INTONATION RESEARCH AND THE AUTOSEGMENTAL-METRICAL MODEL OF INTONATIONAL PHONOLOGY

Amalia Arvaniti
Radboud University

European Research Council
Established by the European Commission
sprint
Speech Prosody in Interaction
ERC-ADG-835263

CLS | Centre for Language Studies
Radboud University

SEMINAR PLAN

This workshop covers theoretical and practical aspects of doing intonation research
It explains the fundamental tenets of the autosegmental-metrical theory of intonational phonology (*AM*)

Part I:

- Challenging features of intonation
- How they are addressed in different models
- How AM comes closer to addressing them

Part II:

- AM and phonetics
- Practical aspects of
  - doing research on intonation
  - using AM principles for intonation analysis

Acknowledgments

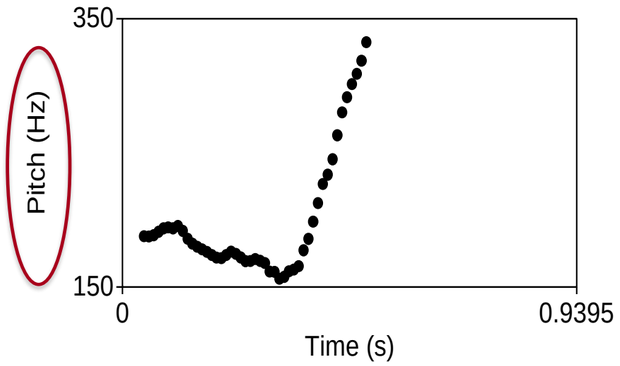
- Evangelia Adamou, CNRS-Lacito
- Mary Baltazani, Oxford
- Stella Gryllia, Leiden & Utrecht
- Argyro Katsika, UC Santa Barbara
- Bob Ladd, Edinburgh
- Georg Lohfink, Kent
- Marzena Żygis, ZAS & Humboldt

- Financial support from the
  - ERC-ADG-835263 (SPRINT)
  - British Academy grant SG160538
  - University of Kent start-up and QR funds

  is hereby gratefully acknowledged

F0, pitch, intonation, and prosody

- These four terms should not be used interchangeably
- F0 (fundamental frequency) is the main (but not the only) exponent of intonation and it is not intonation per se
- F0 is an acoustic property of the speech signal determined by the rate of vibration of the vocal folds
- Vocal fold vibration rates are determined by physiology (sex, age), as well as social factors (e.g. gender and related social norms)
- F0 is measured in Hz (cycles per second)
- F0 gives rise to the percept of pitch
- All languages use and modulate F0; however, intonation is only of the numerous functions of F0; i.e. **not all F0 modulations are intonation**
- Intonation is a component of a language's prosody (which also includes metrical structure)



Uses of F0

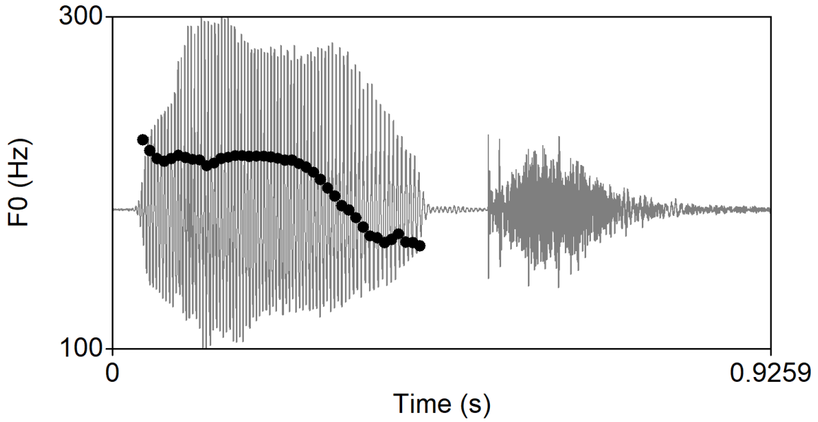
- Languages use F0 lexically, post-lexically, para-linguistically and for sociolinguistic purposes
- Lexical uses of pitch are referred to as tone and pitch accent
- Post-lexical grammatical uses of pitch are referred as intonation
- Para-linguistic uses refers to changes in pitch to show anger, boredom, surprise, excitement and the like; these uses are sometimes referred to as “emotional prosody”.
  - I prefer to use the term prosody for linguistic purposes only
- Changes in **pitch range**, **pitch span** and **dynamism** may also serve sociolinguistic functions, such as performing gender
- All these uses are not mutually exclusive: e.g. languages that use F0 for lexical purposes also have intonation, one can be bored in a tone language, etc.

Female speaker from southern US using extreme span combined with a high pitch range to show excitement

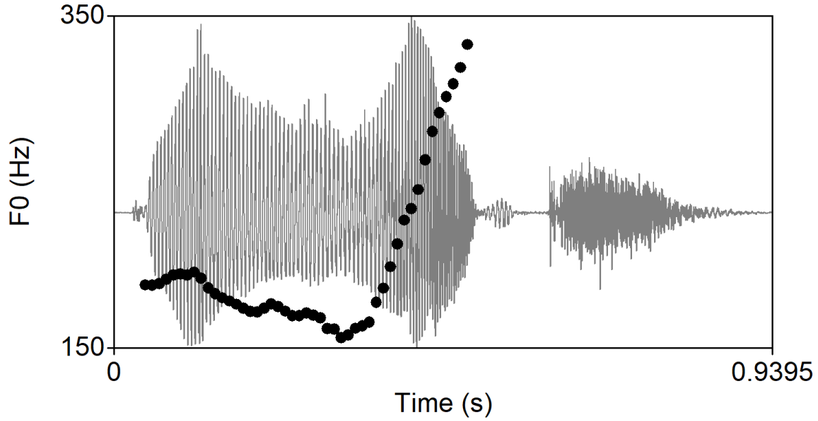
Male and female speaker using the same tunes but with distinctly different pitch range and span

What is intonation?

- Intonation refers to language-specific and systematic modulations of F0 that
  - span entire utterances
  - have grammatical function
    - marking phrasal boundaries
    - encoding pragmatic information (modality, focus, implicatures)



- F0 changes related to intonation are NOT associated with lexical meaning, but with the syntax and pragmatics of the utterance
- cf. ***my ex*** as a response to
  - (i) *who's that?*
  - (ii) being surprised that your new partner has invited their ex to your birthday party



Some myths and misconceptions about intonation

- Intonation is difficult or unsystematic = the “othering” of intonation
  - “half-tamed savage” (Bolinger, 1987)
  - “It certainly cannot be taken for granted that intonation is systematic in any a-priori (grammatical) sense” David Crystal (1969: 385) review of Halliday (1967) in *Language*
- Intonation shares patterns across languages based on biological characteristics (cf. Ohala, 1983; cf. the frequency code of Gussenhoven, 2004)
- Intonation has to do with affect or emotion
- Intonation can be switched on and off (“instruct participants to speak without intonation”)

- Intonation is as systematic as the rest of speech; “othering” intonation comes from confusing intonation and F0

- We do not have sufficiently detailed studies of the intonation systems of enough languages to be certain there are such trends
- Confuses intonation and paralinguistic F0 uses
- It is impossible to speak without intonation

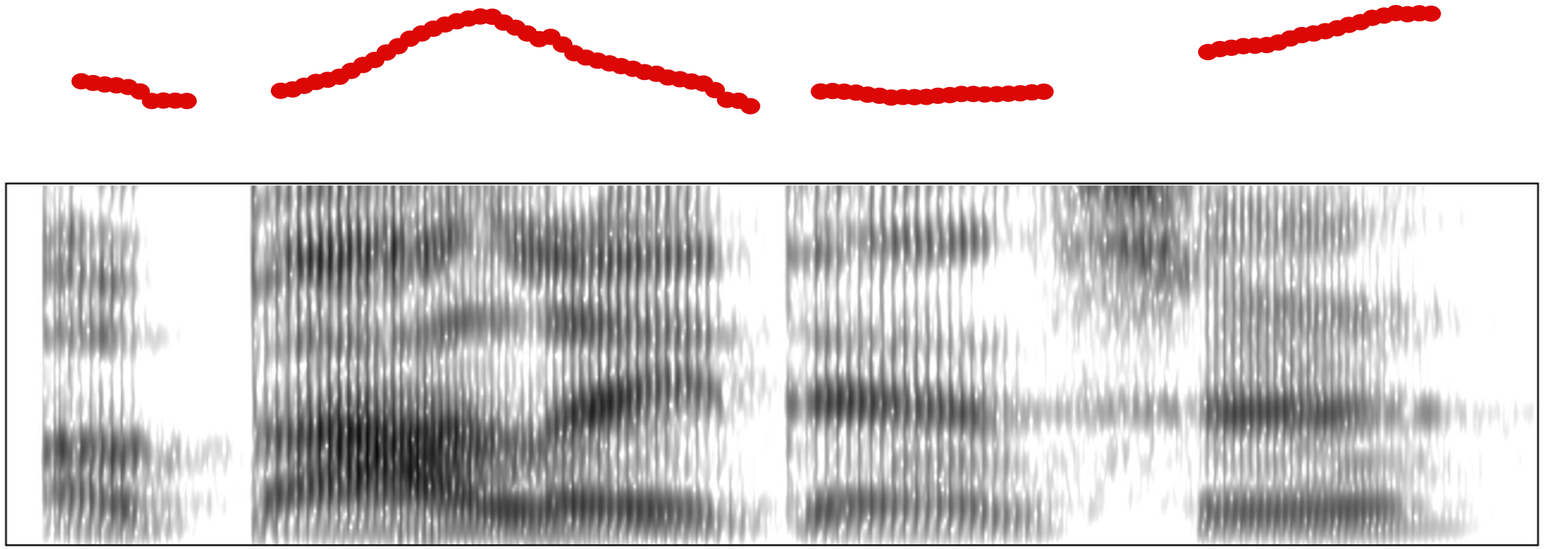


- Particular tunes have particular meanings; e.g. “question intonation”

- The relationship between tune and meaning is many to many (because pragmatics)

A main challenge in intonation research

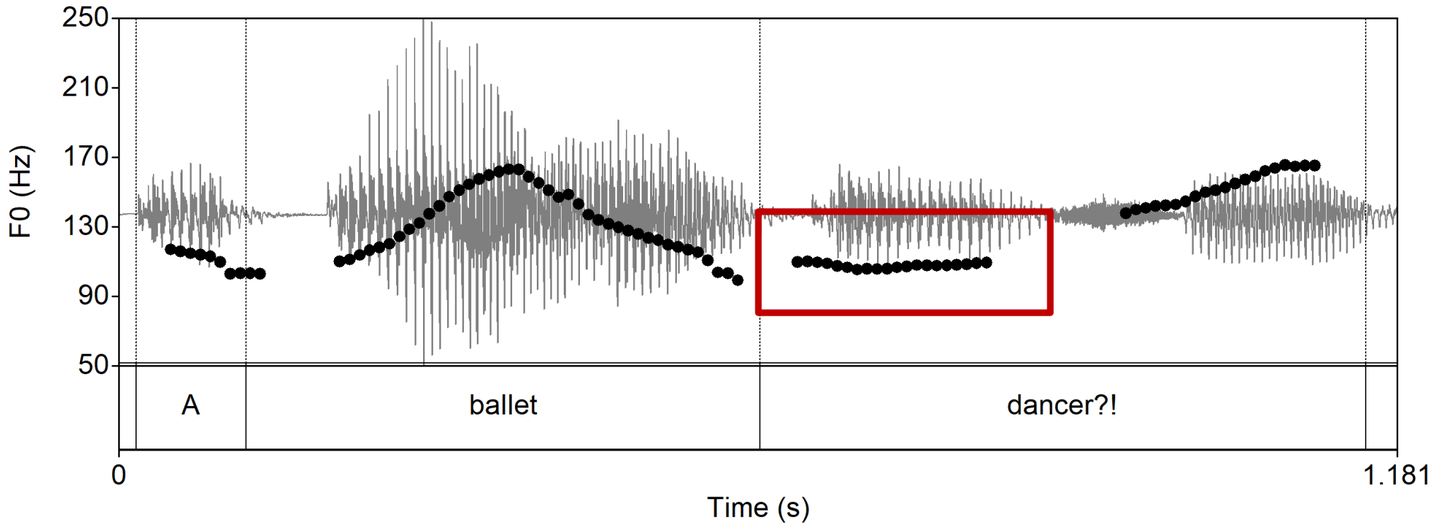
- How to determine the number and nature of a tune's components

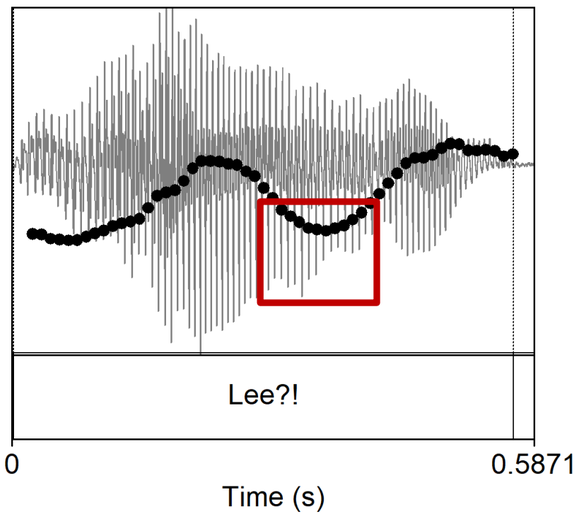


- One option is to assume there are no components, i.e. to assume tunes are configurations

Why pitch contours are not configurations?

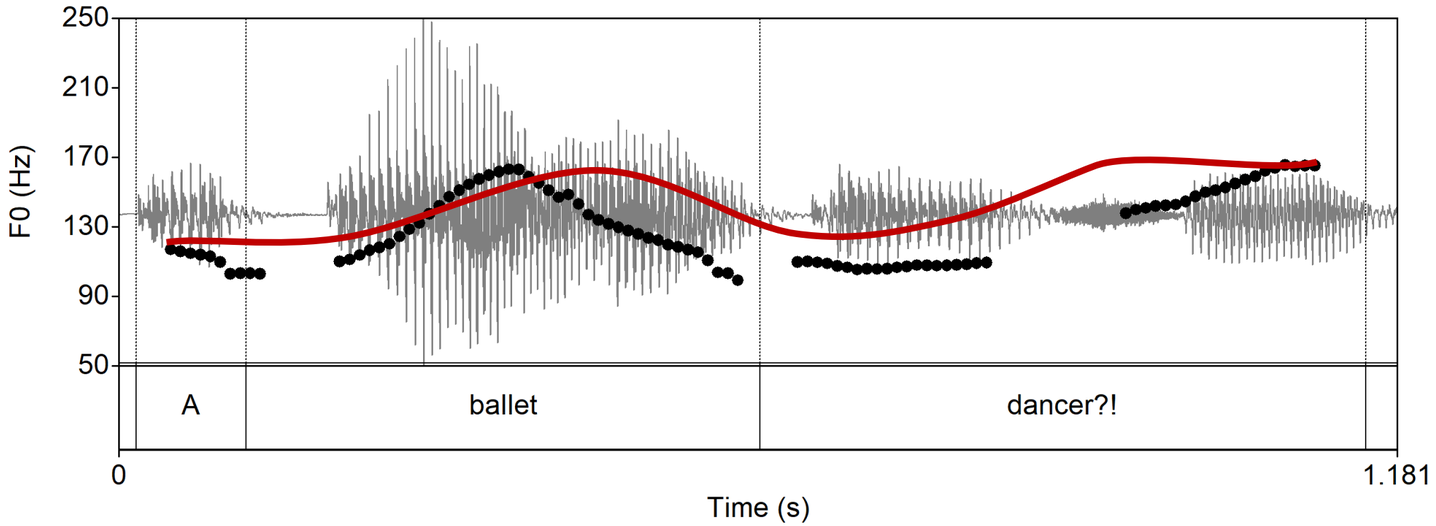
- Pitch contours can vary substantially even when listeners recognize them as instances of the same tune
- Differences are systematic and due to linguistic context: **lawful variability** (Arvaniti & Ladd, 2009; Arvaniti, 2016)

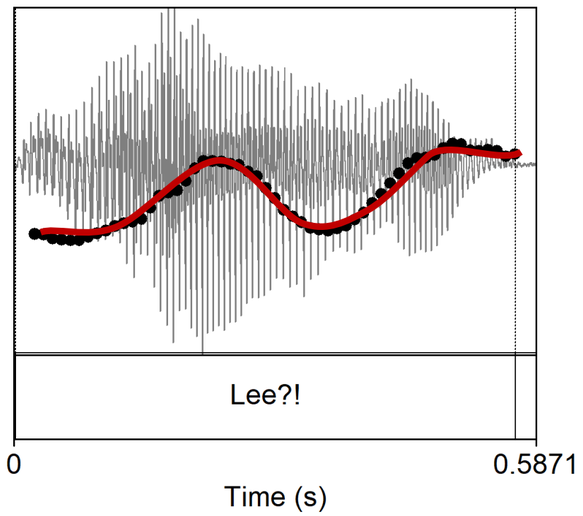




Contours do not behave like an accordion

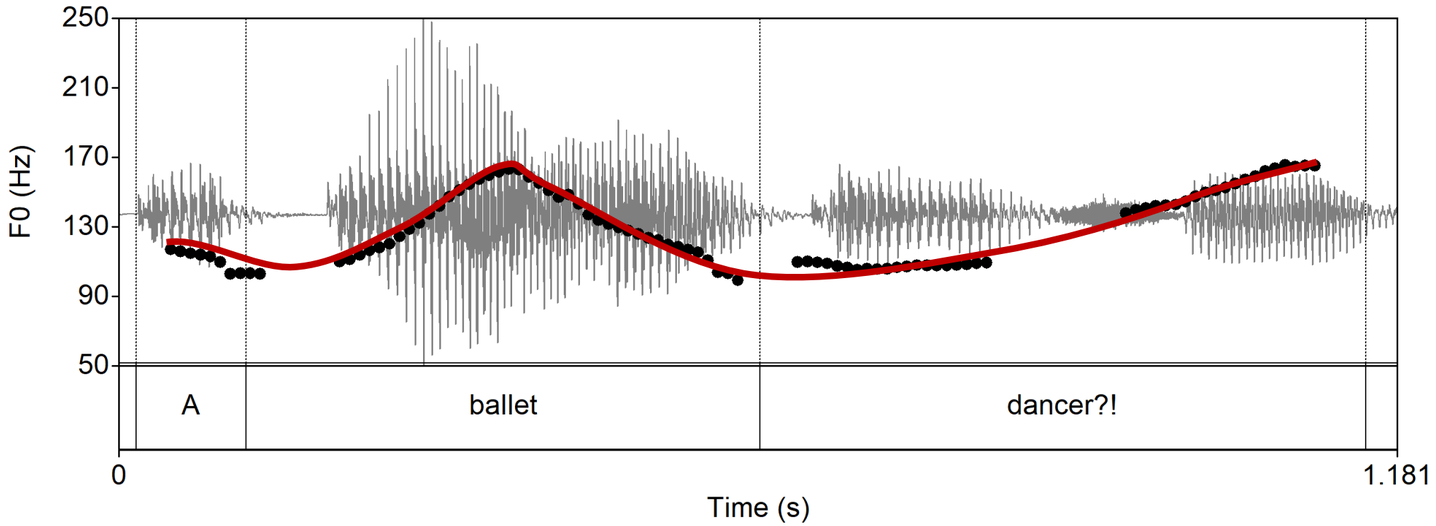
- Contours do not shrink or stretch to fit the segments of an utterance but display systematic changes that can be explained by parameters such as utterance length and the position of stressed syllables

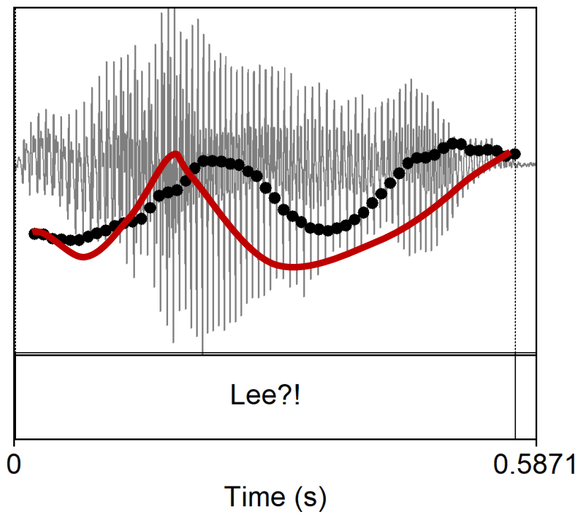




Contours do not behave like an accordion

- Contours do not shrink or stretch to fit the segments of an utterance but display systematic changes that can be explained by parameters such as utterance length and the position of stressed syllables





Essential research questions in intonation

- Our first aim should be to uncover this systematicity in the variation among tunes
- Given the variability in contour shape,
  - how do speakers of a language determine which contours are instances of the same tune?
  - how do they learn how to produce tunes with segmental material of varying lengths and structures?
- How can we capture the above and separate lawful variability from (i) the use of a different tune, and (ii) paralinguistic effects (see Arvaniti, 2019, ICPhS)

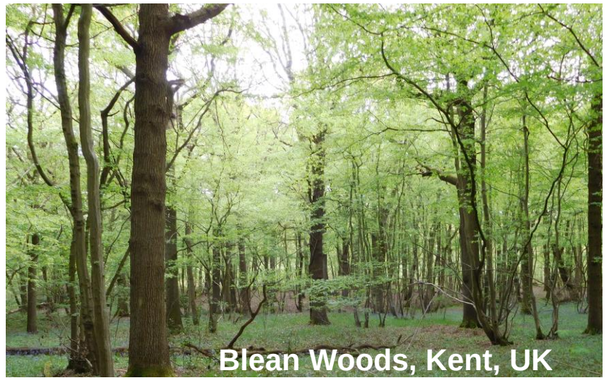
Essential research questions in intonation

- Our first aim should be to uncover this systematicity in the variation among tunes
- Given the variability in contour shape,
  - how do speakers of a language determine which contours are instances of the same tune?
  - how do they learn how to produce tunes with segmental material of varying lengths and structures?
- How can we capture the above and separate lawful variability from (i) the use of a different tune, and (ii) paralinguistic effects (see Arvaniti, 2019, ICPhS)

Prerequisites for a successful theory of intonation

1. account for variability in realization
2. reach abstractions essential for capturing generalizations

Any model that covers only 1 or 2 is unlikely to be sufficient

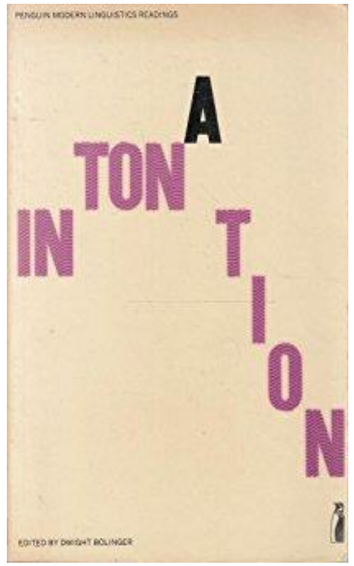


seeing the forest and the trees

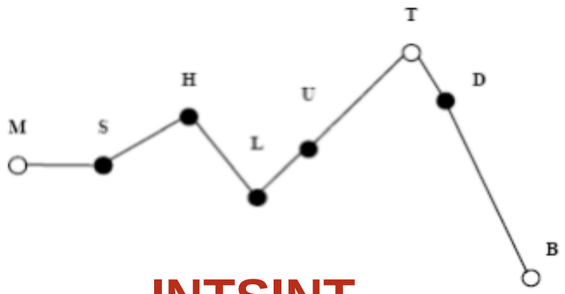
- The debate about this relationship is not new or unique to intonation
- It has been a particularly difficult debate in intonation because of the features and multiple functions of F0
- The general debate on these matters suggests that we need a level of phonetic detail and a level of abstraction

Approaches to intonation: Idealization

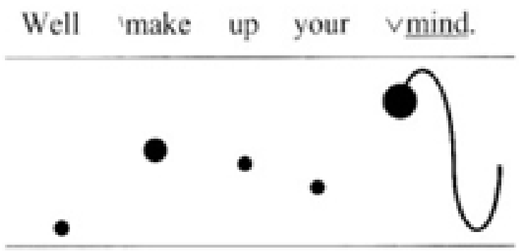
- Some models ignore variability altogether
- Such models cannot deal effectively with the phonetics of intonation



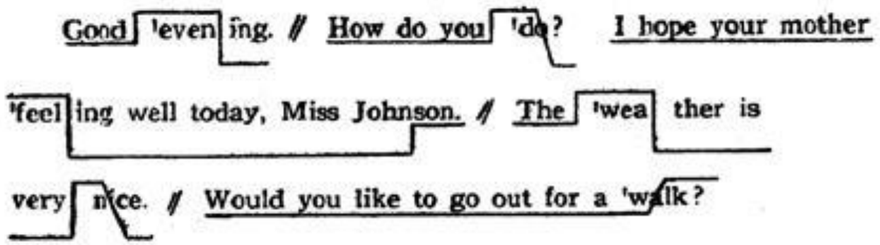
Bolinger



INTSINT



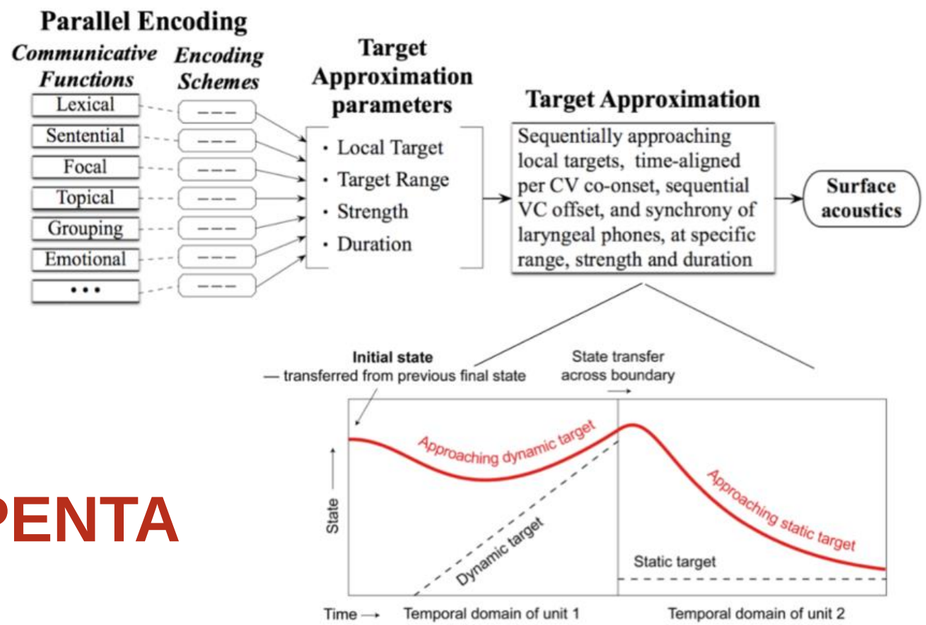
The British School



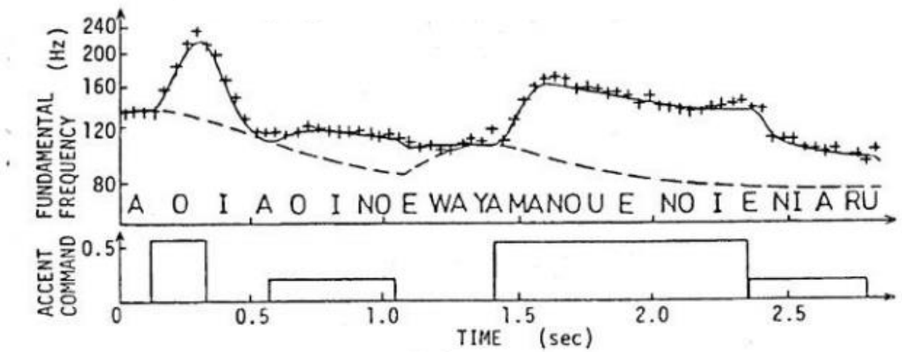
American Structuralists

Approaches to intonation: Focus on phonetic detail

- Some models seek to faithfully represent **F0 contours**
- Such models have difficulties with generalization (Arvaniti & Ladd, 2009; 2015)



PENTA



Fujisaki 1981

What do the idealization and phonetic approaches have in common

- To a very large extent, both types of models, whether they use idealized contour shapes or seek to model these contours in detail, are primarily preoccupied with modelling F0, not intonation
- This is clear for phonetic detail models: they do aim to model F0
- Idealization models abstract away from phonetic detail but not in a meaningful way that
  - allows us to create building blocks of intonation components that give rise to tunes
  - or understand how the meaning of tunes comes about

Both are needed to understand the variability in the shape of F0 curves and the lack of one-to-one correspondence between tunes and meaning

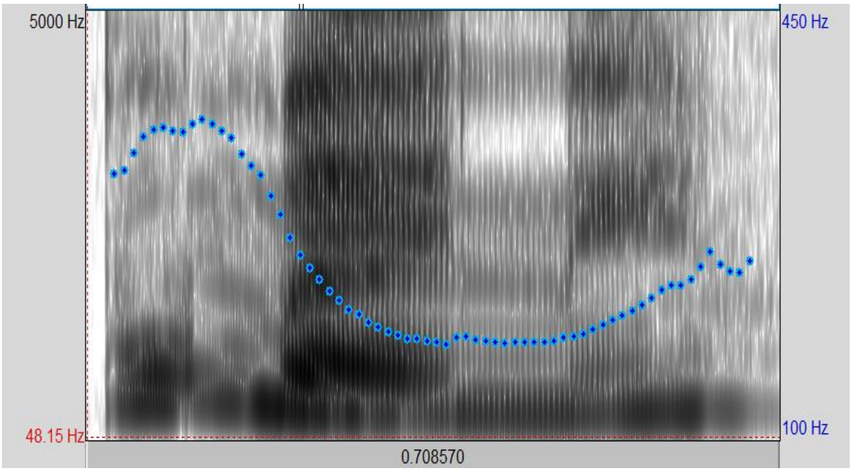
Autosegmental-Metrical Theory of Intonational Phonology (AM)

- Janet Pierrehumbert's thesis (1980): ***The phonetics and phonology of English intonation***
- Bob Ladd (1996) dubbed the theory *autosegmental metrical theory of intonational phonology*

- Why this mouthful of a term?
- Intonation involves the discrete elements called tones, L(ow) and H(igh), which are concatenated in various ways and associated with metrical structure
- **Autosegmental**
  - Tones are autosegments
- **Metrical**
  - Tones associate with structural positions in the metrical structure: heads and boundaries of constituents

What AM is and is not

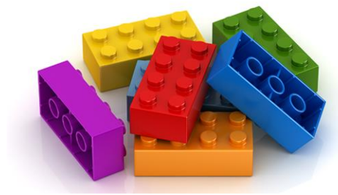
- AM is a phonological theory of intonational structure which also addresses how its phonological representations can be phonetically realized
- The aim of AM is NOT to faithfully present F0 contours – that's what pitch tracks are for
- AM is **not** a transcription system for intonation
- AM is not the same as ToBI, which is a family of systems for prosodic annotation; ToBI requires an existing AM analysis



F0 track, to be used for the phonetic analysis of intonation instead of symbolic representations of F0

Tones as autosegments

- The primitives of intonation are **tones**: low (L) and high (H)
- Tones are represented as a string of **autosegments**: LHLHLH
  - They are independent of vowels and consonants (though they have to co-occur with them to be realized)
  - They are independent of each other
  - They may form bitonal (or even tritonal) groups, e.g. L+H* is a bitonal pitch accent



- Some consequences of viewing tones as autosegments
- The relationship between tones and segments (Tone Bearing Units or TBUs) is NOT one-to-one
- The number of TBUs and tones may match
  - We can have more tones than TBUs
  - We can have more TBUs than tones

Tones as autosegments: an example

- *Etung*

édìmbá	bìsóŋé	ékúé	òbô
pot	*wife*	*forest*	*arm*

(from Gussenhoven & Jacobs 1998)

- How can we account for these patterns, especially the fact that in some words every syllable has a different tone, in others they all have the same tone, and in yet others, some syllables have a complex tone (e.g. last syllable of *arm*)?

édìmbá	bìsóŋé	ékúé	òbô
| | |	| V	↙	| Λ
H L H	L H	H	L HL

Intonation, tones and metrical structure

- Tones create melodies, many of which are very frequent; e.g. **H* L-L%** is the most common melody for English declaratives (what may be described as a fall)
- The autosegmental representation of melodies also includes information as to how the tones associate with the segmental string: this is known as the **tune-text association**
- AM assumes the existence of a metrical structure that is independent of a language's melodies
- This structure provides the necessary information:
  - prominence relations among constituents (*heads*, informally *stresses*)
  - position of phrasal (and other constituent) boundaries
- Tones associate with the **heads** and **edges** of metrical constituents
- Whether tones associate with both heads and edges is language specific;
  - e.g. languages that do not have stress, such as Korean, only show tone associations with phrasal boundaries

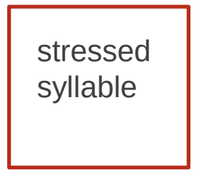
Association by tone type

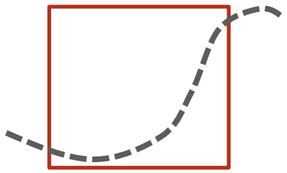
- Tones may associate with the heads of metrical constituents (roughly, stressed syllables):
  - pitch accents, e.g. L*, L+H*, L*+H

- Tones may associated with boundaries:
  - phrase accents, e.g. L-, H-, LH-, associate with the boundaries of intermediate phrases (ip)
    - not all AM analyses involve this level of phrasing
    - phrase accents are also used to demarcate the edges of the Accentual Phrase (e.g. in analyses of Korean and French), a constituent larger than the prosodic word but smaller than the ip
  - boundary tones, e.g. L%, HL%, associate with Intonational Phrase (IP) boundaries
    - typically found in the right boundary (but left boundaries may also associate with boundary tones)
    - analyses with complex BTs have been proposed for Korean, e.g. LHLH% and HLHL% (among other complex boundary tones)

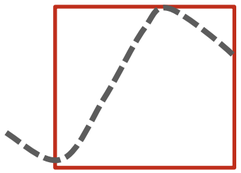
* % + and other AM esoterica

- *, - and % are diacritics that indicate the association of tone with metrical structure
- In bitonal accents the * indicates the **metrically stronger** tone and typically the one that co-occurs with the stressed vowel (but it does not have to be; this is a matter of analysis not phonetics alone)
- When tonal events consist of two (or more) tones, a + may be used to indicate their connection: L+H% or LH%
- Jun & Fletcher (2014) have suggested that + be used only if the tones are independent of each other:
  - L+H* is a weak L tone followed by a strong H tone
  - LH* is a rise
- This practice is not yet widely accepted however
- Brackets are meant to show that the * is related to the unit rather than a particular tone, e.g. (L+H)*
- In some analyses left-associating boundary tones are represented with the symbol before the tone, e.g. %H





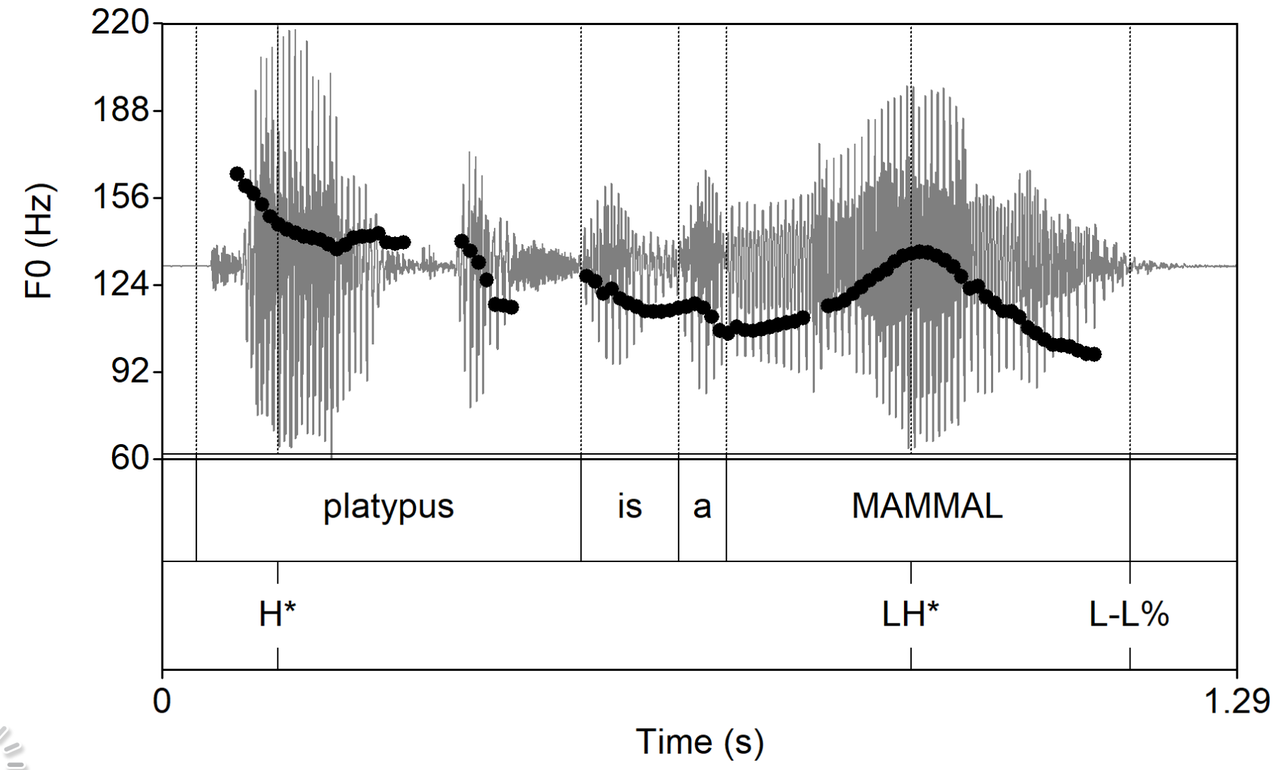
L*+H and



L+H* in English

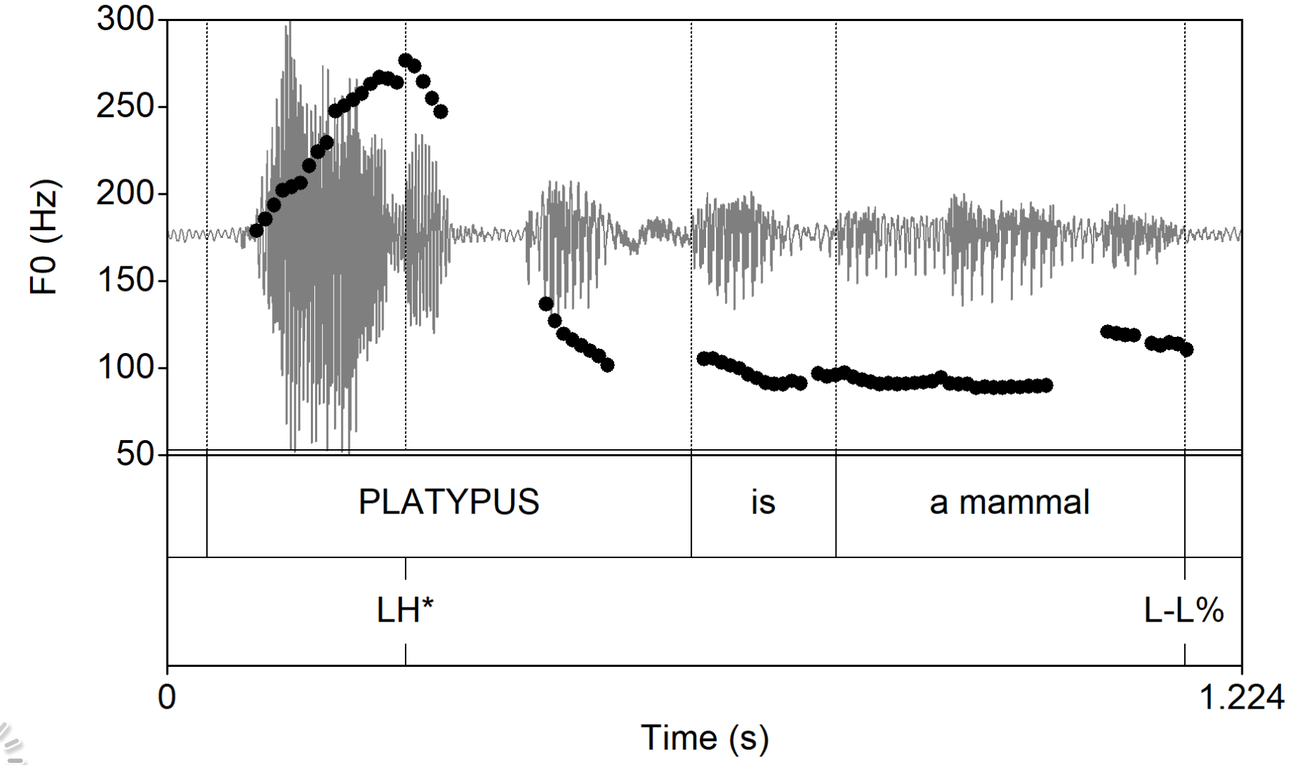
Text-tune association i

-Is platypus a bird?
-Platypus is a MAMMAL

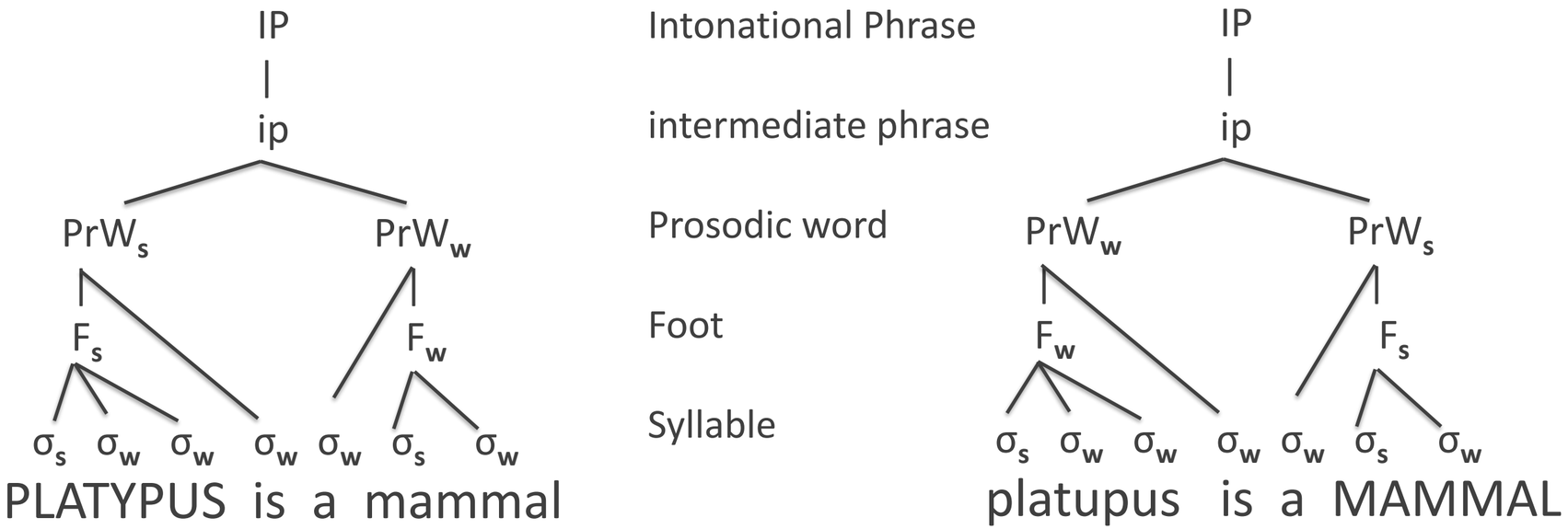


Text-tune association ii

-Name a mammal
-PLATYPUS is a mammal

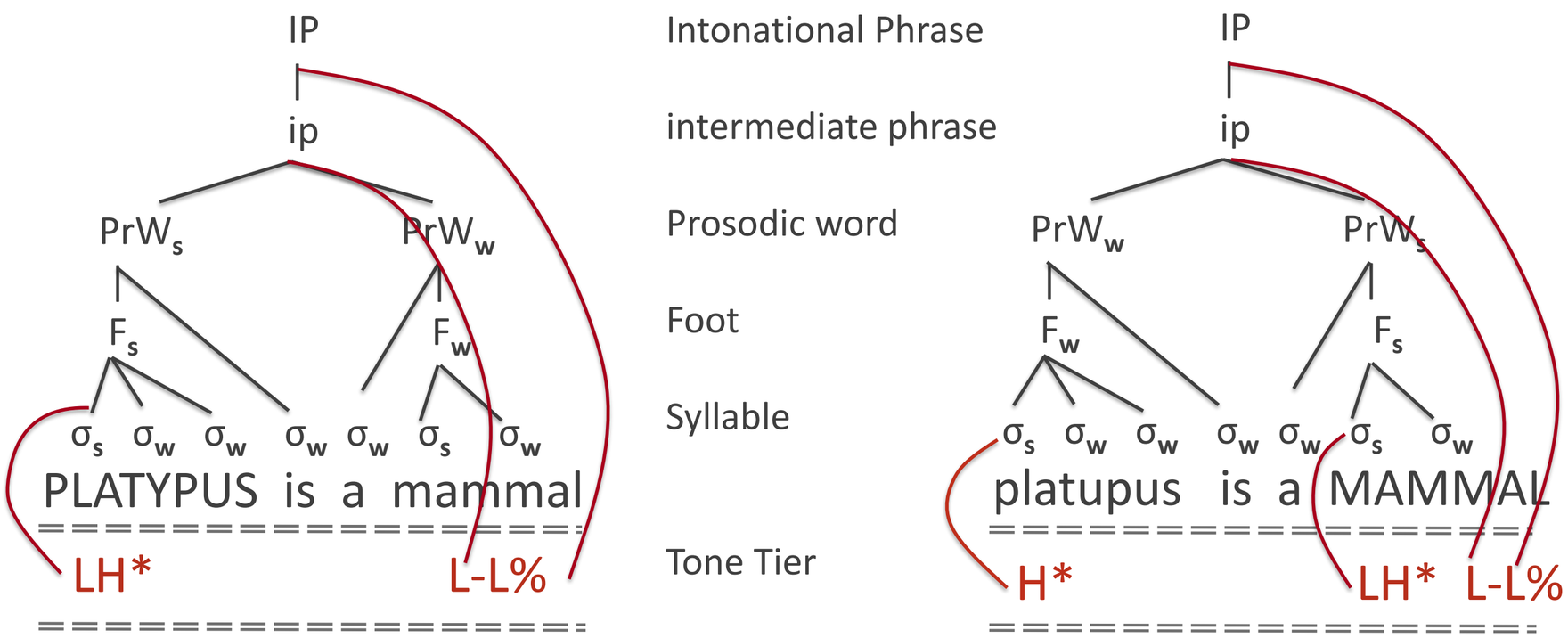


Metrical structures



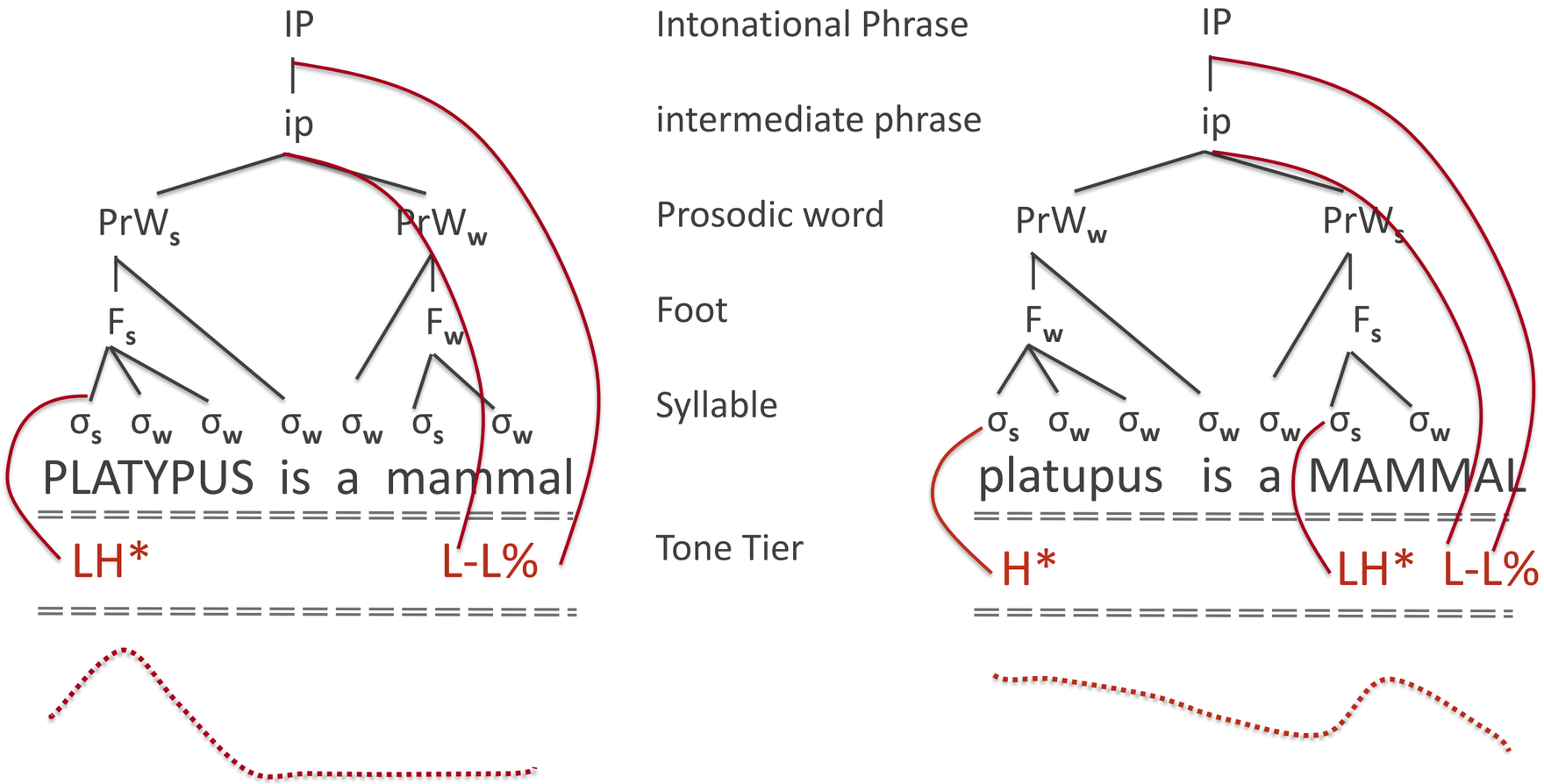
Structures based on Pierrehumbert & Beckman (1988)

Metrical structure and association to tones



Structures based on Pierrehumbert & Beckman (1988)

Metrical structure and association to tones



Structures based on Pierrehumbert & Beckman (1988)

RECAP

- AM is a theory of intonational phonology
- Its aim is to represent the contrastive elements of an intonational system
- These elements are tones, which are considered to form a string of autosegments
- Tones associate with structural positions in metrical structure:
  - heads of constituents (informally stresses)
  - phrasal boundaries
- The overall representation of intonation is seen as part of a language's phonology
- AM is not meant o provide a transcription of actual F0 contours

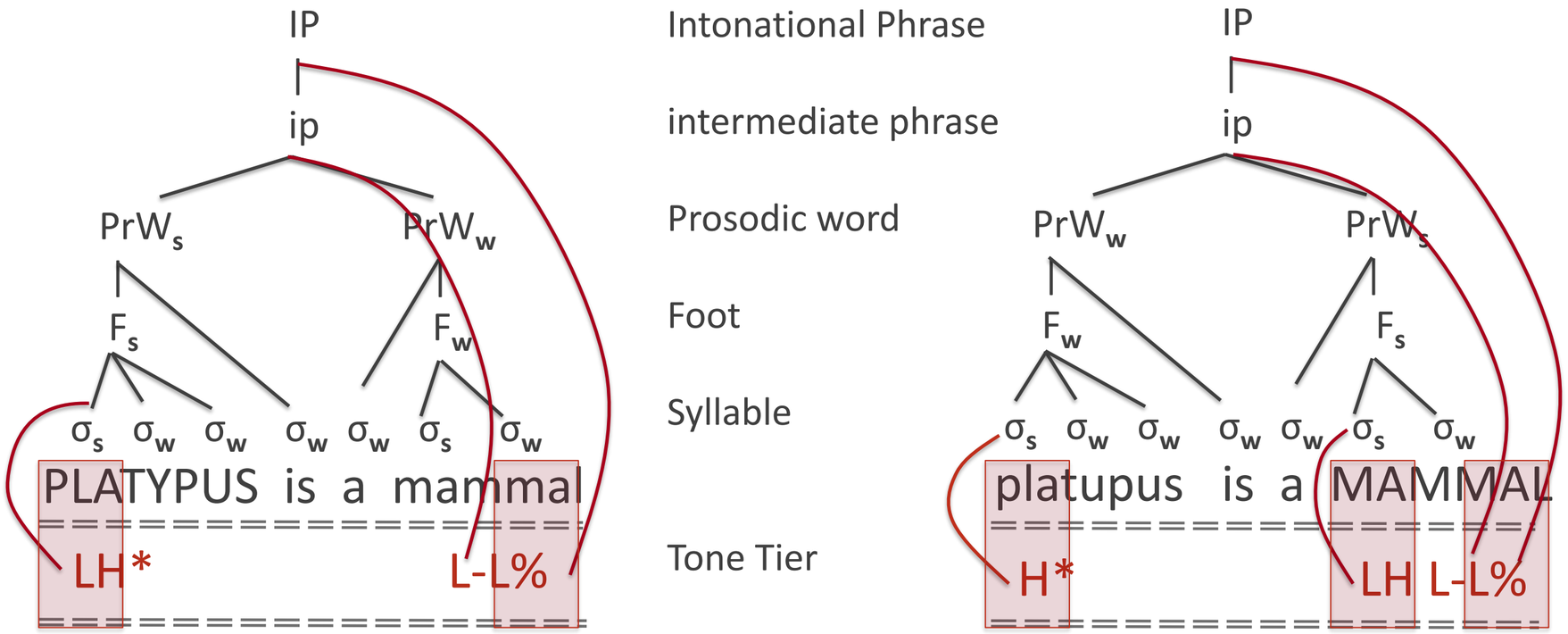
PART II

- AM and phonetics
- Practical aspects of
  - doing research on intonation
  - using AM principles for intonation analysis

Phonetic realization of tones in AM

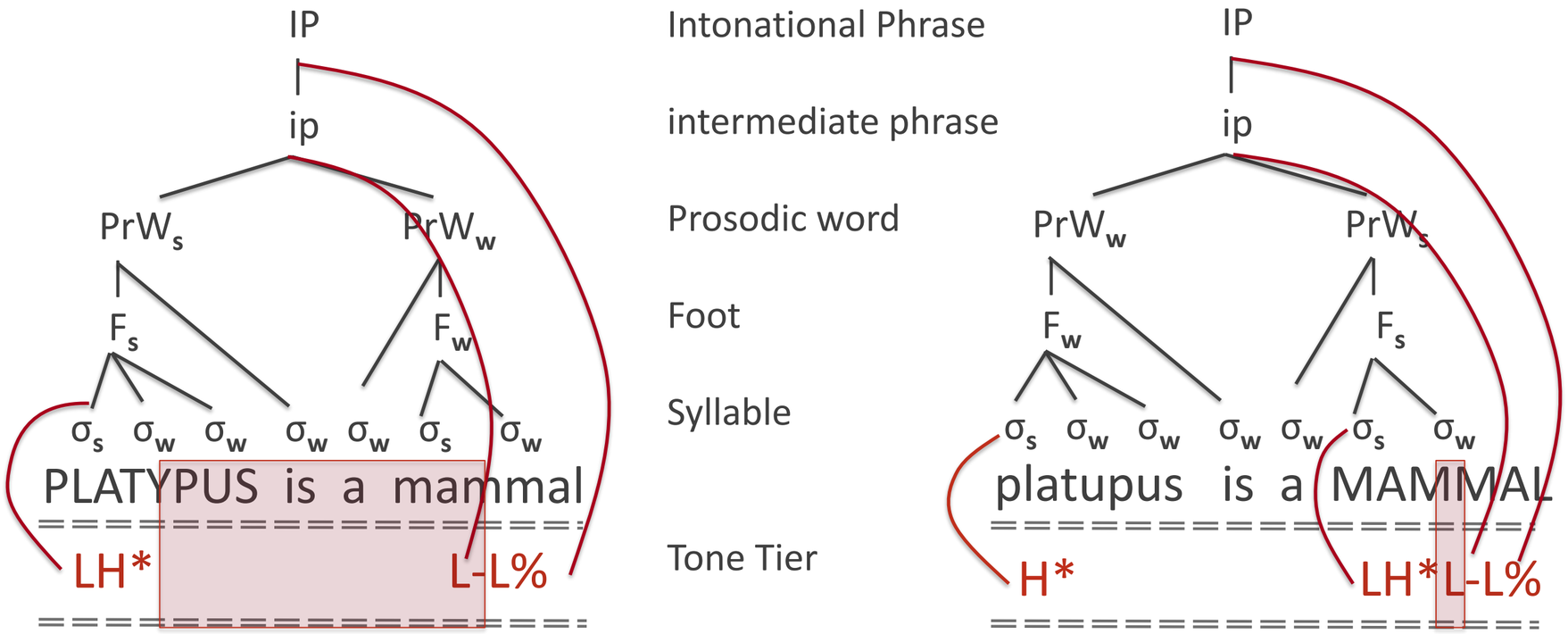
- AM includes a theory of how its abstract representations are phonetically realized
- Phonological tones are realized as tonal targets, specific points in the F0 contour defined along two dimensions:
- **Scaling** = their F0
- **Alignment** = their temporal position with respect to segmental landmarks
- Tonal targets often are local minima and maxima of F0
- Segmental landmarks are likely to be TBUs, syllable and phrasal boundaries
- Landmarks depend on the type of intonational event; roughly
  - pitch accents co-occur with accented syllables
    - the tones of pitch accents may occur before or after the accented syllable
  - boundary tones phonetically align with IP boundary-adjacent syllables
  - phrase accents behave phonetically like boundary tones (but aligning with ip boundaries); phrase accents may also
    - spread
    - be realized on stressed syllables (Grice et al., *Phonology*, 2000)

Metrical structure and association to tones



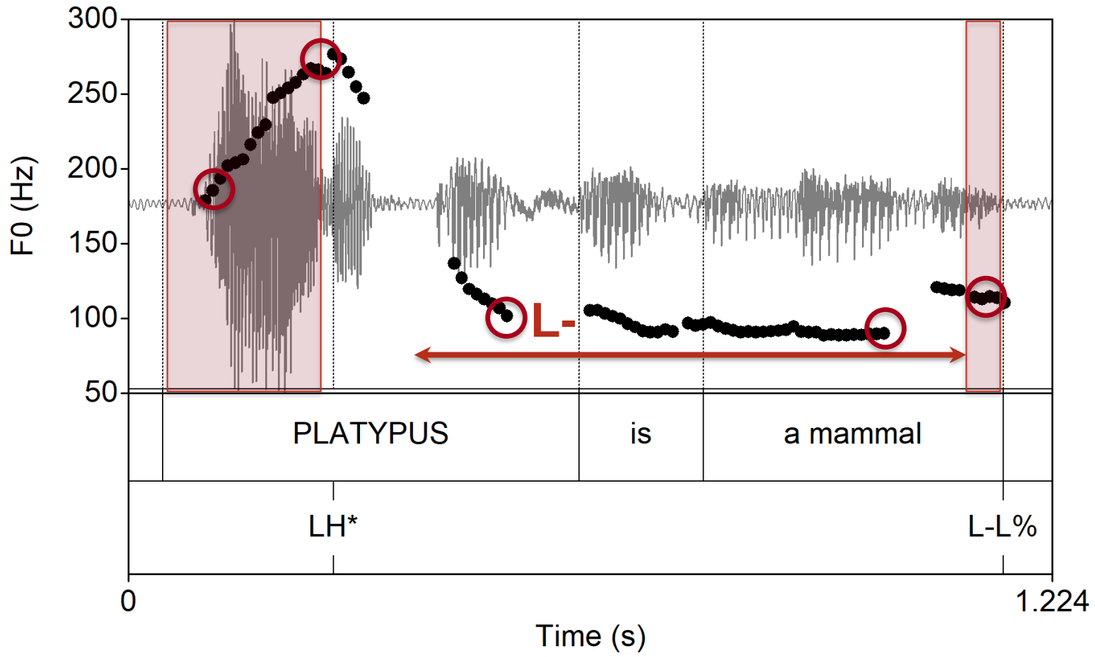
Structures based on Pierrehumbert & Beckman (1988)

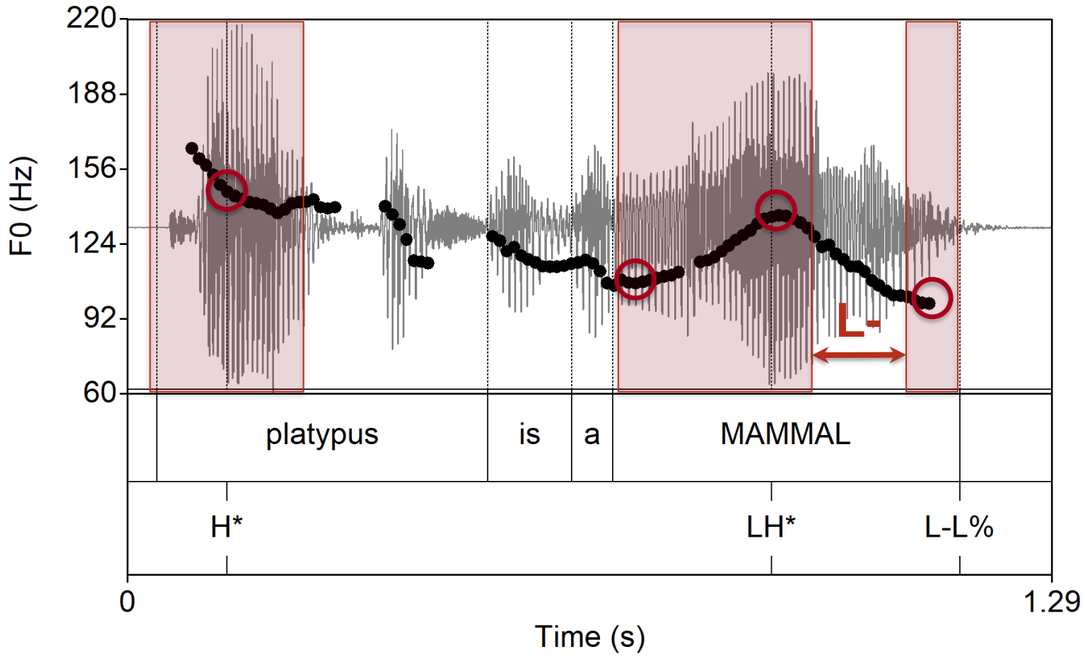
Metrical structure and association to tones



Structures based on Pierrehumbert & Beckman (1988)

F0 and alignment

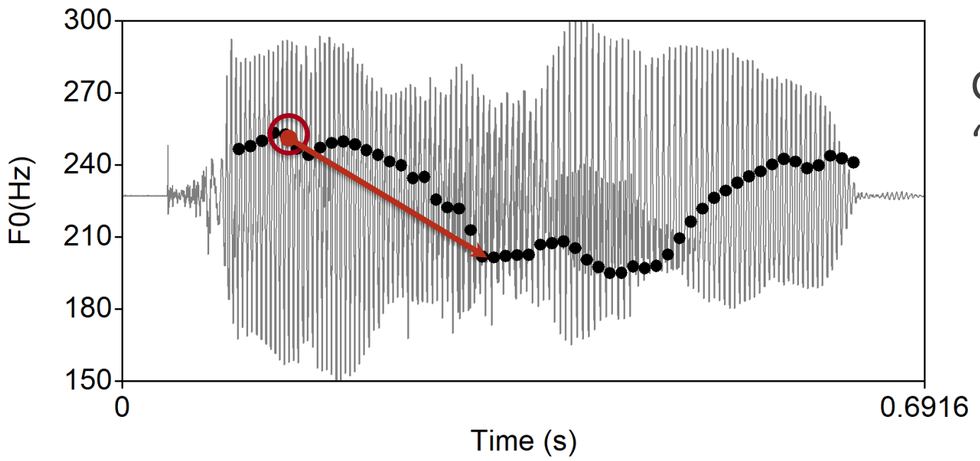




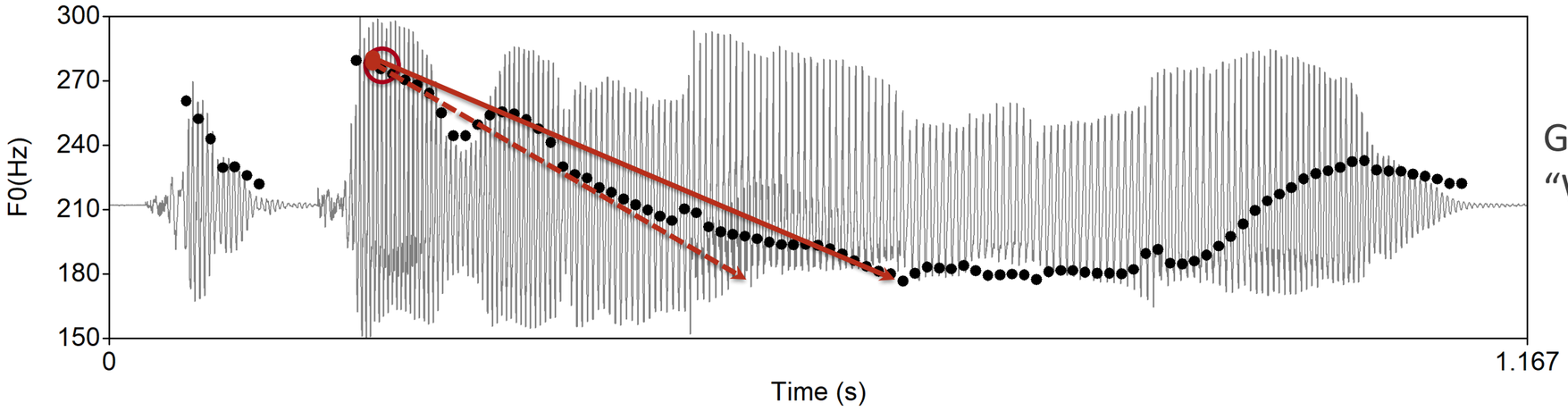
AM and underspecification

- In AM, since tones do not exhaustively represent the course of F0, we can say that representations are underspecified
- This applies at the phonetic level as well: no specifications for F0 are filled in for parts of the segmental tier that are not associated with tones
- There are several consequences of underspecification
  - the F0 of specific stretches may vary from instantiation to instantiation of a tune
  - this is because for segmental material not associate with tones F0 is **derived by interpolation**
  - **tones may spread**

Consequences of underspecification: variable realization



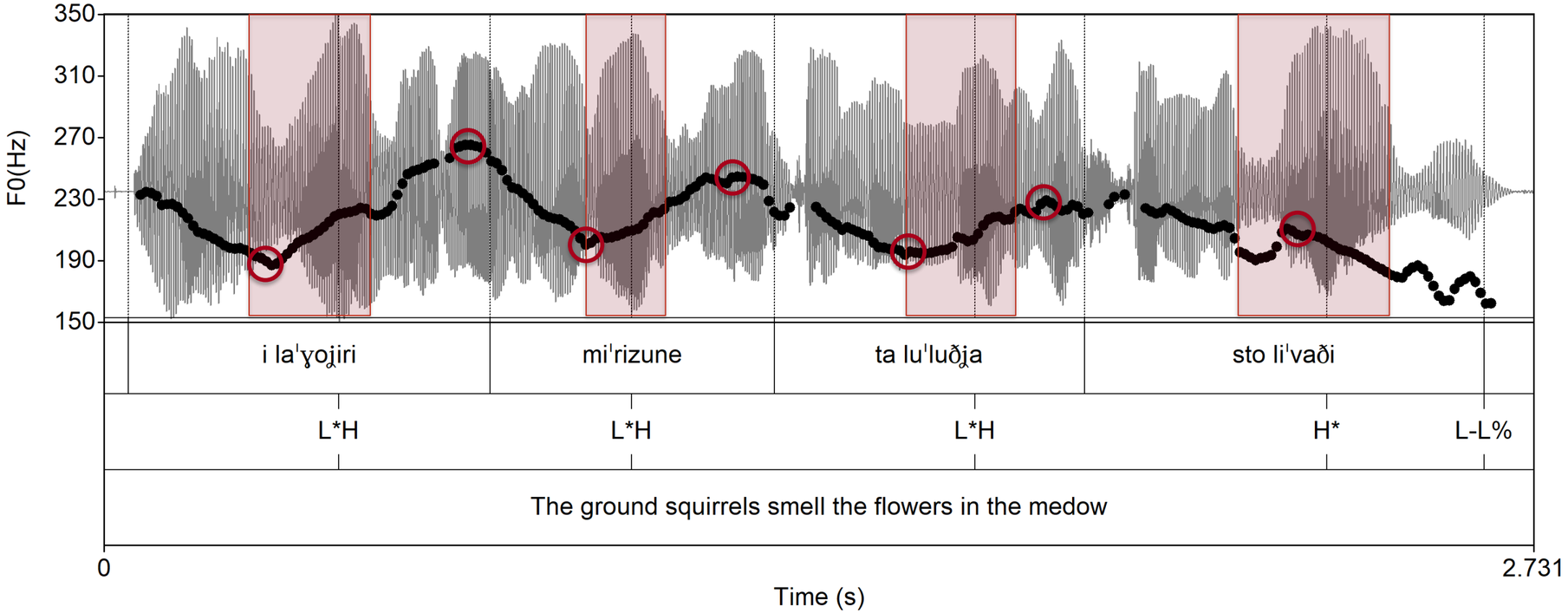
Greek /ˈpu ˈzi/
“Where does s/he live?”



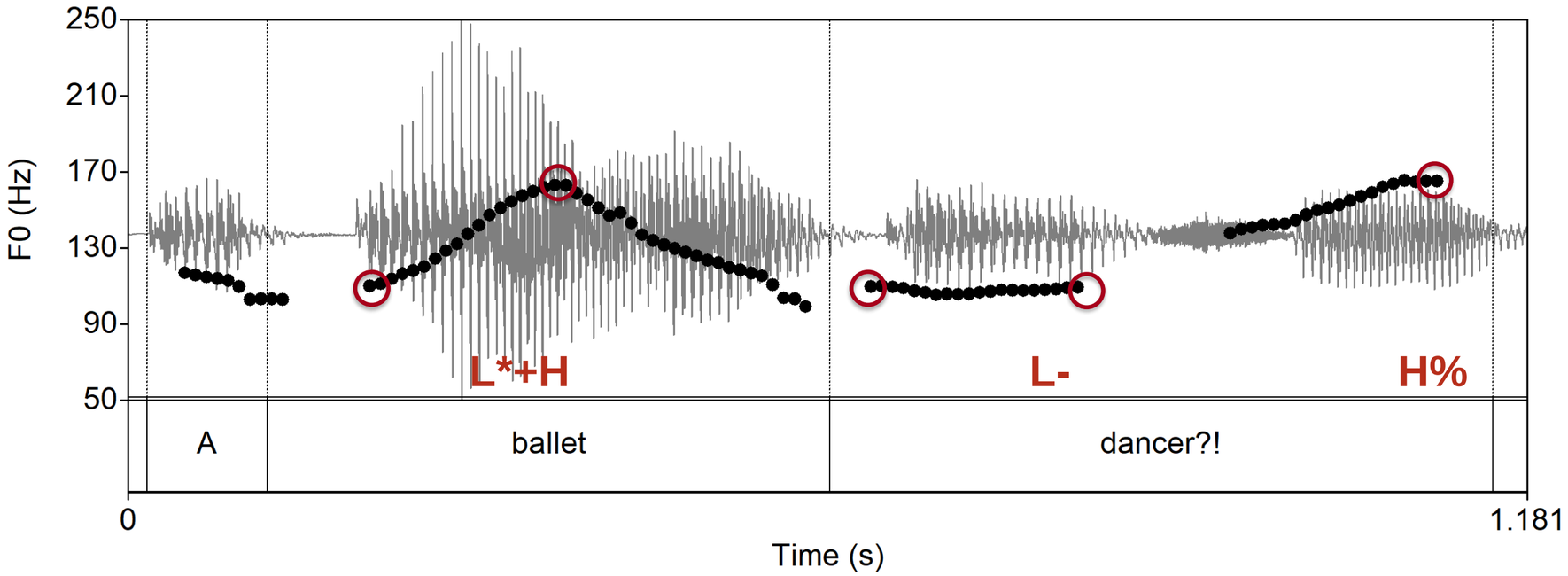
Greek /ˈpu periˈmenune/
“Where are they waiting?”

Arvaniti & Ladd, *Phonology*, 2009

Phonetic consequences of phonological underspecification: some syllables are not specified for tone

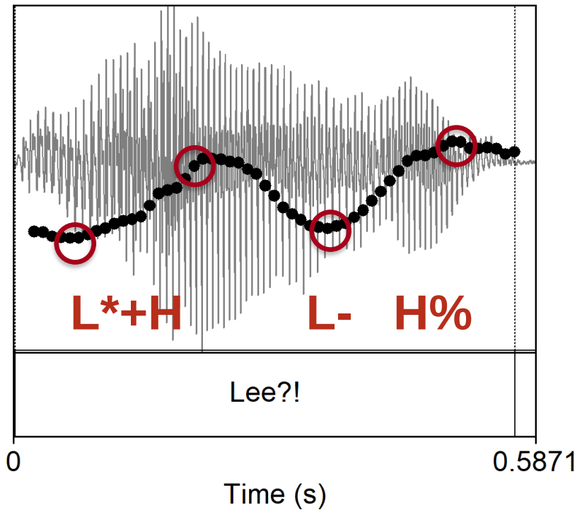


Consequences of underspecification: spreading

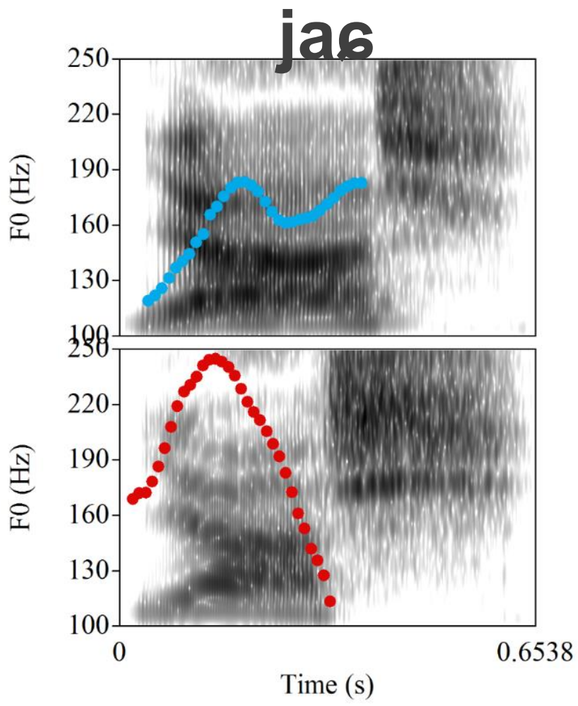


Tonal crowding

- The flip side of underspecification is that in many instances there may be more tones than TBUs
- This is known as tonal crowding
- Languages have a variety of ways of dealing with tonal crowing
  - Deleting or truncating tones (curtailing the tone's excursion)
  - Undershooting (compressing) tones
  - Lengthening the segmental material to fit the tones
  - Using, by preference, a different melody when there is tonal crowding
- No language of those investigated uses one of these methods exclusively (contra Grabe's hypothesis)

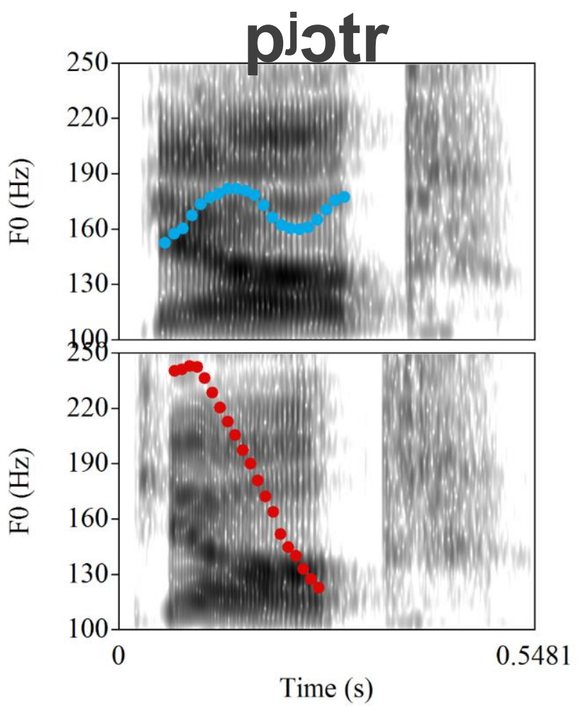


Tonal crowding: Polish calling melodies



Routine call
LH* !H-H%

Urgent call
H* L-L%



The L tone is indispensable in the routine call tune
The L is truncated under extreme tonal crowing in the urgent call, suggesting it is an optional element

Arvaniti, Żygis & Jascula, Phonetica, 2017

What to watch out for when starting a new analysis

- Phonetically, intonation is neither invariable (as some AM analyses suggest), nor more variable and less systematic than segmentals
- If in doubt, think of what you would do if dealing with segments
- No two languages need have the same intonational units
- No two languages need have the same representation for similar pitch contours
  - cf. [k] – a voiceless unaspirated velar plosive – is phonologically classificed as /k/ in French, but as /g/ in English
- The same phonological representation in different languages need not result in identical phonetics
  - cf. the phoneme /s/ is not phonetically realized in the same way in Spanish, English, and Greek
- Languages will show dialectal and contextual variation, that may include neutralization between two tonal categories
- Contrasts in the system (intonational elements) should be posited on the basis of meaning
- The connection between meaning and intonation is arbitrary and many-to-many (so it has to be investigated, not assumed)

Phonological representations, phonetic transparency, and meaning

- Phonological representations of intonation need not always be phonetically transparent
  - the representation of a rising pitch accent need not necessarily involve both a L and H tone
  - different representations are needed if, in the language under analysis, a rising accent conveys different meaning than simply high F0; cf. H* , LH*, L*H in English
  - if such distinctions are not part of the system, the simpler representation should be preferred; in this example, H*
  - Consider: we use /p/, /t/, /k/ in English to phonologically represent the language's voiceless stops, even though we know that their most likely phonetic realization is [p^h], [t^h], [k^h] respectively
- Representations need to capture what is contrastive (no symbolic representations for context-dependent variation; Browman & Goldstein, 1992; Ladd, 2014)
- Criteria for positing phonological categories should include
  - phonetic evidence (cf. Arvaniti, 2016; Arvaniti et al., 2017)
  - phonological considerations (cf. Gussenhoven, 2016)
  - meaning (cf. Prieto & Borràs-Comes, 2018; González et al., 2017), though not loosely applied functional categories or paralinguistic notions used in lieu of pragmatic meaning distinctions

Phonetic realization

- F0 is the main but not the only exponent of intonation
- Intonational categories may be realized by multiple cues in trading relations (a matter of empirical investigation; e.g. Lohfink, Katsika, Arvaniti, 2019)
- Intonational categories are likely to overlap and show extensive phonetic variability
  - across languages (cf. Fletcher, 2015, on Dalabon, Fletcher et al., 2016, on Mawng, Maskikit-Essed & Gussenhoven, 2016, on Ambonese Malay)
  - within languages (cf. Arvaniti, Żygis, Jaskuła, 2017)
  - across speakers (Niebuhr et al., 2011; Grice et al., 2017)

Processing

- The identification of intonation categories during processing is probabilistic and context dependent
  - local phonetic detail and cue-weighting (cf. Barnes et al., 2014; Dorokhova & D'Imperio, 2019)
  - distal tonal context, including the realization of other tonal events in the utterance (cf. Dilley & McAuley, 2008)
  - pragmatic context (e.g. Calhoun, 2010; Baltazani, Gryllia, Arvaniti, 2019; Roettger & Franke, 2019; Gryllia, Baltazani, Arvaniti, 2019)
    - *my name is Amalia* with a final rise, when
      - (i) I am speaking to my new department's secretary
      - (ii) when I wake up from a coma after a car accident

On to Praat!