Introduction to the ToBI transcription system: examples from MAE ToBI

Mariapaola D'Imperio

**Aix Marseille Univ, Laboratoire Parole et Langage, CNRS, Aix-en-Provence (France)*

& Institut Universitaire de France (IUF)

Aix Summer School on Prosody

7/9/2016

The Plan for Today

- Some background on the original TOBI system
  - Reminder about basic AM concepts
  - Motivation, criteria, general aims
  - Tune anatomy
  - Tonal event description
- Then: work on the transcription of some ToBI exercise files from the MIT Open CourseWare.

D'Imperio, Cangemi & Grice 2016 (Laboratory Phonology)

labphon Laboratory Phonology
Start Submission Become a Reviewer

Special Collection

Advancing Prosodic Transcription

Collection launched: 30 Jun 2016

This series aims to stimulate debate on the theory and practice of prosodic transcription, and its role in prosodic typology, phonological theory, second language teaching, as well as in speech synthesis and recognition applications. Papers include selected keynotes from the Advancing Prosodic Transcription Workshops I and II (held in Stuttgart and Lisbon), along with regularly submitted papers inspired by the discussion initiated at these workshops. This collection has been edited by Mariapaola D'Imperio, Martine Grice, and Francesco Cangemi. See the editorial for an overview and discussion of the collection.

Collection Articles

Journal article

Analysis of Intonation: the Case of MAE_ToBI

Carlos Gussenhoven

30 Jun 2016
7 (1):10

Share this:

LATEST ARTICLES

Phonological and Semantic Cues to Learning from Word-Types

Richtsmeier — 09 Aug 2016

Share:

Analysis of Intonation: the Case of MAE_ToBI

Gussenhoven — 30 Jun 2016

Share:

The Importance of a Distributional Approach to Categoriality in Autosegmental-Metrical Accounts of Intonation

Cangemi & Grice — 30 Jun 2016

Share:

Origins of ToBI

- ToBI (Tones and Break Indices system for prosodic transcription)
- Devised by multi-disciplinary team in the U.S.A., 1992-present
- Based on the intonational model of American English developed by Janet Pierrehumbert (1980)

- *http://www.ling.ohio-state.edu/~tobi/ame_tobi/*
- *http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/*

Linguistic Models of intonation

- A linguistic model of intonation must contain two components:
  - a **phonological component**, characterizing melodic contours through a series of elements carrying linguistic meaning
  - a **phonetic component**, explicitly describing the mapping between the underlying form and the melodic continuum (Ladd 1996/2008)

Pierrehumbert 1980

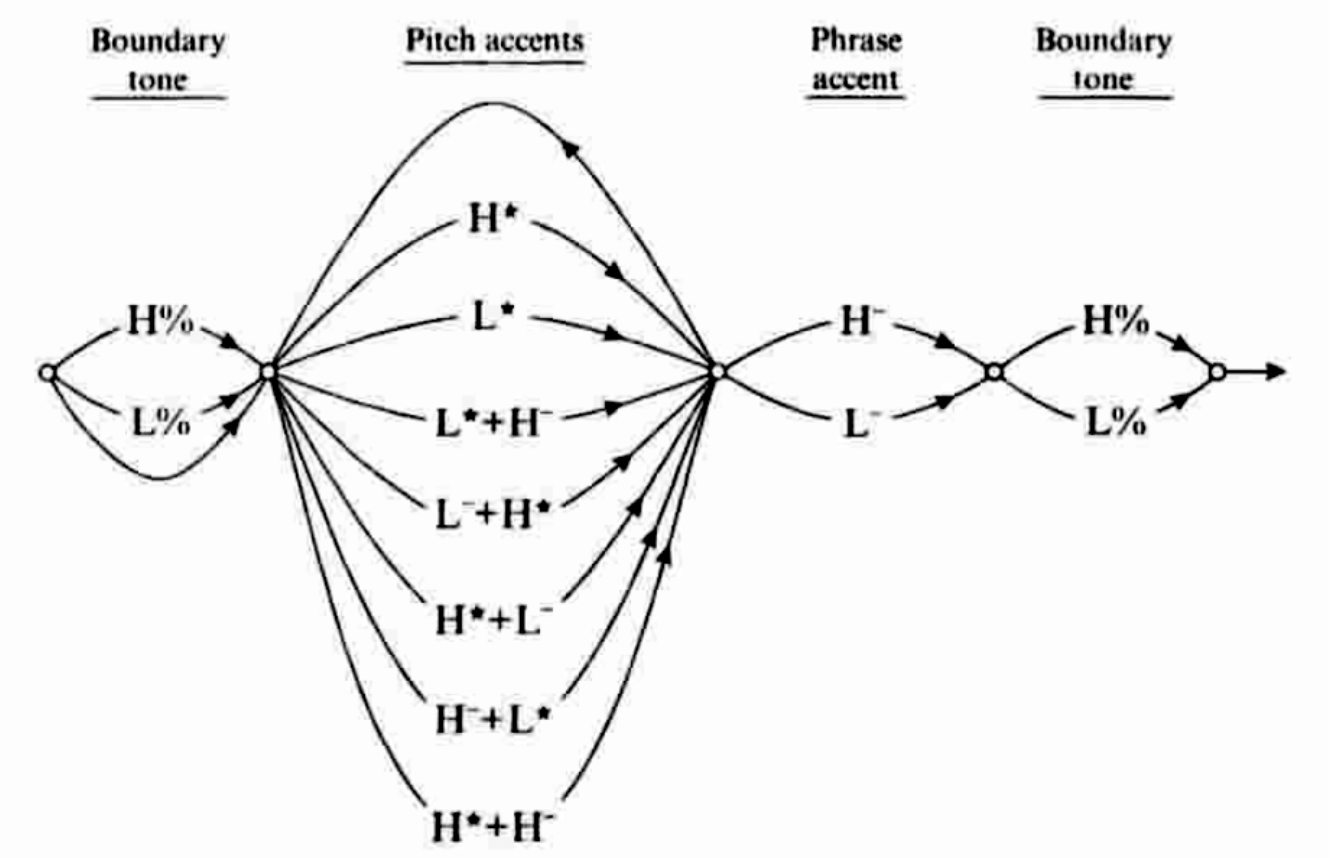
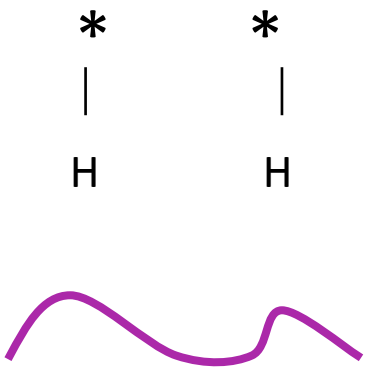


Figure 1. Pierrehumbert's intonational grammar (Ladd, 1996)

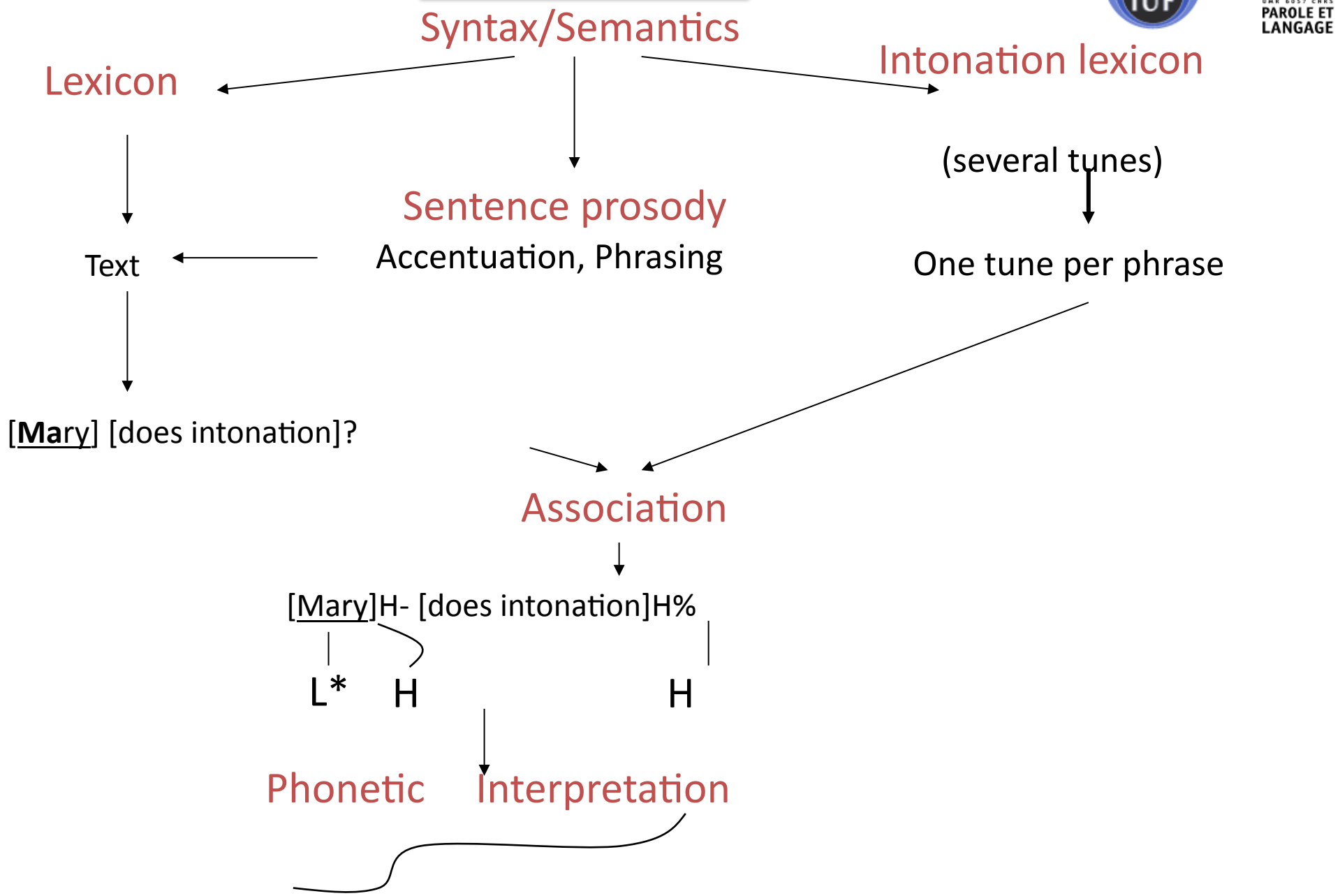
The Autosegmental-Metrical (AM) Model of Intonation: basic concepts

- The intonational pattern of an utterance rests on:
  - A metrical pattern (or **accentual scheme**)
  - A melodic profile (a "**tune**")
  - Rules associating those elements (**tune-text alignment**)



Phonetic implementation

AM model


Syntax/Semantics
Lexicon
Intonation lexicon
(several tunes)
Sentence prosody
Accentuation, Phrasing
Text
One tune per phrase
[Mary] [does intonation]?
Association
[Mary]H- [does intonation]H%
L*
H
H
Phonetic
Interpretation

Pierrehumbert 1980 (continued)

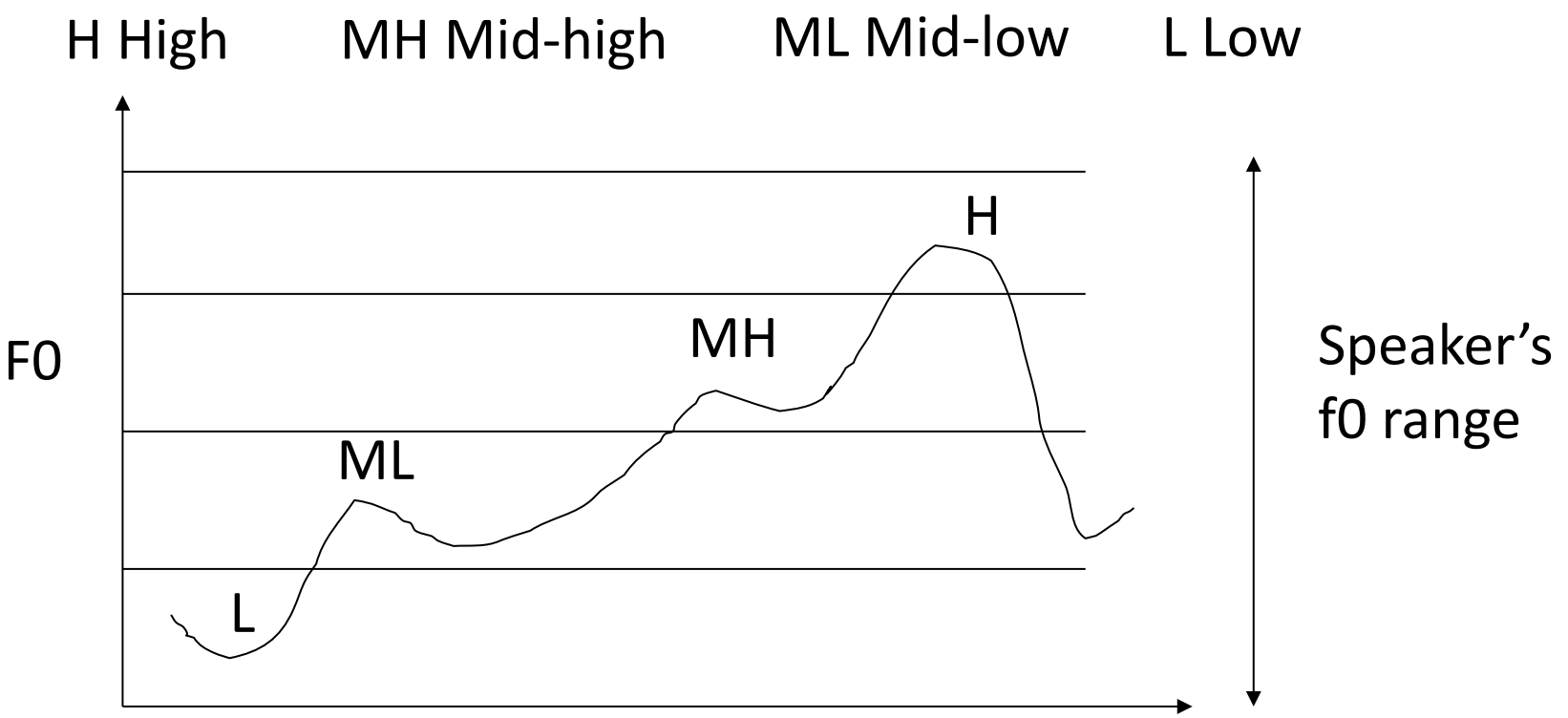
- Intonation patterns are made of sequences of one or more **pitch accents** plus **boundary tones**.
- Pitch accents are associated (then aligned) with **metrically strong syllables** (the stresses) while edge tones align with the edges of melodic units.
- Pitch accents and boundary tones can be adequately represented through **2 tone levels**: H(igh) and L(ow).
- **Phonetic implementation rules** map underlying tones into a continous melodic representation (the actual f_0 curve).

Pierrehumbert 1980 (continued)

- Intonation patterns are made of sequences of one or more **pitch accents** plus **boundary tones**.
- Pitch accents are associated (then aligned) with **metrically strong syllables** (the stresses) while edge tones align with the edges of melodic units.
- Pitch accents and boundary tones can be adequately represented through **2 tone levels**: H(igh) and L(ow).
- **Phonetic implementation rules** map underlying tones into a continous melodic representation (the actual f_0 curve).

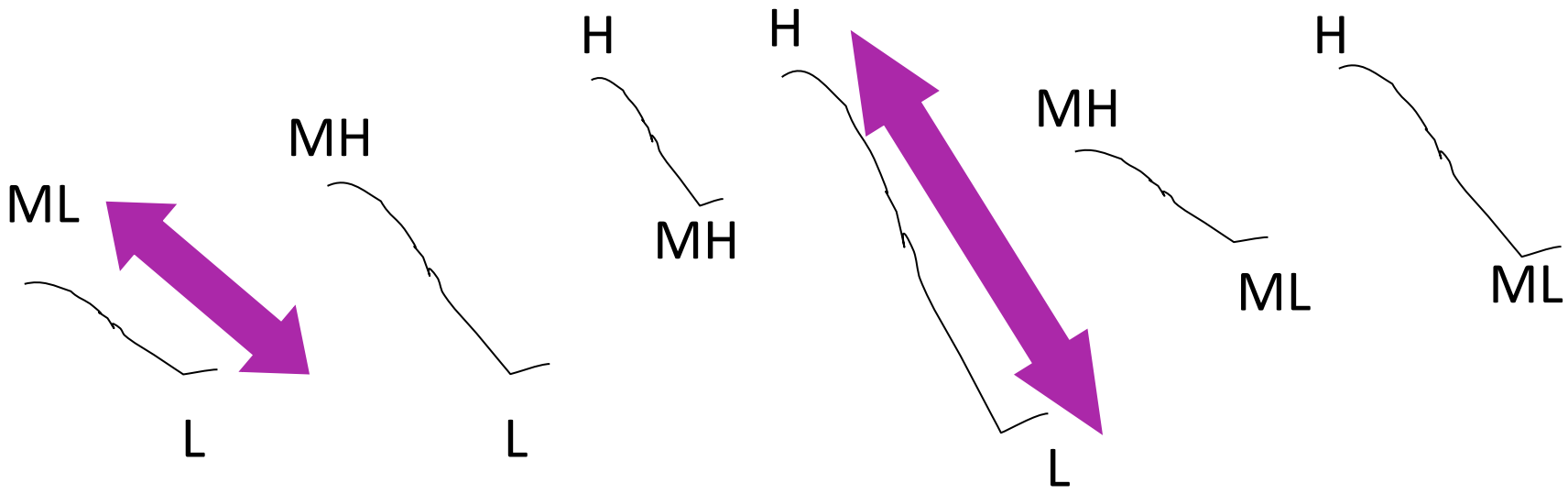
How many levels?

4 Tones in the earlier system of Trager & Smith (1951)



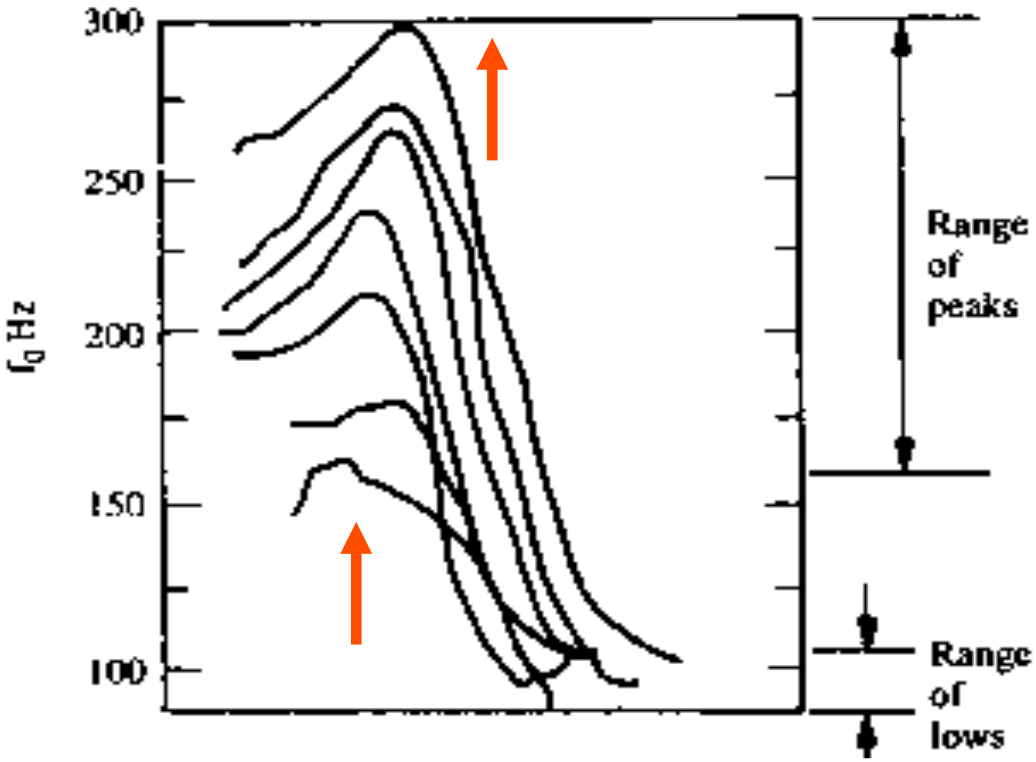
Problem with 4-tone model

1. Too many combinations: There should be 6 different, linguistically relevant, falling contours



There is however no evidence that the differences between e.g. H L and ML L are distinctive.

Pitch range



Degree of Emphasis: Free gradient hypothesis

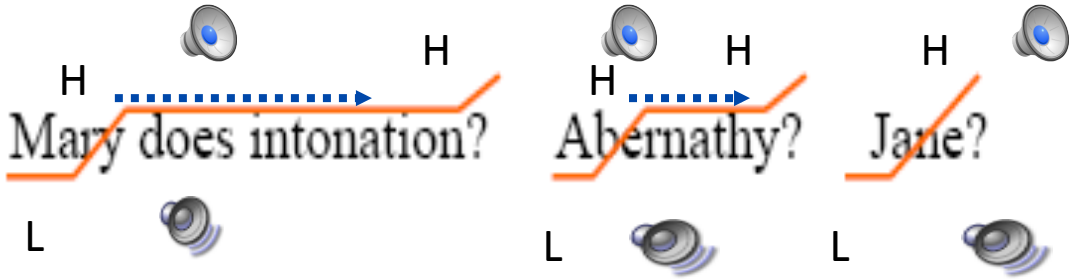
From Liberman & Pierrehumbert 1984

Pierrehumbert 1980 (continued)

- Intonation patterns are made of sequences of one or more **pitch accents** plus **boundary tones**.
- Pitch accents are associated (then aligned) with **metrically strong syllables** (the stresses) while edge tones align with the edges of melodic units.
- Pitch accents and boundary tones can be adequately represented through **2 tone levels**: H(igh) and L(ow).
- **Phonetic implementation rules** map underlying tones into a continous melodic representation (the actual f_0 curve).

Adapting tune to text: notion of target tones

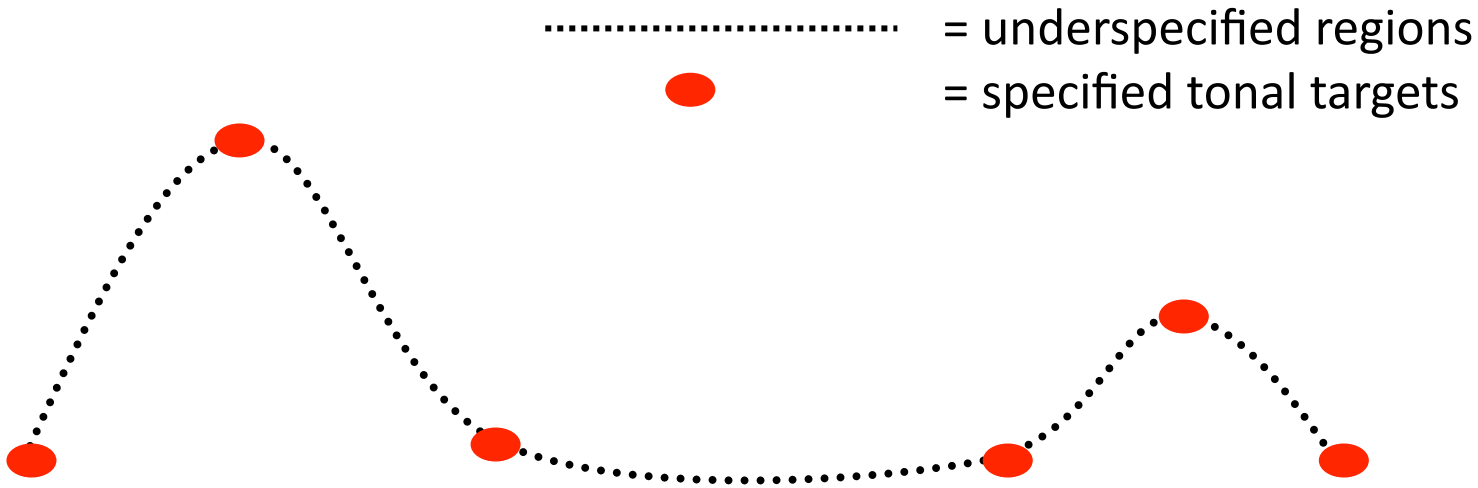
- By representing rises and falls as sequences of targets intonational phonologists could invoke other devices of Autosegmental Phonology



  - Analyzed as **L H H** sequence with second H **spread** over interval between a stressed syllable and phrase end

- Thus the patterns are only rendered equivalent by a representation which distinguishes the contour itself from the way the contour is aligned with the syllables.

Linear interpolation model



Why ToBI?

- To provide a **standardised transcription** tool which is (surface) transparent, and easily learnable
- For researchers from **different theoretical frameworks** working on different types of data (dialectal, speaking style, etc)

- 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: **English varieties!**

What is ToBI (Tones and Break Indices)?

- Framework for developing **community-wide conventions** for transcribing the intonation and prosodic structure of spoken utterances **in a language variety.**
- Should be **grounded in careful research on the intonation system** and the relationship between intonation and the prosodic structures of the language.

What ToBI is not

- **ToBI is not an International Phonetic Alphabet for prosody!**
  - There are many different ToBI systems, each one specific to a language variety and the community of researchers working on that language variety.
  - **each language variety requires a separate ToBI system!**
  - though some varieties can be transcribed using effectively the same or similar annotation criteria (e.g. Australian English, New Zealand English, General American English)

What does a complete ToBI system require?

- Requires commonly agreed conventions
- The conventions are as accurate as possible, based on a large and long-established body of research in **intonational phonology**
- The conventions do not replace a permanent record of the **speech signal** with a symbolic record.
- Conventions are efficient (only **distinctive information** is marked)
- Conventions are easy enough to teach and consistently maintened across transcription sites

What languages and dialects have ToBI systems?

Complete systems, with published standards, training materials, and intertranscriber consistency tests, have been developed for:

English covers Mainstream American, Southern British RP, Australian varieties	German standard variety
Japanese standard (Tokyo) variety	Korean standard (Seoul) variety

Nearly complete systems, with training materials but no published intertranscriber consistency tests as of yet, have been developed for:

Greek standard (Athens) variety

Systems are under development for:

Serbo-Croatian	Cantonese	English - Glasgow variety
Mandarin	Spanish	Taiwanese (Taiwan Min)

Annotation systems with training materials in a similar autosegmental framework for intonational tunes (but without the parallel annotation of prosodic structure) are available for:

Dutch

From: http://www.ling.ohio-state.edu/~tobi/#languages

What does a ToBI analyst need to have?

- Good ears!
- Acoustic waveform
- F0 contour
- Label tiers
  1. tone tier
  2. orthographic tier
  3. a break index tier
  4. a miscellaneous tier

The two main aspects of a ToBI system

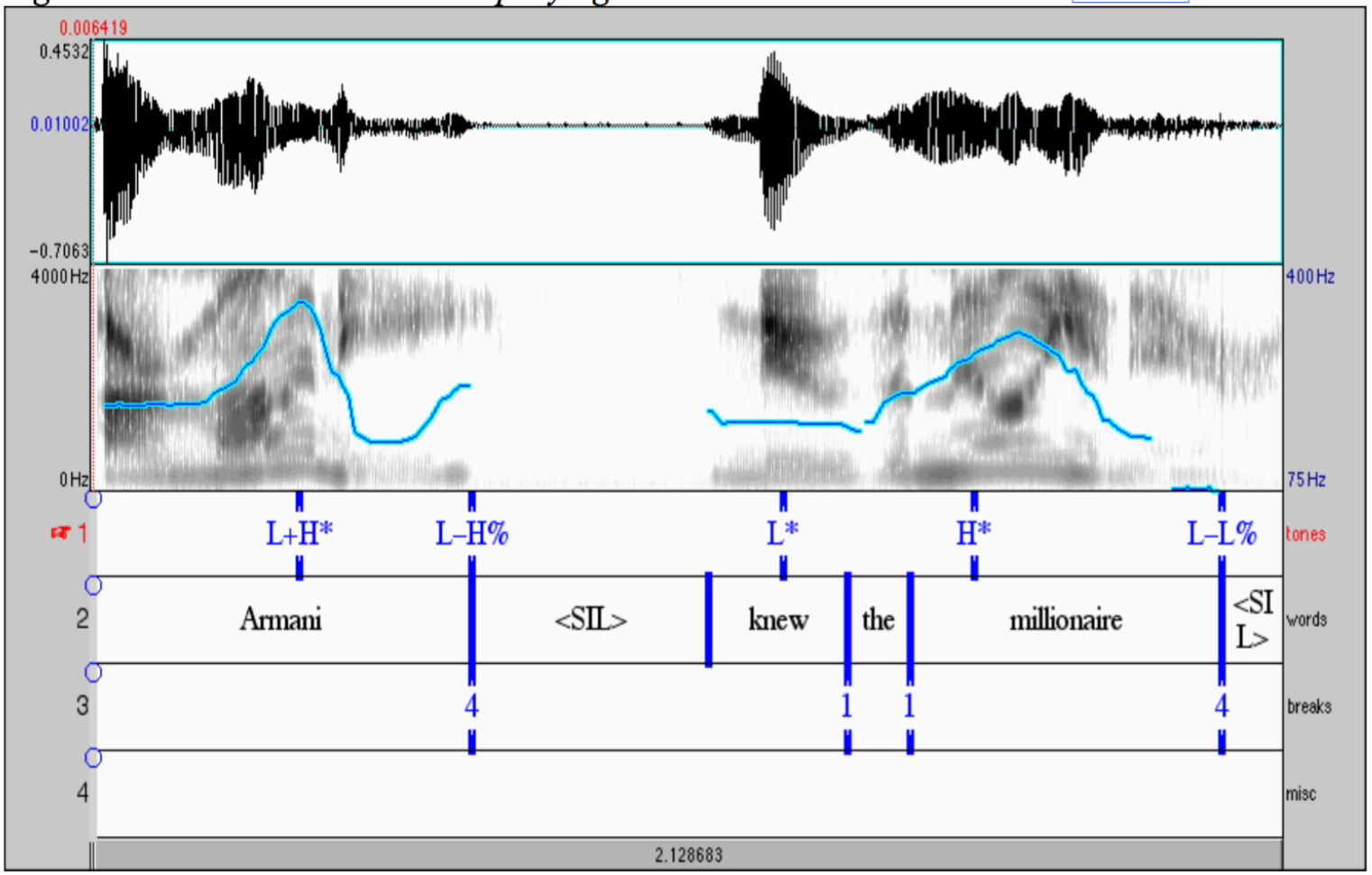
- Tonal tier
  - Pitch targets: the intonation or melody of an utterance is decomposed into component underlying H (high) and L (low) pitch targets
  - These signal two broad functions: **accentuation** and **phrasing**
  - Part of the transcription that **corresponds most closely to a phonological analysis** of the utterance's intonation pattern

The two main aspects of a ToBI system

- Break Index tier
  - Hierarchical prosodic constituency from the word up to the intonational phrase
  - Constituents at different levels are separated by different types of break (from 0 to 4)

An example of a ToBI transcription

Figure 1: armani1.wav with accompanying Praat TextGrid. <armani1>



From: Nanette Veilleux, Stefanie Shattuck-Hufnagel, and Alejna Brugos. *6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI, January IAP 2006*. (Massachusetts Institute of Technology: MIT OpenCourseWare), http://ocw.mit.edu (Accessed). License: Creative Commons BY-NC-SA

The anatomy of a tune

A tune is composed of:

- pitch accents:
- boundary tones: phrase accents and edge tones

Tone Types

- There are two types of tones at play:

1. Pitch Accents
    - associated with a stressed syllable
    - may be either High (H) or Low (L), or a combination of the two
    - marked with a *
2. Boundary Tones
    - appear at the end of a phrase
    - not associated with a particular syllable
    - may be either High (H) or Low (L)
    - marked with a % (or a '-', for the ip level)

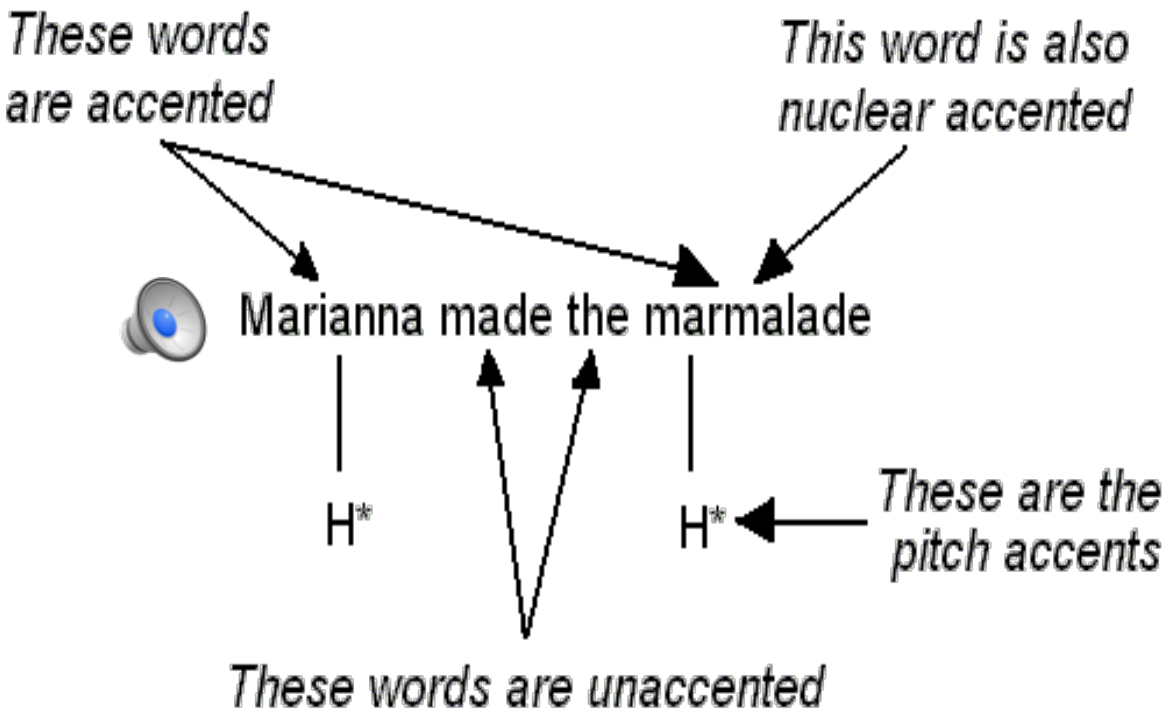
Tonal Association

The association of tune and prosodic phrase

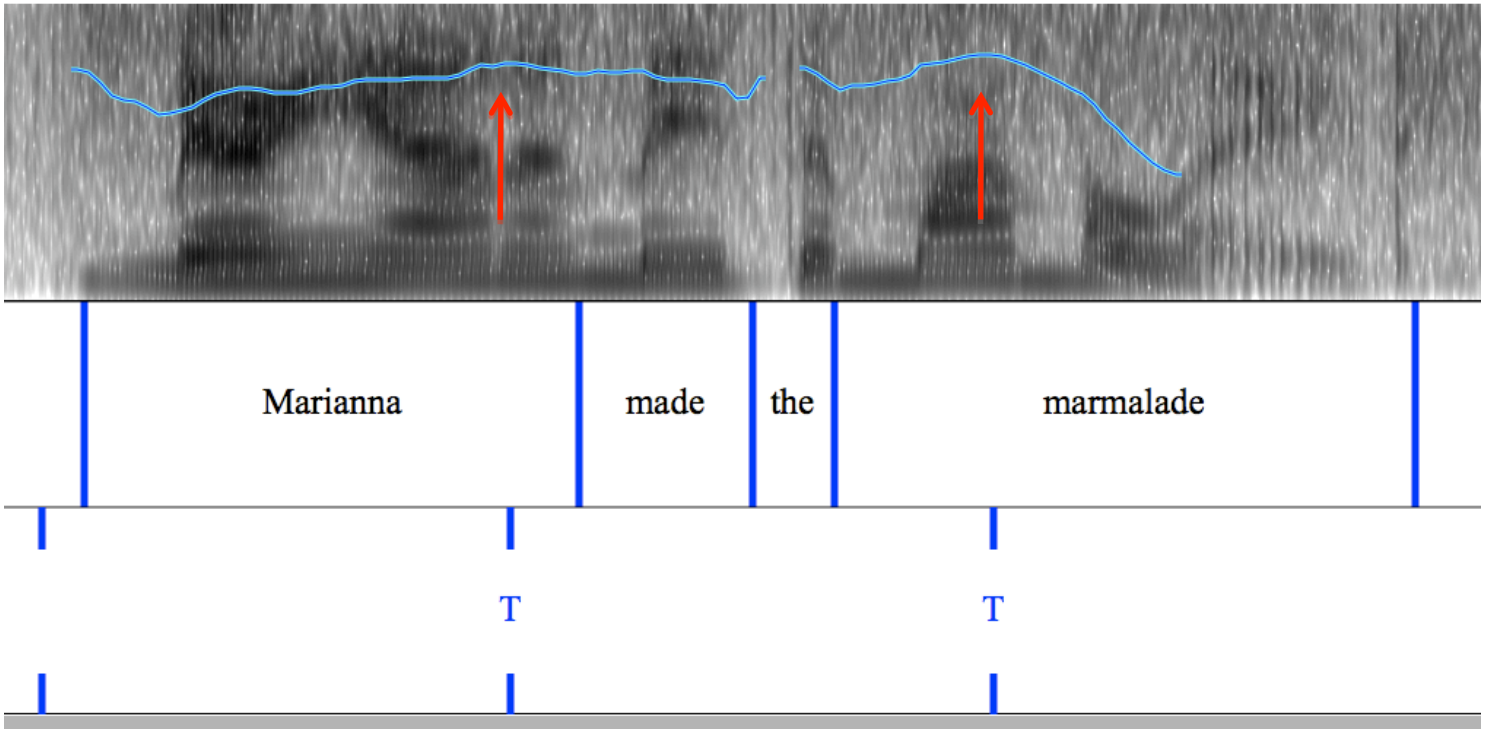
- each pitch accent is associated to each prominent/accented word
- each edge/boundary tone is aligned to the end of each prosodic phrase

How do we hear accented words?

- A speaker places a *pitch-accent* 'in the vicinity of' the *vowel of the primary stressed syllable* of each word that is accented.



Marianna made the **marmalade**



Nuclear Accent

- Every utterance consists of one or more prosodic phrases.
- Within the prosodic hierarchy we can separate prenuclear from nuclear accents
- The last accented word in an intermediate phrase (for MAE at least) is *nuclear accented*.
- In every prosodic phrase, there is one (and only one) nuclear accented word (while there can be more prenuclear ones).

Nuclear Accent and Information structure

- Note that there's a tendency to accent <u>new</u> information in the discourse.
- Nuclear accent can shift from last position: 4 different patterns for 4 different contexts:

H*

H*: Manny came with Anna.

H*

H*: Manny came with Anna.

L*

L*: Manny came with Anna?

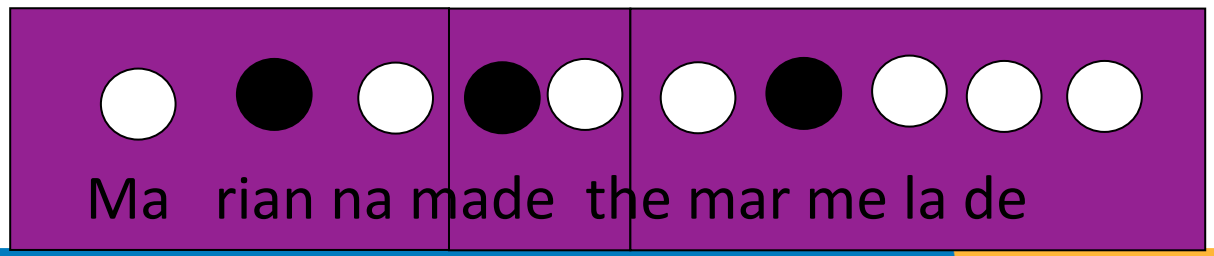
L*

L*: Manny came with Anna?

From: https://webdisk.ucalgary.ca/~swinters public_html/ling441/TOBIlab.html

Pitch accents and „nucleus“

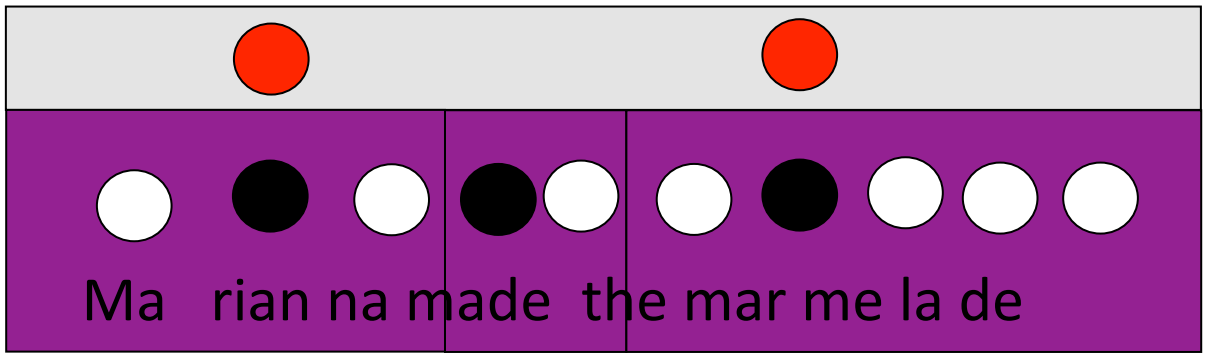
- First level of organisation provided by stress
- Location of stressed syllable specified in the lexicon



Pitch accents and „nucleus“

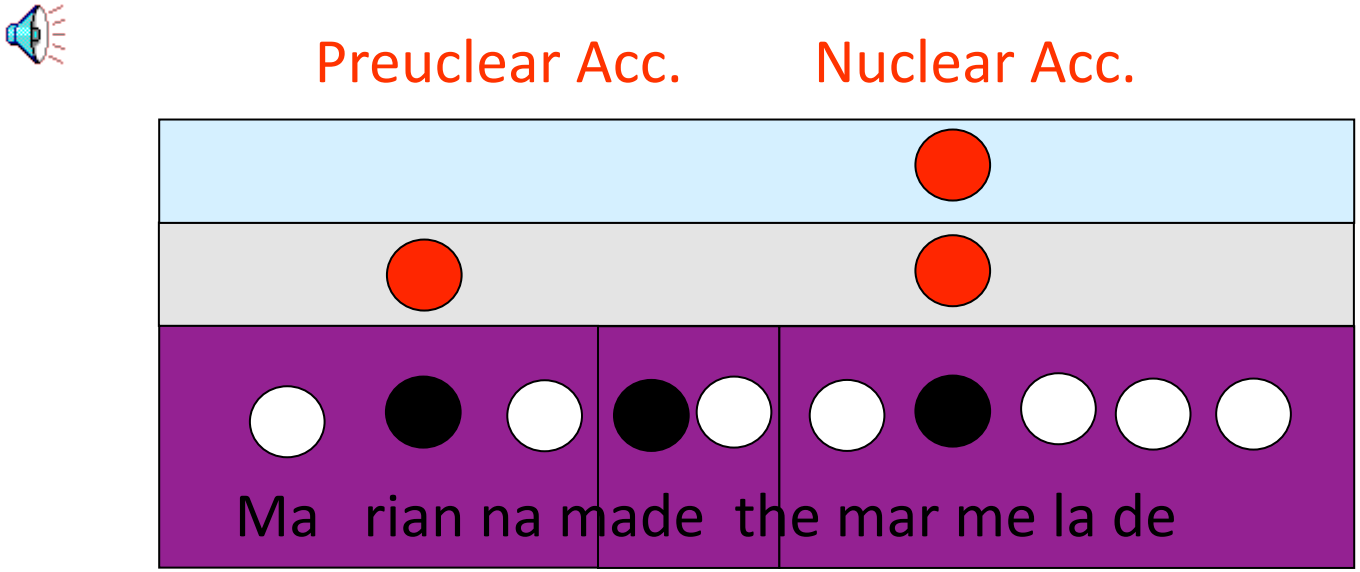
- Second level of organisation: Accent

- Location: stressed syllables can but do not have to be accented; Phonetic correlate: Pitch movement

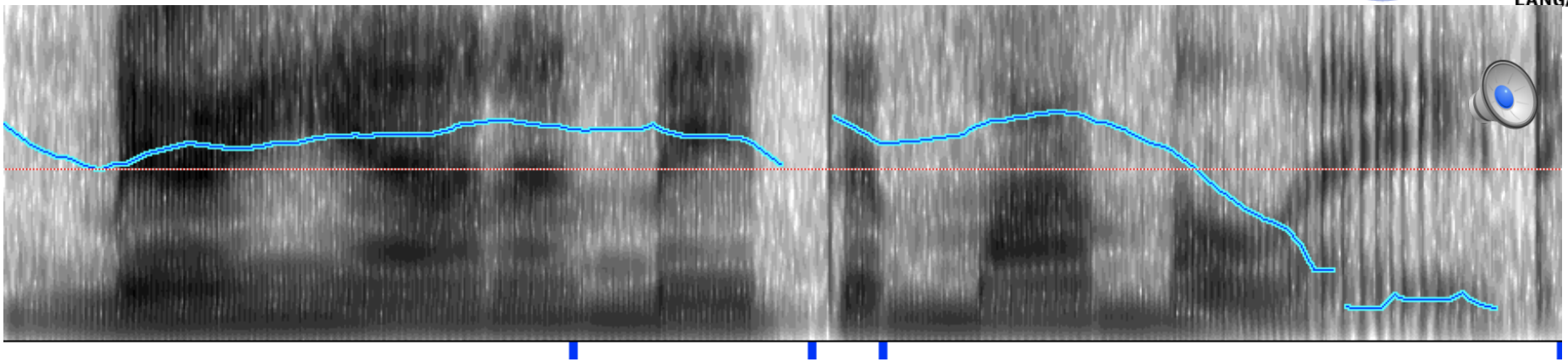


Pitch accents and „nucleus“

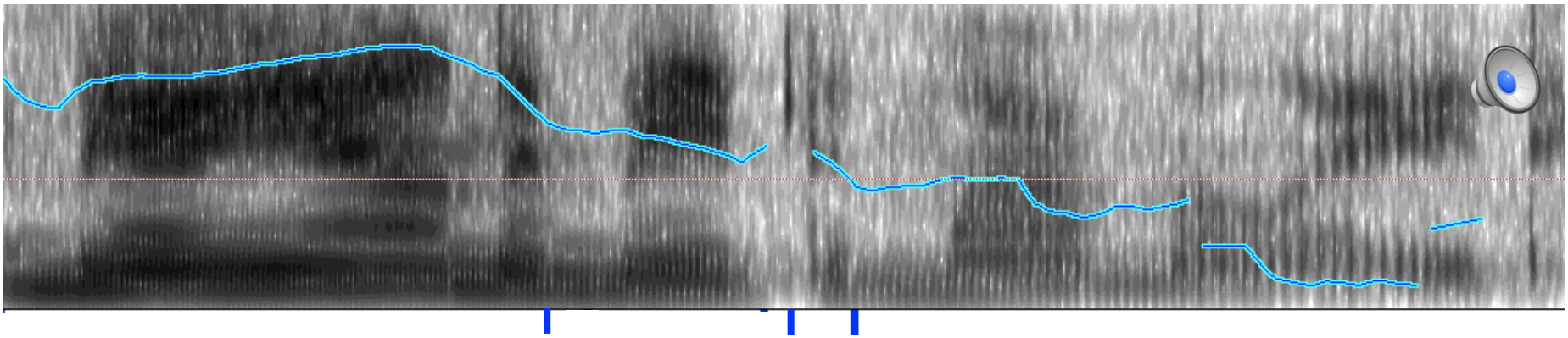
- Third level of organisation: Nuclear accent
- The most prominent accent; in English, it tends to be in the rightmost accent in an intonation group



Examples

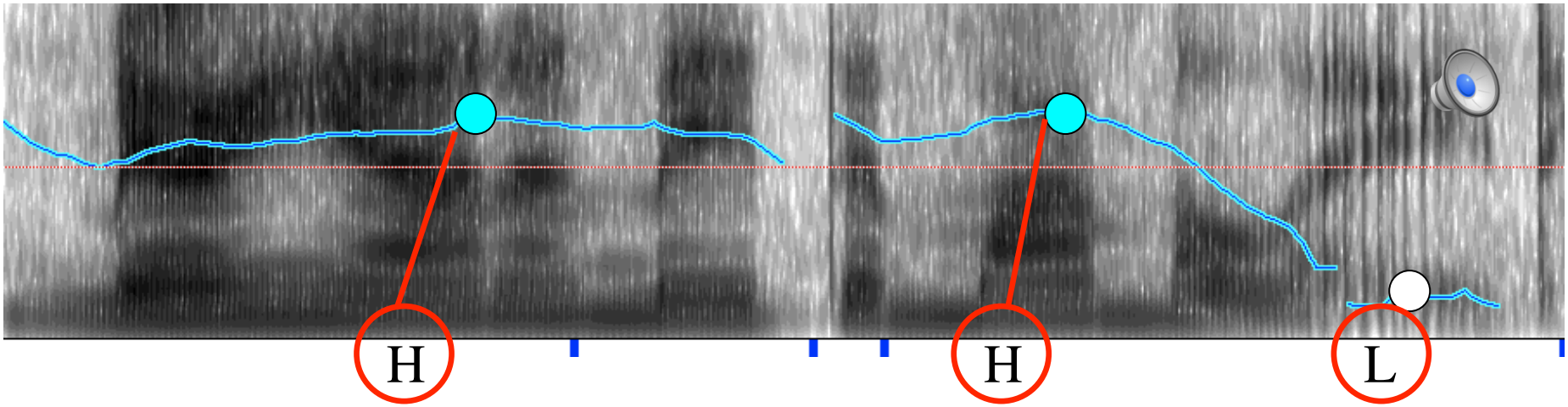


Marianna made the marmalade

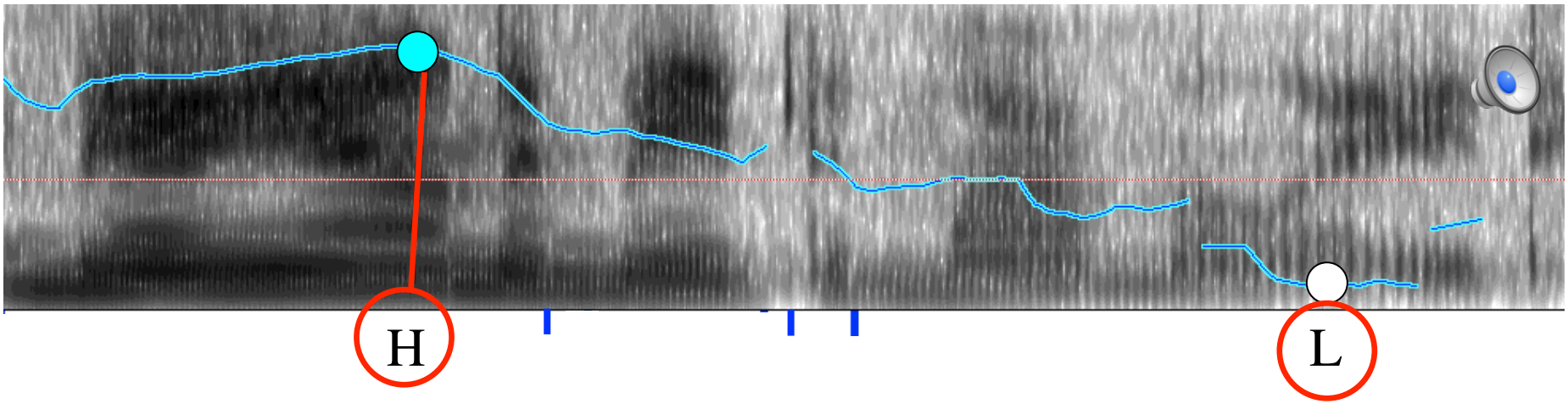


Marianna made the marmalade

Examples

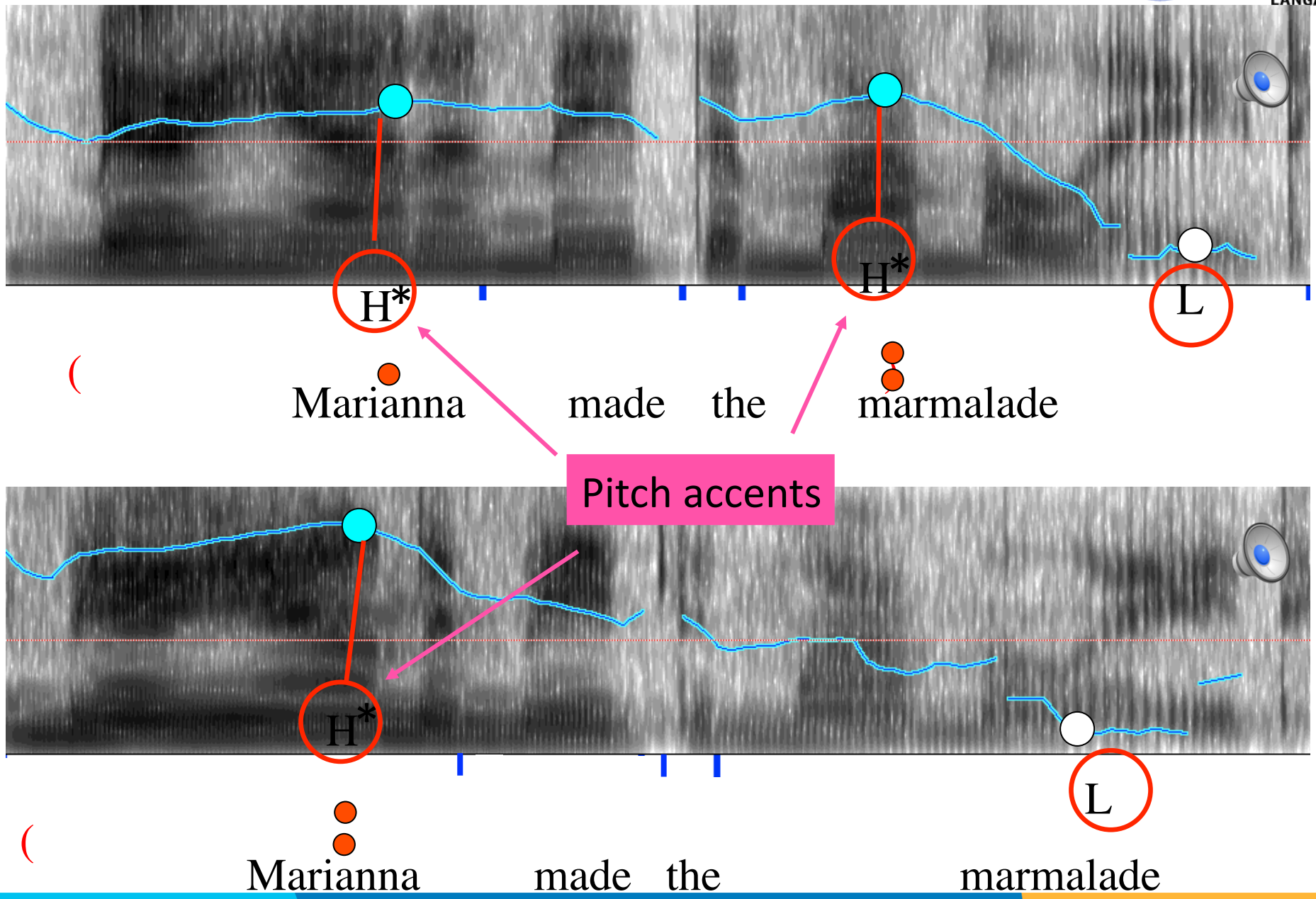


Marianna made the marmalade

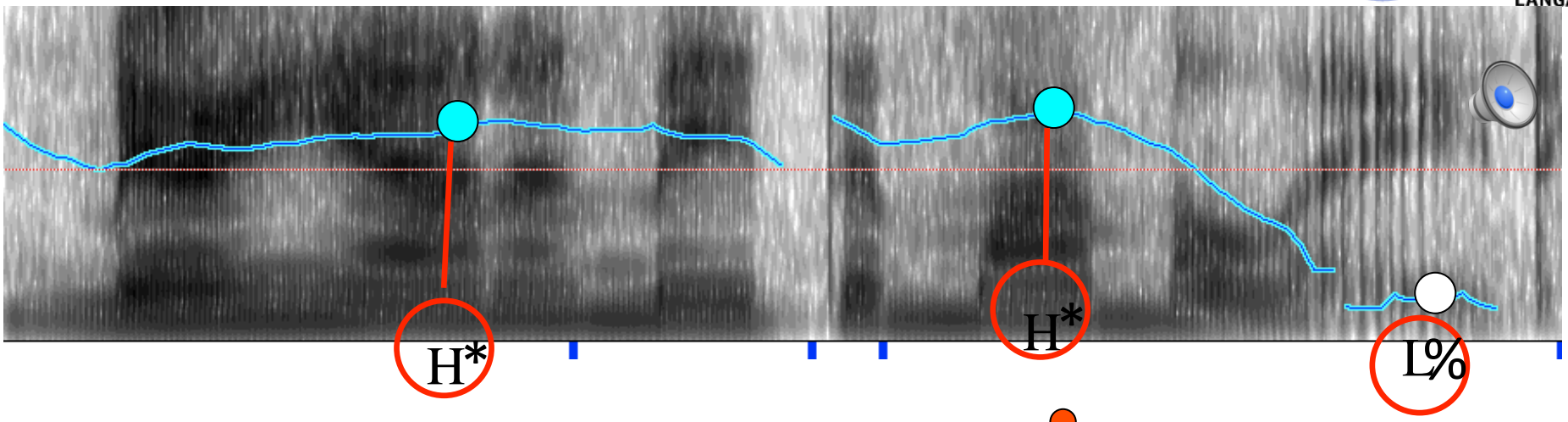


Marianna made the marmalade

Examples

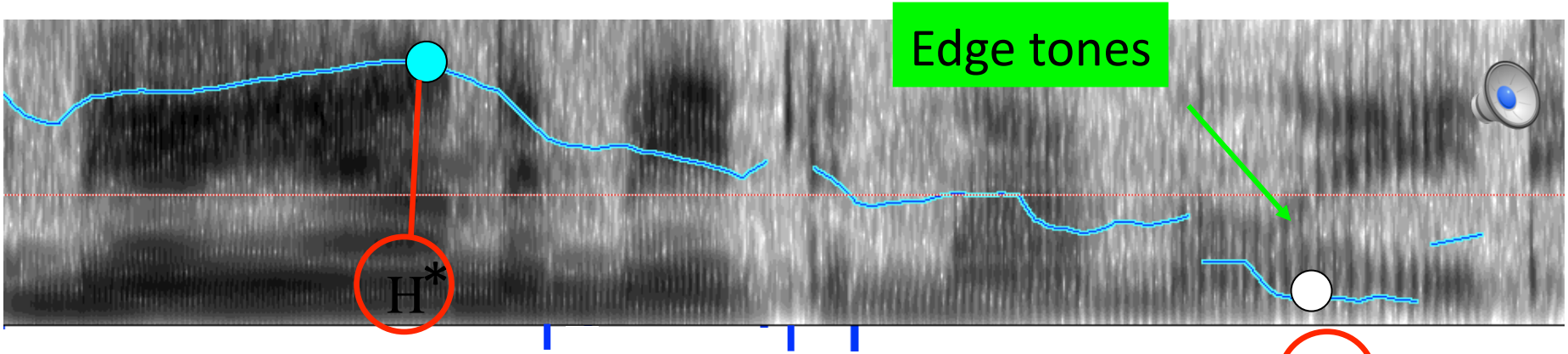

H*
H*
L
Marianna made the marmalade
Pitch accents
H*
L
Marianna made the marmalade

Examples



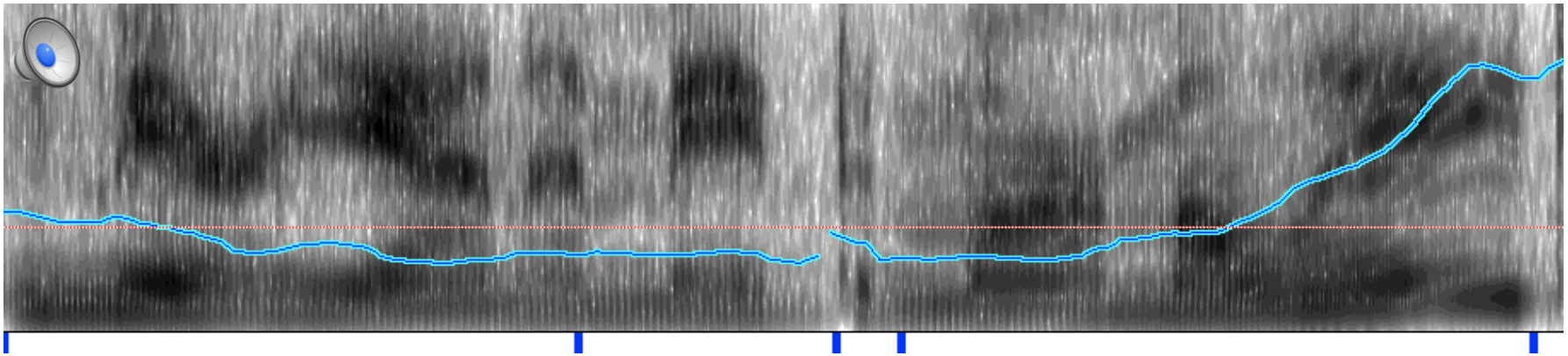
Marianna made the marmalade

Edge tones

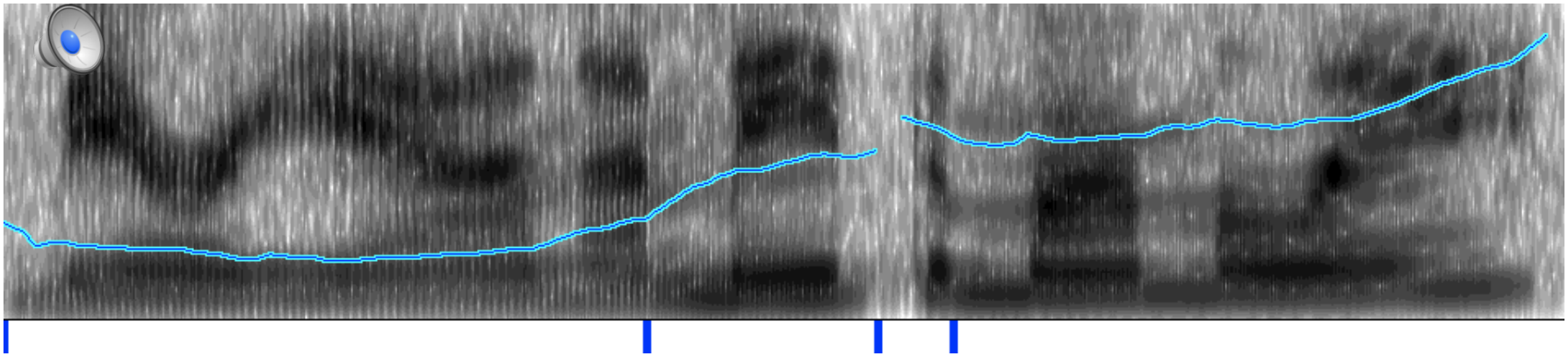


Marianna made the marmalade

Examples

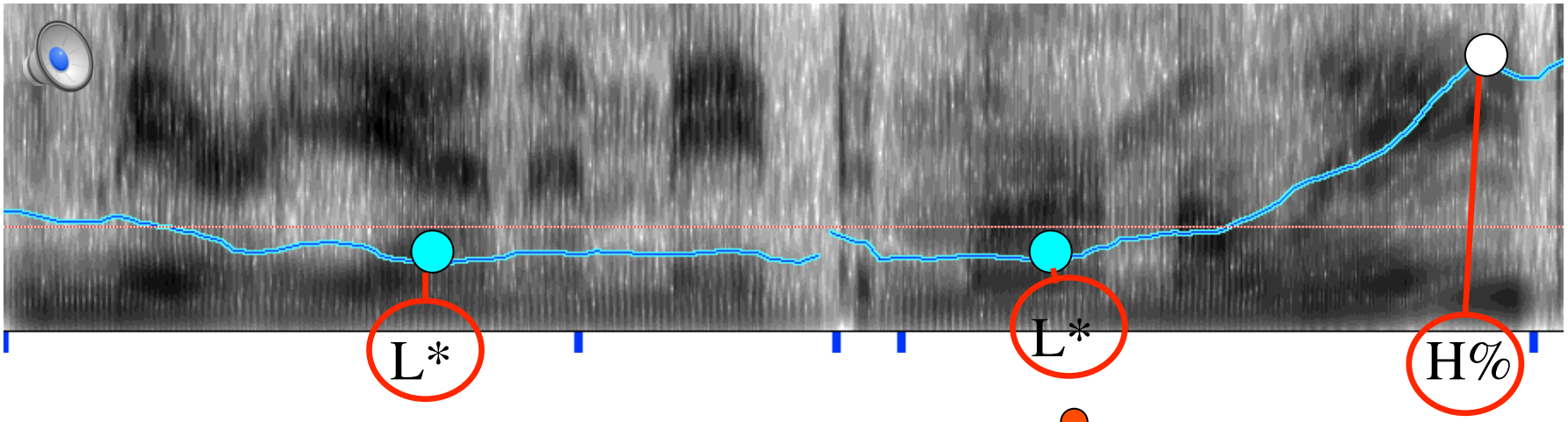


Marianna made the marmalade



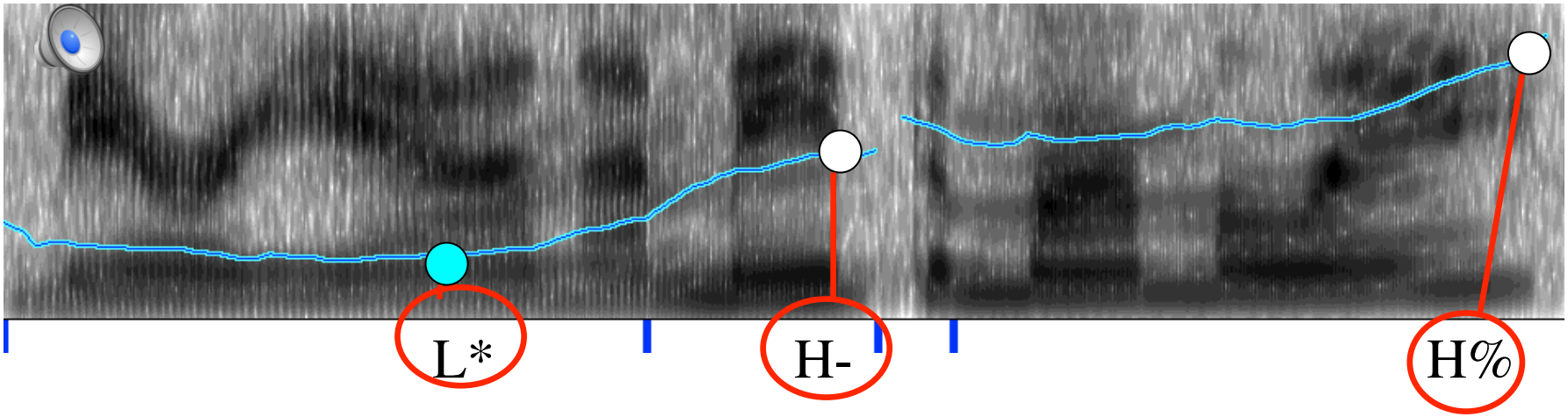
Marianna made the marmalade

Examples



( ) :

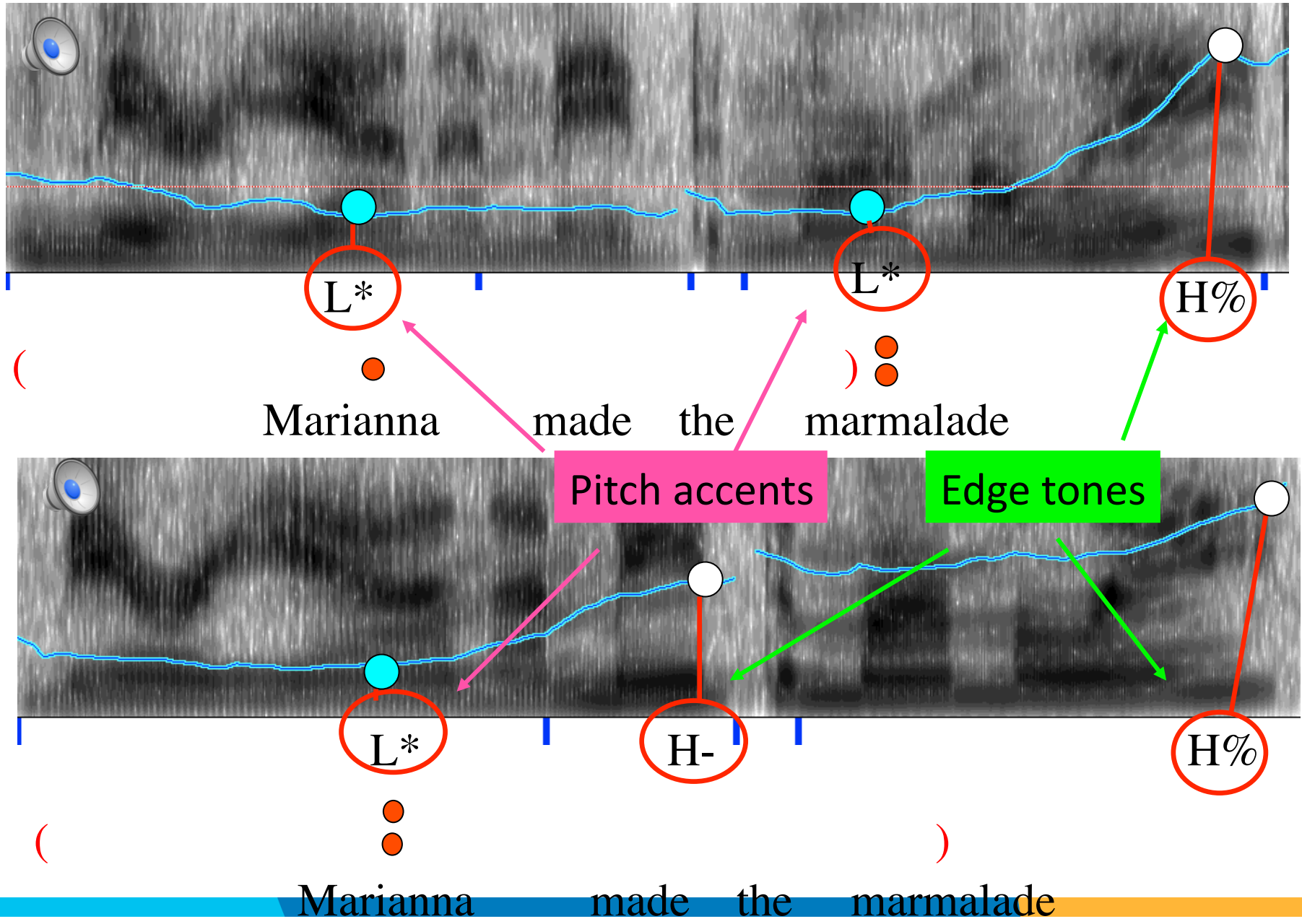
Marianna made the marmalade



( )

Marianna made the marmalade

Examples


L*
L*
H%
(
)
Marianna made the marmalade
Pitch accents
Edge tones
L*
H-
H%
(
)
Marianna made the marmalade

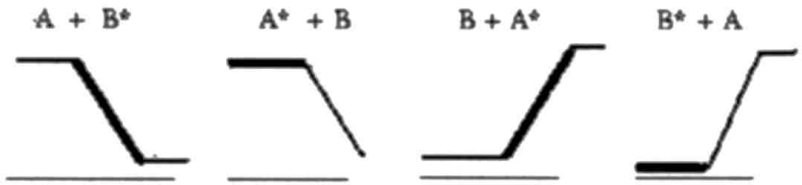
Pitch accents

- Can be **SIMPLE** or complex (**BITONAL**)

H* L+ H*

L* L*+ H

L+ H*



Star expresses the alignment contrast

Pitch Accent Types

- In English, pitch accents can be either high or low
  - H* or L*
- Examples: **High (H*)** **Low (L*)**

Yes. Yes?

H* L*

Magnification. Magnification?

- As with tones in tone languages, “high” and “low” pitch accents are defined relative to a speaker’s pitch range.
  - Steve Winter’s pitch range: H* = 155 Hz L* = 100 Hz
  - Mary Beckman: H* = 260 Hz L* = 130 Hz

From: https://webdisk.ucalgary.ca/~swinters public_html/ling441/TOBIlab.html

Inventories of pitch accents

P80 7 pitch accents	PB86 6 pitch accents	ToBI 5 pitch accents + downstep (!)
H*	H*	H*
L*	L*	L*
H+L*	H+L*	H+!H*
H*+L	H*+L	(H* followed by downstep)
L*+H	L*+H	L*+H
L+H*	L+H*	L+H*
H*+H		

Phrase accents: H-, L- Boundary tones: H%, L%

Monotonal accents

- H*
  There should be a pitch peak on, or near, **the accented word's primary stressed vowel** (if preceding consonant is voiced, e.g. ‘bead’)

- L*
  There should be a pitch trough, on, or near, **the accented word's primary stressed vowel.**

System choices and variability

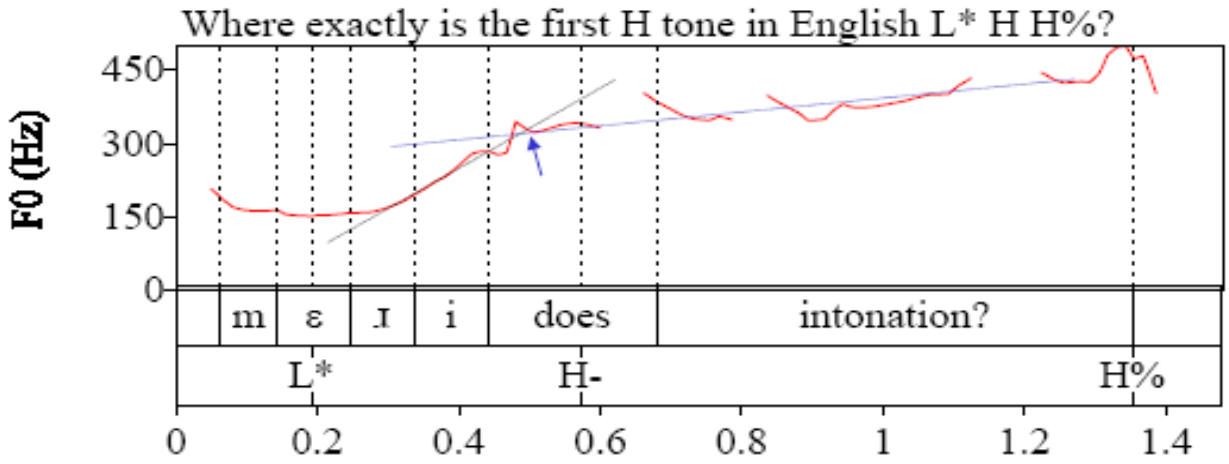
- No complex tonal movement on unstressed syllables and at the end of domains
- Tonal patterns do **adapt to the available segmental material**
- For the H vs. L transcription, one must take into account adjacent tonal values, speaker's pitch range and MANY other variables

Phonetic differences in (prenuclear) H alignment across languages

	Accent type	H Alignment (segment & syllable offset)
Greek (Arvani et al. 1995)	H*	C1
Dutch (Ladd et al. 1998)	H*	S0
Mandarin (Xu, 1998)	Rising rising	S0
English (Ladd et al. 1999)	H*	~C1
Japanese (Ishihara, 2003)	Initial H in accentual phrase	C1 (CV.CV); V0 (CVN.CV); → 2nd mora onset
Spanish	LH*	V0

Segmental and syllabic anchoring of trailing tones across different languages in contexts without tonal coarticulation. Segments: C=consonant, V=vowel, N=coda, S=syllable; Structure: 0=accented syllable, 1=postaccentual syllable (ex: C0V0N0.C1V1).

H tones can be elbows too….



- Arvaniti, Ladd, & Mennen (1998) for Greek L*+H vs. L+H*, etc.
- D'Imperio (2000) for Neapolitan Italian L+H* vs. L*+H
- Dilley (2003) for English H+L* vs. H* L-
- Willis (2003) for focused vs. neutral prenuclear accents in Dominican Spanish
- Welby (2003) for Hexagonal French early vs. late rise

Phonetic alignment and coding

- **Some pitch accent contrasts:**
  - **Dutch** H*, L*, H*+L, L*+H, L* H+L (ToDI system)
  - **Swedish** H*L, HL* (Bruce 1991)
  - **Neapolitan Italian** H+L*, L*+H, L+H*, H* (D'Imperio, 2002)
  - **English** H*, L*, L+H*, L*+H, H+!H* (ToBI)

- Note that most of these **differ in details of alignment** and not just in choice of tone sequence
- Reminiscent of **different VOT** values for same voicing feature (-/+voice, +/-aspirated) in the languages of the world

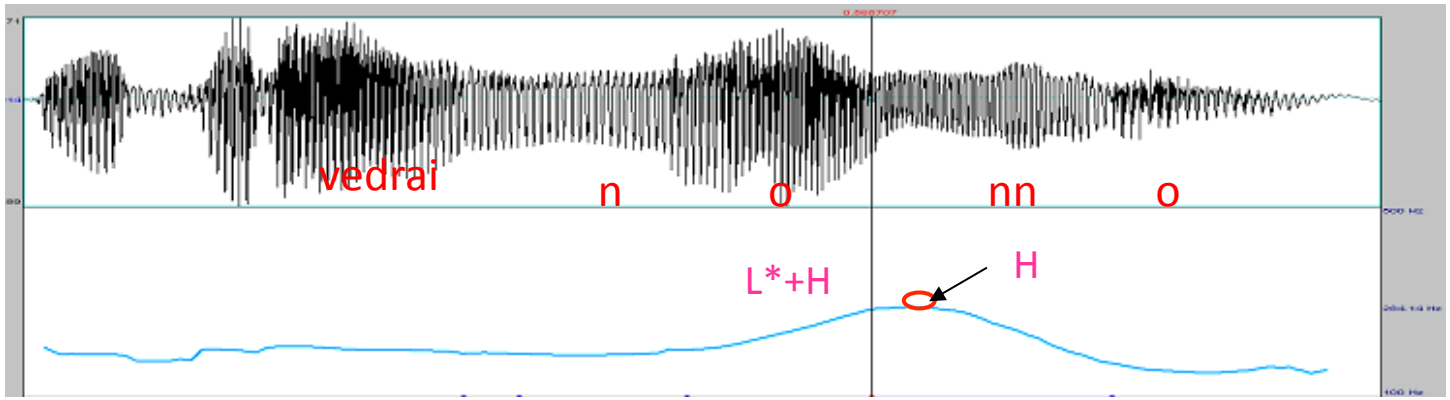
Regularities in tonal alignment

H alignment is quite variable: peaks can variably be positioned within, at the offset of the stressed syllable or even at the onset of the following unstressed syllable

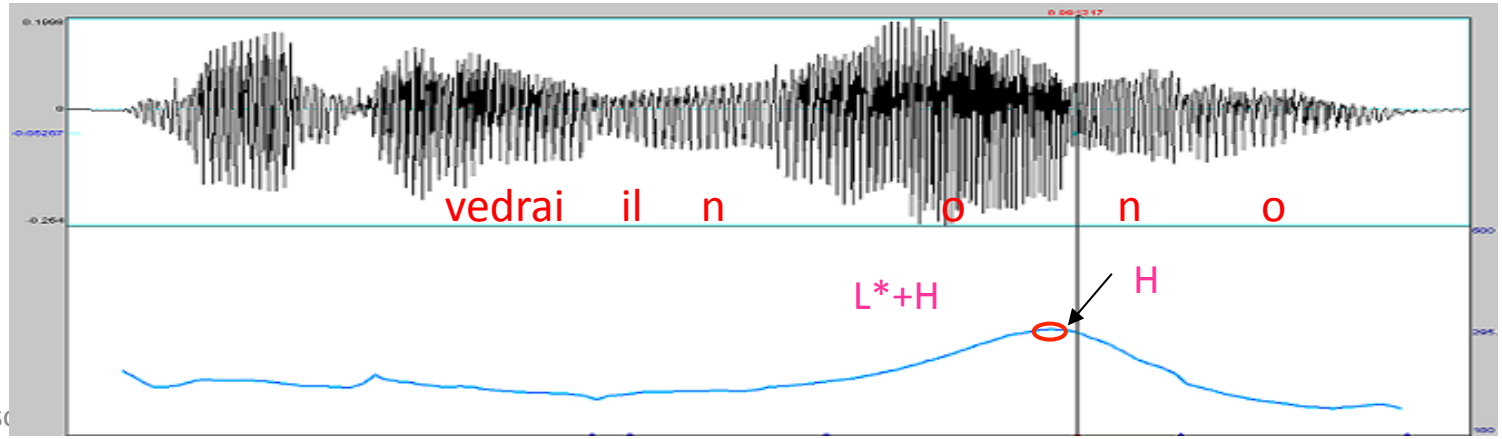
(Silverman & Pierrehumbert 1990 for English; Prieto, van Santen & Hirschberg 1995 for Spanish; Arvaniti, Ladd & Mennen, 1998 for Greek; D'Imperio, 2000 for Italian; Prieto 2005 for Catalan).

Alignment and syllable structure (D'Imperio et al. 2000, 2007)

A. Nonno « grandfather »



B. Nono « ninth »

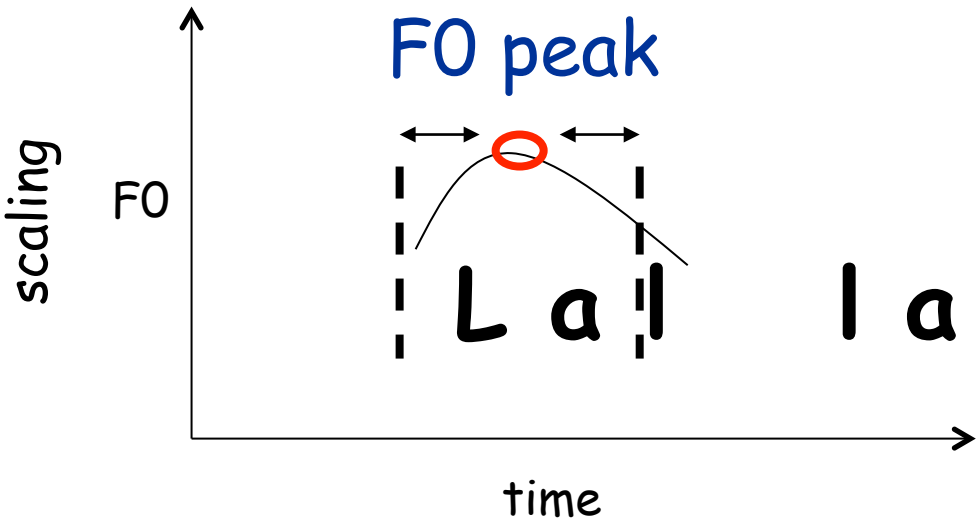


Tonal alignment

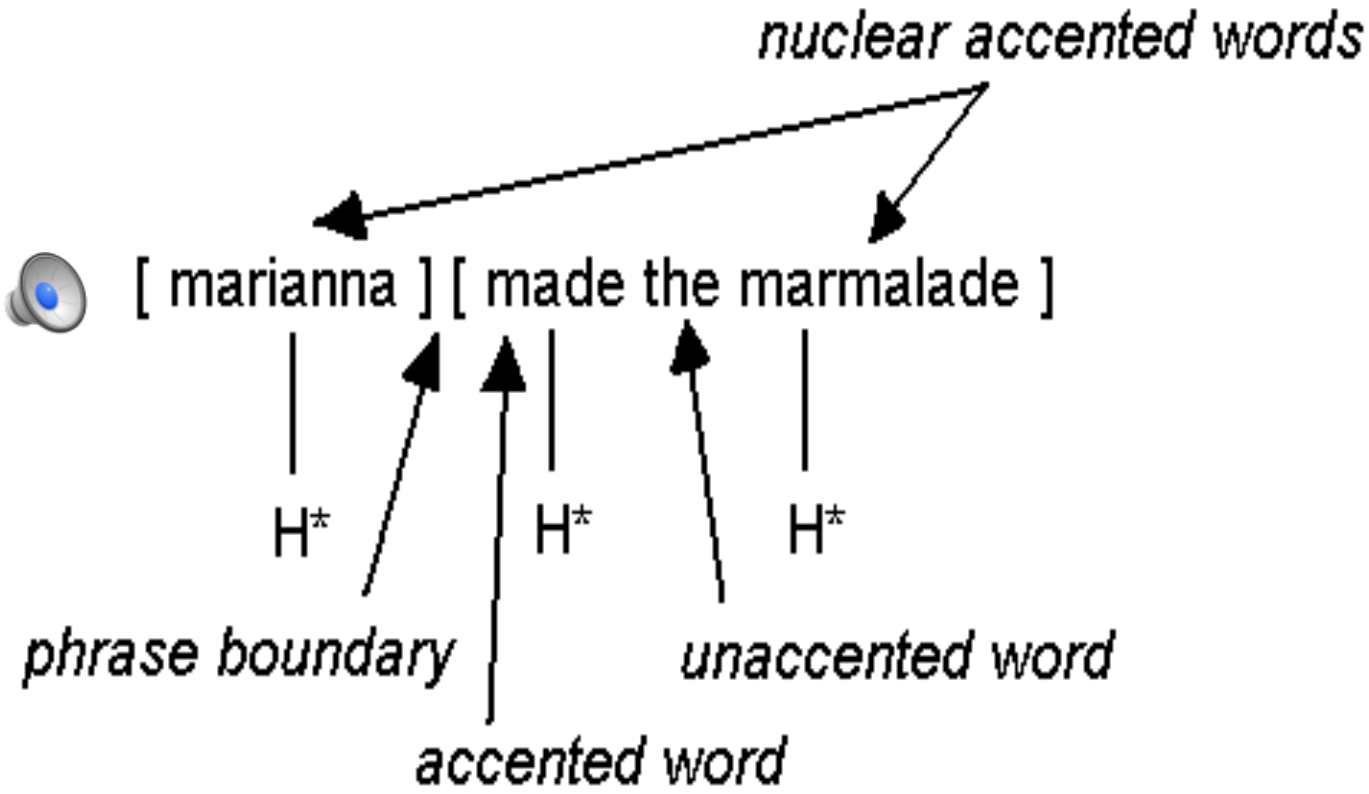
Alignment = synchronization of a target tone with specific segments or segmental boundaries in the string

(Silverman and Pierrehumbert, 1990; Caspers, 1994; Prieto *et al*., 1995; Rietveld and Gussenhoven, 1995; D'Imperio, 1995, 1997, 2001; Arvaniti *et al*., 1998; Ladd *et al*., 1999, 2000; Xu, 1998, 2002.)

Scaling = F0 value of the target.



Prosodic phrases and nuclear accent



Monotonal accents

A. Falling (H* L-L%)

H*

[(M e l a n i e)L] L%

B. Fall-rise (H* L- with H%)

H*

[(M e l a n i e)L-]H%

C. level (H* H- with L%)

H*

[(M e l a n i e)]H-L%

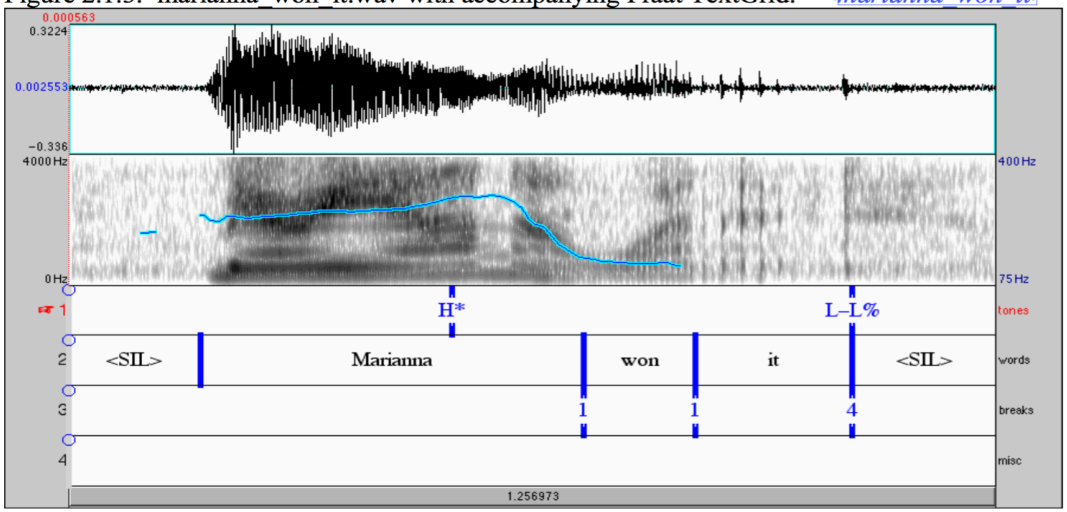
D. Rising (L* H- with H%)

L*

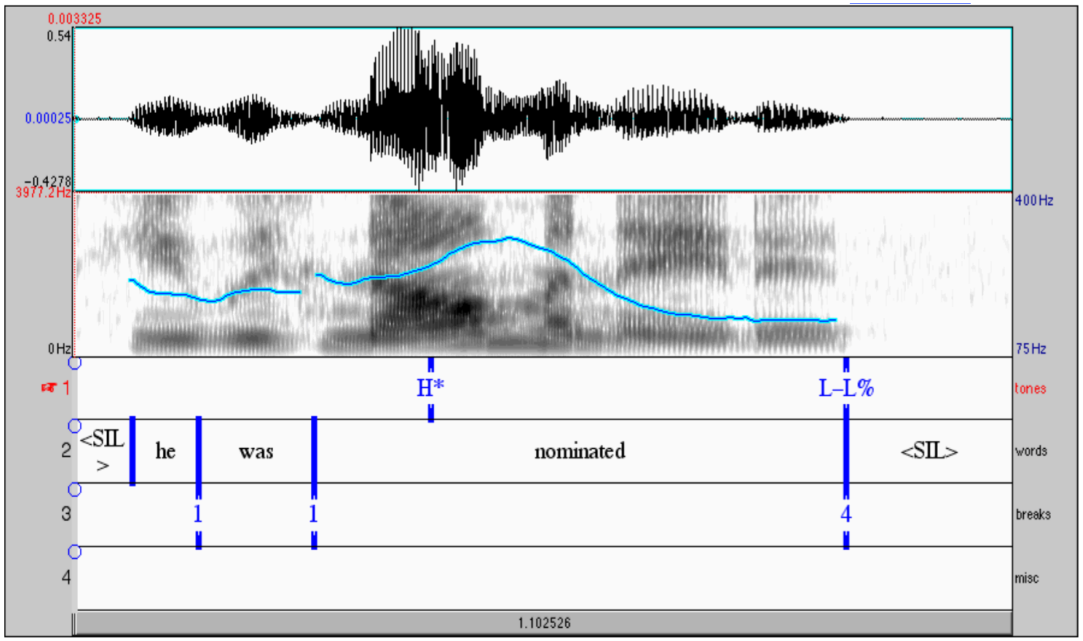
[(Melanie)] H%

EXAMPLES OF L* TONES

Figure 2.1.3: marianna_won_it.wav with accompanying Praat TextGrid. <*marianna_won_it*>



(slow) Rise to a peak

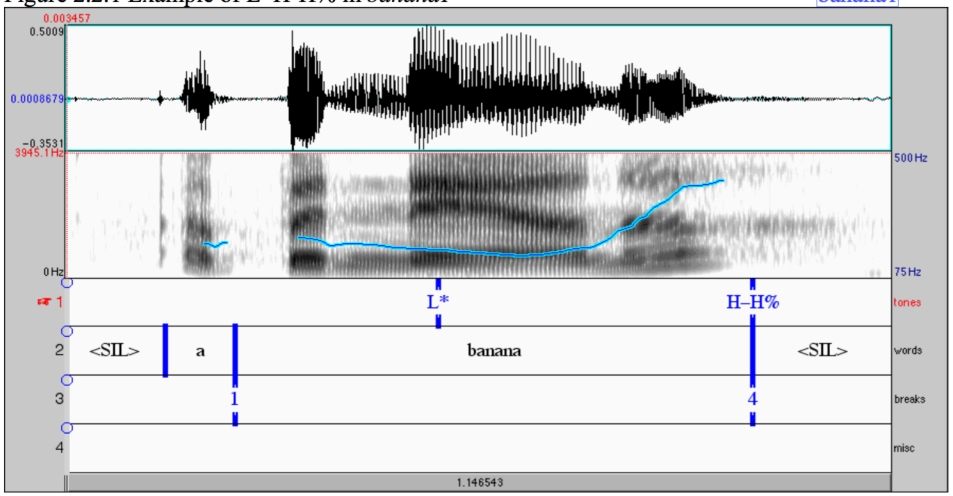


(fast) Rise to a peak

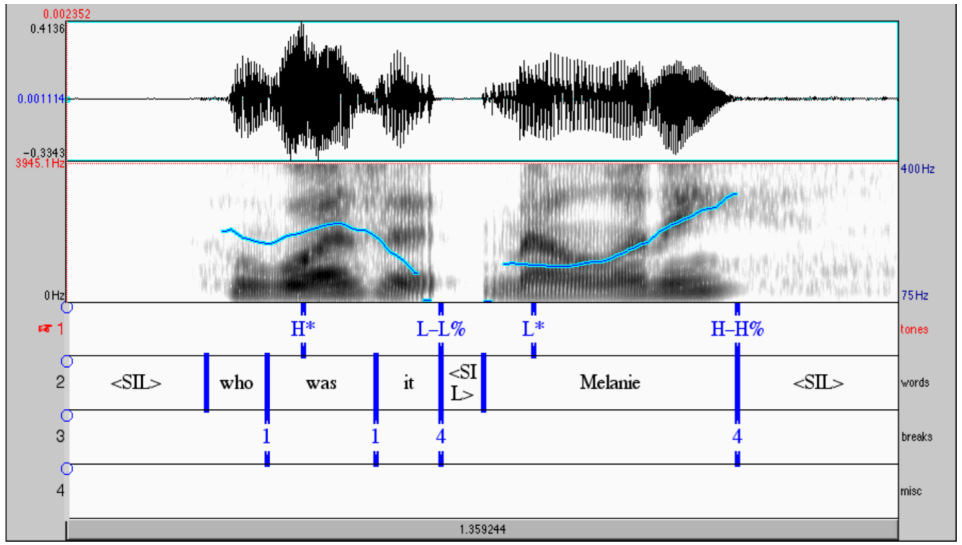
From Veilleux et al 2006

EXAMPLES OF L* TONES

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



F0 local dip or minimum



F0 'elbow'

From Veilleux et al 2006

Bitonal Pitch Accents

- In addition to H* and L*, there are two bitonal pitch accents
  - L + H* (L is leading)
  - L* + H (H is trailing)
- The starred element denotes the tone which is associated with the stressed syllable
- L + H* = high peak on stressed syllable, **preceded** by a sharp rise in pitch
- L* + H = low pitch target on stressed syllable, **followed** by a sharp rise in pitch

From: https://webdisk.ucalgary.ca/~swinters public_html/ling441/TOBIlab.html

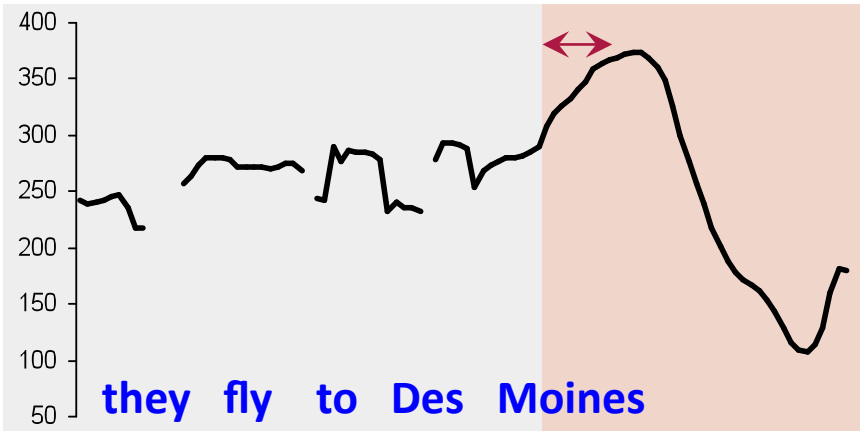
Bitonal Accents

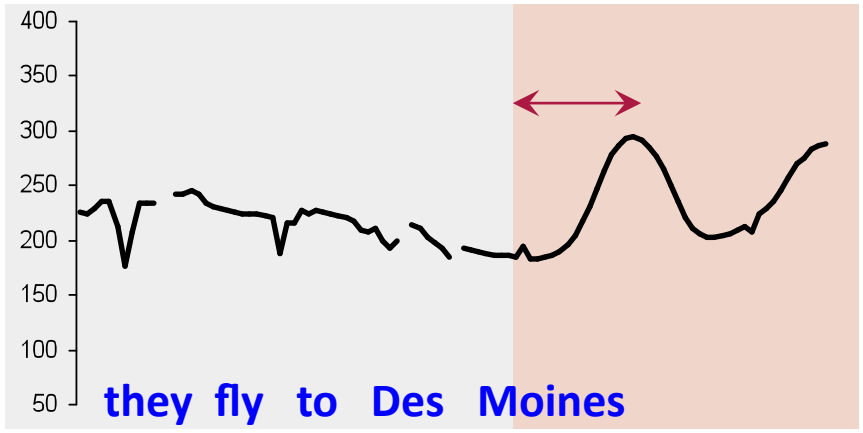
- L+H*
  - The H* tone target is preceded by a rise from a low part of the speaker's range
  - "I assert this"
- L*+H
  - A more "emphatic" accent
  - a L* tone target is followed by a rise to the mid to upper pitch range – sounds like a "scooped" accent.
  - relatively rare compared to H* or L+H* accents.

Alignment differences and pragmatic contrast

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

B: TWA doesn't fly there ...





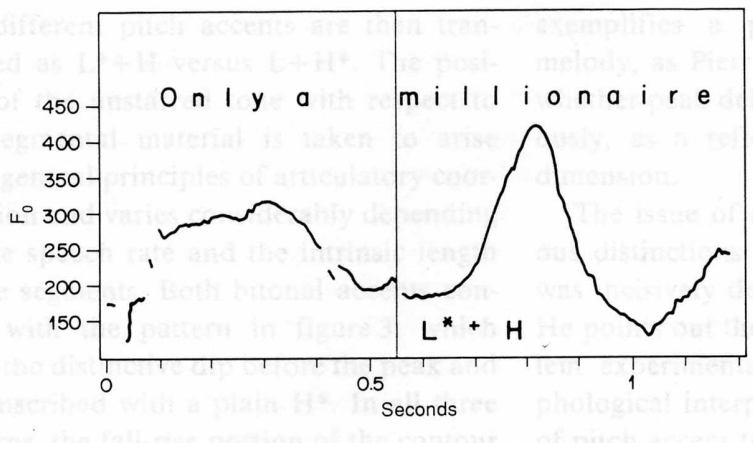
Rise right at start of stressed syllable cues assertion, statement of fact.

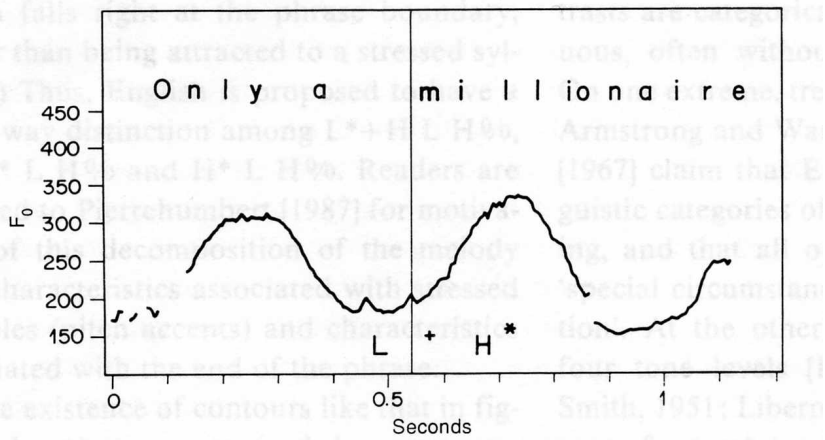
Rise which is delayed somewhat cues suggestion, or uncertainty about whether the statement qualifies as relevant.

(from Venditti 2002)

English

Standard analyses of English rising accents (Pierrehumbert 1980, ToBI):





Criteria for starredness

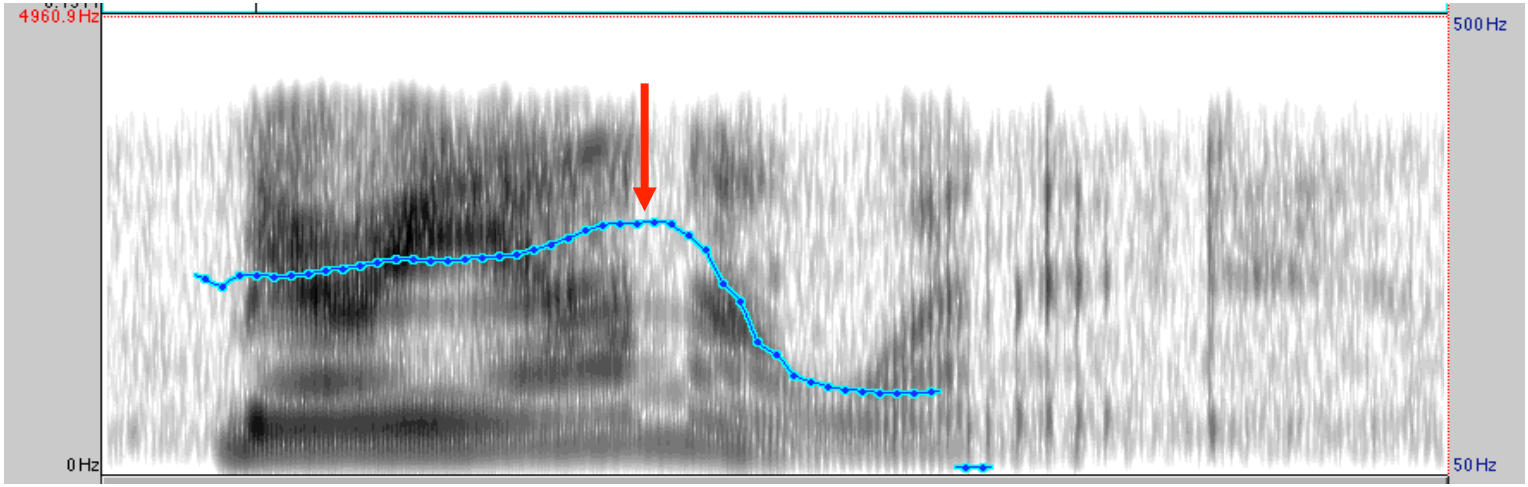
• Arvaniti, Ladd & Mennen (2000:121) present evidence from Greek of the types of problems that arise when we take *phonetic alignment* to the accented syllable to be the exponent of association of tones to segments.

"we show that there exist pitch accents that are clearly bitonal but which neither tones is, strictly speaking, aligned with the accented syllable. We argue from this fact that association cannot be based on phonetic alignment in any straightforward way and that a more abstract and rigorously defined notion of starredness is required."

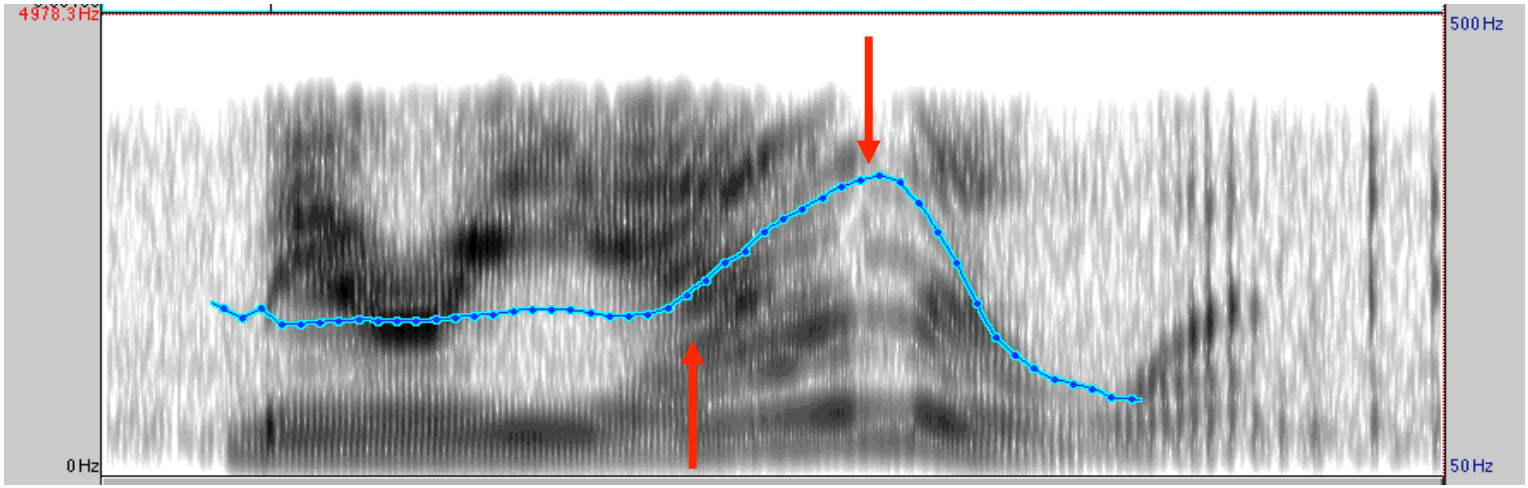
H* vs. L + H*

- Marianna won it.

H*



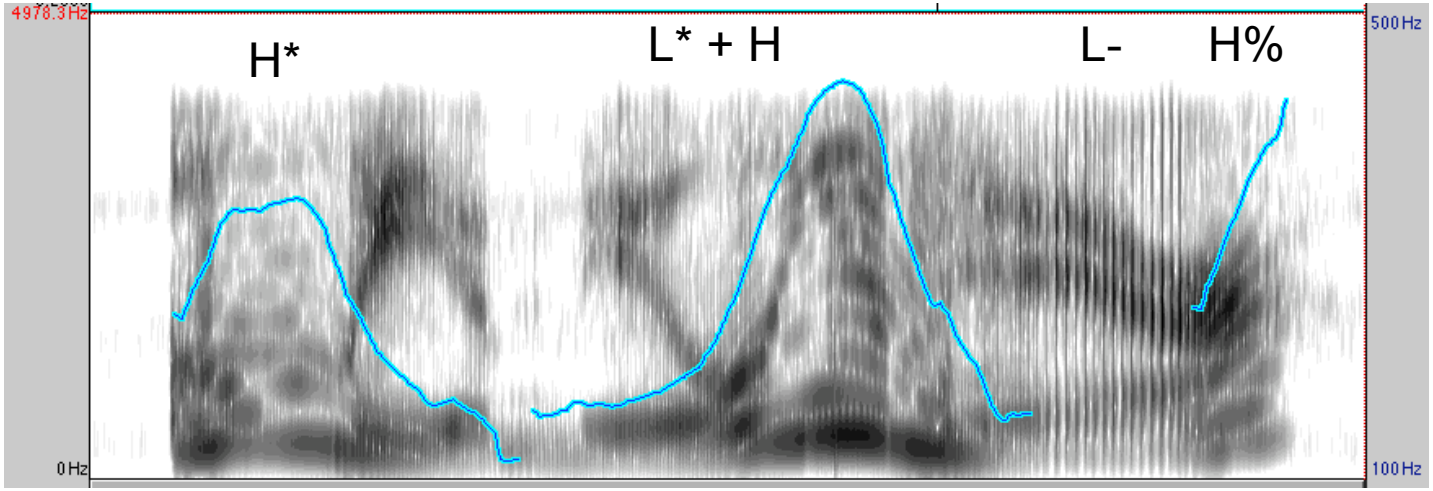
L + H*



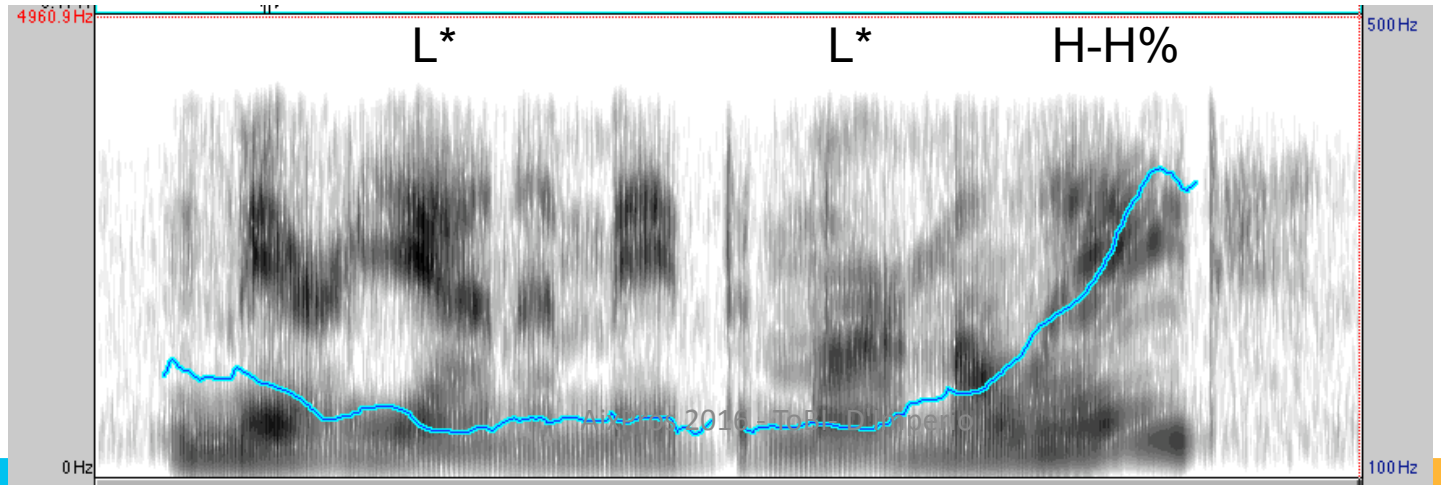
Note: informative vs. contrastive function

L* vs. L* + H

- Only a millionaire.



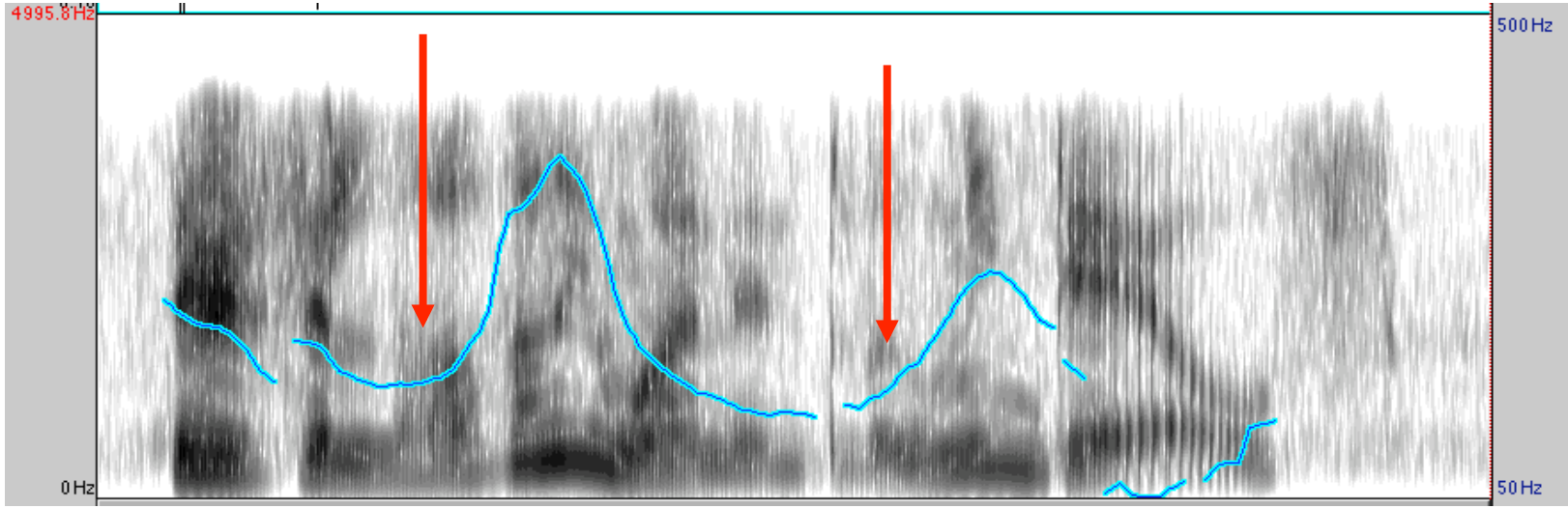
- Marianna made the marmalade.



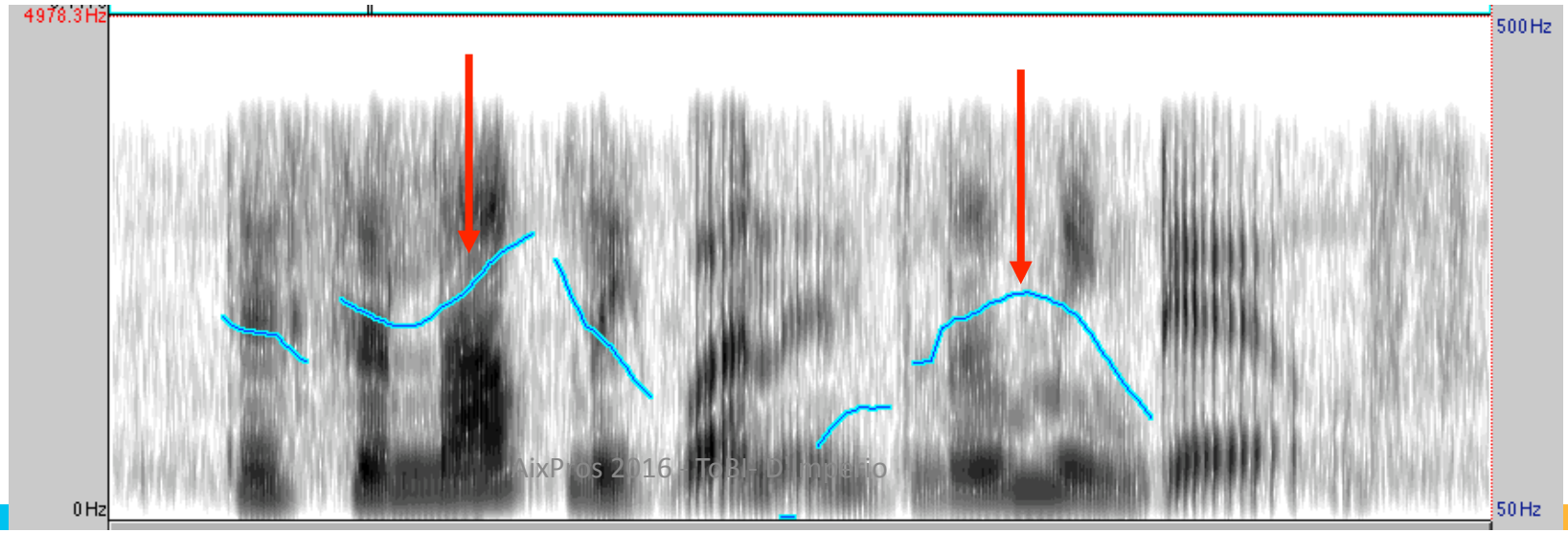
L + H* vs. L* + H

- There's a lovely one at Bloomingdale's.

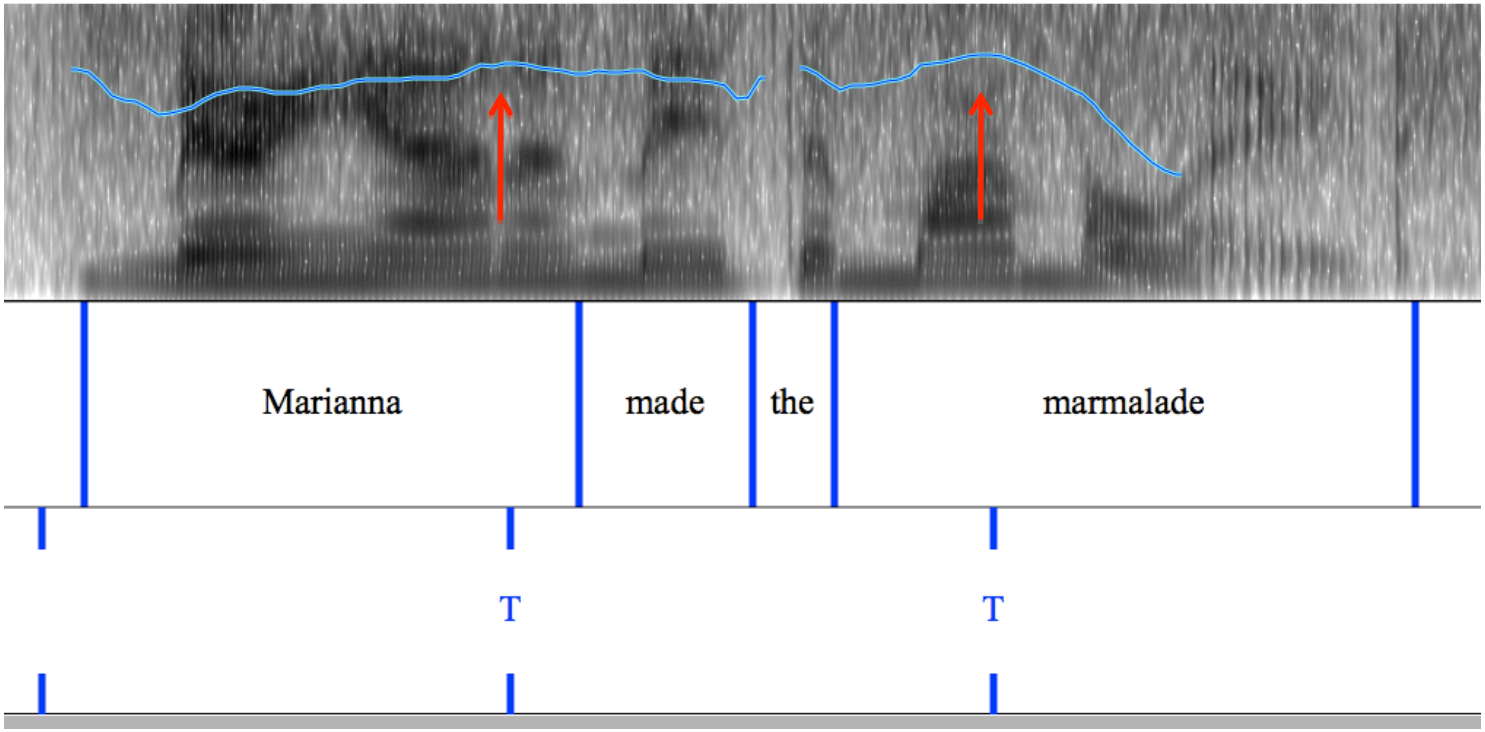
L* + H

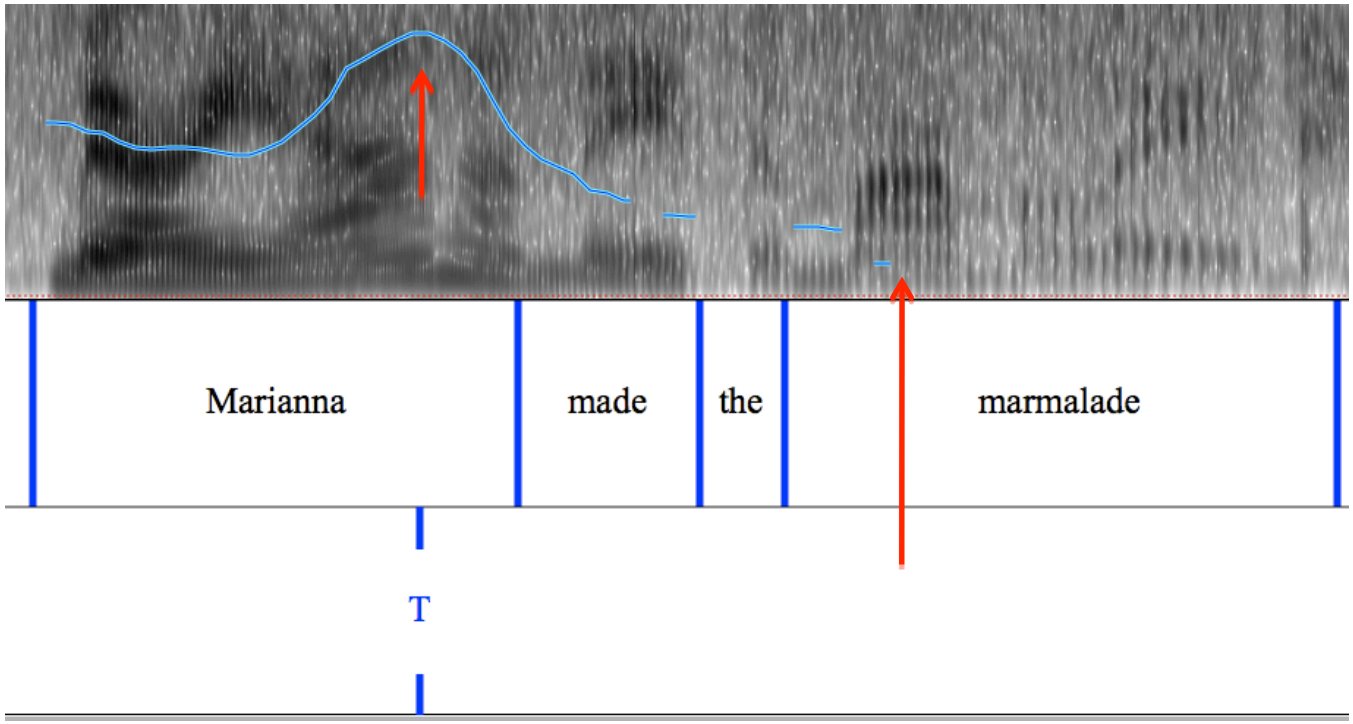


L + H*



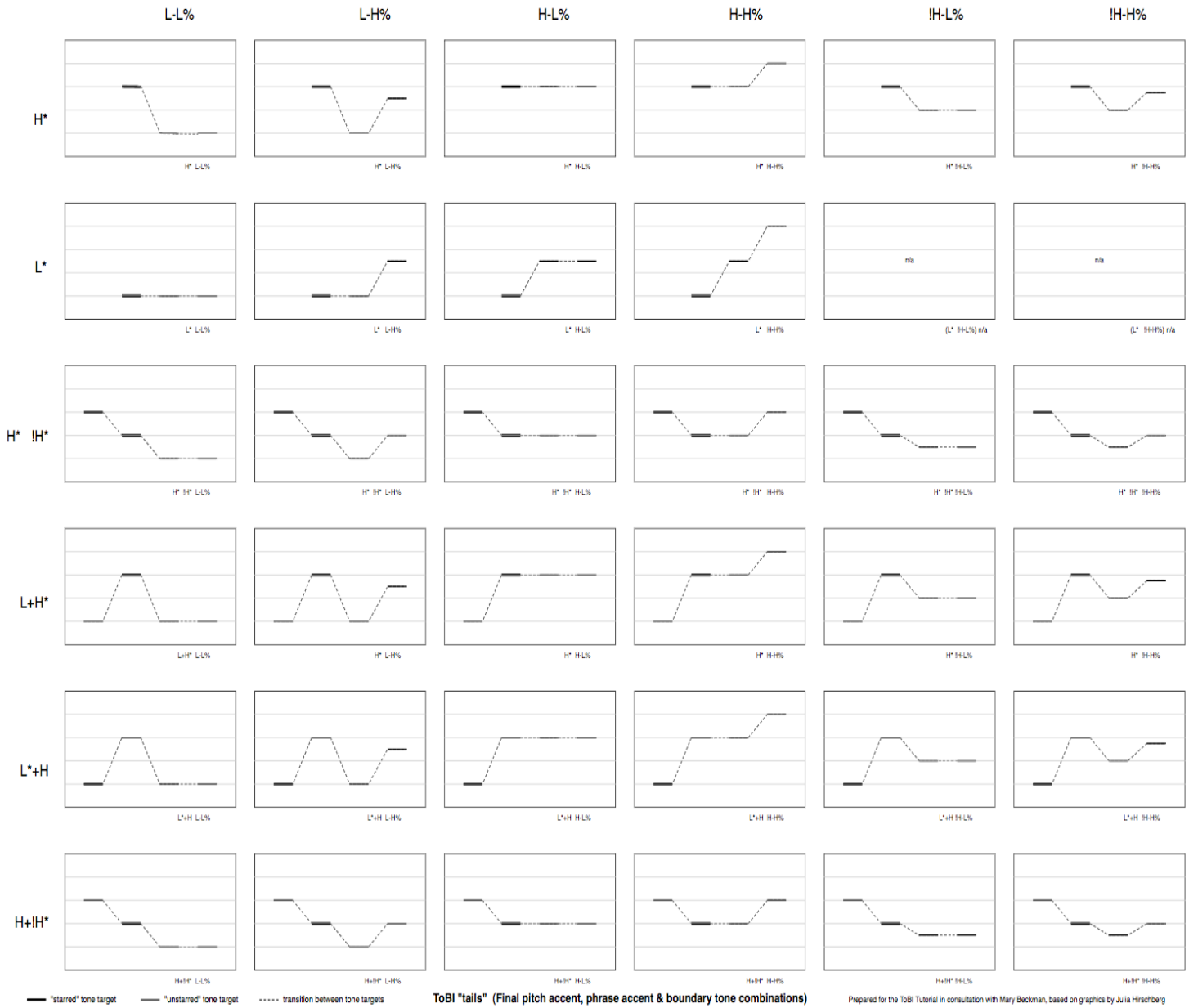
Marianna made the **marmalade**




Marianna
made
the
marmalade
T

Pitch-Accents Round-up

- There are four pitch accents:
- H*
- L*
- L + H*
- L* + H
- They attach to stressed syllables
- The final pitch accent in an intonational phrase is the **nuclear accent**.
    - Generally perceived as more prominent.



From: Nanette Veilleux, Stefanie Shattuck-Hufnagel, and Alejna Brugos. *6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI, January IAP 2006*. (Massachusetts Institute of Technology: MIT OpenCourseWare), http://ocw.mit.edu (Accessed). License: Creative Commons BY-NC-SA (Appendix I)

Practice Time…