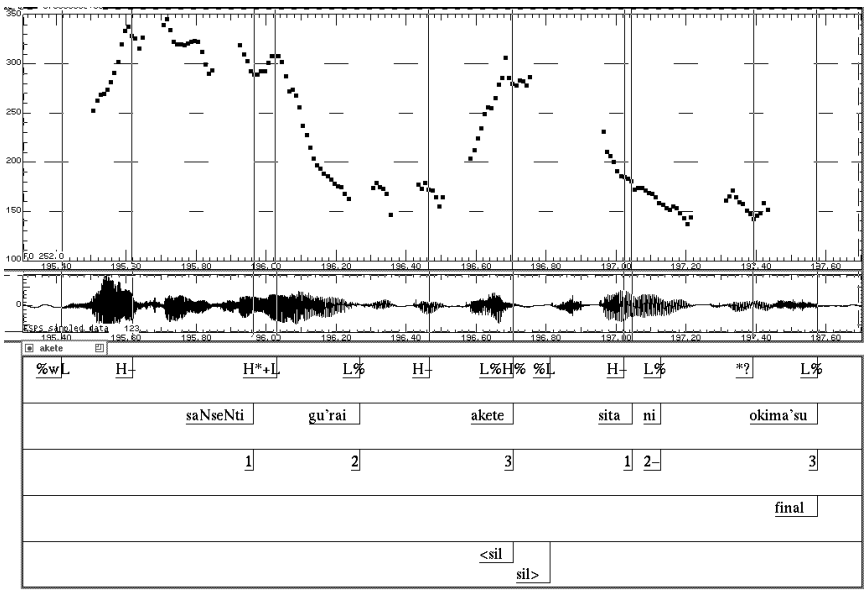

FIGURE 7.7 Sample J_ToBI transcription of the utterance ≪akete≫: 3 cm about open below-LOC put 'I will open up about a 3cm space and put it below there'. (Taken from Venditti 1995.)

accented, but the labeller is uncertain about whether the speaker indeed produced an accent in this case. The '*?' label is used to mark the uncertainty.

In regions of extremely reduced pitch range, not only is the fall of the lexical accent difficult to perceive, but also the signature initial rise of the accentual phrase can be obscured as well. That is, the labeller may find it difficult to judge whether the target word is produced as a separate accentual phrase, or dephrased together with the preceding material to form one single accentual phrase. Such cases lead to break index uncertainty judgements, as shown in Figure 7.7. The labeller is not only uncertain of the accent realization on *okima'su* 'put', but is also uncertain about whether there is an AP break (BI 2) between this and the preceding *sita ni* 'below-LOC'. Break index uncertainty is labelled by adding the diacritic '-' after the break index value, here '2-'.

As with break index judgements themselves, BI uncertainty is highly subjective. Upon careful examination of the sound and Fo records, if the labeller still cannot decide whether or not an accentual or intonation phrase break occurs, the uncertainty label may be used. Uncertainty about whether there is an accentual phrase break (i.e. a medium degree of disjuncture) is labelled by '2-', and uncertainty about larger breaks is labelled by '3-'. That is,

the break index value reflects the highest plausible level of phrasing for that particular juncture, and the '-' diacritic marks the uncertainty.

In Japanese ToBI labelling, uncertainty is a *good* thing. If all breaks were easily categorized, the labelling system would not be as meaningful. The uncertainty labels serve as flags to mark areas of interest for future research using large tagged databases, and as such should be used liberally.

7.3.8. *Finality*

The perceived finality of intonation phrases is marked on a separate finality tier. At present this is a simple binary choice between 'final' and 'not final' (no label is used in non-final cases): a phrase which is judged as 'final' will have at its right edge a strong sense of disjuncture, stronger than that of a non-final intonation phrase boundary. The notion of 'finality' is subjective by nature, and will depend on several acoustic and stylistic factors which, in combination, cue that a given phrase is final. These factors include, but are not limited to: final Fo lowering, segmental lengthening, creaky voice, amplitude lowering, long pauses, stylized 'finality' contours, etc.

The utterance ≪akete≫ shown in Figure 7.7 provides an example of finality marking. Here, the last intonation phrase *sita ni okima'su* 'put it below there' is marked with the finality label at its right edge (in the 4th tier from the top in the label window). This utterance is a good example of the so-called stylized 'finality' contour, which is often employed to signal the end of a turn or unit (common in narrative or instructional sequences). In this type of stylized contour, there is typically an H% prominence-lending rise at the edge of the phrase just before the final predicate (note the H% on *akete* here), followed by an optional pause. The final phrase (i.e. the predicate) is realized in a very reduced pitch range.[14] This particular combination of high pitch immediately preceding a very low predicate is often used in Tokyo Japanese to cue the finality of an utterance.

The 'finality' label was introduced into J_ToBI in order to mark turn or unit-final intonation phrases: the tonal pattern of the IP is the same as in other non-final cases, but it somehow has the sense that the speaker is 'done'. This label is found often in sentence-final contexts, but can also be used on medial IPs, especially in extended monologues, where the speaker composes several higher-level units of thought within one 'utterance'. Sites that choose not to include a finality tier in the J_ToBI transcription may mark the finality

[14] In addition to the H% boundary marking, a very prominent accent or unaccented phrase, followed by a predicate with extremely reduced range, can also serve to cue finality.

of intonation phrases by a break index 4 on the break index tier. This is essentially equivalent to a BI 3 marking on the break index tier and 'final' label on the finality tier. However, we recommend that a separate finality tier be used. Although the notion of 'finality' is at this point only vaguely defined, we anticipate that marking in this tier will be modified and further developed by sites whose focus is on the various degrees of finality in discourse planning and production.

7.4. DIFFERENCES FROM *JAPANESE TONE STRUCTURE*

The J_ToBI model of Japanese intonation borrows heavily from the theory of Japanese tone structure put forth by Beckman and Pierrehumbert more than a decade ago (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988). However, there has been a significant amount of research on Japanese intonation since that time, and these new findings, as well as some reanalyses of previous assumptions, have made their way into the current Japanese ToBI model. This section briefly describes the major differences between the two frameworks.

Probably the most noticeable difference between *Japanese Tone Structure* (henceforth 'JTS', Pierrehumbert and Beckman 1988) and J_ToBI is the reduction in the number of prosodic phrase levels. JTS proposed three levels above the word in the prosodic hierarchy of Japanese: the *accentual phrase* (AP), the *intermediate phrase* (iP), and the *utterance* (utt). The accentual phrase was defined exactly as it is in J_ToBI, as a low-level prosodic grouping delimited by the H- and L% tones. While this level of phrasing made it into J_ToBI virtually untouched, the JTS intermediate phrase and utterance have been merged into one level of phrasing in J_ToBI: the *intonation phrase* (IP).[15]

Arguments in JTS for the utterance level were based on the distribution of final H% boundary tones and final lowering: both said to occur utterance-finally. However, most of the data examined in the JTS experiments were short read speech utterances, which lacked the diverse phrasing patterns found in spontaneous speech. It turns out that H% and other boundary pitch movements are extremely common (even most common) utterance-medially, where they appear at the ends of the JTS intermediate phrases (e.g. Kawakami 1963/1995; Nagahara and Iwasaki 1994; Venditti *et al.* 1998). In addition, the utterance-final Fo lowering phenomenon is seen to occur in other

[15] In addition, J_ToBI has borrowed from the English ToBI system the notion of perceived degree of disjuncture (break indices), which also contributes to the definition of AP and IP levels in J_ToBI. This was not present in the JTS framework.

('utterance-medial') contexts as well. In spontaneous speech, there is no clear notion of an 'utterance', and within a given speaker's turn, there may be a number of instances (and degrees) of 'finality' cued by lowering, as mentioned above in Section 7.3.8. Without these two arguments for a separate utterance level, we are left with the JTS intermediate phrase as the highest level of prosodic organization currently motivated for Japanese.

Japanese ToBI has adopted a slightly revised definition of this intermediate phrase. Specifically, in the new system, boundary tones associate to this level of phrasing, and it is the unit marked with the optional 'final' tag in the finality tier. Since this level is no longer 'intermediate' to anything, and in order to emphasize that its definition has been revised, J_ToBI calls this level the *intonation phrase* (IP). This turns out to be a convenient renaming, since the same name is given to high-level prosodic phrases in other languages (e.g. English or Korean), which are also characterized by boundary tones.

Another difference between JTS and J_ToBI is the inventory of boundary pitch movement types. JTS recognized only the H% high-rise used in question contexts, while the 1995 J_ToBI Guidelines added to this by introducing the H% mid-rise in insisting utterances, and the HL% explanatory pitch movement used most frequently by young speakers. In addition, based on the discussion in Section 7.3.3, this inventory can be supplemented even further with the LH% scooped rise. Therefore, three distinct BPMs, H%, LH% and HL%, are currently included in the J_ToBI tonal inventory.

The Japanese ToBI system also introduces a number of labels and diacritics that are necessary to describe spontaneous speech (and which turn out to be useful for read speech as well). The mismatch label 'm' is an extremely important label in the J_ToBI system, as well is the '*?' and '-' labels to show uncertainty. The 'p' label is useful for disfluent breaks, and the various tags on the miscellaneous tier mark regions of disfluencies or other non-speech phenomena.[16] The late and early Fo event labels (< and >, respectively) are also new to the J_ToBI labelling scheme, and are essential for research on Fo timing, alignment, and pitch range variation.

7.5. AUTOMATIZATION AND LABELLER CONSISTENCY

This last section discusses more practical issues in Japanese ToBI labelling: To what extent do labellers actually agree on the J_ToBI transcription of a given

[16] The labels used in the miscellaneous tier are not described in this chapter. The reader is referred to the original 1995 Guidelines (Venditti 1995) for discussion of these, and for more details about the other labels and tiers.

utterance? Can this time-consuming labelling process be automated, even partially? Computer-guided prosodic labelling can potentially be a valuable tool for tagging large databases.

Fortunately, some parts of a J_ToBI transcription can be easily predicted from text. Since many of the tone labels are either lexically-specified, or are delimitative markings which are fixed (phonologically) in both location and type, they are entirely predictable given an accent-coded dictionary entry, as well as a record of the prosodic phrasing of the utterance. These tones include: the lexical accent H*+L, AP-initial H-, AP-final L%/wL%, and the AP-initial (post-pausal) %L/%wL (six of the nineteen J_ToBI labels).

However, the remaining thirteen of nineteen J_ToBI labels are not easily predictable from text alone. Tones which will be difficult to predict include: the intonation phrase boundary tones H%, LH% and HL%, whose location is predictable from phrasing but whose type is dependent on the meaning of the utterance; the early (>) and late (<) Fo event labels, which surely require human-labelling (or a very clever peak-picking algorithm); and the accent uncertainty '*?' label. In addition, even the predictable tone labels crucially assume that the prosodic phrasing of the utterance is known. However, break indices (and their accompanying diacritics) are not entirely predictable from text. As a first attempt, BI prediction could be facilitated by an algorithm which first assigns BI 1 as a 'default' for all junctures, then tries to determine the other BI values in a variety of ways. BI 0 prediction could be facilitated by comparing spectral slices of the uttered speech to categories of slices stored in a codebook for that speaker. BI 2 and 3 prediction could be facilitated by examining the distribution and degree Fo rising movements in the utterance, or by developing a text analysis model given the factors we know to affect phrasing (see Section 7.2.2).

Campbell (1996) describes an attempt at automatically predicting break indices by using a method whereby the phone sequence of the input text is generated, then aligned with the speech signal using text-to-speech and speech recognition tools. The system uses this alignment of the phones (and their durations) and the original Fo contour as input to a text-to-speech intonation module, in order to predict a number of candidate intonation contours and tone/break parses. The candidates are then compared with the original contour to select the optimal J_ToBI parse. Prediction of human-labelled break indices using such a method yielded promising results in Campbell's study: 68 per cent of the junctures were predicted exactly, 69 per cent were matches if the presence or absence of BI diacritics are relaxed, and the agreement rose to 90 per cent if the predicted break indices

fell within +/− 1 BI of the human-labelled value.[17] The same study examined human-human break index agreement as well. The labels of two expert labellers were compared for a subset of fifty of the 503 utterances used above (containing 282 junctures), again using only BI levels 2–4. Agreement was very high: 92 per cent of labels were an exact match, while 95 per cent matched when relaxing BI uncertainty. Campbell notes that this high degree of human-human agreement could either be due to the uniform reading style of the sentences, or a break index scale which doesn't allow for individual interpretation of juncture strengths, or both.

Another set of human-human labeller consistency data for break indices is also now available. In addition to the fifteen example transcriptions in the 1995 Guidelines, there are also ten un-transcribed practice utterances included, which labellers can use to get acquainted with the system. These utterances contain a total of 89 junctures, which were labelled by five labellers using BI 0–3.[18] Agreement was calculated across all possible pairs of transcribers for each juncture for each utterance, as has been done in English labeller agreement studies (Silverman *et al.* 1992; Pitrelli *et al.* 1994). The 89 junctures examined here do not include utterance-final junctures. The labeller agreement results are reported in Table 7.2.

Results from two subsets of the data are reported, for three separate definitions of what it is to be a break index 'match'.[19] The first row shows results from all (89) utterance-medial junctures, while the second row is a more limited set of cases (55) in which at least one labeller judged the BI value to be different from '1'. BI 1 could be considered a 'default' value (no sign of a higher-level juncture nor of lenition), and is most commonly marked between a noun and its following postposition. This can potentially be confounded by the definition of a 'word' in Japanese, and so it is not of as much interest in judging labeller agreement of higher-level junctures, which are arguably the ones absolutely essential in the characterization of Japanese

[17] These data are of 503 read utterances J_ToBI-labelled by at least one labeller, and contain 3395 human-labelled junctures. BI 4 labels are included in the tabulation, although these occur utterance-finally, and as such should be totally predictable. It is important to note that this prediction and reported agreement is based on label BI values 2–4 only (excluding BI 0 and 1), so that the high performance in prediction in the +/ − 1 BI case is probably a result of most of the data being pooled.

[18] Of the five labellers participating, two were the same expert labellers used in Campbell's (1996) study, one was the author of the Guidelines, and the remaining two had familiarity with Japanese intonation analysis but did not have much hands-on experience with the J_ToBI system. We thank Nick Campbell for providing the time and resources to make this study possible.

[19] We report only on BI agreement here, since a comparison of tonal labels warrants an extended study. Many tones in the J_ToBI system are determined by prosodic phrasing decisions, so it is not useful to compare tonal transcriptions without considering the labeller's prosodic phrase parse of the utterance. A detailed study that takes into account labellers' judgements of breaks, coupled with their tonal markings, is needed. We leave such an analysis of the current data for future work.

TABLE 7.2 Results of the labeller agreement study

data subset	exact match	relaxing diacritics	within +/− 1
all BI	66%	79%	97%
higher-level BI	46%	67%	94%

prosody. Therefore, the 2nd row in Table 7.2 is considered a more revealing estimate of labeller agreement. The first column reports percentage of exact matches, the second column shows the percentage of matches when relaxing the presence or absence of the BI diacritics '-', 'm' and 'p', and the third column shows the percentage of matches when relaxing these and allowing for agreement within +/− 1 break index value. Although results from this comparison cannot be directly compared to Campbell's results or the results for English ToBI agreement (because of differences in materials, BI inventory, tabulation, etc.), they do show that there still is a fair amount of disagreement among labellers. This could be due to a number of things, such as the complexity of the spontaneous speech testing materials themselves, labeller training, or individual differences in BI interpretation. Hopefully, future studies of labeller agreement, using an increased amount of data and number of labellers, will be able to shed more light on the nature of this disagreement.

7.6. SUMMARY AND FUTURE DIRECTIONS

This chapter has presented an overview of Japanese prosodic structure, and has described the tagging of intonational patterns associated with this structure. We have provided details of the labels used in a Japanese ToBI transcription, along with a discussion of the motivation for, and issues concerning, many of the labels. This system was compared with its predecessor, the Beckman–Pierrehumbert model of Japanese tone structure. Finally, we described efforts toward the automatization of J_ToBI, and summarized results of labeller agreement studies.

It is important to reiterate that Japanese ToBI is first and foremost a research tool, intended to be used to tag intonational patterns in databases of both read and spontaneous speech, in order to facilitate and promote continued research on Japanese prosody. The symbolic labels and annotation conventions currently used in J_ToBI are not etched in stone, but rather are open to improvement and revision, based on new insights gained from the ever-increasing amount of data and analyses available from ongoing research

on Japanese intonation. There are many exciting areas of research for which J_ToBI-labelled databases are an invaluable resource. This chapter has mentioned only a handful of such areas: linguistic factors influencing prosodic phrasing, cross-linguistic generalizations, timing and relative height relation of the lexical accent and high phrase tone, boundary pitch movement inventories and their acoustic characteristics, tone/juncture mismatches, stylized (finality) contours, systematic pitch range variation and degrees of finality in discourse, etc. There certainly are many more.

APPENDIX: SUMMARY OF J_ToBI LABELS

H*+L	*Lexical accent*: marked on lexically-accented APs within the accented mora.
<	*Late Fo event*: marked on the actual Fo peak (or start/end of Fo shoulder) when it occurs after H*+L or H-.
*?	*Accent uncertainty*: marked on the lexically-accented mora. Indicates that the labeller is unsure if the accent has been realized.
H-	*AP-initial high phrase tone*: marked on the second mora of the accentual phrase.
L% / wL%	*AP-final low boundary tone*: marked at the right edge of the accentual phrase. The wL% variant is used when the following mora is: (1) heavy and sonorant, or (2) accented.
%L / %wL	*AP-initial low boundary tone*: marked on post-pausal accentual phrases at the leftmost edge. The %wL variant is used when the following mora is: (1) heavy and sonorant, or (2) accented.
H%	*IP-final rise*: marked on the right edge of intonation phrases ending in a prominence-lending or insisting rise.
LH%	*IP-final rise*: marked on the right edge of intonation phrases ending in a question (incredulity or information) rise.
HL%	*IP-final rise-fall*: marked on the right edge of intonation phrases ending in an explanatory rise-fall BPM.
>	*Early Fo event*: marked on the actual Fo peak when it occurs before an H%, LH% or HL%.
0	*Break index*: *strong cohesion*: typical of fast speech or AP-medial lenition processes.
1	*Break index*: *no higher-level boundary*: typical of the majority of AP-medial word boundaries.

2	*Break index*: *medium disjuncture*: typically corresponds to the tonally-defined accentual phrase boundary.
3	*Break index*: *strong disjuncture*: typically corresponds to the tonally-defined intonation phrase boundary.
-	*Break index uncertainty*: marked after the BI value. Indicates that the labeller is unsure of the juncture strength.
p	*Disfluent juncture*: marked after the BI value. Indicates that the juncture is somehow disfluent.
m	*Mismatch*: marked after the BI value. Indicates a mismatch between tones and the degree of disjuncture.

REFERENCES

Beckman, M. E., and Elam, G. A. (1994), 'Guidelines for ToBI Labelling', ms Ohio State University (Version 3.0, March 1997, downloadable from: ling.ohio-state.edu/Phonetics/etobi_homepage.html).

——, and Hirschberg, J. (1994), 'The ToBI Annotation Conventions', ms Ohio State University and AT&T Bell Laboratories.

——, and Pierrehumbert, J. B. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3: 255–309.

Campbell, N. (1996), 'Autolabeling Japanese ToBI', in *Proceedings of the International Conference on Spoken Language Processing* (Philadelphia, PA), 2399–402.

—— (1997), 'The ToBI (Tones and Break Indices) System and Its Application to Japanese [in Japanese]', *Journal of the Acoustical Society of Japan*, 53/3: 223–9.

Fujisaki, H., and Hirose, K. (1984), 'Analysis of Voice Fundamental Frequency Contours for Declarative Sentences of Japanese', *Journal of the Acoustical Society of Japan*, 5/4: 233–42.

——, and Sudo, H. (1971), 'Synthesis by Rule of Prosodic Features of Connected Japanese', in *Proceedings of the International Congress on Acoustics*, 133–6.

Hata, K., and Hasegawa, Y. (1988), 'Delayed Pitch Fall Phenomenon in Japanese', in *Proceedings of the Western Conference on Formal Linguistics*, 87–100.

Hirschberg, J., and Ward, G. (1992), 'The Influence of Pitch Range, Duration, Amplitude and Spectral Features on the Interpretation of the Rise-fall-rise Intonation Contour in English', *Journal of Phonetics*, 20/2: 241–51.

Jun, S.-A. (1993), 'The Phonetics and Phonology of Korean Prosody', doctoral dissertation (Ohio State University).

——, and Fougeron, C. (1995), 'The Accentual Phrase and the Prosodic Structure of French', in *Proceedings of the International Congress of Phonetic Sciences* (Stockholm, Sweden), 722–5.

——, and Oh, M. (1996), 'A Prosodic Analysis of Three Types of Wh-phrases in Korean', *Language and Speech*, 39: 37–61.

Kawakami, S. (1963/1995), 'On Phrase-final Rising Tones [in Japanese]', in *A Collection of Papers on Japanese Accent* (Tokyo: Kyûko Shoin Publishers), 274–98.

Maekawa, K. (1994), 'Is There "Dephrasing" of the Accentual Phrase in Japanese?', in J. J. Venditti (ed.), *Ohio State University Working Papers in Linguistics*, 44: 146–65.

—— (1997), 'The Intonation of Japanese Interrogatives [in Japanese]', in Onsei Bunpô Kenkyûkai (ed.), *Grammar and Sound* (Tokyo Kuroshio Publishers), 45–53.

——, Kikuchi, H., Igarashi, Y., and Venditti, J. (2002), 'X-JToBI: an Extended J_ToBI for Spontaneous Speech', *Proceedings of the 7th International Conference on Spoken Language Processing* (Denver, CO: ICSLP), Vol. 3, 1545–8.

——, and Koiso, H. (2000), 'Design of Spontaneous Speech Corpus for Japanese', in *Proceedings of the Science and Technology Agency Priority Program Symposium on Spontaneous Speech: Corpus and Processing Technology* (Tokyo, Japan), 70–7.

Muranaka, T., and Hara, N. (1994), 'Features of Prominent Particles in Japanese Discourse: Frequency, Functions, and Acoustic Features', in *Proceedings of the International Conference on Spoken Language Processing* (Yokohama, Japan), 395–8.

Nagahara, H., and Iwasaki, S. (1994), 'Tail Pitch Movement and the Intermediate Phrase in Japanese', paper presented at the Linguistic Society of America annual meeting, Boston, MA, 6–9 January.

Pierrehumbert, J. B., and Beckman, M. E. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

——, and Hirschberg, J. (1990), 'The Meaning of Intonation Contours in the Interpretation of Discourse', in P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication* (Cambridge, MA: MIT Press), 271–311.

Pitrelli, J. F., Beckman, M. E., and Hirschberg, J. (1994), 'Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework', in *Proceedings of the International Conference on Spoken Language Processing* (Yokohama, Japan), 123–6.

Poser, W. (1984), 'The Phonetics and Phonology of Tone and Intonation in Japanese', doctoral dissertation (Massachusetts Institute of Technology, Cambridge, MA).

Silverman, K. E. A., Beckman, M., Pitrelli, J. F., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992), 'ToBI: A Standard for Labeling English Prosody', in *Proceedings of the International Conference on Spoken Language Processing* (Banff, Canada), 867–70.

Sugito, M. (1981), 'Timing Relationship Between Articulation and Fo Lowering for Word Accent [in Japanese]', *Gengo Kenkyû*, 77.

Venditti, J. J. (1995), 'Japanese ToBI Labelling Guidelines', ms Ohio State University. (Also printed in K. Ainsworth-Darnell and M. D'Imperio (eds.) *Ohio State University Working Papers in Linguistics* 50: 127–62 (1997), downloadable from: ling.ohio-tate.edu/Phonetics/J_ToBI/jtobi_homepage.html).

—— (2000), 'Discourse Structure and Attentional Salience Effects on Japanese Intonation', doctoral dissertation (Ohio State University).

VENDITTI, J. J., JUN, S.-A., and BECKMAN, M. E. (1996), 'Prosodic Cues to Syntactic and Other Linguistic Structures in Japanese, Korean, and English', in J. Morgan and K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Mahwah, NJ: Lawrence Earlbaum Associates), 287–311.

——, MAEDA, K., and VAN SANTEN, J. P. H. (1998), 'Modeling Japanese Boundary Pitch Movements for Speech Synthesis', in *Proceedings of the 3rd ESCA Workshop on Speech Synthesis* (Jenolan Caves, Australia), 317–22.

——, and VAN SANTEN, J. P. H. (2000), 'Japanese Intonation Synthesis using Superposition and Linear Alignment Models', in *Proceedings of the International Conference on Spoken Language Processing* (Beijing, China).