Introduction; Representation of Intonation

I. Introduction and Background

1. What is intonation? Narrow meaning: pitch modulation over an utterance.

2. Unlike the role of pitch in a tone language, intonational pitch differences do not distinguish words, but rather distinguish semantic/pragmatic meaning, and (along with other aspects of prosodic structure), and determine how information in an utterance relates to a larger discourse.

3. Difficulties associated with the study and representation of intonation:

   - Pitch is relative (no inherent feature) and simultaneously transmits paralinguistic vocal code (emphasis, emotional state) and physiological information about the speaker (i.e., age, sex) along with the linguistic code.

   - In addition to the kinds of information it encodes, pitch has both a categorical (i.e., linguistic/phonological) as well as a gradient/scalar nature.

     o Note: These kinds of information and kinds of uses of pitch line up; paralinguistic “meanings” tend to be “continuous, like their phonetic realization (think about how “excitement” or “fear” might be expressed by prosody). Linguistic meanings, however, tend to be categorical, and these categorical meanings are generally expressed by categorical contrasts in pitch (think about how (some) questions are asked in English—pitch can categorically rise or pitch can categorically fall).

   - Unlike in music, speech prosody generally lacks steady-state pitch over a syllable… it’s always changing/moving.

4. How can intonation be described?

4.1. The two major traditions in the analysis of intonation:

   Instrumental vs. Impressionistic

   a. Instrumental (or phonetic) tradition: by experimental psychologists and phoneticians interested in speech perception and in identifying the acoustic cues to intonational phenomena. For example, identifying i) acoustic cues to interrogation, finality, and continuation; Delattre 1963, Lieberman 1967), ii) emotional states such as anger,
surprise, and boredom (Lieberman & Michaels 1962, Williams & Stevens 1972), and iii) word and sentence stress (e.g., Fry 1958, Lieberman 1960); good review in Lehiste (1970).

- In general, conclusive findings were elusive to these researchers; they were trying to identify the acoustic cues related to perceived prosody, but failed to appreciate/look for the linguistic organization of these cues.

b. Impressionistic (or pre-phonological) approach: by linguists and language teachers. Represented by the American structuralist school (e.g., Pike 1945, Wells 1945, Trager & Smith 1951) and by the British school (e.g., Palmer 1922, Amstrong & Ward 1926, Kingdon 1958, O’Connor & Arnold 1961, 1973, Crystal 1969; starting from Sweet 1892, Jones 1909).

- They treated intonation in terms of a small number of categorically distinct elements such as pitch phonemes and nuclear tones, thus investigating phonology, but their impressionistically driven categories were often hard to verify rigorously.

American tradition:

- The extension of Bloomfieldian principles of phonemic analysis to suprasegmentals: 4 “stress phonemes” (from primary to weak), 4 “pitch phonemes” (from levels ‘1’, low, to ‘4’, high), 4 phonemes of juncture (one internal and 3 final contour, fall, rise, and level).

  o Similar to autosegmental works in 1970’s (Leben 1976, Goldsmith 1976, Liberman 1978), in that the contour was modeled as being, underlying, a sequence of pitch levels.

British tradition:

- Tones: two static (H, L), three kinetic (rising, falling, falling-rising), and two complex (rising-falling, rising-falling-rising). Stress: 4 levels (emphatic stress, full stressed, half stressed, unstressed)

- Intonation, a ‘tune’, consists of two parts: ‘head’ (the part preceding the sentence stress) and the reminder (‘nucleus’, the syllable with sentence stress, and optionally a ‘tail’, any syllables after the sentence stress); this was further refined by Crystal (1969): (prehead)(head)nuclear(tail).

Robert Ladd’s visualization of the British School’s conception of prosodic structure. Note that it prosodic events can be categorized in much the same way modern models do for English: a prenuclear region, and nuclear region/accen, and a post-nuclear stretch.
Narrow phonetic transcriptions of pitch in the British tradition: representations of the relative pitch of each successive syllable, known as a ‘tadpole diagram’.

I hear Sue’s taking a course to become a driving instructor.

Sue!? A driving instructor!?

Figure 2.1 Detailed syllable-by-syllable impressionistic notation of intonation (‘tadpole diagram’) of the sort commonly used by traditional descriptive works of the British school of intonation analysis.

Views on sentence stress

- In the American tradition, stress and pitch are independent phenomena; sentence stress is one level of stress (Chomsky & Halle’s “–stress” or “nuclear stress”), a stress phenomenon which is often associated, but not strictly/systematically with a pitch change.

- In the British tradition, sentence stress, called the “nucleus”, was an intonational (i.e., pitch) phenomenon that has nothing to do with stress at all.

  - Bolinger (1958): sentence stress is neither stress nor intonation, but accent.

    “accent” is syllable prominence signaled by pitch obstruction or pitch change. He treats “stress” as lexical abstraction—the potential location for accent.

    He proposed three types of accent (see Figure 2 in Ladd (1980), p. 18)

    A-accents (falling movement from accented syll),
    B-accents (rising movement to and from the accented syll) and
    C-accents (falling movement into the accented syll).
4.3 A break from the traditional approach

- Two levels of pitch target (High and Low) with a linear representation
- Influenced by Bruce (1977)’s analysis of Swedish intonation.

I. Two-level

a. A large amount of variation can be predicted by assuming pitch levels can (a) be reduced to two distinctive phonological “tones”, L and H, (b) that these tones can be combined, and (c) that the phonetic realizations of these phonological tones can vary quite considerably.

b. The f0 target level is quite consistent. Ex. Bruce’s Swedish word accents are distinguished by the timing of a certain pitch level (=peak) relative to the accented syllable, not the movement itself.

5. Modern American Tradition => autosegmental-metrical (AM) model of Intonational Phonology

cf. definition of intonation (Ladd 1996/2008)

: use of suprasegmental phonetic features to convey ‘post-lexical’ or sentence-level pragmatic meanings in a linguistically structured way.

- suprasegmental: f0 (pitch), intensity (loudness; stress), duration (quantity)
- sentence level or postlexical (larger than a word level)
- linguistically-structured: intonational features are organized in terms of categorically distinct entities (ex. H, L; pitch accent or boundary tones) and relations (ex. stronger/weaker than) but excluding ‘paralinguistic’ features.

Intonation is described by a tune, the prominence relation among words, and phrasing.

A. A tune (pitch contour) is a phrase-level melody conveying a certain pragmatic meaning and sentence type information. It is independent from a text (segmental string) and can be decomposed of multiple pitch/tonal events.

a. Example of the same tune associated with different texts:

Ex.1. Greetings
   Good morning!
   How are you?

Ex.2. Everyone knows that!
   Everyone knows that.
   Duh.
Ex. 3. Sarcasm
   Yeah, right.
   Sure.

Ex. 4. Surprise contour (‘L H L H’ or rise-fall-rise)
   A: I hear Sue got a fellowship to study physics.
   B: Sue?

   A: I hear Sue’s taking a course to become a computer programmer.
   B: A computer programmer?

A tune can be produced over different texts, and has meaning separable from the words it is associated with. The association of its subcomponent pitch event with a text varies for different texts. What’s constant is that the important part of the tune is linked to (or ‘associated with’) the stressed syllable of a word and the boundary of the phrase.

b. Examples of the same text having different tunes (melodies):
   Ex. 1. Anna
      falling (as an answer) or rising (as a question) or high-mid plateau (as calling) – see Figure 2

   Ex. 2. Another orange: - see Figure 3
      ‘another’ can be high or low while ‘orange’ can be rising.

Different tunes can be produced over the same text and this changes the implied meaning of the text.

B. Prominence relation can be different for the same intonation tune.

Prominence can be cued in a variety of ways -- pitch, duration, intensity, vowel quality.

   cf. Japanese accents are signaled by pitch change but English accents are signaled by all cues mentioned above. See Beckman (1986) for perceptual differences between the two languages.

Ex. 1. focus

   (I had enough oranges and he gave me) ANOTHER orange! vs.
   (I thought he said ‘another apple’, but actually he said) Another ORANGE!

   Both phrases have the same tune (rise-fall), but the pitch peak is on a different word. One word is more prominent than the other.

   LEGUMES are a good source of vitamins. – see Fig.4a
Legumes are a GOOD source of vitamins. – see Fig.4b
Legumes are a good source of VITAMINS.

Phrasing:

In addition to the shape of a tune and the location of pitch accent, the phrasing of an utterance changes the interpretation of the phrase/sentence. Pitch accents mark the prominence relation among words, and the boundary tone marks the phrasing of an utterance.

Ex.1. [[Old men] [and women]] vs. [[Old] [men and women]]

Ex.2. [[Sublime mnemonic rhyme] [and free metre]
  [Sublime] [mnemonic rhyme and free metre]]

Ex.3. [[Does Manitowoc have a library]]? (Fig 5a-Manitowoc-1ip-v1-9)
[[Does Manitowoc] [have a library]]? (Fig 5b-Manitowoc-2ip-v1-8)

Ex.4. [[Ramona saw the villain] [with the binoculars]] (Fig 6a-Ramona-binoculars)
  (meaning ‘Ramona had the binoculars’)
[[Ramona saw] [the villain with the binoculars]] (Fig 6b-villain-binoculars)
  (meaning ‘the villain had the binoculars’)

Summary so far:

- A tune and a text are independent, and the association between these two is determined by the semantic and pragmatic meaning of the utterance, i.e., production of the text.
  
  The idea of association between tone and syllable is developed from Autosegmental Phonology

- A tune is composed of a sequence of tones (H and L, and their combinations). Some tones are realized on the stressed syllable, i.e., head, of a word and others are realized on the edge (i.e., the beginning or the end) of a phrase. The tones on the stressed syllable are more prominent than the edge tones. The tone on the stressed syllable is called pitch accent and the tone on the edge of a phrase is called a boundary tone.
Fig. 5a

<table>
<thead>
<tr>
<th>tones</th>
<th>H-H%</th>
</tr>
</thead>
<tbody>
<tr>
<td>words</td>
<td></td>
</tr>
<tr>
<td>Does</td>
<td>Manitowoc</td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
<tr>
<td>350</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
</tr>
</tbody>
</table>

ms: 350 700 1050 1400 1750

Fig. 5b

<table>
<thead>
<tr>
<th>tones</th>
<th>H-L%</th>
<th>L*</th>
<th>H-H%</th>
</tr>
</thead>
<tbody>
<tr>
<td>words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does</td>
<td>Manitowoc</td>
<td>have</td>
<td>a</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>350</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ms: 400 800 1200 1600 2000

Fig. 6a
Fig. 6b

```
<table>
<thead>
<tr>
<th>words</th>
<th>Ramona</th>
<th>saw the</th>
<th>villain</th>
<th>with the</th>
<th>binoculars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ms)</td>
<td>500</td>
<td>1000</td>
<td>1500</td>
<td>2000</td>
<td>25</td>
</tr>
<tr>
<td>Hz</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td>350</td>
<td>400</td>
</tr>
</tbody>
</table>
```