E-ToBI
Intonational Annotation

Janet Fletcher

University of Melbourne

ALTA Summer School, Macquarie University
December 2004

Acknowledgements

- Debbie Loakes, University of Melbourne
- Jonathan Harrington, University of Kiel, Germany

Origins of ToBI

- ToBI (Tones and break indices system for prosodic transcription)
- Devised by multi-disciplinary team in the U.S.A. in 1992-present
- Based on intonational model of American English developed by Janet Pierrehumbert (1980)
- Autosegmental-metrical model of intonation (Ladd, 1996)

Motivation for E-ToBI

- Originally devised to provide a common standard for prosodic annotation of speech databases in U.S.A., United Kingdom (standard Southern British English), and Australia
- **Not** an IPA transcription system – each variety requires a separate ToBI system, although some varieties can be transcribed using the same annotation criteria (e.g. Australian English, General American English)

Goals of half-course

- Flavour of E-ToBI annotation – particularly tone annotation
- Reading F0 tracks and performing skilled auditory analysis
- Brief summary of underlying prosodic model
- Description of Tone tier
- Brief overview of Breaks tier

Materials used in this Half-course

- ToBI training materials (Beckman and colleagues) http://www.ling.ohio-state.edu/~tobi
- ANDOSL (Australian National of Spoken Language) map task corpus http://andosl.anu.edu.au/andosl/
- IViE (Intonational variation in English) corpus (Grabe and colleagues, http://www.phon.ox.ac.uk/~esther/ivyweb/
- EMU Speech Database System http://emu.sourceforge.net/

Why intonation transcription is hard!

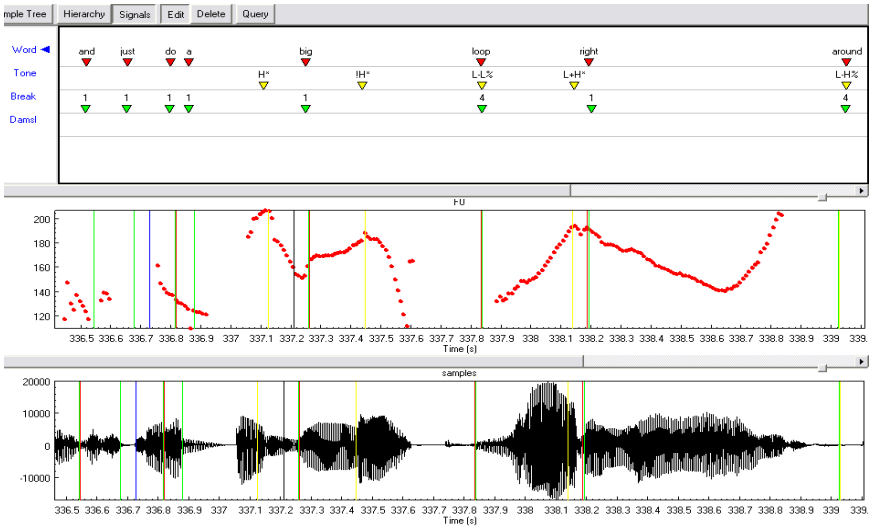
- Same utterance – several melodies
- Same basic melody – variation in phonetic realization due to different speakers' pitch range (e.g. male versus female), fast speech, or compressed pitch range (making F0 interpretation difficult)
- Melodies play a role in determining the expressive force of an utterance – multiple meanings
- Dialectal variation

What does a ToBI annotation include?

- Good ears!
- Acoustic waveform
- F0 contour
- Label tiers
  - tone tier
  - orthographic tier
  - a break index tier
  - a miscellaneous tier

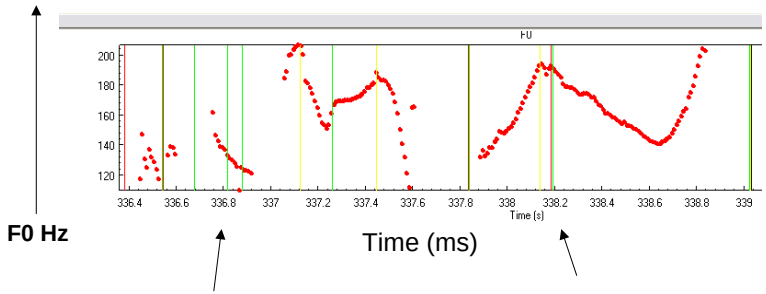
An example of a ToBI transcription

Annotation tiers



The F0 signal

“and just do a big loop right around”

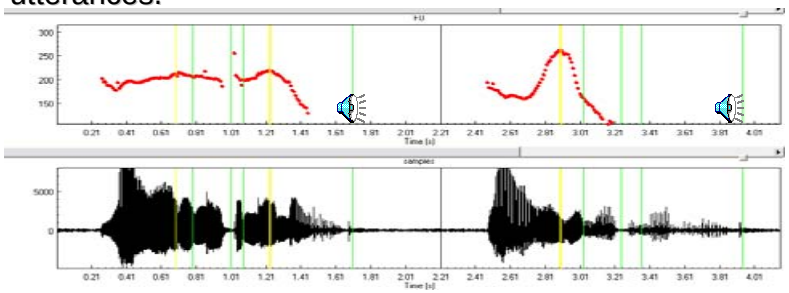


Gaps are generally due to voiceless segments, (in the absence of a silent pause) reflecting the lack of vocal fold vibration

F0 (pitch) shows smoother peaks when segment is a sonorant, falls sharply after a voiceless consonant

Interpreting an F0 contour

- Listeners factor out segmental perturbations from “meaningful” intonational targets that are locally aligned to different structural elements of the text
- Other perturbations, or “gaps” in the F0 signal can be due to “creaky” voice quality – compare these two utterances.



Two major aspects of the ToBI system

- Tonal tier – the intonation or melody of an utterance is decomposed into component underlying H (high) and L (low) pitch targets – signaling two broad functions: phrasing and accentuation
- Break Index tier – the utterance is analysed in terms of its prosodic constituency from the level of the word to the intonational phrase

4 key components to intonational model underlying ToBI

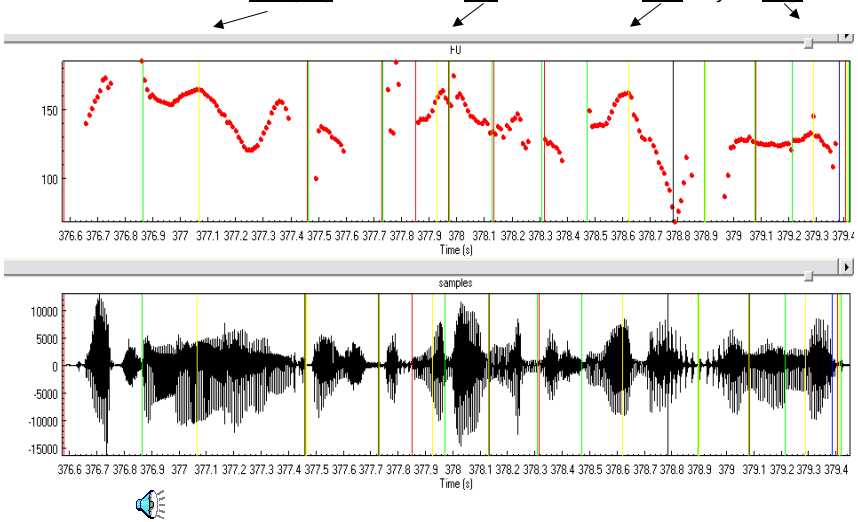
- Stress and accentuation
- Phrasing
- Tune or melody
- Pitch range

Accentuation

- stress - rhythmic properties of the word – the primary stress is the potential location of an **intonational prominence** – ***pitch accent***
- accentuation – property of a word in context – a word is given a ***pitch accent*** to highlight it in the discourse – we hear the word as accented, mainly due to the pitch accent associated with this word – it may be realized as a relatively "high" pitch target, or a relatively "low" pitch target in the speaker's pitch range.
- In addition, the primary stressed syllable of the accented word will be longer, and sound louder than surrounding syllables. The vowel of the accented syllable may also have a more peripheral acoustic target. The entire rhythmic foot may also be longer.

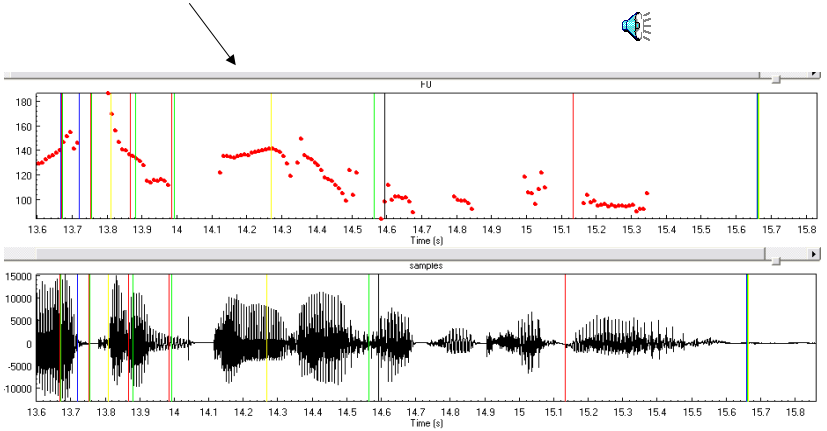
Spot the pitch accents

"what's annoying is that you don't have this river on your map"



What words get accented?

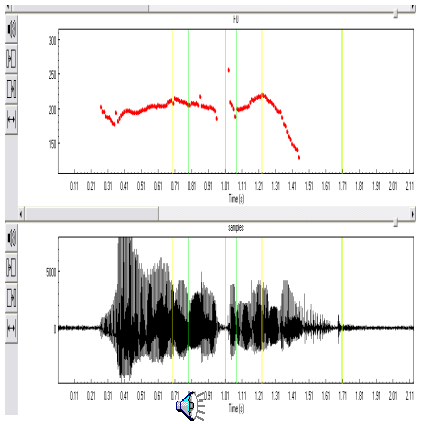
- "Can you see the *galah* open-cut mine?"
- "(No) I've got a *dingo* open-cut mine".



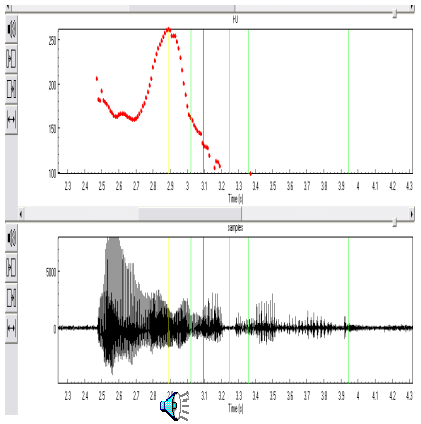
Narrow versus broad focus

- "New" information in the discourse tends to get accented (e.g. pitch accents on *galah, dingo* – sometimes referred to as a "focal" accent)
- When there is "broad" focus, English tends to accent the first and last content words of an intonational constituent.
- Listen to the following examples.

"Marianna made the marmalade."



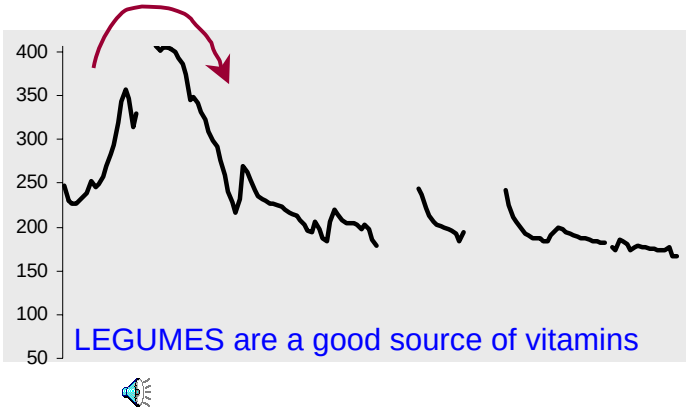
Broad focus – "Flat hat tune"



Narrow focus - Rise-fall tune ("I assert this")

3 examples showing different placement of focal accent (from Venditti, 2002)

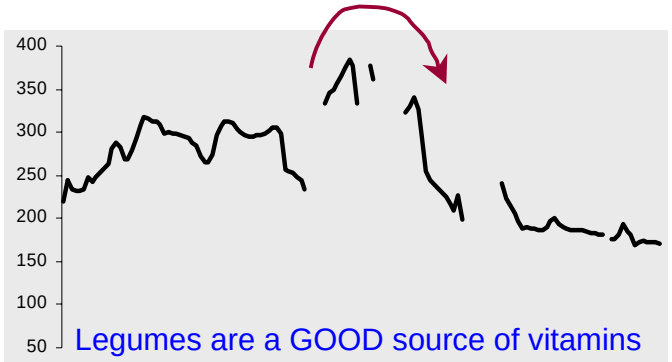
1.



The **rise-fall** tune (= "I assert this") shifts locations.

Placement of focal accent later….

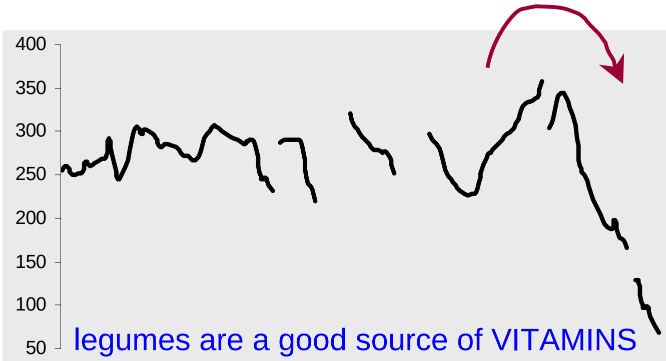
2.



The **rise-fall** tune (= "I assert this") shifts locations.

Placement of focal accent on final content word

3.



The **rise-fall** tune (= "I assert this") shifts locations.

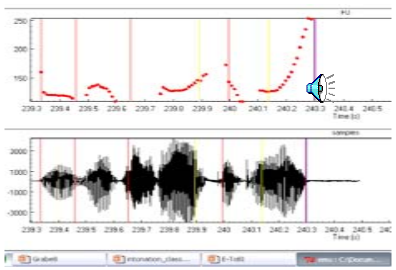
Levels of accentuation and the prosodic hierarchy

Nuclear Accent					x					
Accent					x					
Primary stress					x		x			x
Heavy Syllables	x				x		x		x	x
Syllables	x	x	x	x	x	x	x	x	x	x
	"I've got a dingo					open-cut				mine"

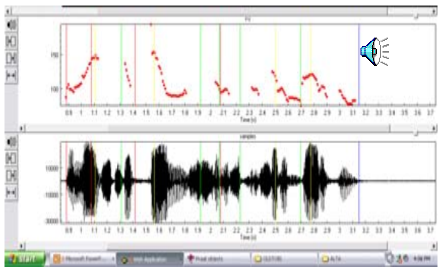
There are several levels of prominence, with rhythmic alternation of strong/weak syllables, and higher levels of prominence signaled by pitch accents. Note the additional prominence of the nuclear accent that is associated with prosodic phrasing.

Exercise 1: Which words are accented?

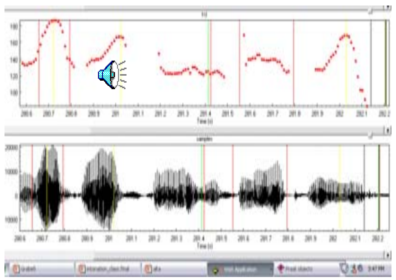
1.



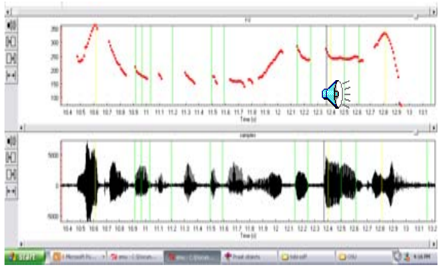
2.



3.



4.



Answers – accented words are in bold italics

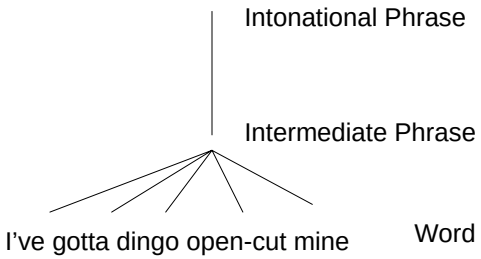
1. You head ***south again***.
2. ***Amongst*** her ***friends***, she was considered ***beautiful***.
3. So ***your gravestones***, is my ***canoes***.
4. ***Legumes*** are a good source of vitamins, and ***so*** are ***greens***.

Phrasing in E-ToBI

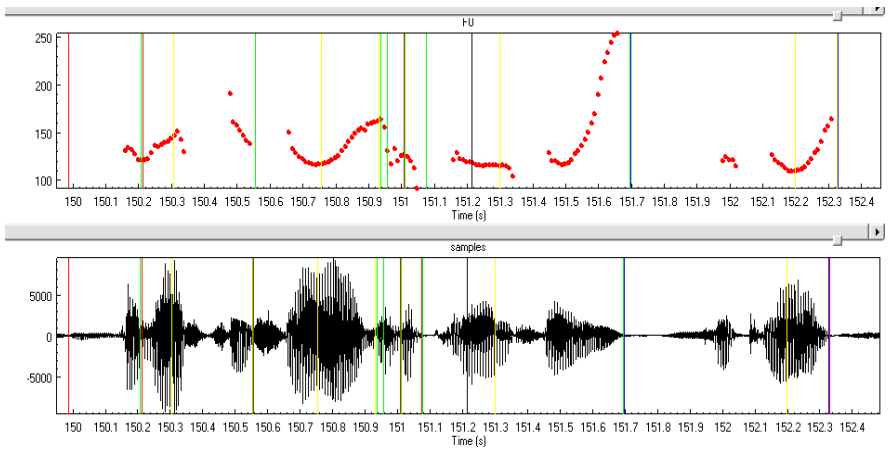
- The prosodic model underlying E-ToBI marks two levels of prosodic phrasing – first level of phrasing is the ***intermediate phrase*** - associated with a second pitch event called a ***phrase tone***
- Words are grouped into intermediate phrases.
- The ***last*** accented word in an intermediate phrase carries the ***nuclear accent*** (called the ***tonic*** in other models of English intonation). Nuclear accent represents the highest level of intonational prominence. The "focal" accent is usually the nuclear accent.

Intonational Phrases

- The highest level of intonational constituency in E-ToBI is the intonational phrase. It is associated with the third pitch event – the ***boundary tone***.



How many intermediate and intonational phrases?



The western side of the pine tree okay

How do we group words into intermediate phrases?

- Phrase-final lengthening – elongation of final syllable, foot or word at the edge of an intermediate phrase, relative to preceding or following speech material
- Some kind of tonal marking at the end or just before the end of an intermediate phrase
- In addition, there *may* be and upwards "pitch reset" and associated with the following word, suggesting the start of a new phrase. Pitch range is always re-set for each new intermediate prhase.
- Intermediate phrases MUST have a nuclear accent.

One representation

Nuclear accented words

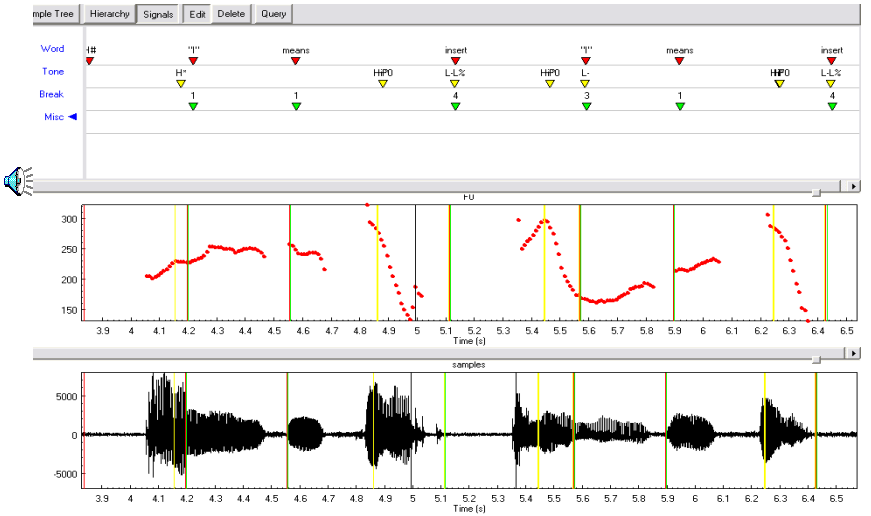
The western *side* of the *pine-tree*? *Ok*.

Intermediate phrase intermediate phrase intermediate phrase

Intonational phrase intonational phrase

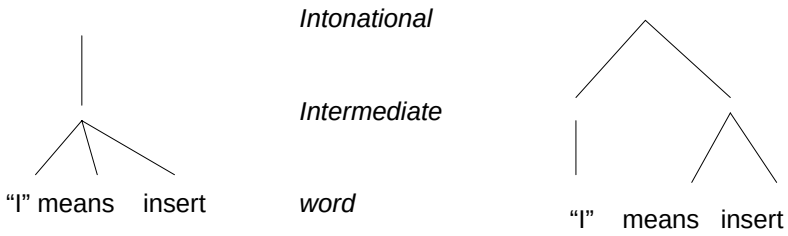
A classic ToBI- training example

- a) [(I means insert)] b) [(I) (means insert)]



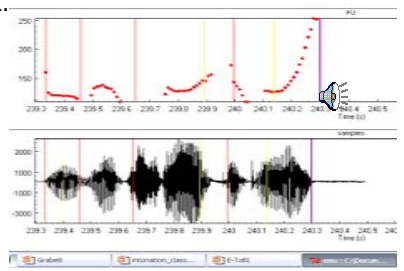
Prosodic representation

a) b)



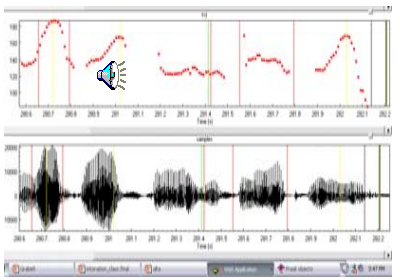
Exercise 2: Which words are nuclear accented? How many intermediate and intonational phrases?

1.

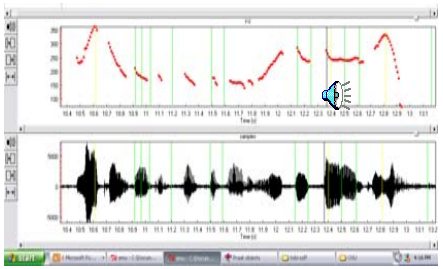


2.

3.



4.



Answers – nuclear accented words are the final accented words in intermediate phrases

1. [(You head ***south)(again***.)]
2. [(***Amongst*** her ***friends*)]** [(she was considered ***beautiful***.)]
3. [(So ***your gravestones)***, (is my ***canoes***.)] (ambiguous phrasing)
4. ***[(Legumes*** are a good source of vitamins)]

[(and so are ***greens***.)]

You can have more than one intermediate phrases in an intonational phrase. An utterance can consist of more than one intonational phrase.

How can we tell the level of phrasing: i.e. intonational or intermediate?

Cues to intonational phrase boundaries:

Same as for intermediate phrases, but more marked.

Also:

- "final lowering" i.e. a final extra low F0 value at the right edge of an intonational phrase.
- silent pause at the right edge of the intonational phrase
- presence of "glottalization" of vowel-onset words at the left edge of intonational phrases, which is not permitted at the onset of intermediate phrases.

TUNE

- the characteristic melody of an utterance
- A tune is broken down into component low (L) or high (H) tone targets
- Radically different annotation of tune from the British School of Intonation models (e.g. Halliday, O'Connor and Arnold; Crystal)

The anatomy of a tune - 3 tone events

- Pitch accents: high or low tones (e.g. H*, L+H*, L*, L*+H) – the tone with a * diacritic is locally aligned with primary stressed syllable of accented word (pitch accent inventory is SAME in nuclear and pre-nuclear contexts)
- Phrase tones: L- H- - describes pitch between nuclear accented syllable and intermediate phrase edge;
- Boundary tones: L% H% - locally aligned with right edge of intonational phrase

The ToBI Tone tier inventory (simplified from Pierrehumbert, 1980)

Pitch accents	H* L* L+H* L*+H	!H* L+!H* L*+!H H+!H*
Phrase tones, or phrase "accents"	L- H-	!H
Boundary tones	L% H%	

Some Common Tunes in AuE

- Declarative hat pattern with final fall (H*) H* L-L%
- Continuation rise or fall-rise (H*) H* L-H%
- Rise-fall ("I assert this") L+H* L-L%
- Rise-fall-rise (emphatic) L*+H L-H%
- Yes-No question (H*) H* H-H%
- High Rising Terminal (statement high rise) (H*) L* H-H% or (L*) H* H-H%)

H* and L* Pitch Accents

- H* (High) and L* (Low) pitch accents - relative to the pitch range for each intermediate/intonational phrase

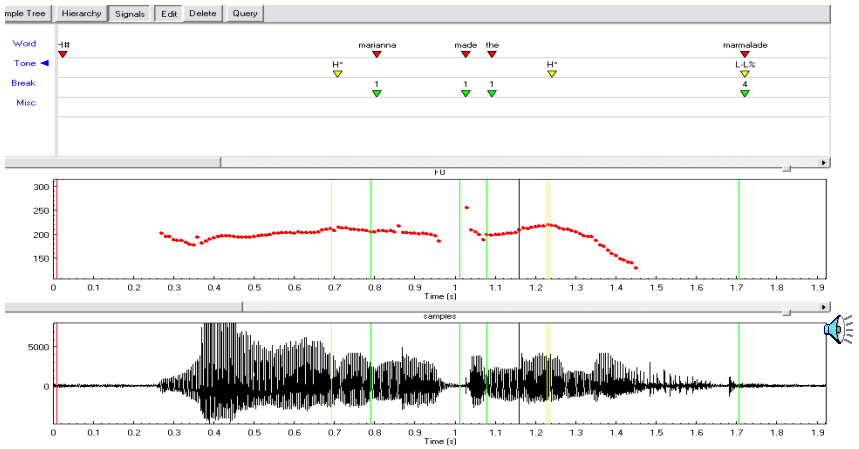
H*

- F0 peak within, or slightly after accented syllable (particularly if syllable coda is a sonorant)
- Tone target in the mid-upper part of the speaker's pitch range
- Realized even higher in expanded pitch range

L*

- Valley or low sustained F0 target in the relatively low part of the speaker's range
- Even lower in expanded pitch range

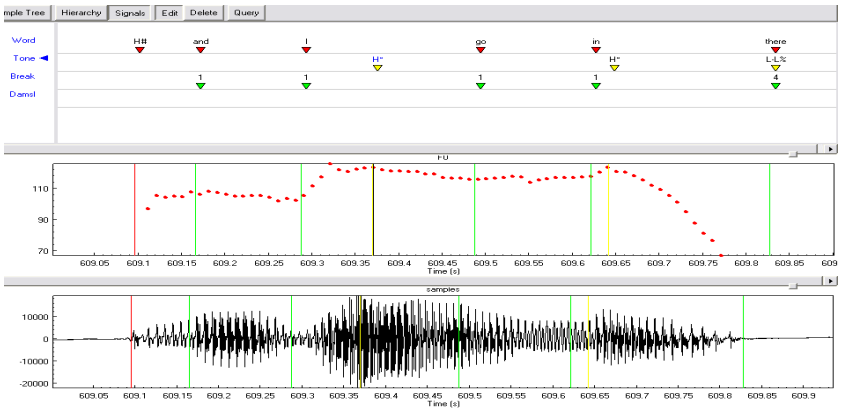
H* Accents – American English



Marianna made the marmalade.

H* H*

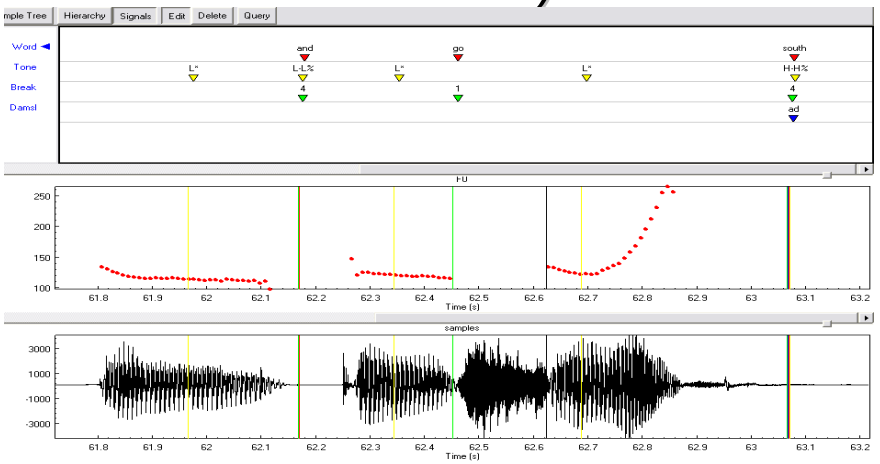
The “Flat Hat” pattern – Australian English



And I go in there

H* H*

L* pitch accent (part of High Rising Terminal)



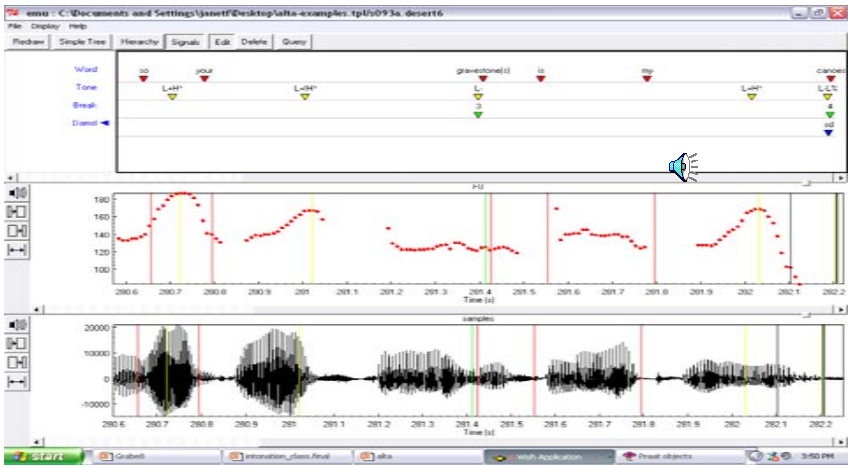
….. And go south

L* L*

Bitonal Accents

- L+H* The H* tone target is preceded by a rise from a low part of the speaker’s range – sounds like a “rising” accent
- L*+H A more “emphatic” accent, where a L* tone target is followed by a rise to the mid to upper pitch range – sounds like a “scooped” accent. This accent is relatively rare compared to H* or L+H* accents.

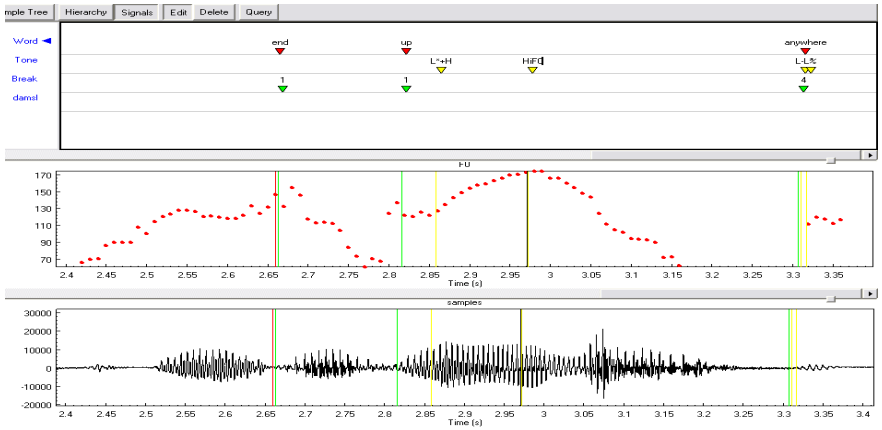
L+H* pitch accents in AuE



So your gravestones, is my canoes”.

L+H* L+!H* L+H*

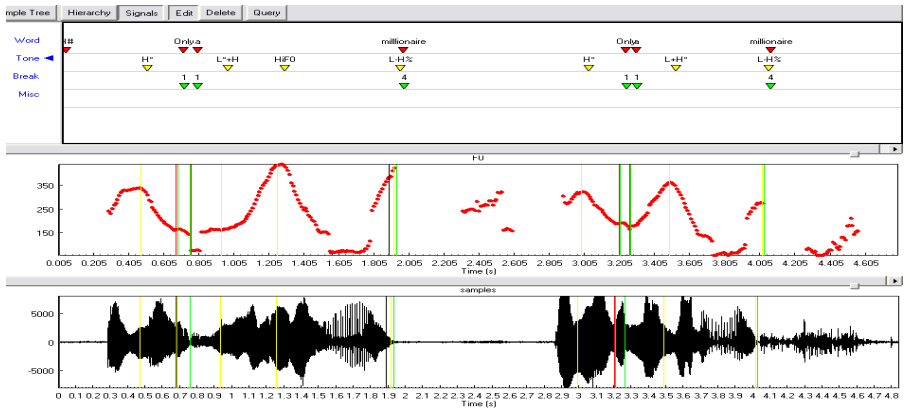
L*+H in AuE ("emphatic" rise-fall)



(By those directions I don't think I'll)

end up anywhere
L*+H

L*+H versus L+H* - a question of * tone alignment (1)

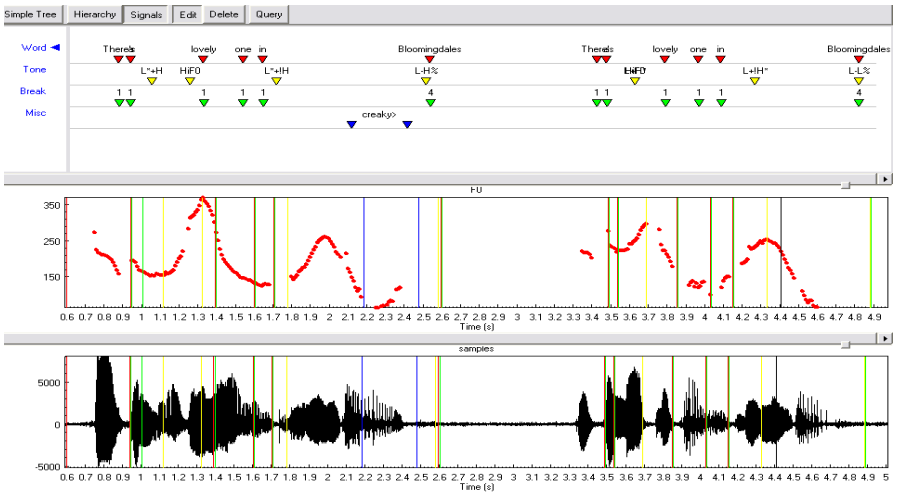


Only a millionaire — L*+H

Only a millionaire — L+H*

BUT…note differences in pitch range!

L*+H versus L+H* (2)



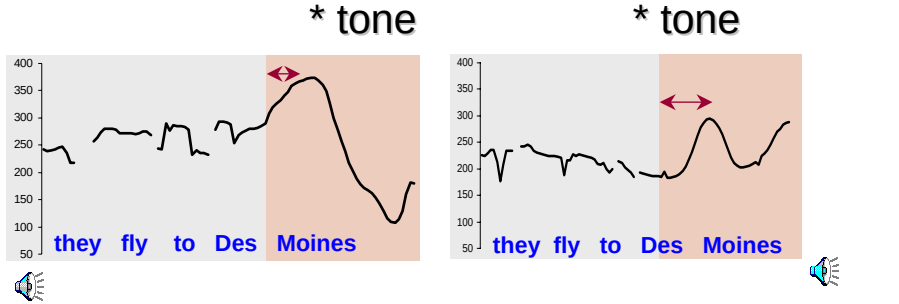
lovely ..Bloomingdales. — L*+H L*+!H*

lovely...Bloomingdales. — L+H* L+!H*

Alignment differences can cue "assertion" vs. "suggestion" in General American English (from Venditti, 2002)

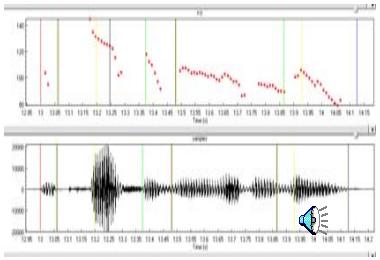
A: I'd like to fly to Davenport, Iowa on TWA.

B: TWA doesn't fly there ..

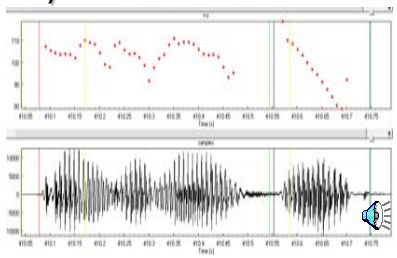


BUT….. Note differences in boundary tones!

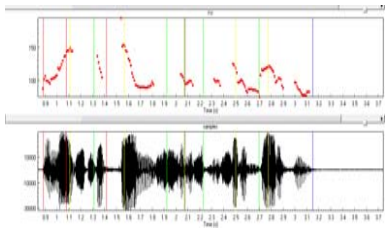
Exercise 3. H* , L+H*?

1. 

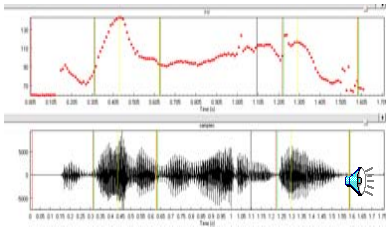
"cross"

3. 

"underneath what"

2. 

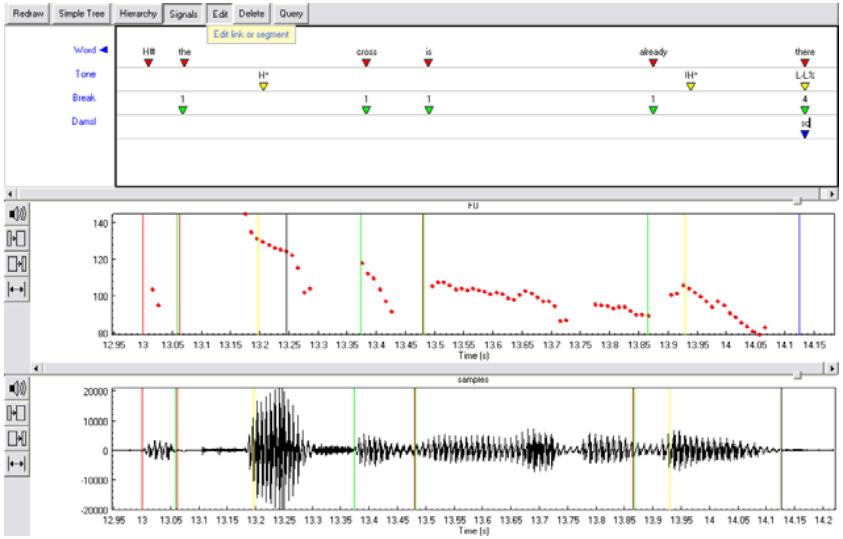
"Amongst ... friends ... beautiful"

4. 

"really……day"

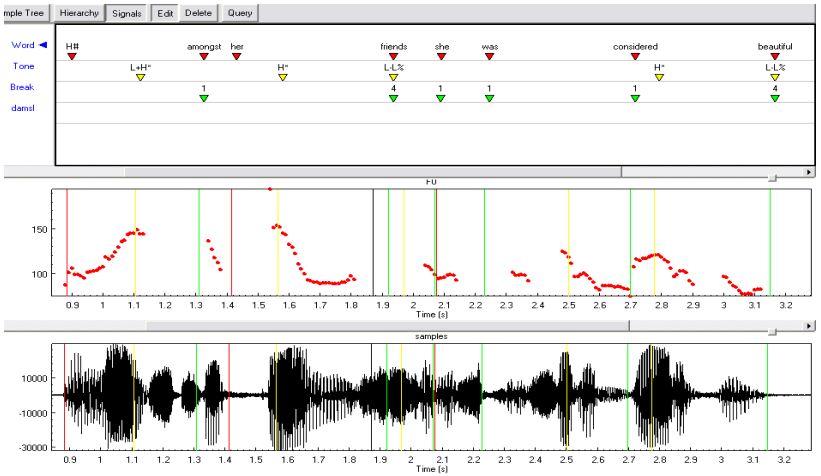
1. The cross is already there.

H*



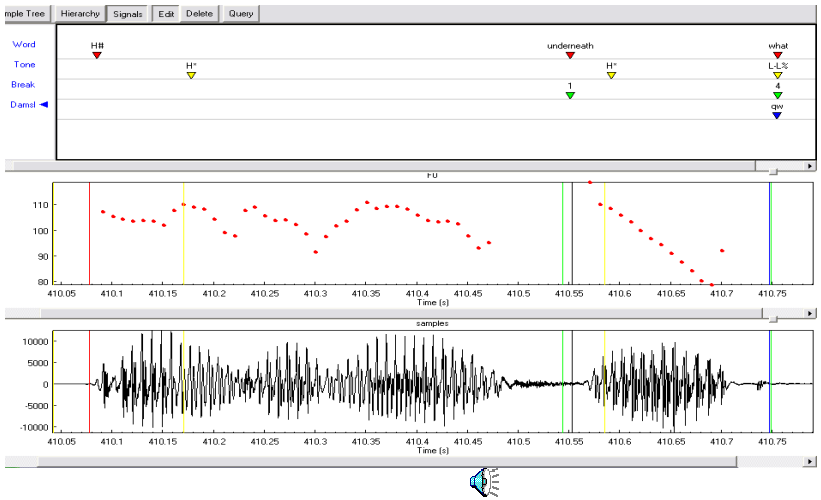
2. Amongst her friends, she was considered beautiful.

L+H* H* H*

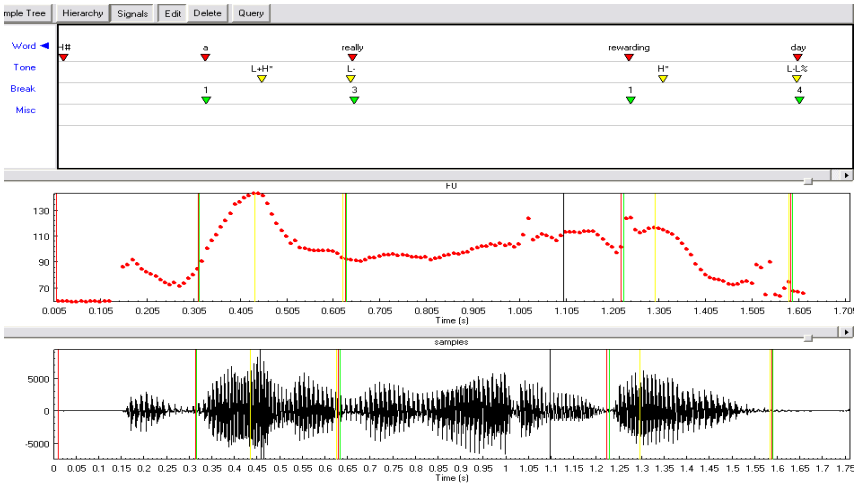


3. underneath what?

H* H*



4. A really rewarding day.
L+H* H*



Downstep

- "Phonologically triggered compression of the pitch range that lowers the F0 targets for any (subsequent) H tones" ((Beckman and Ayers, 1994:section 2.8)
- The tone target has a somewhat lower pitch than the preceding "trigger" tone, or tone combination – this is indicated by a "!" on the downstepped H tone
- Pitch accents can be downstepped

e.g. !H* L+!H* L*+!H H+!H*

- Phrase tones can be downstepped

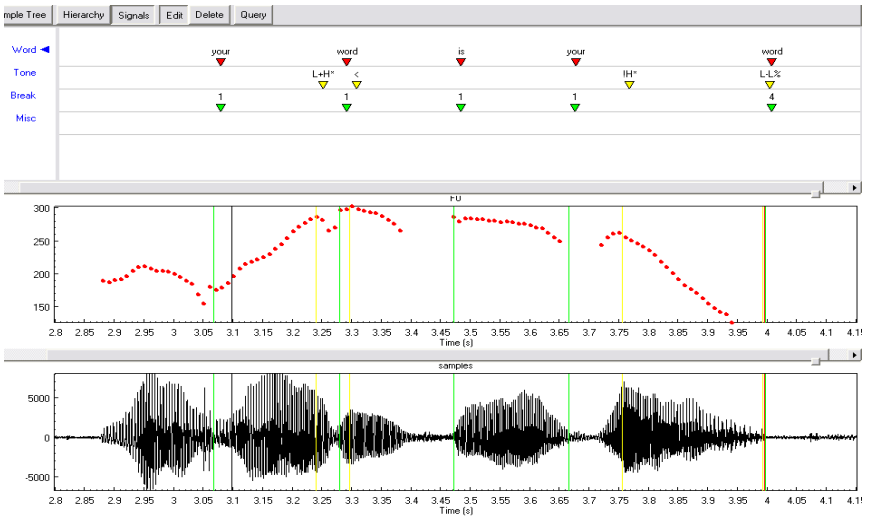
e.g. !H- !H-L% !H-H%

Downstep cont.

- In the original Pierrehumbert model (on which ToBI was based) – the downstepped pitch accent or boundary tone was not explicitly marked.
- Trigger tones were always bitonal accents,

e.g. L+H* L*+H, plus accents not included in ToBI, e.g. H*+L (this accent is "recoverable" from a H* !H* sequence, i.e. H*+L has been collapsed into the H* category)

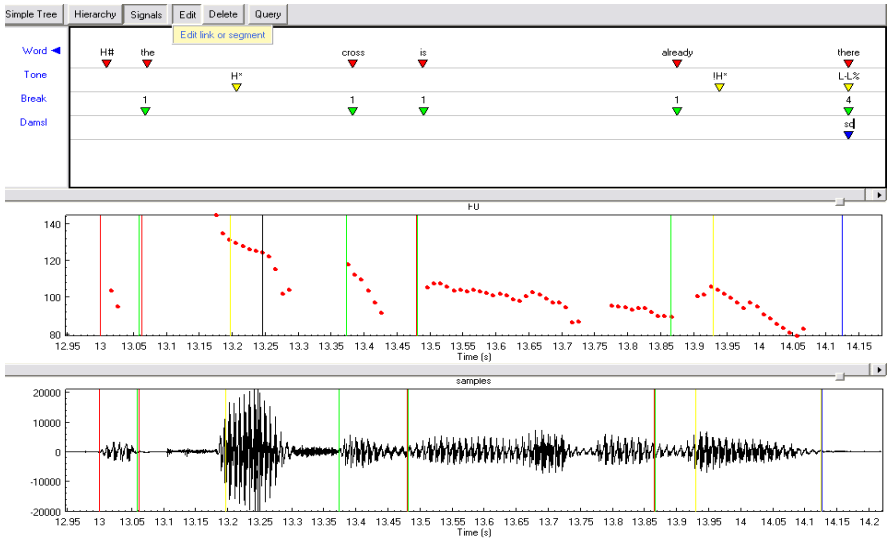
Example 1. Your word is your word
L+H* !H*



Example 2

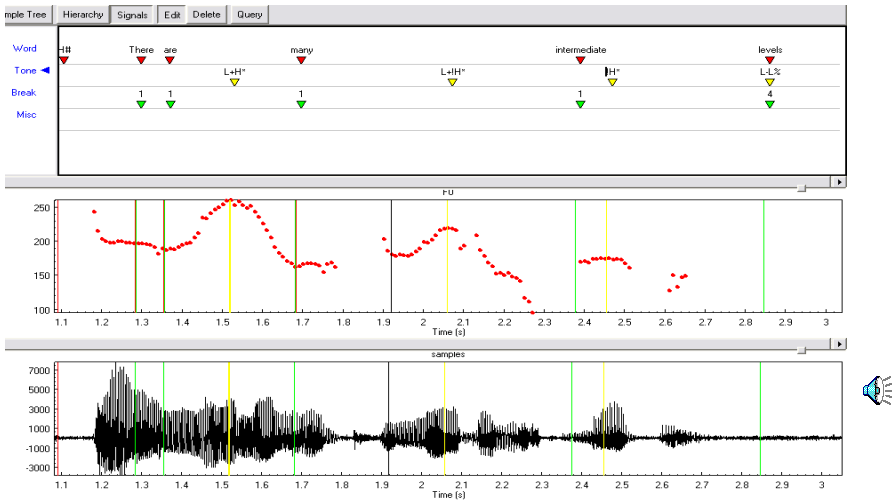
The cross is already there.

H* !H*



Example 3. There are many intermediate levels.

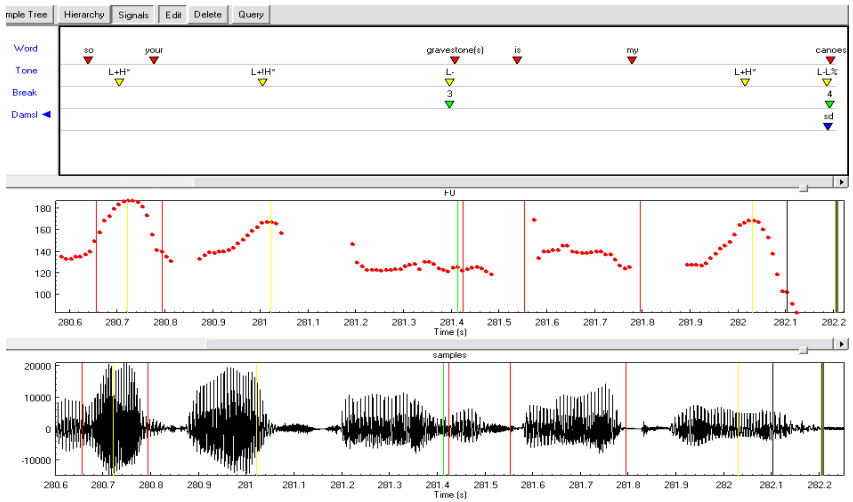
L+H* L+!H* H*



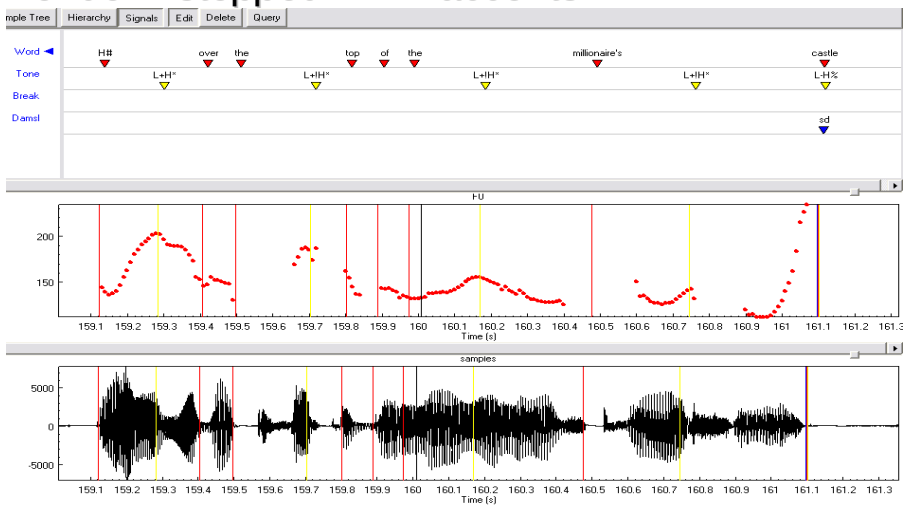
Example 4. Australian English

So your gravestones, is my canoes".

L+H* L+!H* L+H*



Example 5. A "catathesis" chain – sequence of downstepped L+H* accents



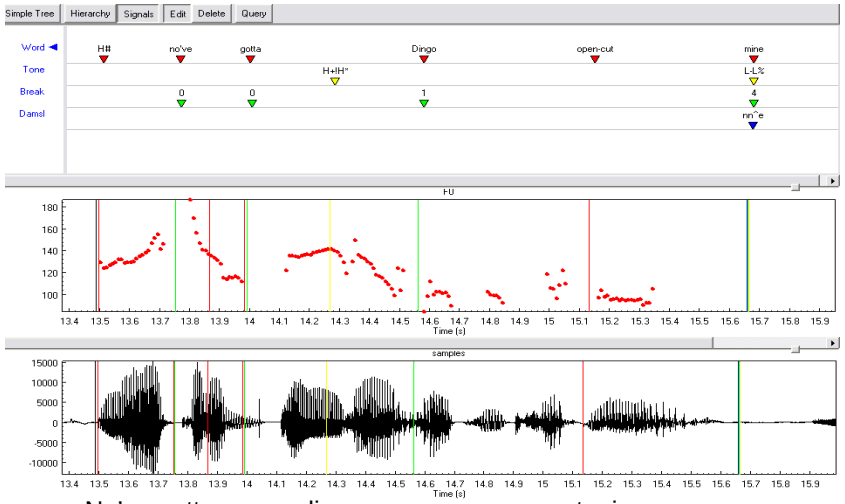
Over the top of the millionaire's castle

L+H* L+!H* L+!H* L+!H*

Another downstepped accent H+!H*

- A clear step down from preceding high pitched unaccented material which is not associated with an initial pitch accent. The tone label should be aligned locally with the downstepped pitch accent.

Example: H+!H*



No've gotta dingo open-cut mine

H+ !H*

E-TOBI- Part 2

Janet Fletcher
University of Melbourne

Phrase tones and Boundary tones

- Phrase tones L- H- !H-
- Influences the pitch after the nuclear accent and the right edge of the intermediate phrase
- H- phrase tone is a trigger for “upstep”, raising the level of the % boundary tone
- Boundary tones H% L% are locally aligned with the right edge of an intonational phrase
- Phrase tones combine with boundary tones define intonational phrase edges

Phrase tone/Boundary tone combinations

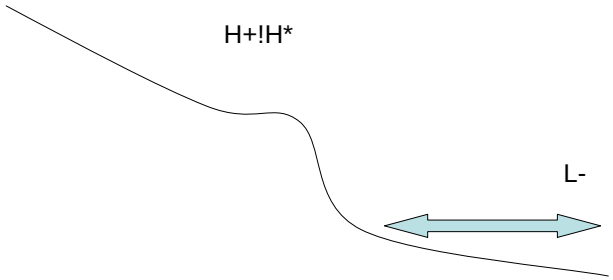
- L-L%
  - Perceived as falling pitch
  - Often, standard declarative sentence
  - F0 contour usually falls sharply
- L-H%
  - Perceived as rising pitch, usually to mid-level or higher
  - often indicates continuation
  - F0 rises usually to mid level or beyond, from a relatively low point in F0 contour

Phrase tone/Boundary tones cont.

- H-H% / !H-H%
  - Perceived as rising pitch, usually into the highest part of a speaker’s pitch range
  - Often associated with yes/no questions or statement high rises
  - F0 rises to a high level in the speaker’s range
- H-L% / !H-L%
  - Perceived as level pitch
  - Common as a “listing” intonation or “enumerations”
  - F0 contour usually sustained at mid-pitch range

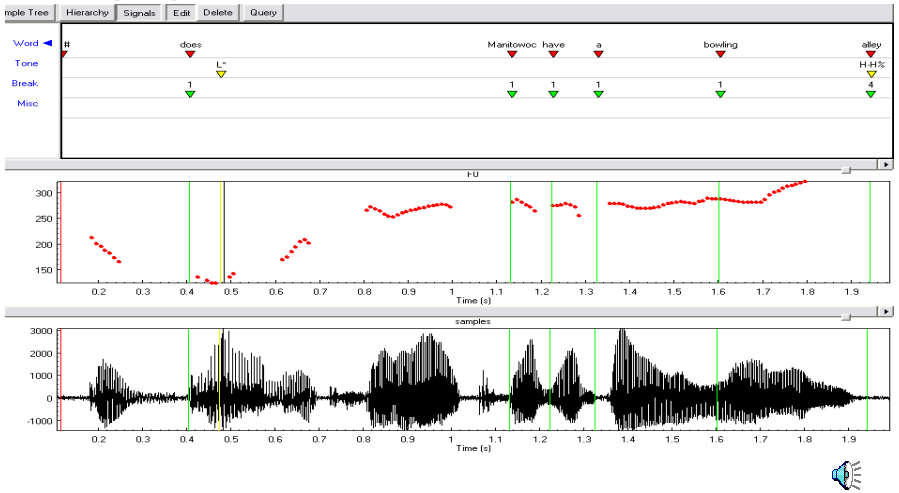
Phrase tones: long vs short tails

- (No I've gotta DINGO open cut mine).



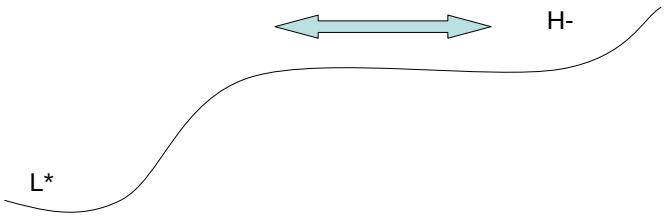
- L- phrase tone accounts for long low stretch after nuclear accent – a long low tail – reaching back from intermediate phrase edge

H- Phrase tone accounts for sustained high pitch level after nuclear accent

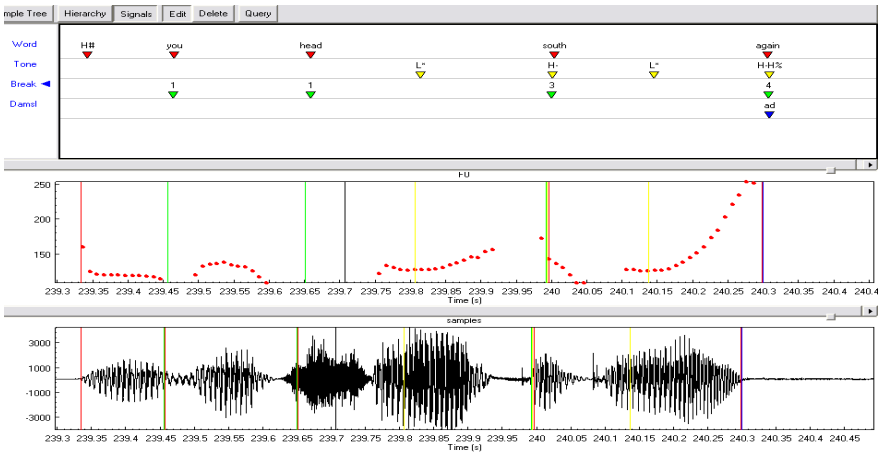


A "long high tail"

- Does MANITOWIC have a bowling alley?



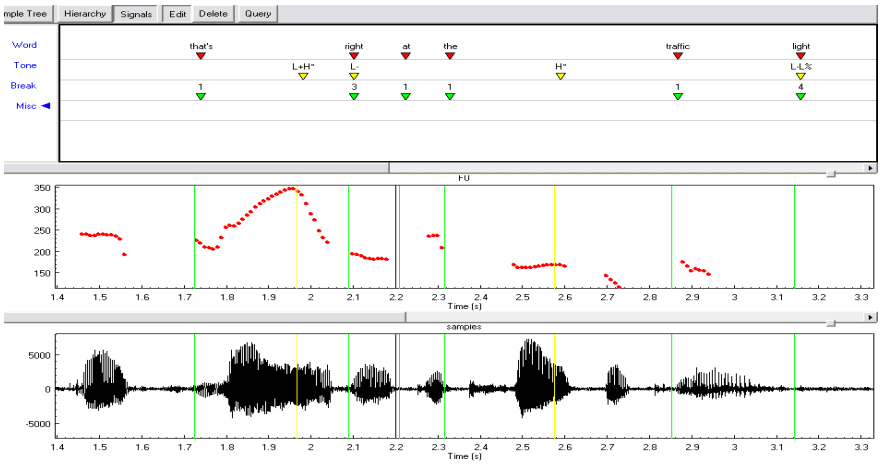
Short tails: intonational phrase internal & vs H-H% combination



you head south again

L* H- L* H-H%

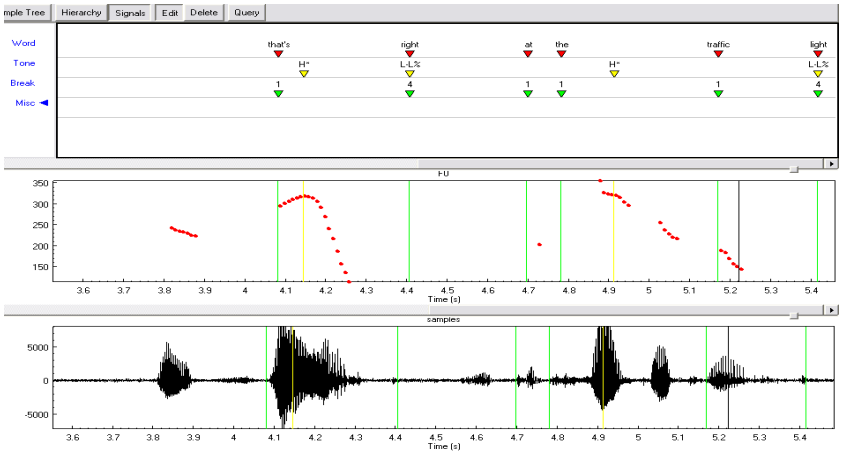
L- tones internal to Intonational phrases vs L-L% configuration



That's right at the traffic light

L+H* L- H* L-L%

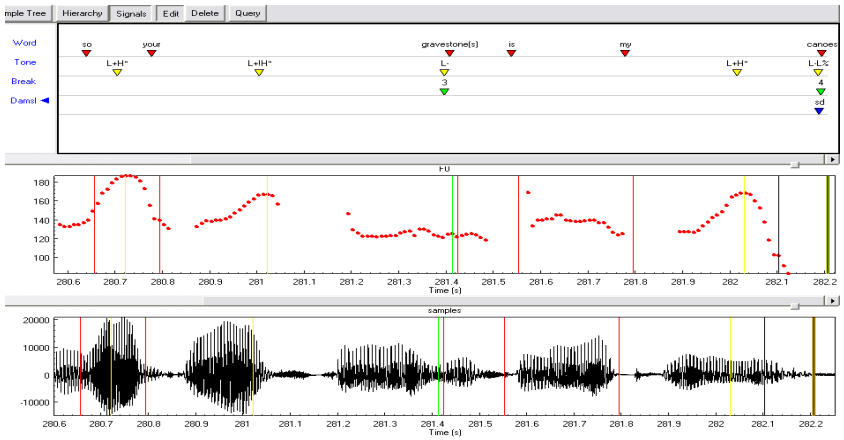
L- as part of L-L% combination



That's right at the traffic light

H* L- L% H* L-L%

L- vs L-L% (with final lowering)

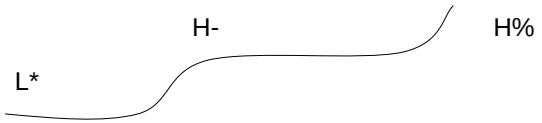


So your gravestones, is my canoes".

L+H* L+!H* L- L+H* L-L%

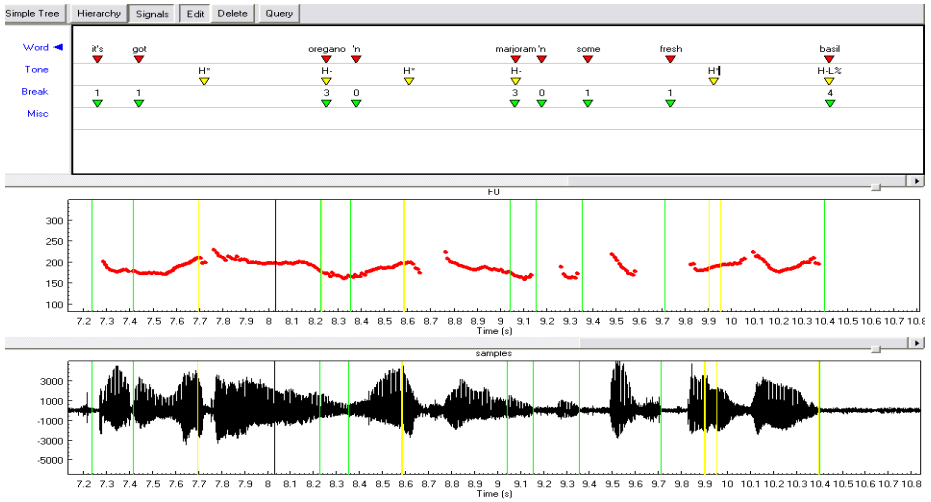
H- and !H- as triggers of "upstep"

- Does Manitowic have a bowling alley



- H- upsteps following boundary tone (H%) – i.e. H% is realized as a higher F0 target

H-L% or !H-L% (controversial!)



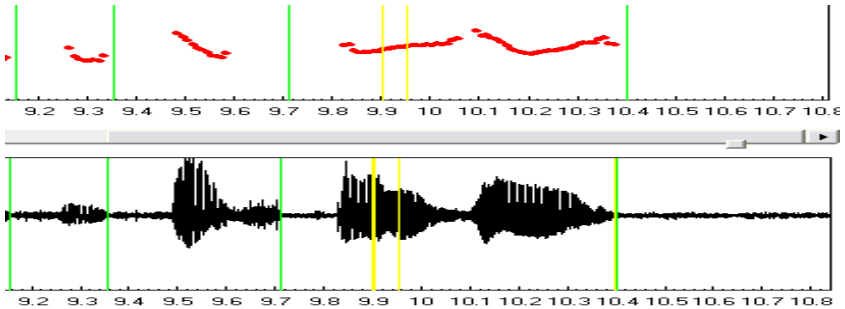
It's got <u>oregano</u> 'n <u>marjoram</u> 'n fresh <u>basil</u>

H* H- H* H- H* H-L%

H- upsteps following L% tone

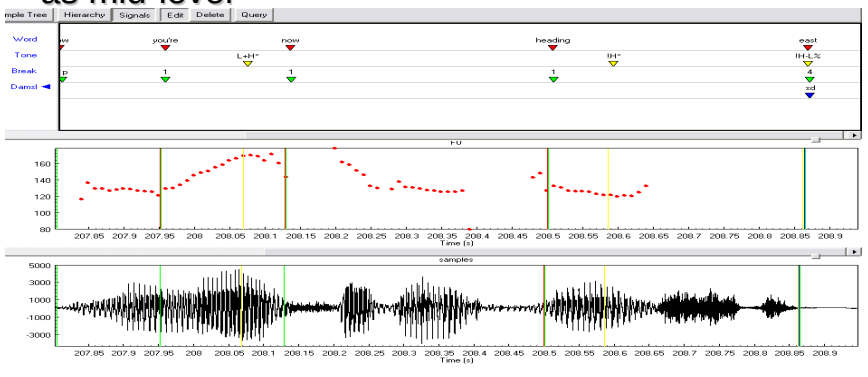
'n some fresh <u>basil</u>

H* H-L%



!H-L%

- After a bitonal accent (L+H*) or underlying H*+L – the !H will RAISE the following L% tone – realised as mid-level

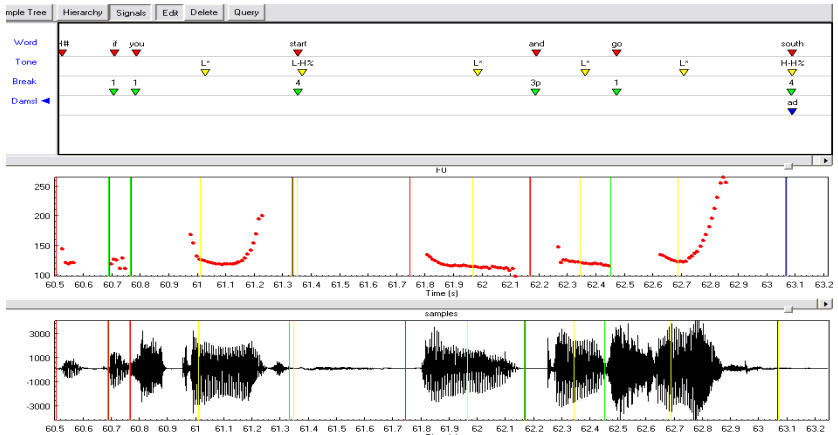


….. You're <u>now</u> heading <u>east</u>

L+H* !H* !H-L%

More rises: L* L-H% vs L* H-H%

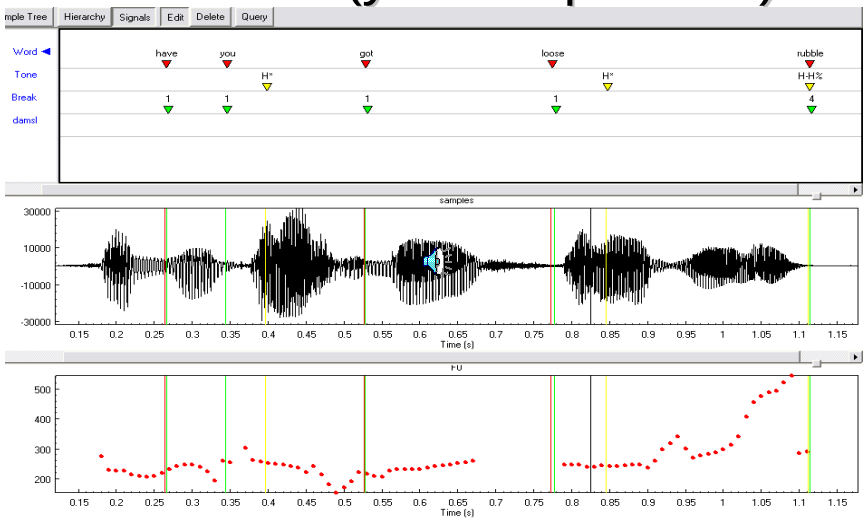
- ToBI allows the annotation of several kinds of final rises e.g. low rise & low-onset high rise (HRT)



if you <u>start</u> <u>and</u> <u>go</u> <u>south</u>

L* L-H% L* L* L* H-H%

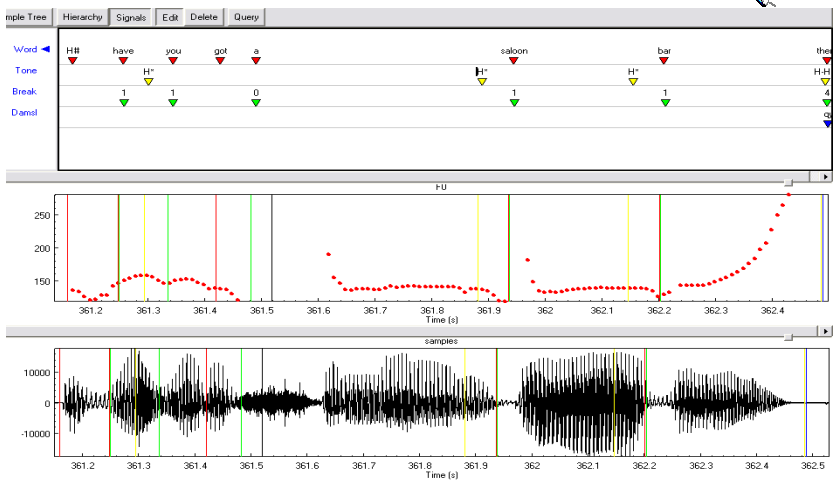
H* H-H% (yes/no question)



Have you got loose rubble?

H* H* H-H%

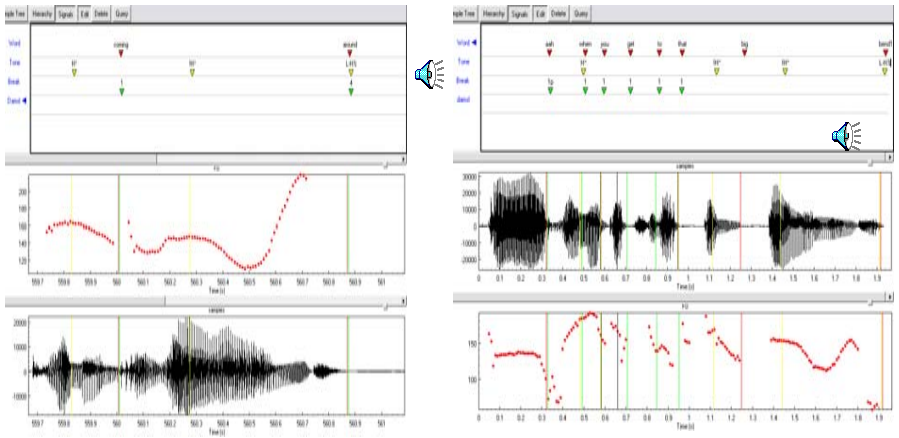
Another yes/no question



Have you gotta saloon bar there?

H* H* H* H-H%

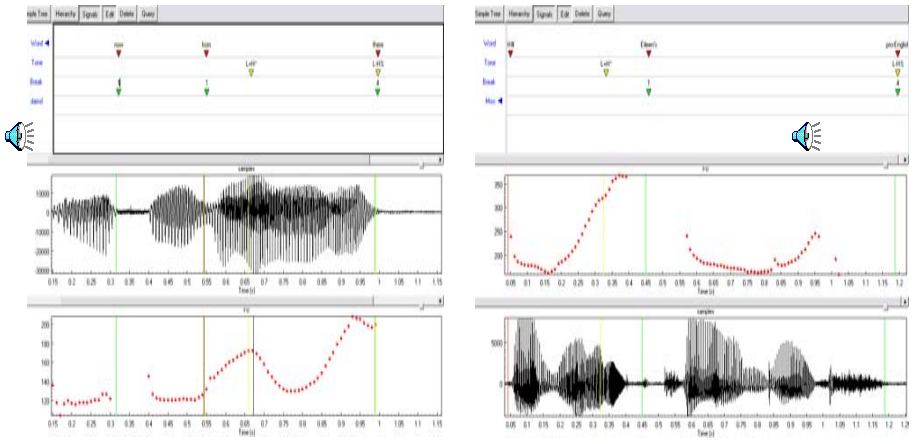
Fall-rise tunes - H*/ !H* L-H% (continuation)



Coming around When you get to that big bend

H* !H* L-H% H* !H* !H* L-H%

Rise-fall-rise tunes L+H* L-H% (“suggestion”)

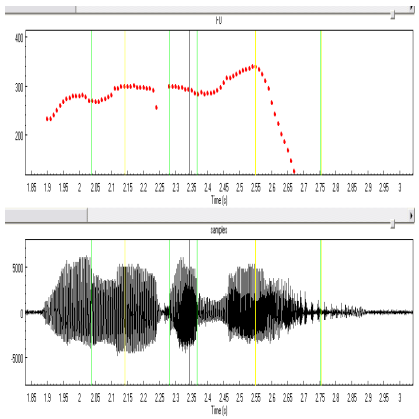


Now from there. Eileen’s pro-English

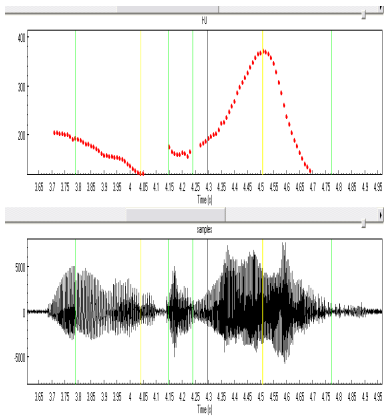
L+H* L-H% L+H* L-H%

Exercise 4 – label the following tunes

1.



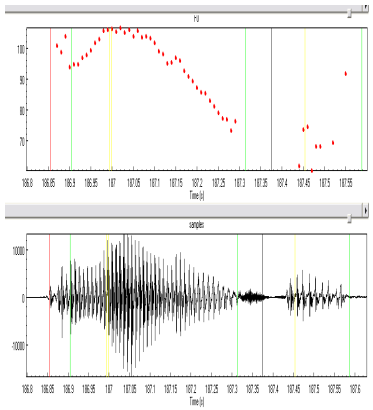
2.



You need a loan.

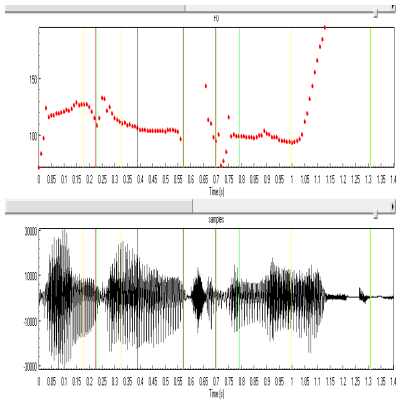
Examples- AuE

3.



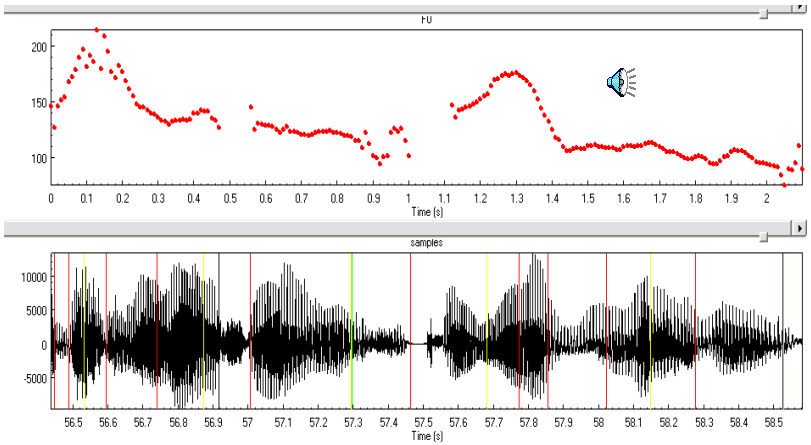
A rolling stone

4.



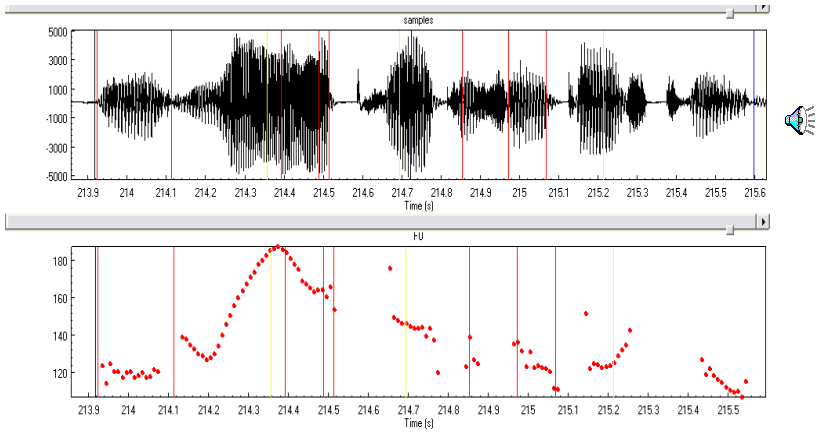
One going to the left…

5.



But that only leaves me a quarter of the way there.

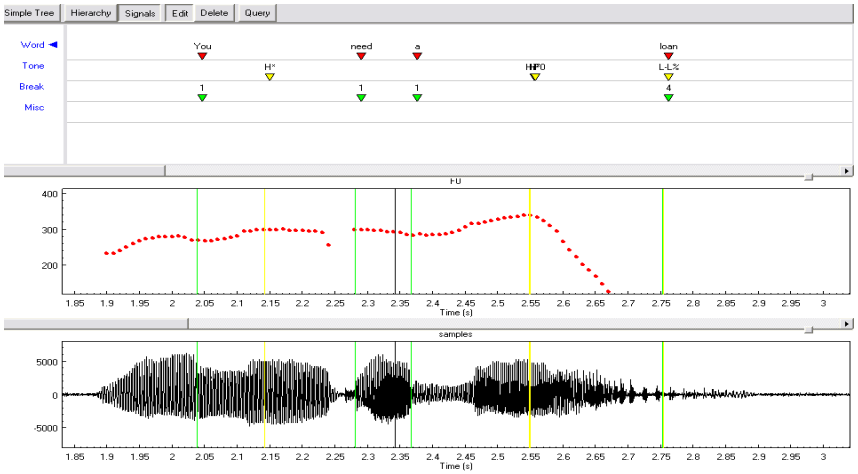
6.



Where on the page is your ghost-town?

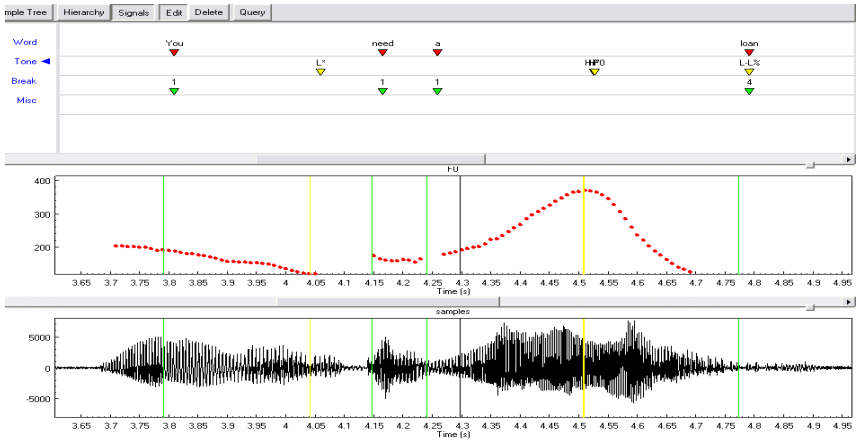
Suggested transcriptions

1.



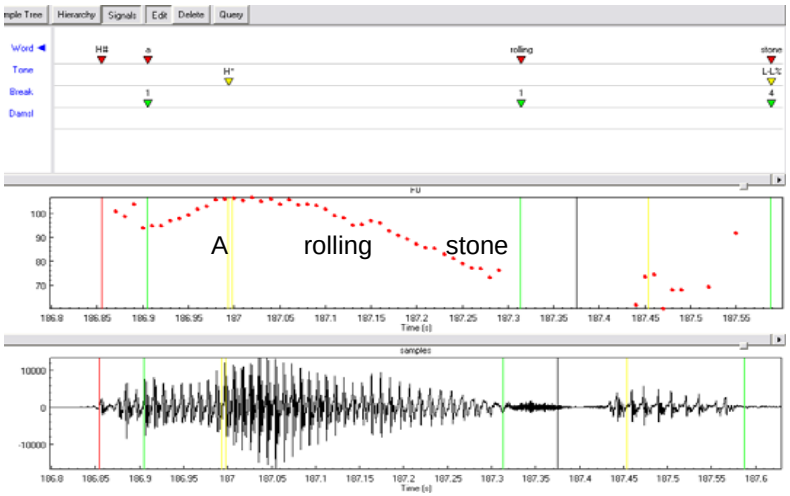
You need a loan.
H* H* L-L%

2.



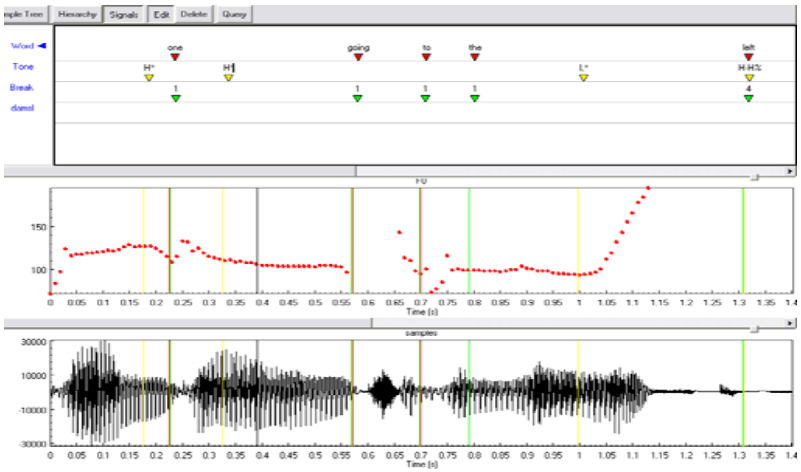
You need a loan.
L* L+H* L-L%

3.



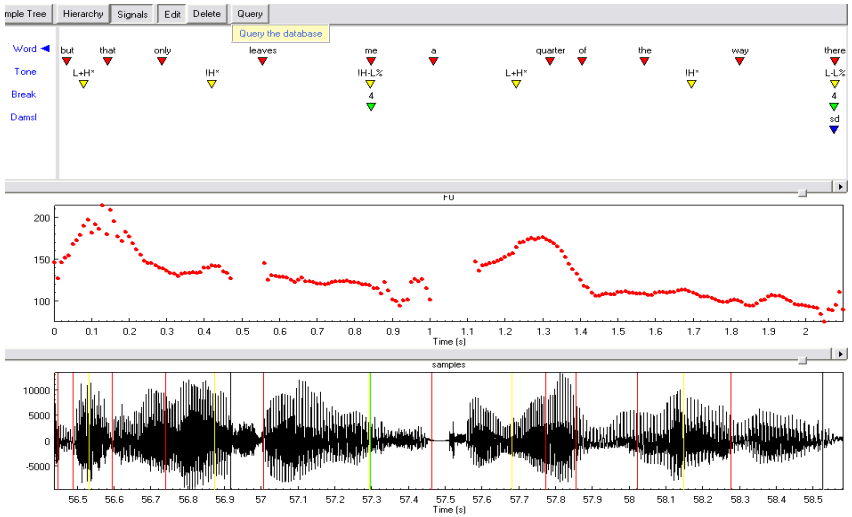
A rolling stone
H* L-L%

4.



One going to the left…
H* !H* L* H-H%

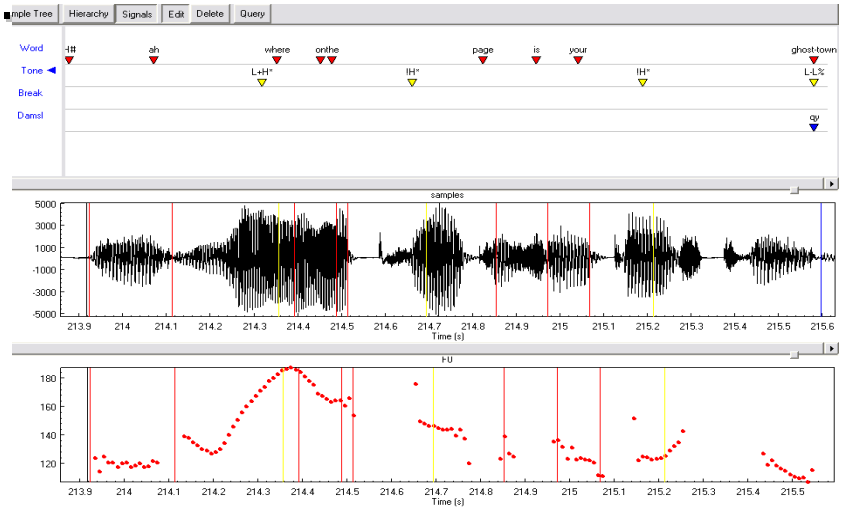
5.



But <u>that</u> only <u>leaves</u> me a <u>quarter</u> of the <u>way</u> there.

L+H* !H* !H-L% L+H* !H* L-L%

6

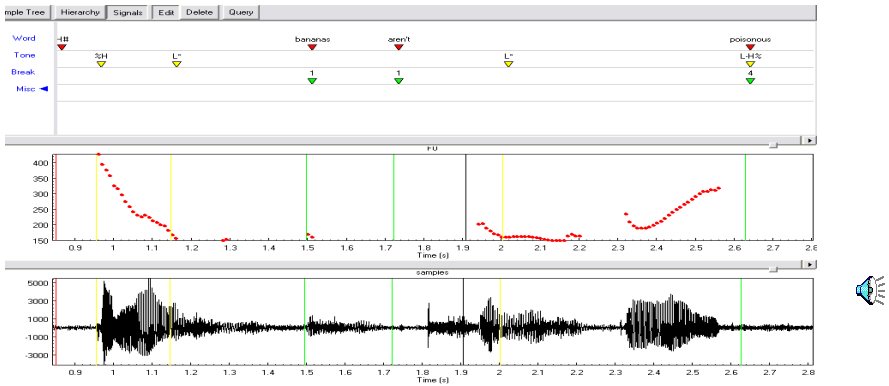


Aaah <u>Where</u> on the <u>page</u> is your <u>ghost-town?</u>

L+H* !H* !H* L-L%

%H boundary tune (part of “surprise-redundancy” tune

- Phrase that begins relatively high, in absence of H* accent or H+!H* accent



<u>Bananas</u> aren’t <u>poisonous</u>

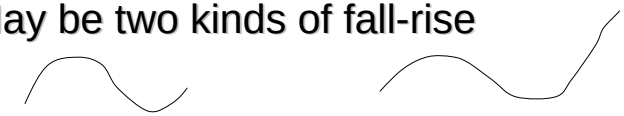
%H L* L* L-H%

Pitch Range HiF0

- Models like ToBI assume that pitch range is “reset” for each intermediate phrase
- HiF0 – additional label associated with one of the H* or L*+H accents
- NOT attached to phrase labels (due to potential effects of upstep (e.g. H% is assumed to be upstepped by preceding H- etc.)
- Other ways of indicating pitch range: H L tone targets; “final lowering” (extra lowering at end of a L-L% boundary, usually marking turn completion); upstep and downstep rules

What's special about Australian English?

- Rising variety: frequency of statement high rises (L* H-H%, H* H-H%;rising (L+H*) accents)
- May be two kinds of fall-rise



(L)+H* L-H% H*+L H-H%

- Caveat – H*+L redundant accent category, but is that a problem?

Break Indices

0 phonetic indication of clitic groups e.g. "gotta" "didja" etc

1 most intermediate or intonational phrase-medial word boundaries

2 strong disjuncture with pause, but no tonal target marking the edge of phrase, or

Weak disjuncture but no clear tonal marking at intermediate or intonational phrase boundaries

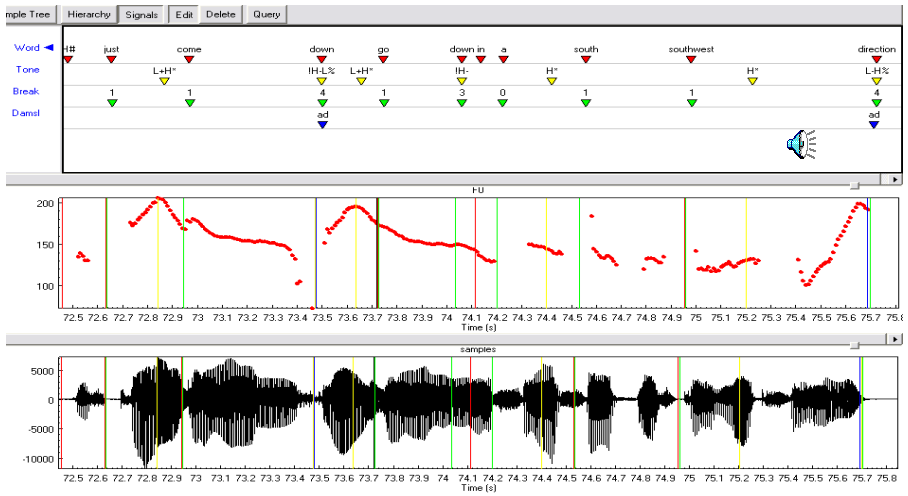
Break Indices cont.

3. Intermediate phrase boundary (marked by phrase tone that influences region from the last pitch accent (i.e. nuclear accent) to the boundary

4. Full intonational phrase boundary – final boundary tone

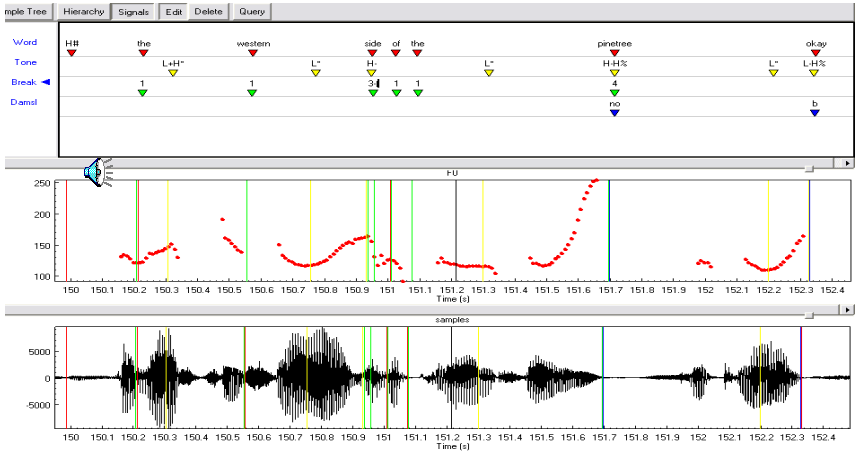
Break Indices – focus on levels of JUNCTURE in the speech stream

Most ambiguities: 3 vs 4



Just	come	down	go	down	in a	south south-west	direction
1	1	4	1	3	0	1 1	4

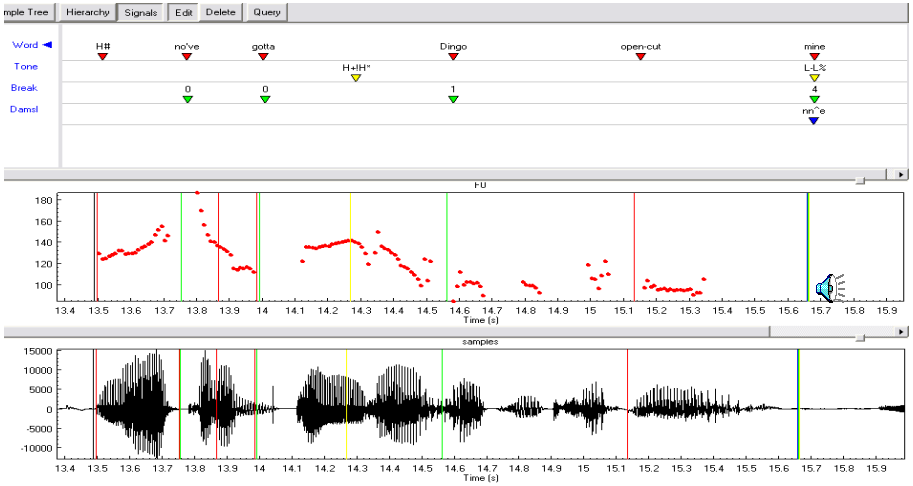
An ambiguous example



the western side of the pine-tree okay

1 1 3- 1 1 4 4

0 vs 1



no've gotta dingo open-cut mine

0 0 1 1 4

Useful diacritics with Break Indices

- Uncertainty 3- 4-
- Disfluencies e.g. prolonged hesitation or hesitation pause - 1p 3p 4p
- There will always be inter-transcriber disagreement!

Conclusions

- Intonation is hard, especially in connected speech
- Disagreement is normal
- Consistency is the key
- ToBI is flexible! Categories can (should!) be modified, added, depending on the English variety, OR language
- ToBI tone transcription can be translated into British School of Intonation categories

Some recommended readings

*Beckman, M. & J. Pierrehumbert (1986) Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255-309.

Bolinger, D. (1972) *Intonation* [introduction and chapter 1]. Penguin Books, Ltd.

Fletcher, J. & Harrington, J. (2001) High Rising terminals and Fall-Rise Tunes in Australian English. *Phonetica,* 58, 215-229

Fletcher, J., Stirling, L., Wales, R., & Mushin, I. . (2002) Rising intonation and dialogue acts in Australian English *Language and Speech*, 45(3), 229-254. London:Kingston Press

*Jun, Sun-Ah (2004) ed. *Prosody and typology: a unified approach.* OUP

*Ladd, D.R. (1996). *Intonational Phonology*, CUP

Pierrehumbert, J. & Hirschberg (1990) The meaning of intonational contours in interpretation of discourse. In Cohen, et al. (eds.) Intentions in Communication. MIT Press.

Venditti, J. (2002). “Intonational meaning in Discourse” (Web tutorial)