



Transcribing Prosodic Structure of Spoken Utterances with ToBI

Dr. Nanette Veilleux

Dr. Stefanie Shattuck-Hufnagel

Dr. Alejna Brugos

1. Overview and some basics

Spoken language conveys not only words, but a wide range of other information about timing, intonation, prominence, phrasing, voice quality, rhythm etc. that is often collectively called spoken prosody. These aspects of an utterance are sometimes called supra-segmental, because they can span regions larger than a single phonemic segment (consonant or vowel). The ToBI (Tones and Break Indices) transcription system is a method for transcribing two distinctive aspects of the prosody of spoken utterances:

- a) accents, which contribute to a word's relative prominence in an utterance and
- b) phrasing, which creates a grouping of words.

These prosodic aspects of spoken language can convey distinctive semantic, syntactic or even morphological facts, and can be useful for tasks such as mapping prominence and grouping patterns to meaning differences, understanding the effects of prominence and grouping on the pronunciation of words, and synthesizing prosodically natural-sounding speech. Speech scientists are interested in annotating the prosodic structure of large numbers of utterances in order to study these phenomena, and ToBI was developed to provide a common transcription system that could be used by many different laboratories, making it easier to share data.

An example of how prosodic prominence and grouping can convey differences in structure and meaning can be seen in two different pronunciations of the word string *It broke out in Washington* (Price et al. 1992, Lehiste 1987). On the one hand, this string can be produced with a prominence on *broke* and a phrase boundary just after that word, corresponding to the orthographic representation *It BROKE, out in WASHington*. <broke1> On the other hand, it can be produced with a prominence on *out* and a phrase boundary just after that word, corresponding to *It broke OUT, in WASHington*. <broke2>.

In this case, the two different patterns that correspond to two different syntactic structures and two different meanings can be easily captured by orthographic conventions and punctuation, i.e. upper case typeface and commas. But other differences, like the distinction between prominence signaled by a high pitch vs. one signaled by a low pitch, or between a prominence signaled with an early pitch peak vs. a later one, cannot be easily captured by conventional orthography. The ToBI system, like other transcription systems, was developed to capture prosodic differences of all types, not just the ones that are easy to write down in English orthography. As you will see, the ToBI framework assumes that the differences that matter are phonological differences, i.e. that the job of the transcription system is to capture the differences between prosodic categories that correspond to differences in function and in meaning, but it is not yet entirely clear how this mapping between prosodic categories and their meanings and functions works. One advantage of developing an explicit system for transcribing the categories is that, when it is applied to samples of natural speech, it provides both a stringent test of the theory behind the transcription system, and a laboratory tool for determining how the theory needs to be extended and revised.

A challenging aspect of transcribing prosody is that there is a substantial level of variability. For example, one important cue to prominence and phrasing is the intonation of an utterance, i.e. changes in the pitch that are caused by changes in the frequency of vibration of the vocal folds, often called f_0 . An f_0 that is high in a speaker's range on a salient syllable can mark a pitch

accent, but so can a lower (but still high-for-this-speaker) f_0 on another salient syllable. Moreover, a high f_0 might mark a pitch accented word or it could mark a phrase boundary tone, as in the pitch rise often heard on the final syllable of certain kinds of questions in English, like *Is it raining?* In order to determine the appropriate ToBI transcription, the entire utterance must be parsed, that is, understood as a whole, to determine how the high and low tones are implemented. This tutorial will begin by introducing utterances with relatively unambiguous ToBI annotations and will gradually move to utterances for which even experienced labellers might disagree about how to transcribe some regions. This will be less disturbing if one recognizes that ambiguous cues are also present in other types of speech annotation. For example, the acoustic cues for a canonical /t/ might not be present, e.g. in some renditions of the word *butter*. However, the existence of a well-established lexicon allows annotators to recognize tokens produced with a flap (sounding more like a /d/) and tokens produced more carefully as variations of a single lexical item, and to label them consistently. At this writing, the nature of variation for prosodic categories (and thus for ToBI labels) is not fully understood, and this results in some reasonable disagreement in prosodic parses in some renditions of some utterances, particularly in spontaneous speech. ToBI labelling is a new endeavor compared to, say, phonetic transcription, and as the nature and range of prosodic variation becomes better understood and documented, we expect the transcription ambiguity to lessen. In the meantime, the system provides a number of tools for recording your uncertainty, to help us understand better where the system can be improved. In fact, although the development of a ToBI system for a particular language requires a certain degree of prior understanding of how prominence and grouping work in that language, a ToBI system can also be viewed as a tool for exploring the prosodic system of a language in greater detail. In this tutorial we introduce you to the ToBI system that has been developed for Mainstream American English; for information about ToBI systems for other languages, and about the general ToBI framework within which these systems have been developed, see Jun (2005), *Prosodic Typology*.

1.1. The basic parts of ToBI¹

A ToBI transcription of an utterance consists minimally of a recording of the speech, its fundamental frequency contour, and (in the transcription proper) symbolic labels for prosodic events. The transcription proper is usually arranged in four time-aligned parallel horizontal panels or tiers, so that the symbolic labels can be easily matched with the corresponding f_0 track and speech waveform. (Other tiers can be added for the needs of particular sites.) The four labelling tiers each appear in their own window:

- (1) the Tone tier, for transcribing tonal events
- (2) the Orthographic tier, for transcribing words
- (3) the Break-Index tier, for transcribing boundaries between words
- (4) the Miscellaneous tier, for recording additional observations

In addition, two new tiers have been suggested for labellers who find them useful. These are not used in the examples in the beginning of this tutorial, but will be explained in Section 2.12:

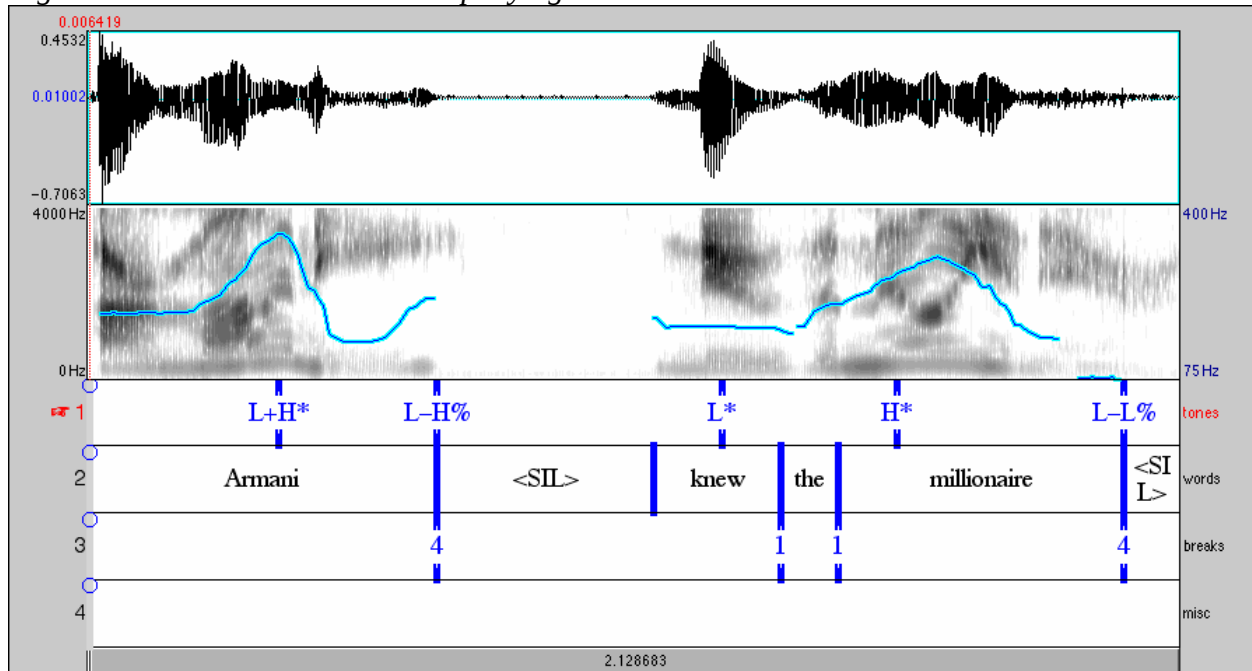
- (5) an Alternative Tier, for transcribing alternative labels in the case of ambiguity
- (6) a Discussion tier, for recording data for selected research issues

¹ Parts of this section and section 1.2 have been adapted from the Guidelines for ToBI Labelling, version 3.0 (1997).

One popular program for labelling and displaying ToBI transcriptions is Praat (available at <http://www.praat.org>). Here is an example using Praat to display the utterance waveform, the f0 contour, the spectrogram and the first four tiers for the utterance *Armani knew the millionaire*.

Figure 1: *armani1.wav* with accompanying Praat TextGrid.

<armani1>



The topmost window in this display shows the waveform of the recorded utterance; later, when you learn how to expand the horizontal time scale, you will see more clearly the vertical striations that indicate the individual pitch pulses made by the vocal folds as they vibrate. The horizontal axis corresponds to time and the vertical axis to the amplitude of the vibration. You can see in the waveform that the amplitude varies roughly with the syllable structure of the utterance: it is large for vowels (where the mouth is open relatively wide), smaller for consonants (which have a constriction in the vocal tract), and zero (or nearly zero) when there is no speech signal.

The rate of vibration of the vocal folds is what we hear as the pitch, and this is represented in the second window as a semi-continuous blue line superimposed on a different representation of the speech signal, the spectrogram (see “What is the spectrogram?” in the grey box below). When you play this utterance, notice where you hear a higher pitch. Do these words and syllables correspond to the places where the F0 contour (also called the pitch track or the f0 track, although f0 and perceived pitch are not exactly the same thing) is higher, as indicated by the blue lines?

What is the spectrogram?

The gray and white pattern in the background behind the f0 track is the spectrogram which is another way of displaying information about the speech signal that you see in the wave form above it. Like the wave form, it shows time on the horizontal axis, but the vertical axis here is frequency. Among other things, the spectrogram shows the frequencies that resonate especially well in the vocal tract as black horizontal bands that appear and disappear, and rise and fall, as the speech articulators move around inside the vocal tract, changing the size and shape of the resonant cavities inside the mouth, nose and throat. This is much like the change in sound that happens when a jug is filling up with water from a faucet: the pitch of the noise rises as the air cavity above the water level gets smaller). The spectrogram often shows the change from a consonant to a vowel especially clearly; notice the sharp boundaries between the /m/s and /n/s and their surrounding vowels in Figure 1.

Experienced ToBI labellers often use the spectrogram to help find the location of successive sounds, syllables and words. This in turn helps them to keep track of where changes in the pitch track occur across the words and syllables of the utterance. Learning how to use this information may take some time, so at the beginning you may want to rely more on listening to make these judgments.

Underneath the spectrogram with the blue pitch contour is the set of four thinner white boxes that make up the four tiers in the ToBI labelling window. Unlike the speech displays, these boxes are text writeable and this is where you will type in the ToBI transcription labels. The top white box is the Tone tier, and the third box is the Break Index tier. These two tiers represent the core ToBI analysis. The Tone tier is the part of the transcription that corresponds most closely to a phonological analysis of the utterance's intonation pattern. It consists of labels for distinctive pitch events, transcribed as a sequence of high (H) and low (L) tones marked with diacritics indicating their intonational function. Tones function either as prominence markers called pitch accents, as parts of pitch accents, or as boundary-related events called phrase accents and boundary tones, that mark the edges of two types of phrases. These categories are based on the work of Janet Pierrehumbert (1980) and joint work by Mary Beckman and Janet Pierrehumbert (1986, 1988).

The Break-Index tier captures the prosodic grouping of the words in an utterance by labelling the end of each word for the subjective strength of its association with the next word, on a scale from 0 (for weakest perceived boundary/strongest perceived conjoining, as in *doncha* for *don't you*) to 4 (for the most disjoint boundary, i.e. at the end of the highest-level intonationally marked phrase). These categories of association strength or “break indices” are based on work by Mari Ostendorf, Patti Price, Stefanie Shattuck-Hufnagel, and their associates (Price et al., 1991). The two highest break indices (3 and 4) are equated with two levels of prosodic groupings (phrases) that are marked intonationally; additional higher-level groupings of these Intonational Phrases are not marked in the ToBI system.

The Orthographic tier is the third white box. It contains a straightforward transcription of all of the words in the utterance, in ordinary English orthography. The word transcriptions are aligned with their locations in the speech waveform. For labellers using Praat or a similar labelling computer application, the convention is to place the orthographic label for a word between two marks that delineate the approximate time interval in the signal that corresponds to the utterance of that word and placing <SIL> to mark silence between words, if any.² The orthographic tier is arguably not part of any core prosodic analysis, except inasmuch as the labels on this tier can be used to interface the transcription to dictionary entries which do indicate such things as which syllable is likely to be most stressed in each word, prosodic information which is not otherwise included in the ToBI system (more on this below). This tier also helps the labeller to keep track of which time regions in the wave form, spectrogram and f0 track correspond to which words in the utterance.

The Miscellaneous tier is the bottom white box in this display. It is essentially a 'comment' tier that can be used to mark events such as breaths, coughs, laughter, long silences and other non-speech events. These are traditionally marked with angle brackets (e.g. <cough>). Like the orthographic tier, it can include events that are arguably not part of prosody per se. However, many events that are typically marked on the Miscellaneous tier are important for interpreting the analyses on the Tone tier and Break-Index tier, because they disrupt the smooth rhythm of the utterance or interrupt the intonation contour. Labels on this tier usually mark the beginning and end of an event interval; one exception is the label 'disfl', which often stands alone to flag the occurrence of a perceived disfluency of some type.

1.2. Guiding principles

As the preceding discussion shows, ToBI does not try to transcribe all aspects of prosody, or even all aspects that are amenable to symbolic transcription. In deciding what to include and what to leave out, we are guided by three principles. First, we want to be able to distinguish in our transcription all of the categorically distinct intonation patterns and prosodic units of the language (in this case, Mainstream American English (MAE), see Jun 2005 for ToBI systems for other languages and dialects). Second, we do not transcribe aspects of prosody which are more amenable to continuous-valued quantitative measures than to the categorical divisions of a symbolic transcription, such as the slope of the changing f0 curve. Finally, we do not want to squander the user's energies in transcribing even categorical aspects of prosody which are predictable from other parts of the transcription or from auxiliary tools, such as dictionaries, that can be used to determine the location of lexical stress within words.

The categorical aspects of prosody which we try to capture completely (according to the first principle) are of two types. The first is the prosodic structure -- the alternating rhythm of more and less prominent words and syllables and the grouping of words into prosodic constituents of various sizes -- and the second is the intonation pattern -- the sequence of contrastive pitch events

² In older transcriptions using xwaves TM, the orthographic label had to be aligned to the end of the word rather than spanning the whole word interval. Other popular programs for displaying and labelling speech are wavesurfer (<http://www.speech.kth.se/wavesurfer>) and emu (<http://emu.sourceforge.net/emu-tobi.shtml>).

that we call pitch accents, phrase accents, and boundary tones, and that determines the f0 contour of the utterance. A basic assumption of the ToBI approach is that both of these goals can be met using the same inventory of elements.

[The next two paragraphs contain further discussion of what is not captured by a ToBI transcription; read them if you are curious about this question. Otherwise, skip directly to section 2.0 below.]

An example of the non-categorical aspects of prosody which we leave out (in accordance with the second principle) is the local tempo of each word in the utterance, which we feel could be more accurately and directly captured by some quantitative measure such as normalized segment duration (e.g., Campbell, 1992) than by any symbolic transcription such as an arbitrary division into, say, categories `1', `2', and `3' (for `slow', `medium', and `fast' tempi).

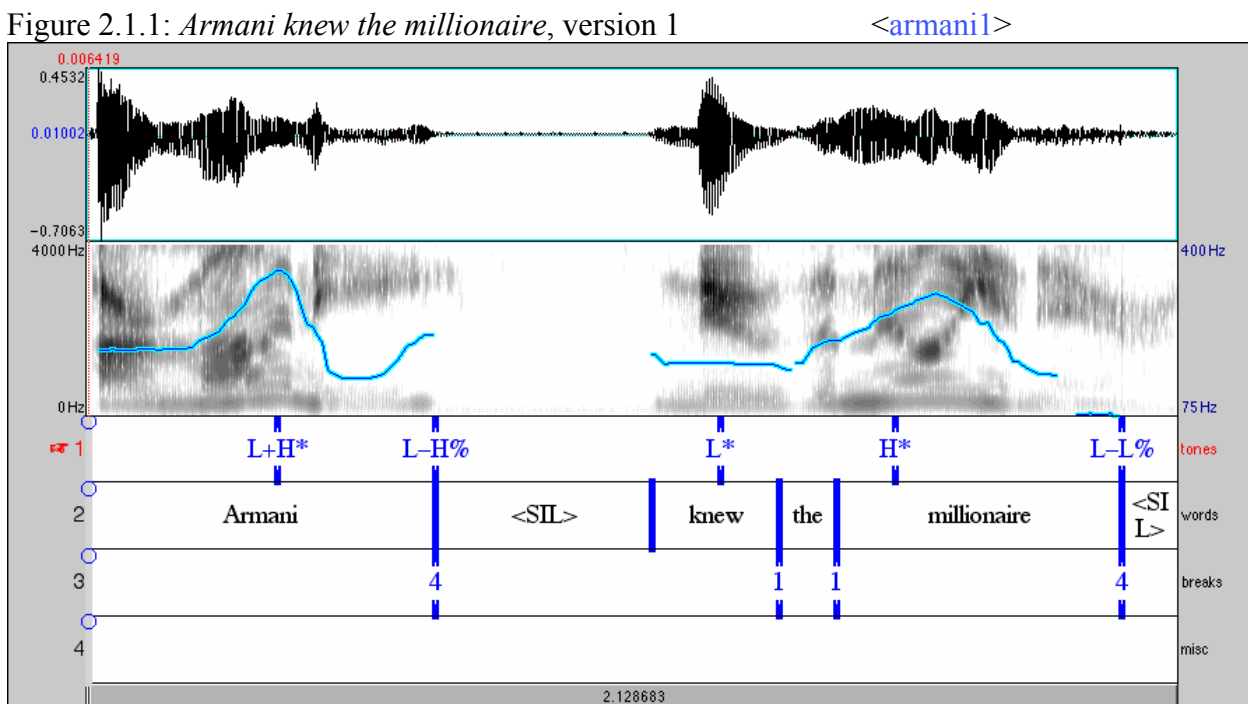
A categorical aspect of prosody which we leave out (in accordance with the third principle) because it should be fairly predictable is the marking of the lexically stressed and unstressed syllables within each word. By this level of stress we mean the word-internal alternation between more stressed and less stressed syllables, where the relative prominence of any pair of syllables is fairly fixed and can be thought of as inherent to the word's dictionary entry.

2.0 ToBI Tutorial

As we have seen, ToBI is a system of prosodic labelling that attempts to capture certain of the prosodic events in a spoken utterance. These prosodic events fall into two categories: 1) events or patterns that mark syllables as more salient or more prominent than neighboring syllables by reason of their intonation, and 2) events or patterns that mark phrasing or grouping within sequences of words.

For example, in the utterance shown in section 1 and repeated here <armani1>, the words *Armani*, *knew* and *millionaire* are more prominent than *the*. Moreover, within the multi-syllable word *Armani*, the syllable *-man-* is more prominent than *Ar-* or *-ni*. (What is your intuition about the relative prominence of *mill-*, *-io-* and *-naire*?) The sentence is also produced as two intonational phrases. That is, the first word, *Armani*, is not grouped with the remaining words in the utterance, *knew the millionaire*.

Figure 2.1.1: *Armani knew the millionaire*, version 1



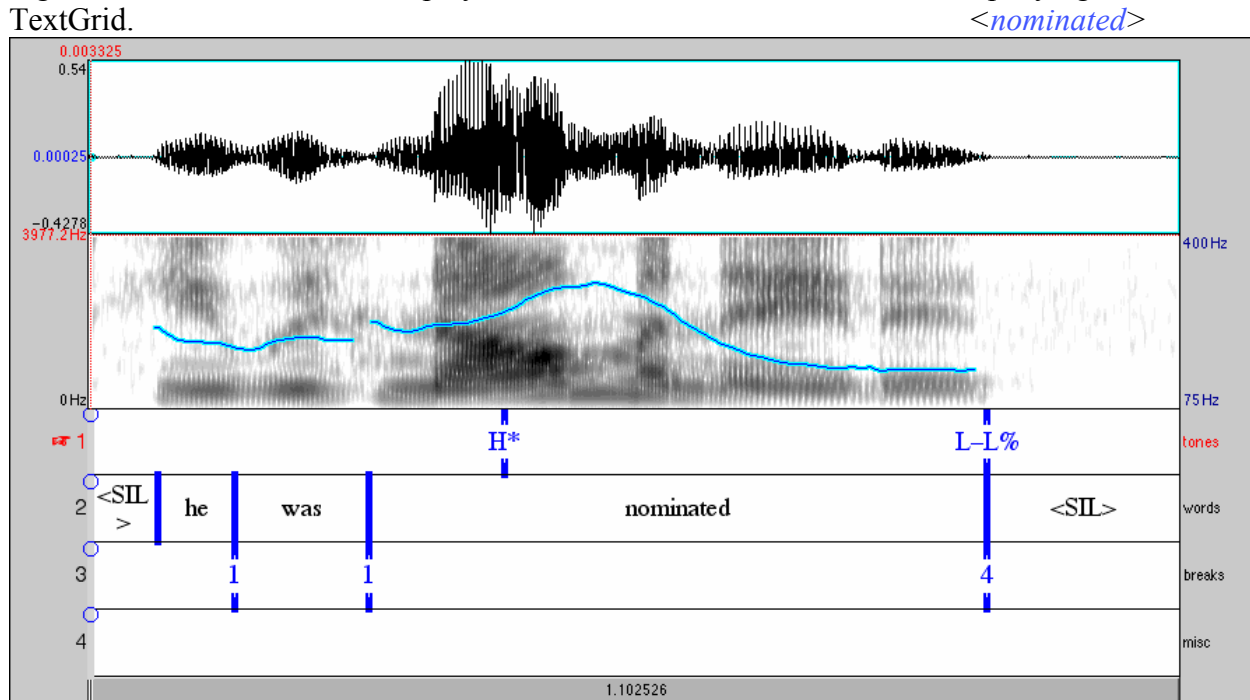
The acoustic correlates that mark these prominences and groupings are typically the f_0 pattern (related to the perceived pitch pattern) and the relative duration; the contributions of other cues such as amplitude (related to perceived loudness), voice quality, pausing, strength of articulation etc. are live topics of research and discussion. The Tone tier captures information about both phrasing and prominence, while the Break Index tier captures primarily phrasing and other grouping information. In the following sections of this introductory tutorial, elements from both of these tiers will be gradually introduced using examples that you can listen to and view labels for.

The first examples you will see are intentionally chosen as straightforward and context independent, to serve as clear illustrations of individual prosodic elements. As the tutorial progresses, more complexity will be introduced, showing the kinds of challenges that can arise in labelling when more elements are present to interact. Unfortunately, isolated examples heard out of context are sometimes missing substantial amounts of important information that can help the listener to parse (interpret) the cues in the acoustic signal in relationship to ToBI labels; examples of later sections will provide more context.

2.1 A first example, introducing the tone labels H*, L-L%, and the break labels 0, 1, and 4

One of the most basic prosodic patterns can be seen in the short utterance in <nominated>, shown in Figure 2, below. (To look closely at this example, open the files nominated.wav and nominated.TextGrid in Praat.)

Figure 2.1.2: nominated.wav displayed with nominated.TextGrid, the accompanying Praat TextGrid.

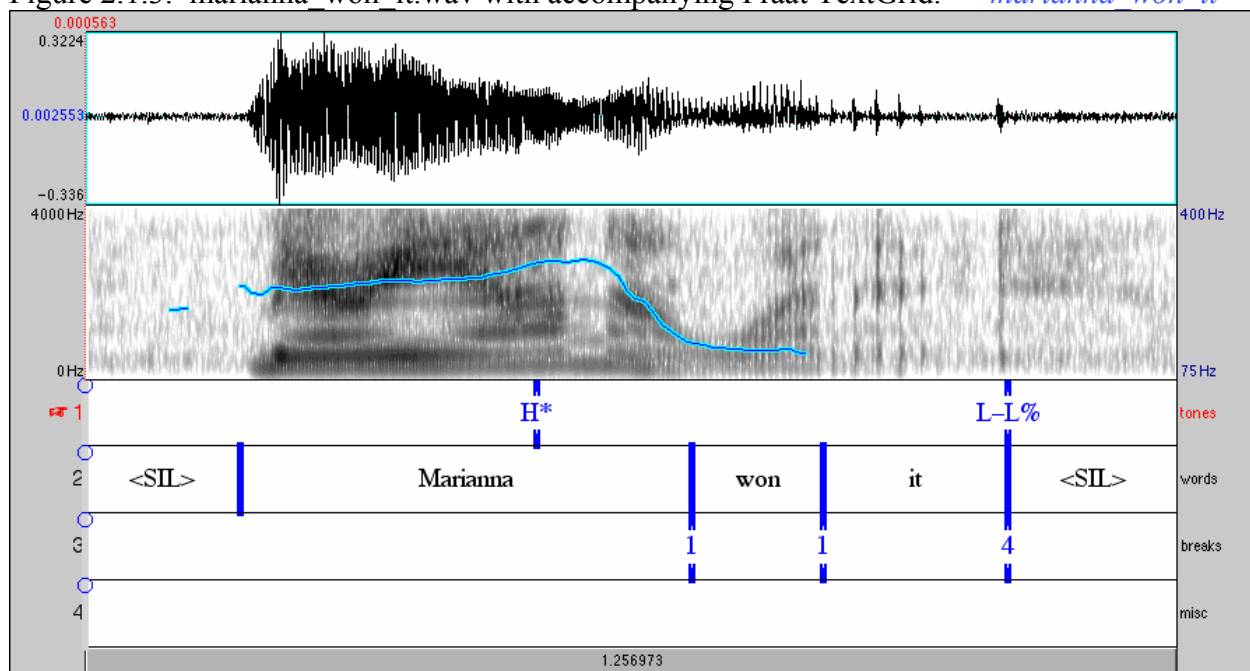


This example has 3 words, and is produced as a single intonational phrase with one pitch accent. Let's look at the transcription for the Tone tier first. As mentioned above, the Tone tier marks two types of events: those associated with accented syllables and those associated with phrasing. The most prominent syllable in this utterance is the first syllable (*nominated*) of *nominated*. The f0 contour shows a rise that peaks at the end of *nom-* and then falls. This is an example of a canonical H* or High pitch accent. In the ToBI transcription, the H stands for a High tonal target and the * indicates that the H tone is associated with an accented syllable. Perceptually, this H* syllable is more salient than other syllables around it. There are other types of pitch accents, including bi-tonal types, that make up the full ToBI inventory. We'll look at these other types in later sections.

The other intonational event that is labelled in the ToBI transcription of this utterance is the falling f0 at the end of the utterance. This low-pitched region is marked with L-L% where the L- indicates that there is a Low phrase accent that is followed by a Low boundary tone (indicated by the L%) on the final syllable. In later sections we'll look at phrase accents and boundary tones separately.

Like all intonation contours, this one can be produced on many different word strings. Another short example is shown in the file <marianna_won_it>. Again, this contour is a single intonational phrase with one pitch accent. The pitch accent is an H* and the phrase accent/boundary tone combination is an L-L%.

Figure 2.1.3: marianna_won_it.wav with accompanying Praat TextGrid. <marianna won it>

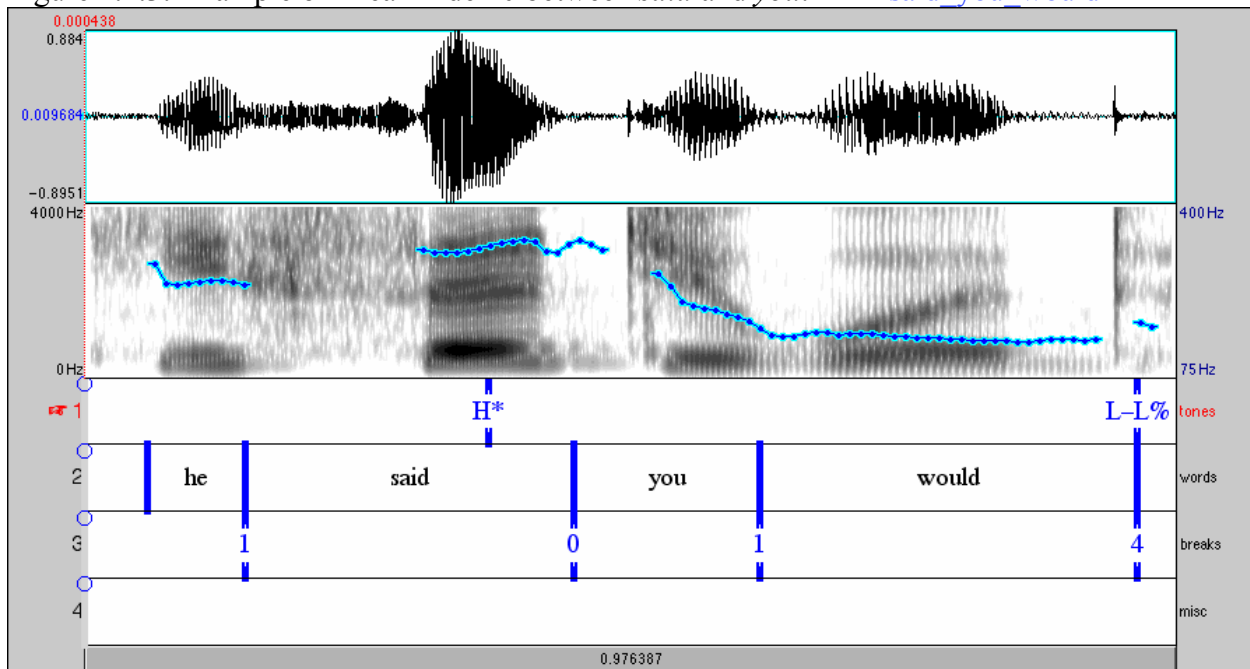


This L-L% sequence is often found at the end of spoken declarative sentences, especially final sentences in a turn or discourse. Later we will see phrase boundaries of a different type that have only a phrase accent. Moreover, just as pitch accents can be combined in different ways within an intonation contour, there are other combinations of high and low tones that make up the full inventory of phrase-boundary-related tonal markers.

Before introducing other kinds of pitch accents, phrase accents and boundary tones, however, let's turn our attention to labels in the Break Index tier. A Break Index marks the level of disjuncture between two words. Since intonation is one of the acoustic cues to disjuncture, there is some redundancy between the Tone and Break Index tiers. For example, an L-L% boundary tone signals the maximal level of disjuncture in this system (i.e. at the end of an Intonational Phrase) and would be marked with a 4. Other boundaries are marked with smaller break indices. For example, typical word boundaries in a fluent sequence of words within a phrase are marked

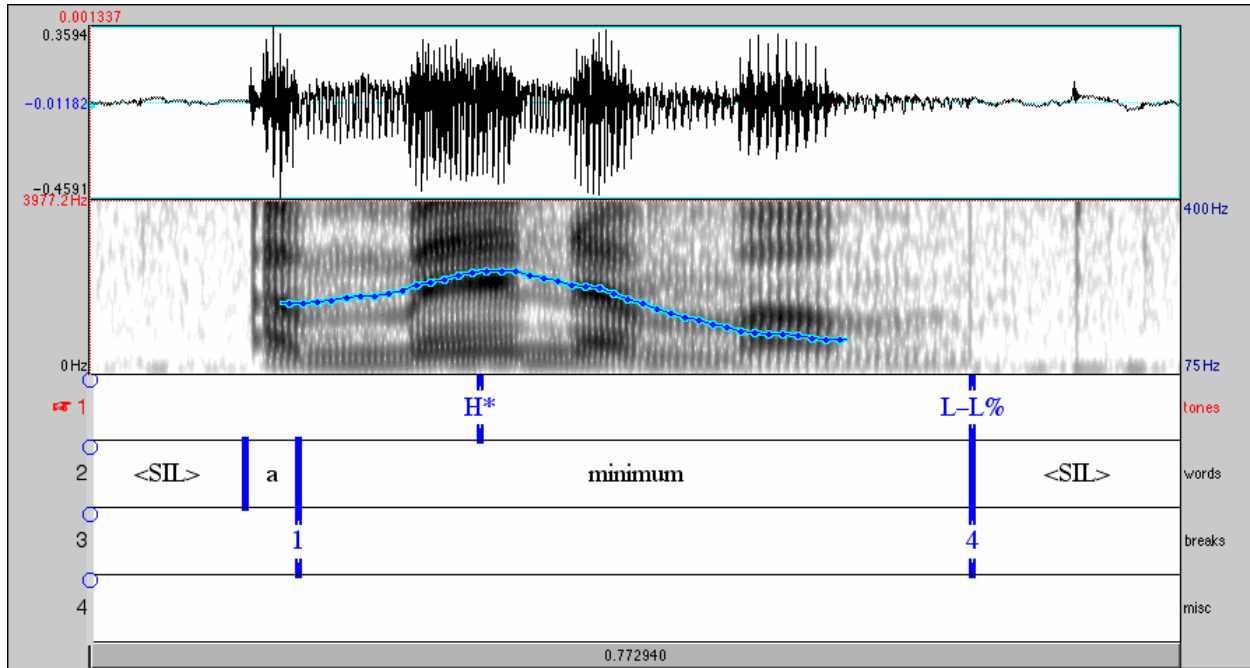
with a 1. In this example, all break indices are 1's except for the final 4 break index. There are levels of intermediate phrasing within intonational phrases that will be marked with break indices 2 or 3, as we will show later. The 0 break index is reserved for the case when two words are produced so that the boundary between them is indeterminate. For example, the word sequence *did you* is frequently produced as *didja*, where the final /d/ of *did* and the initial /y/ of *you* together form a /j/-like sound. (In example <said_you_would> “He said you would”, listen for the merging of the final /d/ of *said* and the initial /y/ of *you*.)

Figure 2.1.3: Example of Break Index 0 between *said* and *you*. <said_you_would>



The following examples show more instances of the same H* L-L% contour on utterances of varying lengths.

Figure 2.1.4: example <minimum1>, a 2-word utterance with H* L-L%. Here, the initial low tone might suggest an alternate pitch accent (L+H*) that will be discussed later. A newly adopted tier, the Alternatives tier, allows labellers to capture these reasonable alternate parses of the intonation contour.



To summarize the inventory that we've seen so far:

Tones:

H* high pitch accent

L-L% low phrase accent, low boundary tone

Break indices:

0: word boundary apparently erased

1: typical between-word disjuncture within a phrase

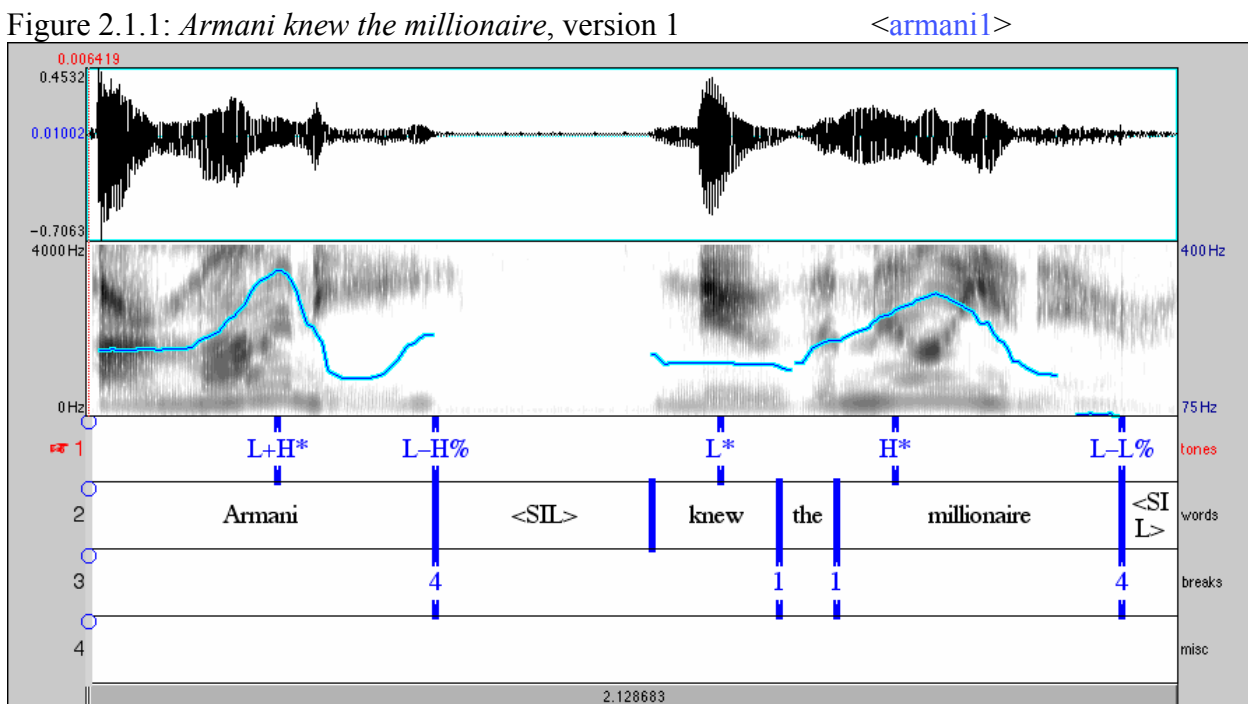
4: end of an intonational phrase

2.0 ToBI Tutorial

As we have seen, ToBI is a system of prosodic labelling that attempts to capture certain of the prosodic events in a spoken utterance. These prosodic events fall into two categories: 1) events or patterns that mark syllables as more salient or more prominent than neighboring syllables by reason of their intonation, and 2) events or patterns that mark phrasing or grouping within sequences of words.

For example, in the utterance shown in section 1 and repeated here <armani1>, the words *Armani*, *knew* and *millionaire* are more prominent than *the*. Moreover, within the multi-syllable word *Armani*, the syllable *-man-* is more prominent than *Ar-* or *-ni*. (What is your intuition about the relative prominence of *mill-*, *-io-* and *-naire*?) The sentence is also produced as two intonational phrases. That is, the first word, *Armani*, is not grouped with the remaining words in the utterance, *knew the millionaire*.

Figure 2.1.1: *Armani knew the millionaire*, version 1



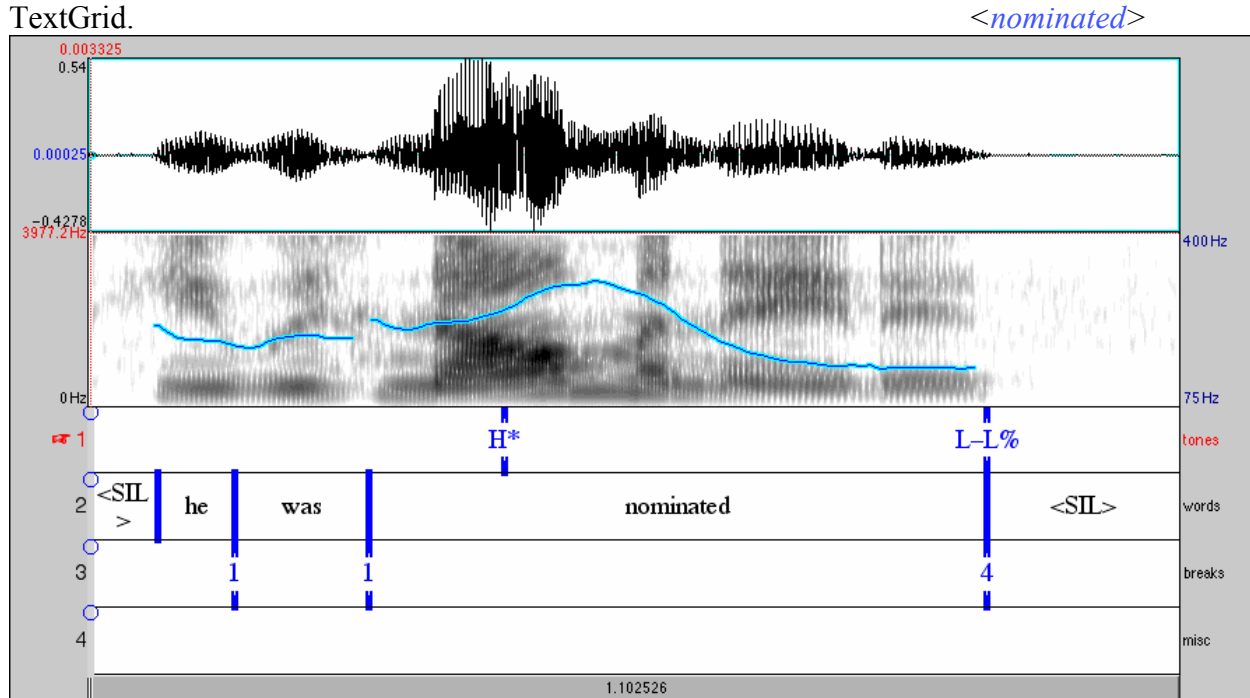
The acoustic correlates that mark these prominences and groupings are typically the f_0 pattern (related to the perceived pitch pattern) and the relative duration; the contributions of other cues such as amplitude (related to perceived loudness), voice quality, pausing, strength of articulation etc. are live topics of research and discussion. The Tone tier captures information about both phrasing and prominence, while the Break Index tier captures primarily phrasing and other grouping information. In the following sections of this introductory tutorial, elements from both of these tiers will be gradually introduced using examples that you can listen to and view labels for.

The first examples you will see are intentionally chosen as straightforward and context independent, to serve as clear illustrations of individual prosodic elements. As the tutorial progresses, more complexity will be introduced, showing the kinds of challenges that can arise in labelling when more elements are present to interact. Unfortunately, isolated examples heard out of context are sometimes missing substantial amounts of important information that can help the listener to parse (interpret) the cues in the acoustic signal in relationship to ToBI labels; examples of later sections will provide more context.

2.1 A first example, introducing the tone labels H*, L-L%, and the break labels 0, 1, and 4

One of the most basic prosodic patterns can be seen in the short utterance in <nominated>, shown in Figure 2, below. (To look closely at this example, open the files nominated.wav and nominated.TextGrid in Praat.)

Figure 2.1.2: nominated.wav displayed with nominated.TextGrid, the accompanying Praat TextGrid.

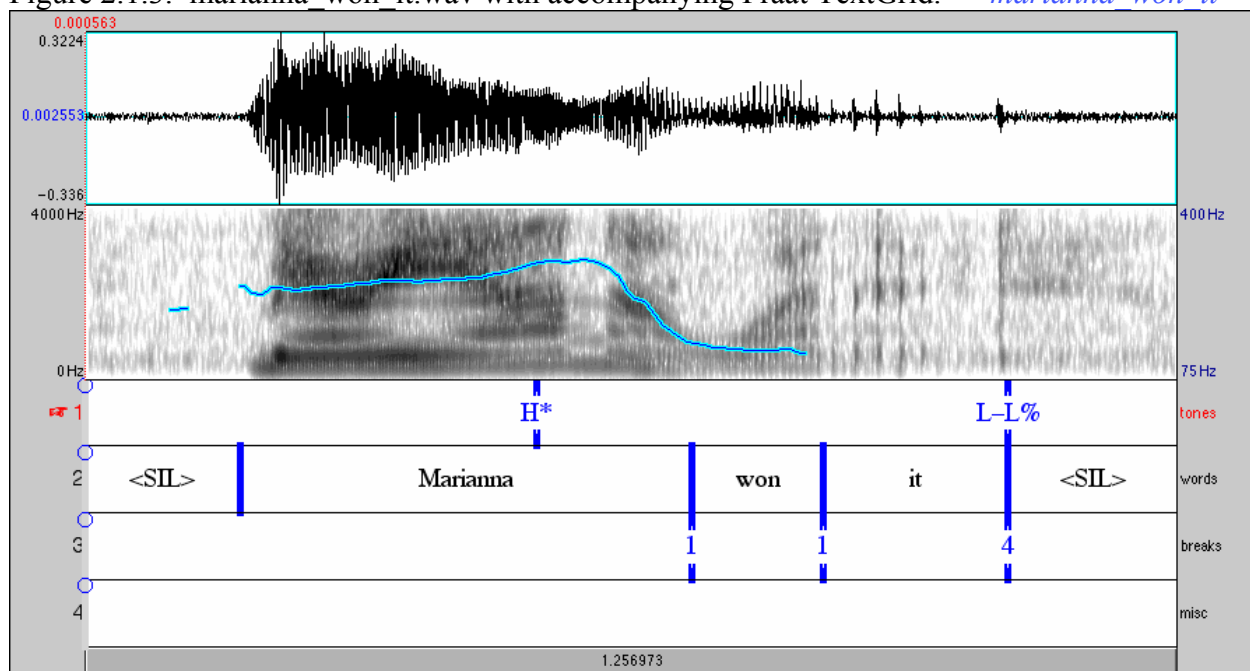


This example has 3 words, and is produced as a single intonational phrase with one pitch accent. Let's look at the transcription for the Tone tier first. As mentioned above, the Tone tier marks two types of events: those associated with accented syllables and those associated with phrasing. The most prominent syllable in this utterance is the first syllable (*nominated*) of *nominated*. The f0 contour shows a rise that peaks at the end of *nom-* and then falls. This is an example of a canonical H* or High pitch accent. In the ToBI transcription, the H stands for a High tonal target and the * indicates that the H tone is associated with an accented syllable. Perceptually, this H* syllable is more salient than other syllables around it. There are other types of pitch accents, including bi-tonal types, that make up the full ToBI inventory. We'll look at these other types in later sections.

The other intonational event that is labelled in the ToBI transcription of this utterance is the falling f0 at the end of the utterance. This low-pitched region is marked with L-L% where the L- indicates that there is a Low phrase accent that is followed by a Low boundary tone (indicated by the L%) on the final syllable. In later sections we'll look at phrase accents and boundary tones separately.

Like all intonation contours, this one can be produced on many different word strings. Another short example is shown in the file <marianna_won_it>. Again, this contour is a single intonational phrase with one pitch accent. The pitch accent is an H* and the phrase accent/boundary tone combination is an L-L%.

Figure 2.1.3: marianna_won_it.wav with accompanying Praat TextGrid. <marianna won it>

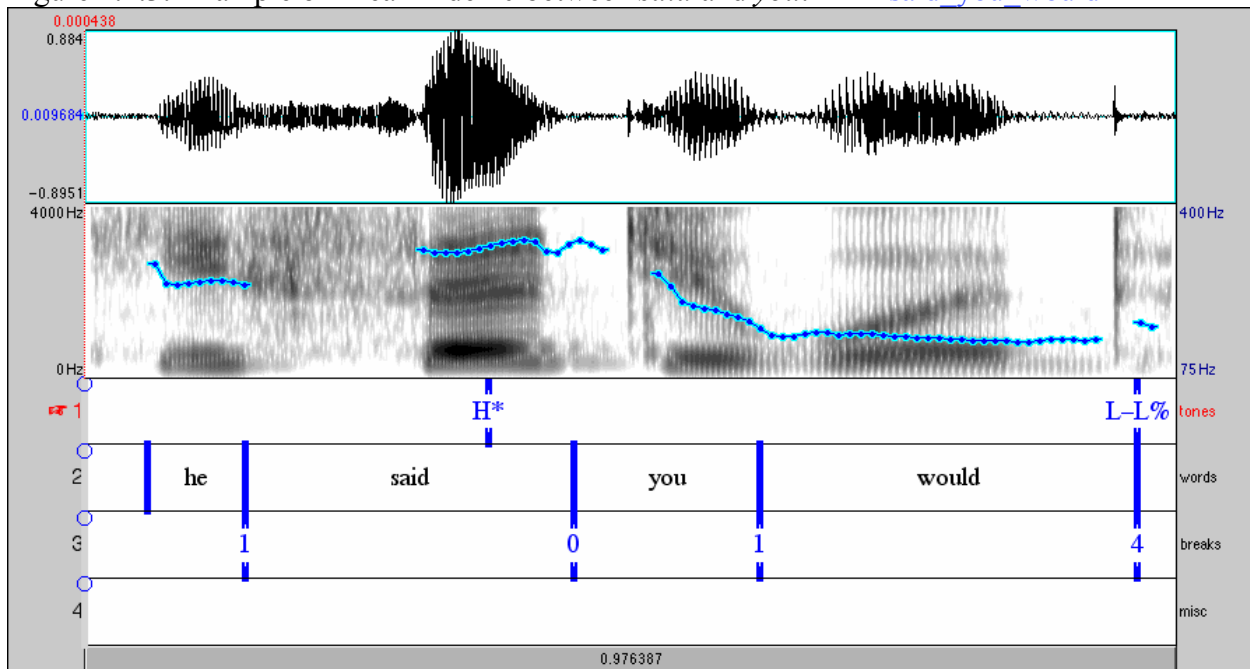


This L-L% sequence is often found at the end of spoken declarative sentences, especially final sentences in a turn or discourse. Later we will see phrase boundaries of a different type that have only a phrase accent. Moreover, just as pitch accents can be combined in different ways within an intonation contour, there are other combinations of high and low tones that make up the full inventory of phrase-boundary-related tonal markers.

Before introducing other kinds of pitch accents, phrase accents and boundary tones, however, let's turn our attention to labels in the Break Index tier. A Break Index marks the level of disjuncture between two words. Since intonation is one of the acoustic cues to disjuncture, there is some redundancy between the Tone and Break Index tiers. For example, an L-L% boundary tone signals the maximal level of disjuncture in this system (i.e. at the end of an Intonational Phrase) and would be marked with a 4. Other boundaries are marked with smaller break indices. For example, typical word boundaries in a fluent sequence of words within a phrase are marked

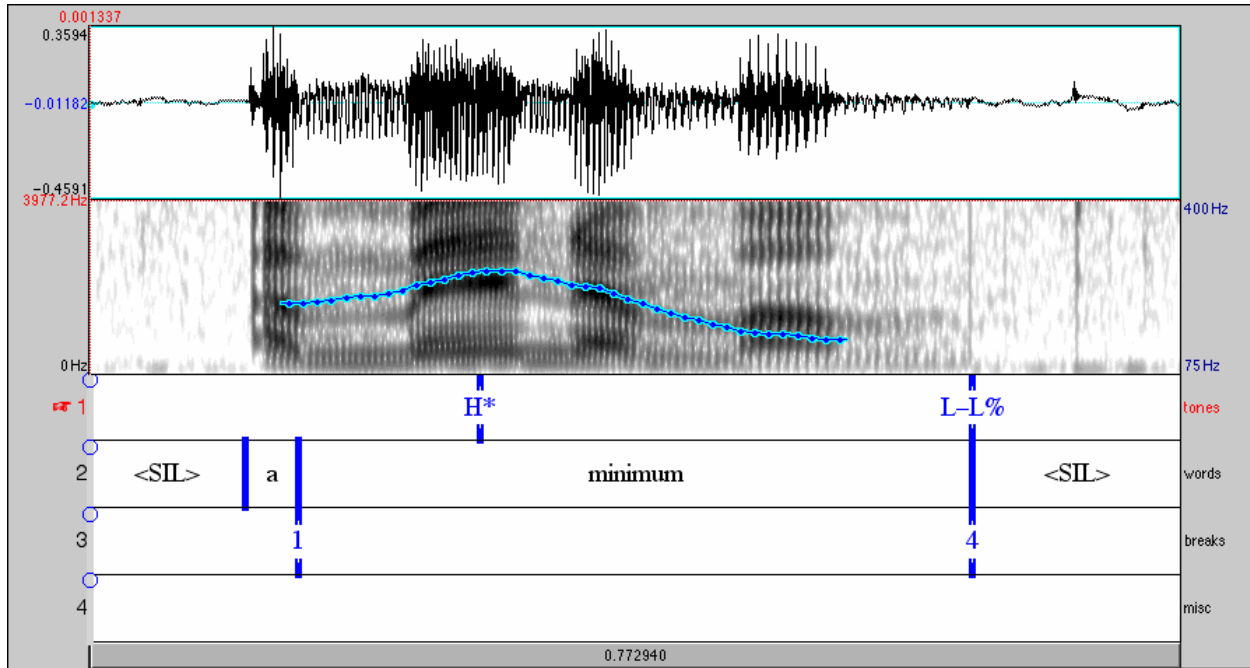
with a 1. In this example, all break indices are 1's except for the final 4 break index. There are levels of intermediate phrasing within intonational phrases that will be marked with break indices 2 or 3, as we will show later. The 0 break index is reserved for the case when two words are produced so that the boundary between them is indeterminate. For example, the word sequence *did you* is frequently produced as *didja*, where the final /d/ of *did* and the initial /y/ of *you* together form a /j/-like sound. (In example <said_you_would> “He said you would”, listen for the merging of the final /d/ of *said* and the initial /y/ of *you*.)

Figure 2.1.3: Example of Break Index 0 between *said* and *you*. <said_you_would>



The following examples show more instances of the same H* L-L% contour on utterances of varying lengths.

Figure 2.1.4: example <minimum1>, a 2-word utterance with H* L-L%. Here, the initial low tone might suggest an alternate pitch accent (L+H*) that will be discussed later. A newly adopted tier, the Alternatives tier, allows labellers to capture these reasonable alternate parses of the intonation contour.



To summarize the inventory that we've seen so far:

Tones:

H* high pitch accent

L-L% low phrase accent, low boundary tone

Break indices:

0: word boundary apparently erased

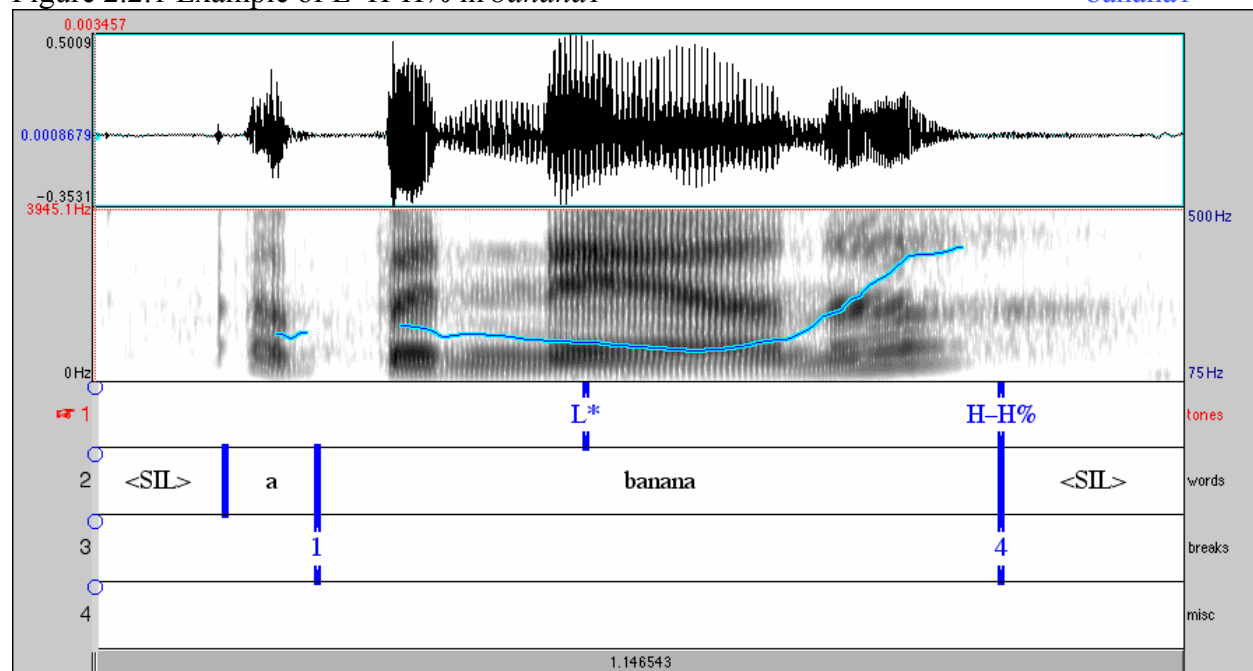
1: typical between-word disjuncture within a phrase

4: end of an intonational phrase

2.2 Adding to the inventory: tones L* and H-H%

Tonal events that cue phrasally-prominent syllables can be high or low. The following example has a falling f0 that reaches a minimum on the *-nan-* syllable in *banana1*. This is a Low pitch accent on *banana*, marked with L*. Again, the L stands for a low tonal target and the * means that the low tone is associated with a prominent syllable. Perceptually, you can hear that the *-nan-* in *banana* is more salient than other syllables in this utterance. Also notice that the end of the utterance has a sharply rising f0. This intonational phrase has a high phrase accent (H-) and a following high boundary tone (H%), hence it is marked as H-H%. In Mainstream American English (MAE), this intonational contour is one typical way of indicating that an utterance is a question for which the speaker expects a yes-or-no answer.

Figure 2.2.1 Example of L*H-H% in *banana1*

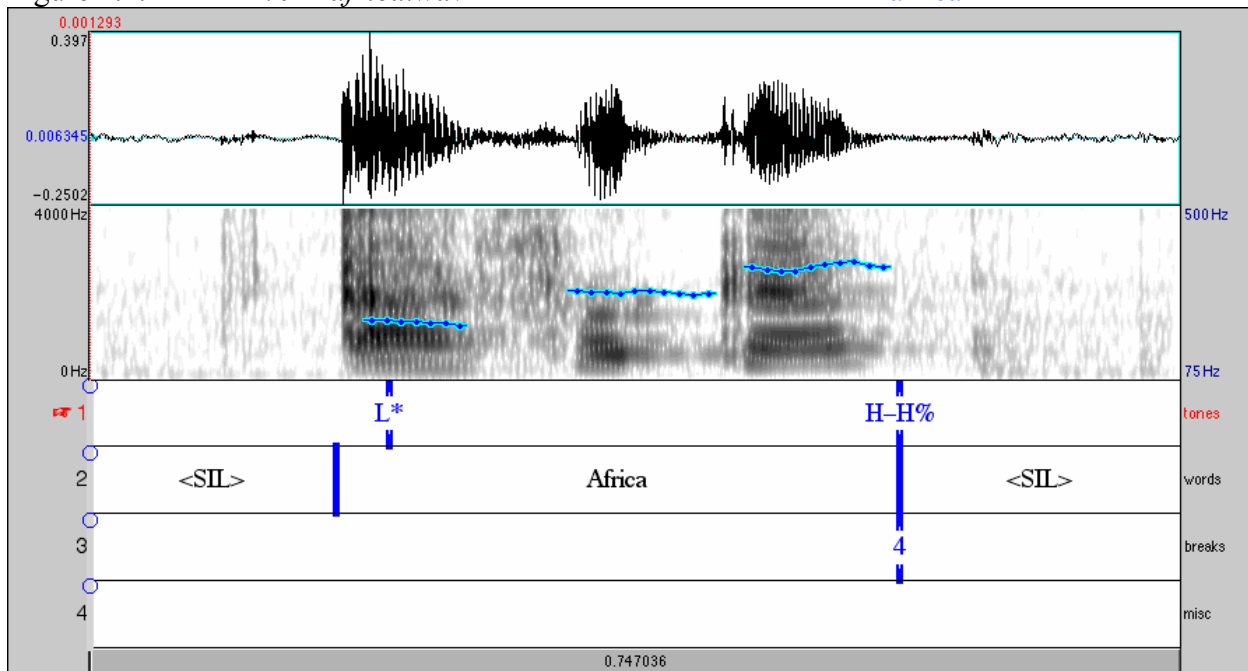


The distribution of break indices in this example is the same as many of those seen in the examples in Section 2.1: there are 1's between each pair of words and a 4 at the end of the Intonational Phrase (IP).

In the next example, <africa>, a chance placement of unvoiced segments /f/ and /k/ (sounds for which the vocal folds don't vibrate, so there is no f0) separates the three tonal elements L* H-H%. This results in a step-like appearance that, while convenient to illustrate the three different elements, is not typical. In other segmental contexts, the f0 for the successive tonal elements will flow smoothly from one to the next. By the way, the L* may seem to be rather high if examined out of context. However, this speaker has a relatively high pitch range.

Figure 2.2.2 L*H-H% in *africa.wav*

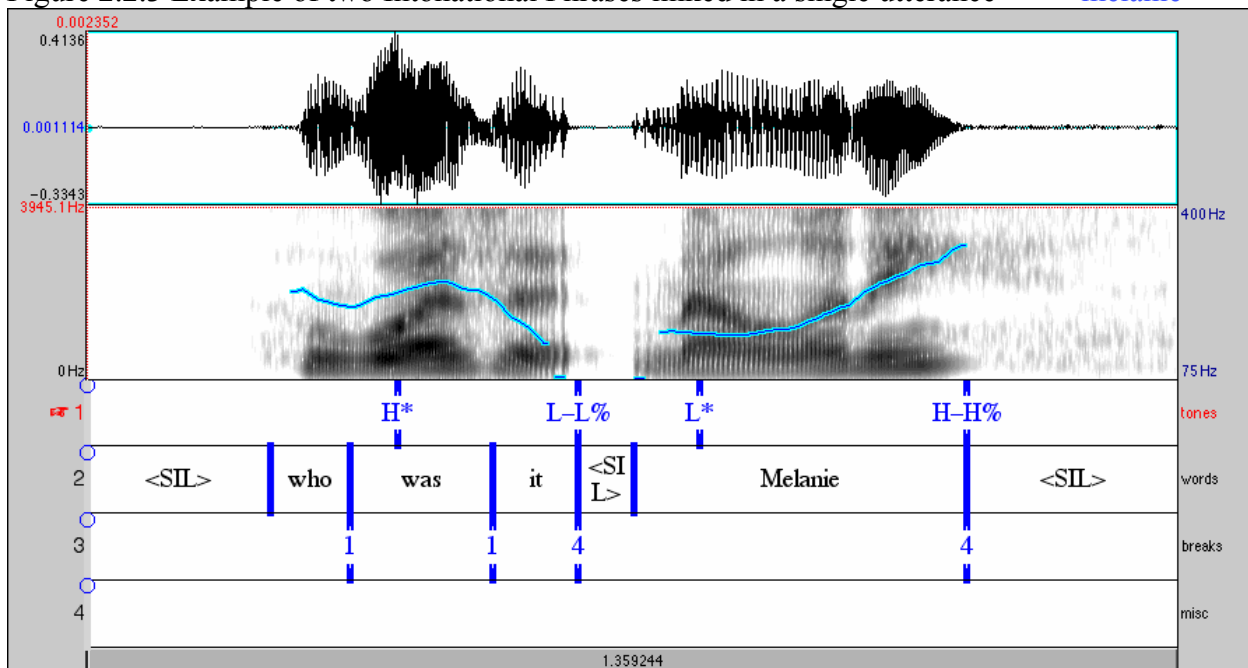
<africa>



Most of the examples shown here and earlier have only a single IP, so that the 4 at the end of the IP is also at the end of the utterance. In cases like this, disjuncture denoted by the 4 Break Index is between the last word of the IP and the silence that follows. Perhaps the idea of “disjuncture” is more comfortably reflected in the context of a sequence of intonational phrases, which

Figure 2.2.3 Example of two Intonational Phrases linked in a single utterance

<melanie>



frequently occurs in conversational or narrative speech contexts, but not so far in this tutorial. If there are several IP's in sequence, the 4 at the end of the each IP describes the disjuncture between the last word in the first phrase and the first word in the next phrase, as shown in the example <melanie.wav>. The words in the first part of this larger utterance (*Who was it*) group together into a single intonational phrase, separate from the word in the second part of the utterance (*Melanie*). The disjuncture between the two IP's, marked with the break index 4, is cued by the phrase accent (L-), boundary tone (L%) and relative lengthening on the final syllable of the first intonational phrase.

Inventory so far:

Tones:

H* high pitch accent

L* low pitch accent

L-L% low phrase accent, low boundary tone

H-H% high phrase accent, high boundary tone

Break indices:

0: word boundary erased

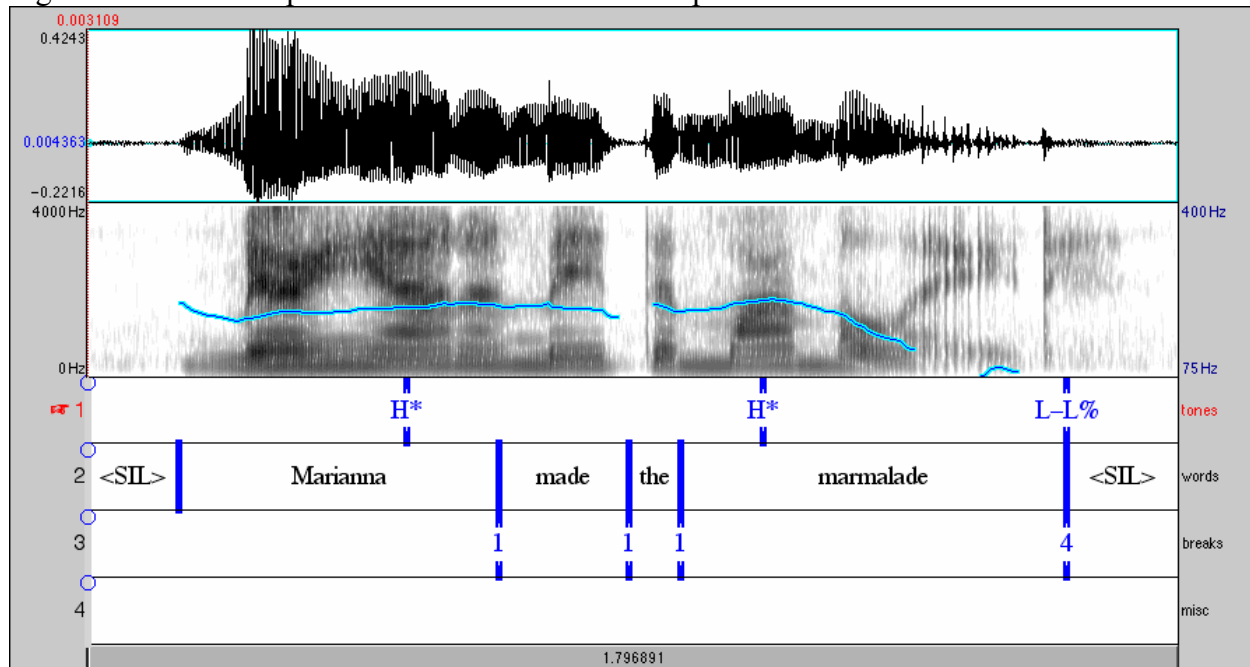
1: typical inter-word disjuncture within a phrase

4: end of an intonational phrase

2.3 Two accents in an intonational phrase

In the English prosodic system, every intonational phrase contains at least one pitch accent (refs) and many intonational phrases contain more than one pitch accent. For example, the following utterance, <marmalade1>, contains two H*'s, followed by the boundary tone sequence L-L%.

Figure 2.3.1 Two H* pitch accents in an intonational phrase <marmalade1>



In this example, the speaker produced the sentence *Marianna made the marmalade* with a particular prosody. As we'll see later, this is just one of the many ways this string of words could be produced. But let's look in some detail at the prosody of this one; in particular, let's look at the prominence of the various syllables. Notice, for example, that two of the syllables sound more prominent than other syllables in the phrase: the *-an-* of *Marianna* and the *mar-* of *marmalade*. These are the pitch accented syllables, each associated with H*, like many that we have seen already. For this phrase, just as for the examples in Section 2.1, the phrase-final tonal markers are a Low phrase accent followed by a Low boundary tone, i.e. L-L%. You may have the intuition that the last of the two H* pitch accents is the strongest one; this final pitch accent in the intonational phrase is called the nuclear pitch accent (for historical reasons; see references to British intonation studies in the 1950s by Halliday and others).

There are other syllables that sound quite strong, but not as strong as the pitch accented syllables: e.g. the *Mar-* of *Marianna*, the word *made* and the *-lade* of *marmalade*. We can contrast several levels of prominence, from the weakest (i.e. reduced vowels like the *a-* of *about*, the *-a-* in the middle syllable of *marmalade* or the *-o-* in *mammoth*, written as a schwa in the International Phonetic Alphabet) to the intermediately strong full vowel that lacks any special phrase level prominence (e.g. non-pitch accented vowels like the *Mar-* of *Marianna*, mentioned above) to full vowels with phrase-level (pitch-accented) prominence, like the second *-a-* of *Marianna* and the first *-a-* of *marmalade*. This illustrates an important fact: not every full vowel that is marked in

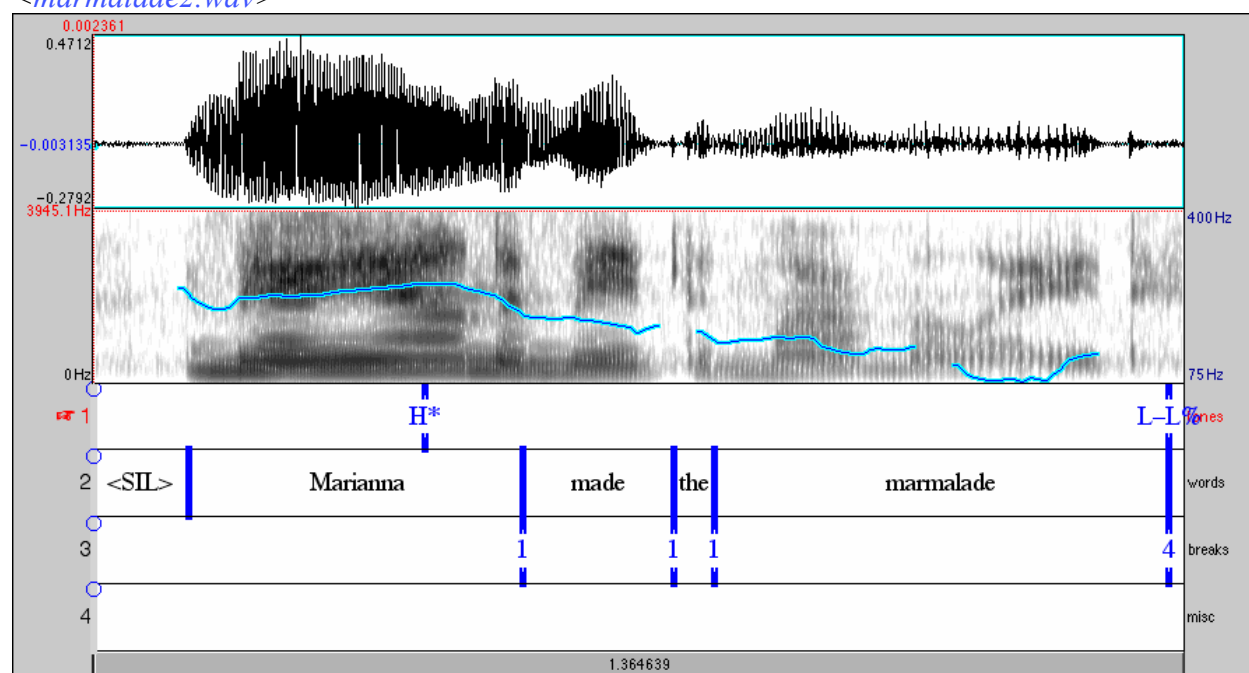
the lexicon as lexically stressed receives a pitch accent in every utterance, and full vowels that don't receive a pitch accent are less prominent, perceptually, than those that do. Also note that durational lengthening often accompanies pitch-accented syllables but can also occur at phrase boundaries. For example, the *-lade* of *marmalade*, in addition to being a full-vowel syllable, is also quite long because it is followed by the phrase boundary, but this extra duration is not associated with a pitch accent.

Finally, some of the syllables in this utterance are very weak: the *-ri-* and *-na* of *Marianna*, the word *the*, and the midword syllable *-ma-* of *marmalade*. These syllables have weak vowels, and sound much less prominent than either the pitch-accented or the non-accented full-vowel syllables.

This example illustrates the fact that the syllables of an utterance vary in their prominence, i.e. in how strong they sound compared to other syllables. We could rank the various degrees of syllable prominence, with pitch accented syllables highest, full-vowel lexically-stressed syllables lower, and reduced-vowel unstressed syllables lower still (Beckman and Edwards 1994). There is more to say about these issues, but we will save further discussion for a later section of the tutorial.

Compare the utterance (<marmalade1>) with <marmalade2> below, where only the syllable *-an-* in *Marianna* is prominent, or with the earlier examples in sections 2.1 and 2.2 that have only one pitch accent.

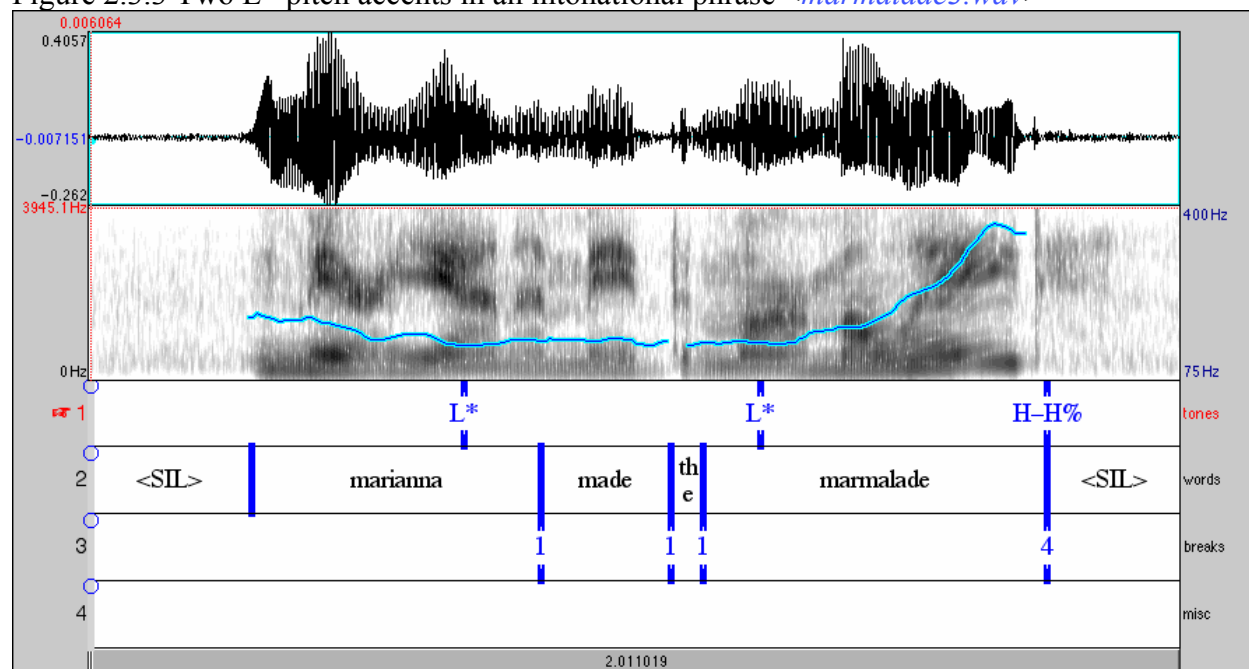
Figure 2.3.2 A single H* in an intonational phrase with the same words as Figure 2.3.1
<marmalade2.wav>



In the examples we have seen so far, an f0 peak or valley occurs on the accented syllable. This is often the case, but not always; we'll look at this question more closely later in the tutorial. But whether there is a sharp peak/valley and whether it is aligned with the accented syllable or not, the perception is that the syllable is more prominent than other syllables (or than it would be if it were not accented), and that this prominence is associated with an H* or a L* tonal event.

The next example (<marmalade3>) shows that more than one L* pitch accent can occur in a single intonational phrase, just as more than one H* can:

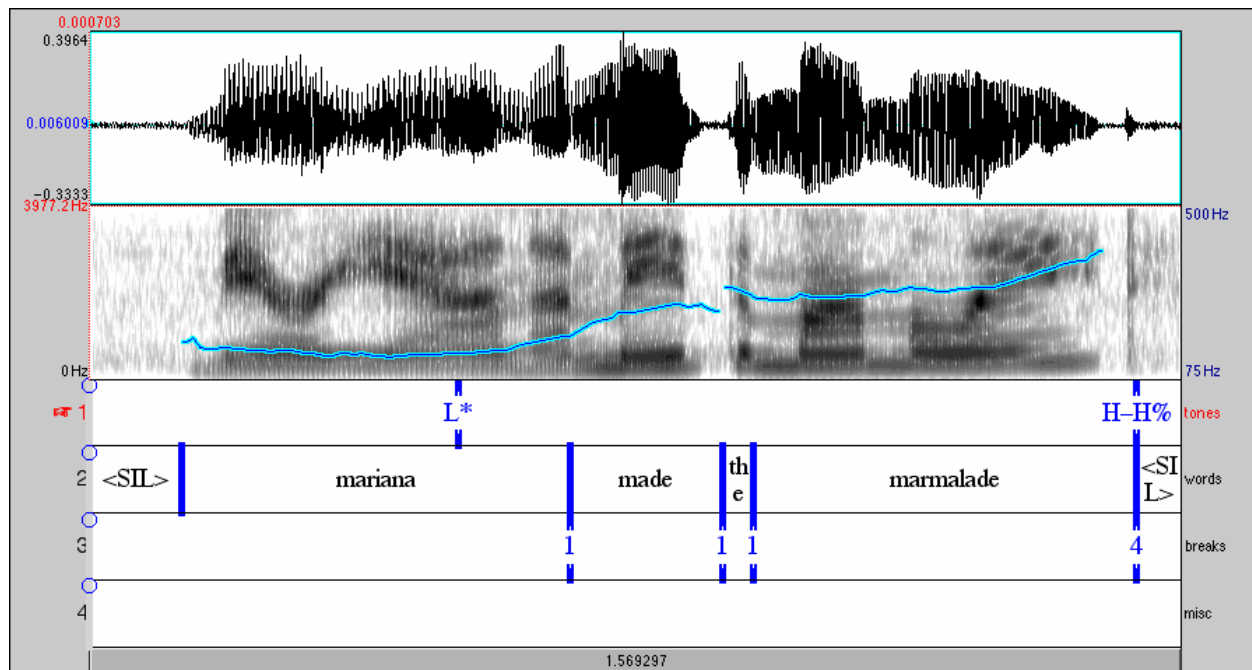
Figure 2.3.3 Two L* pitch accents in an intonational phrase <marmalade3.wav>



In this example, there are two prominent syllables: the *-an-* of *Marianna* and the *mar-* of *marmalade*, just as in the first example in 2.3. But in this case, both syllables carry low pitch accents and are labeled with L*. Because there are two L*'s in a row, the f0 appears to be somewhat flat. However, notice a slight rise between the two pitch accents and then a subsequent slight fall to the second L*. Later, we will give you suggestions on how to adjust the f0 display settings to more closely examine smaller f0 differences than those we've shown in these straightforward examples. This example in particular was chosen because it ends with H-H%: this provides a maximal contrast with the last L*, which makes it easier to recognize. The examples in sections 2.1 and 2.2 were chosen for similar, purely illustrative, reasons. But as we'll see later, H* and L* can occur with other final tones as well.

Compare the same sentence produced with a single L*, in <marmalade4>:

Figure 2.3.4 An example of an IP with a single L*. <marmalade4.wav>

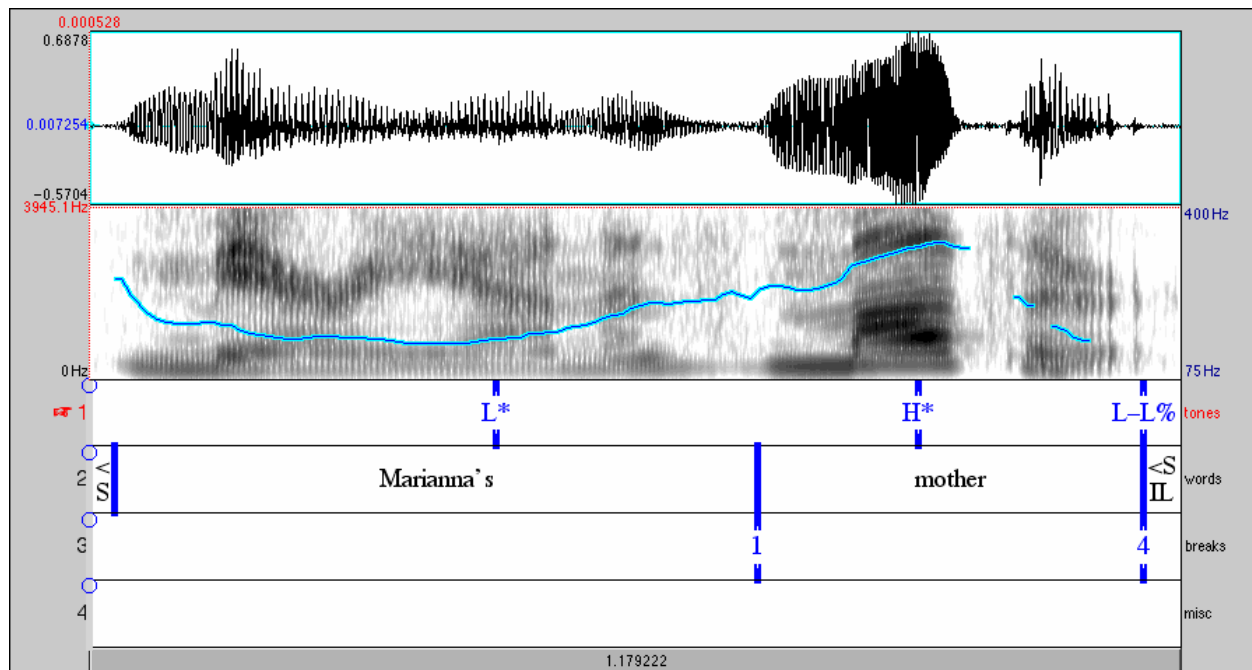


Here, there is only one pitch accented syllable; unlike in the previous example, the *mar-* in *marmalade* does not have an L*. The f0 rises in response to the H-H% at the end of the intonational phrase.

Combinations of High and Low pitch accents can occur within the same Intonational Phrase as well. In the following example, <mother>, both *-an-* in *Marianna* and *moth-* in *mother* are prominent syllables. However, the tonal marker for *-an-* is Low, and therefore labeled with an L*, while the tonal marker on *moth-* is High and marked with an H*. The f0 after the last accent indicates a Low phrase accent succeeded by a Low boundary tone (L-L%).

Figure 2.3.4 Example of an L* followed by an H* in a single IP

<mother.wav>



As more tonal markers are produced in an intonational phrase, the f_0 contour becomes more complex, because tonal elements can affect each other (see section 2.x+ for discussion). Another source of complexity is that the particular sound segments that make up the words of an utterance can sometimes have an effect on the f_0 contour. For example, some segments (like /p/, /t/, /k/, /s/, /f/) are unvoiced, which means that the vocal folds do not vibrate for some period of time, so the f_0 track is interrupted. You've seen this phenomenon in the example <afrika> in Section 2.2. Some segments are articulated in such a way that the vibration of the vocal folds is disturbed, so that rapid rises or falls may occur which are not part of the intended tone contour and would not occur if the same contour were produced on a different set of words. You may have noticed that in the *made the* portion on the examples *Marianna made the marmalade*, the f_0 disappears and seems to show local aberrations. A similar disturbance interrupts the f_0 track at the -'s and the -th- of *Marianna's mother*. Finally, some segments have a naturally higher f_0 than others (compare 'ee' with 'aw', 'ah' with 'm'). Learning to ToBI-label requires learning which aspects of the f_0 track are informative about the intonational contour that is intended by the speaker, and which aspects are due to other causes.

Introduced here:

Tonal patterns:

H* H* L-L%

L* L* H-H%

L* H* L-L%

Inventory introduced so far (no new elements were fully introduced in this section):

Tones:

H* high pitch accent

L* low pitch accent

L-L% low phrase accent, low boundary tone

H-H% high phrase accent, high boundary tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

4: end of an intonational phrase

2.4 Other full intonational phrase boundary tone combinations: H-L% and L-H%

So far, we've introduced just two intonational phrase boundary tone contours: H-H% and L-L%. We deliberately chose these contours to maximally contrast with the pitch accents that we were introducing (L* and H* respectively). Two other combinations are possible and both occur frequently: L-H% and H-L%. These boundary tone sequences, like L-L% and H-H%, can occur with any final pitch accent (e.g. L*, H*). Again, it is sometimes easier to distinguish which part of the f0 contour corresponds to the final pitch accent, the phrase tone and the boundary tone when the phrase tone contrasts with both the preceding nuclear pitch accent and the following boundary tone (e.g. as in the sequence H* L-H%), so we'll consider examples of these first.

2.4.1 Illustrative Examples: H* L-H% and L* H-L%

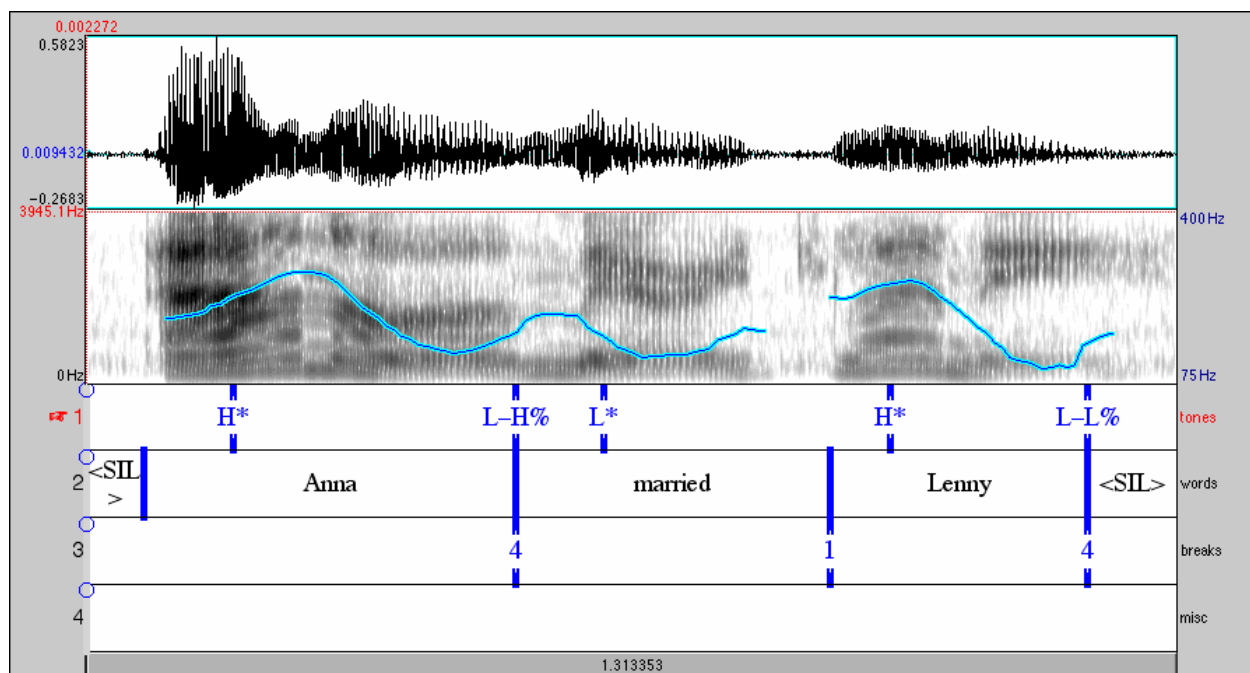


Figure 2.4.1 Example of H*L-H%

<annal>

In the example <annal>, the high pitch accent (H*) on *An-* in *Anna* is followed by a low phrase tone (L-) that rises again into a high boundary tone (H%). Even though the f0 is smooth across the word boundary, the changes in f0 and the lengthening of *-na* give the perceptual impression of a boundary between *Anna* and *married*. Notice that the second intonational phrase in this utterance is the now familiar L*H*L-L% (compare example <mother>). As we have seen, the Intonational Phrases are separated by the 4 Break Index. The juncture between *married* and *Lenny* is the default phrase-medial inter-word boundary and is labelled with a Break Index 1.

The next example of H*L-H% occurs on a single word intonational phrase <dan>. Note the rise in f0 on the vowel of *Dan* (H*) followed by a fall (L-) that ends in a small rise (H%).

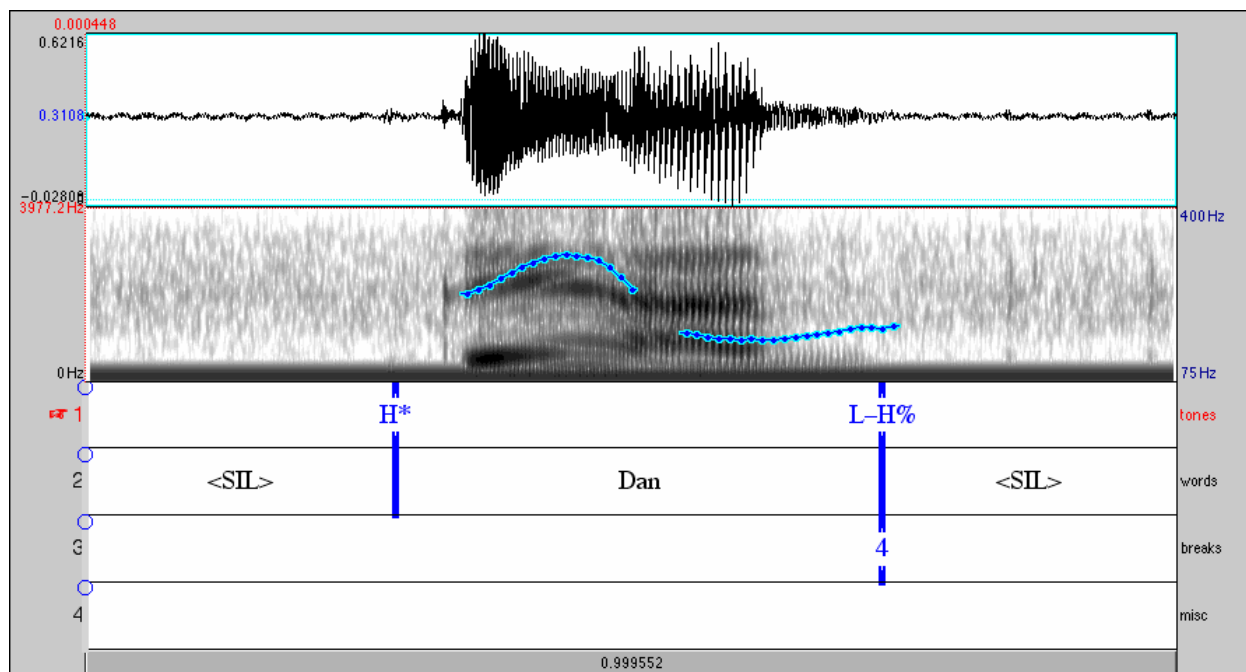


Figure 2.4.1 Example of H*L-H% on *Dan*

<dan>

Often, mixed phrase tone/boundary tone combinations (e.g. L-H% as shown in these two examples above, or H-L% described below) do not end in the more extreme high or low levels that are typical of H-H% or L-L% respectively. In this example, the L- makes it unlikely that the H% will rise to the top of the speaker's range. Without the extreme f0 excursions, other boundary cues, like the lengthening at the end of the last syllable in the intonational phrase, becomes more important for identifying the occurrence of the phrase boundary.

The last phrase tone/boundary tone sequence we'll discuss here is H-L%, which we show preceded by an L* (H* H-L% sequences occur as well). In the example <alejna1>, the L* on *-le-* (*Alejna* is pronounced 'uh-LAY-na') is followed by a non-pitch-accent-leading f0 movement up to an H- that is then followed by a slight fall (L%). Other examples of H-L%, including <banana2> below, do not end in a falling f0; although the phrase tone/boundary tone sequence is also an H-L%, the final f0 is nearly flat in a middle range. This pattern is more typical of an H-L% than a slight fall is.

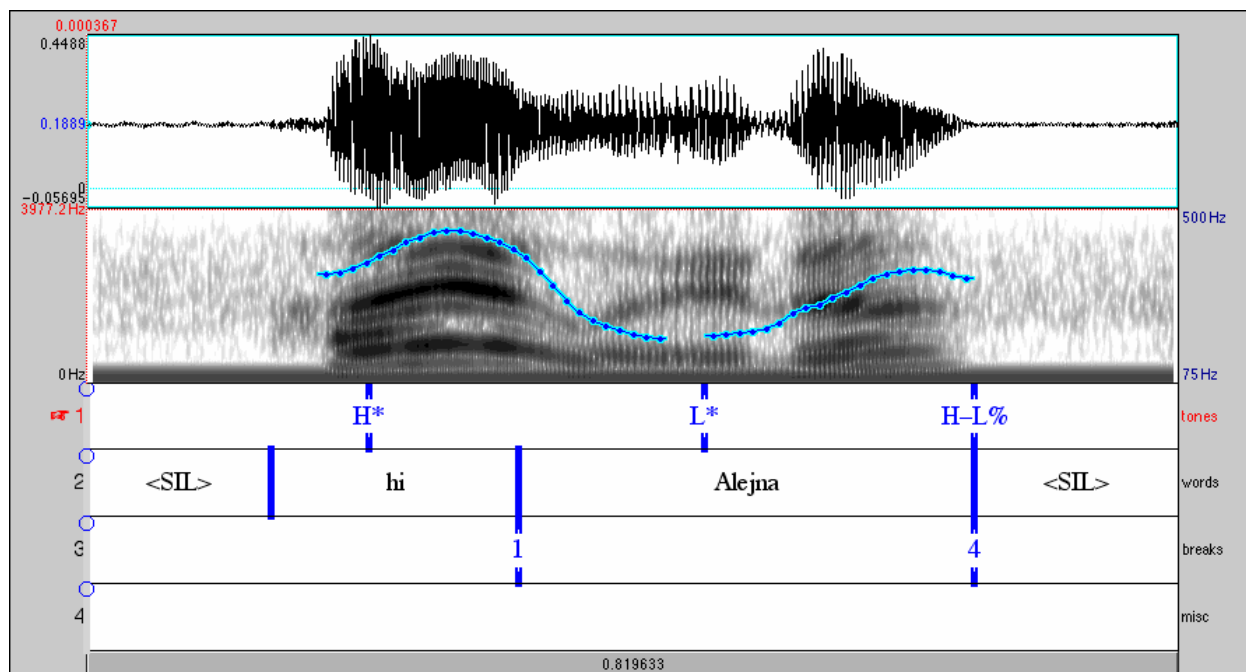


Figure 2.4.2 Example of L* H-L%:

<alejna1>

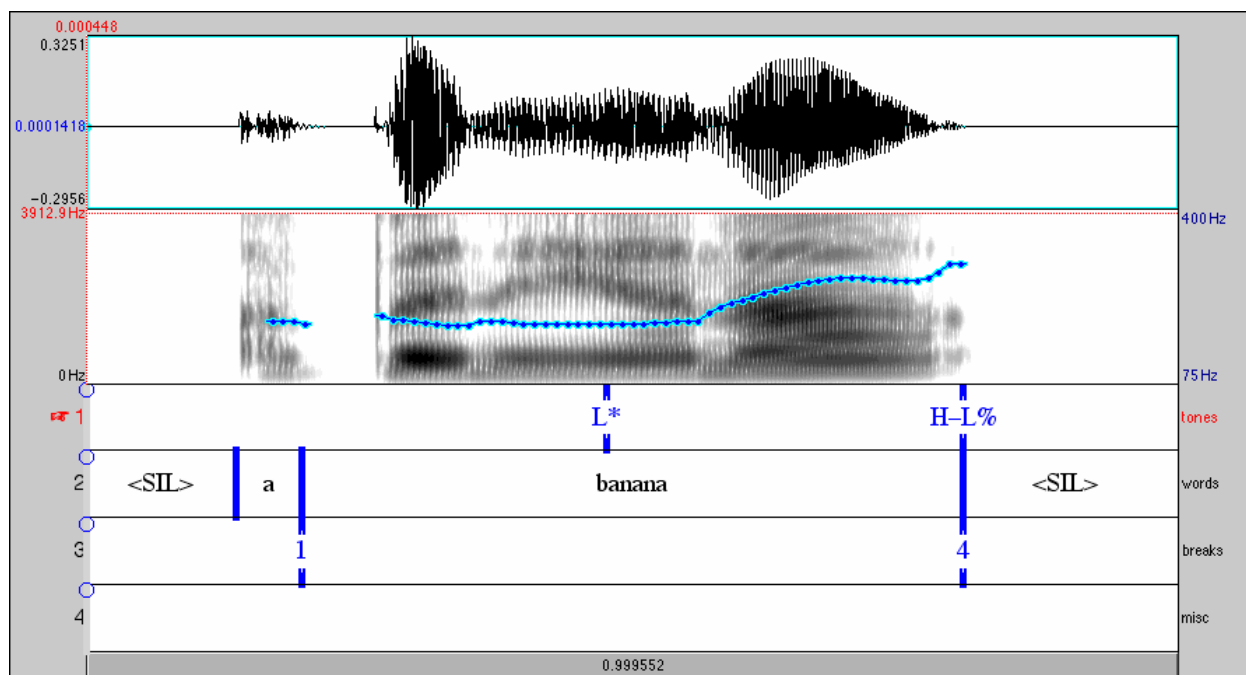


Figure 2.4.3 Example of L* H-L%:

<banana2>

Because the two tones (in these H-L% and L-H% examples) are being realized over a potentially short duration, sometimes only over the end of the final syllable, it is often difficult to see the short realizations of the final boundary tone. However, sometimes a labeler will get a perceptual impression even if it isn't visible in the F0 track, or can hear that it is not another category even if she can't hear or see a clear rise and fall.

The next example is cautionary: the pitch tracking software indicates a rise in f_0 at the end of the utterance. However, the pitch periods have become irregular, which often happens as a speaker ends an utterance, causing a recurring spurious small second peak in the waveform. The software that calculates the f_0 mistakes this small peak for a pitch period and determines the pitch period to be shorter; as a result, the f_0 frequency is calculated as higher than it should be. The best test in cases like this is to compare f_0 from earlier in the utterance with the end of the utterance by listening carefully for which sounds lower. The most reliable measuring tool in these cases is the listener's auditory system: listen carefully to whether it sounds as if the speaker's pitch were intended to go up at the end (specifying an H%) or to stay low (specifying an L%).

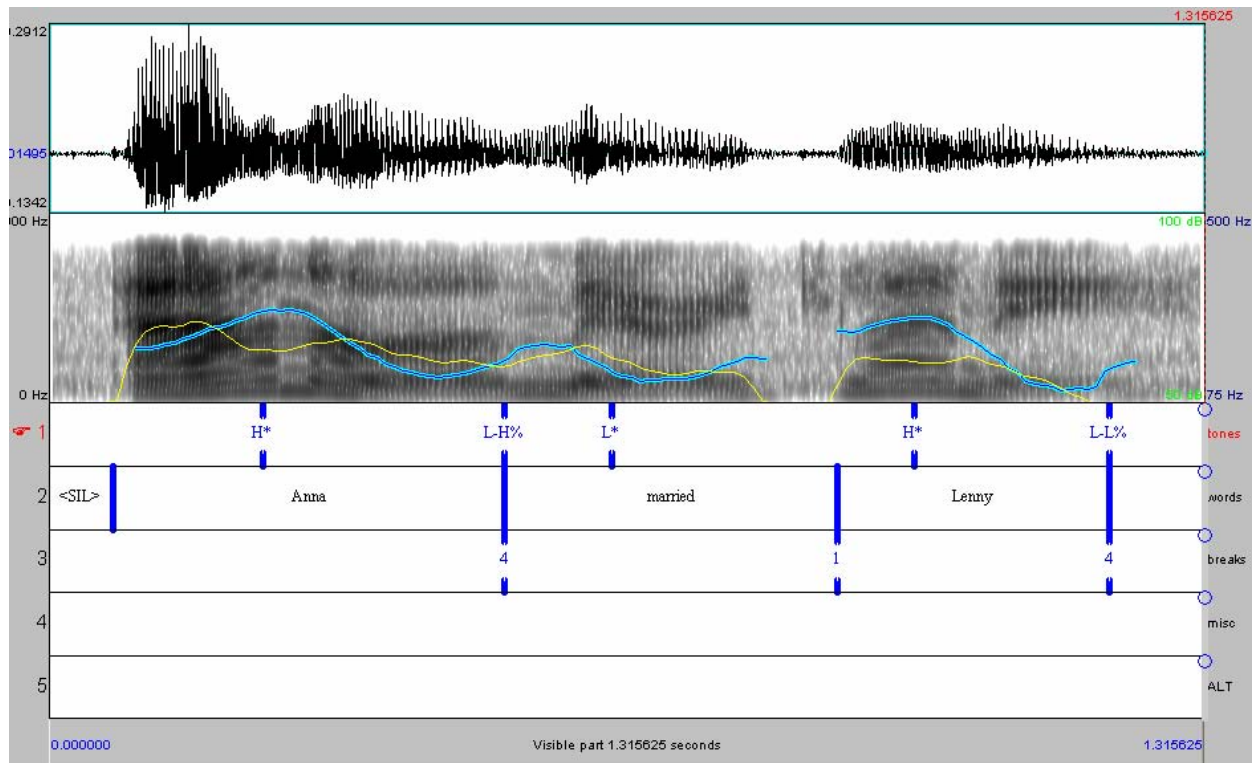


Figure 2.4.4 Example of H* L-L% with a spurious rise at the end Excerpt of <anna1>

Like the boundary tones introduced earlier, there is no constraint on what pitch accent can precede which of these boundary tones, so that more challenging sequences like L*L-H% or H*H-L% are also possible. The examples above, H* L-H% and L* H-L%, were introduced first in this tutorial because they are maximally contrastive, so it is easier to see the two different tones in the final sequence. Next, we'll show examples where the pitch accent choice does not provide maximal contrast.

2.4.2 Other combinations: L* L-H% and H* H-L%

L* and H* accents can be combined freely with various <phrase tone + boundary tone> sequences. For example, the contour produced on *banana* in the file <banana3> is L* L-H%.

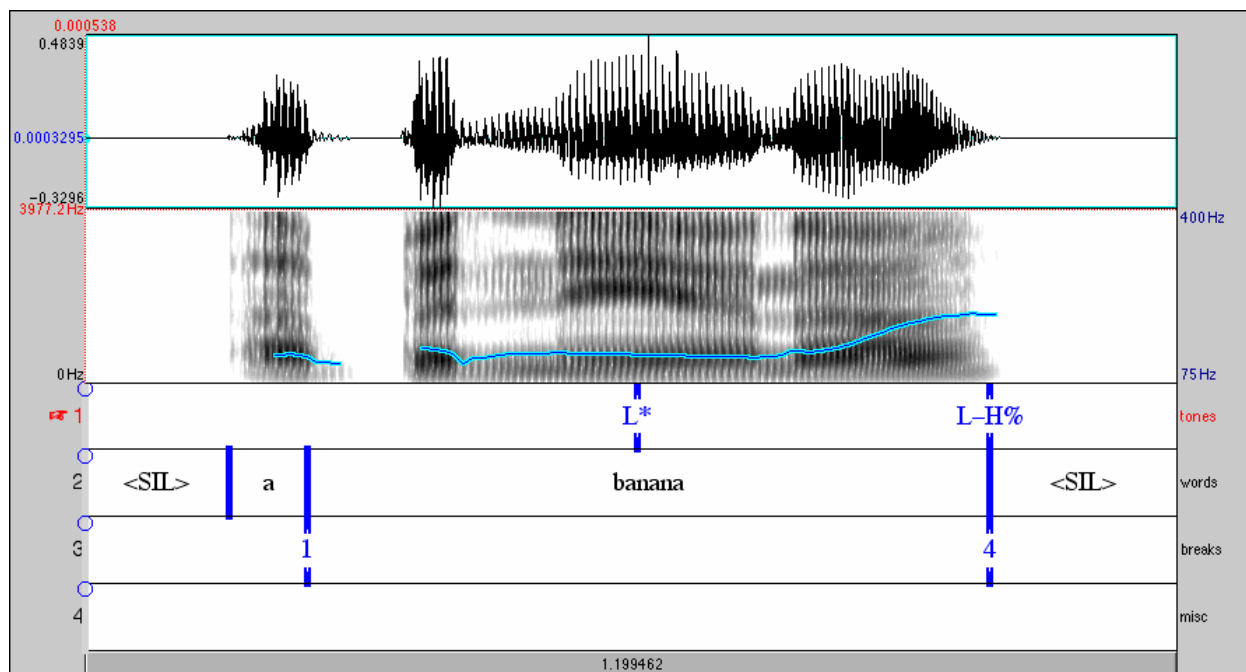


Figure 2.4.5 An example of L* L-H%

<banana3>

Compare the L-H% contour in <banana3> to both the L-L% and the H-H% in the following three renditions of *Marianna's money*. In the first (left-most) phrase the f0 rises to produce the H* on *-anna*, then falls for the L- phrase tone on *-na's money* until it culminates in a rise for the H% boundary tone on *-ney*. The H% tone is clearly higher than, e.g. the f0 in the L-L% of the middle rendition. On the other hand, the H% in this L-H% combination does not rise as high as it does in the last (right-most) rendition of *Marianna's money*, where the L* is followed by an H-H%. Here, the f0 rises very high in the speaker's range, as is frequently seen in productions of H-H%. Understanding these interactions between phrase tones and boundary tones, as well as changes in voice quality at the end of phrase and their effects on the visible f0 track, is an important part of learning to use the ToBI system.

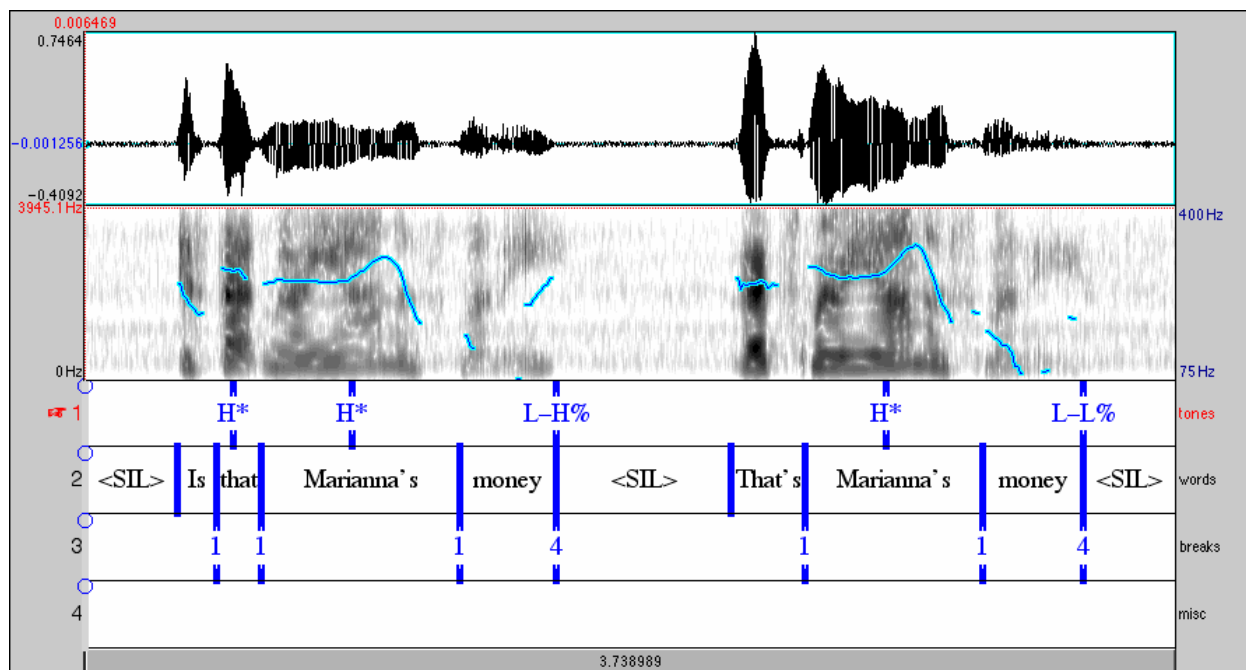


Figure 2.4.6 Examples of H* L-H% and H*L-L%, to contrast with H* H-L% in Figure 2.4.7
 <money1>

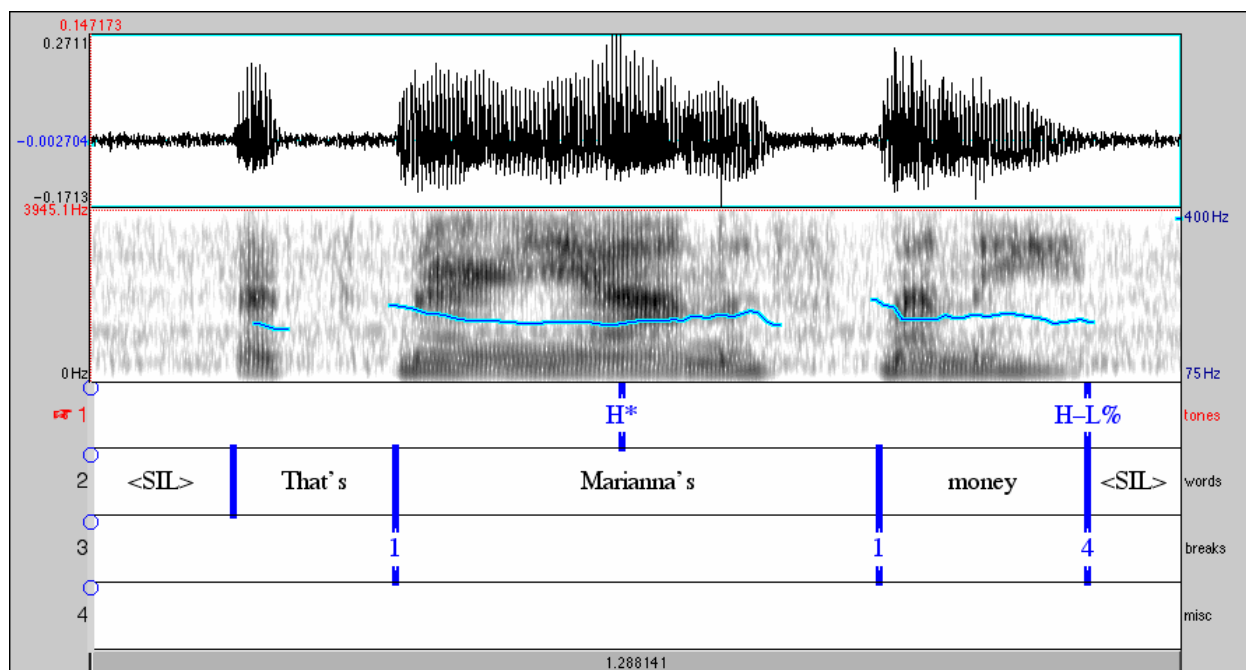


Figure 2.4.7 Example of H* H-L% <money2>

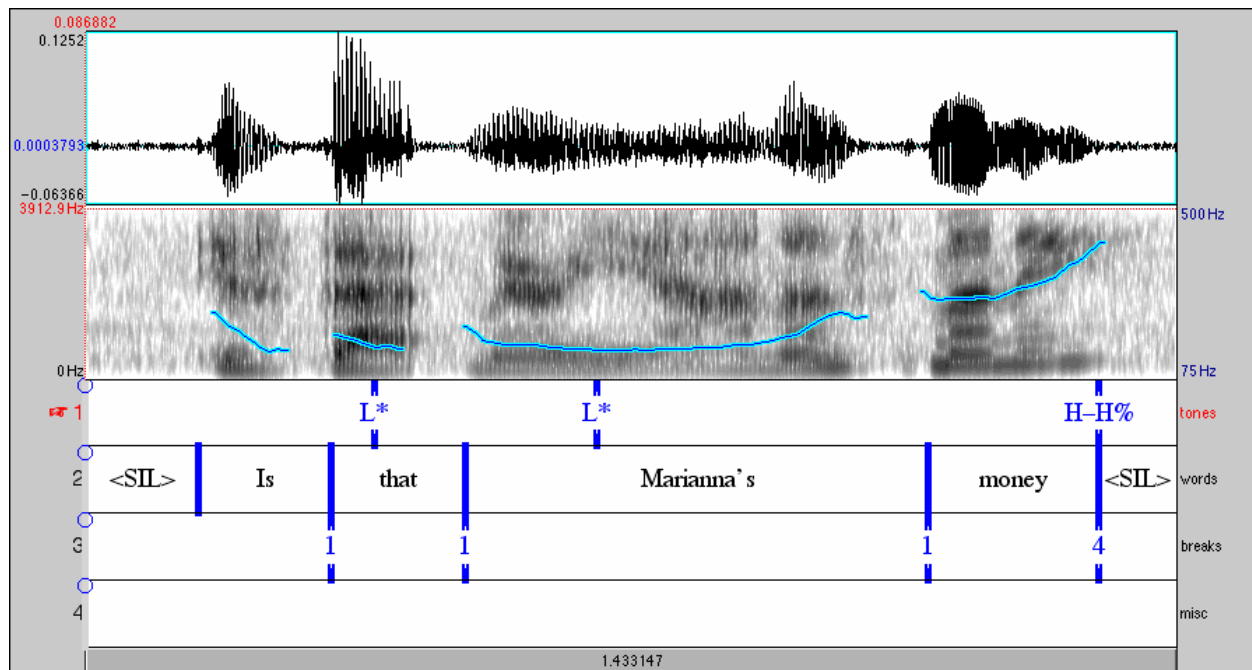


Figure 2.4.8 Example of L* L*H-H%

<money3>

Summary of ToBI labels introduced so far:

Tones:

H* high pitch accent

L* low pitch accent

L-H% low phrase tone, high boundary tone

H-L% high phrase tone, low boundary tone

L-L% low phrase tone, low boundary tone

H-H% high phrase tone, high boundary tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

4: end of an intonational phrase

2.5 Pitch Accents with more than one tone: Bi-tonal accents L+H* and L*+H

All pitch accents described so far have been produced with a single tonal element, either a High (H*) or a Low (L*) aligned with the accented syllable. There are other choices open to speakers however, to convey the prominence (greater relative salience) of words and syllables in an utterance. There are times where two tones, both a High and a Low, are associated with the same prominence-marking event. Such events are called bitonal pitch accents. This section will describe two bitonal pitch accents which have in common a Low tone followed by a High tone as part of the same pitch accent: L+H* and L*+H. The plus symbol (+) is used to indicate that the two tones are associated, and form a single unit: a complex (bitonal) pitch accent. The notational difference between these two pitch accent labels is that the star (*) symbol immediately follows the H symbol in one, and the L symbol in the other. This difference in notation reflects 1) a perceptual difference in which of the two tones lends more prominence to the pitch accent and 2) alignment characteristics of f0 movement in relation to the pitch accented syllable.

2.5.1 L+H* vs L*+H

The examples <amelia1> and <amelia2> illustrate the difference between the L+H* and L*+H. In both of these files, the same speaker produces an L H L H sequence. However, the two utterances have two different intonational contours, where the labels differ only by the location of the star (*) symbol. That is, both contain a bitonal pitch accent on the same syllable (the –*mel-* of *Amelia*), followed by a Low phrase accent / High boundary tone sequence (L-H%), the same break labels, and the same words (*Amelia knew him*), but in one case the prominence is associated with the H* and in the other case with the L*.

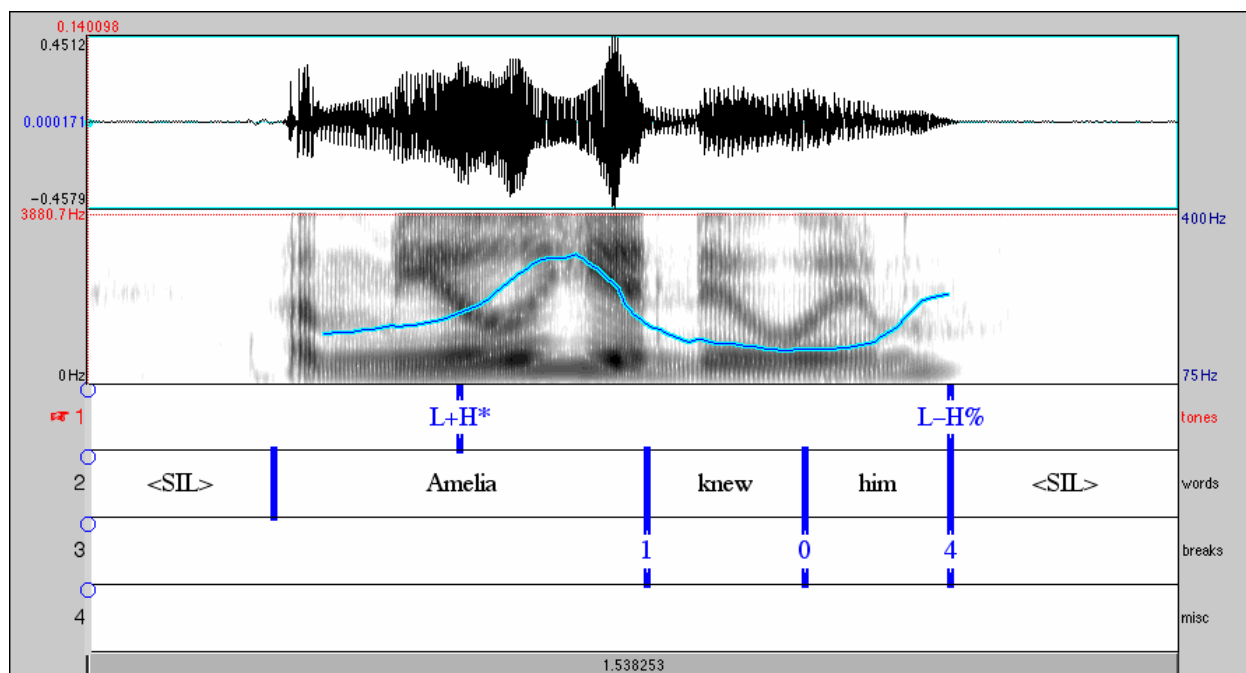


Figure 2.5.1 L+H*

<amelia1>

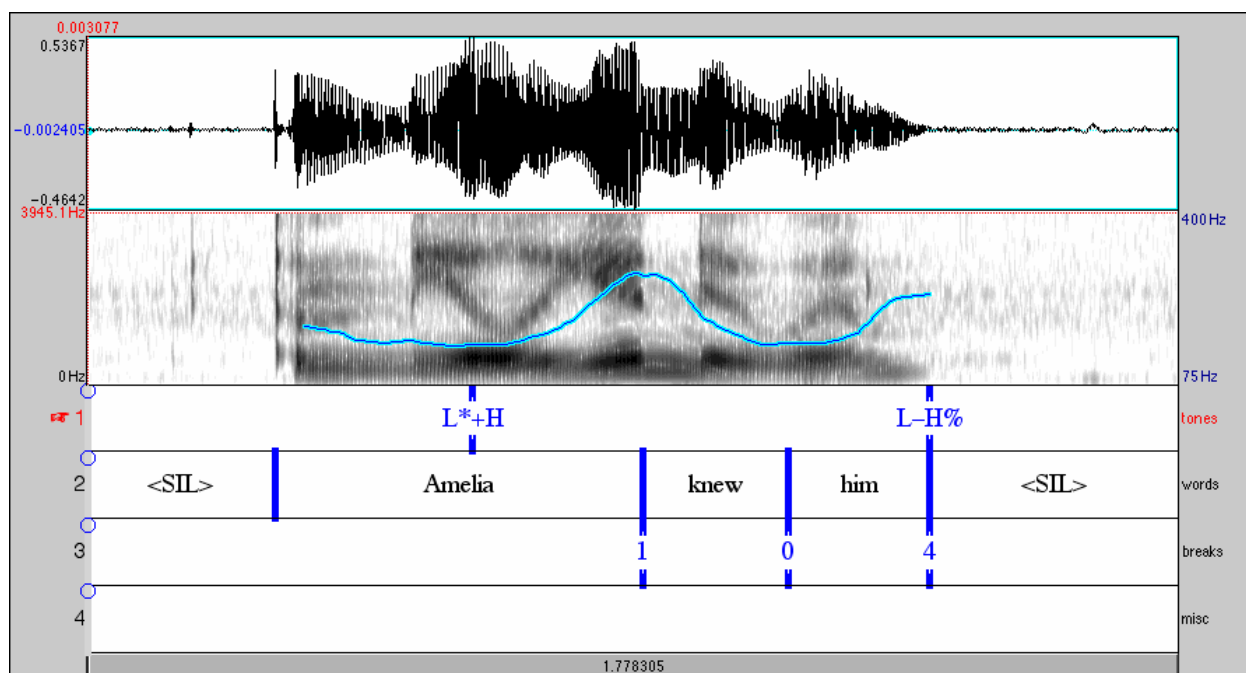


Figure 2.5.2 L*+H

<amelia2>

In both files, there is a stretch of low f0 followed by a rise to a high f0 peak, then a fall to another low stretch, and finally a rise to a higher f0. Look closely at how the pitch track aligns with the syllables of the word *Amelia*. In the L+H* example, the H tone aligns with the prominent syllable *-mel-*; the f0 rises through the prominent syllable, with the peak occurring at the end of the syllable. (Note how useful the spectrogram is for figuring out which syllable the f0

movement occurs on, especially when combined with listening to selected time segments of the utterance.) This alignment of the H tone with the prominent syllable corresponds to the perception that pitch accent is one of a High tone prominence—the “starred” tone in this bitonal pitch accent is the H (L+H*). In the L*+H example, the tone aligning with the prominent syllable *-mel-* is Low: while there is a peak corresponding to a High tone, this H tone occurs at the very end of the non-prominent syllable *-ia* (at the end of *Amelia*). The alignment of the L tone with the prominent syllable corresponds to the perception that pitch accent is one of a Low-tone prominence—the “starred” tone in this bitonal pitch accent is the L (in an L*+H). We’ll see later that f0 peaks and valleys do not invariably occur on the accented syllable; here we simply illustrate that the L*+H / L+H* contrast is generally reflected in peak alignment differences.

2.5.2 Bitonal vs single-tone pitch accents

The bitonal pitch accents L+H* and L*+H differ from the single-tone H* and L* accents by virtue of Tone events that precede or follow the (starred) H or L target of the pitch accent. Specifically, the L+H* differs from the H* primarily by a more substantial rising pitch movement leading up to the H* target, i.e. the presence of a preceding L target. The L*+H differs from the L* primarily by a rising pitch movement that follows the L* target, i.e. the presence of a following H target.

2.5.2.1 L* vs. L*+H

The difference between the single-tone L* pitch accent and the bitonal L*+H pitch accent is highlighted in the files <amelia3> and <amelia2>. These two files, like the pair of files discussed above, have labels that differ only in the type of pitch accent, here, L* and L*+H. In the L*+H example, shown above in figure 2.5.2, the Low target of the prominent syllable (the *-mel-* of *Amelia*) is followed by a rise to a peak, corresponding to the +H portion of the pitch accent. (The subsequent Low and the final rise of the contour are due to the Low phrase accent and High boundary tone, L-H%). The example with the single-tone L*, <amelia3>, also shows a low f0 on the pitch-accented syllable (again, the *-mel-* of *Amelia*). However, the f0 stays low and flat following the prominent syllable all the way through the end of the word *Amelia*, and then through the word *knew*, until the rise to the H% boundary tone at the end of the word *him*. This difference between the two files reflects the difference between L* L-H% and the bitonal-containing sequence L*+H L-H%.

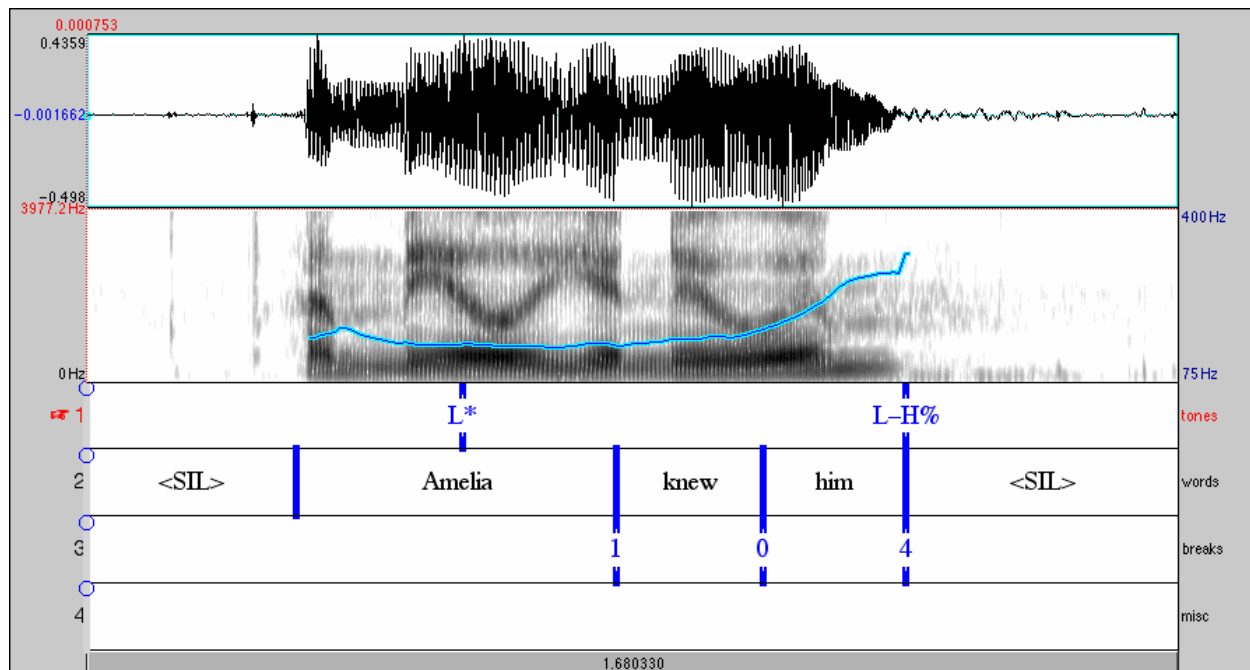


Figure 2.5.3 L*

<amelia3>

2.5.2.2 H* vs. L+H*

In the file <amelia4> the now familiar contour H* L-L% is contrasted with a contour containing the bitonal pitch accent L+H*. In both cases, the prominent syllable, the *-mel-* of the word *Amelia*, has a high tone associated with it, and end of the two contours is very similar, as the pitch falls from the high tone into the L-L%. The difference between the two contours is apparent in the pitch events leading up to the peak associated with the High tone of the prominent syllable. In the second intonational phrase of the file, with the tone labels L+H* L-L%, notice the sharp rise in pitch at the beginning of the vowel of the prominent syllable, *-mel-*, towards the peak of the High tone. The f₀ at the beginning of the word, while starting at a mid-range, falls through the word-initial vowel *A-* and the following /m/ to a fairly low level. This Low tone is the L of the bitonal pitch accent L+H* in this example. The first intonational phrase in the same file <Amelia> gives a contrasting example of the single-tone H* pitch accent on the same word. Compared to the L+H* version, the rise into the peak of the prominent High tone is gradual from the word onset. While there appears to be a slight fall of the f₀ into the /m/ of *Amelia*, this most likely a segmental effect. In any case, the f₀ before the H* is not as low as it is for the L+H*.

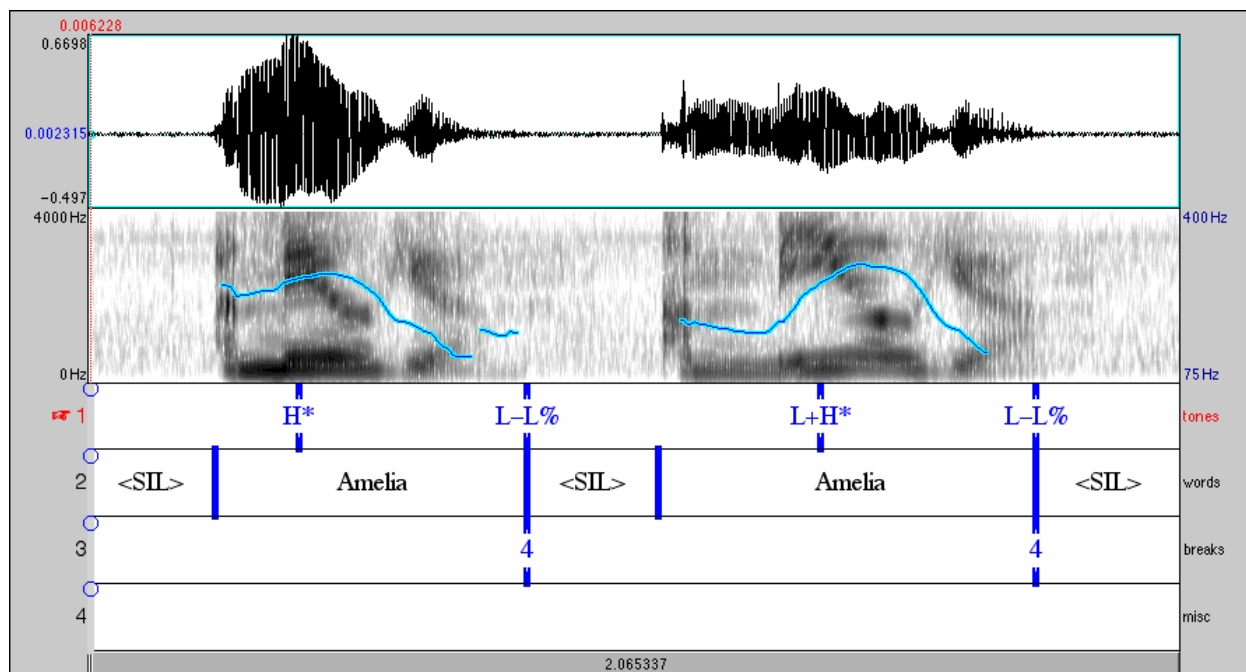


Figure 2.5.4 H* and L+H*

<amelia4>

The file <marmalade5> shows another example of the L+H* L-L% contour, this time on the text *Marianna made the marmalade*. Compare this to the earlier example <marmalade2> which has the tone labels H* L-L%. In both files, <marmalade2> and <marmalade5>, the pitch accent is aligned with the prominent syllable, orthographically the second -a- of *Marianna*. In <marmalade5>, with the bitonal pitch accent L+H*, the pitch track shows a gradual fall from the beginning of the word *Marianna*, reaching a low in the -ri- syllable, then a sharp rise into the High tone of the prominent syllable. In <marmalade2>, with the H* pitch accent, the rise to the peak of the H tone on the pitch-accented syllable is more gradual, starting from the beginning of the intonational phrase and across the first two syllables of the word. (The slightly higher f0 at the very beginning of the word *Marianna* is due to effects of the segment /m/, which the labeller should try to ignore in interpreting the tones of the intonational contour. Since the labeller can generally hear the intended contour “through” these segmental effects, this is an example of how listening trumps looking in labelling.)

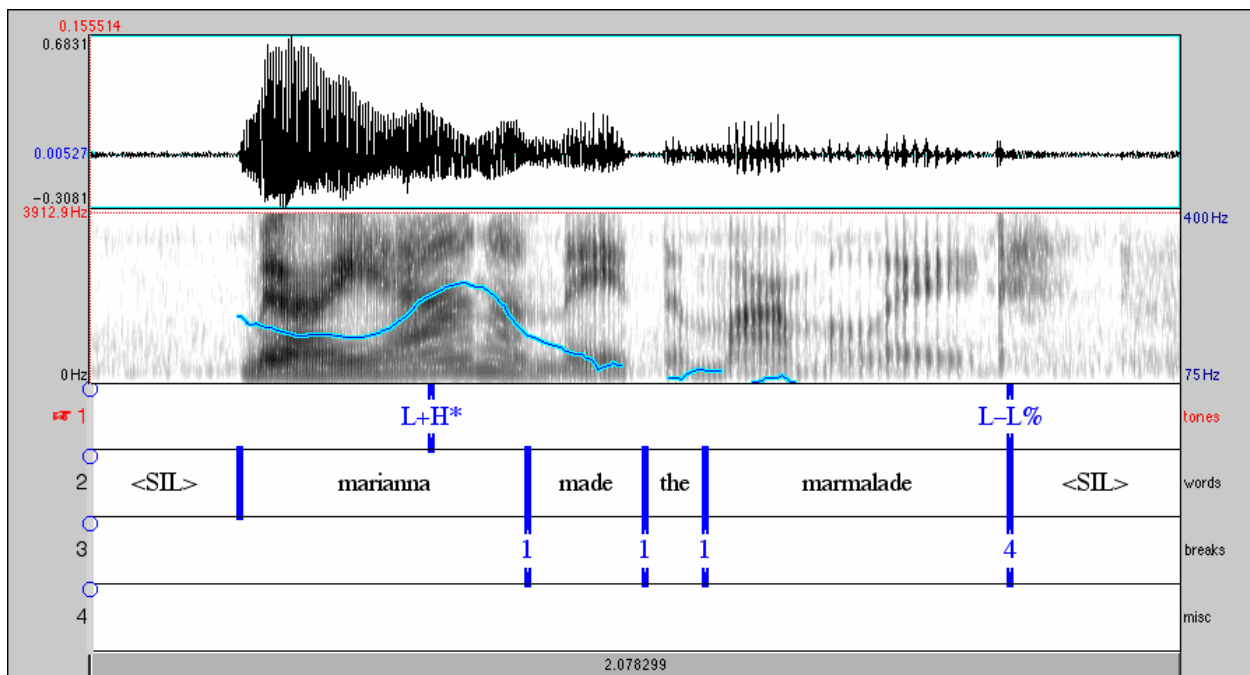


Figure 2.5.5 L+H* L-L%

<marmalade5>

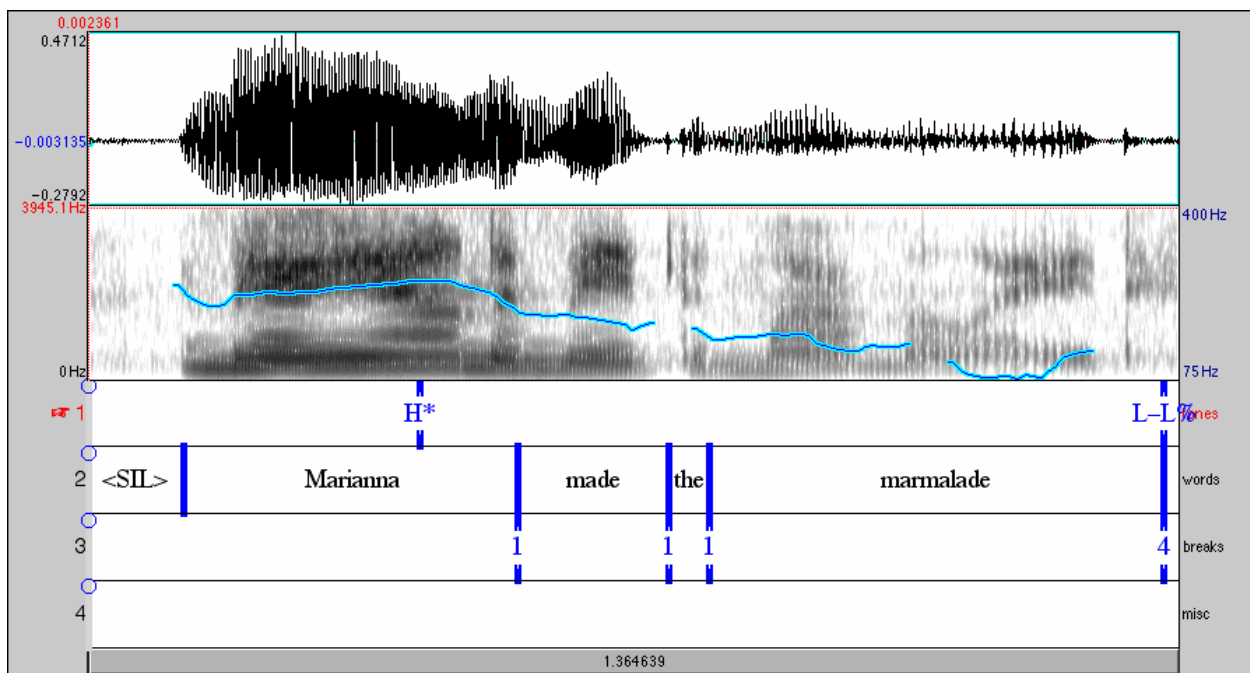


Figure 2.5.6 H* L-L%

<marmalade2>

Pitch Accent and Meaning

Speakers often use the L+H* pitch accent in contexts of contrast, although its use is not limited to such cases. For example, when saying the word sequence *Marianna made the marmalade*, the contour with L+H* might be used in a context where the speaker is trying to make it clear that the person who made the marmalade was Marianna, as opposed to some other person. In contrast, the example with a H* pitch accent could be expected in response to a question about who made the marmalade:

*L+H**, as in <marmalade5>

Speaker A: Bob made the marmalade.

Speaker B: (No.) Marianna made the marmalade.

*H**, as in <marmalade2>

Speaker A: Who made the marmalade?

Speaker B: Marianna made the marmalade.

However, it is important to remember that both contours can be used in a variety of contexts, and a specific context will not necessarily lead all speakers to select the same intonation contour.

2.5.2.3 Constraints on the use of the L+H* label

The use of the label L+H* is constrained to places where a low f₀ cannot be accounted for by some other tonal event. For example a rise from a Low tone to a High-toned prominent syllable can sometimes be accounted for by a preceding L-L% phrase accent/boundary tone combination, or a preceding L* pitch accent. Cases of L+H* are clearest when there is at least one non-pitch-accented syllable preceding the prominent syllable, so that the L tone of the pitch accent can be realized on preceding syllable(s). Such is the case in the examples discussed above. However, when the pitch accented syllable is the first syllable of an Intonational Phrase, and there is less clear evidence of an initial Low tone (i.e., the f₀ rise is small), the conventions of ToBI labelling prescribe using the simple H*. For example, in the file <anna2>, the tone on the pitch-accented syllable *An-* of *Anna* in the first Intonational Phrase of the file is perceptually quite similar to the L+H* on the *Le-* of *Lenny* in the second IP. However, the pitch accent labelled on *Anna* is H*, because there is not sufficient evidence of a Low tone to justify the use of L+H*.

You may have noticed the pitch track at the end of the word *Lenny* in this file <anna2> shows strange squiggles and spikes after the initial smooth fall from the peak into the L-L% phrase accent/boundary tone sequence. This region corresponds to the place in the sound file where the speaker's voice growls or sounds irregular or creaky, which is a common occurrence when speakers produce a pitch that is at the lower end of their pitch range, such as with L-L%. In these regions, pitch periods come at irregular intervals, so there is no regular period for the f₀ tracker to track. This irregularity in the pitch periods produces irregularity in the pitch track, i.e. a pitch tracking error, which is a reflection of the difficulty of measuring f₀ in regions where a speaker's

voice becomes creaky. Interestingly, you may find that you can nevertheless get a perceptual impression of falling f0 in such regions.

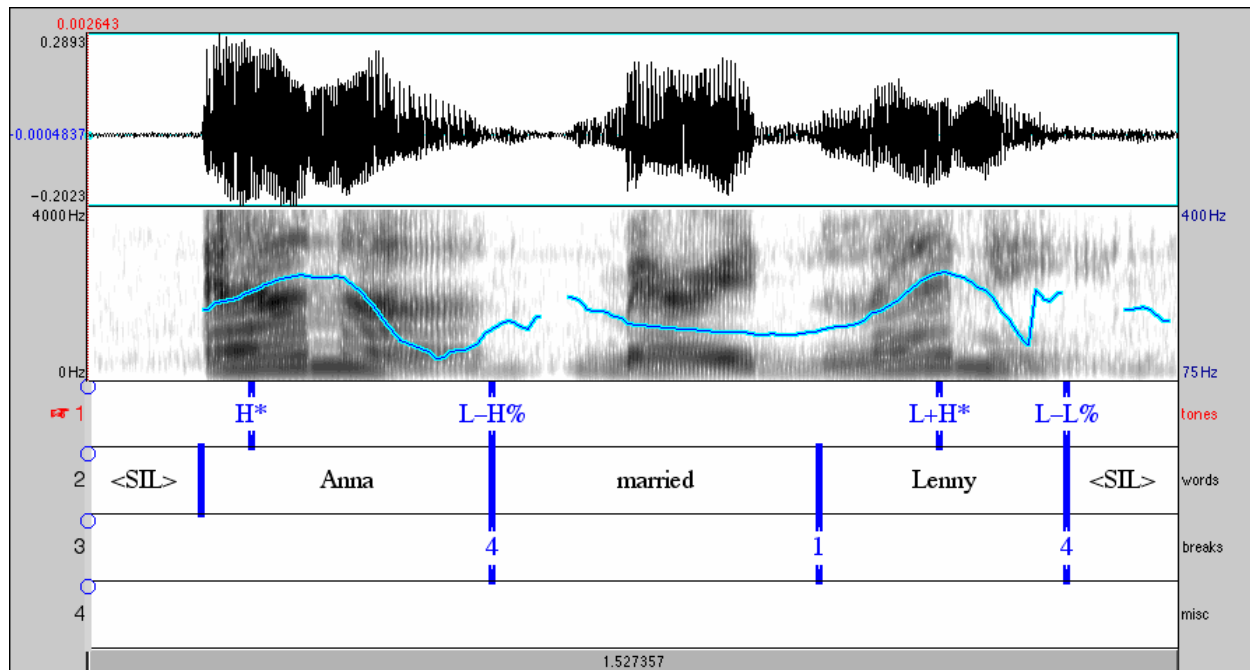


Figure 2.5.7 H* vs. L+H*

<anna2>

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal low tone with high tone on accented syllable

L*+H: bitonal high tone with low tone on accented syllable

L-L%: low phrase accent, low boundary tone

H-H%: high phrase accent, high boundary tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

4: end of an intonational phrase

2.6 Relationships among pitch accents within a phrase: !H*

As we have seen, Intonational Phrases frequently contain more than a single pitch accent. In previous sections we have seen four pitch accent labels: the two single-tone accents (H* and L*) and two bitonal accents (L+H* and L*+H). Theoretically, there are no constraints within the ToBI system on how these pitch accents may combine in an Intonational Phrase: any pitch accent label of the set seen so far may precede or follow any of the pitch accent labels in that same set. For example, H* may be followed or preceded by another H*, or by L* or L+H* or L*+H. In fact, as we will see later, multiple prenuclear pitch accents are possible within a single Intonational Phrase.

However, the !H* ('downstepped H star') pitch accent label is dependent on the immediately preceding labels. It indicates that there is a specific tonal relationship between the prominence labelled !H* and the preceding pitch accent. Specifically, a downstepped H pitch accent indicates that the tone of the prominent syllable is realized by a perceptually lower f₀ than that of an immediately preceding High tone: the tone has 'stepped down' from the preceding High. While the !H* is realized by a lower pitch than the preceding High, it is distinct from the L* pitch accent, which is characterized by a pitch excursion down towards the bottom of the speaker's pitch range (for that utterance). The !H* does not necessarily approach the bottom of the speaker's range. For example, in <banana4>, the H* on the *–noth–* syllable of *another* is followed by !H* on the *–nan–* of *banana*. The phrase ends with a Low phrase tone and a Low boundary tone (L-L%), which are realized by additional lowering of the f₀ in the final syllable (*–na* of *banana*). Note that the pitch track shows what looks to be a rise in the f₀ at the very end of *banana*, but this is a pitch-tracking error; it occurs well after the regular vertical striations in the spectrogram (corresponding to regular vowel-like pitch periods in the final syllable) have ended.

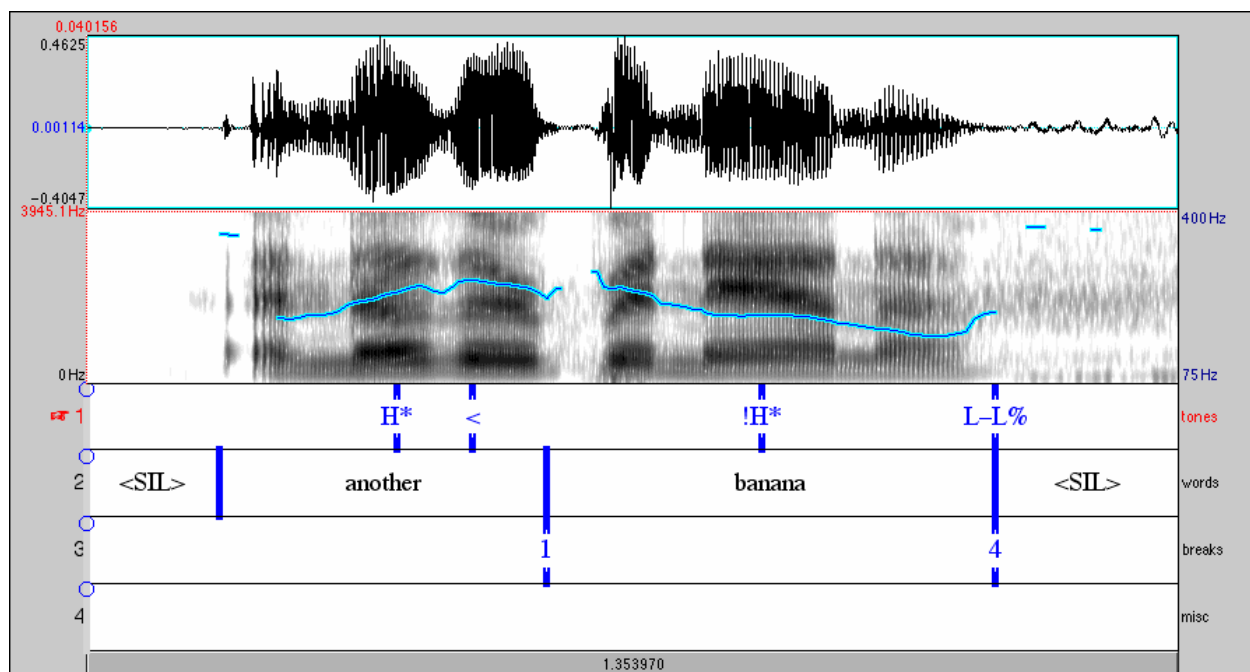


Figure 2.6.1 H* !H* L-L%

<banana4>

The example <clean_slate> below shows another instance of the same tone sequence (H* !H* L-L%), this time realized on two single-syllable words, *clean slate*. Here, the pitch track shows a nice ‘step down’ from the word *clean*, labelled with H*, to the word *slate*, labelled with !H*. (The audible fall in pitch into creaky voice at the end of the vowel in the word *slate*, which is the realization of the L-L% phrase tone-boundary tone combination, is not captured by the pitch tracker, and the voiceless release noise of the final /t/ is without any pitch track at all.)

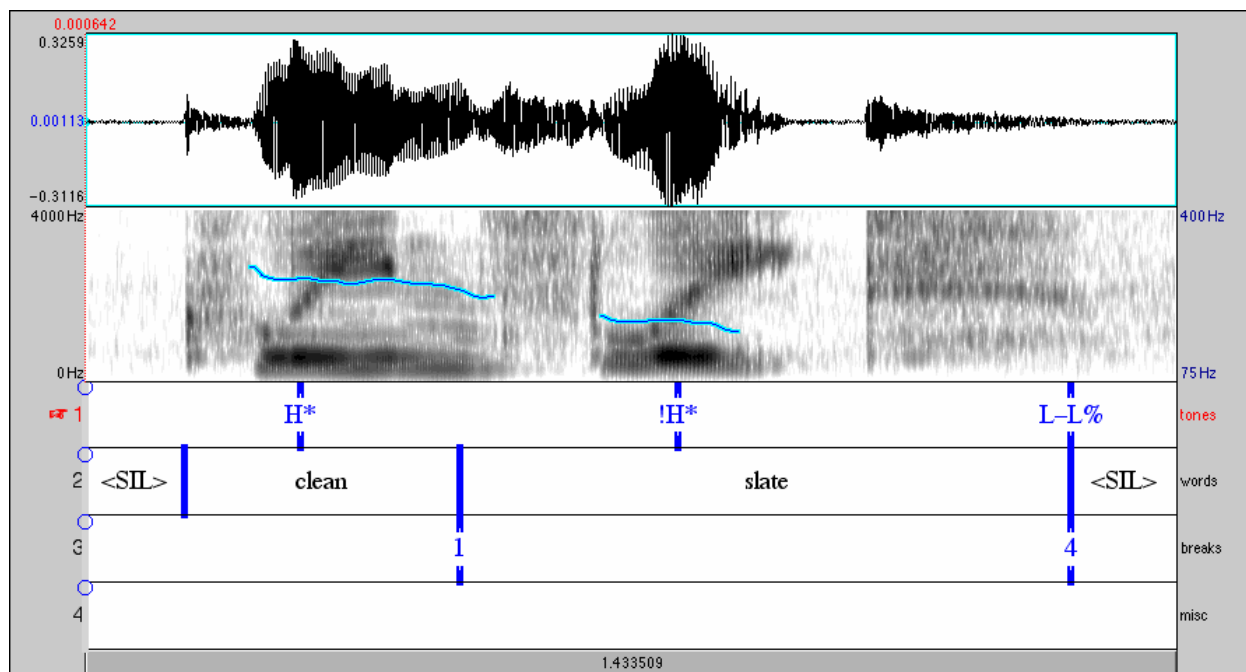


Figure 2.6.2: another example of H* !H* L-L% <clean_slate>

The !H* label may follow any of the pitch accents which contain an H target, such as H*, L+H* or L*+H. The !H* label will therefore never immediately follow a (single-tone) L* pitch accent, as that pitch accent has no High tone target. The !H* label will likewise never be the first pitch accent in a phrase.

In general, pitch accents labelled !H* are realized in the middle of the speaker's pitch range for that Intonational Phrase, but in this system they are not considered 'mid' tones. Like L+H* and H*, the !H* is a pitch accent for which the prominence is considered to be signalled by a High tonal target, which is why the H symbol is used with the star symbol. As with the other pitch accents with H* targets, the actual realization of the High tone in a !H* can vary quite a lot (i.e. speakers can speak in a relatively low pitch range or a high pitch range, but still produce High tones). This is particularly apparent in cases where one !H* pitch accent is followed by another !H*. The second !H* means that the target of this prominence is lower relative to the target of the first !H*.

For example, in *Figure 2.6.3* <private_ryan1> (*Saving Private Ryan*) the last prominent (pitch-accented) syllable of the phrase (the *Ry-* of *Ryan*) is produced with a lower f₀ than that of the previous prominence (on the *Pri-* of *Private*), which is in turn lower than the f₀ of the first prominent syllable of the phrase (the *Sav-* of *Saving*). This relationship in the pitch height is reflected in the use of the !H* labels on the last two pitch accents of the phrase, giving the tone sequence H* !H* !H* L-L%. Notice that the pitch range (50-200 Hz, indicated on the right vertical axis in the f₀ pane) for this speaker is much more compressed than the range seen in other examples. It is useful to adjust the range for each speaker in order to see the f₀ excursions. If the range selected for display is too large (e.g. 75-400 Hz for this example), then the f₀ may appear nearly flat, as is shown in *Figure 2.6.4*, below.

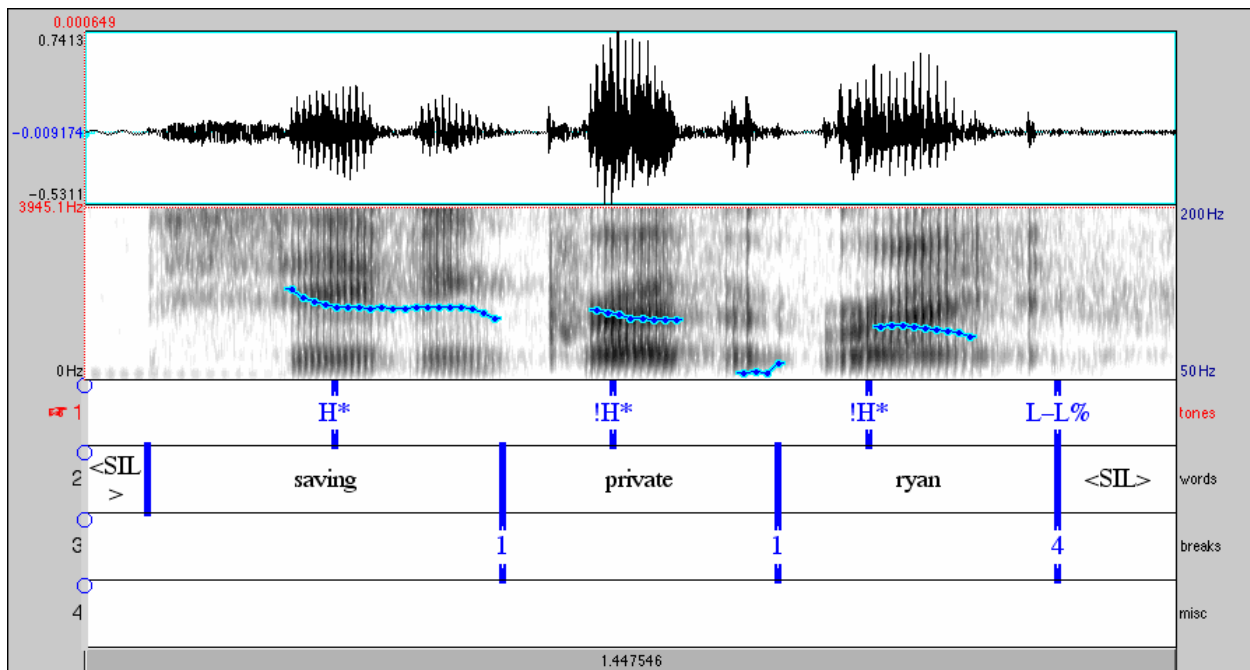


Figure 2.6.3a Two sequential !H* pitch accents

<private_ryan1>

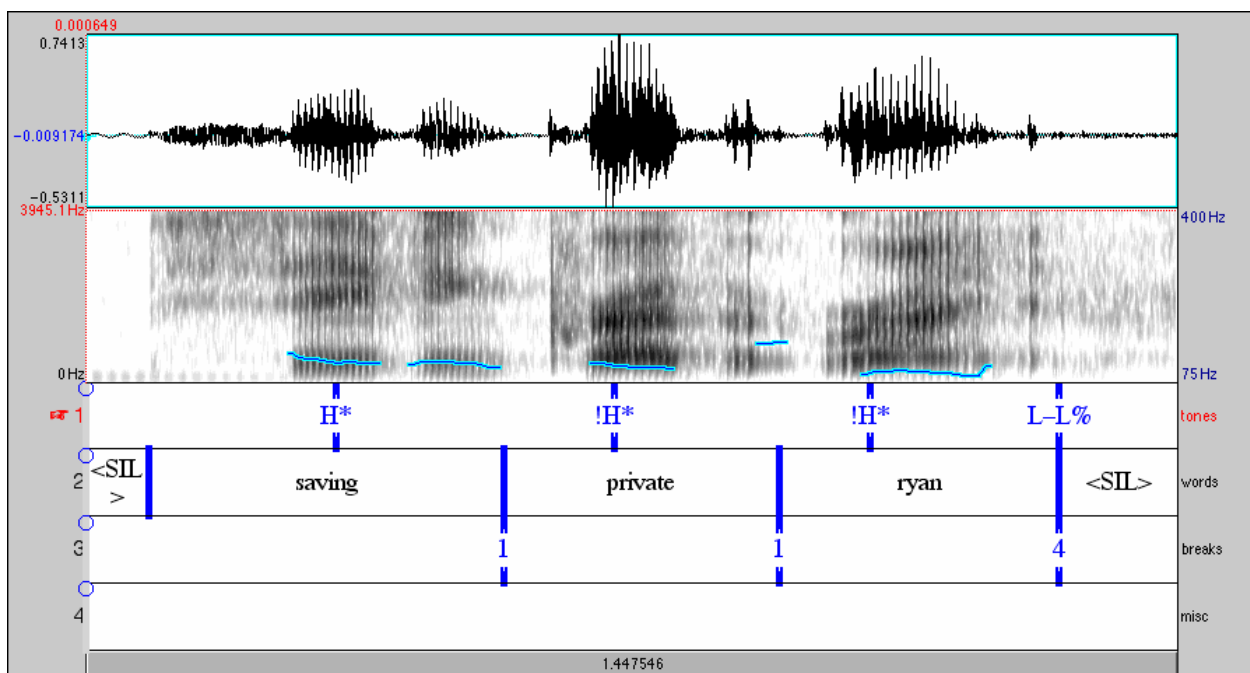


Figure 2.6.3b The example private_ryan1 shown with the pitch range in Praat set too wide for this speaker.

<private_ryan1>

The same tone labels (an H* pitch accent followed by !H* followed by another !H*) can be seen in the example *Figure 2.6.5* <fenced_meadow>. There are no theoretical limits within the ToBI system as to how many times such downstepping can occur within a phrase.

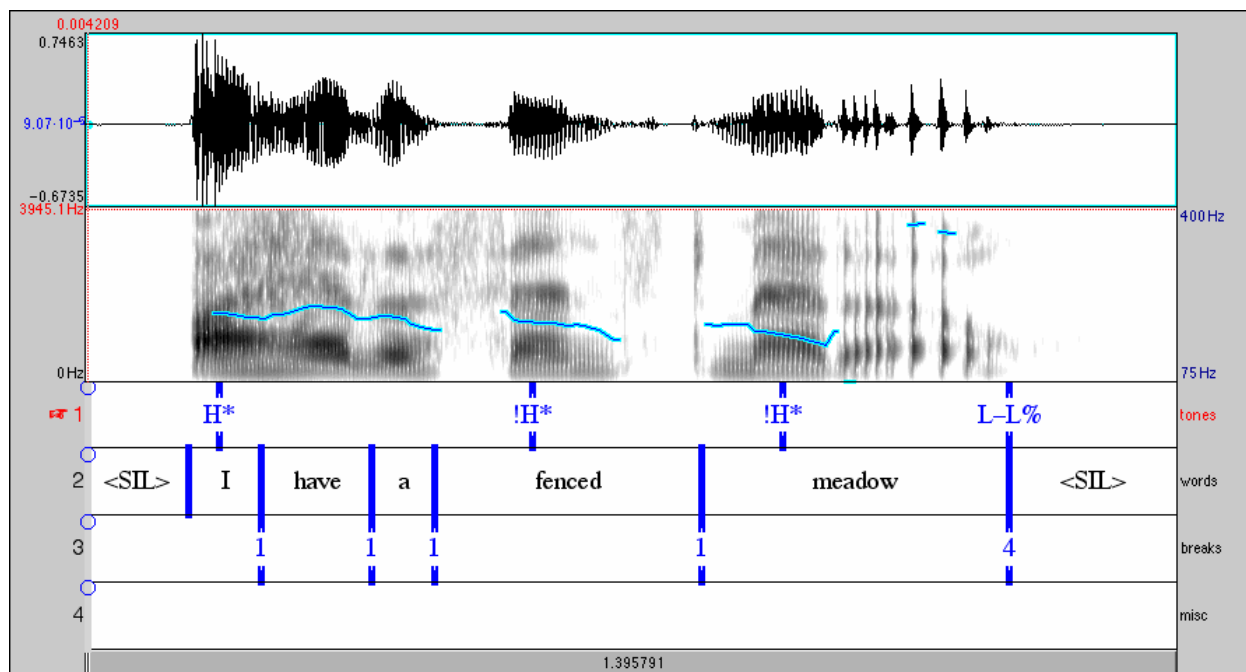


Figure 2.6.4 Another example of two sequential !H* pitch accents <fenced_meadow>

On the other hand, there is no special term or diacritic for a second High pitch accent that is realized with a higher f0 than its immediate predecessor. In this case, the second, higher pitch accent would be simply labeled as another H*, as in the following examples:

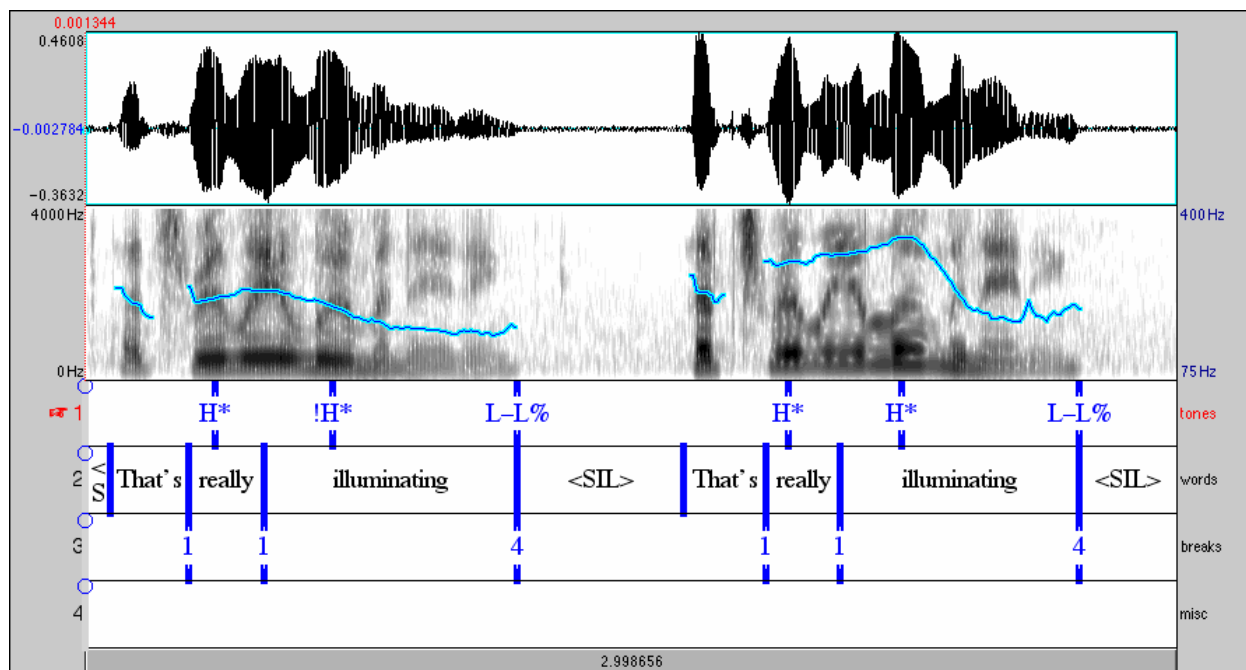


Figure 2.6.5 H* !H* L-L% compared to H* H* L-L% <illuminating1>

In this example, the first rendition of *That's really illuminating* has two pitch accents, an H* on *real-* and a !H* on *-lum-*, similar to the examples of downstep shown above. However, the second rendition has an H* on *real-* followed by a higher H* on *-lum-*. The differences in height of H*'s in an intonational phrase are not distinctive; that is, H*'s with medium-high f0 don't convey different information than H*'s with higher f0. However, the term 'upstep' is sometimes used casually or in other contexts (see box).

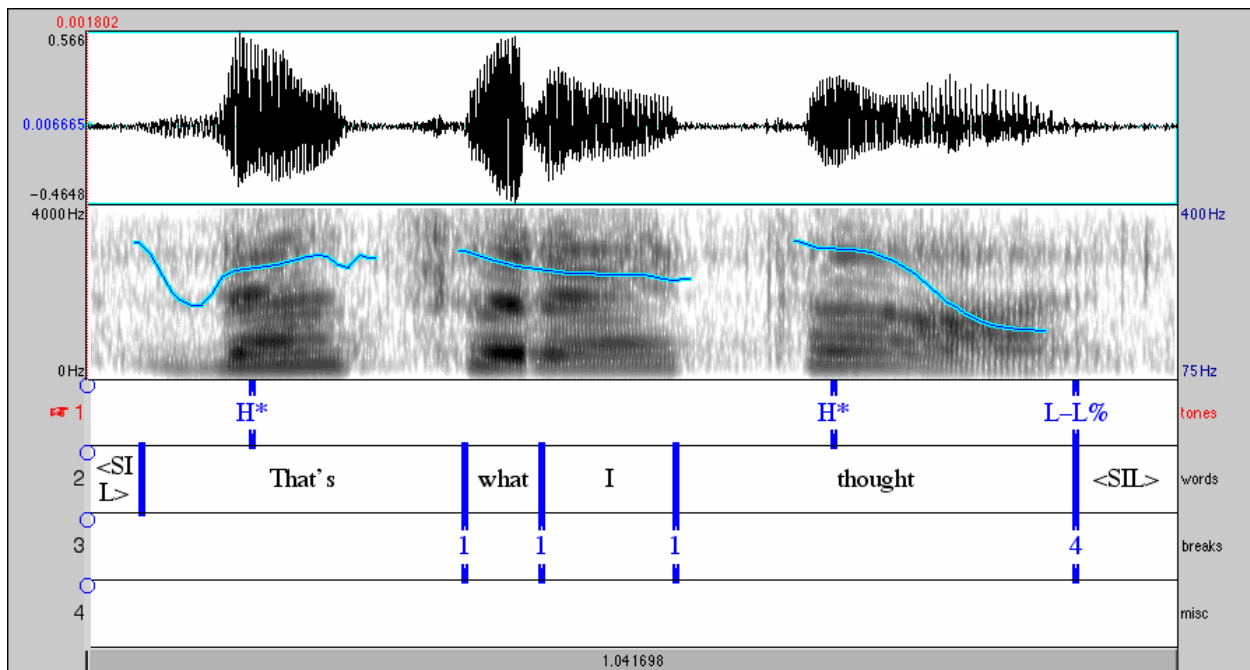


Figure 2.6.6 An example of H* followed by another H* realized with a higher f0: *<thought>*

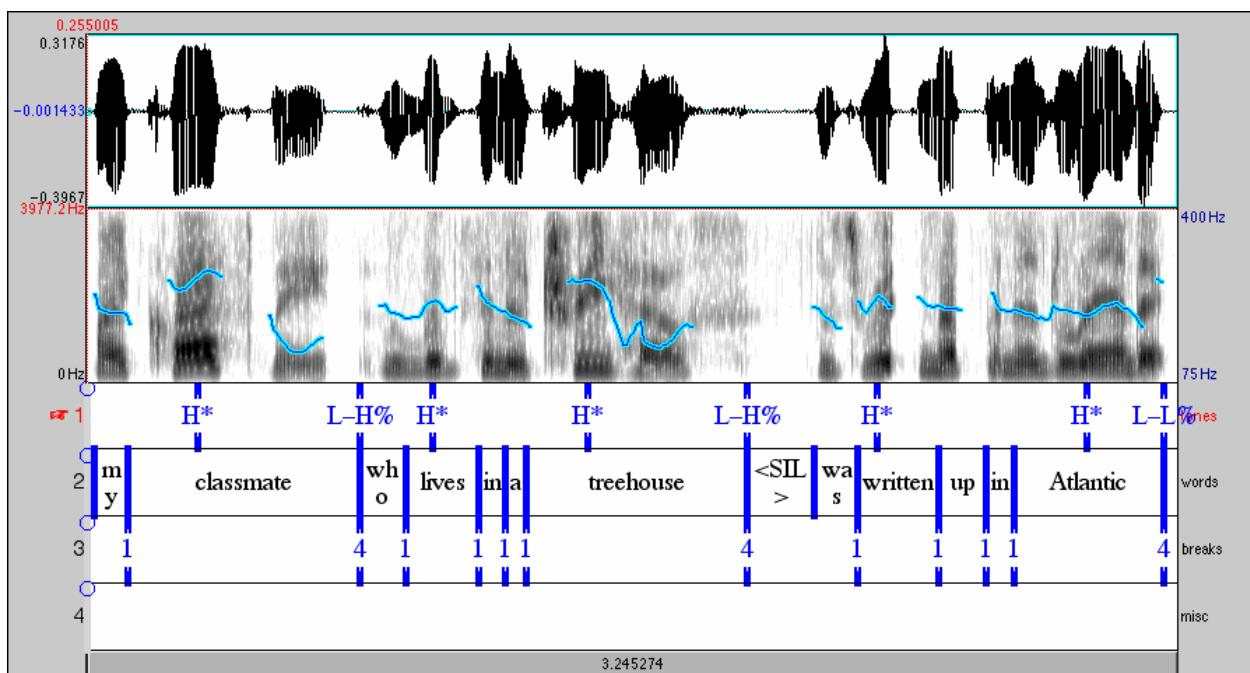


Figure 2.6.7 Another example of H* followed by another H* realized with a higher f0:
< treehouse >

In this last example, note the middle Intonational Phrase (*who lives in a treehouse*) where the H* on *treehouse* is higher than the H* on *lives*.

What about 'Upstep':

The term 'upstep' has sometimes been used in ToBI literature (e.g. the Labelling Guide), but not in a way corresponding to the term 'downstep.' Specifically, it has been used not for pitch accents, but in relation to phrase tone-boundary tone combinations. The High phrase tone (H-) is said to 'upstep' a following boundary tone. For example, the Low boundary tone (L%) in a H-L% combination is not realized as a lowering in pitch, as the H- phrase tone has 'stepped it up.' The High boundary tone (H%) is also considered 'upstepped' after a H phrase tone (H-): the rise in pitch at the boundary for H-H% (i.e. when following a High phrase tone) is higher than at the boundary for L-H% (i.e. when following a Low phrase tone).

In a sequence of two pitch accents with H targets, we have seen that it is often the case that the second pitch accent will be realized with a higher f0 than the first. While downstep is explicitly marked with the ! symbol, there is no analogous marking of 'upstep' in a sequence of two pitch accents

In addition, once a speaker has produced a !H*, there is no requirement to continue to downstep for the rest of the Intonational Phrase. A speaker can produce a !H* and follow it with an H*, as in the following example:

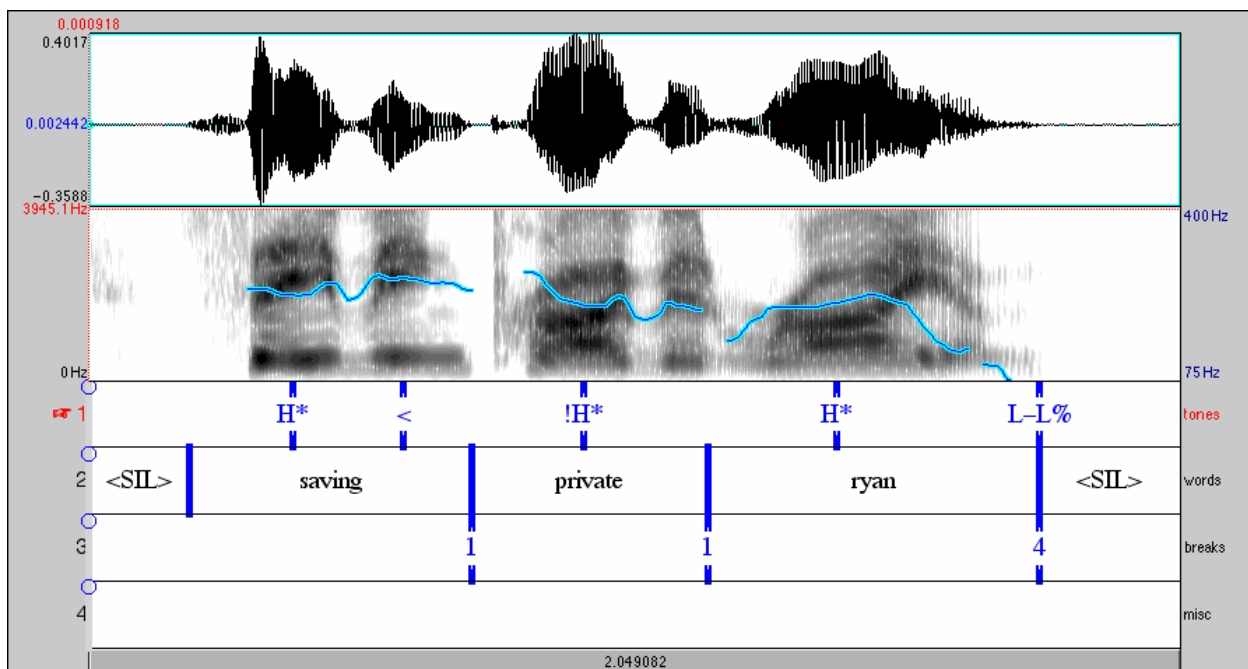


Figure 2.6.8 H* following !H* on *saving private ryan*

<private_ryan2>

Here, the speaker produces an H* on *Sav-*, a !H* on *Pri-* but follows with an H* on *Ry-*. Although the last H* (on *Ry-*) is produced with a lower f0 than the first H* (on *Sav-*), downstep is relative to the preceding pitch accent only, so here it would be relative to the !H* on *Pri-*. Of course, the sentence can be produced in many different ways; for example, it could be produced with a series of !H*, as in earlier examples and the example below.

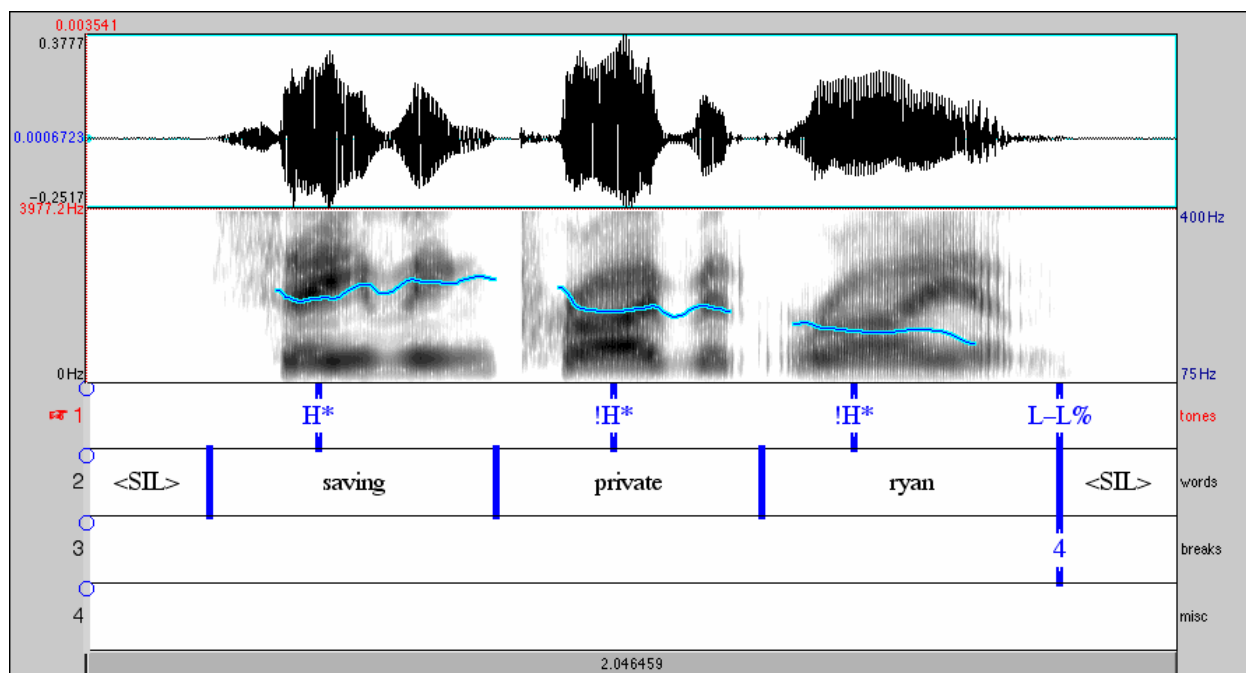


Figure 2.6.9 a sequence of !H* produced by the same speaker as the previous example.
<private_ryan3>

One other optional ToBI label was used in <private_ryan3>: the < in *Saving*. As you may have noticed, the peak of the High Tone doesn't always align with, e.g. the center of the accented syllable's vowel. Many labelers wish to note when a peak is particularly late with respect to the vowel and will mark the < to indicate a late peak. This label is optional and therefore not consistently marked throughout the ToBI community; the question of whether this alignment distinguishes different pitch accent categories in Mainstream American English is still under investigation. (A good way to remember which symbol to use is that the sharp end of the angled bracket points in the direction of the associated accented syllable.)

In earlier discussions of pitch accent (whether in the single-tone H* or L*, or in the bitonals L+H* and L*+H), we have tried to give examples where the pitch movement associated with the pitch accent is clearly visible in the pitch contour. However, it is not the case that each pitch accent will have a clearly distinguishable pitch movement. Sometimes the pitch event associated with a pitch accent is not a movement, such as a High or Low culminating in a "peak" or "valley," but an absence of movement toward another target. For example, in an Intonational Phrase with an H* ending in a Low phrase tone-Low boundary tone combination (L-L%), the fall into the L phrase tone is expected immediately after the high pitch associated with a High pitch accent (H*). This is the case in <armani2>, shown below, where the F0 begins to fall rather sharply immediately following the end of the pitch-accented syllable (the *-man-* of *Armani*). Notice that by the time the speaker reaches the word *millionaire*, the f0 has nearly reached the bottom of the speaker's pitch range for that utterance. In contrast, <armani3> the f0 between the H* on the *-man-* of *Armani* and the !H* pitch accent on the *mil-* of *millionaire*, stays comparatively high. This second pitch accent, while it doesn't have a 'peak' of its own,

effectively prevents the f0 from falling into the L- phrase tone until later, i.e. until after the second pitch accent. In this example, even the first H* pitch accent does not have a clear ‘peak’ in the f0. Many single-tone H* pitch accents are realized with fairly flat (but still high) f0 regions.

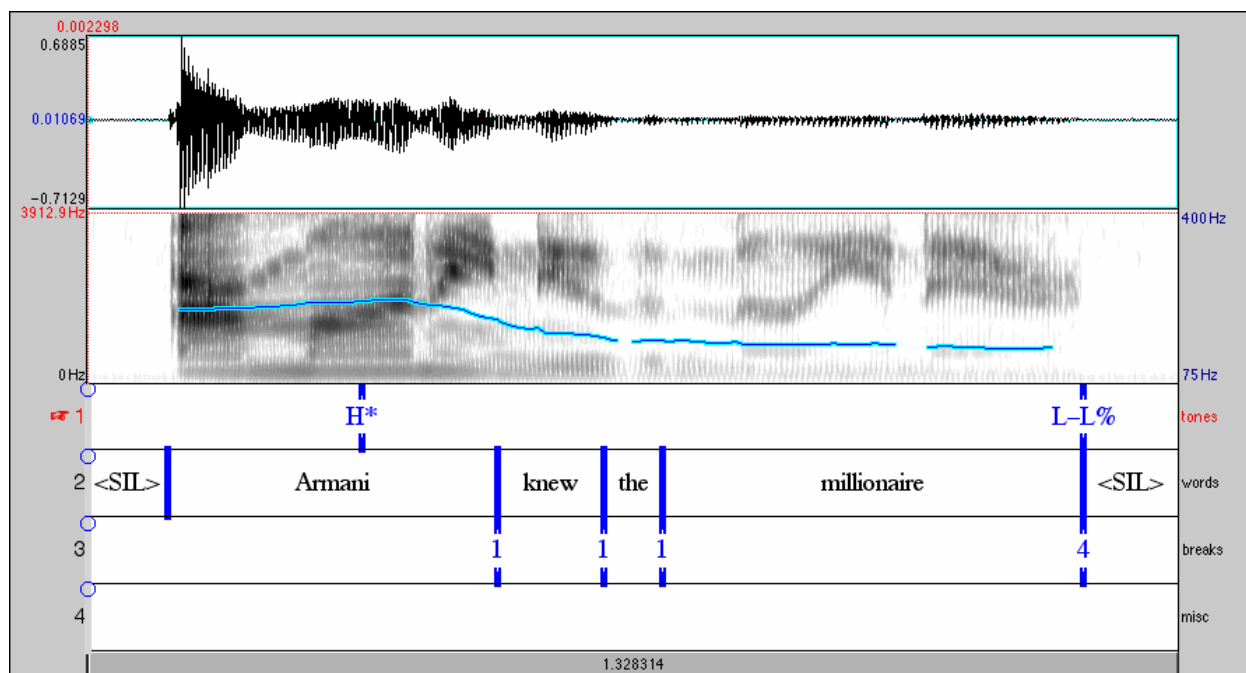


Figure 2.6.10a and 2.6.10b: H* L-L% vs. H* !H* L-L%

<armani2>

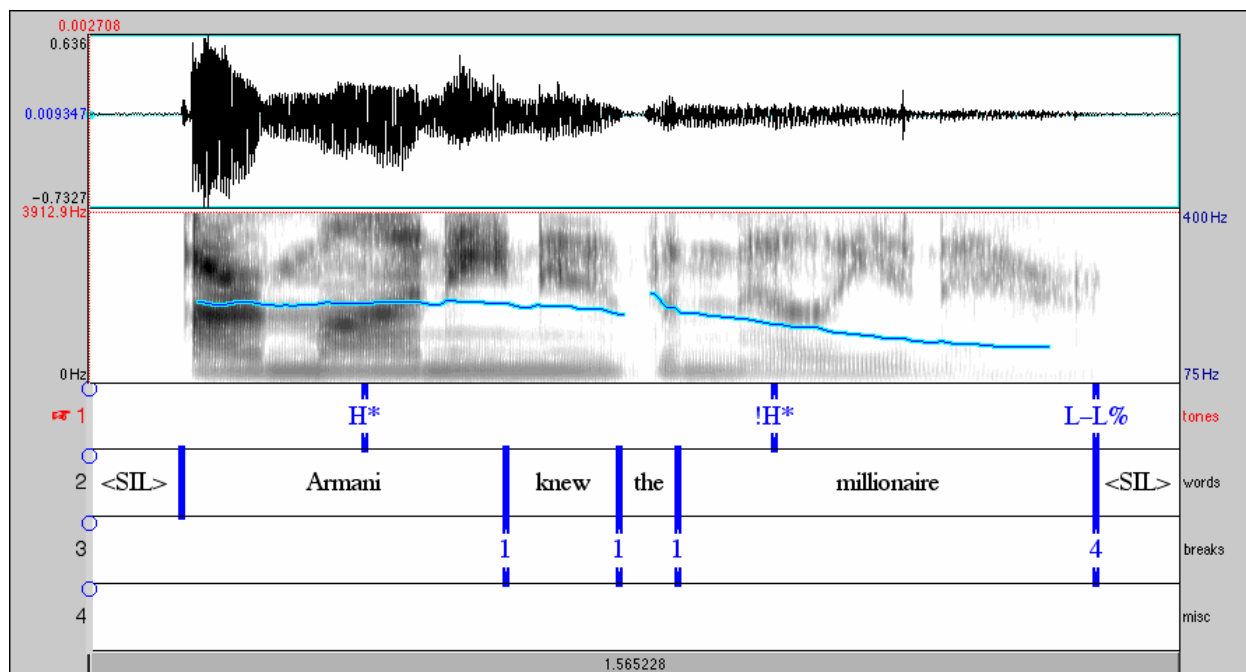


Figure 2.6.10b

<armani3>

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone

!H*: downstepped high pitch accent

L-L%: low phrase tone, low boundary tone

H-H%: high phrase tone, high boundary tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

4: end of an intonational phrase

Optional labels:

<: late High Tonal peak

End list must be added.

<heyAlejna-L>

What triggers downstep?

This section is taken from notes by Mary Beckman, used at the ToBI clinic at Ohio State University.

The trigger for downstep in English is very controversial. Pierrehumbert (1980) first proposed that downstep is triggered by any sequence of alternating L and H tones (as in the African tone languages where it was first described), whereas Beckman & Pierrehumbert (1986) proposed that it is only the alternating sequences of bitonal pitch accents that triggers it. Note that in both of these models, there is a bitonal accent H^*+L which is missing in ToBI, for which the trailing "L" target is not very low in the pitch range. So an Intonational Phrase contour that in ToBI would be transcribed as $H^* !H^* L-L\%$ would be represented in the Pierrehumbert (1980) or Beckman & Pierrehumbert (1986) models as $H^*+L H^* L-L\%$, with the downstep that is explicitly marked in ToBI being understood as an automatic consequence of the choice of a bitonal H^*+L accent.

Ladd (1983, 1986) objected to this characterization of downstep, because it meant positing this H^*+L accent that typically doesn't have an obvious L target. Gussenhoven and his colleagues have also objected to this, and have always maintained that downstep is a facultative choice that a speaker can make to signal greater cohesion within a prosodic phrase (see, e.g. van den Berg, Gussenhoven, & Rietveld (1992) in LabPhon2). They have a way of representing it that differs from ToBI, in terms of a combination of a H^*L accent type for which the L tone typically represents the L- phrase tone, but in cases where the accent is not the nuclear tone, it detaches and reattaches to the following H^*L accent, or simply detaches and is a 'floating' tone that is realized as downstep. In this system, downstep is represented not as a mark on the accent tones, but as the choice of 'float' among the three-way opposition among ways to deal with the L tone (i.e., 'spread' for a nuclear accent vs. 'reattach' for a what we would transcribe as the leading L tone of a $L+H^*$ accent vs. 'float' for what we would transcribe as a H^* or $L+H^*$ accent before a downstepped $!H^*$ accent).

Ladd (1986) proposed a different representation again -- namely that downstep be modeled as one term in a binary opposition between two different "register tone" relationships -- adopting Clements's tree model of downstep in African tone languages. So the short story is that no one is sure what triggers downstep. There are competing theories, but no definitive data to decide among them.

Since there is this controversy about the theory, what we decided to do back in 1990, at the first workshop where we started to devise the English ToBI system, was to sidestep the controversy completely, by explicitly marking downstep on the first H tone that is realized in the downstepped pitch range. So downstep can follow any pitch accent that has a H target in it. If you want to read the description of how we decided this, by the way, you can look at the Beckman, Hirschberg, & Shattuck-Hufnagel chapter Jun's (2005) book.

2.7 More on downstep: bitonal pitch accents with a downstepped H element (!H)

In the previous section, we introduced the downstepped H pitch accent (!H*) [and a downstepped High phrase accent, shown with a Low boundary tone (!H-L%)]. We will see in this section additional contexts for which the downstep diacritic is used. Namely, there are three bitonal pitch accents which have a downstepped H element: L+!H*, L*+!H and H+!H*.

2.7.1 Examples of L+!H* and L*+!H

As was the case for the single-tone !H* pitch accent, the L+!H* and L*+!H pitch accent labels are only used to mark a relationship between the high element of the pitch accent under consideration, and an immediately preceding pitch accent containing a High tone element. And much like the two non-downstepped bitonals we have looked at previously (L+H* and L*+H), the L+!H* and L*+!H indicate that there are two tonal targets associated with the prominence: a Low and a High (this time a downstepped High).

The file <bloomingdales1>¹ is an utterance of the sentence *There's a lovely one in Bloomingdales*. This Intonational Phrase contains two bitonal pitch accents, and culminates in a Low phrase accent-Low boundary tone sequence. The first pitch accent is the bitonal L+H*, such as we have seen in previous sections. Notice how the f0 does not begin to rise (i.e. it stays relatively low) until the first syllable, *love-* of the word *lovely*, peaking around the end of the prominent syllable, again *love-*. This f0 alignment reflects the L+H* pitch accent. The second pitch accent, on the syllable *Bloom-* of *Bloomingdales*, shows similar behavior: the f0 is low leading up to the prominent syllable, *Bloom-*, and then rises sharply to a peak around the end of that syllable. What makes this pitch accent L+!H*, as opposed to L+H* (with no downstep) is that the High target of the bitonal pitch accent on *Bloom-* is perceptually lower than the High target of the preceding pitch accent (the pitch accent on *love-* of *lovely*). This relationship among perceptual pitch heights is comparable to that between the single-tone !H* pitch accent and the pitch accent with a High target that immediately precedes it. (The example <illuminating3> shows another example of the same pattern of tones, L+H* L+!H* L-L%, on the words *that's really illuminating*.) In these two examples, you may have noticed that the leading tone of the L+!H* (on *Bloom-* of *Bloomingdales* and *-lum-* of *illuminating*) is also lower than the leading L of the first pitch accents (the L+H* on *love-* and *real-*). However, this is not necessarily always the case; the L+!H* does not reflect any height relationship with the Low of the preceding pitch accent.

¹ From this point in the tutorial, we will include the spectrograms (see Chapter 1) in our displays because they are useful in aligning the intonational contour with the sounds in a syllable. You may notice that you can locate, e.g. where the /l/ in *lovely* begins and ends from the light band in the spectrogram of <bloomingdales1>.

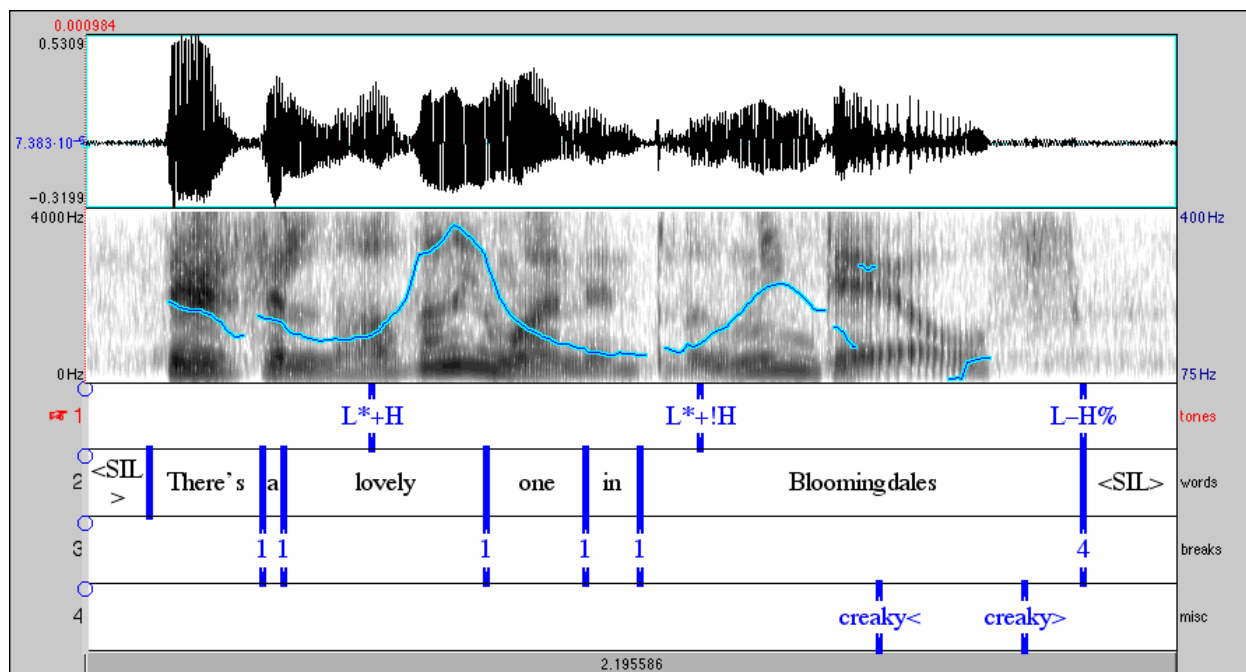


Figure 2.7.1 L+H* L+!H* L-L%

<bloomingdales1>

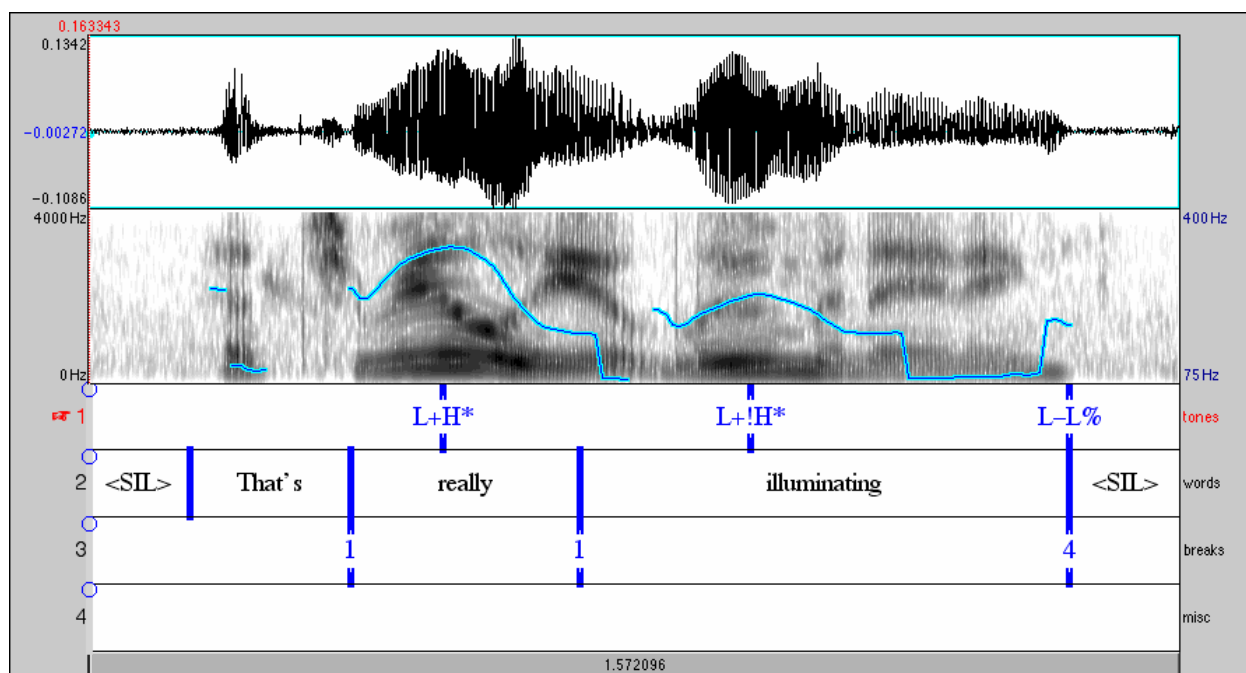


Figure 2.7.2 Another example of L+H* followed by L+!H

<illuminating3>

The example <yellow_mini>, shown below for comparison, has a sequence of two L+H* pitch accents with no downstep relationship between them; the peak of the second pitch accent, on the *yel-* of *yellow* is approximately the same height as that of the first pitch accent, on *John*.

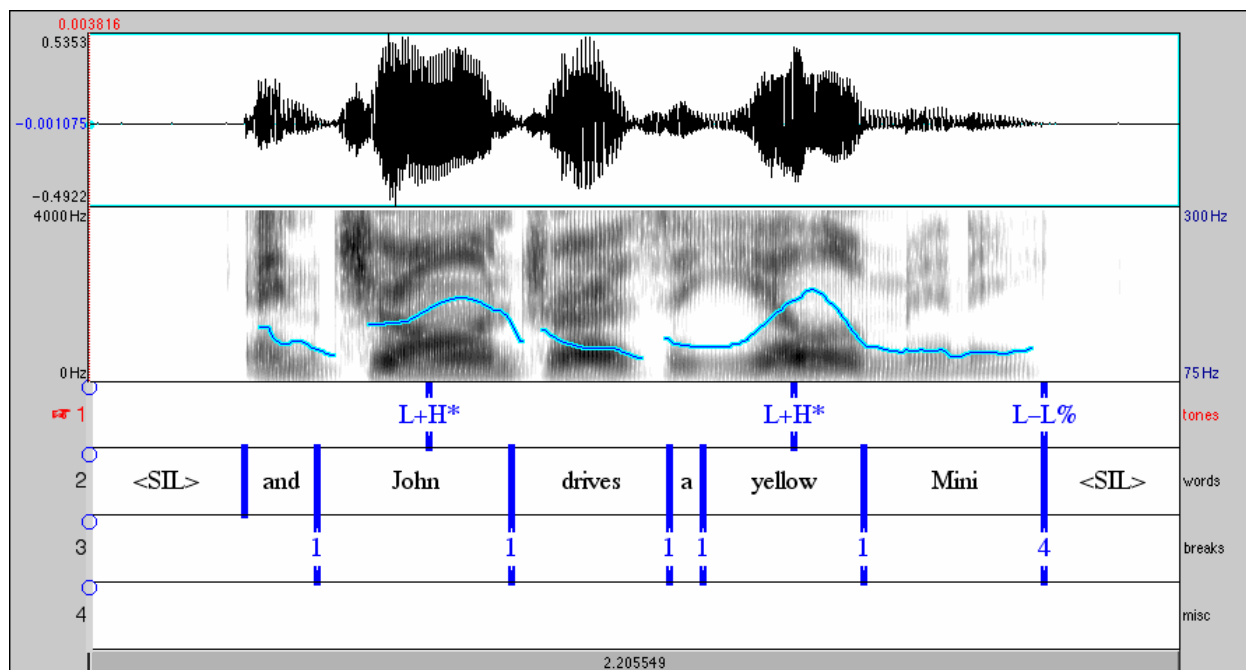


Figure 2.7.3 L+H* L+H* (no downstepping between the two accents) <yellow_mini>

Another example is shown in <marianna>, where both L+H* accents occur in the small word. The High tone in the second L+H* is actually higher than the first.

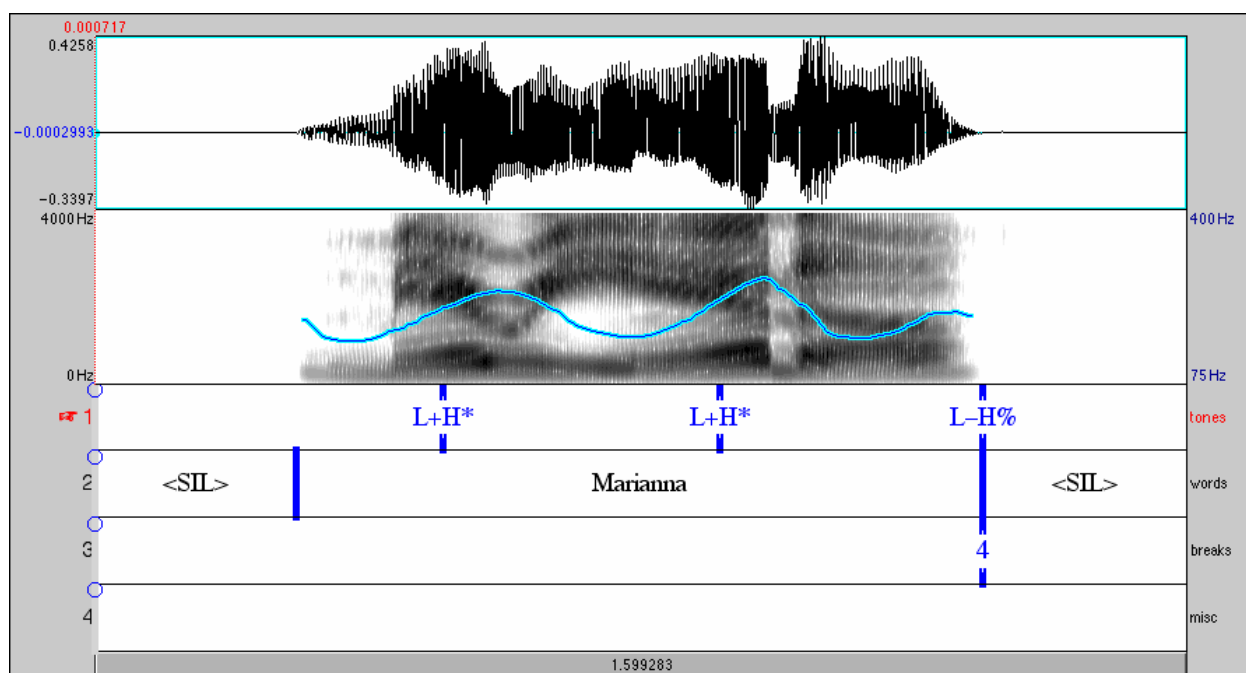


Figure 2.7.4 L+H* L+H* (again, no downstepping between accents) <marianna>

The relationship between the L*-based bitonals (L*+H vs. L*+!H) is quite similar to that of the H*-based bitonals (L+H* vs. L+!H*): the downstep diacritic before the H symbol always indicates that the High tone of the pitch accent in question is realized with a perceptually lower

pitch range than the High tone target of the immediately preceding pitch accent. However, note that in the case of the L*+!H pitch accent, the H symbol associated with the ! diacritic is not the starred tone.

The example <bloomingdales2>, shown in *Figure 2.7.4* below, has an instance of L*+!H on the syllable *Bloom-* of *Bloomingdales*. In this example, the L*+!H follows a (non-downstepped) L*+H pitch accent. (Again, note that while the L target of the second pitch accent in this sequence is also perceptually lower than the L target of the preceding pitch accent on the *love-* of *lovely*, this is not necessary to the use of the L*+!H label, nor is this pitch difference between the two Low tones indicated in ToBI.)

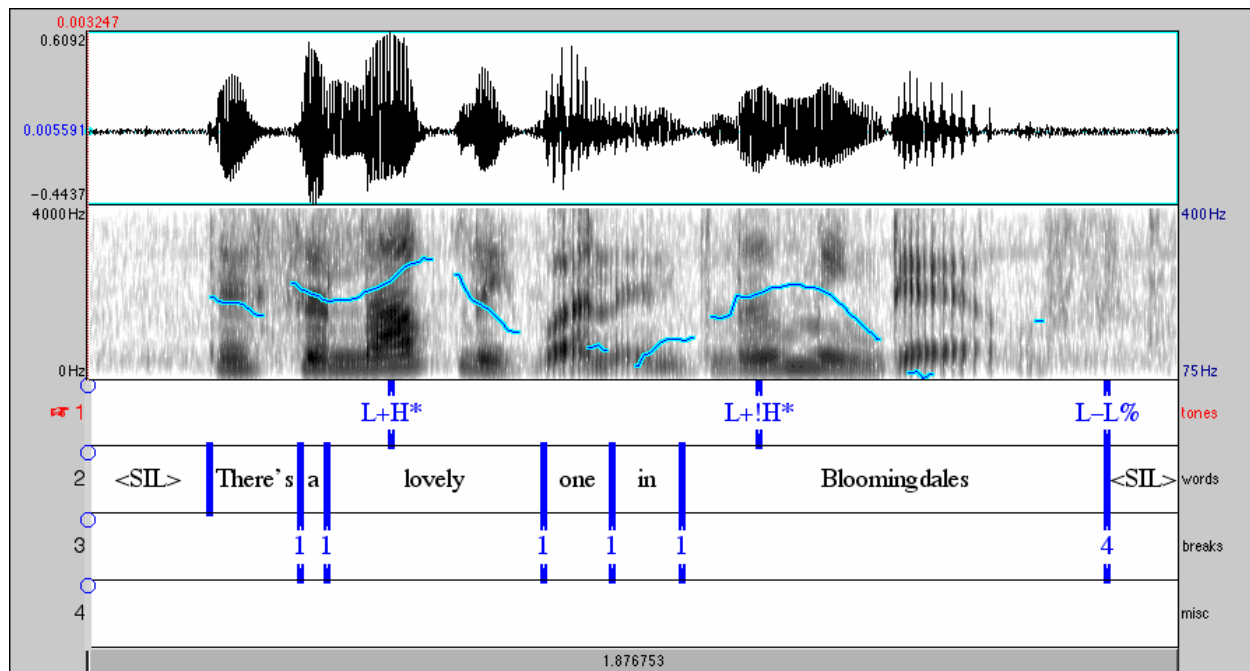


Figure 2.7.5 L*+H L*+!H L-L%

<bloomingdales2>

Like the single-tone !H* pitch accent, these two bitonal pitch accents with a downstepped H tone can follow any pitch accent with a High tone target. Specifically, this means that they can follow H*, L+H*, L*+H but also other pitch accents with a downstepped High target: !H*, L+!H* or L*+!H. The example <joeey> shows H* followed by L*+!H, and <plenty> shows H* followed by L+!H*.

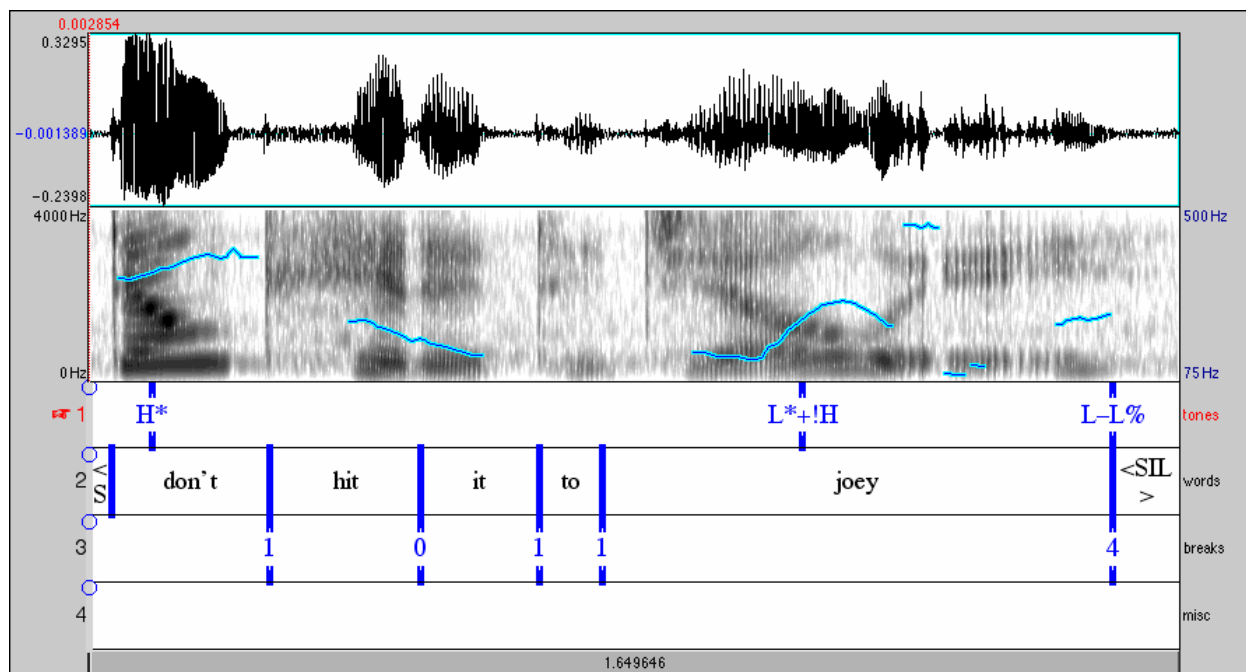


Figure 2.7.5 $L^*+!H$ following a single-tone H^* pitch accent

<joey>

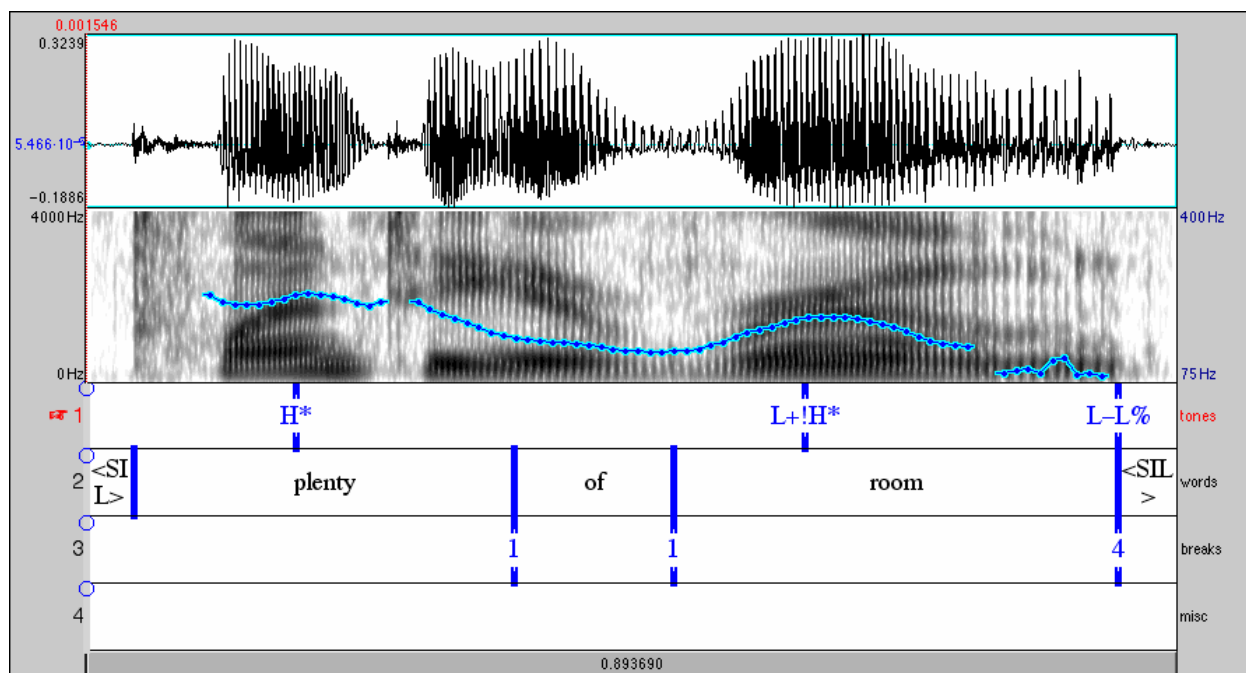


Figure 2.7.6 $L+!H^*$ following a single-tone H^* pitch accent

<plenty>

2.7.2 The $H+!H^*$ pitch accent

There is one final pitch accent type that has not been introduced: the $H+!H^*$ pitch accent. The $H+!H^*$ pitch accent (also referred to as “H plus downstepped H star”) is different from other pitch accents in the ToBI inventory in two ways. First, the downstep symbol (!) is used somewhat differently than in other pitch accent events. While it still demarcates a tone that is

lower in relation to a preceding High, in this case, the preceding High is not that of a preceding pitch accent. Instead, the downstep symbol marks a relationship between tones within a single prominence-lending tone event; the !H* part of the pitch accent is downstepped in relation to the leading High tone (the first H of the H+!H*) of the same pitch accent.

The other way in which the H+!H* pitch accent differs from others in the ToBI inventory is that it is a bitonal pitch accent that does not have contrasting High and Low targets. Each of the previously introduced bitonal pitch accents (L+H*, L*+H, L+!H* and L*+!H) has a Low target, followed by a High target. The H+!H* has two High targets.

The example <mother_theresa>, shown below, contains two full Intonational Phrases, the second of which, on the text *Mother Theresa*, contains an H+!H* pitch accent. The pitch accent in question is on the prominent syllable *-re-* of *Theresa*, and is followed by the phrase accent-boundary tone combination L-L%. Look carefully at the pitch track over the words *Mother* and *Theresa*. Notice how the highest F0 in the pitch track (for the Intonational Phrase on the words *Mother Theresa*) occurs at the beginning of the word *Theresa*: this is a reflection of the first H tone (the unstarred H) of the H+!H*. The pitch is falling through the word *Theresa*, and is quite a bit lower by the time it reaches the prominent syllable *-re-*, a reflection of the !H* portion of the bitonal pitch accent H+!H*. The pitch then continues to fall through the end of the word and the end of the intonational phrase, as it heads towards the Low targets of the phrase accent and boundary tone. (As we have seen before, the pitch track disappears due to pitch tracking errors, as the speaker reaches the bottom of her pitch range; notice that there is no reasonable pitch track for the final syllable *-sa* of *Theresa*.)

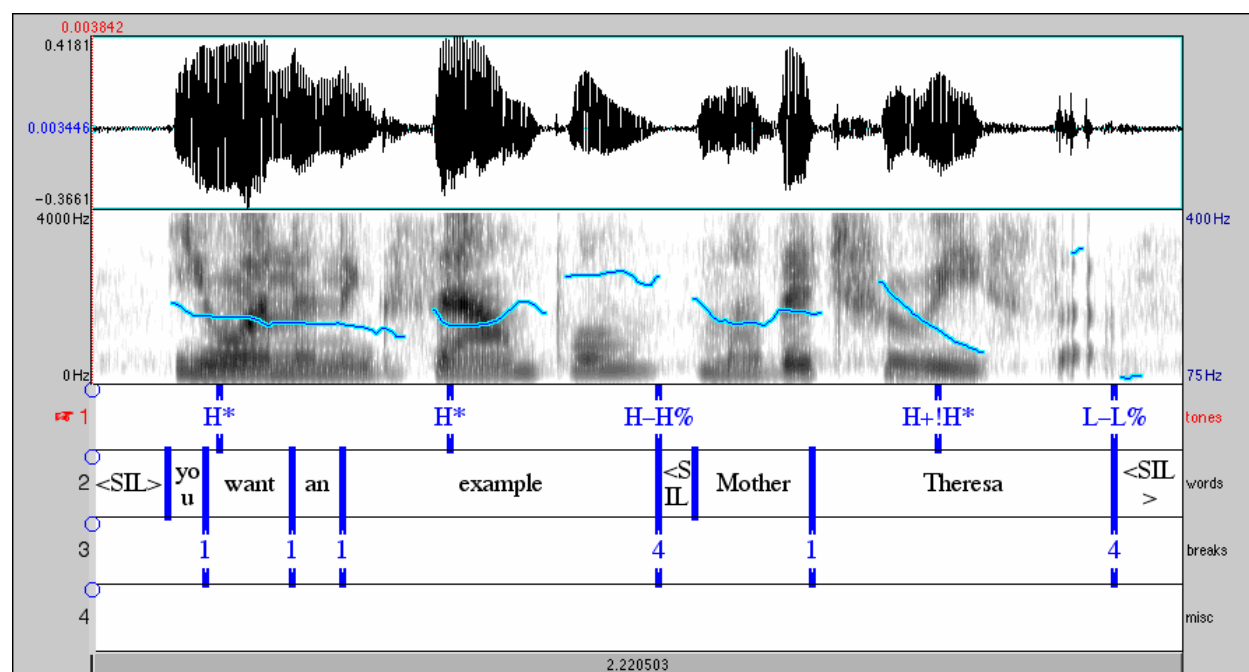


Figure 2.7.7 H+!H* followed by L-L%

<mother_theresa>

The example <design>, shown below, includes another instance of the H+!H* pitch accent, also followed by L-L%. In this example, the pitch accent in question is on the syllable *-prove-* in the

word *improvements*. As with other bitonal pitch accents we have seen, the unstarred tone is often realized on syllables adjacent to the prominent syllable. Here, the highest pitch associated with the H+!H* pitch accent again occurs in the syllable before the prominent syllable: in this case the highest F0 actually occurs on the *im-*.

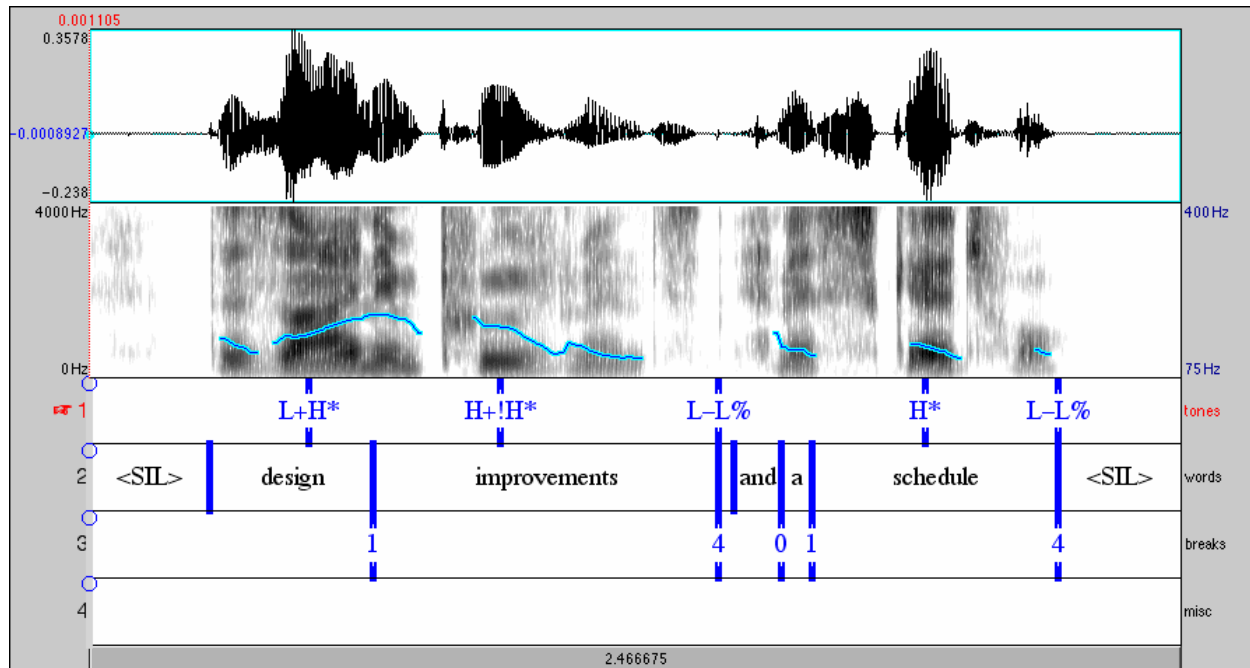


Figure 2.7.8 Another example of H+!H* L-L%

<design>

As with other pitch accents in the ToBI inventory, the H+!H* pitch accent can occur with a range of other pitch accents and phrase accent-boundary tone combinations. The example <minimum2> contains an H+!H* pitch accent followed by a High phrase accent-Low boundary tone combination (H-L%). In this example, the pitch starts high on the word *a* at the beginning of the Intonational Phrase, and falls slightly into the syllable *min-* of *minimum*, and then stays flat through the end of the Intonational Phrase. The change from the high pitch on *a* to the lower (downstepped) high pitch on *min-* is captured in the pitch accent label H+!H*. The following high flat stretch through the final two syllables of the word *minimum* reflects the H-L%. This flat, but high, f0 stretch is comparable that seen in previous examples where the H-L% combination follows a High-tone pitch accent, such as H* or !H*. Two examples of other high, flat f0 stretches associated with the H-L% phrase accent-boundary tone combination are reproduced below. The example <money2>, originally seen in section 2.4, shows H-L% following a (plain) H* pitch accent, and <alejna2> shows H-L% following a downstepped H* (!H*) pitch accent.

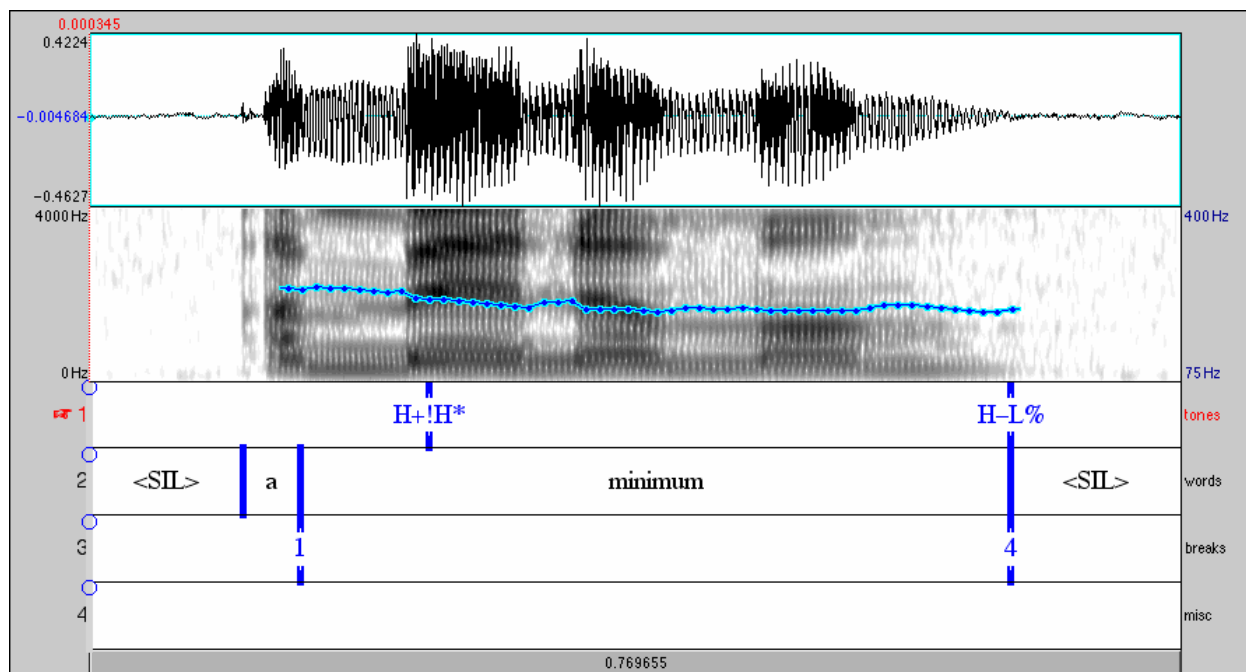


Figure 2.7.9 H+!H* followed by H-L%

<minimum1>

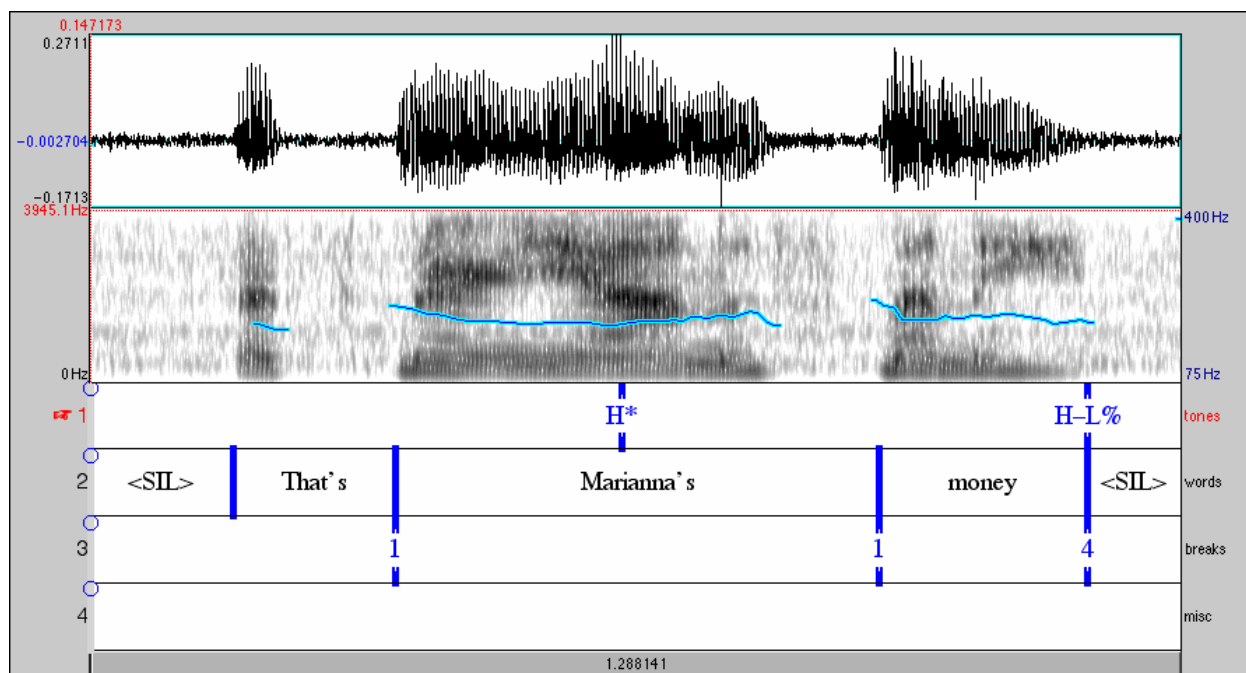


Figure 2.7.10 H* H-L% (shown for comparison with H+!H* H-L%)

<money2>

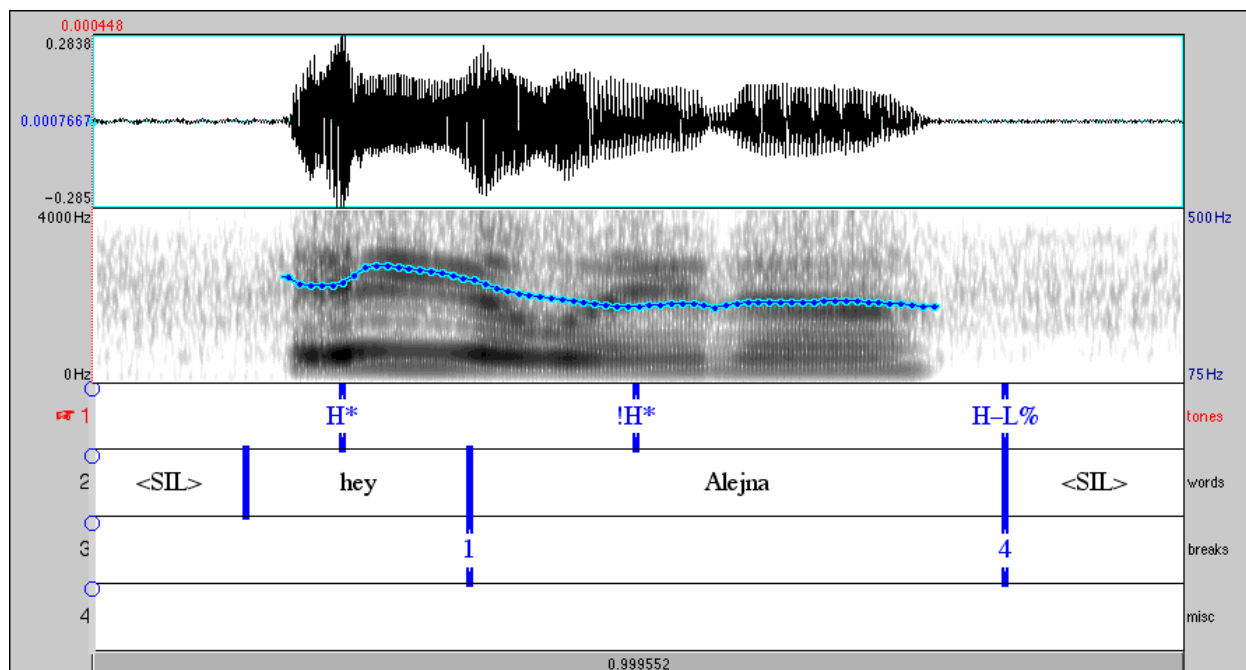


Figure 2.7.11 H !H* H-L% (shown for comparison with H+!H* H-L%)

<alejna2>

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone

!H*: downstepped high pitch accent

L+!H*: bitonal pitch accent with low tone followed by a downstepped high tone prominence

L*+!H: bitonal pitch accent with low tone prominence followed by downstepped high tone

H+!H*: bitonal pitch accent with high tone followed by downstepped high prominence

L-L%: low phrase accent, low boundary tone

H-H%: high phrase accent, high boundary tone

L-H% low phrase accent, high boundary tone

H-L% high phrase accent, low boundary tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

4: end of an intonational phrase

Optional labels:

<: late High Tonal peak

Apparent gaps in the inventory of pitch accents

All pitch accents (and other tones) in ToBI are composed of one or more High (H or !H) or Low tone elements. In addition to the three single-tone pitch accents (H*, L* and !H*), there are five bitonal pitch accents, which are composed of two tones: L+H*, L+!H*, L*+H, L*+!H and H+!H*. You may have realized that there are other logical possibilities for tone combinations, such as H*+L, H*+H, L*+L or L+L*. Some of these possibilities have been proposed by researchers as additional pitch accent categories for ToBI, but to date are not included in the official inventory. Pitch accent labels that have been proposed include: H*+L, H+L* and H*+H.

Some background on the 3 break

Pierrehumbert (1980) postulated the Intonational Phrase as the only intonational constituent in American English, but later work with Beckman (Pierrehumbert and Beckman 1986) suggested the need for a second level of intonationally-defined constituent, the Intermediate Intonational Phrase. This constituent, whose boundary corresponds to a level 3 Break Index in the ToBI system, is defined as an intonation contour with one or more pitch accents and a phrase accent, but no (final) boundary tone. Once Intermediate Intonational Phrases had been proposed, many people began using the term Full Intonational Phrase for the higher-level constituent which contains one or more Intermediate Intonational Phrases plus a right Boundary Tone (realized on the final syllable); this allowed the term 'Intonational Phrase' to be used as a more general term for both kinds of intonationally-defined constituent.

FN: Not everyone follows these terminological conventions, so you will sometimes see the term Intonational Phrase used to refer to the Full Intonational Phrase, and terms like Intermediate Phrase or even Phonological Phrase sometimes refer to what Pierrehumbert and Beckman mean by an Intermediate Intonational Phrase (see terminological table, Shattuck-Hufnagel and Turk 1996). (There is not even complete agreement on whether intonationally defined constituents belong in the same hierarchy of structures as other prosodic constituents; see the work of Gussenhoven and colleagues for a different view.) In addition, the terms major and minor phrase are sometimes used to denote a Full Intonational Phrase (4) and an Intermediate Phrase (3), respectively.

2.8 Intermediate Intonational Phrases: 3 breaks and phrase accents

As we have seen, the break index 4 delimits the end of the highest-level Intonational Phrase, but this is not the only level of intonationally-marked phrasing: the break index 3 delimits the end of a lower-level constituent called the Intermediate (Intonational) Phrase (see box for terminology variations). An Intermediate Phrase represents a level of disjuncture that is less disjoint from the next phrase than a Full Intonational Phrase (labeled with Break Index 4) is. In terms of the hierarchy of prosodic constituents, every Full Intonational Phrase contains at least one Intermediate Intonational Phrase. Each Intermediate Intonational Phrase contains at least one Pitch Accent (and so therefore does every Full Intonational Phrase).

Like the Intonational Phrase, the Intermediate Phrase is tonally marked after its final pitch accent, but not as strongly as the Intonational Phrase. An Intermediate Phrase is labeled with a phrase accent (H-, L- or, as we will see in the next section, !H-), but not with a boundary tone. This distinguishes the Intermediate Phrase's tonal marking from that of an Intonational Phrase, which is marked with boundary tone as well (e.g. L-L%).

In addition to tonal marking, the Intermediate (Break Index 3) Phrase has other cues associated with phrasing: there is durational lengthening in the phrase-final syllable. Like tonal marking, this duration tends to be less strongly marked than the lengthening at the end of an Intonational

Phrase. NB: the final pitch accent in the Intermediate Phrase can occur on any accentable syllable in the phrase. Because our examples tend to be short utterances, they tend to show the final pitch accent on the phrase-final word. However, this does not have to be the case.

It is easiest to understand perceptually what makes a 3 different from a 1 or a 4 in cases where the final pitch accent is an H* and the phrase accent is L-, or vice versa, i.e. where there is a tonal contrast between the final accent in the phrase and the phrase-related tonal marking. For example, consider the example in <design2>. In this example, the f0 curve as well as the pitch perception suggests that there is a Low tonal marker after the H* pitch accent on the word *and*. This evidence, combined with the duration lengthening on the final syllable, which is too strong for a level 1 Break Index, suggests that this syllable ends an intonational constituent. But the degree of duration lengthening and the f0 fall are not substantial enough for a 4/L-L%, so this must be a 3/L-, i.e. an Intermediate Intonational Phrase with an L phrase accent. Contrast this *and* production with both the degree of the L tone as well as the degree of lengthening on the Intonational Phrase final word *schedule*.

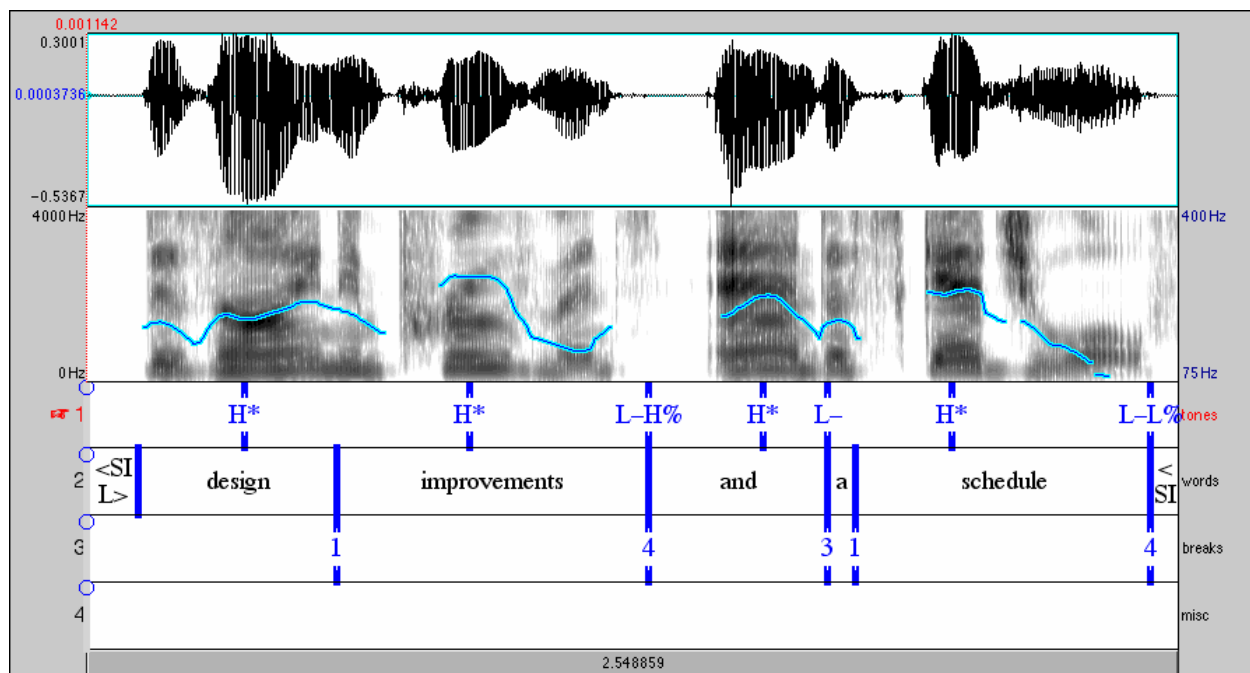


Figure 2.8.1 H* L-[3] on *and* and *a* in *design improvements and a schedule*

<design2>

Another case where the effect of the phrase accent is easy to perceive occurs when the final pitch accent is L+H* and is followed by an L-. Consider the following example <dimmer>, where the L+H* on the first *much* is followed by a falling f0 track which strongly suggests an additional L tonal target, but there is not enough duration lengthening or f0 fall on the final syllable to correspond to an L-L% (with a 4 Break Index). There is also more lengthening than would be appropriate for a level 1 Break Index. Therefore, the best label here is Break Index 3 with an L-tonal marker.

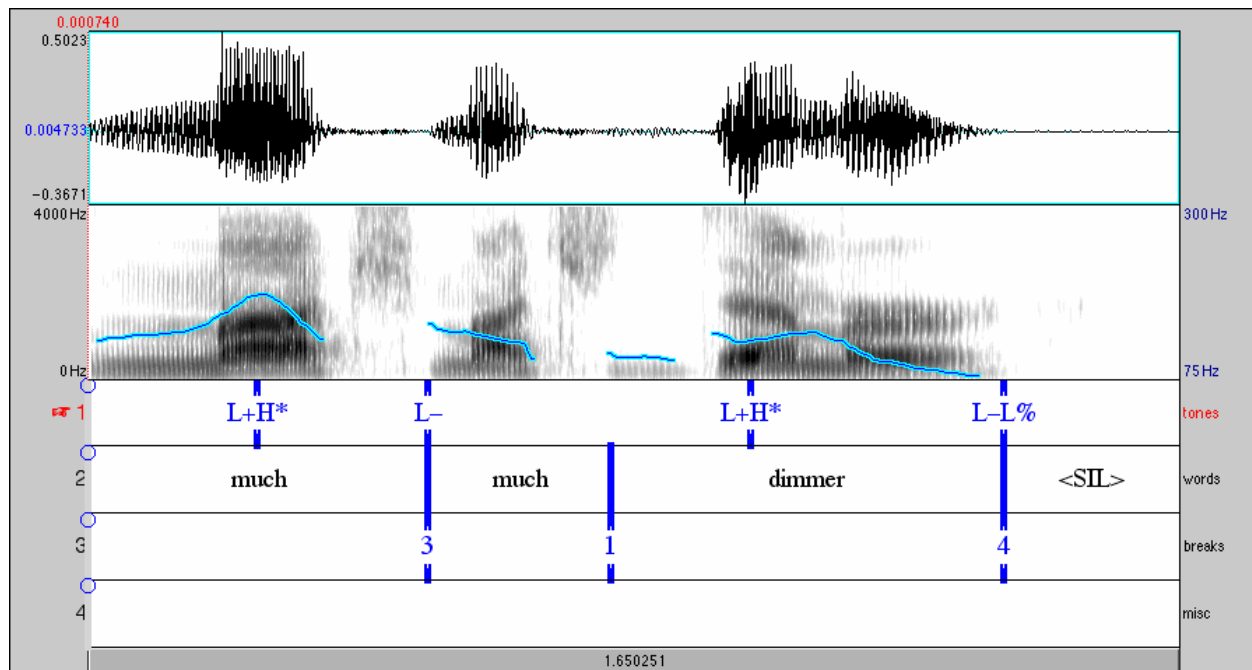


Figure 2.8.2 H* L-[3] on *much* in *much, much dimmer*

<dimmer>

Cases where the final pitch accent is Low (L*) and the phrase accent is H- are likewise relatively easy to perceive. Consider example <argument>, where the L* on *can* is followed by an f0 track which suggests an additional H tonal target. Similar to the examples above, there is again not enough duration lengthening or f0 rise on the final syllable to correspond to an H-H% (with a 4 Break Index). However, in addition to the otherwise unexplained H tone, there is more lengthening than would be appropriate for a level 1 Break Index. Therefore, the best label here is Break Index 3 with an H- tonal marker.

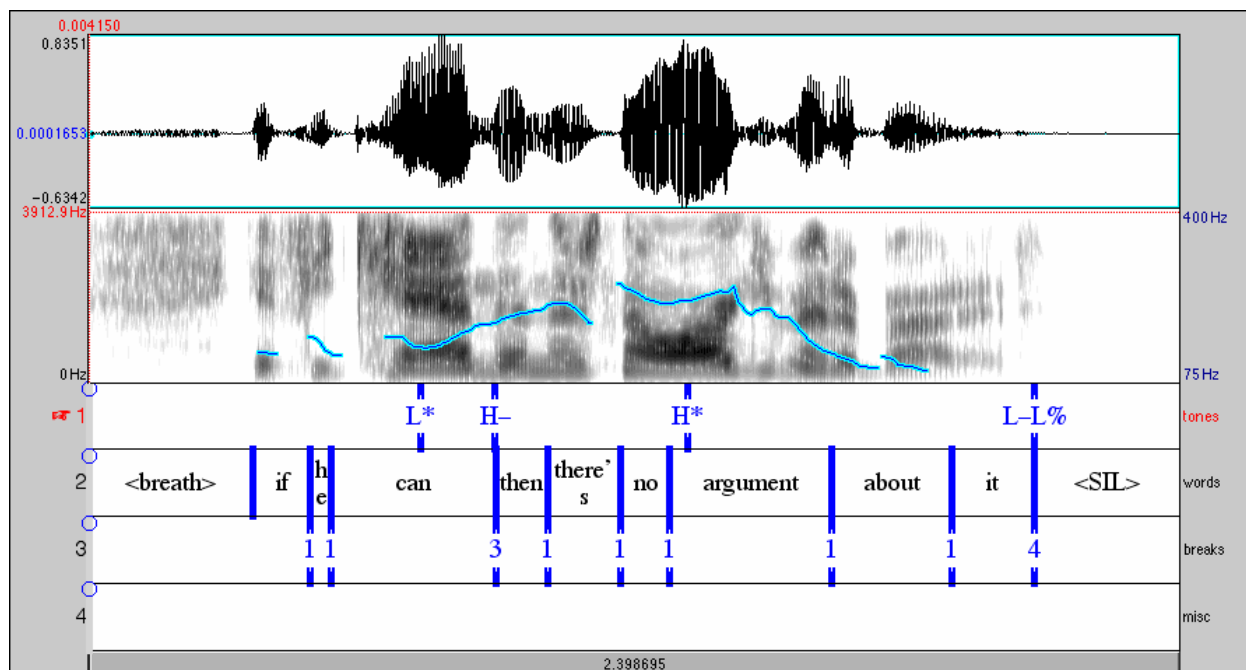


Figure 2.8.3 L^*H^- on *can* in *if he can then there's no argument about it*

<argument>

In the following example, two different productions of *I was wrong* demonstrate the differences between an Intermediate Phrase break (Break Index 3, shown below on the left with an H^-) on *I* vs. an Intonational Phrase break (Break Index 4, on the right, produced here with an $H-H\%$) also on *I*. Although there is a High tonal marker at the end of both phrases, the High tone is higher and the durational lengthening is stronger at the Intonational Phrase Break (4, right example) than on the Intermediate Phrase (3, left example)

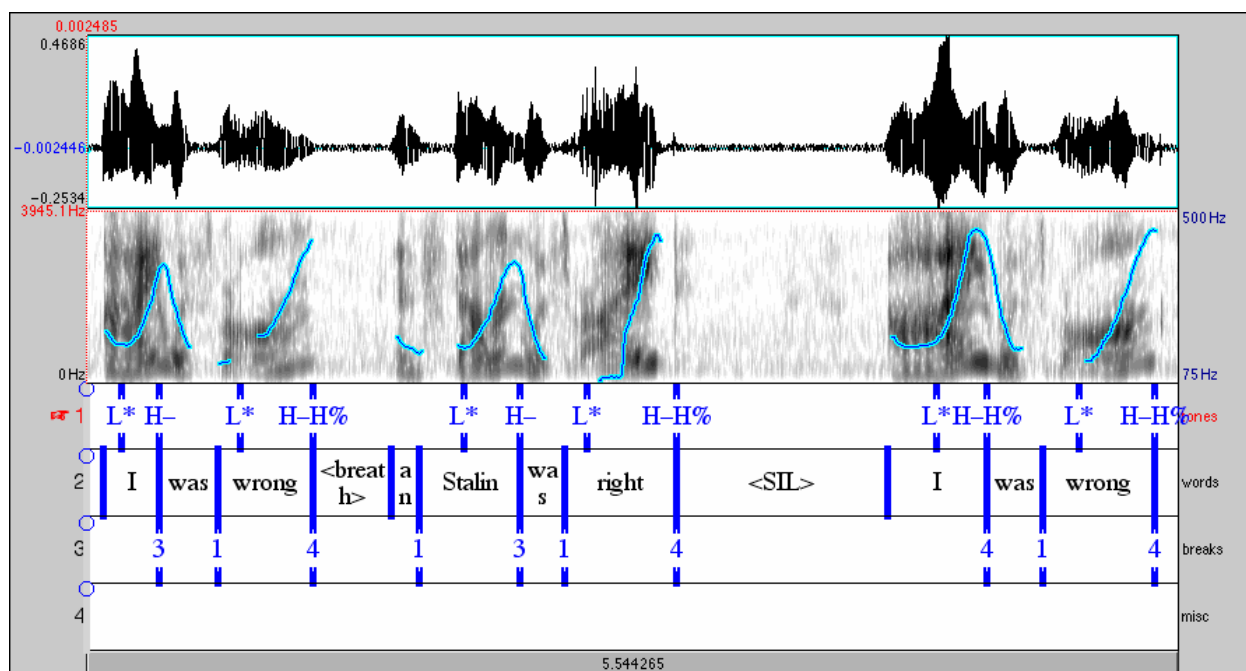


Figure 2.8.4 $L^* H^-$ vs. $L^* H-H\%$ on *I*

<wrong>

Of course, there is no constraint on what phrase accent can follow which pitch accent¹. The examples above were chosen for contrastive clarity. However, there are H* H- and L* L- productions that are less easily identified but no less likely. For example, the H* on *right* in the <abandoned> example below is followed by a High phrase accent (H-). There is a sustained High tone after the pitch accent and some durational lengthening which create a level of disjunction greater than one would see in the case of a 1 Break Index. On the other hand, the durational lengthening and the tonal marking is not as strong as, e.g. that on *cottage*.

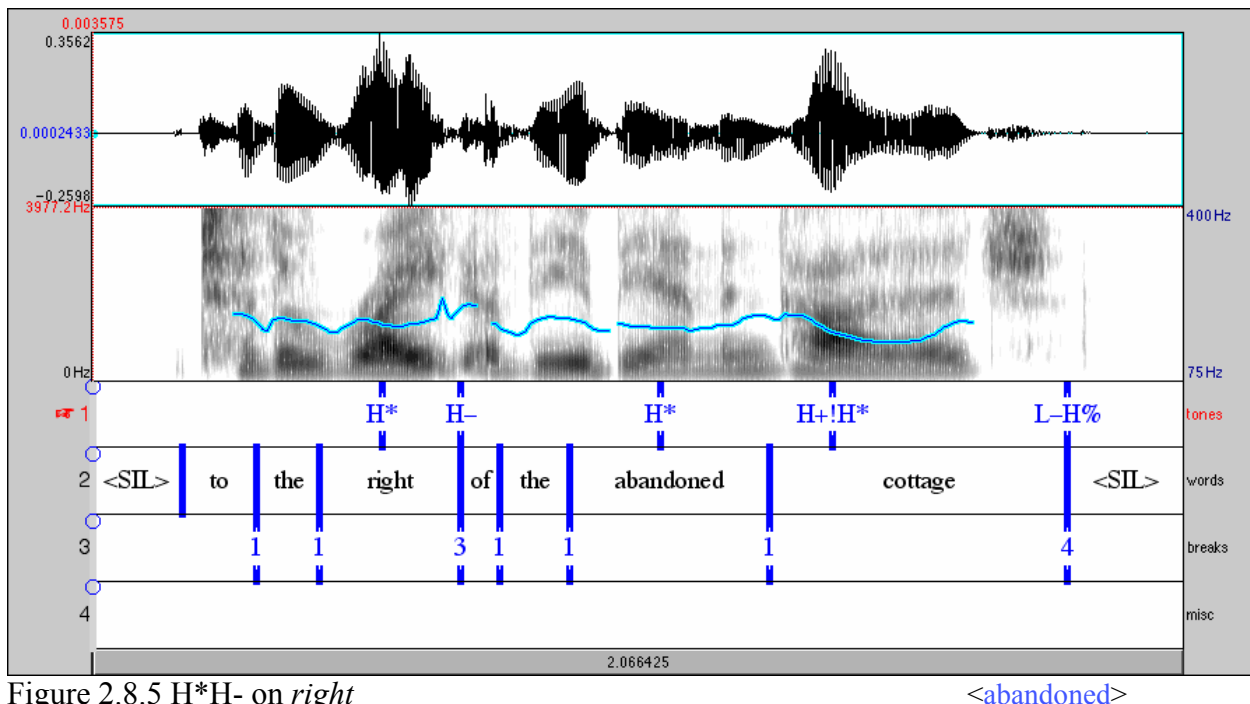


Figure 2.8.5 H*H- on *right*

<abandoned>

Another case of a less obvious Intermediate Phrase occurs when the speaker produces the phrase as an L* pitch accent followed by a L- phrase accent. In <marmalade6>, the flat f0 track might tempt one to not label any prosodic events until *marmalade* at the end of the utterance. However, *Marianna* is perceptually salient and carries a L* tone. Furthermore, there is a sense of disjuncture between *Marianna* and *made* cued by durational lengthening and a sustained Low tone.

¹ The exception to this is with the !H- phrase accent, which will be shown section 2.9, for which there are constraints similar to the use of downstepped pitch accents.

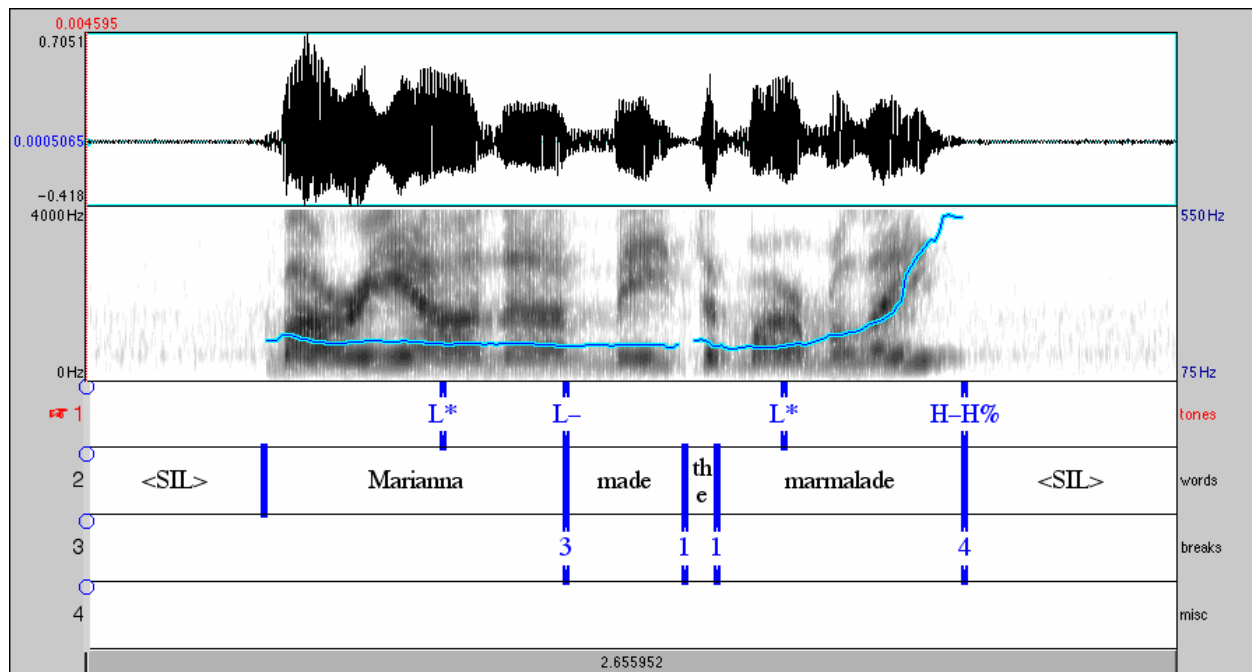


Figure 2.8.6 L*L- on *Marianna made*

<marmalade6>

In summary, the 3 Break Index marks the end of an Intermediate Intonational Phrase, which contains at least one Pitch Accent followed by a phrase accent. The final syllable of this constituent is lengthened more than at a typical phrase-medial word boundary and less than at a Full Intonational Phrase boundary.

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone

!H*: downstepped high pitch accent

L+!H*: bitonal pitch accent with low tone followed by a downstepped high tone prominence

L*+!H: bitonal pitch accent with low tone prominence followed by downstepped high tone

H+!H*: bitonal pitch accent with high tone followed by downstepped high prominence

L-L%: low phrase accent, low boundary tone

H-H%: high phrase accent, high boundary tone

L-H%: low phrase accent, high boundary tone

H-L%: high phrase accent, low boundary tone

H-: high phrase accent

L-: low phrase accent

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

3: end of an intermediate phrase

4: end of an intonational phrase

Optional labels:

<: late High Tonal peak

2.9 The downstepped High Phrase accent: !H-L% and !H-

The downstepped High phrase accent (!H-), as its name suggests, is a High phrase accent realized in a downstepped pitch range. Just as in pitch accents with a downstepped H element (!H*, L+!H*, L*+!H, and H+!H*), the ! diacritic is used with the H symbol in phrases accents to indicate a “downstepped” High tone. And like pitch accents with a downstepped H element, the “downstepping” in a phrase accent is in relation to a preceding pitch accent with a High tone element. (See section 2.6 for review.) Like the previously introduced phrase accents, H- and L-, the !H- phrase accent can either occur with a boundary tone (in cases where there is a full Intonational Phrase and a 4 break) or without (in cases of an Intermediate Phrase and a 3 break).

2.9.1 H* !H-L%

The !H-L% phrase accent-boundary tone combination is similar to H-L%: it is characterized by a high flat stretch of pitch following the final pitch accent of the phrase. However, rather than being realized at the same high pitch level as the High of the final pitch accent, the phrase accent is realized as a stepped-down high flat.

The example <crazy>, shown in Figure 2.9.1, below, shows a single Full Intonational Phrase with a High pitch accent (H*) followed by a downstepped High phrase accent-Low boundary tone combination (!H-L%). The H* pitch accent is realized on the first syllable, *cra-*, of *crazy*, and the phrase accent and boundary tone are realized on the final syllable, *-zy*. The pitch of the final syllable of *crazy* is perceptibly lower than that of the first syllable, but the pitch remains somewhat high, and does not approach the bottom of the speaker’s pitch range as we would expect from a Low phrase accent (such as in L-L%). The pitch stays flat, at this fairly high level, through the end of the word, forming a “plateau” that is very similar to those seen in examples of the (non-downstepped) H-L% phrase accent-boundary tone combination. (Note that the appearance of a small rise at the end of the phrase is due to a pitch-tracking error. Similar “false highs” occur at the onset of the word, immediately following the /k/ sound and during the /z/ sound in the middle of the word *crazy*.)

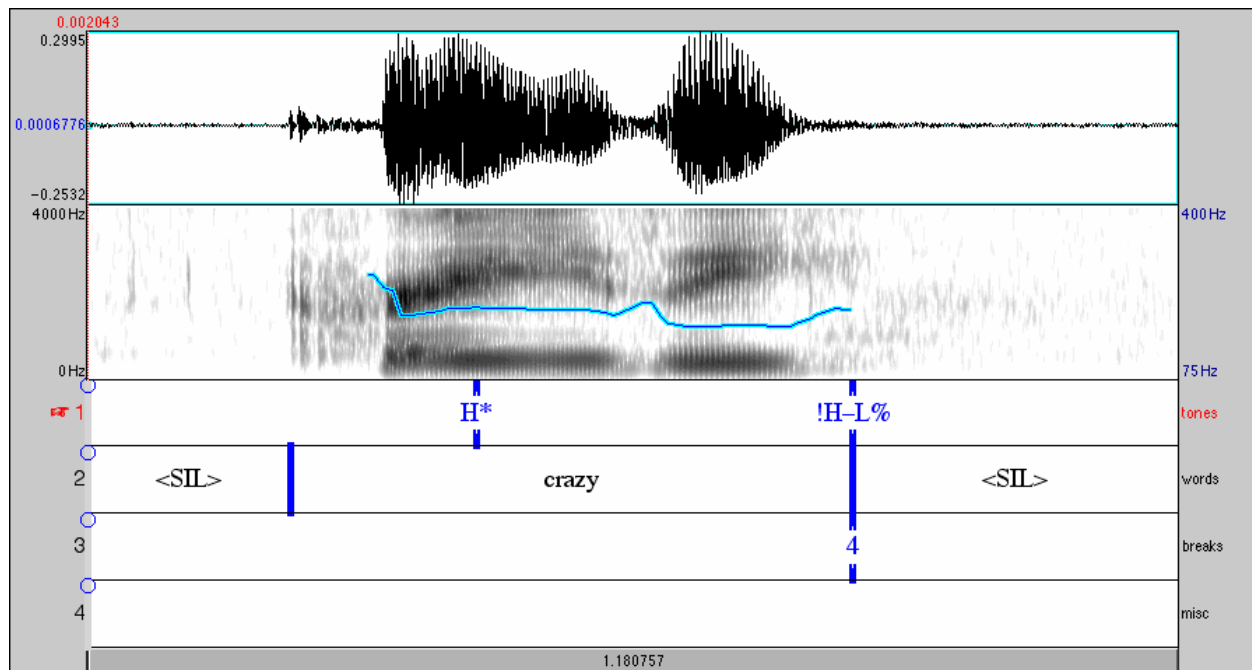


Figure 2.9.1 H* !H-L% on *crazy*

<crazy>

2.9.2 Patterns which are similar to the H* !H-L% sequence

2.9.2.1 H* H-L% compared to H* !H-L%

For contrast with H* !H-L%, an example of H* H-L%, <money2> is reproduced below, in Figure 2.9.2. Notice how, when H* is followed by H-L% rather than a !H-L% (such as can be seen in the final IP of the file <money2> on the words *and money*), the pitch stays quite level from the pitch accented syllable, the *mo-* of *money*, through the final syllable *-ney*. Unlike with the related !H-L% phrase accent-boundary tone combination, the pitch does not fall to a lower level with H-L%.

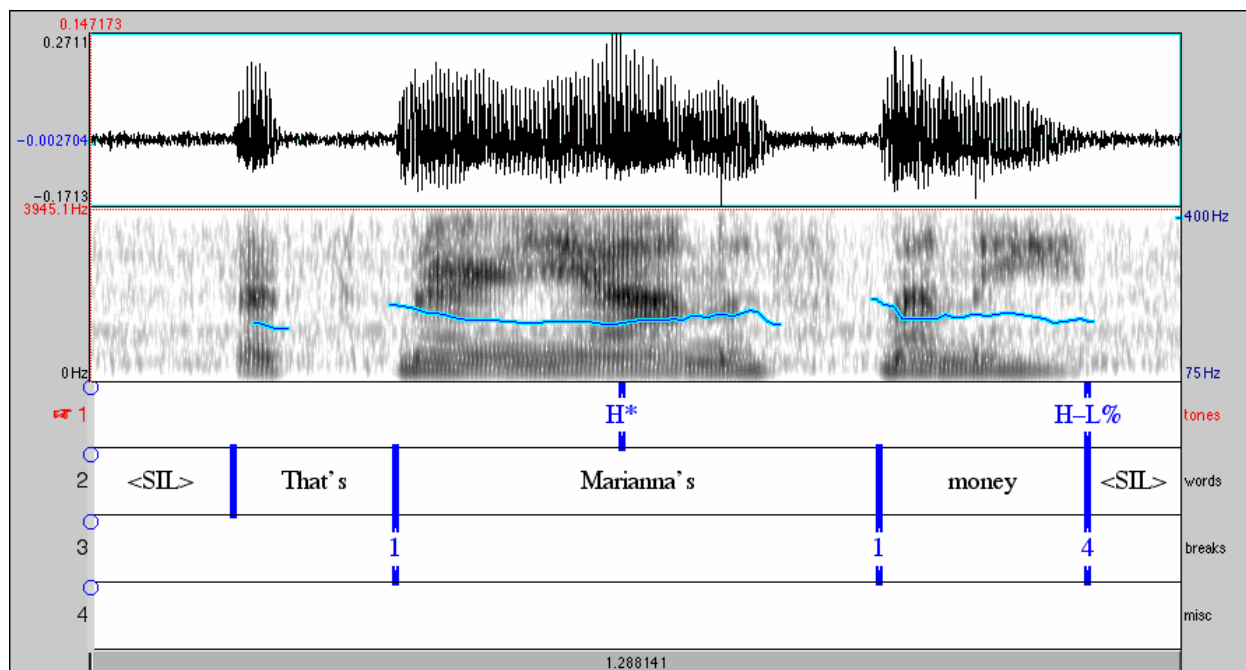


Figure 2.9.2 H* H-L% (shown for comparison with H* !H-L%)

<money2>

2.9.2.2 H* !H* H-L% compared to H* !H-L%

The pitch contour produced in examples of H* !H-L% can also be quite similar in appearance to another previously seen pattern: H* !H* H-L%. Both involve a high pitch associated with a H* pitch accent, followed by a step down to a lower-pitch flat stretch. In H* !H* H-L% sequences, such as in the previously seen example <alejna2> (from section 2.7, reproduced below in Figure 2.9.3), the step down to a lower pitch occurs with a pitch accent; the fall is from the High of the first pitch accent to the lower High of the second pitch accent. The pitch stays flat after the second pitch accent, characteristic of the H-L%. In H* !H-L% sequences, though, there is no other pitch accent-marked prominence between the H* and the phrase accent. This is particularly clear when the pitch accent occurs in the final syllable of the phrase, such that the pitch accent, phrase accent, and boundary tone are all realized in the same syllable: in these cases, there is generally no “room” for a second (!H*) pitch accent. The example <bye> shows H* !H-L% produced on a single syllable, the word *bye*. (Again, the apparent small rise in the f0 at the very end of the word *bye* is due to a pitch-tracking error.)

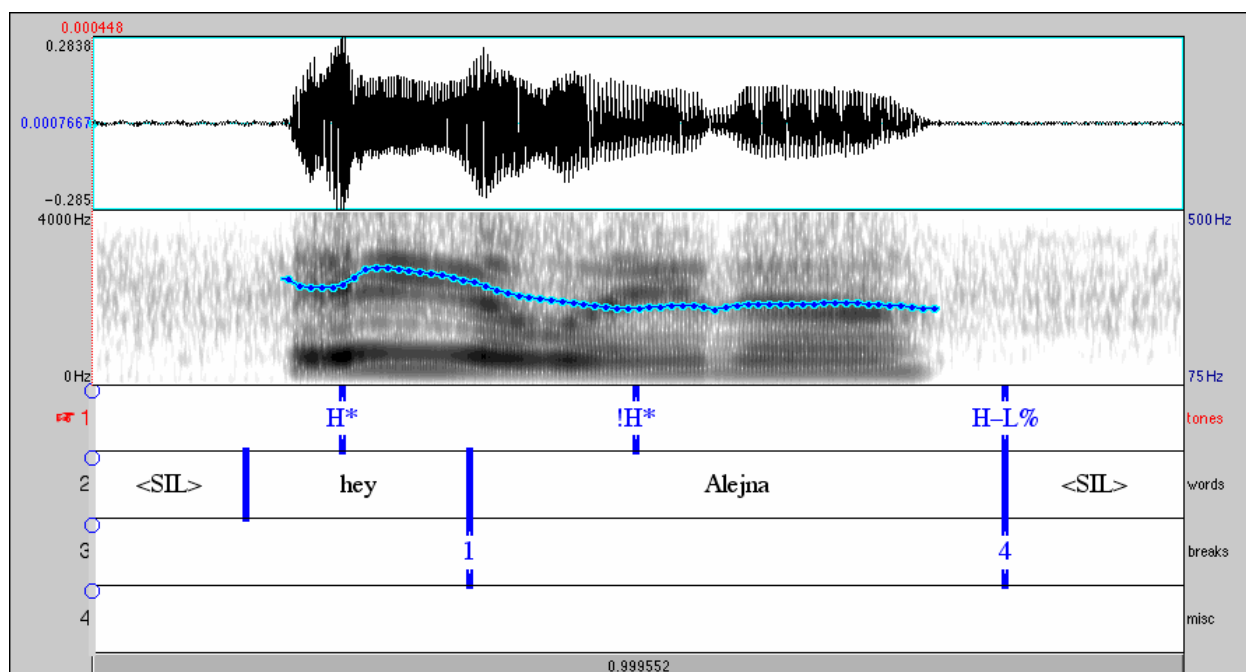


Figure 2.9.3 H !H* H-L% (shown for comparison with H* !H-L%)

<alejna2>

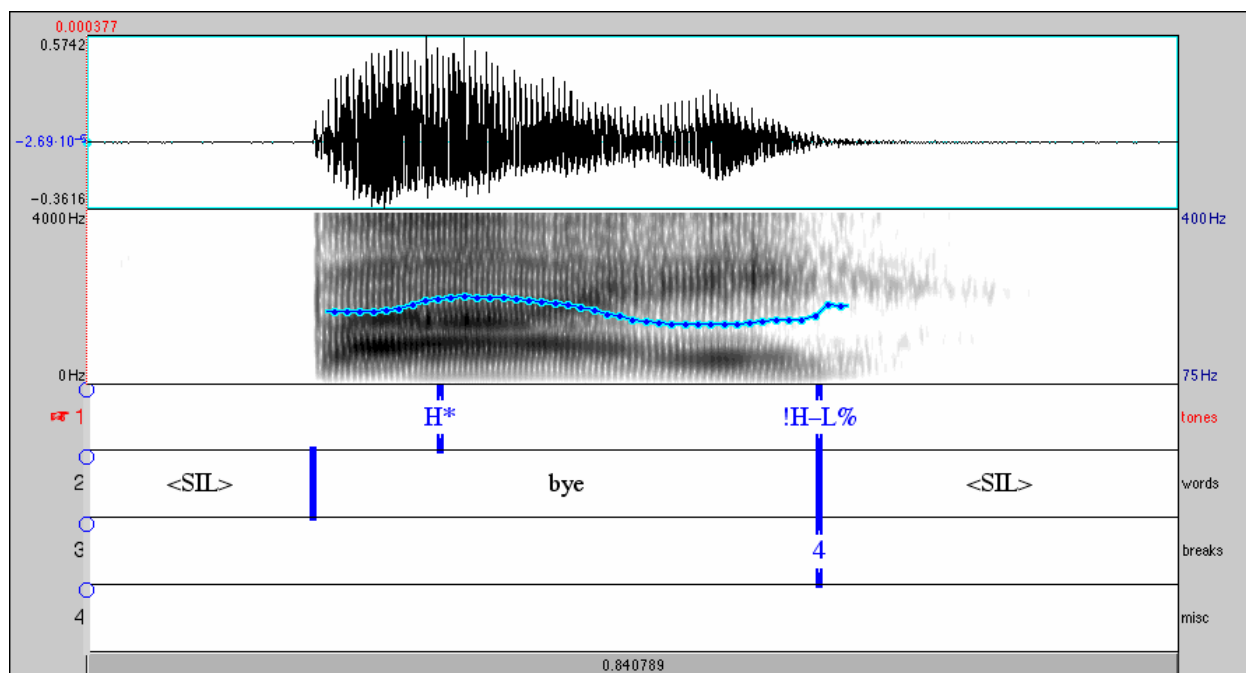


Figure 2.9.4 H* !H-L% on a single syllable

<bye>

The distinction between !H- and !H* is sometimes unclear from the pitch track alone; the labeller must rely on listening, and particularly on perceived prominence when deciding if the step down is to a pitch accent (i.e. !H*) or to the phrase accent (!H-). The example <green_tea> is such a case: there is an H* pitch accent on *green*, and no pitch accent on *tea*. This is an example where it would have been possible for the speaker to have put a second pitch accent on the word *tea*. However, listening to this example shows that the word *tea* is not produced here with a pitch

accent. Therefore, the fall from the High tone of the pitch accent on *green* to the downstepped High on *tea* is attributed to the presence of a !H- phrase accent (and not a downstepped second pitch accent).

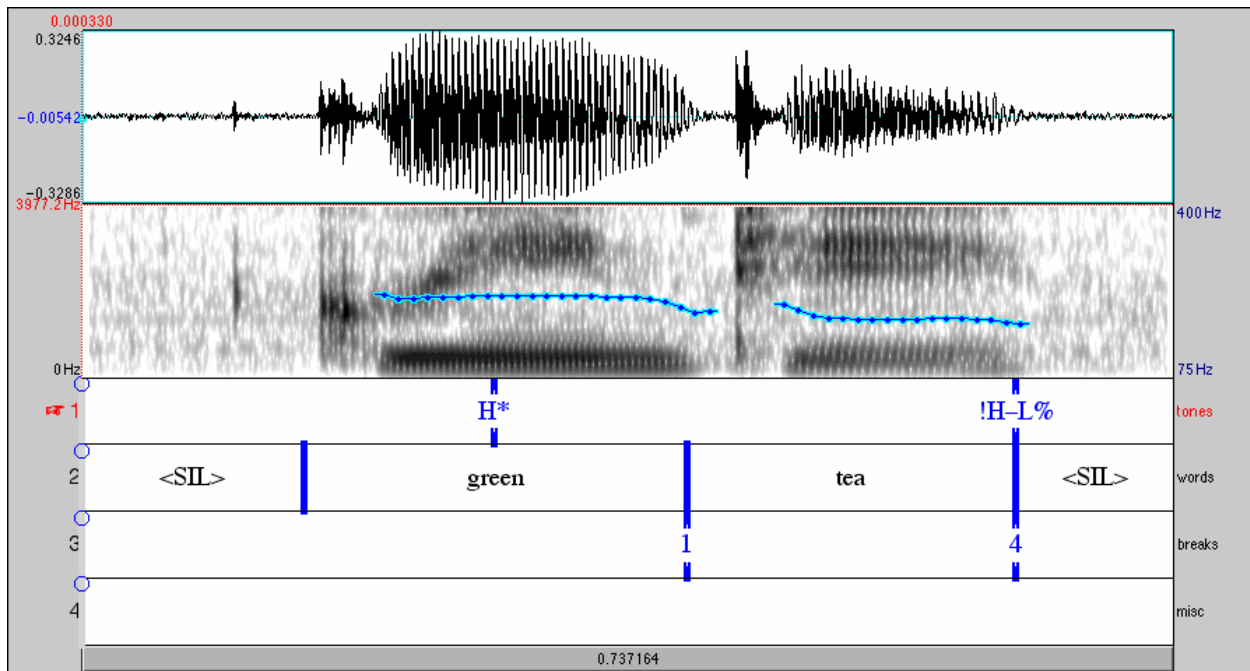


Figure 2.9.5 Another example of H* !H-L%

<green_tea>

2.9.2.3 H+!H* H-L% compared to H* !H-L%

Yet another (previously introduced) contour which is similar to H* followed by !H-L% is when H-L% follows the bitonal pitch accent H+!H*. The H+!H* H-L% sequence, like H* !H-L% or H* !H* H-L%, is characterized by a high pitch followed by a stepped-down high flat stretch. However, in this case, the fall occurs going into the pitch accent, characteristic of H+!H*, not after it, as would be appropriate for a !H- phrase accent. The file <minimum2> is reproduced below, in Figure 2.9.6, as an example of the H+!H* H-L% sequence.

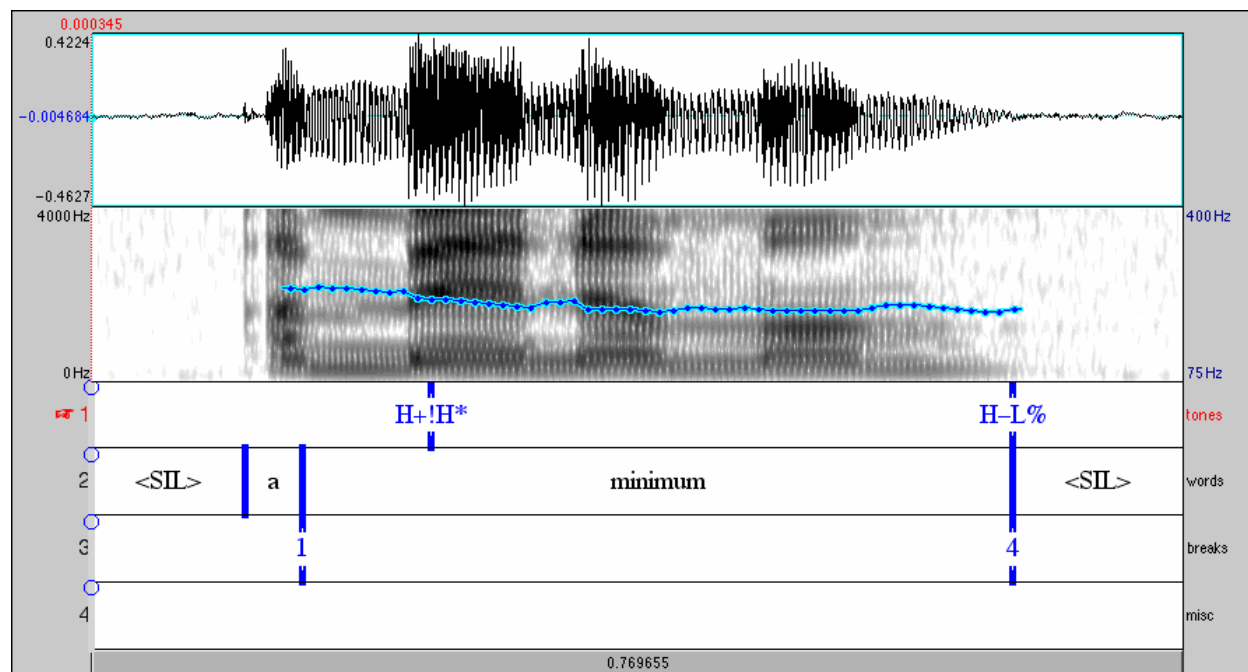
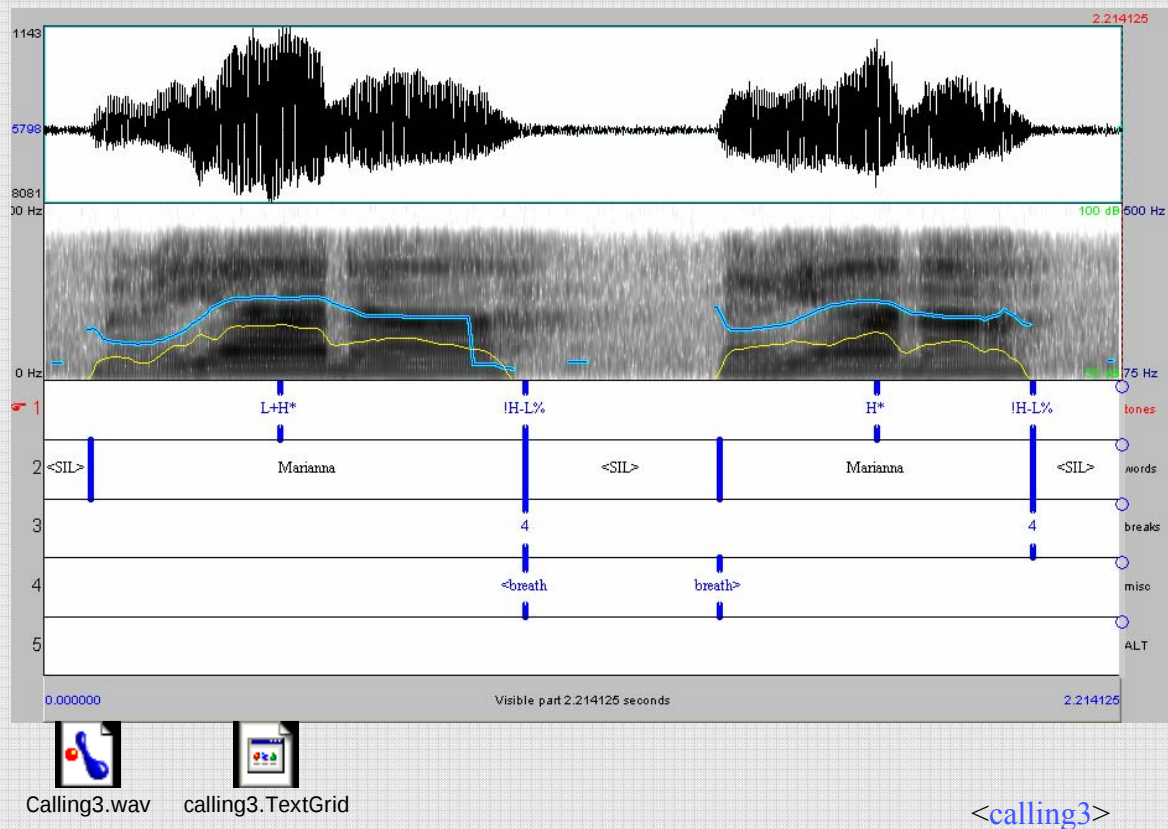


Figure 2.9.6 H+!H* H-L% (shown for comparison with H* !H-L%)

<minimum2>

The Calling Contour:

The !H-L% phrase tone-boundary tone combination is familiar to many as part of what are known as “calling contours” (often used when pronouncing a name with the purpose of calling the person named). The example <calling3> shows two variations of calling contours on the name “Marianna,” one with H*, one with L+H*, and both with !H-L%.



2.9.3 !H-L% following other pitch accents

The !H-L% combination can theoretically occur after any pitch accent with a High tone element. The example <in_here> has !H-L% following an instance of the bitonal pitch accent L+H*. While the pitch in this example has an initial low period, and a sharp rise into the pitch accented syllable (characterizing the L+H* pitch accent), the realization of the !H-L% afterwards is very similar to that of !H-L% following the single-tone H* pitch accent; the pitch drops slightly to a still fairly high level, and then stays flat at that high level through the end of the phrase.



The !H- phrase accent may also occur at the end of an intermediate phrase, with a 3 break, just like H- and L- (introduced in section 2.8). As in all other !H cases, the !H- phrase accent associated with a 3 break follows a pitch accent with a High tone element (such as H*, L+!H* or L*+H) and is realized with a pitch that is lower than that preceding pitch accent. And like other !H cases, the !H- phrase accent associated with a 3 break is still realized in a relatively high pitch range for that speaker, especially compared to Low tones, such as L* or L-. The !H- phrase accent is generally realized as a fairly flat stretch in pitch, much like a shortened version of !H-L%. The example <heavy_rain>, shown in Figure 2.9.8, below, shows an intermediate phrase on the words “heavy rain,” where the phrase accent is a downstepped High phrase accent, !H-, realized at the end of the word *rain*. Here, the !H- is downstepped with respect to the L+H* pitch accent immediately preceding it, also on the word *rain*. The degree of disjuncture is compatible with the end of an intermediate phrase and a 3 break. There is neither enough lengthening nor intonational marking to indicate a larger break (a 4 Break Index), as in a full intonational phrase. Looking closely at the pitch of this example shows that after the peak associated with the L+H* pitch accent early in the word *rain*, the pitch lowers a bit and continues fairly flat through the end of the word. (The final sharp drop in the pitch track at the end of the word *rain* is not audible, and is most likely a pitch-tracking error due to segmental effects.)

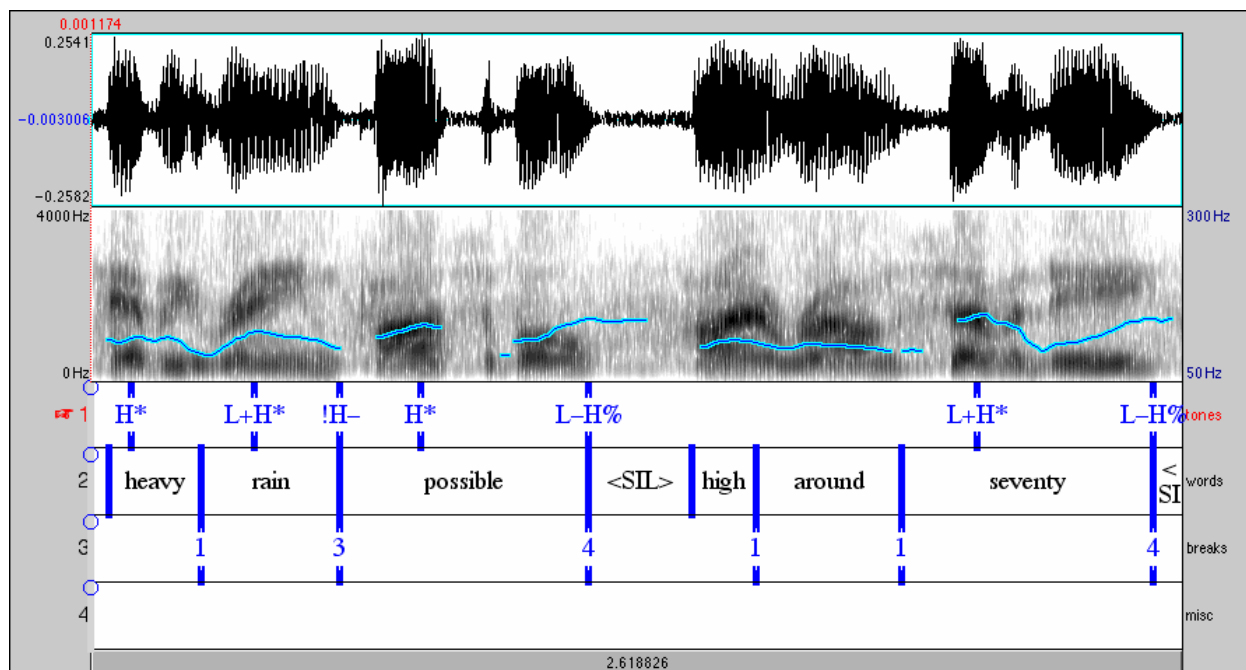


Figure 2.9.8 an example of !H- with a 3 break

<heavy_rain>

The !H- phrase accent may often be confusable with L-; both are characterized by a lowering of pitch after the last pitch accent of an intermediate phrase. However, !H- is generally used for cases where the pitch does not reach as low a level as might be expected from L-. Labellers shouldn't agonize over whether to use !H- vs. L-: when in doubt, use L-, and save !H- for cases where L- does not capture a sustained high pitch (that is relatively lower than the preceding pitch accent) of the phrase accent.

2.9.5 Other logical possibilities with downstep

The examples above of the !H- phrase accent have all been either with a Low boundary tone (!H-L%), or as the phrase accent in an intermediate phrase (!H- with no boundary tone). You may have realized that another possibility exists, namely !H- followed by a High boundary tone, or !H-H%. This phrase accent-boundary tone combination is theoretically possible, and may be encountered in labelling. However, it may also be hard to distinguish from the L-H% phrase accent-boundary tone combination. Labellers may use the !H-H% label combination when they feel that it best captures the tone pattern of a phrase.

The !H- phrase accent is the last downstepped tone type used in ToBI. The label !H% does not exist, nor are there downstep labels with Low tones. (There are no !L* or !L- labels in ToBI.)

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone
!H*: downstepped high pitch accent
L+!H*: bitonal pitch accent with low tone followed by a downstepped high tone prominence
L*+!H: bitonal pitch accent with low tone prominence followed by downstepped high tone
H+!H*: bitonal pitch accent with high tone followed by downstepped high prominence
L-L%: low phrase accent, low boundary tone
H-H%: high phrase accent, high boundary tone
L-H%: low phrase accent, high boundary tone
H-L%: high phrase accent, low boundary tone
!H-L%: downstepped high phrase accent, low boundary tone
H-: high phrase accent
L-: low phrase accent
!H-: downstepped high phrase accent

Break indices:

0: word boundary erased
1: typical inter-word disjuncture within a phrase
3: end of an intermediate phrase
4: end of an intonational phrase

Optional labels:

<: late High Tonal peak

2.12 Uncertainty in labelling

While many stretches of speech will be straightforward to label with ToBI, many others may prove more challenging to the labeller. Naturally-produced speech contains an enormous amount of variation, and there are cases where a labeller may be uncertain what labels to use. There are times when the interpretation of the prosody will be ambiguous, and more than one ToBI label may seem plausible.

Table 2.12.1 sources of variation that may lead to uncertainty

• individual differences between speakers (obvious differences like pitch, high-pitched or low-pitched voice)
• differences even for the same speaker (changes in rate, loudness, affective differences—boredom vs. excitement)
• dialectal differences: speakers from different parts of the country, while still speaking what is considered Mainstream American English, may pronounce some words differently. Likewise, there may be some dialectal variation in intonation
• stylistic differences (news broadcasters or lecturers vs casual conversation)
• a host of other confounding factors (which will be addressed in a new chapter in a future version of this tutorial): speech rate, segmental effects, ambient noise if the recording was not done under ideal conditions like static or background noise, signal dropout, other speakers, laughter, yawning, coughing, irregularities of voicing due to illness (laryngitis or a cold) or even tiredness.

The ToBI system provides a variety of conventions for addressing ambiguity and other sources of labeller uncertainty, including uncertainty labels and the alternatives tier.

2.12.1 The alternatives (alt) tier (Note for 2006 draft: all content on the alt tier is still under development, and the conventions described are subject to change)

The alt tier, a recent addition to the ToBI system, allows the labeller to indicate places where more than one label was seriously considered. There may be times when a labeller spends a comparatively long time determining which of two possible labels to use for a particular tone or break. The alternatives tier allows an outlet for the labeller to include both labels, so that she can move forward with more speed, and less frustration.

Use of the alternatives tier, along with other markers of uncertainty which will be discussed later in this section, also provides a record of which regions and points in time in the speech signal are harder to label, so that researchers may investigate the sources of ambiguity. Without uncertainty markers, labels would not differentiate between those stretches of speech which are straightforward to label, and those which contain ambiguities.

2.11.2 Uncertainty regarding pitch accent type

A common situation of uncertainty arises when the labeller has determined that a syllable is pitch-accented, but has trouble deciding which pitch accent type best accounts for the pitch pattern observed. Most often, the labeller is considering between two specific pitch accent types, such as H* vs. L+H*. The labeller might feel that the best label to use would be H*, but that L+H* is a reasonable alternative. In such a case, the labeller should use the label H*? in the tones tier, and put L+H* in the alt tier. The ? symbol indicates, simultaneously, that the labeller was somewhat uncertain about the choice of pitch accent label, and that the alternatives tier has been used to indicate the other pitch accent that the labeller considered. The example <indeed>, shown in Figures 2.12.1 & 2.12.2, contains two instances where the labeller was not certain about which pitch accent label to use: first on the syllable –*deed* of *indeed*, and then later in the file, on the monosyllabic word *clouds*. In the first full intonational phrase of the file (shown in Figure 2.12.1) with the words *and indeed*, there is a strong prominence that is marked by a peak in the pitch track on the syllable –*deed* of *indeed*; the syllable is clearly pitch-accented. However, the labeller was not certain whether the pitch accent on –*deed* should be labelled with the single-tone H*, or with the bitonal L+H*. The pitch at the beginning of the phrase starts out moderately high in the speaker's pitch range, and rises slightly into the first syllable of *indeed*, –*in*, and then rises to a peak in the pitch-accented syllable. There is an apparent fall in f0 shown in the pitch track at the end of the syllable *in*-, but this can be attributed to segmental effects. In fact, these segmental effects between the two syllables of the word *indeed* make the f0 tricky to interpret. The rise into the pitch accented syllable appears to be quite sharp, as would be expected from the bitonal pitch accent L+H*. However, the labeller here felt that the pitch accent sounded more like an exaggerated or emphatic production of a single-tone H* pitch accent.

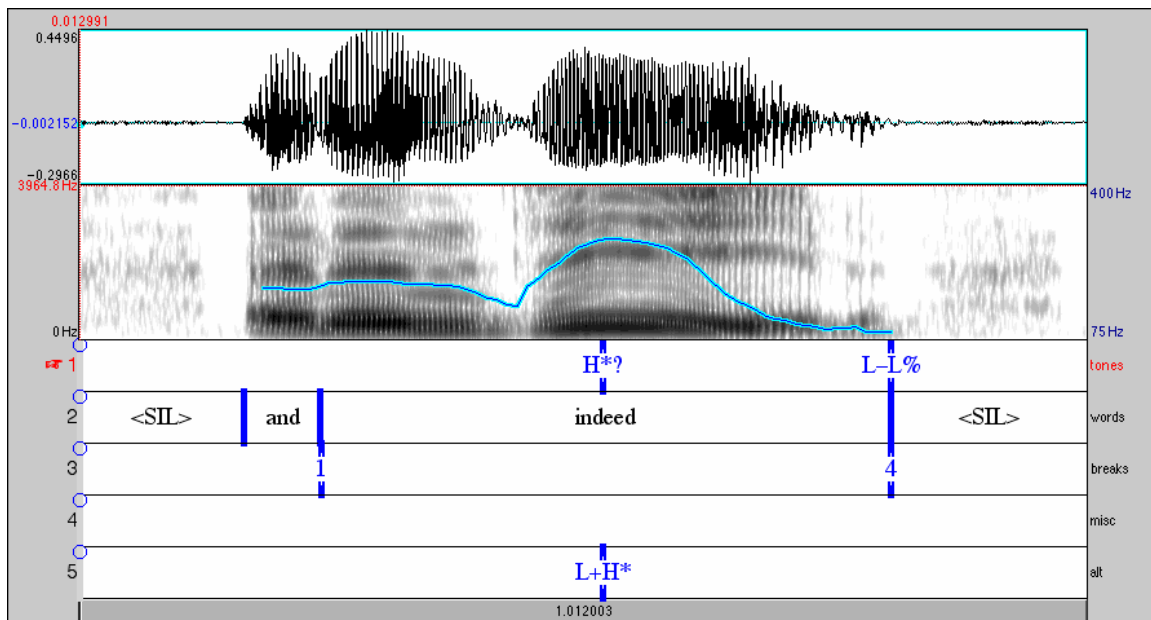


Figure 2.12.1 Using the alt tier and the ? symbol to indicate uncertainty about the choice of pitch accent label in the first intonational phrase of the file *indeed.wav* <indeed>

In the second Intonational Phrase of the file <indeed>, shown in full in Figure 2.12.2, below, the labeller has used $L+H^*?$ on the word *clouds*. In this case, the second option considered by the labeller was $L+!H^*$ (where the bitonal has a downstepped High). In this case, the labeller perceived the pitch height of the peak of the pitch-accented word *clouds* as being about the same as the pitch height reached in the preceding pitch accent on the *-lec-* of *electron*. However, the pitch track shows that the f_0 of the peak on *clouds* as lower than the peak of the first pitch accent. The labeller gave precedence to the choice of $L+H^*$ (with no downstep), and put that label in the tones tier, along with the “?” marker. Then, she put her second choice, $L+!H^*$, in the alt tier.

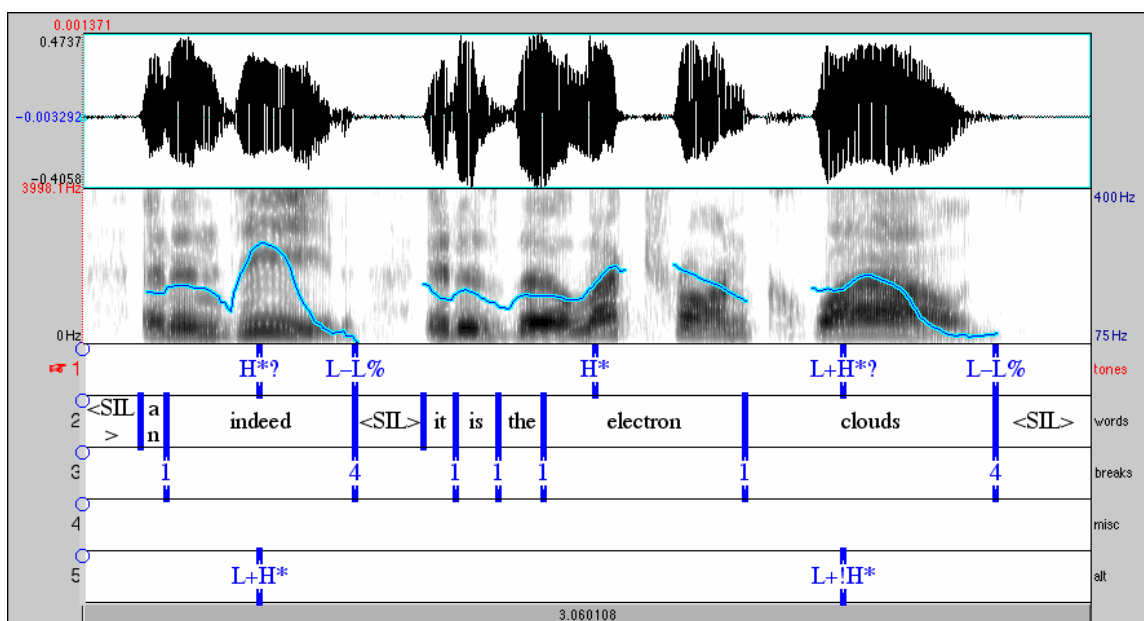


Figure 2.12.2 Another example of using the alt tier on *clouds* in the second intonational phrase
<indeed>

2.12.3 The X*? label:

While a labeller will usually have a first choice of label when considering two alternatives, there may be times when the labeller will be unable to decide which pitch accent type to put in the tones tier. The label X*? can be used for such cases. Whenever possible, the labeller should still use the alternatives tier, and list both alternatives there. For example, if the labeller can't decide between L+H* and H*, and isn't happy committing to either one for the tones tier, the labeller can put X*? in the tones tier, and both L+H* and H* in the alt tier. The labeller should align a single point in the alt tier with the X*? in the tones tier, and list both alternatives at that point, separated by a "pipe" ("|") symbol, eg. "H*|L+H*".

Table 2.12.2 Commonly confusable pairs of pitch accents

H* vs L+H*:
The pitch accented syllable clearly has a High prominence, but the rise may be more gradual than expected for the bitonal L+H*, but steeper than would be expected for a single-tone H*, and there may be varying degrees of evidence for a preceeding L target.
L* vs H* (in a compressed pitch range):
The syllable is clearly prominent, but the pitch track is more or less flat, with no clear peaks (as often occurs with H*) or dips (as one would expect for L*), and the labeller is uncertain whether the region in question is low enough in the speaker's pitch range to justify Low tones (L*) or in a mid-range, such as would happen with H*.
L+H* vs L*+H:
There is a distinct low and a distinct high associated with a pitch accent, but the labeller is uncertain which tone, L or H is more strongly associated with the prominence
H* vs !H*:
A pitch accent following another High tone pitch accent (such as H* or L+H*) is

somewhat lower in f0 than that preceding accent, but the labeller is uncertain that the tone is perceptually low enough to be a !H*.

H+!H* vs !H*:

The labeller is uncertain whether a peak leading to the !H* prominence is associated with that pitch accent, or if it is associated with some preceding High tone pitch accent (such as a late peak after a H*).

2.12.4 Uncertainty about whether or not a syllable is pitch-accented

There will be cases of uncertainty relating to pitch accent that are not only about which type of pitch accent to use, but about whether to use a pitch accent label at all. The label *? is used for just such cases, and its use will be discussed below.

Levels of prominence:

As discussed in section 2.3, syllables in spoken language can be produced with a range of degrees of prominence: it is not the case, for example, that all syllables which do not bear a pitch accent are equally weak. Some syllables can be prominent compared to neighboring weaker syllables, but still less prominent than a syllable bearing a pitch accent. To make matters more complex, not all pitch-accented syllables are equally prominent either. For example, it may be the case that in an intonational phrase with two pitch accents, one pitch-accented syllable may be much perceptually stronger, and more salient than the other. The less-strong pitch-accented syllable is still pitch-accented, nevertheless.

There may be times, though, where it is not clear whether a syllable, which has some degree of prominence, is in fact pitch-accented. The example <august> shows an instance where the labeler was not certain whether a particular syllable was pitch-accented or not. In this example, shown in Figure 2.11.3, the labeler has used the *? label on the word *how* in the second intermediate phrase, *for how long* in that file. In this example, the labeler's use of the L* pitch accent symbol on the word *long* shows that she was certain that the word was pitch accented, and that the pitch patterns were best captured by the Low pitch accent. However, the labeler felt that the word *how* could also be pitch-accented. While not as prominent as *long*, the word seems to be fairly prominent. The pitch track also shows some ambiguity: there is a slight rise in f0 at the beginning of the word *how*. While this is potentially the result of a pitch-tracking error, the labeler may have been uncertain whether the rise in pitch, combined with the sense that the word *how* is fairly prominent, was evidence of a pitch accent on that syllable. The labeller addressed this uncertainty by using the *? label in the tones tier. Then, in the alt tier, the labeller put the type of pitch accent label that she would use had she felt confident that indeed the syllable was pitch-accented.

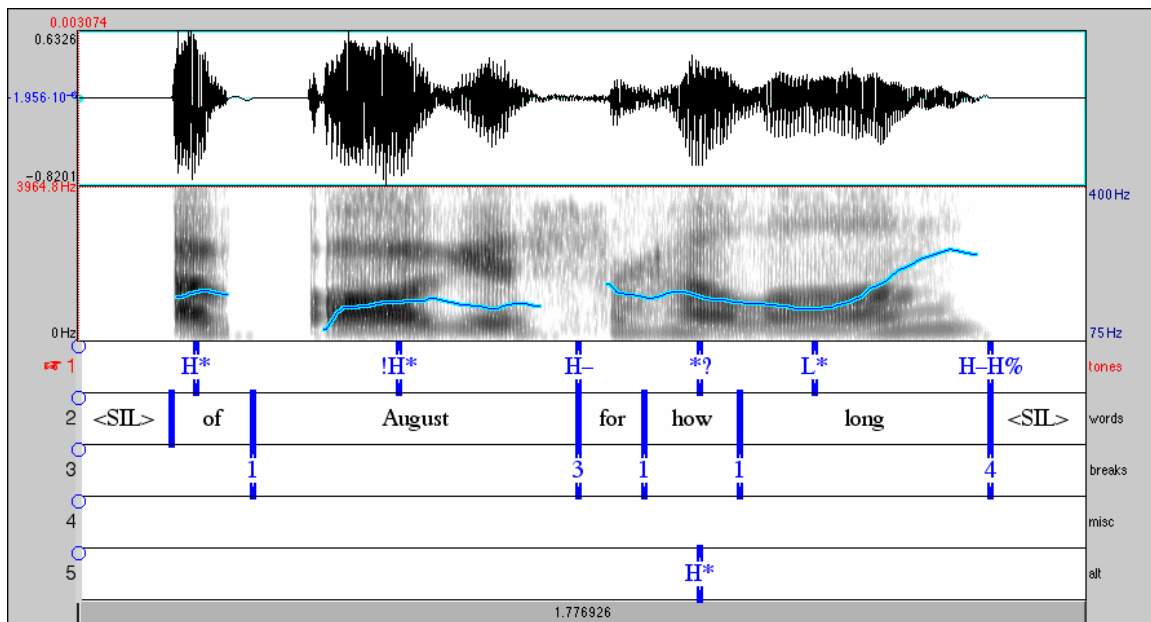


Figure 2.12.3 Using the alt tier and the *? symbol to indicate uncertainty about the choice of pitch accent label on *how* <august>

Sources of uncertainty/ambiguity:

There are a number of fairly common phenomena that make it difficult to determine whether or not a certain syllable is pitch-accented. It is sometimes the case that a labeller will be uncertain whether a strong syllable between two (clearly) pitch-accented syllables also contains a pitch accent. This can happen when the syllable in question occurs during a stretch of high pitch between two H* pitch accents, or comparably during a stretch of low pitch between two L* pitch accents.

The example <marmalade7> shows an example of a *? label in a region between two L* pitch accents. In this example, there is a Low-tone pitch accent on the *-an-* of *Marianna*, and then another Low-tone pitch accent on the *mar-* of *marmalade*. The pitch lowers into the first of these two pitch accents, and stays low through the second. Notice how one of the syllables between these two pitch accents, the monosyllabic word *made*, is also fairly strong, and could potentially also have a Low (L*) pitch accent: the signal is ambiguous here. The labeller has captured this ambiguity, and resulting uncertainty, by using the *? label in the tones tier, and L* in the alt tier.

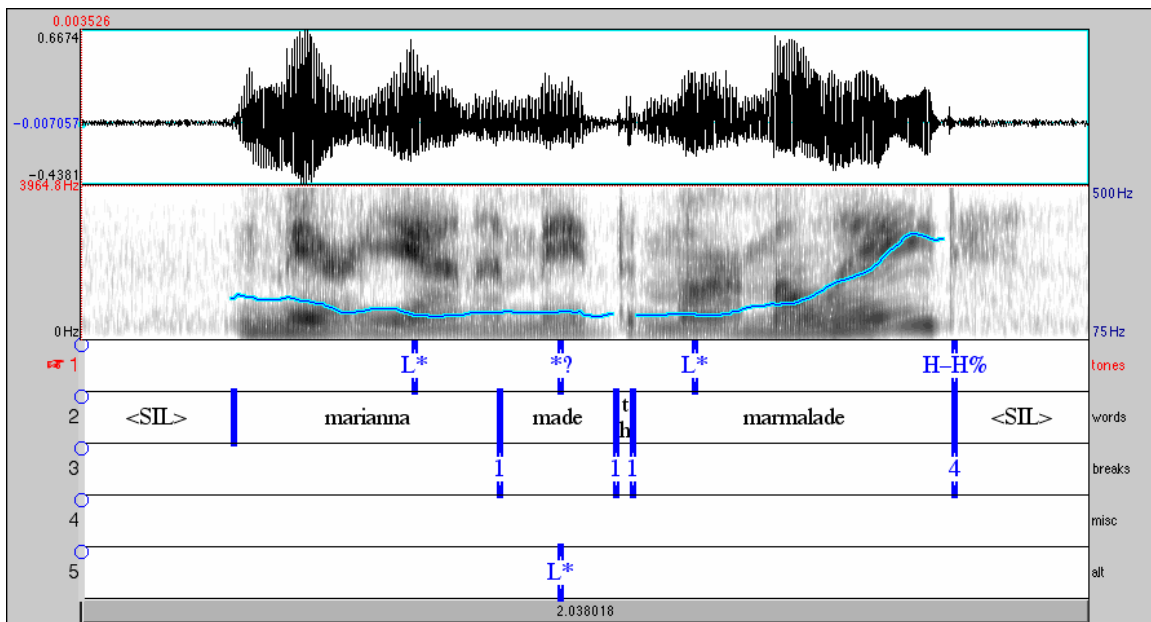


Figure 2.12.4 Using *? to mark uncertainty on *made*

<marmalade7>

The example <anna3>, shown in Figure 2.12.5, shows a parallel case to the one discussed above: in this case, a strong syllable produced in a stretch of high pitch between two syllables with High (H*) pitch accents could potentially also be pitch-accented with H*. Here, while the *An-* of *Anna* and the *Len-* of *Lenny* both are clearly pitch-accented (with H* pitch accents) the syllable *mar-* of *married* could also potentially be pitch accented; the label *? is used for this syllable in the tones tier, and H* in the alt tier.

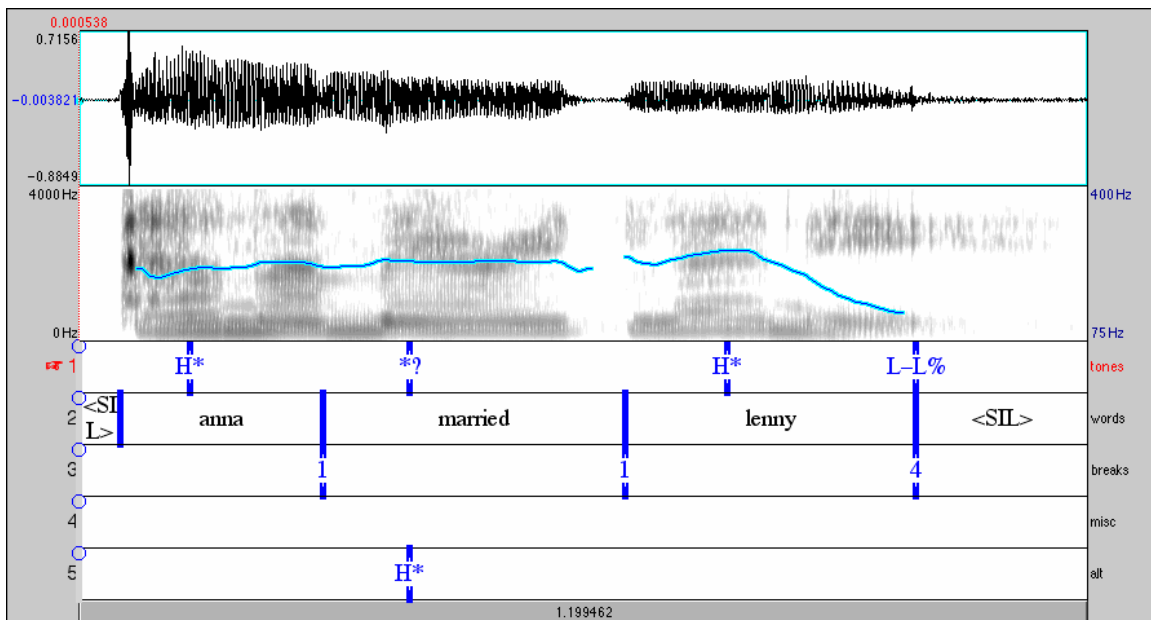


Figure 2.12.5 Using the alt tier and the ? symbol to indicate uncertainty about the choice of pitch accent label on *mar-* in *married*

<anna3>

Table 2.12.3 Other common contexts for *? (in addition to those discussed above):

• after one or more !H* pitch accents, such that the local pitch range has been greatly compressed
• when the speaker is generally using reduced pitch range
• when there is a phrase final full-vowel syllable, especially a monosyllabic word, where the duration cues to accent and boundary may be confounded.
• a phrase-initial full-vowel syllable, especially starting with a vowel (glottal stop)
• 2 consecutive full-vowel syllables: both seem strong, and it may be hard to tell which of the two is pitch accented—there may be times when there is pitch movement in the area of these two syllables that could be associated with an accent on one or the other or both.

2.12.5 Uncertainty regarding level of disjuncture

Labellers may also encounter times when they are uncertain about the appropriate break index. There are two ways which the labeller may signal uncertainty about level of disjuncture, or type of break index, which will be discussed in the following sections: 1) using the “minus” diacritic (“-”) and 2) using the alternatives tier.

2.12.5.1 The “minus” diacritic:

The minus diacritic is used in conjunction with a break index number to indicate uncertainty about break index level where there are 2 clear alternatives, such as about whether break index 4 or break index 3 best captures the degree of disjuncture. Conventions for use of the “-” diacritic prescribe that in such cases, the higher break should be used in the breaks tier, with the “-” diacritic. So, uncertainty between break index level 4 and level 3 can be indicated with the break index of “4-”. It is not necessary to use the alt tier for such cases, as the two alternative labels are understood to be 4 vs. 3. In such cases, the labeller must mark the phrase accent and boundary tone associated with the higher break level, the 4. In the example <sure>, shown in Figure 2.12.16, below, the labeller was uncertain whether the disjuncture between the words *sure* and *of* would best be captured by break index 4, indicating the end of a full intonational phrase, or whether break index 3, indicating an intermediate phrase boundary, would best fit. The lengthening in the word *sure* seems a bit short for a 4, but perhaps longer than might be expected for a 3. The labeller captured this uncertainty, and the ambiguity in the signal, by using the label 4- in the breaks tier. The labeller also put the corresponding phrase accent-boundary tone combination of L-L% in the tones tier, corresponding to a 4 break. The use of the 4- break index with L-L% in the tones tier indicates that the labeller considered that a break index 3, and a corresponding L- phrase accent, was a possible alternative.

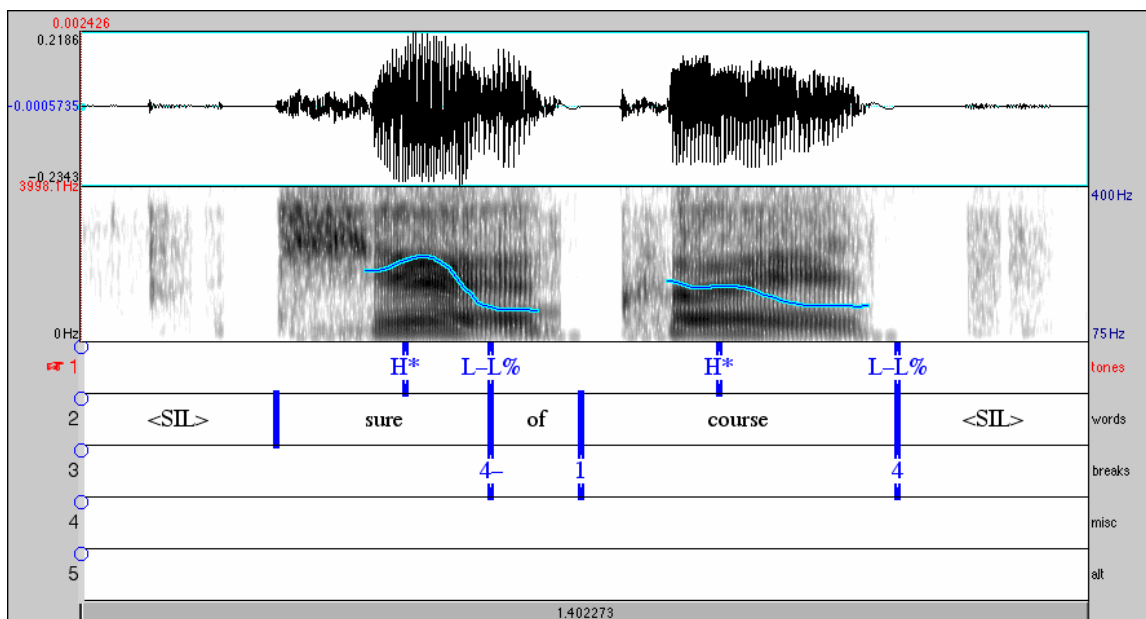


Figure 2.12.16 Using a 4- diacritic to indicate uncertainty about the level of break index on *sure* at the end of the first intonational phrase <sure1> (version 1)

2.12.5.2 Using the alt tier for uncertainty about break index level

The use of the “minus” diacritic does not make it explicit which of the two alternatives the labeller preferred. For cases where the labeller wishes to indicate a strong preference for one break label over the other alternative, the labeller may make use of the alt tier. Figure 2.12.16, below, shows the same file, <sure>, with the alt tier used instead of the “minus” diacritic. In this case, the labeller shows a stronger preference for considering that the end of the word *sure* corresponds to the end of an intermediate phrase (rather than a full intonational phrase). The labeller has put break index 3 in the breaks tier, followed by the “?” diacritic, and the phrase accent L- in the tones tier, also followed by the “?” diacritic. Aligned to the time of both of these labels, the labeller puts a mark in the alt tier with both the alternative break index, and the corresponding phrase accent-boundary tone combination with the “&” (“ampersand”) symbol between them: 4&L-L%.

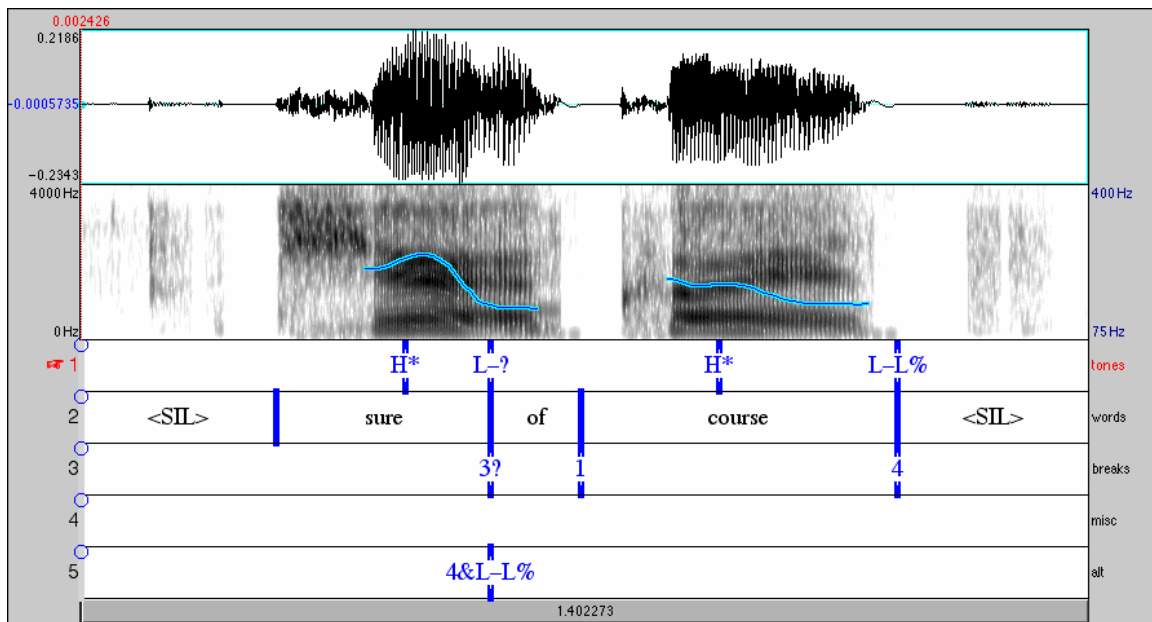


Figure 2.12.16 Using the alt tier and the ? symbol to indicate uncertainty about the choice of break index on *sure* at the end of the first intonational phrase
 <sure2> (version 2)

2.12.6 Uncertainty about phrase accent or boundary tone type:

There will be times when the labeller is uncertain what the appropriate phrase accent or boundary tone is for cases of intermediate and full intonational phrases. In such cases, labellers can again use the “?” diacritic with the tone markers in the tones tier, and list alternatives in the alt tier. In the example <diagonal>, the labeller indicated uncertainty about which boundary tone best would capture the tone patterns at the end of the final full intonational phrase of the file, realized on the word *mountain*. Here, the labeller was uncertain that the appropriate phrase accent was H-, but less certain about the appropriate boundary tone. While the pitch track seems to show a final rise, as would be characterized by H-H% following a H* pitch accent, the labeller perceived the pitch as being fairly flat, and not especially rising, as would be expected from the sequence H* H-L%. To capture that the pitch appeared to be rising more than expected, the labeller used the “?” diacritic after the phrase accent-boundary tone combination in the tones tier, and put H-H% in the alt tier. (Note that the labeller repeated the phrase accent part of the combination in the alt tier, rather than simply listing the boundary tone alone.)

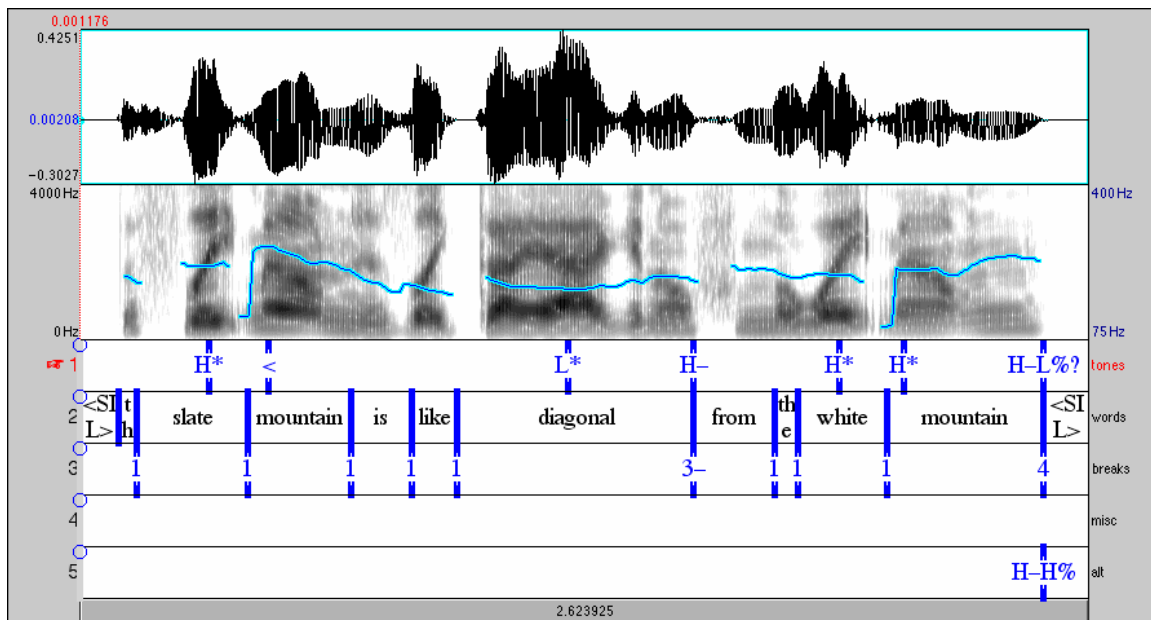


Figure 2.12.17 Using the alt tier and the ? symbol to indicate uncertainty about the choice of phrase accent and boundary tone at the end of the intonational phrase
<diagonal>

In cases where there is uncertainty about the phrase accent portion of a phrase accent-boundary tone combination, the labeller should put the “?” diacritic following the boundary tone in the tones tier, and list the alternative phrase accent type with boundary tone in the alt tier. (For example, if uncertain whether a phrase accent should be !H- or L- in a phrase accent-boundary tone combination, the labeller should put, for instance, !H-L%? in the tones tier and L-L% in the alt tier.)

2.12.6.1 X- and X%

ToBI has an additional two uncertainty markers, which can be used where the labeller has no “best guess” about which phrase accent or boundary tone to use: X-? and X%? respectively. These should be used only as a “last resort,” such as in cases where the recording quality is very poor.

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone

!H*: downstepped high pitch accent

L+!H*: bitonal pitch accent with low tone followed
by a downstepped high tone prominence

L*+!H: bitonal pitch accent with low tone prominence followed

by downstepped high tone
H+!H*: bitonal pitch accent with high tone followed by
downstepped high prominence
L-L%: low phrase accent, low boundary tone
H-H%: high phrase accent, high boundary tone
L-H%: low phrase accent, high boundary tone
H-L%: high phrase accent, low boundary tone
!H-L%: downstepped high phrase accent, low boundary tone
H-: high phrase accent
L-: low phrase accent
!H-: downstepped high phrase accent

Break indices:

0: word boundary erased
1: typical inter-word disjuncture within a phrase
3: end of an intermediate phrase
4: end of an intonational phrase

Optional labels:

<: late High Tonal peak

Uncertainty markers:

*?	uncertainty marker used on the tones tier to indicate that the labeller is not sure whether the specific syllable is pitch-accented
X*?	uncertainty marker used on the tones tier to indicate that the syllable in question is definitely pitch-accented, but the labeller is unable to determine or decide which type of pitch accent
?	used in the tones or breaks tier to indicate that the Alternatives (alt) tier has been used for a particular label
-	used after a break index (on the breaks tier) number to indicate uncertainty between two levels of disjuncture (4-, 3-, 1-)
4-	uncertainty marker to indicate that the labeller is uncertain whether the break in question is a 3 or a 4. Both the phrase accent and boundary tone must be marked in these cases, and the break index should be considered a 3.
3-	uncertainty marker to indicate that the labeller is uncertain whether break should be 3 or a smaller index. The phrase accent must be marked.
1-	1 or 0
X-?	uncertainty marker to indicate that the labeller is unable to determine the type of phrase accent present at the end of an Intermediate Phrase (or in conjunction with a boundary tone, at the end of a full Intonational Phrase)
X%?	uncertainty marker to indicate that the labeller is unable to determine the type of boundary tone present at the end of a full Intonational Phrase
X-H%?	unable to determine whether L-H%, H-H% or !H-H%
X-L%?	unable to determine whether L-L%, H-L% or !H-L%

H-X%?	H-L% or H-H%
L-X%?	L-L% or L-H%
!H-X%?	!H-L% or !H-H%

2.10 Break index 2

Note for this 2006 ToBI tutorial:

There is one remaining break index that we have not yet discussed: break index 2. The 1993 ToBI tutorial (Ohio State) noted that “In the course of developing the ToBI transcription system, we encountered several utterances in which we felt a strong sense of disjuncture at a boundary between two words where the pitch pattern showed no evidence of the necessary tonal events for either of these (3 or 4 break level) levels of intonational constituency,” As described in detail below, break index 2 was devised to label these mismatched cues that provide conflicting evidence either for an intermediate phrase break or an ordinary intra-word break. However, since then it has not proved to be a simple matter to divide the set of potential 2’s into the two kinds of mismatch described below in the 1993 ToBI Labelling Guide. Cases like ‘Six southern Iraqi cities’ in the discussion below, where timing cues suggest 3-breaks but there are no compelling intonational cues to a Phrase Tone, are not unusual, but cases like ‘Quincy’, where timing cues are consistent with an ordinary 1-break but intonational cues suggest a Phrase Tone, are often parsed differently by different transcribers. In addition, some labelers find the 2 diacritic useful for cases in which there is no actual mismatch, but the labeler perceives a boundary between two words that is larger than a 1 but smaller than a 3. We are currently discussing options for re-phrasing the discussion of Break Index 2, and it is anticipated that changes may be made to this section to reflect evolving understanding of this level of disjuncture. For now, we include the current sections of the 1997 3rd revision of the ToBI Labelling Guide.

From the 1997 3rd revision of the ToBI Labelling Guide

(http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf):

As noted in the previous section, each 3 on the break index tier must correspond to the marking of a phrase accent for the intermediate phrase on the tone tier, and each 4 must correspond to the marking of a boundary tone. The implication is that any other interword juncture will be something that can be transcribed on the break index tier with either a 0 or a 1. However, the subjective impression of boundary strength does not always allow such a neat correspondence. In the course of developing the ToBI transcription system, we encountered several utterances in which we felt a strong sense of disjuncture at a boundary between two words where the pitch pattern showed no evidence of the necessary tonal events for either of these two levels of intonational constituency. We also encountered the converse case: utterances in which the pitch pattern at a boundary between two words clearly indicated an intermediate or intonation phrase boundary with none of the preboundary lengthening or other cues that support the subjective sense of a strong disjuncture. Break index 2 was devised to mark cases of these two types of ‘mismatch’ between the subjective boundary strength and the intonational constituency. These two types are described in the ToBI Annotation Conventions as follows:

a strong disjuncture marked by a pause or virtual pause, but with no tonal marks; i.e. a well-formed tune continues across the juncture.

OR

a disjuncture that is weaker than expected at what is tonally a clear intermediate or full intonation phrase boundary.

Example utterance <<iraqi>> illustrates the first type of mismatch, and example utterance <<quincy>> illustrates the second. In <<iraqi>>, the smooth sequence of apparent downstepped peak accents with no clear intervening phrase accent suggests that the words "six", southern, "iraqi", and "cities" all belong to the same intermediate phrase, yet there is an intonation phrase sized pausing between each adjacent pair of these words. In <<quincy>>, the clear tonal markings for at least an intermediate phrase boundary are unaccompanied by any clear preboundary lengthening, making some transcribers uncomfortable in labelling this juncture with a 3.

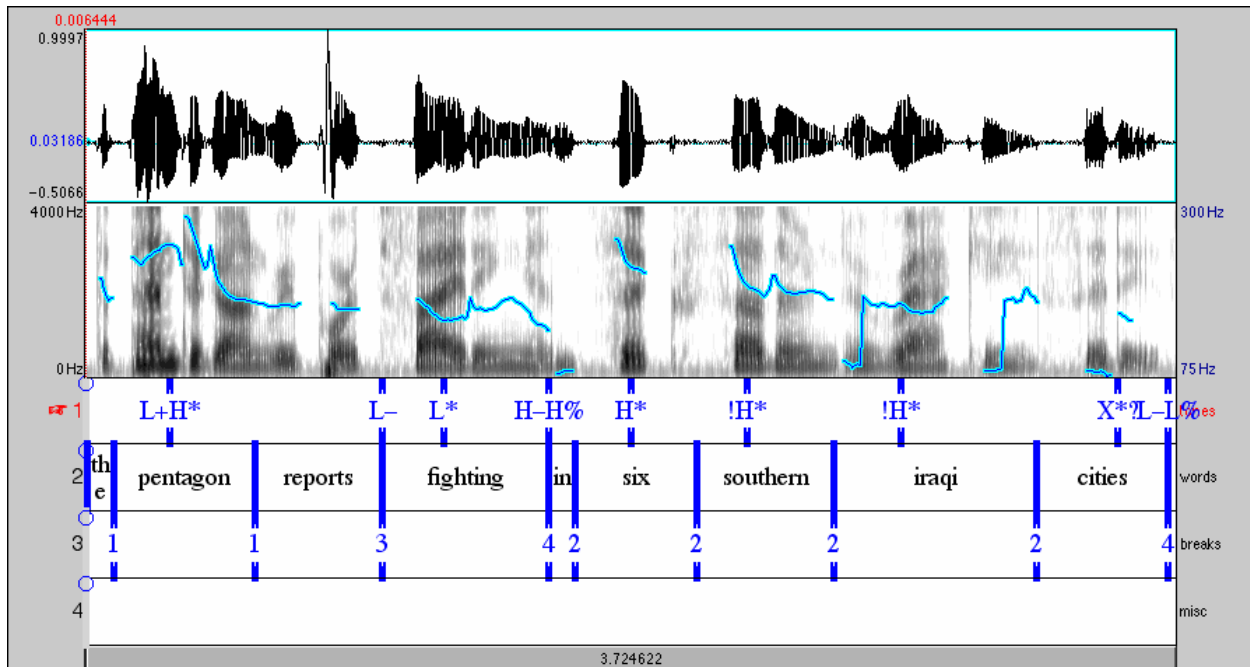


Figure 2.10.1

<iraqi>

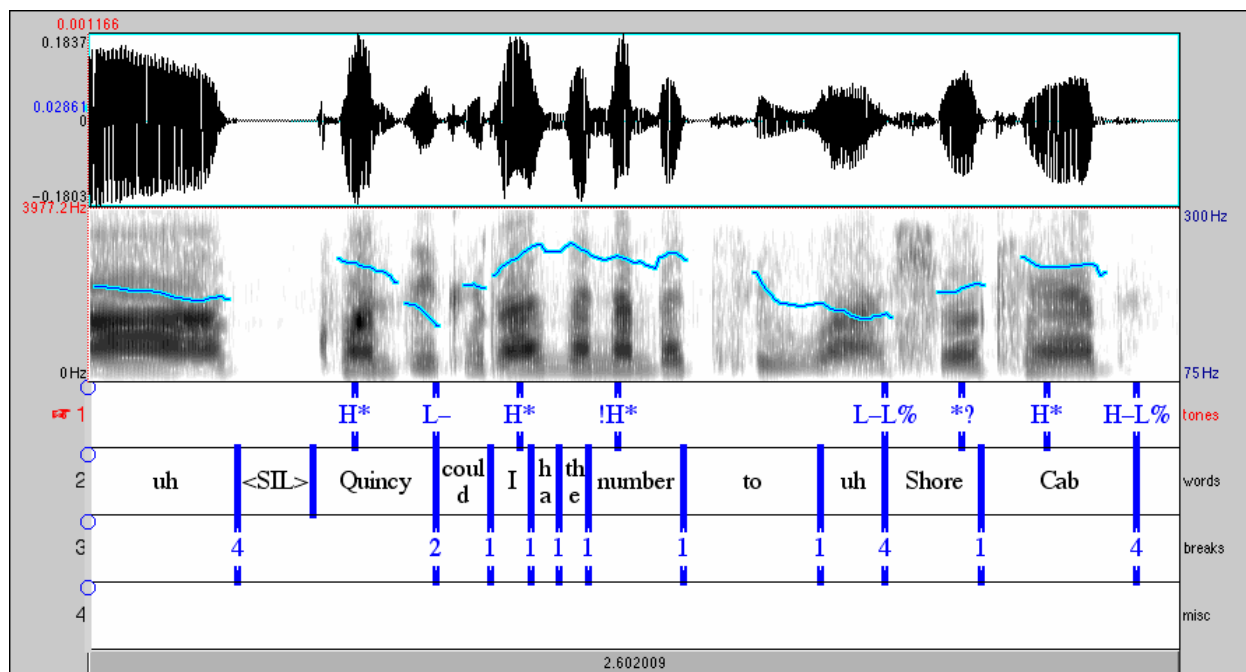


Figure 2.10.2

<quincy>

Break index 2 was devised for cases where the mismatch between the tonal marking and the disjuncture is not accompanied by any sense of hesitancy or disfluency. When 2 is used in the first way (to indicate a stronger sense of disjuncture than 1 even while producing a coherent contour for an uninterrupted intermediate phrase), it can have the rhetorical effect of careful deliberation, as in the <<iraqi>> example. In the opposite case (when 2 is used to mark intermediate phrase boundaries which do not have a very strong sense of disjuncture) the speaker may be speaking quickly to hold the floor or to convey a sense of urgency, while using the tonal marks necessary to convey attentional focus on several closely placed words. We suspect that both types of 2 will be explained ultimately by a better understanding of the complexities of discourse structure, an understanding that can best be achieved by the transcription and analysis of many occurrences in natural dialogue.

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone

!H*: downstepped high pitch accent

L+!H*: bitonal pitch accent with low tone followed by a downstepped high tone prominence

L*+!H: bitonal pitch accent with low tone prominence followed by downstepped high tone

H+!H*: bitonal pitch accent with high tone followed by downstepped high prominence

L-L%: low phrase tone, low boundary tone

H-H%: high phrase tone, high boundary tone

L-H%: low phrase tone, high boundary tone

H-L%: high phrase tone, low boundary tone

H-: high phrase tone

L-: low phrase tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

2: mismatched inter-word disjuncture within a phrase

3: end of an intermediate phrase

4: end of an intonational phrase

2.11 The p diacritic (and the %r tone label)

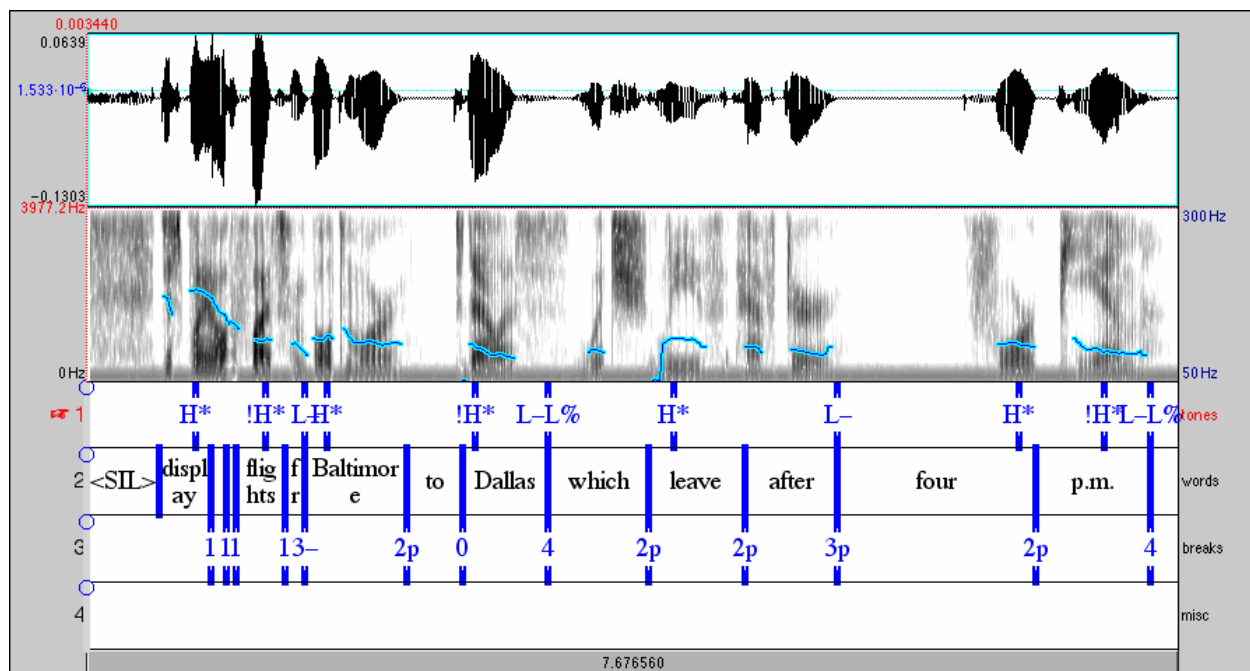
Note for this 2006 ToBI tutorial: Prosodic labelling of disfluencies has turned out to be one of the biggest challenges to the ToBI system, and because these events happen often in spontaneous speech, extending the system to label them appropriately and reliably is a pre-requisite for labeling large spontaneous corpora of spontaneous speech. This is a topic of current research (e.g. Arbisi-Kelm and Jun (2005)¹) It is anticipated that changes may be made to this section to reflect evolving understanding of the prosody of disfluencies. For now, we include the section that describes the p diacritic for disfluent regions, and the %r tone label for a restart from the 1997 ToBI Labelling Guide.

From the 1997 3rd revision of the ToBI Labelling Guide

(http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf):

[Christine Nakatani and Elizabeth Shriberg contributed greatly to the preparation of this and the following sections.]

There are other cases of mismatch between tone tier and segmental rhythm, however, where break index 2 does not seem to be appropriate. For example, in utterance <<display>>, the pauses after "Baltimore", "which", and "leave" do not have the feel of a speaker striving for an effect of judicious deliberation, as in the "six southern Iraqi cities" phrase of the <<iraqi>> example, but rather sound disfluent, as if the speaker were hesitating as he searches for the next word. Such cases can be distinguished from fluent cases of 2 by the use of the p diacritic.



EXAMPLE <display>: Display all the flights from Baltimore to Dallas

¹ Timothy Arbisi-Kelm and Sun-Ah Jun (2005) "A comparison of disfluency patterns in normal and stuttered speech" in *Proceedings of Disfluency in Spontaneous Speech Workshop*, Aix-en Provence, France.

1 1 1 1 3- 2p 0 4
 which leave after 4:00 p.m.
 2p 2p 3p 2p 4

The p diacritic is used in conjunction with a break index 1, 2, or 3, to indicate a disfluency in the timing or separation of words across a break. The notation 'p' was chosen initially to denote the prolongation of the hesitation pause with break indices 2 and 3, but we have since extended the diacritic's usage to cover also abrupt cutoffs before restarts and repairs, which are often but not necessarily separated from the disfluent stop by a pause. In this case, the appropriate break index is 1. Thus the inventory of combinations of break index and p diacritic is:

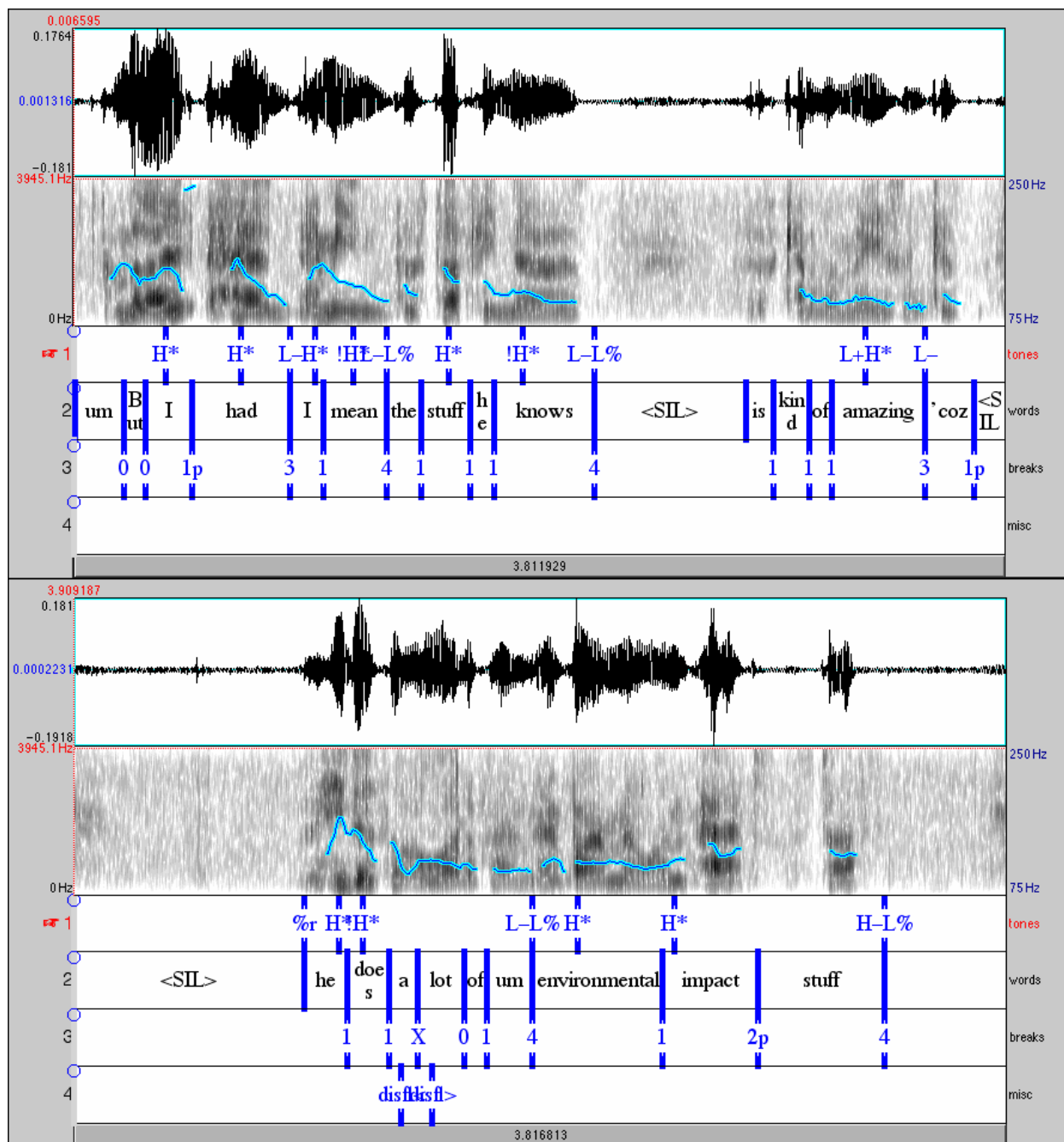
1p -- an abrupt cutoff before an actual repair, or as if stopping to permit a repair or restart of some kind

2p -- a hesitation pause or prolongation of segmental material where there is no phrase accent perceived in the intonation contour

3p -- a hesitation pause or a pause-like prolongation where there is a phrase accent in the tone tier.

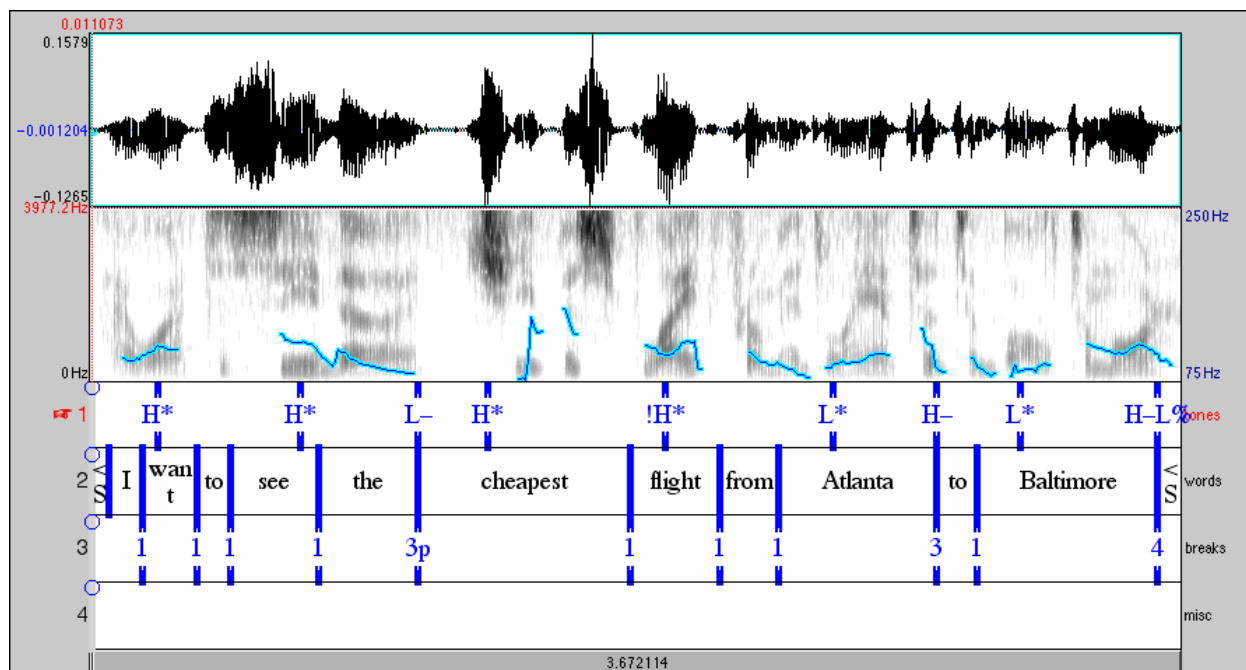
The p diacritic is not used with break index 4, because it is difficult to reliably identify hesitations between two full intonational phrases. Example utterances <<amazing>> and <<cheapest>> illustrate the use of the diacritic with break indices 1 and 3. Example <<display>> also had an occurrence of 3p. Note the presence of the phrase accent distinguishing this interword juncture from the surrounding cases of 2p.

EXAMPLE <amazing>: um But I had I mean the stuff he knows is kind of
 0 0 1p 3 1 4 1 1 1 4 1 1 1
 amazing 'coz he does a lot of um environmental
 3 1p 1 1 X 0 1 4 1
 impact stuff
 2p 4



EXAMPLE <cheapest>: I want to see the cheapest flight from Atlanta

1 1 1 1 3p 1 1 1 3
to Baltimore
1 4

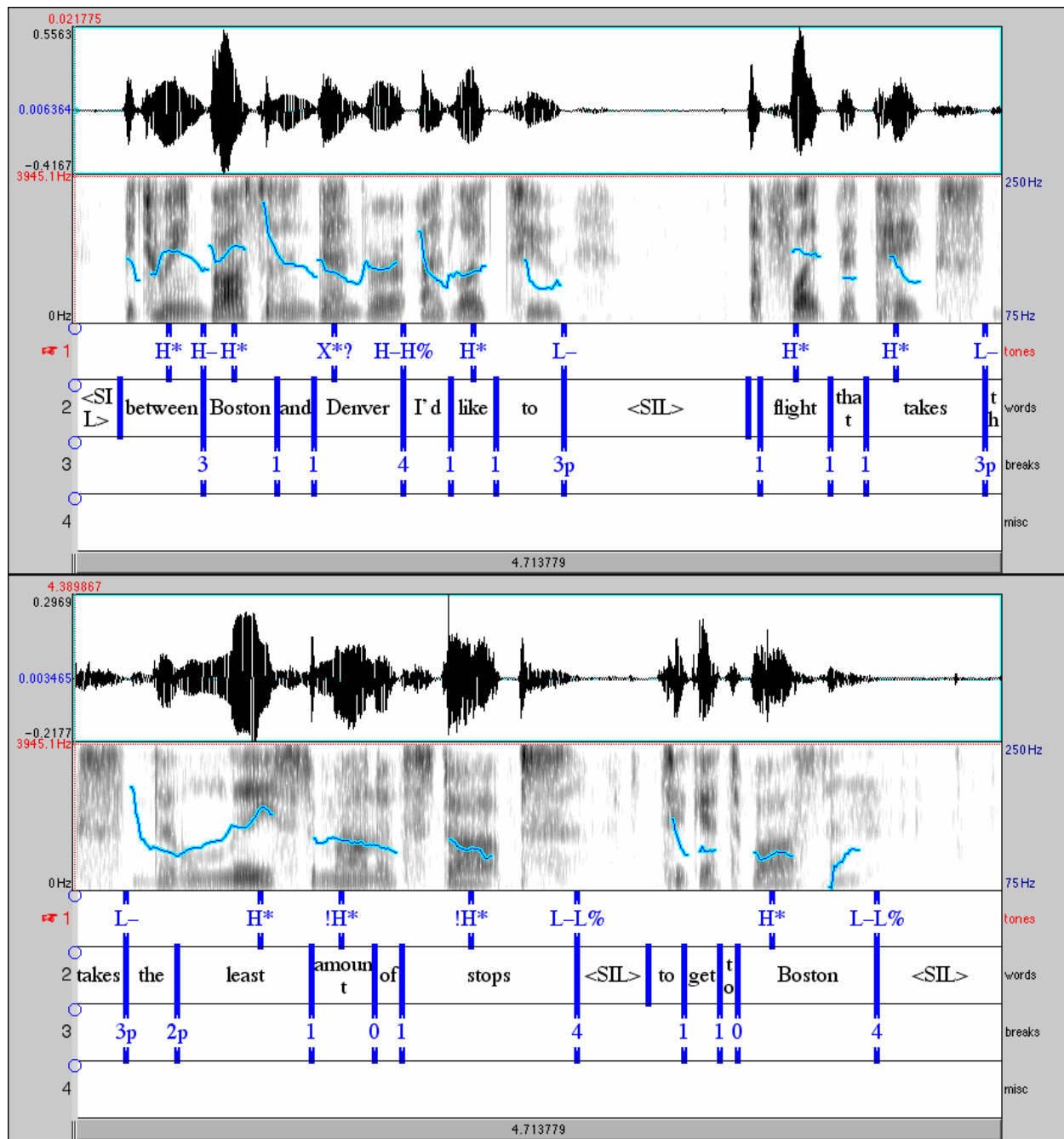


In general the p diacritic should be used conservatively, and should not become a substitute for 2. A good test for appropriateness is to imagine whether the break would have been the same if the speaker were asked to repeat the utterance with the same intonation, but more 'fluently'. If the break were the same upon repetition, it should probably not get the p.

Note also that the prolongation of segmental material for a 2p label can physically occur at the beginning of a word rather than at the end, as in example <<least>>, where the hesitation lengthens the [l] of "least" rather than the vowel of "the".

EXAMPLE <least>: Between Boston and Denver I'd like to a flight that

3 1 1 4 1 1 3p 1 1 1
 takes the least amount of stops to get to Boston
 3p 2p 1 0 1 4 1 1 0 4



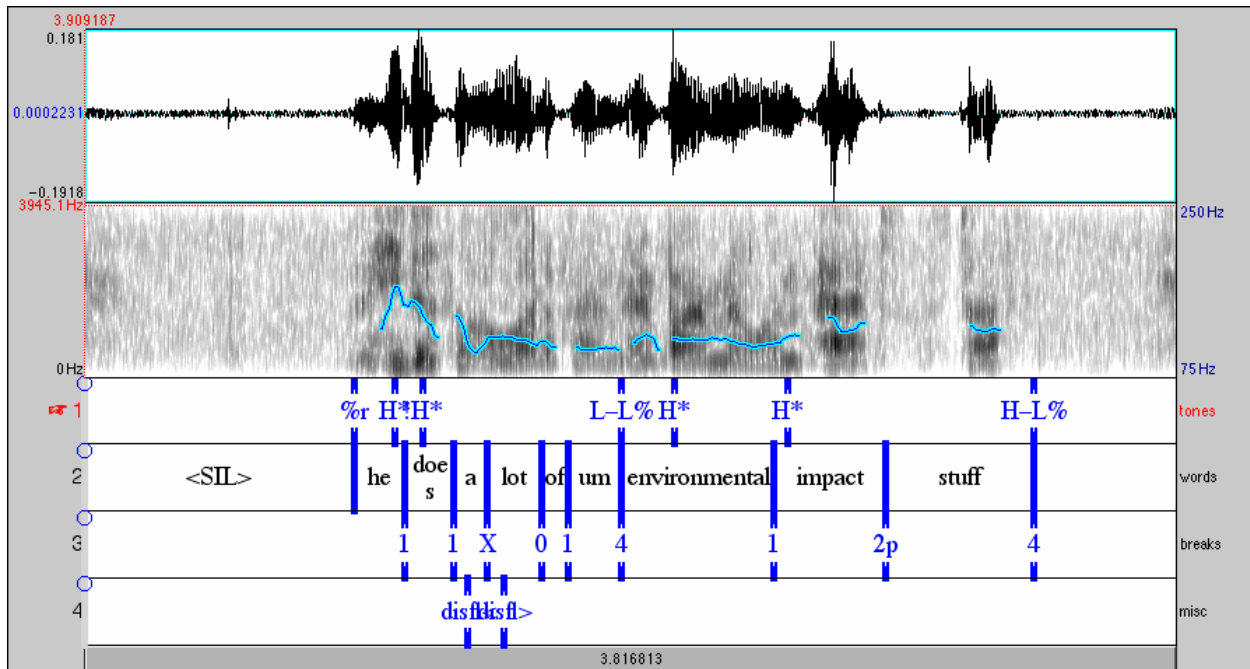
Comment (7/7/05): One point to keep in mind is that the 'cutoff' described for 1p is not equivalent to the loss of phonological material from the end of a word. For example, speakers often apparently delete the final stop from a consonant cluster, as in 'los' an' found' for 'lost and found'. In these cases there is clearly no disfluency and 1p should not be used. Conversely, an abrupt cutoff that needs a 1p can occur when no phonological material is missing, as when a final vowel is abruptly interrupted, e.g. in an utterance like "I-I don't know". In these cases there is a perceived disfluency i.e. a sudden and unnatural interruption (often at the glottis), despite the fact that the final segment of the word was produced, and a 1p would be appropriate. Perhaps the best definition of a cutoff that requires a 1p label is that the word ends in a way that

sounds interrupted.

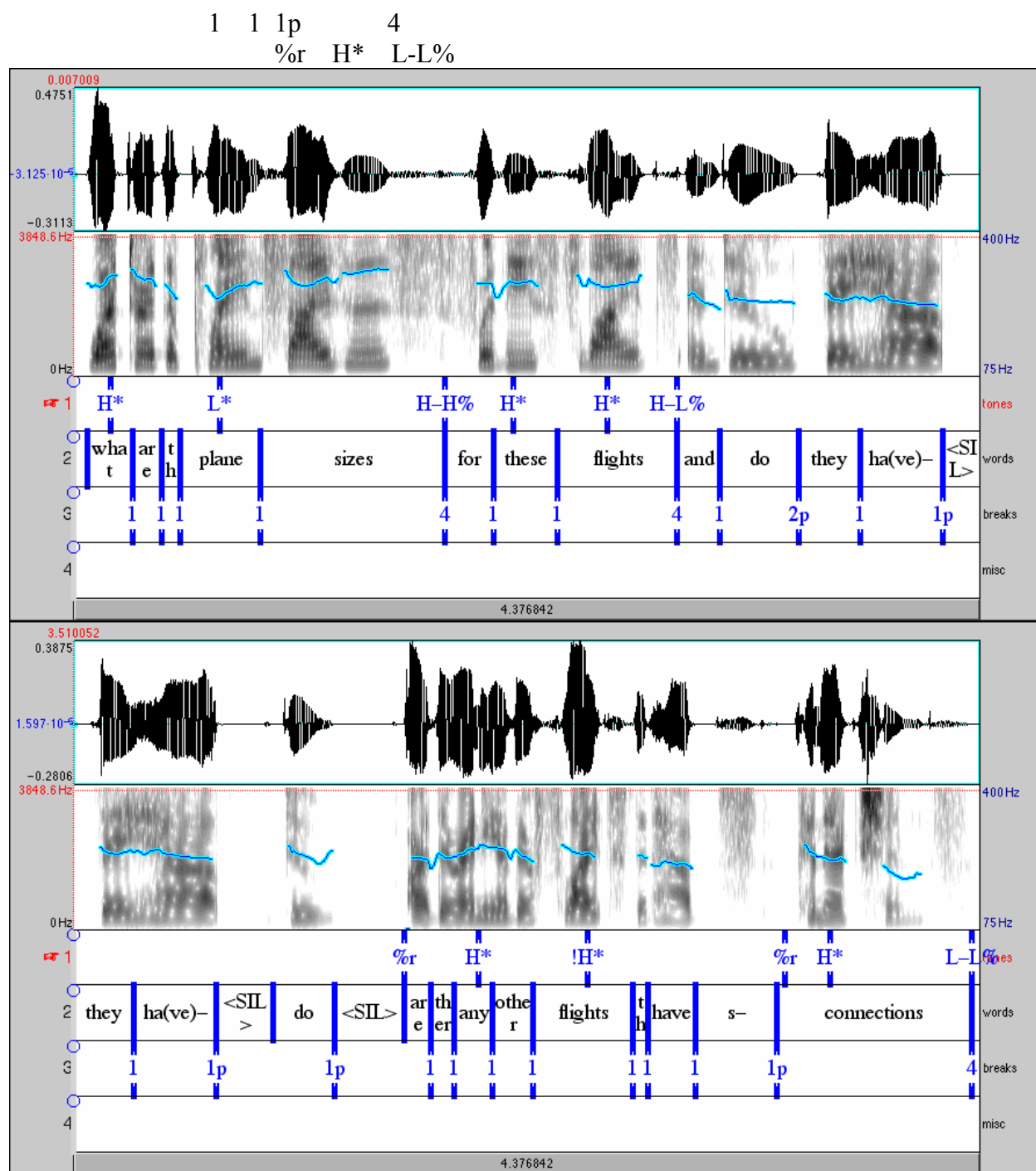
Here is what the labeling guide says about %r:

Closely associated with these definitions of 1p, 2p, and 3p in the break index tier is the tone tier label %r, for restarting with a brand new intonation contour when the last contour was interrupted without being finished by some disfluency. This is most common at a 'repair', where the speaker abruptly stops and begins again with the intended or 'repaired' material, as in example utterance <<amazing>>, already cited above, and in example <<connections>>, below.

EXAMPLE <amazing>: um But I had I mean the stuff he knows
 0 0 1p 3 1 4 1 1 1 4
 H* H* L- H* !H* L-L% H* !H* L-L%
 is kind of amazing 'coz he does a lot of um
 1 1 1 3 1p 1 1 X 0 1 4
 L+H* L- %r H* !H* L-L%
 environmental impact stuff
 1 2p 4
 H* H* H-L%



EXAMPLE <connections>: What are the plane sizes for these flights and
 1 1 1 1 4 1 1 4 1
 H* L* H-H% H* H* H-L%
 do they ha(ve)- do are there any other flights
 2p 1 1p 1p 1 1 1 1 1
 %r H* !H*
 that have s- connections

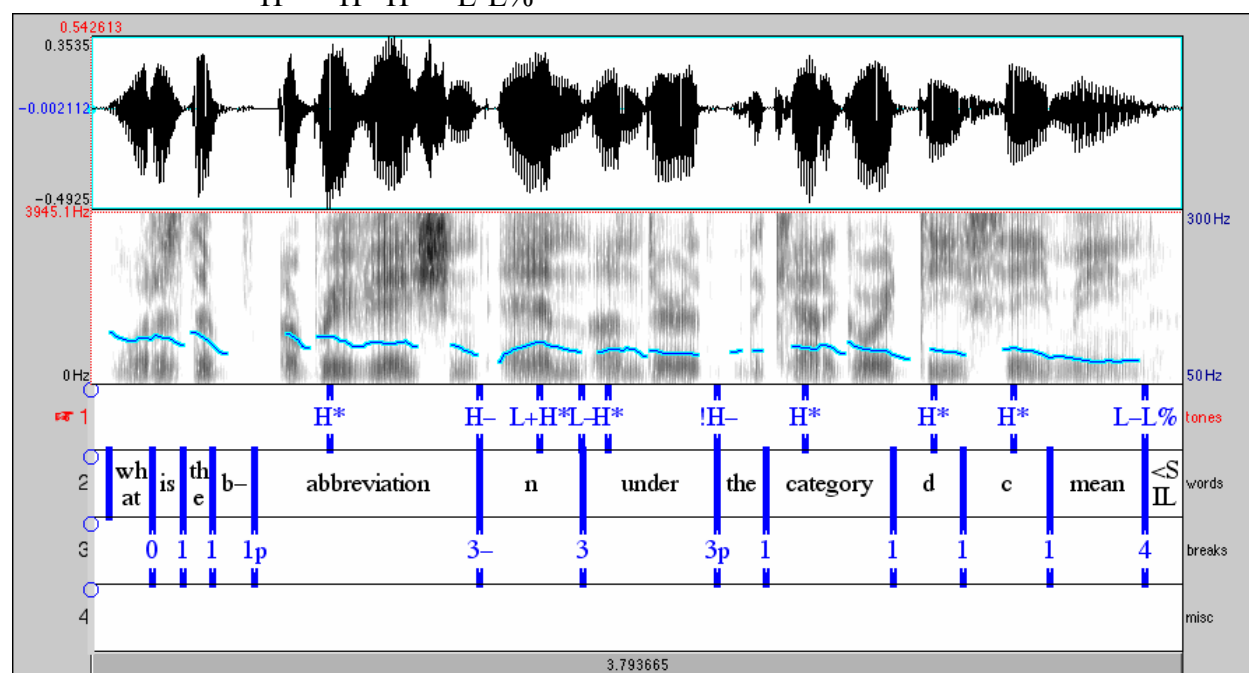


As with the use of the p diacritic, one should be conservative in using %r. It is needed only if there is good evidence that a new intonational phrase has begun after disfluent pause, evidence such as a notable change in f0 range or amplitude. It should not be used in cases such as the "had" after the first 1p in <<amazing>>, which continues with a fluent H* accent in the same pitch range (unlike the H* !H* on "he does" after the second 1p in this utterance, which is in a new pitch range). Nor should %r be used in example utterance <<abbreviation>>, where after

the speaker stumbles and pauses momentarily around the end of "what is the", the intonation on "abbreviation" continues as if there had been no interruption.

EXAMPLE <abbreviation>: What is the b- abbreviation n under

0 1 1 1p 3- 3 3p
H* H- L+H* L- H* !H-
the category d c mean
1 1 1 1 4
H* H* H* L-L%



Especially, %r should not be used after a 3p, where the (re)start of a new intonation contour is already implicit in the break index for the intermediate phrase.

Summary of ToBI labels introduced so far:

Tones:

H*: high pitch accent

L*: low pitch accent

L+H*: bitonal pitch accent with low tone followed by high tone prominence

L*+H: bitonal pitch accent with low tone prominence followed by high tone

!H*: downstepped high pitch accent

L+!H*: bitonal pitch accent with low tone followed by a downstepped high tone prominence

L*+!H: bitonal pitch accent with low tone prominence followed by downstepped high tone

H+!H*: bitonal pitch accent with high tone followed by downstepped high prominence

L-L%: low phrase tone, low boundary tone

H-H%: high phrase tone, high boundary tone

L-H%: low phrase tone, high boundary tone

H-L%: high phrase tone, low boundary tone

H-: high phrase tone

L-: low phrase tone

Break indices:

0: word boundary erased

1: typical inter-word disjuncture within a phrase

2: mismatched inter-word disjuncture within a phrase

3: end of an intermediate phrase

4: end of an intonational phrase

Disfluency marked break indices:

1p: an abrupt cutoff before an actual repair, or as if stopping to permit a repair or restart of some kind

2p: a hesitation pause or prolongation of segmental material where there is no phrase accent perceived in the intonation contour

3p: a hesitation pause or a pause-like prolongation where there is a phrase accent in the tone tier.

Other disfluency markers:

%r: marks the well formed beginning (restart) of an intonational phrase after a disfluent ending of the previous intonational phrase

Optional labels:

<: late High Tonal peak



———— "starred" tone target

———— "unstarred" tone target

----- transition between tone targets

ToBI "tails" (Final pitch accent, phrase accent & boundary tone combinations)

Prepared for the ToBI Tutorial in consultation with Mary Beckman, based on graphics by Julia Hirschberg

MIT OpenCourseWare
<https://ocw.mit.edu/>

6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI
IAP 2006

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.