**ToBI (Tones and Break Indices): A standard for labelling prosody**  
*(English ToBI is now MAE_ToBI)*


a. The **ToBI conventions** are a consensus system for labelling spoken utterances that segregates tags for different types of phonological events and structures into parallel quasi-independent tiers. Most notably, the conventions specify a way to mark the *phonologically* contrastive intonational events (*tones*) separately from the hierarchy of inter-word junctures (*breaks*) with which some of these pitch events are associated.

b. Originally, ToBI was the name of an annotation system, developed in the period 1991 to 1994, which was designed for use in labeling intonation and prosody in databases of spoken Mainstream American English (Beckman & Hirschberg, 1994), thus it was English-specific. The original ToBI for Mainstream American English is now named MAE_ToBI.

c. Very quickly, however, it came to refer to a general framework for the development of prosodic annotation systems in other varieties of English (e.g., Mayo, Aylett, & Ladd, 1997 [Glasgow]) and in other languages (e.g., German, Japanese, Korean, Mongolian, W. Greenlandic, Basque, Greek, Dutch, Serbo-Croatian, Cantonese, Mandarin, Chickasaw, Italian, Spanish, Catalan, European Portuguese, Bengali, Tamil, Arabic, Bininj Gun-Wok, Dalabon (the last two are aboriginal languages of Australia), Jamaican English Creole, Papia)ntu). This makes it possible to compare prosodic systems across languages using a common vocabulary, and to search for universals.

d. The immediate antecedents of the MAE_ToBI model are Pierrehumbert (1980), Beckman & Pierrehumbert (1986) for H vs. L tones; Ladd (1983) for the treatment of downstep; and Price et al. (1991) and Wightman et al. (1992) for the treatment of juncture.

e. ToBI has at least six parts; two are continuous phonetic records (audio recording and record of the f0 trace) and four are symbolic strings (words, tones, break indices, miscellaneous). Cf. recent addition: ‘Alt’ tier (Brugos et al. 2008)

f. The tones in a ToBI framework are comparable to a broad *phonemic* transcription of consonants and vowels, and not to a narrow phonetic one (e.g., INTSINT (International Transcription System for Intonation) by Hirst & Di Cristo is a transcription model of intonation comparable to a narrow/phonetic transcription.)
Note: ToBI is not an International Phonetic Alphabet for prosody. Because intonation and prosodic organization differ from language to language, and often from dialect to dialect within a language, a ToBI system of one language is different from that of others. Each ToBI system is specific to a language variety and the community of researchers working on that language variety.

Cf. Workshop on Developing an International Prosodic Alphabet (IPrA), a satellite meeting of International Congress of Phonetic Sciences, Aug. 2015, Glasgow, Scotland.

Motivation: it may be useful to have tonal symbols and other diacritics to represent surface f0 patterns that (a) appear to be categorical in nature but not necessarily distinctive or (b) whose distinctiveness is not yet known (at the beginning stages of documenting/analyzing the language’s prosody/intonation). Furthermore, studying prosodic typology based on phonological labels cannot capture similarities in realizations. It would be helpful to have tonal labels/diacritics reflecting surface f0 shapes in addition to the underlying/distinctive f0 categories.

g. ToBI is an ongoing research program rather than a set of ‘rules’ cast in stone for all time. Its value: Useful as a communal corpus creation tool even in its current state. Through labeled corpus and inter-transcriber agreement studies, it can test and build theoretical claims (e.g., stress shift, accent and semantic weight of a word, the role of accent in pronoun resolution, discourse structure and its markers) as well as to train intonation recognizer and synthesis system, and understand better the strength and weakness of the underlying phonological theories.

What does a complete ToBI system require?

Because a ToBI-framework system is a community-wide standard, it requires a community of users who have agreed to adopt the conventions in database development and related research. It must also conform to the following five principles:

1. The conventions are as accurate as possible, given the current state of knowledge. Ideally, they will be based on a large and long-established body of research in intonational phonology, dialectology, pragmatics and discourse analysis for the language variety, but at the very least, they are based on a rigorous analysis of the intonational phonology.

2. The conventions do not replace a permanent record of the speech signal with a symbolic record. An electronic recording of the transcribed utterance is an essential component of a complete ToBI framework transcription.

3. The conventions are efficient. They do not waste transcriber time by requiring the transcriber to symbolically mark non-distinctive pitch rises and falls that can be extracted from the signal automatically, or anything else that could be extracted from resources such as online pronunciation dictionaries. Do not transcribe predictable categorical aspects of prosody -- ex. Stress, Nuclear pitch accent
4. The conventions are easy enough to teach (and learn) that their use is not limited to a few experts to do the transcription. Therefore, there must be a freely available manual for teaching the system to new transcribers, with many recorded examples of transcribed utterances graded from easy to difficult.

5. The conventions are used and maintained consistently across transcription sites. Therefore, in the course of developing a ToBI framework system, there must be rigorous tests of intertranscriber consistency, and there should be an agreed-upon center for maintaining the standard with periodic rechecks and evaluation of any proposed revisions.

Guiding principles of ToBI (MAE_ToBI) by Beckman & Hirschberg (1997)
Webpage for ToBI guidelines: http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/

A full ToBI annotation consists of audio recordings, with some software used to create four parallel tiers, time aligned with the sound file. (Praat is the most widely used these days, but this will surely change eventually. Many of the articles by Beckman and colleagues that you’ll see you the program xwaves software; the homework printouts and examples in the handout mostly use the PCQuirer software).

Words tier –

- An Orthographic tier
- Word it written such that the last segment of the word comes on the right edge of speech signal. [in Praat, the orthographic tier is an ‘interval’ tier, where each word is located in the middle of the interval.]
- As noted previously, I recommend marking syllable boundaries in the word tier as well (of course, you could also do this separately as a “syllable tier”). If the fine details of tonal alignment are crucial to your research question, you will likely want to label each individual segment, which is easy to do with a forced aligner (e.g., the PENN Phonetics Lab Forced Aligner, or the ProsodyLab-Aligner (McGill)).

Tone tier –

- A tier for making tones aligned with both lexically stressed syllables (i.e., pitch accents) and tones aligned with a two-level phrase structure (i.e., phrase accents and boundary tones).
- In Praat, it makes more sense to use a ‘point tier’ rather than an interval one, for the simple reason that tones are modeled as point-like targets.
- Place your pitch accent label at the place in the syllable where it is phonetically occurring (which may not always be the exact center); place your phrase accents and boundary tones at the end of the word interval that bears this tone.
- Said otherwise:
  - pitch accents (peaks/valleys) are placed somewhere within the accented syllable, preferably within the interval that can be identified with the syllable’s vowel.
  - ‘<’ is used to mark a tonal realization as delayed (‘>’ is used for early tonal target).
- **phrasal tones**: L-, H-, !H- intermediate Phrase (ip) phrase accent and L% & H%
  
  Intonational Phrase (IP) boundary tone H% is labeled at the end of an IP, and %H at the beginning of an IP.

- This results in 4 possible combinations, i.e., ways that a IP can end:
  - L-L%, L-H%, H-L%, or H-H% (in addition to !H-H%)
  - In practice, for labeling:
    - …an IP-medial phrase accent, write the phrase accent at the end of the ip.
    - …an IP-final phrase accent, write phrase accent and boundary tone together at the end of the IP
  
  cf. The reason of writing a phrase accent at the end, not the beginning, of an ip: the beginning of an ip does not correspond to the degree of a phrase juncture, i.e. break index ‘3’.

  cf. Pitch range: for marking local pitch range for each intermediate phrase, write ‘HiF0’ at the highest pitch accent within ‘ip’ (not H% or H- or high f0 triggered by pitch perturbation)

**Break-index tier**

- Transcribe the strength of association (coherence or disjuncture) between adjacent words.
- In *Praat*, the BI tier is also a ‘point’ tier where each BI label is located at the time point you specify.
- Write a number 0 to 4 at the right edge of each word. This numbering captures the hierarchical nature of the prosodic groupings:
  - ‘0’ for clear phonetic marks of clitic groups (multiple words behaving segmentally like single prosodic words).
  - ‘1’ for most phrase-medial word boundary
  - ‘2’ for a mismatch between tone and juncture. ex. strong disjuncture but no tonal marks, or a weak disjuncture but clear tonal marks
  - ‘3’ for a juncture corresponding to the intermediate phrase boundary
  - ‘4’ for a juncture corresponding to the Intonation Phrase boundary
  
  ➢ ‘2’ may be modified to ‘1m’ (break index of ‘1’ but phrase accent) and ‘3m’ (break index of ‘3’ but no phrase accent)

**Miscellaneous tier**

- for any comments or markings about the utterance (e.g. silence, audible breaths, laughter, false start, hesitation, disfluencies, and other spontaneous speech effects.)
- when events have a duration, mark the starting point with ‘<‘ and the ending point with ‘>‘.
[in Praat, Misc tier is an ‘interval’ tier where each comment is located in the middle of the interval.]

Ex. laughter< ... laughter>

**Notations for uncertainty or underspecification**

*  underspecified pitch accent (accented but the tonal value has yet to be assigned)
-  underspecified phrase accent.
%  underspecified boundary tone

**? Not sure if the syllable is pitch accented or not; uncertainty of accentedness.
Same with phrase accent (-?) and boundary tone (%?).

X**? Now sure what the pitch accent type of this accented syllable is; uncertainty of accent type. Same with phrase accent (X-?) and boundary tone (X%?).

x  underspecified break index
#-  break uncertain between # and #-1 level (ex. 2- means not sure of 2 or 1)
#p  disfluency after this level of juncture; 1p for abrupt cutoffs of a word; 2p, 3p for prolongation, but if there is a phrase accent, label 3p and if not, label 2p.

**Examples for bitonal pitch accents and downstepped pitch accents:**

**Below, speech file names are in italics, and the words showing the example pitch accent are underlined**

**L*+H:** millionr (first): Only a millionaire.

stein: Stein’s not a bad man. Ligamarole is monomorphemic.

**L+H*:** millionr (second): Only a millionaire.

yellow2: There’s a lovely yellowish old one.

blooming(dale), (second): There a lovely one in Bloomingdales.

yellow3: It’s lovely and yellowish and it’s an old one.

levels: There are many intermediate levels.

capote: Capote died Saturday at the Bell-Air home of Joanne Carson, estranged wife of talkshow host Johnnie Carson and she was among those who eulogized him.

hennessy: Hennesy is widely respected for his legal scholarship and ..

onions: Now, be careful!. Okay, chop the onions and put them into that bowl.

**L+!H*:** yellow2: There’s a lovely yellowish old one.
yellow3: It’s lovely and yellowish and it’s an old one.
levels: There are many intermediate levels.
blooming(dale), (second): There a lovely one in Bloomingdales.

L*+!H: flap: Don’t hit it to Joey (Don’t may show pitch halving)
blooming(dale), (first): There’s a lovely one in Bloomingdales.

!H* (=after H*+L in 1986): oregano (don’t know, basil),
nose : Oh, don’t nuzzle me you marmalade nose.
anna2 (2nd): Anna married Lenny.
capote: Capote died Saturday at the Bell-Air home of Joanne Carson, estranged wife of talkshow host Johnnie Carson and she was among those who eulogized him.
hennessy: Hennessy is widely respected for his legal scholarship and ..
philadel(phia): from Philadelphia to Dallas.

onions: Okay, now chop the onions. Now, be careful!. Okay, chop the onions and put them into that bowl.
capote: Capote died Saturday at the Bell-Air home of Joanne Carson, estranged wife of talkshow host Johnnie Carson and she was among those who eulogized him.

More examples

!H*: blond-baby1: what’s the difference among my long memory, your blond baby, and the pink carpeting?
democrat: the chairman Wendell Ford democrat of Kentucky,
Iraqi: The pentagon reports fighting in six southern Iraqi cities.
least: between Boston and Denver (X*?) I’d like to a flight that takes the least amount of stops to get to Boston.
amazing: but I had I mean the stuff he knows is kind of amazing ‘coz he does ....

L+H*: pigs: They’ve eaten the pigs. They’ve eaten the pigs.
wellies2: No I think I’ll wear my hiking boots.
memphis2: my mother lives in Memphis.
noodle3: we have a lean mini noodle dish.
blond-baby1: what’s the difference among *my* long memory, *your* blond baby, and the pink carpeting?
park2: definitely the shortest and probably the pleasantest way to go is through the park.
democrat: the **chairman** Wendell Ford democrat of Kentucky,
amazing: but I had I mean the stuff he knows is kind of **amazing** ‘coz he does ...

L+!H*: **wellies2**: No I think I’ll wear my **hiking** boots. Bloomingdales.)