

Japanese ToBI Labelling Guidelines*

Jennifer J. Venditti
venditti@ling.ohio-state.edu

1 What is Japanese ToBI?

The Japanese ToBI labelling scheme (J.ToBI) is a method of prosodic transcription for Tokyo Japanese utterances which is consistent with the design principles of the ToBI system for English (see Silverman et al. 1992, Beckman and Hirschberg 1994, and Beckman and Ayers 1994). The purpose of the Japanese ToBI system is to provide a systematic phonological transcription of Japanese prosody which can be used to consistently label corpora at different sites. The J.ToBI system should be able to accurately describe the phonological events in pitch contours of spontaneous speech as well as read lab speech.

A J.ToBI transcription consists of the speech waveform and F0 contour for the utterance and a set of symbolic labels. The mandatory labels are divided into 5 separate label tiers in which labels of the same type are marked: tones, words, break indices, finality and miscellaneous. Other optional user-defined tiers can be added, as appropriate for the focus of research at each particular site. In fact, a separate tier containing the labeller's own comments and flags (e.g. for difficult areas, etc.) is recommended.

The software currently used in making a J.ToBI transcription, and that which is used in the figures in this text, is *Waves+* by Entropic Research Laboratory. However, in theory any speech analysis software may be used, as long as it has the capabilities to align and mark labels on separate tiers. It is also possible to make a J.ToBI transcription in a non-*Waves+* ASCII format (labels and timepoints), as given in Appendix B.

The following sections outline the basics of a J.ToBI transcription and give examples of labelled utterances. This is intended to be a complete and self-contained guide to the prosodic labelling of Japanese, so that anyone with a knowledge of Japanese may be able to use J.ToBI to label their own databases. The entire purpose of Japanese ToBI is to provide a standard for prosodic labelling of diverse speech data. With such a tool in hand, we will be much more prepared to examine current issues in Japanese prosody.

2 Word Tier

The word tier in J.ToBI corresponds to the "orthographic tier" in English ToBI. In this tier, words may be marked using either Japanese orthography or romanization, depending on which one the transcriber is more comfortable with, or which is most appropriate for exporting to relevant applications. The romanized transcription used in the examples in this paper is outlined in Appendix A.

In either type of transcription, lexical accent (where applicable) should be marked on the relevant mora. If labelling words in Japanese orthography, *furigana* should also be provided in order to precisely mark the location of the *akusento kaku*. Accent should be labelled according to the dictionary entry for the word, taking into consideration accent shift or compound formation rules that may apply (usually described in detail at the back of an accent dictionary). Any inconsistencies between the dictionary entry and the speaker's actual production will be marked in the tone tier (see section 3).

What constitutes a minimal "word" in Japanese is a matter of some debate, and will be discussed in more detail below in section 4.2. Filled pauses should also be marked in this tier, and should have some consistent form of transcription, which is agreed upon by labellers at each particular site.

*Copyright 1995 by Jennifer Venditti. This document is an exact reprint of the original "Japanese ToBI Labelling Guidelines" distributed in October 1995.

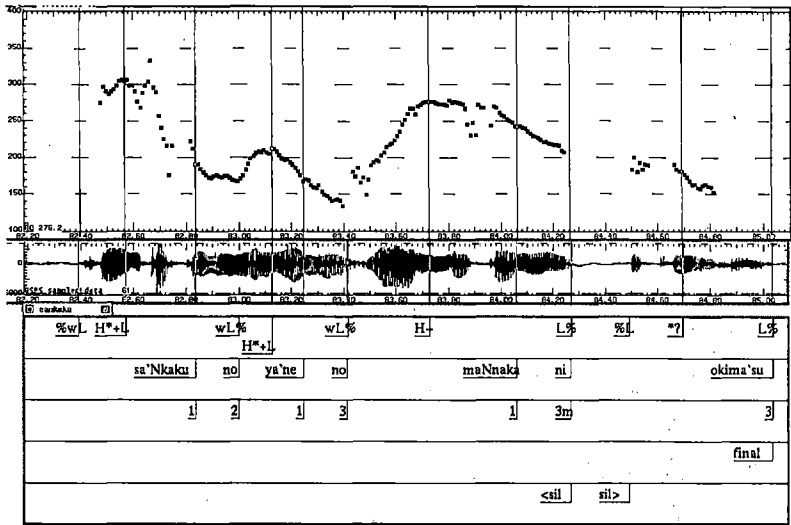


Figure 1: «sankaku» "I will place it right in the center of the triangle roof."

In the word tier, the label for each word is placed at its right edge, according to waveform or spectrogram segmentation.

3 Tone Tier

This tier contains the tones of Tokyo Japanese, as in the analysis initially proposed by Beckman and Pierrehumbert (see Beckman and Pierrehumbert 1986, Pierrehumbert and Beckman 1988), and developed further in work by Venditti (see Venditti (forthcoming)). The following subsections describe the inventory of tones for Tokyo Japanese, and give instructions on how to mark them in the J.ToBI tone tier.

3.1 Accent H*+L

This label marks the lexical accent, and should be placed within the accented mora. In many cases, the position of this H*+L accent label will coincide with the location of the actual F0 maximum. However, it is not uncommon to see the peak of the accent (and the fall) occur after the accented mora (e.g. *ososagari*, see Sugito 1981). In such cases, the H*+L label should be placed within the accented mora, and an additional "<" label should be placed at the actual peak to mark the late F0 event (see section 3.3.3 for use of the ">" label in marking an early F0 event). It is essential for the point of F0 maximum to be marked, either with the H*+L label or the < label.

The example utterance «sankaku» shows the marking of the H*+L pitch accent in cases where the peak occurs within the accented mora. (For now, concentrate on the /sa'Nkaku no ya'ne no/ portion only, and ignore the tiers other than the tone and word tiers. We will return to the other parts of the transcription below. Word-for-word English glosses for each example utterance in this guide are given in Appendix B.) In this utterance, the H*+L accent labels on /sa'Nkaku/ 'triangle' and /ya'ne/ 'roof' are placed at the F0 peak, which falls within the accented mora.

Utterance «yane» gives an example of an accent peak occurring after the accented mora. The nouns /ya'ne/ 'roof' and /ma'do/ 'window' are initially accented, but clearly the F0 maximum and accentual fall in each word occur well into the second syllable (see the dip in amplitude in the

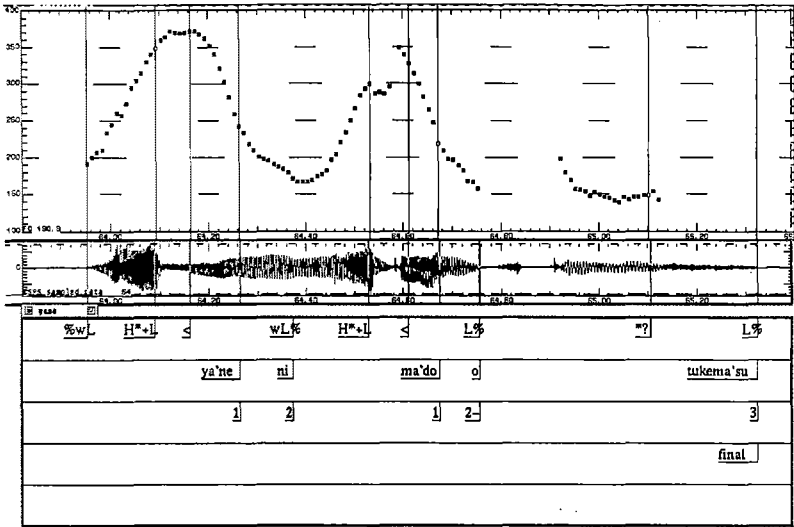


Figure 2: <yane> "I will put a window on the roof."

waveform for the /n/ of /ya'ne/ and the /d/ of /ma'do/). Here, the H*+L accent label is placed within the accented mora, and the actual F0 peak is marked with the < label (placed at the start of the precipitous fall in pitch).

Labellers should be aware that there may be movements or peaks in the F0 contour which are not genuine tonal events, but rather are segmental perturbations or mistrackings due to creaky voice, etc. For example, the /d/ of /ma'do/ in example utterance <yane> causes a considerable F0 perturbation. The English ToBI Labelling Guidelines (Beckman and Ayers 1994) provides an extensive discussion of such segmental effects on the F0 contour.

Careful labelling of the actual F0 event in J.ToBI transcribed databases will facilitate future research on the timing of F0 peaks relative to the accented mora (for example, see Sugito 1981, Hata and Hasegawa 1988). Marking the actual high F0 event will also surely help research on the relationship between discourse structure and pitch range or local prominence.

3.1.1 No Mark for Downstep

Unlike in English intonation, where the use of a downstepped accent (IH* or H+IH*, as opposed to H*) is a paradigmatic choice made by the speaker, downstep in Japanese is completely predictable from the lexical accent specification of the preceding phrase. An accentual phrase (whether itself accented or not) will be downstepped if (1) the preceding accentual phrase bears an accent, and (2) both phrases are in the same intonation phrase. (The terms "accentual phrase" and "intonation phrase" are described in more detail below in the sections on tones 3.2, 3.3.1, and sections on break indices 4.3, 4.4.) Downstepping is seen, for example, in the utterance <sankaku>, in which the second phrase /ya'ne no/ 'roof-GEN' is downstepped relative to the preceding accented phrase /sa'Nkaku no/ 'triangle-GEN'. Since the presence or absence of downstep is predictable from the information given in the word tier (i.e. the lexical accentuation of words) and the break index tier (i.e. the type of prosodic boundaries), there is no need to mark downstep in the tone tier of J.ToBI.

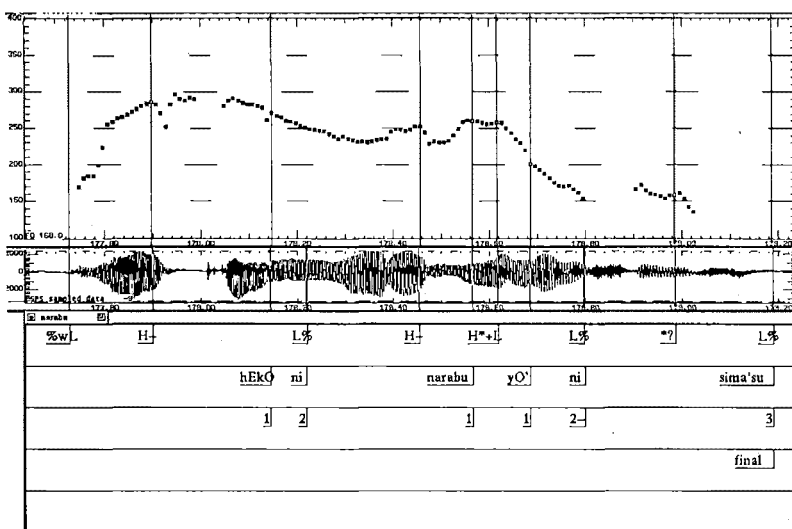


Figure 3: <narabu> "I will make it so that they line up level with each other."

3.2 Phrasal H-

This label marks the H- phrasal tone of the accentual phrase, the level of the prosodic hierarchy above the word. At this level, words may group together into prosodic units delimited by two tones: a H- phrasal tone and L% boundary tone. That is, there is a H- phrasal high near the beginning of the phrase, and a final L% boundary tone marking the end (more on the boundary tone in section 3.3.1). The H- phrasal tone is marked on each unaccented accentual phrase, and on any accented accentual phrase where the H- is distinguishable from of the high tone of the lexical accent (i.e. the shoulder for the H*+L).

The H- label should be placed within the second mora of the phrase. In Tokyo Japanese, this H- phrasal tone is associated phonologically to the second mora. The peak F0 in unaccented phrases (and at the end of the rise in accented phrases where the the H- is distinguishable) should occur around this point. The example <narabu> shows the marking of this tone in both unaccented and accented phrases. In this utterance, the H- label is marked on the second mora in both phrases. This location coincides with the peak F0 (disregarding segmental perturbations) in the first unaccented phrase /hEkO ni/ 'level-PART', and at the end of the rise in the second accented phrase /narabu yO ni/ 'line up so that'. Example <sankaku> also shows the marking of the H- label, placed on the second mora (at the high F0) of the phrase /maNnaka ni/ 'middle-LOC'.

In some utterances the second mora may not coincide with the actual high F0 of the phrase (or the end of the rise in accented phrases). In such cases, the label "<" (late F0 event) should be used to mark the actual high F0 point. This is the same labelling convention used for the H*+L accent label, i.e. the H tone label (H*+L or H-) is placed on the relevant mora, and the < label is used to mark the actual event if it occurs later. The example <kazumi> shows an utterance in which the high of the phrase is realized after the second mora (after the first word, in fact).

Marking the actual high F0 event associated with the phrasal H- will make it possible to automatically extract an estimate of the pitch range or prominence of unaccented phrases (and even accented phrases in which the H- is higher than the shoulder of the accentual H*+L) for use in future research. Thus, as with H*+L, the labeller should take care to place the H- label (or the < label) at a reliable point in the F0 contour.

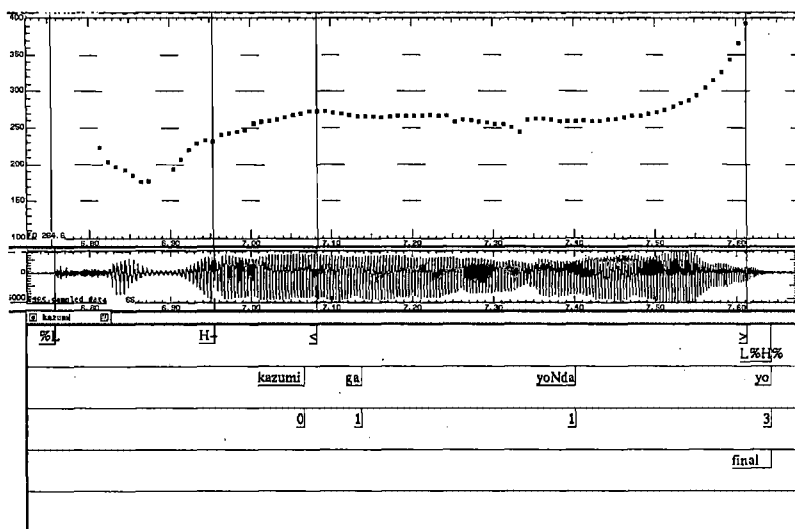


Figure 4: «kazumi» “Kazumi called her.”

Words accented on the last mora (e.g. /kami/ ‘paper’) which are intonation phrase final do not contain an accentual fall (in Tokyo Japanese), and thus are not distinguishable from unaccented words. In such cases where no fall is observed (e.g. when a pause or some disfluency follows), J.ToBI prescribes that the high F0 of these words be marked using the phrasal H-, and not with the accentual H*+L.

3.3 Boundary Tones

3.3.1 Final L% and wL%

As mentioned in the previous section, the accentual phrase in Tokyo Japanese is delimited by two tones: the H- phrasal tone and a final L% boundary tone. The L% boundary tone is marked in J.ToBI at the right edge of the accentual phrase (see also break index 2 in section 4.3). This L% tone label is aligned with the word and break index labels, and is marked at the end of every accentual phrase, even if there is an additional rise due to a H% or HL% boundary tone (see sections 3.3.3 and 3.3.4 for descriptions of these tones). Example utterance «kazumi» and others later in this paper show the marking of a L% before the rise to the H%. Example utterances «sankaku» and «yane» show the L% boundary tone on utterance final accentual phrases with no rise.

If the immediately following phrase (with no intervening pause) is initially accented, or begins with a long syllable, a wL% (“weak” low) boundary tone is used instead of the L% (“strong” low) tone. In such cases, the L% does not have enough time to be realized fully due to the immediately following high tone, and is undershot, resulting in a weak low (wL%). The utterance «sankaku» contains two wL% boundary tones marking the edges of the accentual phrases /sa’Nkaku no/ ‘triangle-GEN’ (occurring before the initially accented word /ya’ne/ ‘roof’), and /ya’ne no/ ‘roof-GEN’ (occurring before a word with a long first syllable /maNnaka/ ‘middle’). Example «yane» also has a wL% boundary tone marking the edge of the first phrase before the initially accented word /ma’do/ ‘window’.

3.3.2 Initial %L and %wL

A low boundary tone is also marked at the beginning of utterance initial or post-pausal medial phrases. As with the final low boundary tone, this initial boundary tone has "strong" and "weak" variants depending on the characteristics of the initial part of the phrase, and should be labelled in the same manner as described in section 3.3.1 above. The label should be aligned exactly with the start of phonation, according to the waveform or spectrogram.

Example <<kazumi>> shows the marking of %L utterance initially. The first mora of the initial word /kazumi/ is short and bears no accent, hence the use of the "strong" variant of the %L boundary tone. In examples <<sankaku>>, <<yane>> and <<narabu>>, the utterance initial words have either an accented first mora (<<sankaku>> and <<yane>>) or begin with a long syllable (<<narabu>>). In these cases, the "weak" variant of the initial boundary tone, %wL, is used.

This initial low boundary tone provides an anchor from which the F0 rises at the beginning of an utterance or after utterance medial silent intervals. In the case of utterance medial phrases with no preceding pause, the final L% of the preceding phrase serves this purpose (and thus an additional initial %L or %wL boundary tone is not necessary). However, there is one case in which it is necessary to mark an initial low boundary tone even though no silence precedes it. This is when an utterance medial phrase follows a H% boundary tone, with no intervening pause. In such cases an F0 fall is observed from the high boundary tone to a low point at the start of the next phrase. The marking of %wL after H% is shown in the first part of example utterance <<nibanme>>. (For now, concentrate only on the %wL boundary tone before the word /siNsitu/ 'bedroom'. The rest of this utterance will be discussed in detail below.)

3.3.3 Final H%

This label marks a final high boundary tone for an intonation phrase (see section 4.4 below for discussion of this level of the prosodic hierarchy). This boundary tone typically occurs finally in interrogatives, such as in example <<nara_quest>>, but is also often found at the end of declarative sentences such as example utterances <<kazumi>> and <<mayumi>>, and also on utterance-medial phrases such as <<nibanme>> and others in this guide.

The H% mark should be placed at the right edge of the intonation phrase, aligned exactly with the word and break index marks. In cases where this location does not correspond to the actual maximum value in the F0 contour, labellers should use the early F0 event label (">") to pinpoint the actual event. Before pauses, very often the maximum F0 value of this H% boundary tone will occur before the cessation of phonation (i.e. before the word boundary), due to mistrackings of F0 caused by the rapid decrease in amplitude. The use of the early F0 event label to mark the F0 maximum is shown in example utterances <<kazumi>> and <<mayumi>>.

When marking a H% boundary tone on the right edge of an intonation phrase, labellers should also not forget to label the L% boundary tone of the final accental phrase (the H- and L% tones delimit each accental phrase). The combined tone label L%H% (or L%HL%) is thus used for convenience. In example utterances <<nara_quest>> and <<mayumi>>, the rise to the H% from the previous L% is obvious. However, in the utterance <<kazumi>>, it may not be so apparent. In this utterance, the pitch rises to the H- phrasal tone, then rises a second time to the high boundary tone utterance finally. We know that there is a L% present since the F0 does not continuously rise after the phrasal H-, but is leveled out by the L%, then rises again to the H%.

In addition to utterance-final rises, utterance-medial H% boundary tones also occur, as seen in examples <<nibanme>> and <<pinku_mado>>. In <<pinku_mado>>, the accental phrases /pi'Nku no/ 'pink-GEN' and /ma'do o/ 'window-ACC' both form their own intonation phrase (see section 4.4 below), marked by a H% boundary tone and following pause (see Nagahara and Iwasaki 1994 for other examples of utterance-internal high boundary tones).

One last issue that deserves mention regarding the H% boundary tone is the qualitative difference in height between pitch rises. In examples <<kazumi>>, <<nara_quest>>, <<mayumi>>, and <<pinku_mado>> we observe fairly high F0 excursions to the top of the phrase's pitch range. However, there are also cases which appear to be qualitatively different from these, in which the pitch rises only part of the way. Compare the utterance in <<nara_quest>> to the so-called "insisting" declarative shown in <<nara.insist>> (see also the boundary tone on /siNsitu no ma'do wa/ 'bedroom-GEN window-TOP' in example utterance <<nibanme>>).

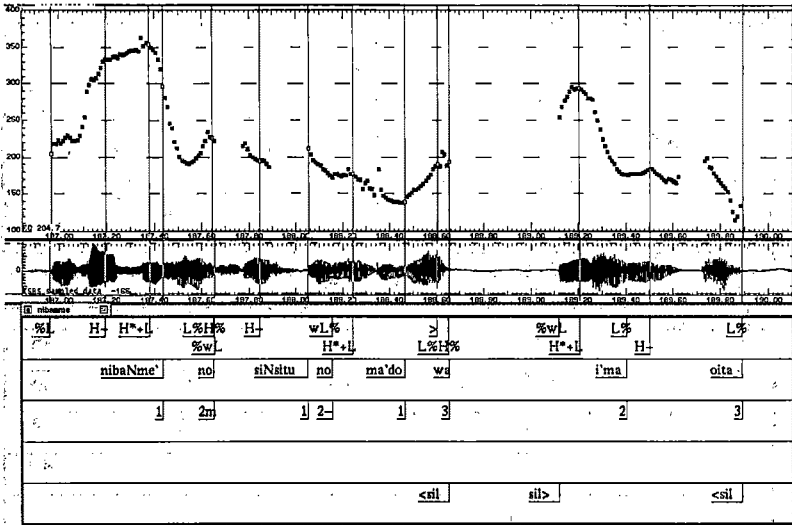


Figure 5: «nibanme» (part 1) "I will put the second bedroom window below the first window which I just layed down."

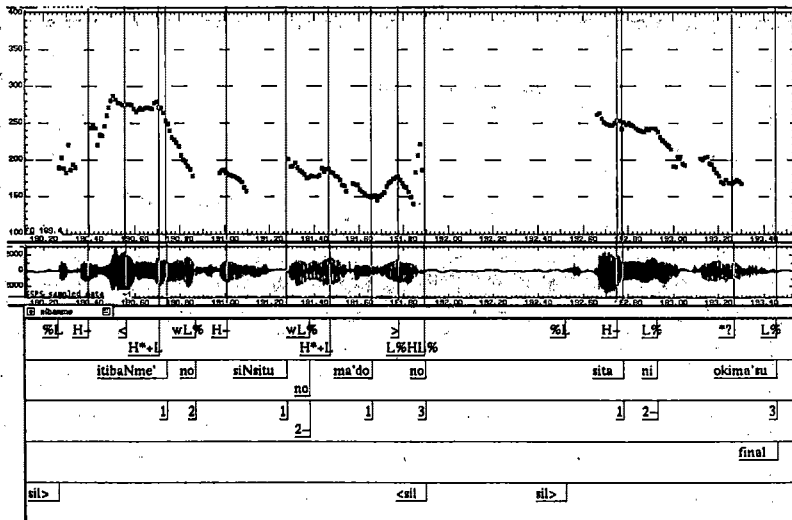


Figure 6: «nibanme» (part 2) "I will put the second bedroom window below the first window which I just layed down."

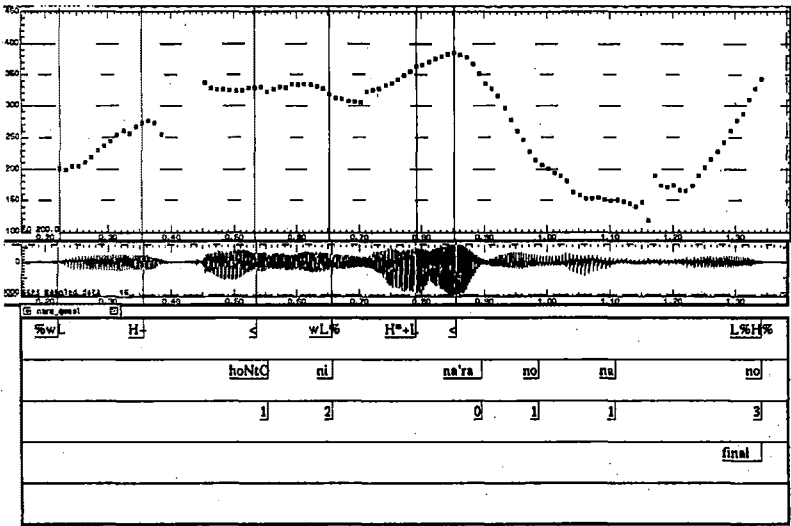


Figure 7: <nara_quest> "Is that really the one from Nara?"

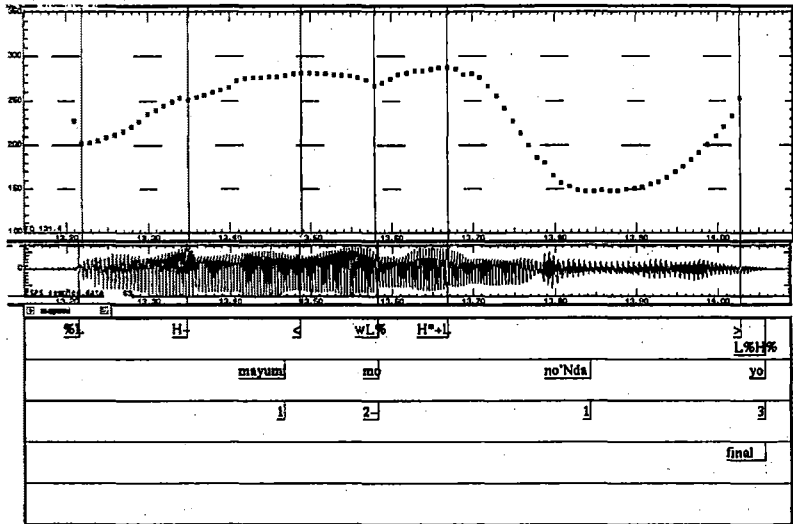


Figure 8: <mayumi> "Mayumi drank too."

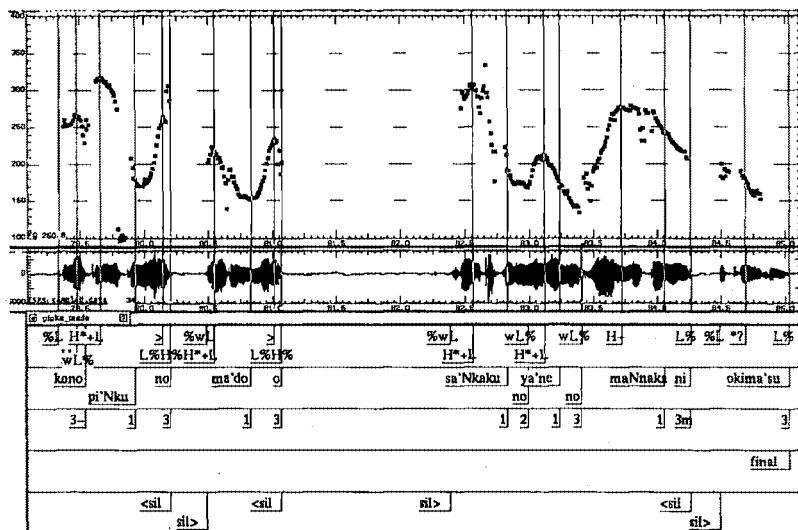


Figure 9: <<pinku_mado>> "I will place this pink window right in the center of the triangle roof."

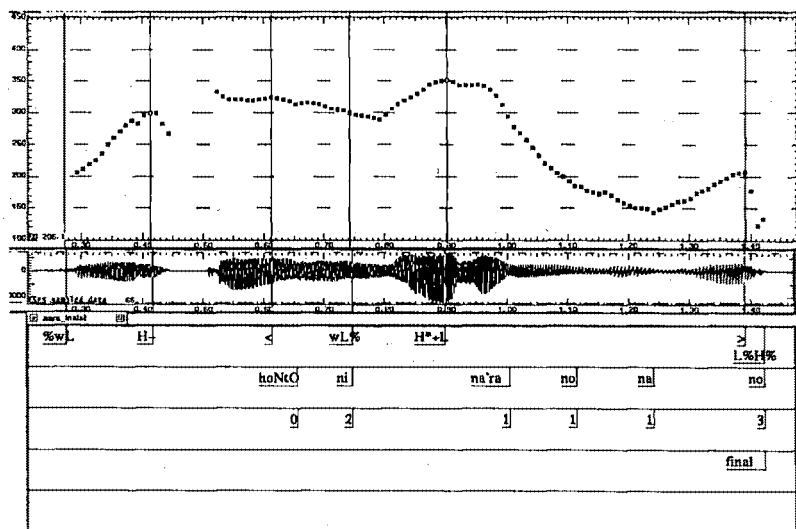


Figure 10: <<nara_insist>> "That's really the one from Nara." (insisting)

Since the tone tier in a Japanese ToBI transcription describes only shapes of contours (phonological tonal events), the amplitude of those F0 movements is not documented. There is therefore no way to distinguish between a high-rise and a mid-rise, which are both labelled as H%. For labellers interested in this difference in boundary tone height, a site-specific tier could be added to mark the distinction. This is a good example of how a basic J.ToBI transcription can be expanded to incorporate the research topics of a particular site.

3.3.4 Final HL%

This label marks the high-low boundary tone found at the end of intonation phrases in some speaking styles. Like the H% boundary tone, it should be aligned exactly with the word and break index marks.

The utterance <<nibanme>> gives an example of this boundary tone. The F0 contour at the end of the phrase /itibaNme/ no siNsi tu no ma'do no/ 'first-GEN bedroom-GEN window-GEN' has a marked rise-fall pattern (occurring within the final mora /no/). Here, a L%HL% label is aligned with the edge of this intonation phrase: the L% of the final accentual phrase, and the HL% boundary tone. Note that the ">" early F0 event label is used here to mark the F0 peak of the HL% tone.

3.4 Accent Uncertainty (*?)

As mentioned above, a given string of words may form separate accentual phrases or may group together to form one large accentual phrase. It is most common for unaccented words to group together with adjacent words, while accented words tend to form separate phrases. However, there are cases in which even accented words can join with adjacent words in the same phrase. In such cases, the left-most accented word retains its accentual fall, and the accents to the right in the phrase are totally deleted. Since the difference between a very subordinate accent and deletion of an accent can be subtle, it may be difficult for the labeller to decide whether there is indeed an accent present on a given word (and thus forms a separate phrase), or if the accent has been deleted (i.e. "totally dephrased" with the preceding words). This is especially the case when the pitch range in which the word is realized is very reduced.

In these cases where a word is lexically specified as accented, but upon consideration of both the sound and F0 records the labeller is uncertain whether the speaker indeed produced an accent, a "*?" label should be marked in the tone tier. This label simply means "I don't know if the speaker actually produced an accent", and thus should not be used in cases where the labeller feels that the accent has been totally deleted. Like the H*+L label, *? should be placed within the mora marked for lexical accent.

Utterance <<narabu>> shows an example of labeller uncertainty about the accentuation of the final verb /sima'su/ 'do'. It is common in Tokyo Japanese for final predicates to be produced in a reduced pitch range, thus making it difficult to see or hear the accentual fall. It is ambiguous whether the verb has been dephrased together with the preceding words (and thus the accent deleted), or has been produced in a separate phrase with a very narrow pitch range. Here, the *? is marked to indicate this ambiguity. (Section 4.6 will discuss in detail the break index label "2-" which accompanies this tonal uncertainty.)

In example utterance <<curtain>>, on the other hand, the labeller is not uncertain about the accentuation of the final verb /tukema'su/ 'attach', but rather feels that the accent has been totally deleted. Therefore, no marking (H*+L nor *?) appears on the verb, and the break index (BI 1) indicates that the verb belongs to the same phrase as the preceding words (see section 4.2 below).

The use of the *? uncertainty label is highly subjective, and labeller opinions about whether a word has an accent or not may vary. Since there is no right or wrong answer, labellers should not hesitate to mark *? if they are uncertain. The "degeneration" and total dephrasing of accented words in Japanese is an interesting area of research (see Maekawa 1994), and with these relevant locations flagged by *?, it will then be possible to search through large labelled databases to pull out examples for further research.

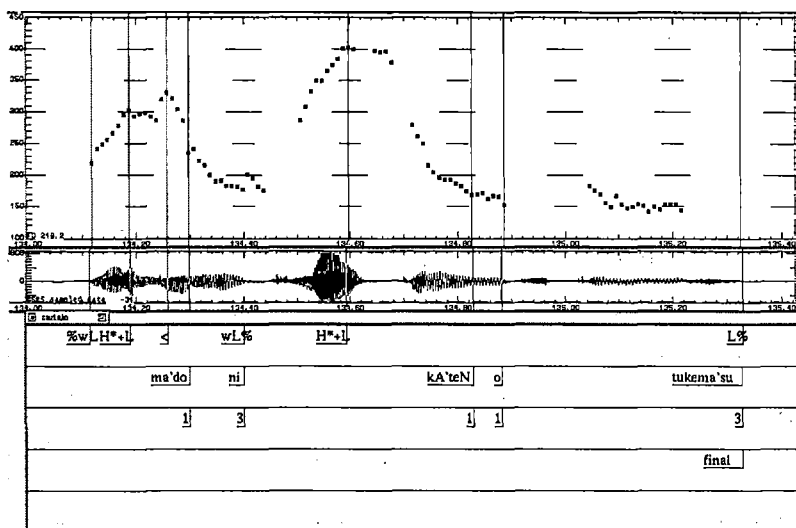


Figure 11: <<curtain>> "I will attach a curtain to the window."

4 Break Index Tier

Break indices are labels indicating degree of prosodic association between two sequential units on the word tier. They are markers which show the prosodic grouping of words at various levels. These are subjective values — measures of perceived juncture between adjacent words — and should therefore be labelled upon careful consideration of the sound record. In addition, they will typically have observable physical correlates, such as tonal markings (but see section 4.5 for examples of mismatch). J.ToBI currently distinguishes 4 degrees of disjuncture (on a scale from 0 (weak) to 3 (strong)) in the prosodic structure of Japanese.

All junctures (including filled pauses, cut-off words before restarts, etc.) should be assigned a break index value. The break labels should be aligned exactly with the word labels.

4.1 Break Index 0

This break index marks junctures which are common in fast speech processes, in which there is a very small sense of disjuncture between adjacent words. This may include phenomena such as weakening of velar stops into approximants across word boundaries, or various contracted forms such as /kore+wa/ → [korya] 'this-TOP', /yatte+simau/ → [yattya] 'do completely', or /no'Nde+iru/ → [no'Nderu] 'is drinking'.

Break index 0 is marked in the example utterance <<kazumi>> between the words /kazumi/ 'Kazumi (proper name)' and /ga/ 'NOM'. The velar stop is weakened to an approximant, giving the sense of hardly any separation between the words (there is no audible nasalization of the stop here, which is common for some speakers of the Tokyo dialect).

Example utterance <<zettai>> shows break index 0 marked at the boundaries between the verb /kuru/ 'come' and the quotative particle /to/, and between the verb /itte/ 'say' and the perfect progressive marker /ita/. The contracted forms [kurutte] and [itteta] are indicative of the small degree of disjuncture at these boundaries.

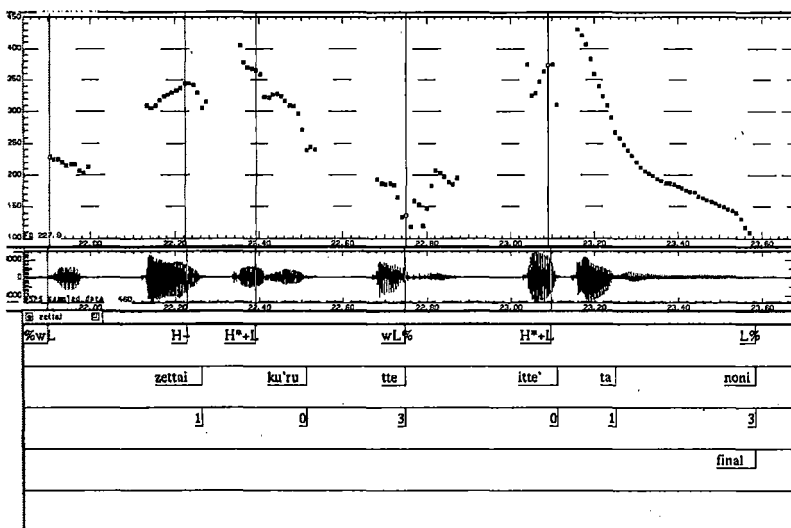


Figure 12: «zettai» "But she SAID that she would definitely come ..."

4.2 Break Index 1

Break index 1 marks the juncture between two consecutive "words", with no higher-level prosodic boundary. The question of what is a "word" in Japanese is a difficult one, especially concerning the status of postpositions as separate words. J.ToBI does not provide a definitive answer to this question, and currently the dictionary entry (including any inflectional or derivational endings) is taken as the working definition of a "word". Postpositions and sentence particles are treated as separate words in the word and break index tiers. Boundaries marked with a break index 1 have a stronger sense of disjuncture than BI 0, but a smaller disjuncture than BI 2 marking a full accentual phrase break (see section 4.3 below). All of the example utterances shown so far contain examples of the BI 1 label.

4.3 Break Index 2

This break index marks a medium degree of disjuncture between adjacent words. The boundary marked by BI 2 is stronger than that marked by BI 1, but it lacks the cues (e.g. lengthening, pauses, etc.) common to an even stronger boundary marked by BI 3 (see section 4.4).

In most cases the unit marked by a BI 2 at its boundary is characterized by the H- and L% delimiting tones of the accentual phrase, the level of the prosodic hierarchy above the word. However, this perceptually-defined unit (BI 2) and the tonally-defined unit (accentual phrase) are not *always* identical. There may occasionally be cases of mismatch between the perceived juncture and the tonal characteristics (this is described in detail in section 4.5 below).

The utterance «sankaku» contains a good example of break index 2. The words /sa'Nkaku/ 'triangle' and /no/ 'GEN' are grouped together into one tonally-defined unit (accentual phrase), and the following words /ya'ne/ 'roof' and /no/ 'GEN' into another. The perceived separation between words within each phrase (BI 1) is smaller than the separation between words belonging to adjacent phrases (BI 2). The rise to the H- of the second accentual phrase gives the sense of the beginning of a new unit, and thus we perceive the boundary between the two units as stronger than that within units, or stronger than if there was no phrase break present. In this utterance, the unit marked by BI 2 is identical to the tonally-defined accentual phrase.

Example utterances <<yane>> and <<narabu>> also show break index 2 marking a medium degree of disjuncture, which in these cases also is identical to an accentual phrase break.

4.4 Break Index 3

Break index 3 marks a strong degree of disjuncture between adjacent words, or between a word and following silent interval. This is the strongest boundary marked in the break index tier of a J.ToBI transcription.

This break index often corresponds to the boundary of the tonally-defined intonation phrase, the highest level of the prosodic hierarchy of Japanese (see Venditti (forthcoming)). The intonation phrase is the prosodic domain within which pitch range is defined and thus within which downstep occurs. At an intonation phrase boundary, the speaker resets to a paradigmatically contrastive new pitch range value for the next phrase. This also is the unit at whose edge a H% (or HL%) boundary tone may occur (see section 3.3.3 above). However, while the unit marked by a break index 3 often corresponds to the intonation phrase, the two are not always identical (see section 4.5 below). As break indices are primarily subjective evaluations of perceived disjuncture, labellers should evaluate the strength of each juncture by carefully considering the sound record, and not only by looking at the F0 contour.

Utterance <<sankaku>> also shows an example of the strong disjuncture between adjacent words marked by break index 3. As noted above, this utterance begins with two accentual phrases /sa'Nkaku no/ 'triangle-GEN' and /ya'ne no/ 'roof-GEN'. The boundary between these phrases is marked by BI 2 (medium disjuncture). The boundary between the second phrase /ya'ne no/ and the next phrase beginning with /maNnaka/ 'middle' has an even stronger sense of disjuncture, which is marked with BI 3. Tonally speaking, the first two accentual phrases of the utterance form one larger intonation phrase unit, with downstep causing the second phrase to be lowered with respect to the first (see section 3.1.1 above). The pitch range is then reset on the word /maNnaka/, which is the beginning of the next intonation phrase. The large pitch rise here between intonation phrases gives the sense that a new unit has begun, and thus there is a strong sense of disjuncture between /no/ and /maNnaka/. Example utterances <<curtain>> and <<zettai>> also show a pitch range reset associated with a strong disjuncture marked by BI 3.

A large F0 rise (e.g. pitch range reset) is one factor that can make labellers sense a strong disjuncture between words. In addition, there may be other factors involved too, including segment lengthening, F0 lowering, decreased amplitude, pauses, etc. The speech signal is full of information that can contribute to the subjective evaluation of disjuncture, and the F0 contour is only one thing.

Example utterances <<pinku_mado>> and <<nibanme>> (and others later in this guide) show the marking of break index 3 before long pauses. Inserting a pause is one way for a speaker to indicate a separation of information (i.e. a boundary) in the stream of speech. In addition, all of the examples presented here are marked with BI 3 at the end of the utterance. This indicates a strong disjuncture between the final word and the following silent interval (see section 5 below for discussion of marking the degree of finality of these pre-pausal and utterance-final boundaries).

4.5 Mismatch ("m")

All of the example utterances in sections 4.3 and 4.4 above show cases in which the units defined by break indices 2 and 3 corresponded to the prosodic units accentual phrase and intonation phrase, respectively. It was noted that while these perceptually-defined units and tonally-defined units coincide in most cases, there may be instances of mismatch between them, and thus they are not totally redundant to one another. In such cases of mismatch, the labeller should mark the break index according to her/his evaluation of the degree of disjuncture (exactly as in non-mismatch cases), but should then also add the diacritic "m" following the break index value.

Example utterance <<nibanme>> shows the marking of 2m. Here, the speaker has produced a H% boundary tone at the edge of the first phrase /nibaNme' no/ 'second-GEN'. As mentioned above in section 4.4, H% and HL% tones are found at the edges of intonation phrases, whose boundaries are most commonly characterized by a strong degree of disjuncture (i.e. BI 3). Yet the disjuncture between the /no/ of the first phrase and the following /siNsitu/ 'bedroom' is clearly not a strong one, but more like a medium disjuncture. The 2m label reflects the fact that the perceived juncture

is similar to that commonly associated with an accentual phrase (i.e. BI 2), but that there is a mismatch with the tonal pattern.

The utterance <<sankaku>> also shows an example of a mismatch. In this utterance there is a pause between the phrase /maNnaka ni/ 'middle-LOC' and the verb /okima'su/ 'put'. This gives the sense of a strong disjuncture (BI 3). However, the pitch range on the following verb seems reduced, as if the verb has been downstepped. If this is the case, the string /maNnaka ni okima'su/ would form one intonation phrase, within which downstep applies. The mismatch arises from the fact that there appears to be a strong disjuncture marked with BI 3 (which is most often associated with an intonation phrase) *within* an intonation phrase. Therefore, the break has been labelled by 3m to mark this mismatch.

With instances of mismatch flagged using the "m" diacritic, it will enable researchers to search through large labelled databases and investigate these cases further. We will then be in a better position to say how common mismatches are, and whether they are in any way related to certain speaking styles.

4.6 Break Index Uncertainty ("–")

In marking break indices, there may be cases in which the labeller is uncertain about the strength of the juncture. Specifically, it may be difficult to decide between two similar levels, such as 1 and 2. In such cases, the higher break index value should be chosen, and the diacritic "–" should be marked after it. Note that this "–" label does not mean that the strength of the juncture is somewhere in between BI 1 and BI 2, for example. It simply indicates that the labeller is uncertain — (s)he is just not sure of the boundary strength.

Utterance <<yane>> shows an example of labeller uncertainty between BI 1 and BI 2. In this utterance, it is not clear if the verb /tukema'su/ 'attach' has been totally dephrased (accent deleted) or if it is just very subordinate and thus realized in a reduced pitch range (see also section 3.4 on accent uncertainty). The break index 2– is used here to indicate this uncertainty about whether the adjacent words are grouped together into a single unit, or remain separate. Since break index 2 often corresponds to the boundary of an accentual phrase (unless there is reason to suspect mismatch), the label 2– indicates uncertainty about the presence of an accentual phrase break. (Note that when the 2– is used, the labeller should also mark the appropriate accentual phrase tones (H- and L% or wL%) in the tone tier.)

Example utterances <<narabu>> and <<nibanme>> also show the use of 2–. In each of these example utterances, the labeller is uncertain of the boundary strength right before the final verb. Since utterance-final verbs are often realized in a very reduced pitch range, it is ambiguous whether the verb has joined together with the preceding words or forms a separate unit. The break index 2– marks labeller uncertainty due to this ambiguity.

Utterance <<pin ku_mado>> shows an example of labeller uncertainty between BI 2 and BI 3. Here, the labeller is unsure of the boundary strength between /kono/ 'this' and /pi'Nku/ 'pink'. The rise in pitch on the word /pi'Nku/ indicates that it may be the start of a new intonation phrase. However, the labeller is uncertain that there is such a strong boundary, and chooses to mark this break with 3–.

As with accent uncertainty described in section 3.4 above, there is no right or wrong answer in using the "–" uncertainty diacritic. It simply allows more freedom to the labeller to express her/his commitment to the break index value assigned. Therefore, labellers should not hesitate use this label liberally. It is only by flagging these uncertain areas that we will be able to go back and take a closer look at them in future research on phrasing.

4.7 Disfluencies ("p")

It is common in spontaneous speech for the speaker to hesitate, stop abruptly and restart, or produce other similar disfluencies. Since the aim of J.ToBI is to describe spontaneous as well as read lab 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. Since this "p" label is reserved for **disfluent** junctures only, labellers should ask

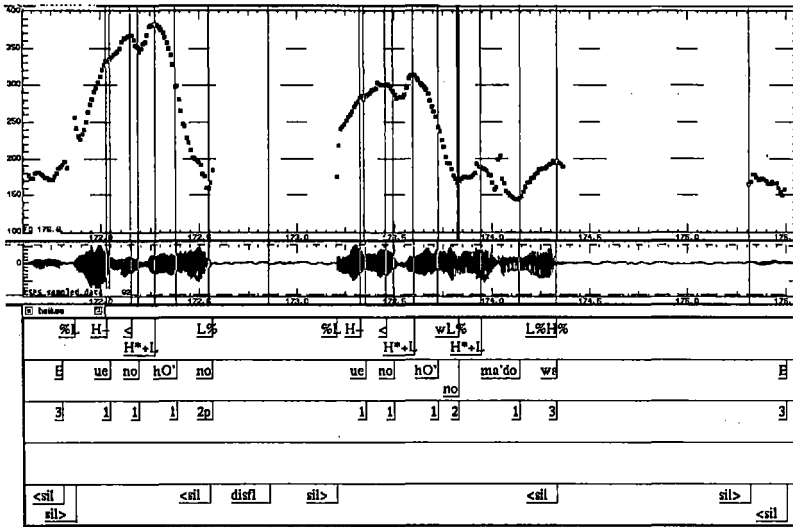


Figure 13: <<heikoo>> (Part 1) "Um, the one on top, the window on top, um, I will make it so that it lines up level with the livingroom window."

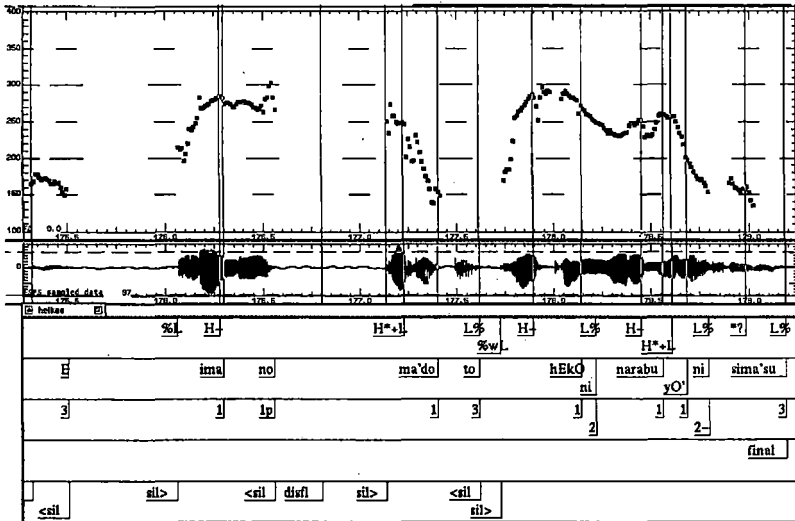


Figure 14: <<heikoo>> (Part 2) "Um, the one on top, the window on top, um, I will make it so that it lines up level with the livingroom window."

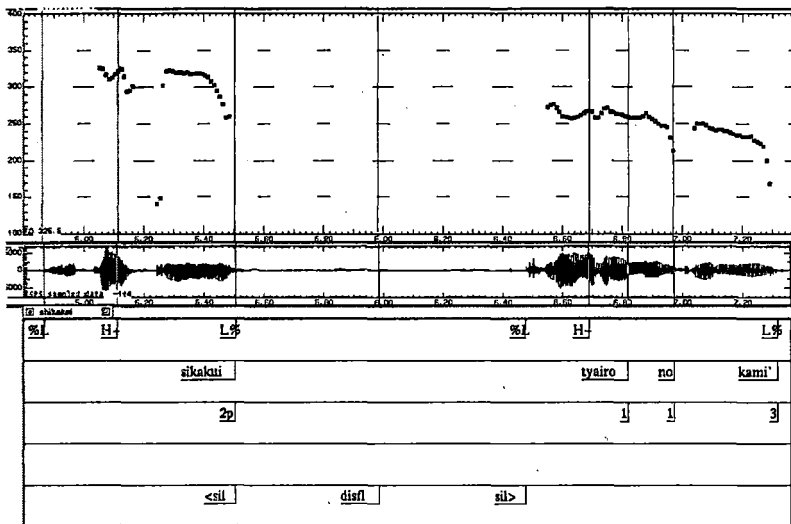


Figure 15: <<shikakui>> "The square, brown paper ..."

themselves whether the utterance might have been produced differently (more fluently) if the speaker was given a second chance to produce it.

A 1p marking on the break index tier indicates cases of abrupt cut-off in which there is no sense of the L% boundary tone which accompanies an accentual phrase juncture (BI 2). Utterance <<heikoo>> shows an example of 1p marking. Here, the speaker stops abruptly after the words /ima no/ 'livingroom-GEN' but then continues on with the following /ma'do to/ 'window-with' as if no disfluency had occurred (without restart). Tonally, the string /ima no ma'do to/ constitutes a well-formed accentual phrase (which also happens to be a single intonation phrase). The break index value 1 marked after the /no/ reflects the fact that this juncture falls inside a larger unit (accentual phrase), and the "p" diacritic flags the disfluency (see also section 6 for discussion of the "disfl" label on the miscellaneous tier).

A 2p marking on the break index tier, on the other hand, marks a disfluent juncture which is accompanied by the sense of a L% accentual phrase final boundary tone. Example utterance <<shikakui>> shows 2p marking. In this utterance, the speaker hesitates after the word /sikakui/ 'square', but then continues on after a moment with the rest of the phrase /tyairo no kami/ 'brown-GEN paper'. The downtrend of the words (not downstep here, since the words are unaccented) gives the sense that they are grouped into a single intonation phrase, and that no reset has occurred after the disfluent pause. The boundary after /sikakui/ is a medium disjuncture (BI 2), and indeed if the pause is cut out entirely the utterance sounds like a fluent intonation phrase, with no strong boundary intervening (i.e. BI 3). Therefore, the break index marked here reflects the medium disjuncture (BI 2), as well as the fact that there is a disfluency due to hesitation.

The utterance <<heikoo>> also gives an example of break index 2p. There is a disfluent break after the first phrase /ue no hO' no/ 'the one toward the top', and the speaker chooses to restart the utterance after this point. However, the break after /no/ of the first phrase does not have the sense of a strong disjuncture (BI 3), but rather, it sounds as if the speaker would continue on with the utterance, despite the disfluency. Thus, a medium disjuncture is marked (BI 2), along with the "p" diacritic.

Word-internal breaks such as in [tya-tyairo] 'br-brown' should not be indicated on the break index tier, but only by a "disfl" label in the miscellaneous tier (see section 6 below). However,

word-internal breaks followed by a restart, such as in [tya- ore'Nzi no kami] 'the br- orange paper' should be marked using 1p on the break index tier, as well as with a "disf" label in the miscellaneous tier.

5 Finality Tier

This tier provides a measure of perceived finality of each intonation phrase (break index 3). At present this is a simple binary choice between "final" and "not final". A phrase which is judged as "final" will have at its edge a strong sense of disjuncture, stronger than that of a non-final intonation phrase boundary. This percept of finality should be marked in this tier with the "final" label, which is aligned with the break index label. Those phrases which the labeller perceives as non-final should have no marking.

The notion of "finality" is subjective by nature, and will depend on several acoustic and stylistic factors which, in combination, cue that a phrase is final. The labeller should take into consideration the following phenomena (and possibly others too) when assessing the finality of a phrase:

- final FO lowering
- segmental lengthening
- creaky voice
- amplitude lowering
- long pauses
- stylized "finality" contours

The labeller should listen the phrase in question and ask her/himself the simple question: "Does the speaker sound done?". If the waveform were to be cut immediately after that break, would it sound as if the speaker had finished her/his turn (or completed an information unit)? Undoubtedly the meaning of the words in the phrase will also play some role in making this judgement, but labellers should concentrate on the sound.

The utterance «akete» provides an example of finality marking. Here it is the last intonation phrase /sita ni okima'su/ 'below-LOC put' which is marked with the finality label at its right edge. This utterance also provides 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 a H% boundary tone at the edge of the phrase just before the final predicate (note the H% on /akete/ 'open up' here), followed by an optional pause. The final phrase (i.e. predicate) is realized in a very reduced pitch range. This particular combination of tone, pause, and pitch range serves to cue the finality of an utterance.

Another example of finality marking is given in «ueshita». This utterance shows that fragments, and not just sentence-final phrases containing verbs, can also carry the percept of finality. At three points (aside from the actual end of the utterance) the speaker uses cues such as lowering, lengthening, etc. to signal finality.

Labellers should keep in mind that utterance final intonation phrases are not *always* marked with a final label. In the utterance «shikakui», the final (and only) intonation phrase is not marked with a "final" label — as the listener will notice, this phrase was cut out of a larger utterance, and thus lacks the finality which is characteristic of end-of-turn phrases.

Sites which choose not to include this a finality tier in the J.ToBI transcription may mark the 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. We anticipate that marking in this tier will be modified and further developed by sites whose focus is on the various degrees of finality and relationship with discourse structure.

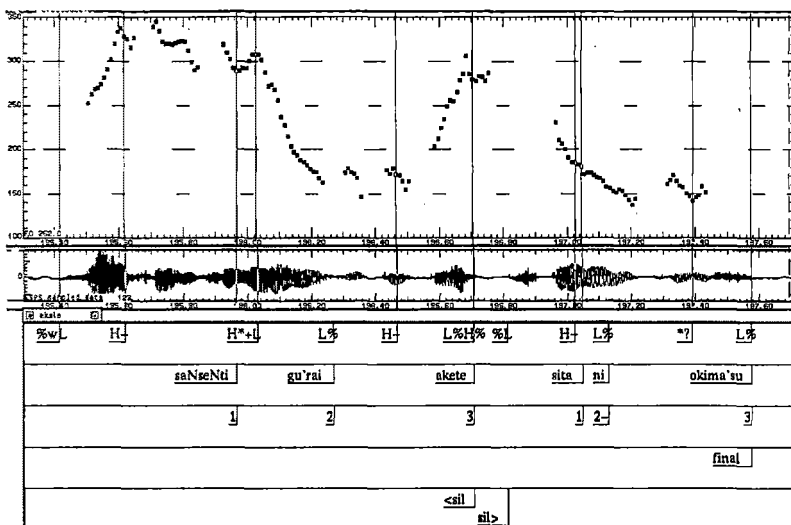


Figure 16: «akete» "I will open up about a 3cm space and put it below there."

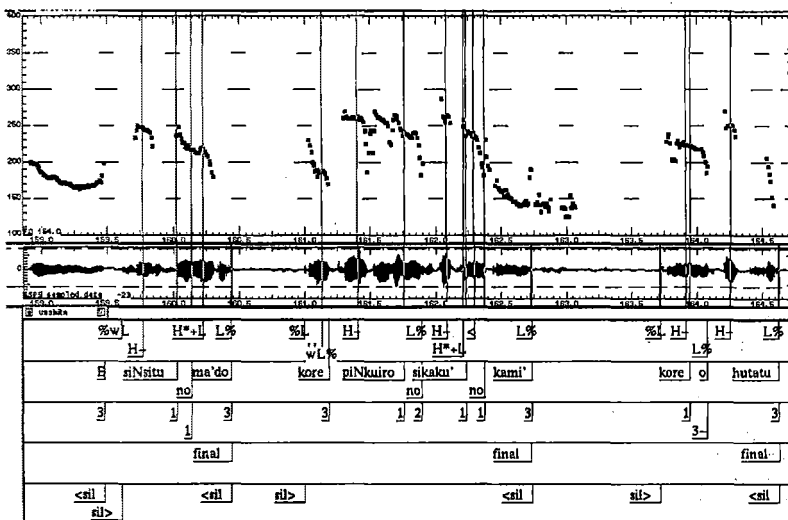


Figure 17: «ueshita» (Part 1) "Uh, the bedroom window, this pink square paper, I will put two of them, line them on top of one another."

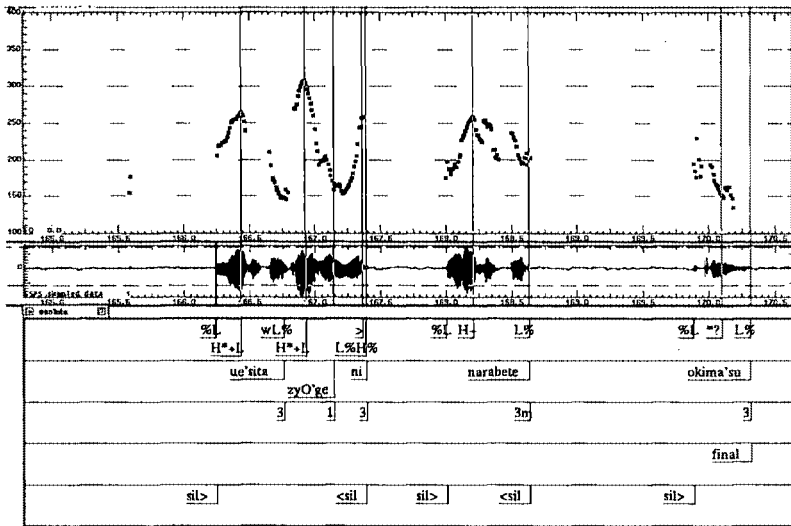


Figure 18: «ueshita» (Part 2) “Uh, the bedroom window, this pink square paper, I will put two of them, line them on top of one another.”

6 Miscellaneous Tier

This tier is reserved for other phenomena present in the speech signal which cannot be properly described by the phonological events marked in the tone and break index tiers. Such phenomena include repairs, disfluencies, silences, laughing, etc. Those phenomena which span clearly defined intervals (such as silences, laughing, etc.) should be marked by a pair of labels at their temporal beginnings and ends (e.g. laugh< laugh>). Other phenomena which are less clearly defined temporally should be marked by a single flag (e.g. “disfl”) at the approximate location. Since disfluencies often effect the corresponding tonal events and break index marks, labels in the miscellaneous tier will serve as flags to identify possible unfinished or ill-formed sequences on the other tiers.

Labeller comments may also be included in the miscellaneous tier, or additional tiers can be added to fit the needs of research at each particular site.

7 Online Data Files and Future Versions

As more speech data becomes available, the Japanese ToBI labelling guidelines may be reformulated and refined. Please check the following World Wide Web pages for revised versions:

http://ling.ohio-state.edu/Phonetics/J_ToBI/jtobi_homepage.html

<http://www.itl.atr.co.jp/ToBI/jtobi.html>

A postscript version of this guide and example utterances are also available on these WWW pages. This allows non-*Waves+* users to have access to the sound and FD files. In addition, there is also a link to the ftp site containing the *Waves+* formatted data files.

If you would like to be placed on the J_ToBI mailing list to receive notices of updates and participate in discussion of J_ToBI labelled utterances, please follow these steps:

- 1) send e-mail to: majordomo@ling.ohio-state.edu
- 2) body of the message should say only: `subscribe jtobi`
(the contents of the subject line do not matter)

Other comments or questions concerning J.ToBI are also welcome at:
venditti@ling.ohio-state.edu

Acknowledgements

The work reported in this paper was supported in part by a Title VI Foreign Language Area Studies Fellowship, and by a university relations summer fellowship from AT&T Bell Telephone Laboratories. This guide was completed while I was a visitor at ATR Interpreting Telecommunications Research Laboratories and Osaka Gakuin University. I would like to thank Mary Beckman, Nick Campbell, Toshio Hirai, Sun-Ah Jun, Y. Ohta, Marc Swerts, Junko Tsumaki, and Shigemi Yamato for their helpful discussions and insights. However, all mistakes are my own.

References

- Beckman, M. E., and G. M. Ayers. 1994. Guidelines for ToBI Labelling. Unpublished manuscript, Ohio State University. [Send e-mail to tobi@ling.ohio-state.edu for ordering information, or visit the English ToBI homepage at <http://ling.ohio-state.edu/Phonetics/etobi.homepage.html>].
- Beckman, M. E., and J. Hirschberg. 1994. The ToBI Annotation Conventions. Unpublished manuscript, Ohio State University and AT&T Bell Telephone Laboratories.
- Beckman, M. E., and J. B. Pierrehumbert. 1986. Intonational Structure in Japanese and English. *Phonology Yearbook* 3:255-309.
- Hata, K., and Y. Hasegawa. 1988. Delayed pitch fall phenomenon in Japanese. In *Proceedings of the Western Conference on Formal Linguistics*, 87-100.
- Maekawa, K. 1994. Is there 'dephrasing' of the accentual phrase in Japanese? In *Working Papers in Linguistics: Papers from the Linguistics Laboratory*, ed. J. J. Venditti, Vol. 44, 146-165. Ohio State University.
- Nagahara, H., and S. Iwasaki. 1994. Tail pitch movement and the intermediate phrase in Japanese. Paper presented at the Linguistic Society of America annual meeting, January 1994.
- Pierrehumbert, J. B., and M. E. Beckman. 1988. *Japanese Tone Structure*. MIT Press.
- Silverman, K. E. A., M. Beckman, J. F. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. 1992. TOBI: A standard for Labeling English Prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, Vol. 2, 867-870. Banff, Canada.
- Sugito, M. 1981. Timing relationship between articulation and F0 lowering for word accent. *Gengo kenkyuu* 77.
- Venditti, J. J. (forthcoming). Developments in a theory of Japanese intonation: From *Japanese Tone Structure* to J.ToBI. Unpublished manuscript, Ohio State University.

Appendix A Romanization

The following specifies how to transcribe words on the word tier with the romanization system used in the examples in this paper. It is only slightly different from the phonemic *kunreisiki* style of romanization. Accent location is marked by an apostrophe (') after the accented mora.

CONSONANTS

p	/patiNko/	'pachinko'
b	/basyo/	'place'
t	/tabe'ru/	'eat'
	/ti'zu/	'map'
	/tugi'/	'next'
d	/dame'/	'useless'
k	/kagi'/	'key'
g	/ga'maN/	'perseverance'
h	/ha'ru/	'spring'
	/hito'tu/	'one'
	/huzi/	'wisteria'
s	/sake/	'sake'
	/siro'i/	'white'
z	/zabu'toN/	'sitting cushion'
	/zibuN/	'one's self'
	/tezu'kuri/	'hand-made'
r	/raineN/	'next year'
m	/makeru/	'lose'
n	/na'be/	'pot'
N	/oNseN/	'hot spring'
y	/yawaraka'i/	'soft'
w	/waka'ru/	'understand'
pp	/kappukE'ki/	'cupcake'
tt	/waratta/	'laughed'
kk	/mikka/	'three days'
dd	/be'ddo/	'bed'
ss	/massu'gu/	'straight'
	/massi'ro/	'pure white'
Nn	/koNnitiwa'/	'Hello.'
py	/happyaku'/	'eight hundred'
by	/byOiN/	'hospital'
ty	/tyairo/	'brown'
ky	/okyakusan/	'guest'
gy	/gyaku/	'opposite'
hy	/hyaku'/	'one hundred'
sy	/sya'mozi/	'rice paddle'
zy	/zyama/	'hinderance'
ry	/ryukkusa'kku/	'backpack'
my	/myO'zi/	'surname'
ny	/nyU'su/	'news'

VOWELS

a	/ka'ge/	'shadow'
i	/inu'/	'dog'
u	/ue/	'top'
e	/eho'N/	'picture book'
o	/koke'/	'moss'
A	/bAgeNsE'ru/	'bargain sale'
I	/kawaI'/	'cute'
U	/kU'ki/	'air'
E	/seNsE'/	'teacher'
O	/hOritu/	'law'

Identical vowels belonging to adjacent syllables are distinguished from long vowels by being written with two sequential letters, as in:

/baai/	'occasion'
/midoriro/	'green'
/kooni/	'imp'

Diphthongs are distinguished from otherwise identical sequences of heterosyllabic vowels by putting a hyphen between the vowels in the heterosyllabic sequence, as in:

/ko-inu/	'puppy'	(heterosyllabic)
/ko'i/	'love'	(diphthong)
/kusa-iro/	'(darkish) green'	(heterosyllabic)
/kusa'i/	'stinking'	(diphthong)
/usu-iro/	'pale (of color)'	(heterosyllabic)
/su'iro/	'water conduit'	(diphthong)

Appendix B Non-Waves+ J_ToBI transcriptions

The following are the Non-Waves+ ASCII versions of the J_ToBI transcriptions for utterances shown in this paper in order of appearance. Non-Waves+ transcriptions are divided into the following fields:

- 1 words
- 2 ^tones
- 3 \$break index
- 4 @BI timepoint
- 5 &finality
- 6 ;miscellaneous
- 7 #English gloss

```

<<sankaku>>
sa'Nkaku  ^%wL H**+L  $1  @82.836998  &  ;  #triangle
no        ^wL%         $2  @82.999528  &  ;  #GEN
ya'ne     ^H**+L      $1  @83.247598  &  ;  #roof
no        ^wL%         $3  @83.415830  &  ;  #GEN
maNnaka   ^H-         $1  @84.063096  &  ;  #middle
ni        ^L%         $3m  @84.271248  &  ;<sil  #LOC
                                ;sil> 84.496926
okima'su  ^%L *? L%    $3  @85.041123  &final ;  #put
  
```

```

<<yane>>
ya'ne     ^%wL H**+L < $1  @64.264530  &  ;  #roof
ni        ^wL%         $2  @64.373958  &  ;  #LOC
ma'do     ^H**+L <    $1  @64.676112  &  ;  #window
o         ^L%         $2-  @64.757775  &  ;  #ACC
tukema'su ^*? L%      $3  @65.326151  &final ;  #attach
  
```

```

<<narabu>>
hEk0      ^%wL H-      $1  @78.143512  &  ;  #level
ni        ^L%         $2  @78.218332  &  ;  #PART
narabu    ^H-         $1  @78.566405  &  ;  #line up
yo'       ^H**+L      $1  @78.686767  &  ;  #so that
ni        ^L%         $2-  @78.800623  &  ;  #PART
sima'su   ^*? L%      $3  @79.190986  &final ;  #do
  
```

```

<<kazumi>>
kazumi    ^%L H-      $0  @7.066480  &  ;  #(name)
ga        ^<         $1  @7.138377  &  ;  #NOM
yoNda     ^         $1  @7.402000  &  ;  #called
yo        ^> L%H%     $3  @7.643574  &final ;  #SEN_PART
  
```

```

<<nibanme>>
nibaNme'  ^%L H- H**+L  $1  @187.437953  &  ;  #second
no        ^L%H%       $2m  @187.653405  &  ;  #GEN
siNsitu   ^%wL H-    $1  @188.053084  &  ;  #bedroom
no        ^wL%       $2-  @188.159249  &  ;  #GEN
ma'do     ^H**+L     $1  @188.468375  &  ;  #window
wa        ^> L%H%     $3  @188.655725  &  ;<sil  #TOP
                                ;sil>189.122388
i'ma      ^%wL H**+L L% $2  @189.410530  &  ;  #now
oita      ^H- L%      $3  @189.896873  &  ;<sil  #put
                                ;sil>190.258653
itibaNme' ^%L H-< H**+L $1  @190.741026  &  ;  #first
no        ^wL%       $2  @190.869560  &  ;  #GEN
  
```

siNsitu	~H-	\$1	@191.276004	&	;	#bedroom
no	~wL%	\$2-	@191.376747	&	;	#GEN
ma'do	~H**L	\$1	@191.658131	&	;	#window
no	~> L%H%	\$3	@191.897829	&	<sil	#GEN
					;sil>192.524955	
sita	~%L H-	\$1	@192.781298	&	;	#below
ni	~L%	\$2-	@192.929732	&	;	#LOC
okima'su	~*? L%	\$3	@193.458247	&final	;	#put

<<nara_quest>>

hoNtD	~%wL H- <	\$1	@0.553271	&	;	#really
ni	~wL%	\$2	@0.657389	&	;	#PART
na'ra	~H**L <	\$0	@0.896246	&	;	#(pl. name)
no	~	\$1	@0.986889	&	;	#GEN
na	~	\$1	@1.109380	&	;	#COP
no	~L%H%	\$3	@1.342113	&final	;	#SEN_PART

<<mayumi>>

mayumi	~%L H-	\$1	@13.466608	&	;	#(name)
mo	~< wL%	\$2-	@13.582692	&	;	#also
no'Nda	~H**L	\$1	@13.844117	&	;	#drank
yo	~> L%H%	\$3	@14.058353	&final	;	#SEN_PART

<<pinku_mado>>

kono	~%L H- wL%	\$3-	@79.540736	&	;	#this
pi'Nku	~H**L	\$1	@79.927887	&	;	#pink
no	~> L%H%	\$3	@80.206763	&	<sil	#GEN
					;sil>80.490035	
ma'do	~%wL H**L	\$1	@80.834235	&	;	#window
o	~> L%H%	\$3	@81.072629	&	<sil	#ACC
					;sil>82.392984	
sa'Nkaku	~%wL H**L	\$1	@82.836998	&	;	#triangle
no	~wL%	\$2	@82.999528	&	;	#GEN
ya'ne	~H**L	\$1	@83.247598	&	;	#roof
no	~wL%	\$3	@83.415830	&	;	#GEN
maNnaka	~H-	\$1	@84.063096	&	;	#middle
ni	~L%	\$3m	@84.271248	&	<sil	#LOC
					;sil>84.496926	
okima'su	~%L *? L%	\$3	@85.041123	&final	;	#put

<<nara_insist>>

hoNtD	~%wL H- <	\$0	@0.656519	&	;	#really
ni	~wL%	\$2	@0.747603	&	;	#PART
na'ra	~H**L	\$1	@1.005674	&	;	#(pl. name)
no	~	\$1	@1.115734	&	;	#GEN
na	~	\$1	@1.242239	&	;	#COP
no	~> L%H%	\$3	@1.424407	&final	;	#SEN_PART

<<curtain>>

ma'do	~%wL H**L <	\$1	@134.299321	&	;	#window
ni	~wL%	\$3	@134.401580	&	;	#LOC
ka'teN	~H**L	\$1	@134.830093	&	;	#curtain
o	~	\$1	@134.883657	&	;	#ACC
tukema'su	~L%	\$3	@135.330491	&final	;	#attach

<<zettai>>

zettai	~%wL H-	\$1	@22.262873	&	;	#absolutely
--------	---------	-----	------------	---	---	-------------

ku'ru	~H**L	\$0	@22.512168	&	;	#come
tte	~wL%	\$3	@22.750384	&	;	#QUOT
itte'	~H**L	\$0	@23.114171	&	;	#said
ta	~	\$1	@23.252668	&	;	#PROGRESS
noni	~L%	\$3	@23.586909	&final	;	#PART

<<heikoo>>

E	~	\$3	@171.803163	&	;	#um
					;<sil	
					;sil>171.865027	
ue	~%L H-	\$1	@172.043665	&	;	#upper
no	~<	\$1	@172.190412	&	;	#GEN
h0'	~H**L	\$1	@172.381998	&	;	#
no	~L%	\$2p	@172.565247	&	;	#GEN
					;disfl	
					;sil>173.203382	
ue	~%L H-	\$1	@173.352157	&	;	#upper
no	~<	\$1	@173.498904	&	;	#GEN
h0'	~H**L	\$1	@173.727177	&	;	#
no	~wL%	\$2	@173.833161	&	;	#GEN
ma'do	~H**L	\$1	@174.142959	&	;	#window
wa	~L%H%	\$3	@174.334545	&	;	#TOP
					;sil>175.322155	
E	~	\$3	@175.506354	&	;	#um
					;sil>176.060629	
ima	~%L H-	\$1	@176.299048	&	;	#livingroom
no	~	\$1p	@176.559290	&	;	#GEN
					;disfl	
					;sil>177.139433	
ma'do	~H**L	\$1	@177.405076	&	;	#window
to	~L%	\$3	@177.619776	&	;	#with
					;sil>177.727658	
hEk0	~%wL H-	\$1	@178.143512	&	;	#level
ni	~L%	\$2	@178.218332	&	;	#ADV
narabu	~H-	\$1	@178.566405	&	;	#line up
y0'	~H**L	\$1	@178.686767	&	;	#so that
ni	~L%	\$2-	@178.800623	&	;	#PART
sima'su	~*? L%	\$3	@179.190986	&final	;	#do

<<shikakui>>

sikakui	~%L H- L%	\$2p	@5.502376	&	;	#square
					;disfl	
					;sil>6.477005	
tyairo	~%L H-	\$1	@6.820425	&	;	#brown
no	~	\$1	@6.976449	&	;	#GEN
kami'	~L%	\$3	@7.319700	&	;	#paper

<<akete>>

saNseNti	~%wL H-	\$1	@195.964593	&	;	#3cm
gu'rai	~H**L L%	\$2	@196.270458	&	;	#about
akete	~H- L%H%	\$3	@196.707749	&	;	#open up
					;sil>196.814808	
sita	~%L H-	\$1	@197.049458	&	;	#below
ni	~L%	\$2-	@197.130703	&	;	#LOC
okima'su	~*? L%	\$3	@197.579943	&final	;	#put

<<ueshita>>

E	~	\$3	@159.476999	&	;	#um
---	---	-----	-------------	---	---	-----

```

;sil>159.605087
siNsitu    ~%wL H-    $1    @160.025312    &    ;    #bedroom
no         ~         $1    @160.141215    &    ;    #GEN
ma'do     ~H**L L%   $3    @160.435432    &final ;<sil    #window
;sil>160.997869
kore      ~%L H- wL%  $3    @161.188805    &    ;    #this
piNkuiro ~H-        $1    @161.763865    &    ;    #pink
no        ~L%        $2    @161.897600    &    ;    #GEN
sikaku'   ~H- H**L   $1    @162.236395    &    ;    #square
no        ~<        $1    @162.379046    &    ;    #GEN
kami'     ~L%        $3    @162.740130    &final ;<sil    #paper
;sil>163.726736
kore      ~%L H-    $1    @163.956116    &    ;    #this
o         ~L%        $3-   @164.088244    &    ;    #ACC
hutatu    ~H- L%    $3    @164.633626    &final ;<sil    #two
;sil>166.243984
ue'sita   ~%L H**L wL%  $3    @166.764472    &    ;    #above/below
zy0'ge    ~H**L     $1    @167.145487    &    ;    #above/below
ni        ~> L%H%    $3    @167.393008    &    ;<sil    #PART
;sil>168.002199
narabete  ~%L H- L%   $3m   @168.636177    &    ;<sil    #line up
;sil>169.890971
okima'su  ~%L *? L%   $3    @170.320860    &final ;    #put

```

Appendix C J_ToBI Labelling Conventions

This Appendix is intended to serve as a reference to the J_ToBI labels and conventions introduced in this guide, and therefore should be consulted only after the labeller has reviewed the examples and explanations in the preceding sections.

Synopsis

The Japanese ToBI labelling scheme (J_ToBI) is a method of prosodic transcription for Tokyo Japanese utterances which is consistent with the design principles of the ToBI system for English (see Silverman et al. 1992, Beckman and Hirschberg 1994, and Beckman and Ayers 1994). The purpose of the Japanese ToBI system is to provide a systematic phonological transcription of Japanese prosody which can be used to consistently label corpora at different sites.

A J_ToBI transcription consists of the speech waveform and F0 contour for the utterance and a set of symbolic labels. The mandatory labels are divided into 5 separate label tiers in which labels of the same type are marked: tones, words, break indices, finality and miscellaneous. Other optional user-defined tiers can be added, as appropriate for the focus of research at each particular site. In fact, a separate tier containing the labeller's own comments and flags (e.g. for difficult areas, etc.) is recommended.

Word Tier

This tier contains the individual words of the utterance, transcribed using either Japanese orthography or some conventional romanization. Word labels should be marked at the right edge of each word, according to waveform or spectrogram segmentation. We currently take a minimal dictionary entry as the working definition of a "word", and as such we mark postpositions and particles as separate words. Accented words are marked with an apostrophe (') after the relevant mora.

Tone Tier

This tier contains the tones of Tokyo Japanese, as in the analysis initially proposed by Beckman and Pierrehumbert (see Beckman and Pierrehumbert 1986, Pierrehumbert and Beckman 1988), and developed further in work by Venditti (see Venditti (forthcoming)). The J_ToBI tone labels can be divided into three groups: those concerning lexical accent, accentual phrase tones, and intonation phrase boundary tones.

Lexcial Accent

H*+L This bitonal pitch accent marks the lexically specified accent of accented phrases. This label should be placed within the accented mora. In cases in which the F0 peak occurs after the accented mora (e.g. *ososagari*), an additional label "<" (late F0 event) should be placed at the actual F0 peak.

***?** This accent uncertainty label marks labeller uncertainty about whether an accentual fall is present on a word marked as accented in the lexicon. This label is commonly used in phrases with a narrow pitch range (e.g. utterance-final verbs), where it is difficult to distinguish a very subordinate accent from a totally deleted accent.

Accentual Phrase

H- This phrasal high tone is marked on the second mora in unaccented phrases, and also in accented phrases in which the H- is distinguishable from the shoulder of the H*+L accent. It is one of the two tones which delimit the accentual phrase in Japanese (see also break index 2). In cases in which the F0 peak in unaccented phrases (or the end of the rise in accented phrases) occurs after the second mora, the label "<" (late F0 event) should be used to mark the actual high F0 point.

- L%** Along with the phrasal H-, this final low boundary tone characterizes the accentual phrase. It should be placed at the right edge of each phrase, aligned with the word and break index labels. There is also a "weak" variant of this tone (wL%) used in cases where the next phrase (with no intervening pause) begins with a long syllable, or is initially accented.
- %L** This initial low boundary tone is marked at the beginning of post-pausal phrases. It provides an anchor from which the F0 rises at the beginning of utterances and after pauses. As with the final low boundary tone, this initial tone also has a "weak" variant (%wL), used in the same contexts.

Intonation Phrase

- H%** This intonation phrase final high boundary tone marks the final rise common in interrogative utterances and also in some declaratives. It should be marked at the right edge of the intonation phrase (see also break index 3), aligned exactly with the word and break index labels. In cases in which the high F0 at the end of the rise occurs before the right edge of the phrase, an additional label ">" (early F0 event) should be used to mark the actual high F0 point.
- HL%** This final high-low boundary tone marks the rise-fall contour found at the end of some intonation phrases. This label should also be placed exactly at the right edge of the phrase, and the actual F0 peak should be marked using the ">" early F0 event label.

Break Index Tier

Break indices are labels indicating degree of prosodic association between two sequential units on the word tier. They are markers which show the prosodic grouping of words at various levels. These are subjective values — measures of perceived juncture between adjacent words — and should therefore be labelled upon careful consideration of the sound record. In addition, they will typically have observable physical correlates, such as tonal markings. J.ToBI currently distinguishes 4 degrees of disjuncture (on a scale from 0 (weak) to 3 (strong)) in the prosodic structure of Japanese.

All junctures (including filled pauses, cut-off words before restarts, etc.) should be assigned a break index value. The break labels should be aligned exactly with the word labels.

- 0** This break index marks junctures which are common in fast speech processes (e.g. /kore+wa/ → [korya] 'this-TOP').
- 1** This break index marks the juncture between two consecutive "words", with no higher-level prosodic boundary.
- 2** This break index marks a medium degree of disjuncture between adjacent words. The boundary marked by BI 2 is stronger than that marked by BI 1, but it lacks the cues (e.g. lengthening, pauses, etc.) common to an even stronger boundary marked by BI 3. In most cases the unit marked by a BI 2 at its edge corresponds to the tonally-defined unit *accentual phrase* (but see below for cases of mismatch).
- 3** This break index marks a strong degree of disjuncture between adjacent words, or between a word and following silent interval. BI 3 often corresponds to the boundary of the tonally-defined *intonation phrase* (but see below for cases of mismatch).

Diacritics Attached to Break Index Values

- This diacritic marks labeller uncertainty about the degree of disjuncture between adjacent words. In such cases of uncertainty, the higher level break index should be chosen, and this diacritic should be affixed directly after it (e.g. "2-" indicates uncertainty between "1" and "2").
- m** This diacritic marks mismatch between the subjective degree of disjuncture (break index value) and the tonal characteristics (e.g. boundary tones, downstep, etc.) In such cases of mismatch, the labeller should mark the break index according to her/his evaluation of the degree of disjuncture, and should affix this diacritic directly after it. ("2m" indicates, for example, that there is a sense of medium disjuncture that typically corresponds to an accentual phrase break, but with the tonal markings of a full intonation phrase).

p This diacritic marks a disfluent juncture marked by an abrupt cut-off or lengthening due to hesitation. This diacritic should be affixed directly after the break index value. ("1p" marks an abrupt cut-off which lacks the sense of an accentual phrase final L%, and "2p" marks a disfluent juncture accompanied by this sense of L%).

Finality Tier

This tier provides a measure of perceived finality of each intonation phrase (break index 3). At present this is a simple binary choice between "final" and "not final" (additional labels will be added once we have a more complete model of discourse finality and its relation to the structuring of information or turn-taking). A phrase which is judged as "final" will have at its edge a strong sense of disjuncture, marked by F0 lowering, segmental lengthening, creaky voice, etc. This percept of finality should be marked in this tier with the "final" label, which is aligned with the break index label. Those phrases which the labeller perceives as non-final should have no marking.

Miscellaneous Tier

This tier is reserved for other phenomena present in the speech signal which cannot be properly described by the phonological events marked in the tone and break index tiers. Such phenomena include repairs, disfluencies, silences, laughing, etc. Those phenomena which span clearly defined intervals (such as silences, laughing, etc.) should be marked by a pair of labels at their temporal beginnings and ends (e.g. laugh< laugh>). Other phenomena which are less clearly defined temporally should be marked by a single flag (e.g. "disfl") at the approximate location. Labeller comments may also be included in the miscellaneous tier, or additional tiers can be added to fit the needs of research at each particular site.

Appendix D Practice Utterances

Ten unlabelled utterances are included at the end of the guide for labellers to practice transcribing using the J.ToBI system. These data files are also available online in various formats on the WWW page mentioned in section 7.

Since our goal is to develop a system for labelling Japanese prosody that can be used to quickly and consistently transcribe Japanese utterances, we would like to check whether it is possible, after going through the examples and discussions in this guide, for labellers to do just this. We would therefore like to ask for your cooperation in helping us check labeller consistency of the 10 utterances included here. We would like to ask those interested to label these few utterances, and share their labels with us (either by sending us the ASCII label files, or a faxed copy of the labelled utterances). It is only by comparing the transcriptions of many people that we will be able to get an idea of how well the J.ToBI system can be used, and in which areas the difficulties lie.

Please contact the author at: venditti@ling.ohio-state.edu for further information. We thank you for your cooperation.

List of Practice Utterances: (in order of difficulty)

<<door>
<<nondenai>
<<futatsu>
<<tugi_ni>
<<ima_double>
<<tree>
<<nokotta>
<<hachiue>
<<migi>
<<kasanete>

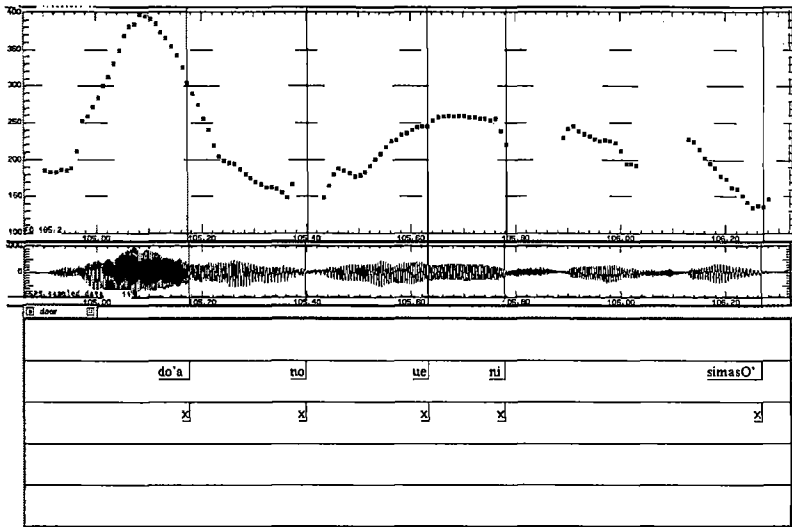


Figure 19: «door» "Let's put it above the door."

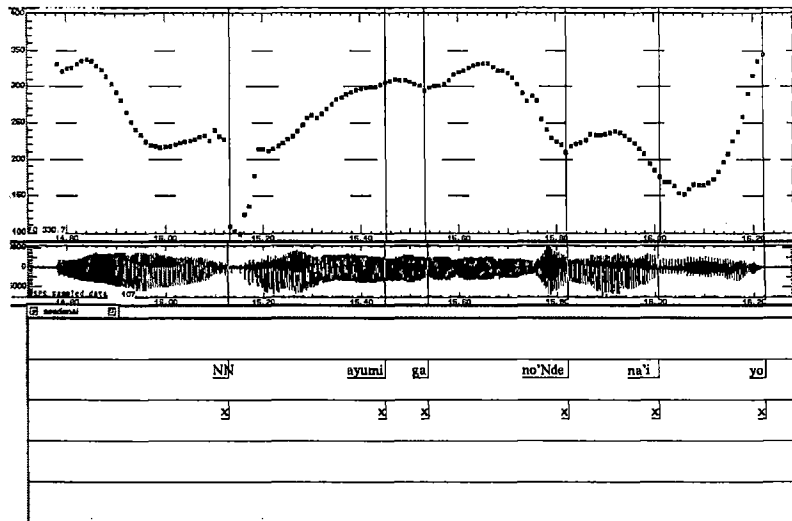


Figure 20: «nondenai» "No, Ayumi isn't drinking." [Don't worry about the tones on /NN/.]

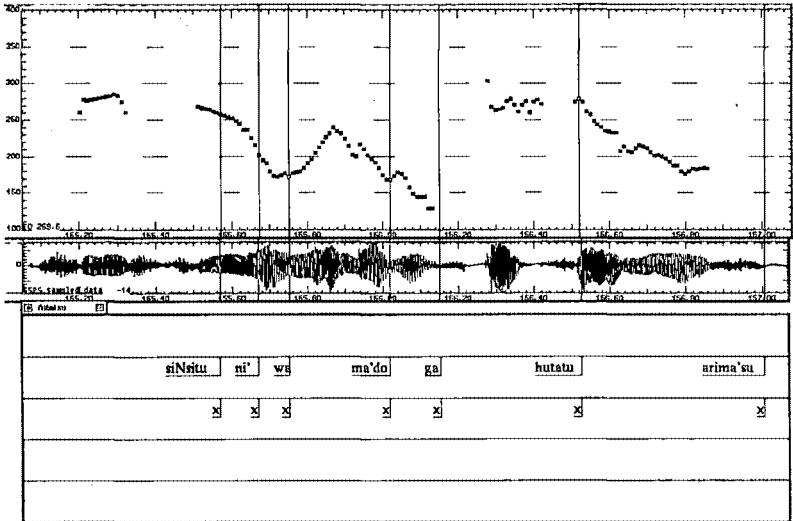


Figure 21: «futatsu» "There are two windows in the bedroom."

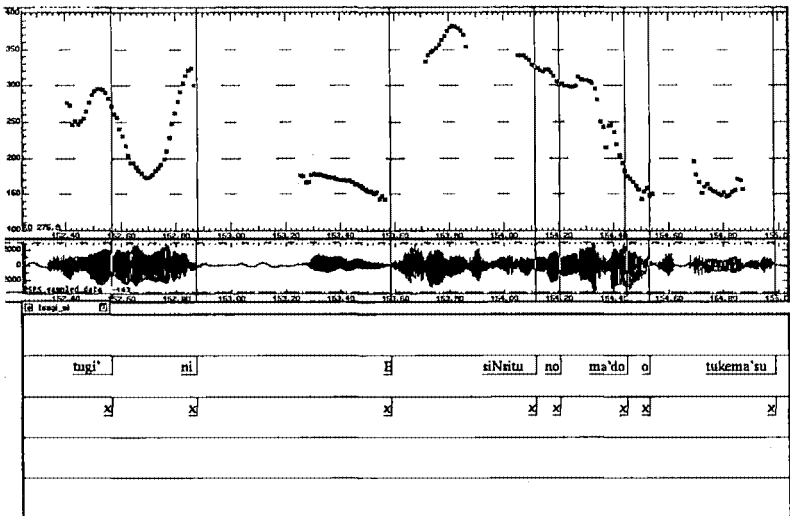


Figure 22: «tugi.ni» "Next, um, I'll attach the bedroom window."

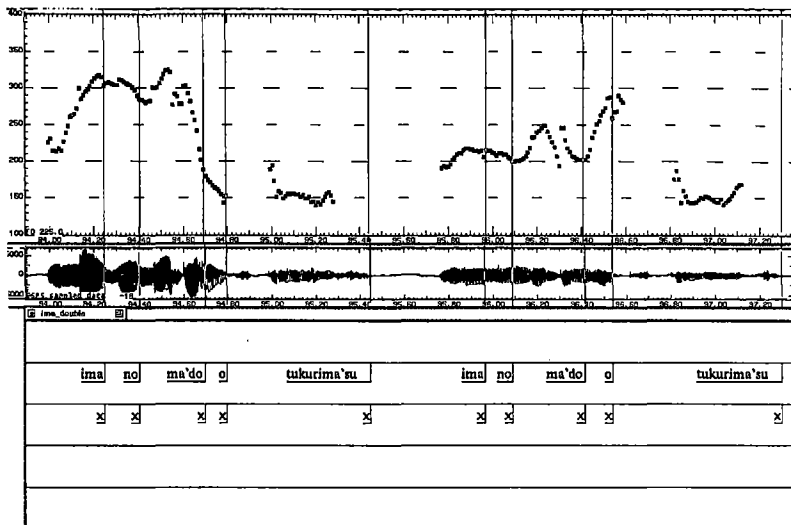


Figure 23: <<ima_double>> "I'll make the livingroom window. I'll make the livingroom window."

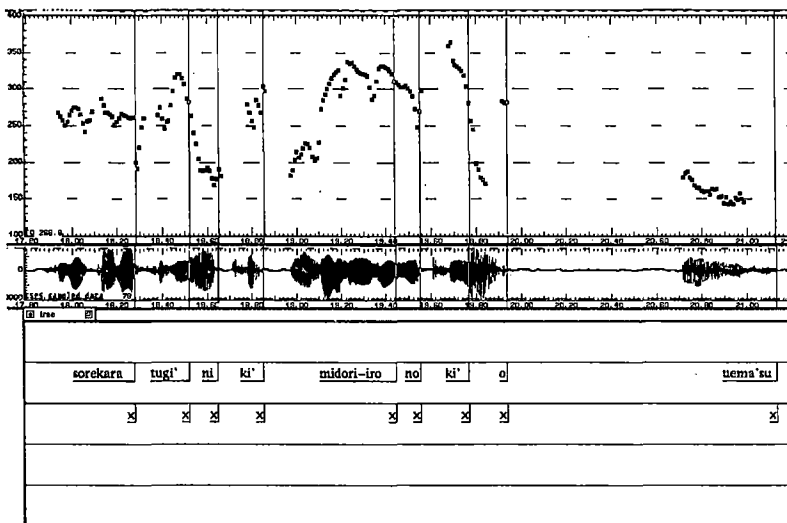


Figure 24: <<tree>> "Then next, I will plant a tree, a green tree."

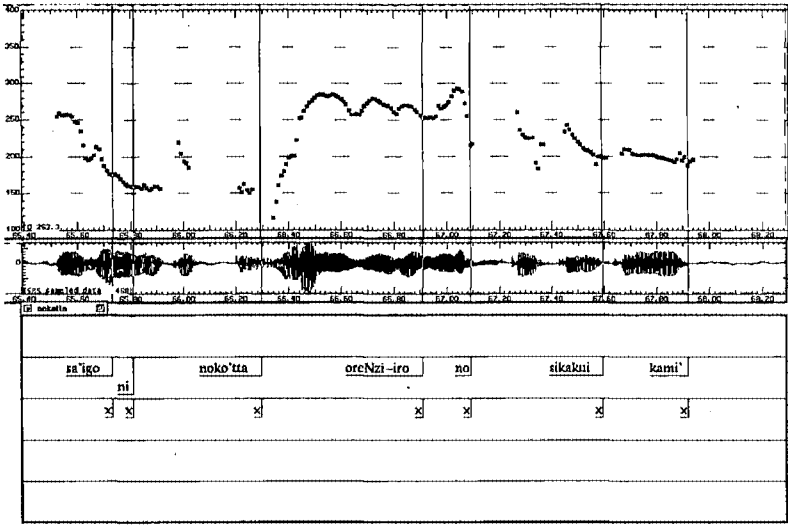


Figure 25: «nokotta» "the last remaining orange paper"

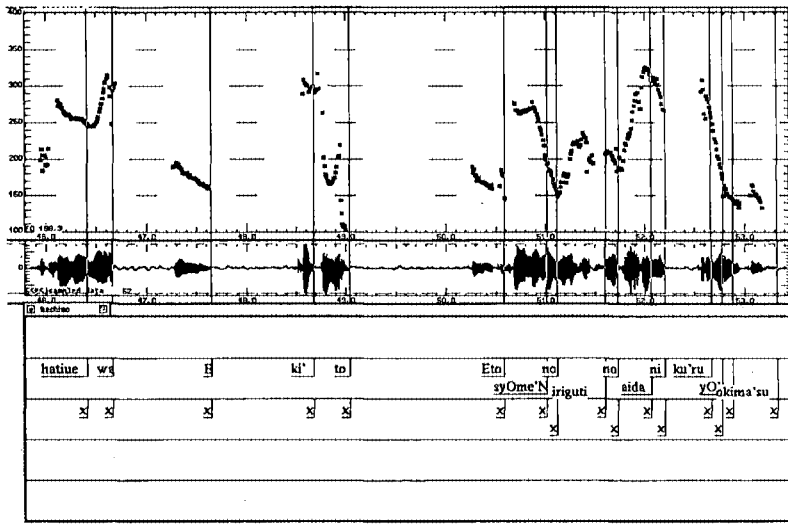


Figure 26: «hachiue» "I'll put the plant, um, so that it comes right in between the tree and, um, the front entrance."

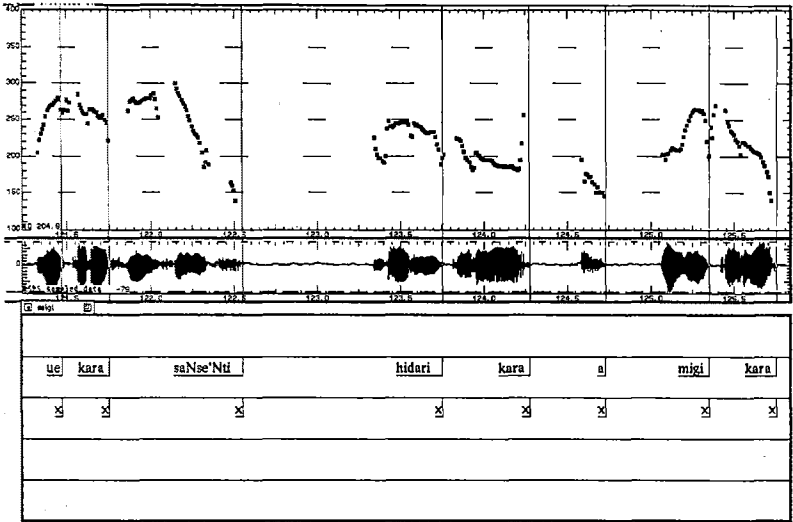


Figure 27: «migi» (Part 1) “3cm from the top, from the left ... oh, 1cm from the right, I’ll attach the livingroom window here.”

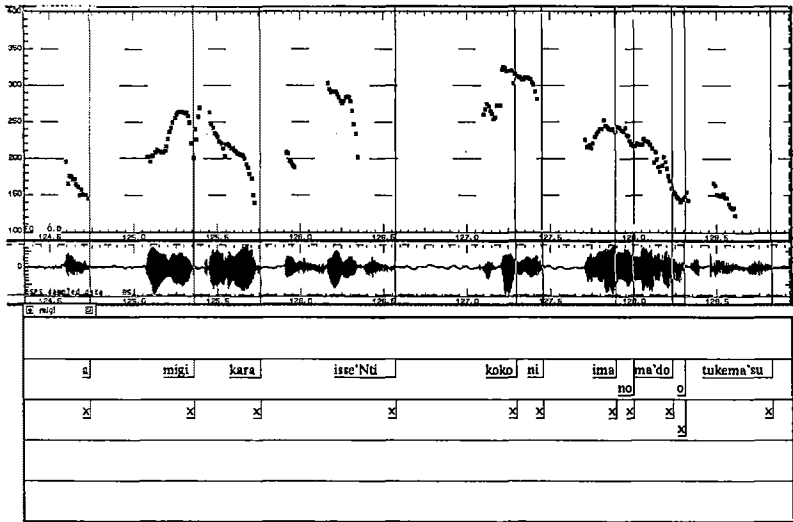


Figure 28: «migi» (Part 2) “3cm from the top, from the left ... oh, 1cm from the right, I’ll attach the livingroom window here.”

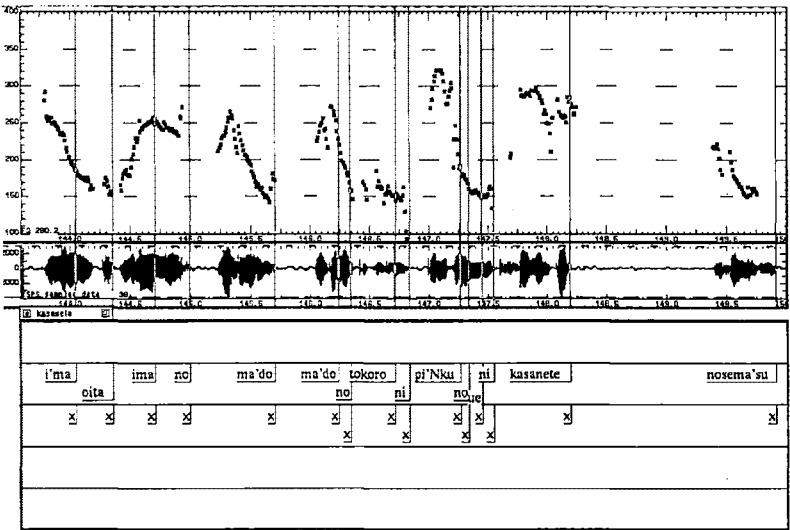


Figure 29: «kasanete» "I'll lay it on top of the pink one, on top of the livingroom window that I just put down."