Prosody: speech rhythms and melodies

6. Speech Timing

Dafydd Gibbon

Guangzhou Prosody Lectures, November 2016

Schedule

Week 1:

01 *Forms and functions of prosody: models and methods*

Nov. 2 (Wednesday) 2:30pm--4:30pm

02 *Forms and functions of prosody: prosodic semantics*

Nov. 4 (Friday) 10am--12pm

Week 2:

03 *Basics of digital phonetics*

Nov. 8 (Tuesday) 10am--12pm

04 *Pitch Stylisation: tone and intonation*

Nov. 8 (Wednesday) 10am--12pm

Week 3:

05 *Syllables and Prosody Modelling*

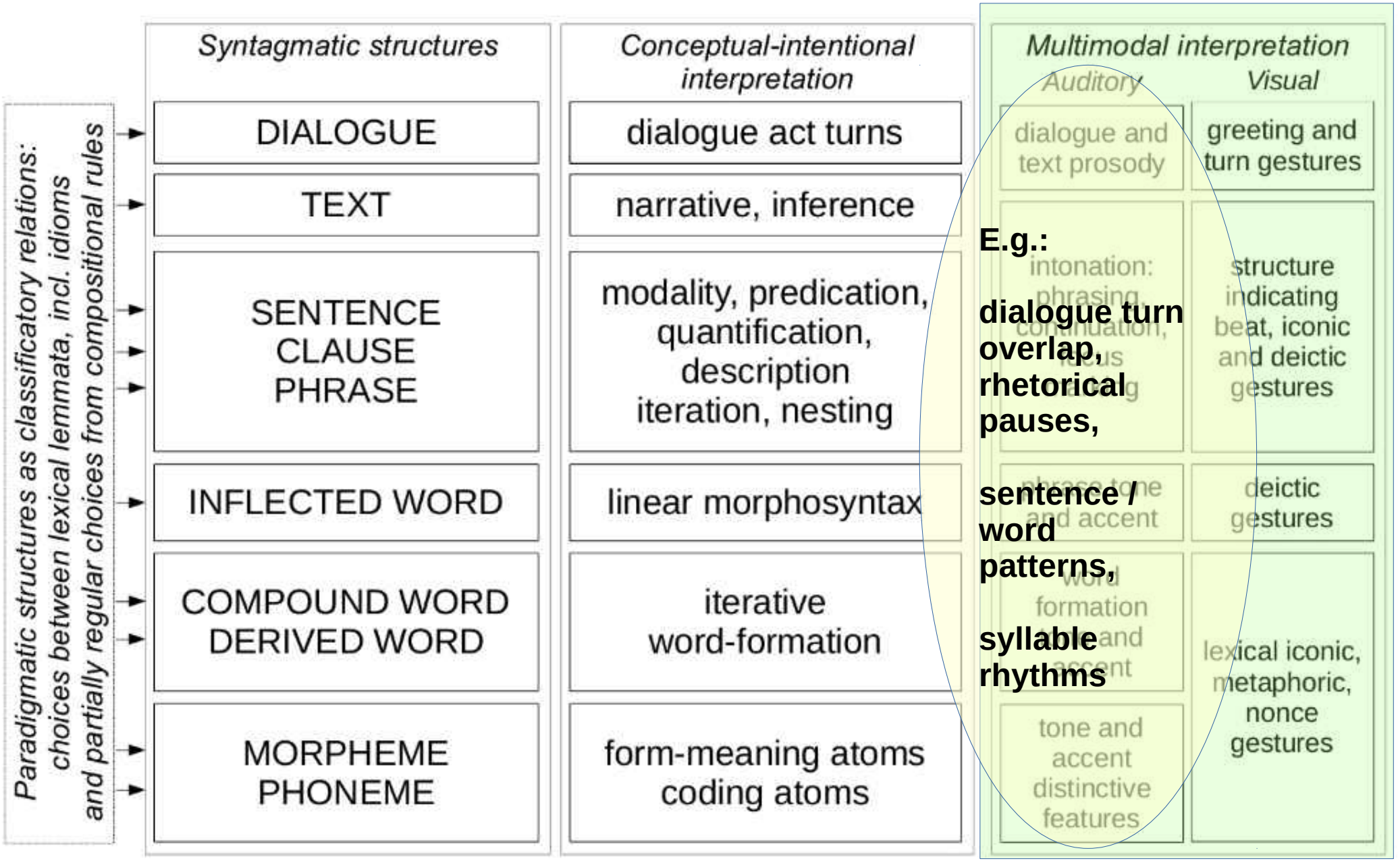
Nov.15 (Tuesday) 10am--12pm

06 ***SpeechTiming: durations and rhythm***

Nov.15 (Tuesday) 2:30pm--4:30pm

Speech Timing: Ranks from Dialogues to Phonemes

Multilinear Grammar – Ranks and Interpretations (MLG-RI) Architecture



Time concepts in linguistics

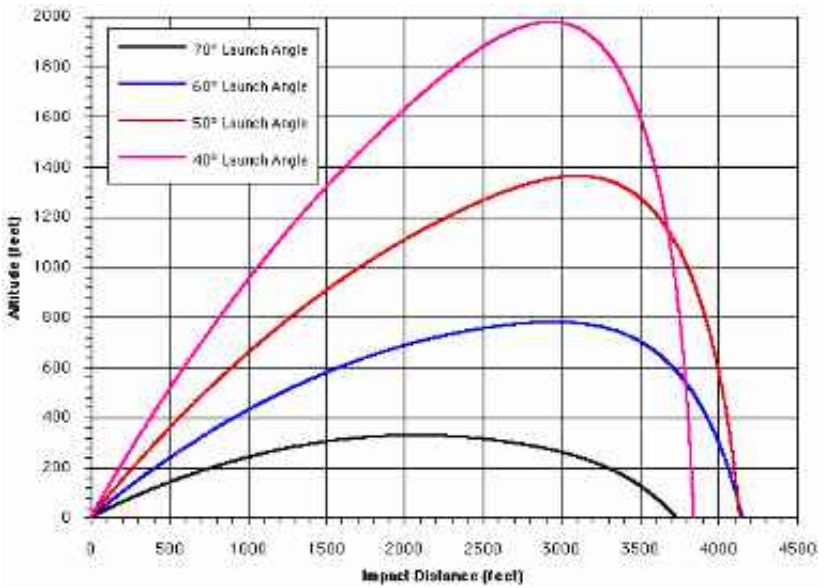
- The time dimension (1)
  - Time in pragmatics:
    - pause type (filled – unfilled), length:
      - disfluency
      - hesitation
      - rhetorical
    - tempo (speed) of speech
    - synchronisation: sequence and ovelap of dialogue contributions
  - Time in semantics:
    - tense, aspect,
    - temporal adverbs and prepositional phrases
    - temporal nouns (including dates) and verbs
  - Time in discourse, utterance, sentence and word grammar:
    - word order
    - parallel patterns of words, prosody, gestures

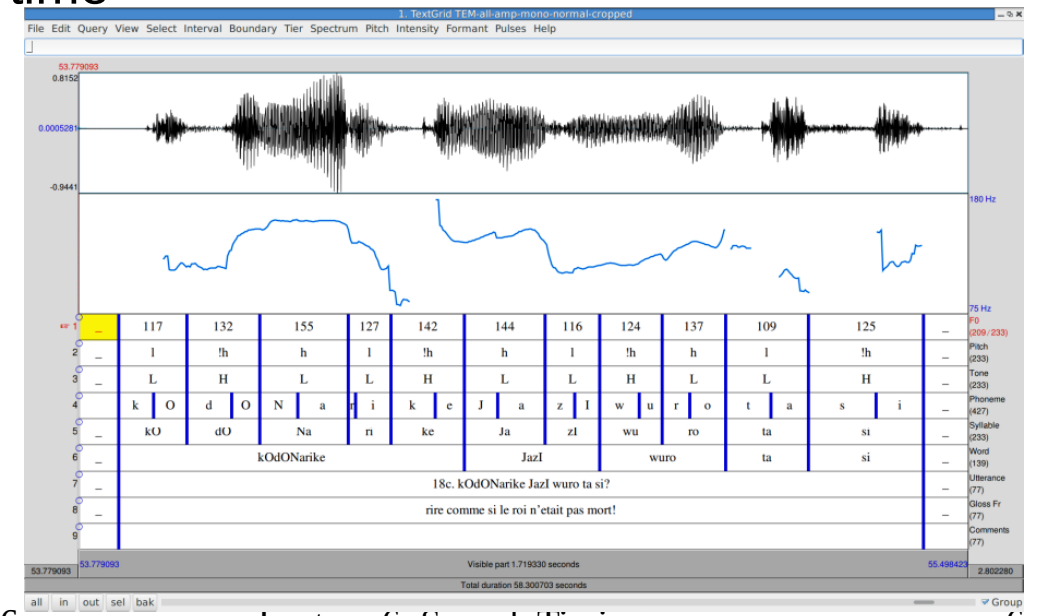
Time concepts in linguistics

- The time dimension (2)
  - Time in phonology:
    - lexical phonemic contrast: long-short vowels and (geminate) consonants
    - phoneme, syllable, foot, intonation phrase order
    - parallel patterns in prosodic phonologies (e.g. autosegmental phonology)
    - regular rhythm-like patterns in metrical phonology
  - Time in phonetics:
    - sequence of phones
    - coarticulation of phones with neighbouring phones (assimilation, dissimilation)
    - parallel features:
      - consonants: place of articulation, manner of articulation, nasality, voicing
      - vowels: high-mid-low, front-centre-back, rounded, nasal
      - articulation rate – fundamental frequency – pitch
- Methodological perspectives:
    - static (study of features at points/intervals in time)
    - dynamic (study of trajectories of features in time)

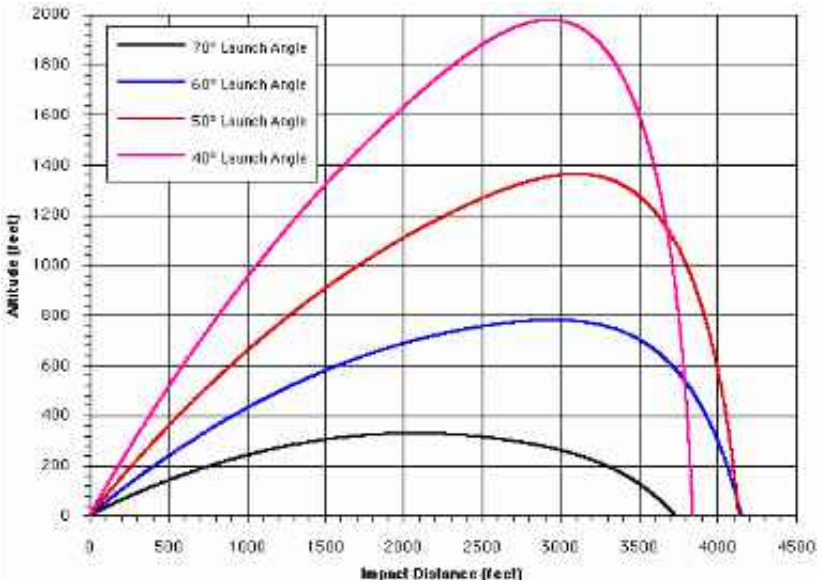
Perspectives on time in phonetics

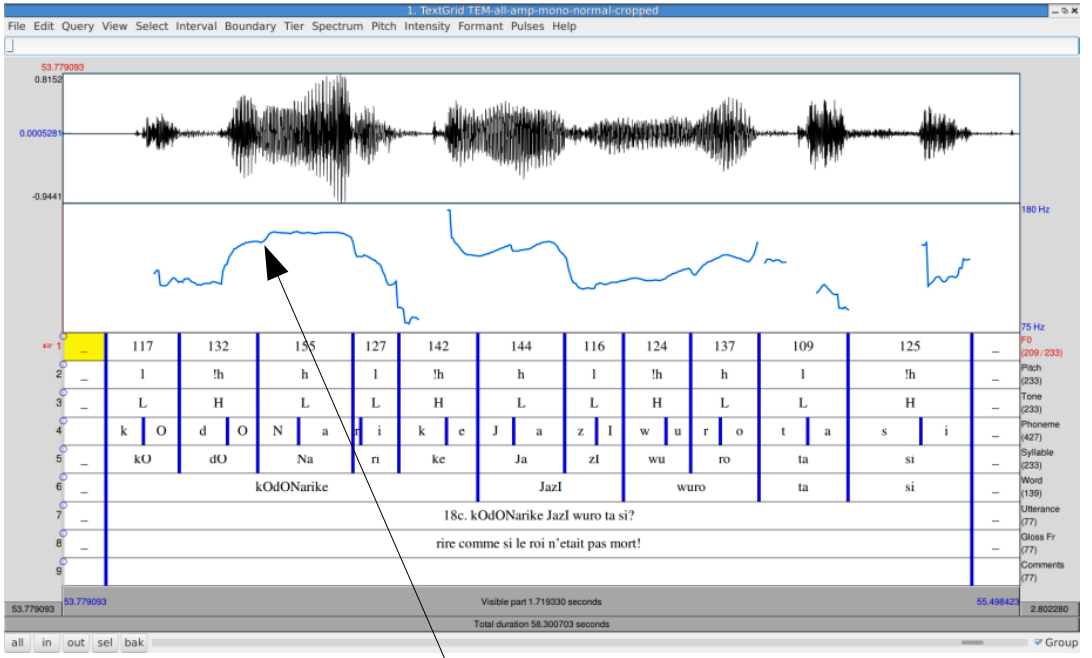
- The time dimension (3)
  - Methodological perspectives:
    - **static**
      - study of feature spaces
        - at points in time
        - in intervals in time
    - **dynamic**
      - study of trajectories through feature spaces
        - through intervals in time

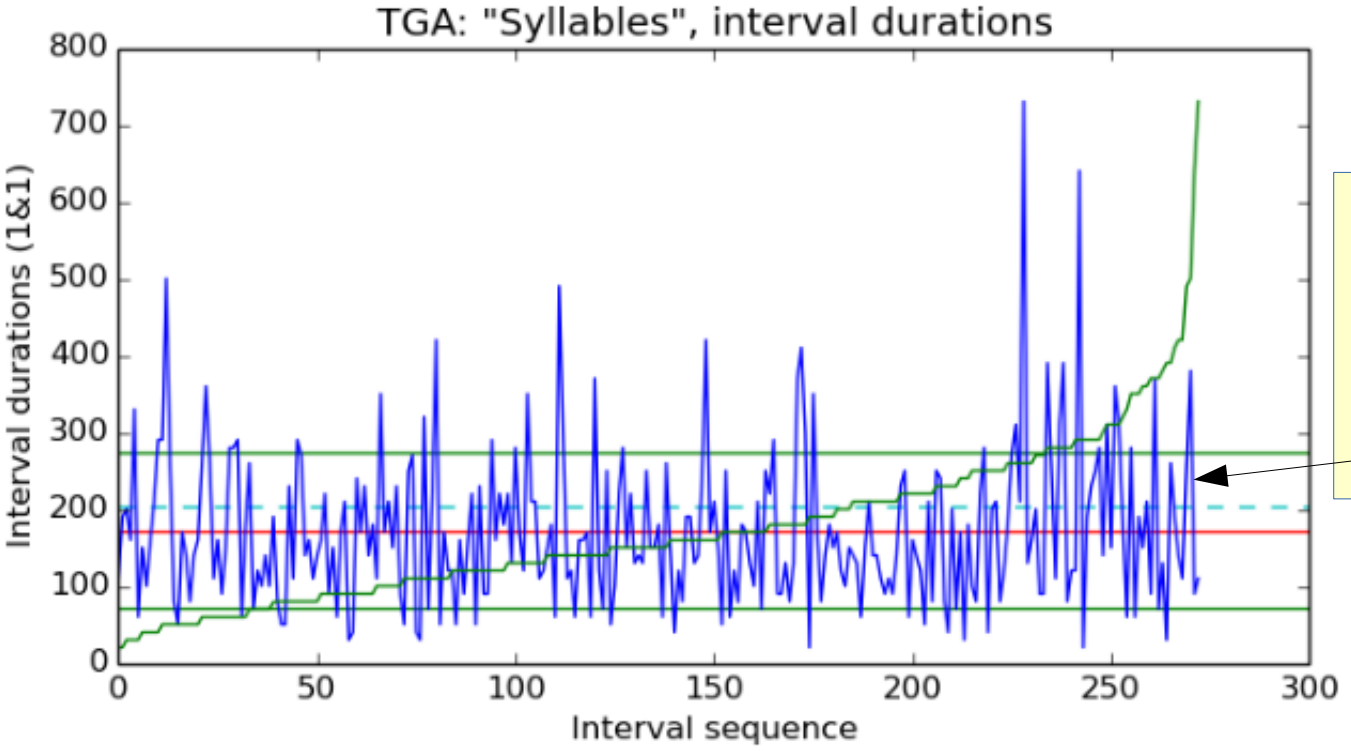




Perspectives on time in phonetics







Trajectory visualization

pitch trajectory

syllable duration trajectory

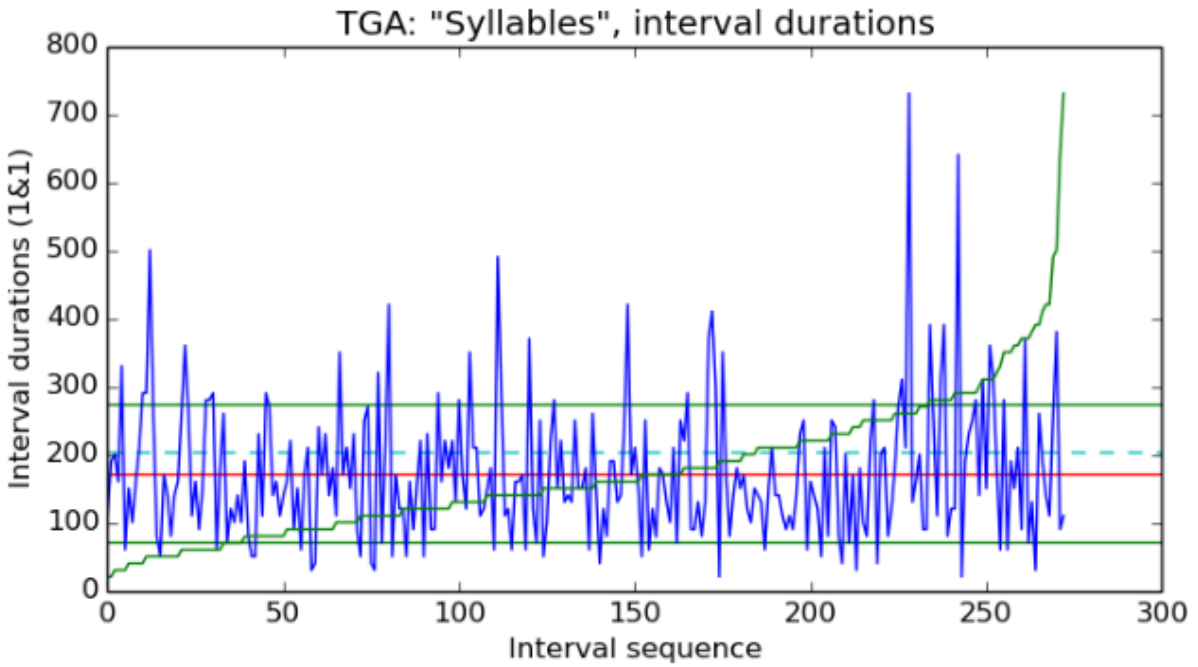
Perspectives on time in phonetics

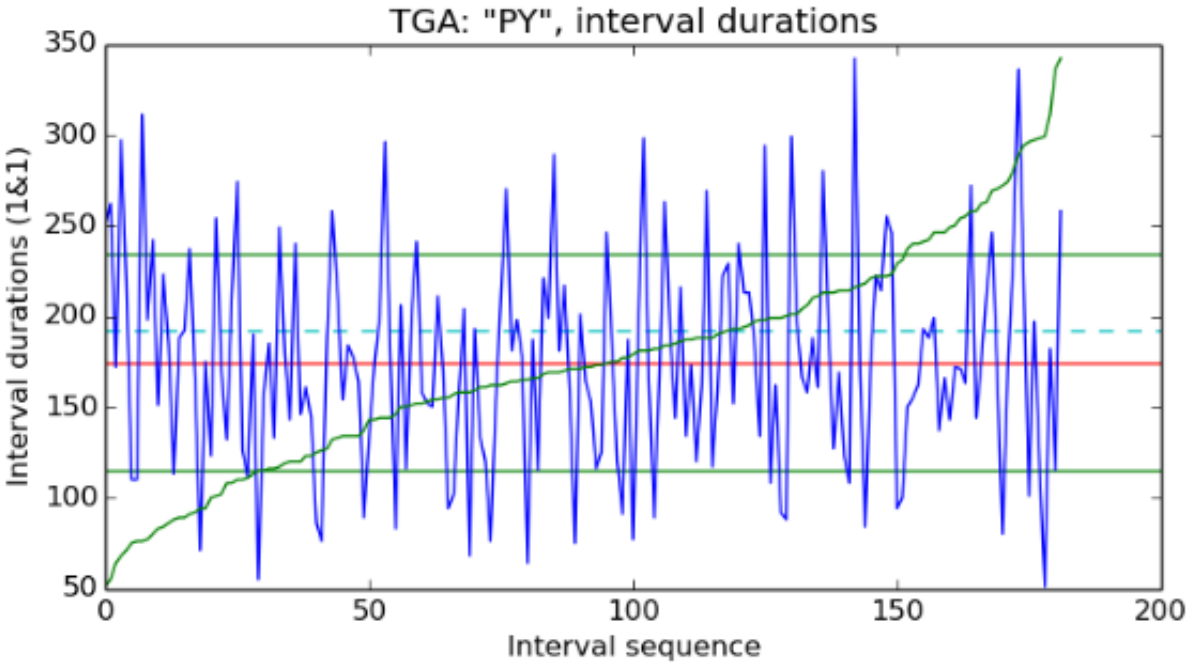
English

Trajectory visualization

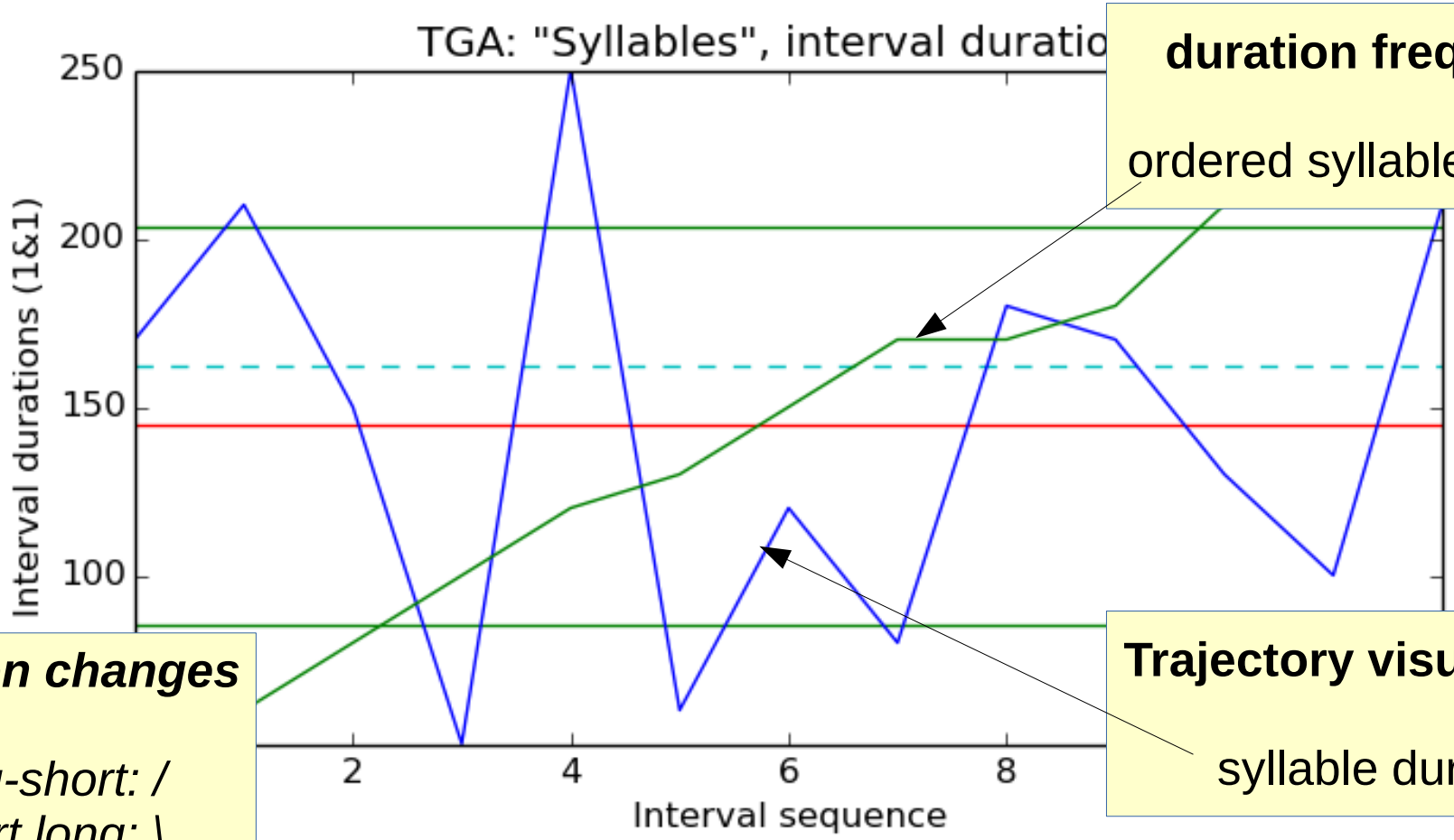
syllable durations

Mandarin





Perspectives on time in phonetics



duration frequencies

ordered syllable durations

duration changes

long-short: /
*short long: *
equal:

Trajectory visualizations

syllable durations

/ / \ / \ / \ = / / \

I t s ' m e n t @ z @ ' s I m b l @ v , r e k @ n , s I l I ' eI S n

170 210 150 50 250 60 120 79 180 171 130 100 210

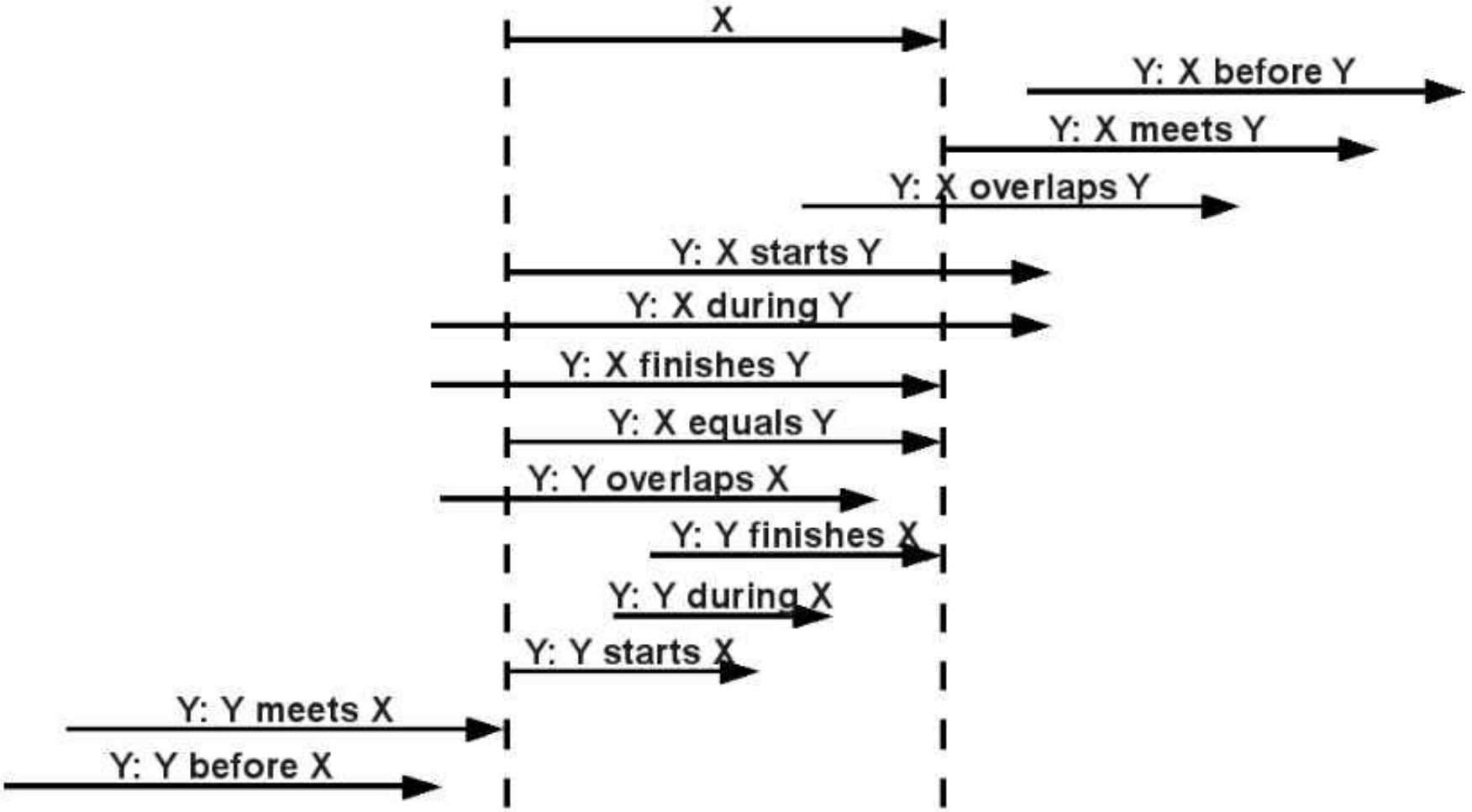
Timing: Basic Characteristics

- Timing:
  - either a property of an event
    which is a relation between a state and an interval:
    event = <state, interval>
    - e.g.
      - the state and the interval of speaking an utterance
      - the state and the interval of producing a vowel
  - or one of 13 possible relations
    between two (or more) intervals / events
    - which are in sequence
    - which overlap

The Allen Interval Calculus

- Timing:

  13 possible relations between the intervals of all kinds of events:



The Allen Interval Calculus

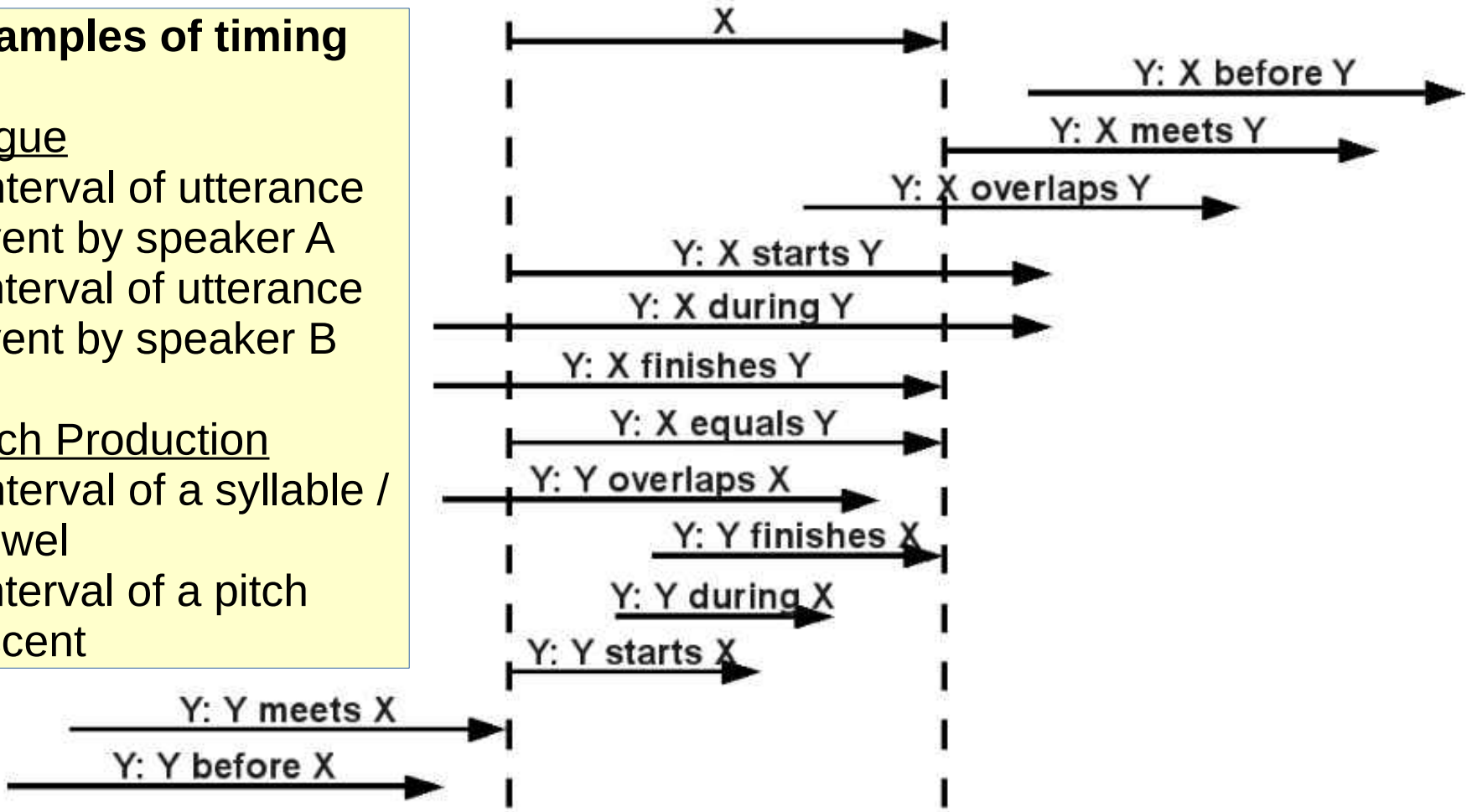
- Timing:

  13 possible relations between the intervals of all kinds of events:

Examples of timing

Dialogue
X = interval of utterance event by speaker A
Y = interval of utterance event by speaker B

Speech Production
X = interval of a syllable / vowel
Y = interval of a pitch accent



Timing and Rhythm

What is Rhythm?

- Oscillation
- Regular timing
  - syllable timing
  - foot / stress timing
  - mora timing
- Variation of regular timing:
  - irregularity
  - syncopation
  - acceleration/deceleration
- Strong – Weak alternation
- Speech-Pause ratios

A fundamental issue

Is rhythm an objective, physical fact?

Or is rhythm a pattern which our perception imposes on physical events?

An abstract model of rhythm

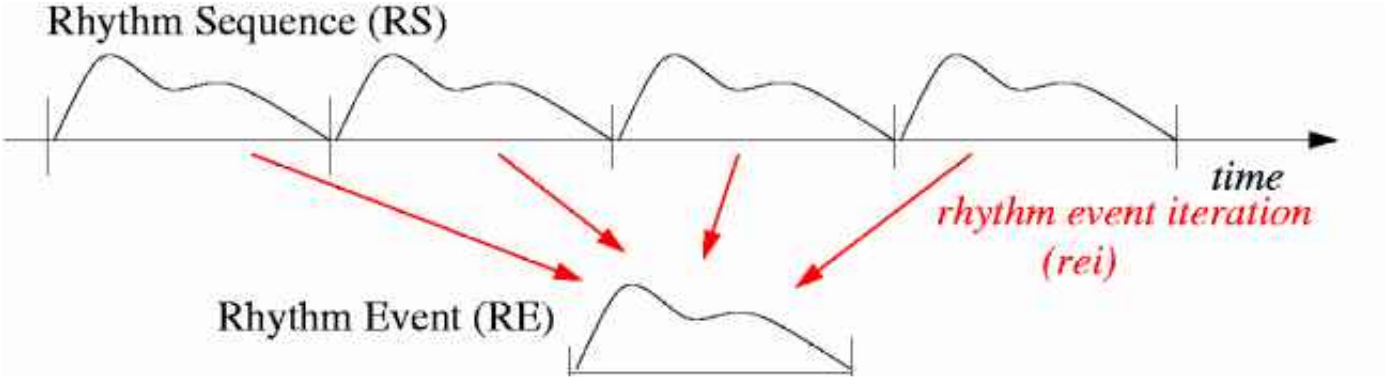
Components of rhythmic oscillation of syllables:

- State:
  - Strong-Weak state (e.g. C-V segment; high-low or level-contour pitch)
  - Long-Short interval

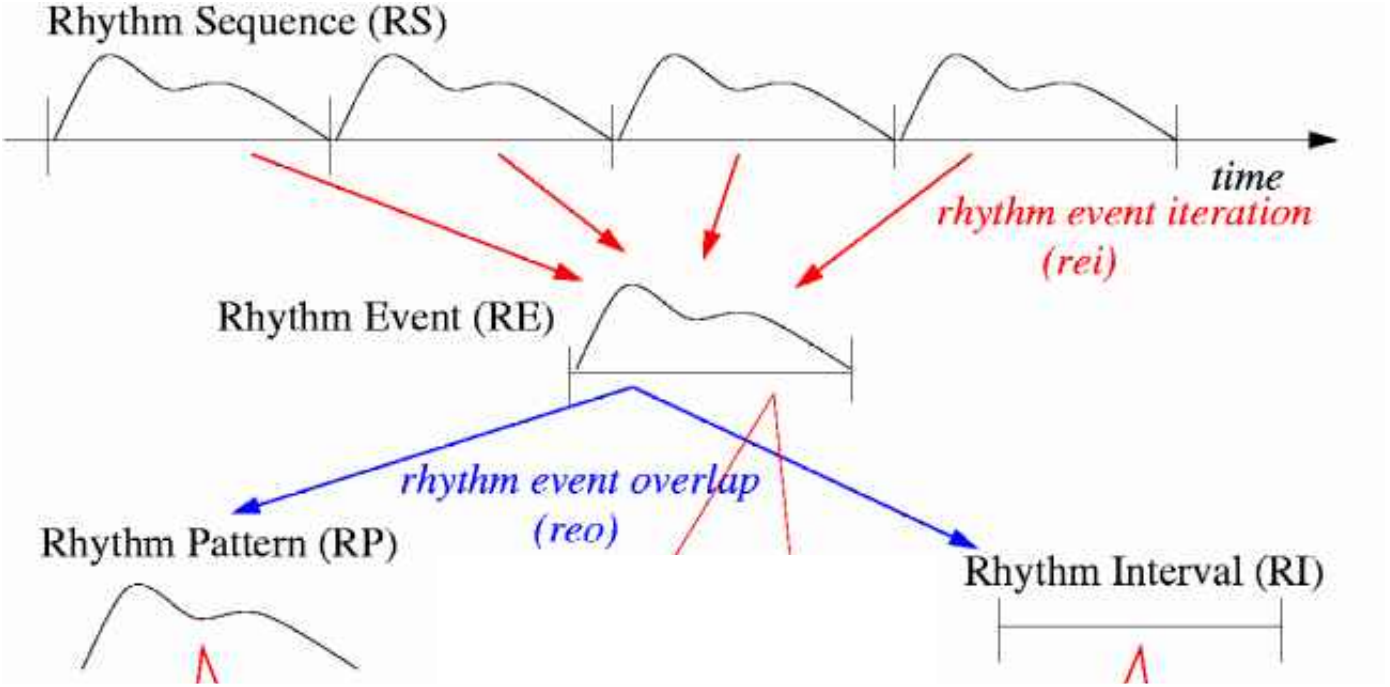
- Sequence of states:
  - Strong-Weak state sequence
  - Long-Short interval sequence

Gibbon, Dafydd (2006). Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In: Sudhoff, Stefan, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter and Johannes Schließer, eds. (2006). *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, 281-209.

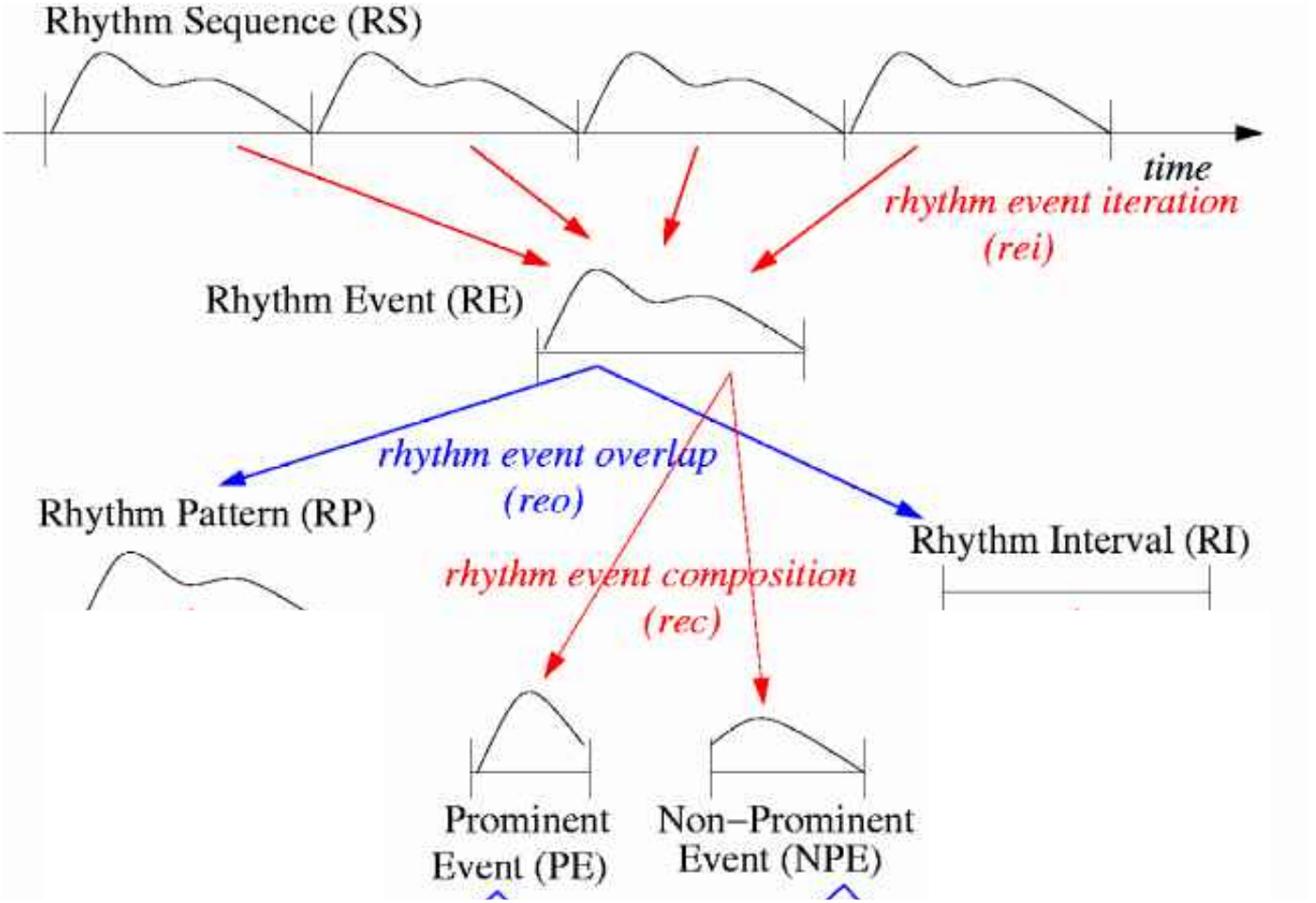
An abstract model of rhythm



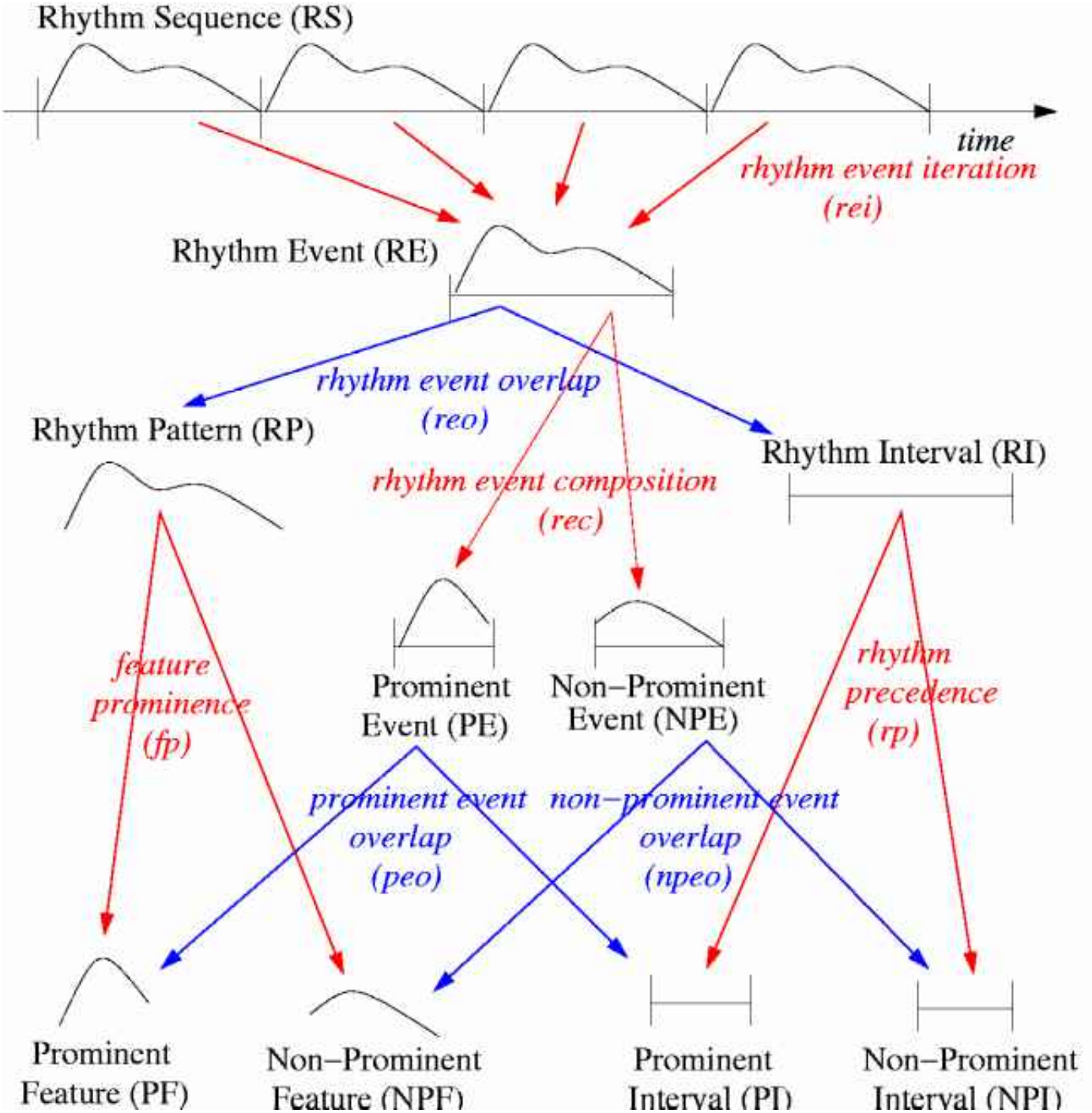
An abstract model of rhythm



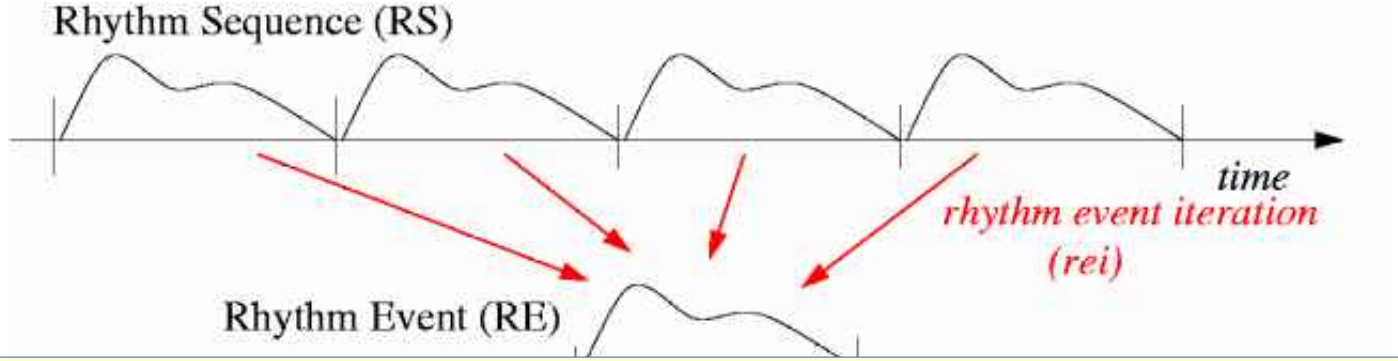
An abstract model of rhythm



An abstract model of rhythm



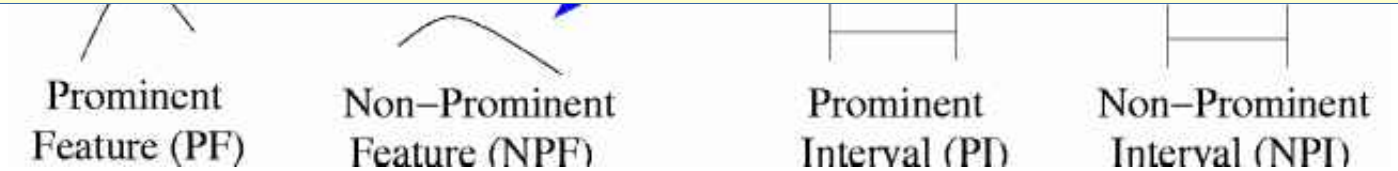
An abstract model of rhythm



Components of rhythmic oscillation of syllables:

- State:
  - Strong-Weak state (e.g. C-V segment; high-low or level-contour pitch)
  - Long-Short interval
- Sequence of states:
  - Strong-Weak state sequence
  - Long-Short interval sequence

Gibbon, Dafydd (2006). Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In: Sudhoff, Stefan, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter and Johannes Schließer, eds. (2006). *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter, 281-209.



Timing domains

and the prosodic hierarchy

Domains of speech timing: from dialogues to phonemes

- Dialogue timing
- Utterance timing
- Phrase timing
- Word timing
- Syllable timing
- Phoneme timing

X

- Duration
  - contrastive
  - rhetorical
- Rate
  - speed
  - acceleration
- Rhythm
  - internal
  - external

Hierarchical domains of speech timing: a closer look

Dialogue

- turn sequence & overlap
- synchronisation / entrainment)

Utterance (Utt)

Intonational Phrase (IP)

Phonological phrase (PhP)

Foot (φ)

Syllable (σ)

Mora (μ)

Phoneme (‘segment’, seg)

Conventional Prosodic Hierarchy domains

MLG-RI domains

Subphoneme

- phases of voiceless plosive/stop production: closing – silence – opening; affricates; diphthongs

MLG: Multilinear Grammar
RI: Rank Interpretation architecture

Hierarchical domains of speech timing: a closer look

- Dialogue timing:
  - Dialogue act sequences, turn sequences
  - Turn sequence and overlap, synchronisation and entrainment
- Utterance timing:
  - Phones/syllable rate (standard deviation, Pairwise Variability Index
  - Acceleration and deceleration of speech rate; rhythm
  - Pauses; timing of information structure - focus, emphasis, contrast
  - Turn structure: rhetorical pause, hesitation, disfluency
- Phrase timing:
  - Timing of parts of speech (lexical words, functional words); boundary effects, grammatical structure
  - Timing of pitch accents, ‘phrase accents’, ‘sentence stress’
- Word timing:
  - Strong and weak syllables, primary, secondary ‘word stress’; ‘foot timing’ rhythms
  - Prosodic morphology: compound words and stems and affixes of derived words
- Syllable timing:
  - Onsets, Nuclei and Codas; Moras (*morae*); contrastive V and C length
  - syllable rate, syllable timing rhythms
  - Phones and sub-phone units: rate; stop closure-pause-release; Voice Onset Time

The Prosodic Hierarchy: a finite depth grammar

Utterance (Utt): constituent of turn-taking, Q&A etc.

Intonational Phrase (IP): boundary tones, association with grammatical phrase

Phonological phrase (PhP), Intermediate Phrase (ip): phrase boundary tone, domain of phrase stress

Phonological word, Prosodic Word (PW, PrWd, ω): domain of word stress, prosodic morphology, clitics

Foot (φ): Domain of primary, secondary, fixed stress, prosodic morphology

Syllable (σ): main local timing unit; phonotactic patterns, stress-bearing unit

Mora (μ): tone placement, phonotactic patterns

Segment: smallest 'leaf' element in prosodic hierarchy

Subsegment: affricates, diphthongs; (phonetic: stop closure-pause-release, Voice Onset Time)

The Prosodic Hierarchy: a finite depth grammar

Prosodic Hierarchy (PH) inventory and ordering:

PC = {***Utt***, ***IP***, ***PhP***, ***PrWd*** (**ω**), ***Ft*** (**φ**), ***syll*** (**σ**), **mora** (**μ**), **seg**}

Structural constraints on PH:

< l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8> (note also subsegmental units)

Strict Layering Hypothesis: ***PC***s at L_i dominate only ***PC***s at L_{i+1}

- Fixed depth (no recursion): No ***PC*** at L_i dominates a ***PC*** at L_i
- Exhaustivity: All ***PC***s at L_i are dominated by a single ***PC*** at L_{i-1}

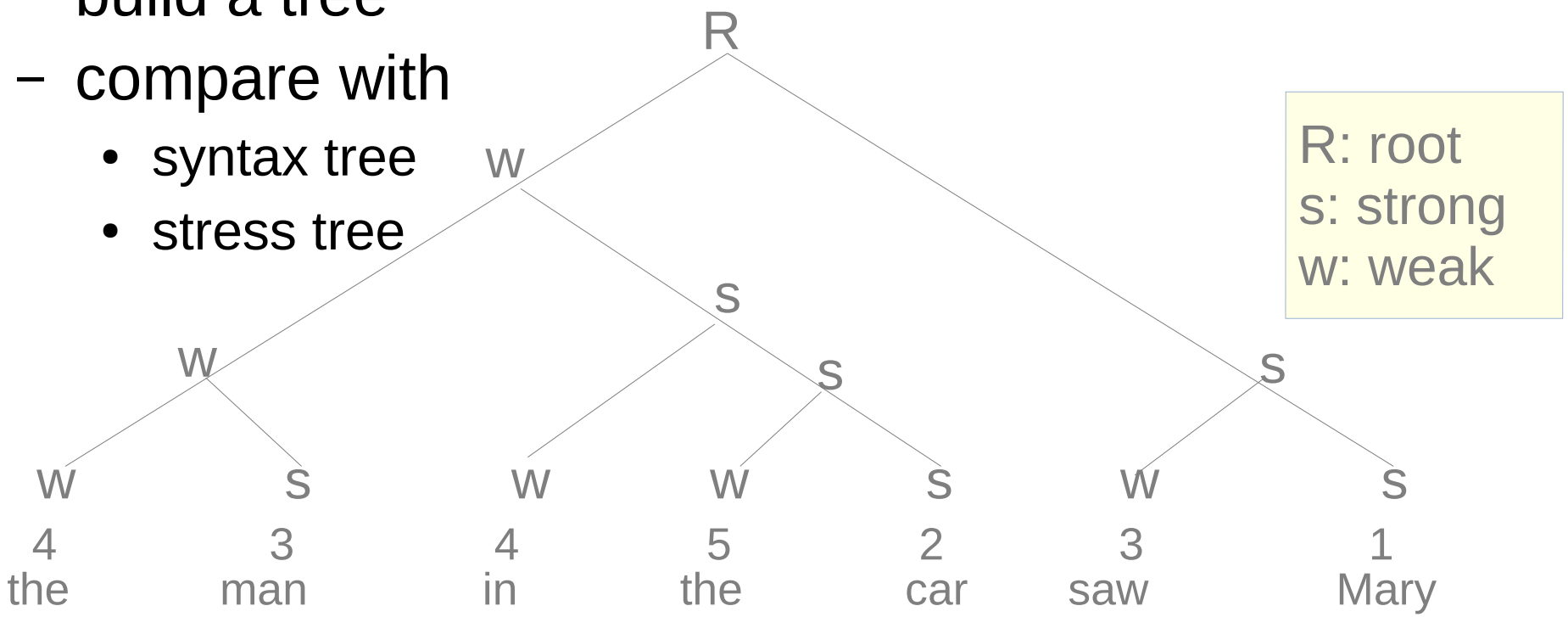
Headedness:

- Every ***PC*** at L_i immediately dominates a ***PC*** at L_{i+1}

Selected top-down domains: sentence

Timing and stress hierarchies

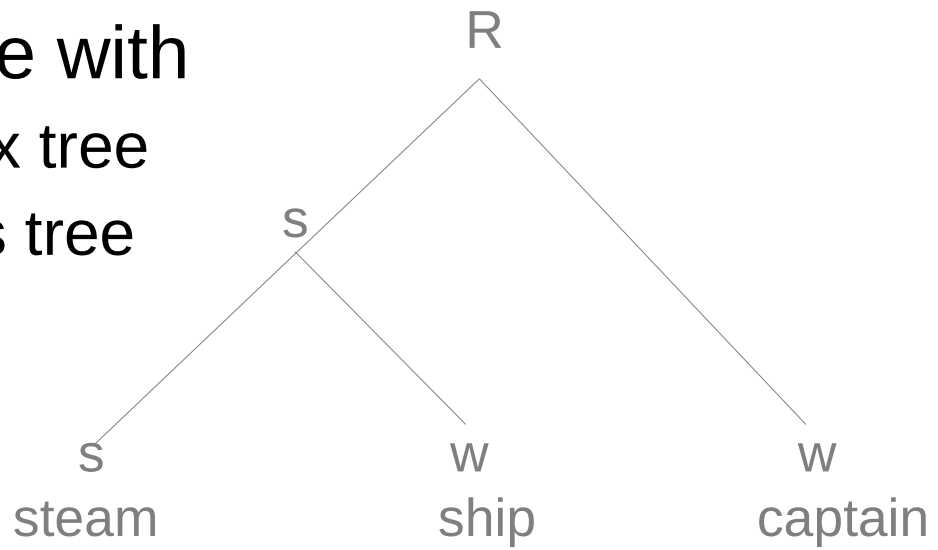
- Syllable timing tends to follow the Nuclear Stress hierarchy
- Time Tree analysis:
  - compare durations of neighbouring syllables
  - select criterion (longest/shortest item last/first)
  - build a tree
  - compare with
    - syntax tree
    - stress tree



Selected top-down domains: word

Timing and stress hierarchies

- Syllable timing also tends to follow the Compound Stress hierarchy
  - approximately: reversal of Nuclear Stress hierarchy
- Time Tree analysis:
  - compare durations of neighbouring syllables
  - select criterion (longest/shortest item last/first)
  - build a tree
  - compare with
    - syntax tree
    - stress tree



R: root
s: strong
w: weak

PHONETIC APPROACHES: STATIC TIMING MODELS

1. Tones and syllable durations

What are we looking for?

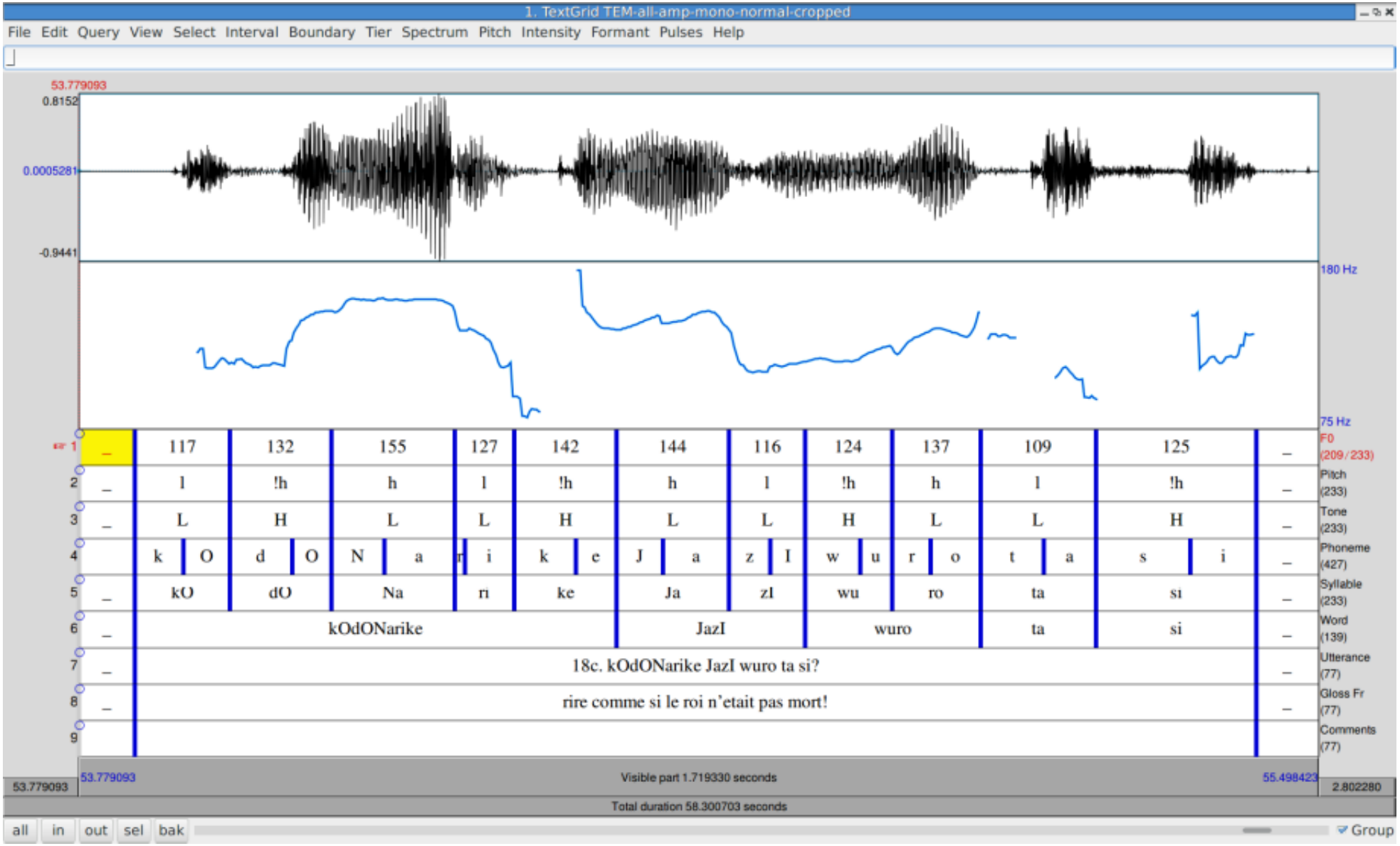
- Try to prove or disprove a ‘null hypothesis’

- Null hypotheses:
  - if
    - all syllables have the same duration
      (syllable timing)
    - all tones are associated with syllables
      (the syllable is the tone-bearing unit, TBU)
  - then
    - **all tones have the same duration**

- Right or wrong?

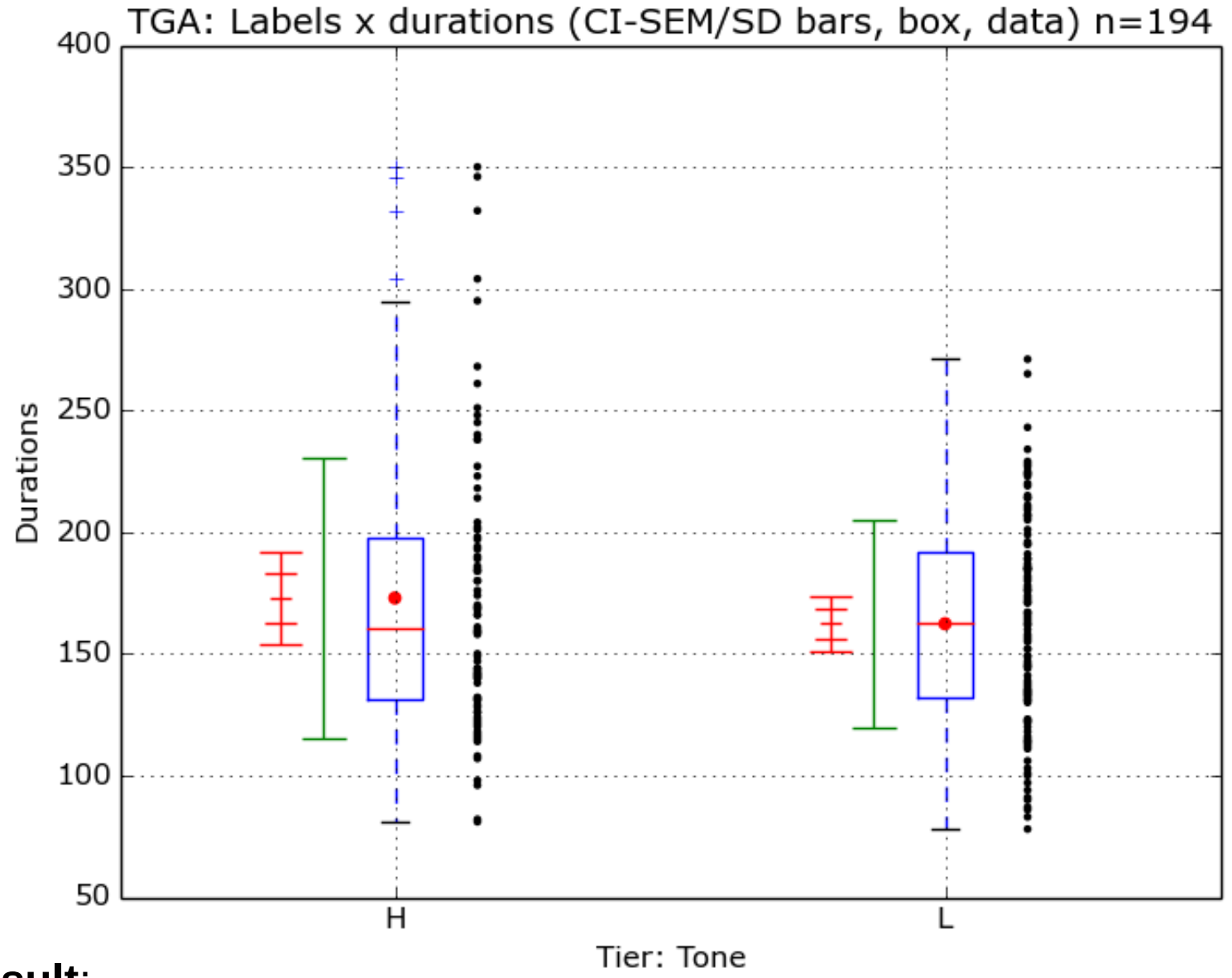
1.1. The H and L tones of Tem (Niger-Congo > Gur)

The H and L tones of Tem (Niger-Congo > Gur)



TEM kodoNa

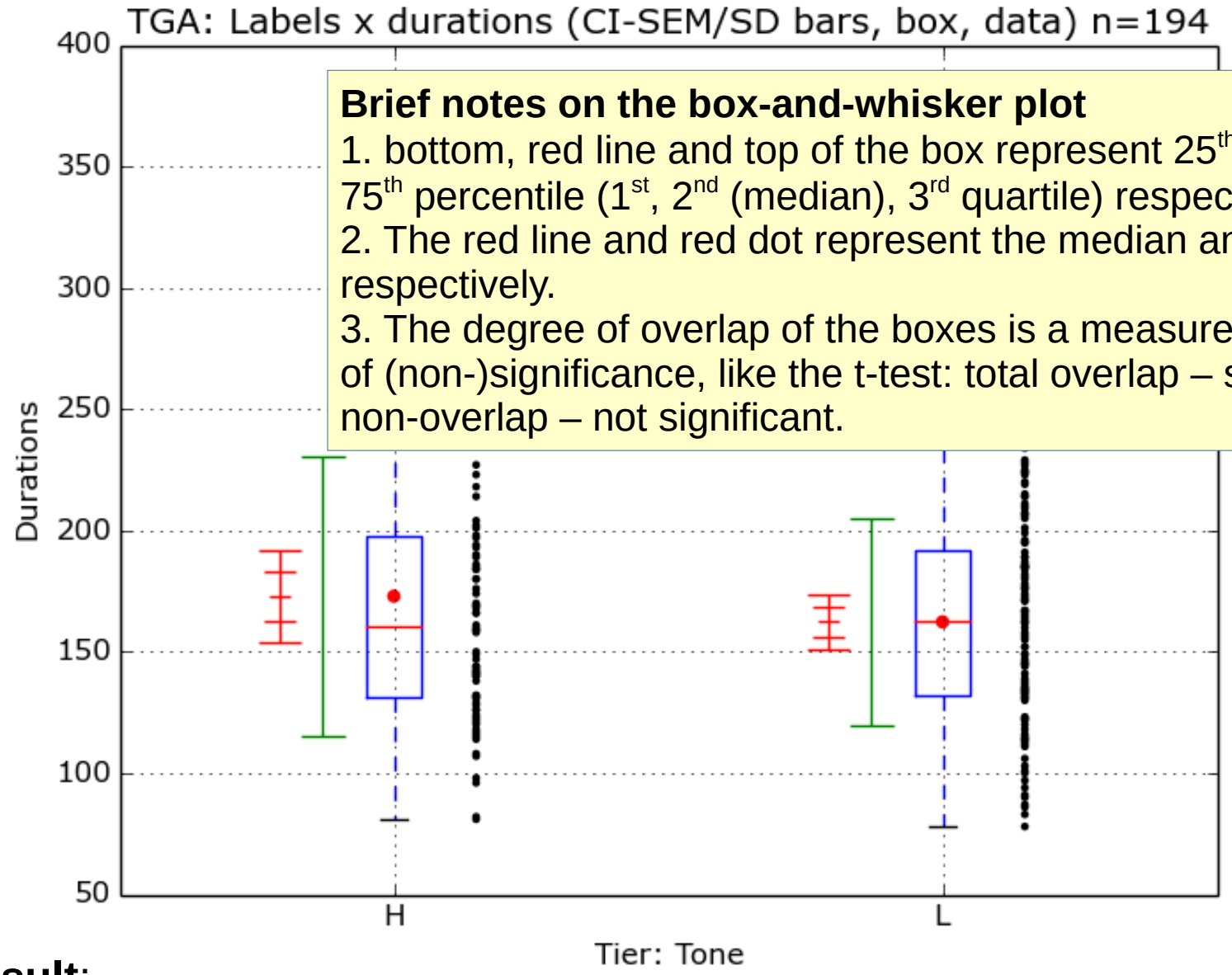
The H and L tones of Tem (Niger-Congo > Gur)



Result:

- Boxes overlap strongly, thus no significant difference.
- So the null hypothesis is not refuted:
  H and L syllables tend to have the same length.

The H and L tones of Tem (Niger-Congo > Gur)



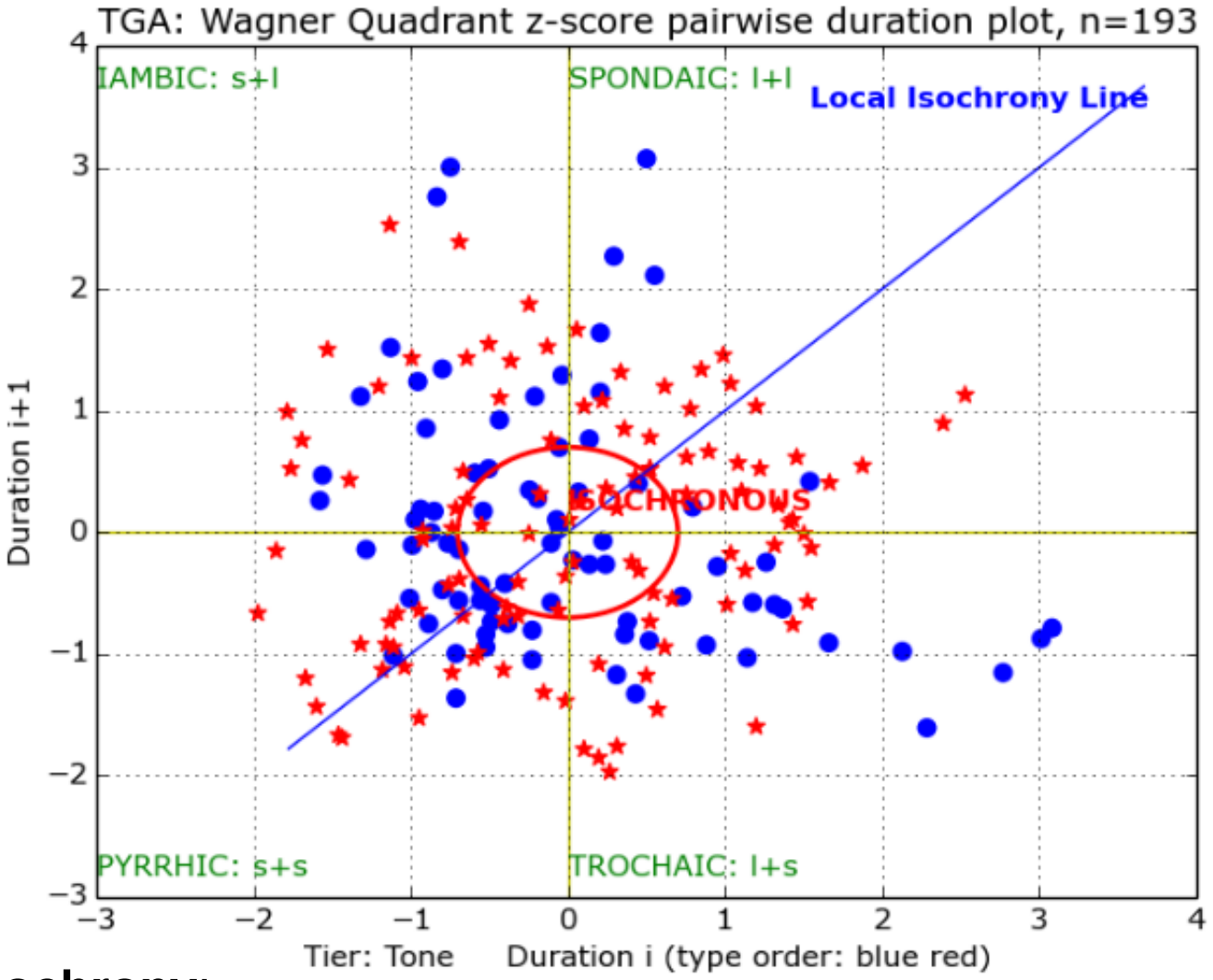
Brief notes on the box-and-whisker plot

1. bottom, red line and top of the box represent 25th, 50th (median), 75th percentile (1st, 2nd (median), 3rd quartile) respectively.
2. The red line and red dot represent the median and the mean respectively.
3. The degree of overlap of the boxes is a measure of the degree of (non-)significance, like the t-test: total overlap – significant; total non-overlap – not significant.

Result:

- Boxes overlap strongly, thus no significant difference.
- So the null hypothesis is not refuted:
  H and L syllables tend to have the same length.

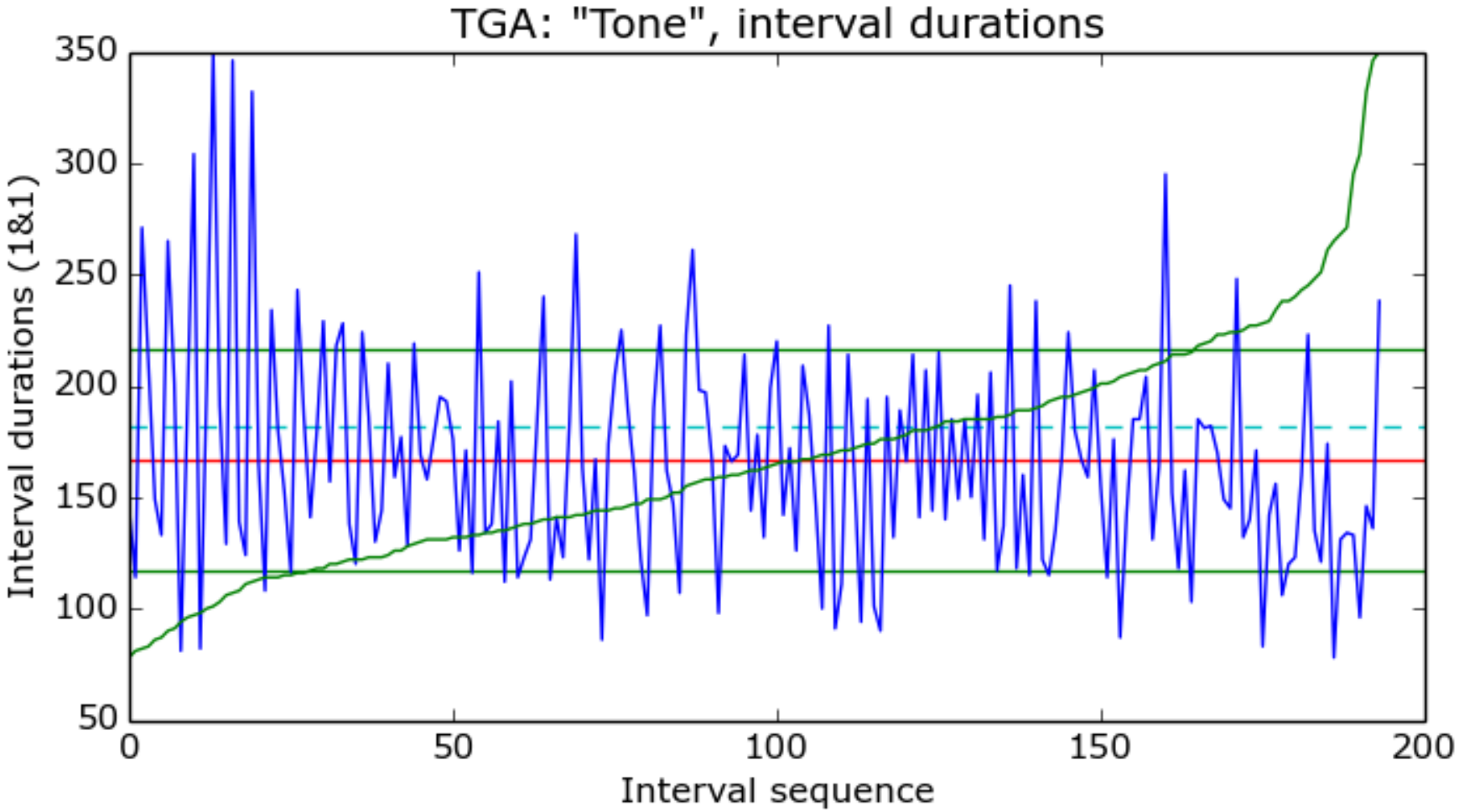
Wagner Quadrant for Tem tones (Niger-Congo > Gur)



Quasi-isochrony:

- L tone: blue, H tone: red
- Durations of neighbouring syllable pairs tend to be scattered randomly close to zero in the normalised z-scores (zero is the mean duration)
- For details see the box plot.

The H and L tones of Tem (Niger-Congo > Gur)

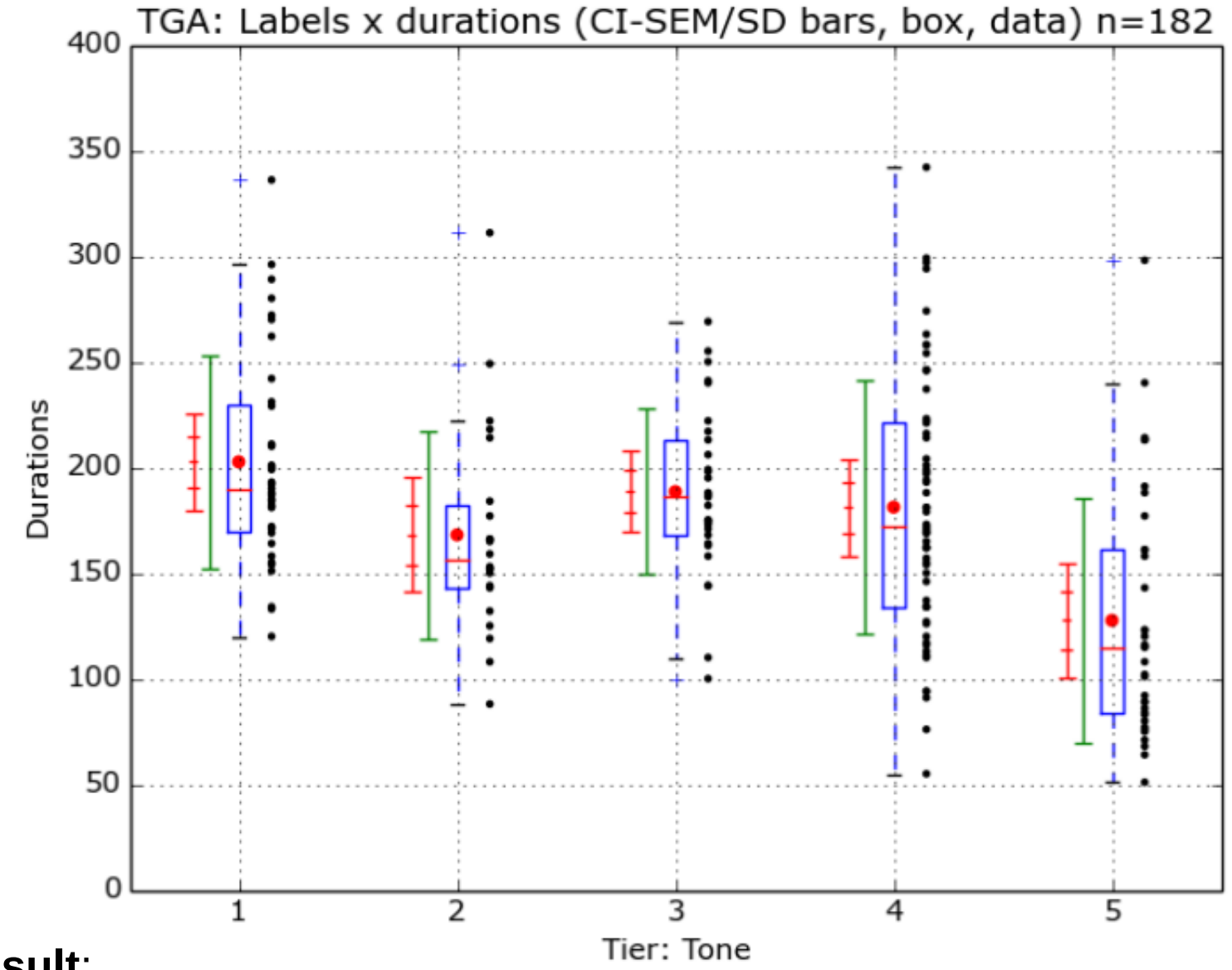


Durations:

- Durations are similar and vary randomly around the mean.
- As the speech session continues, syllables get shorter and shorter (i.e. the syllable rate gets faster)

1.2. The 5 tones of Mandarin (1, 2, 3, 4 and neutral)

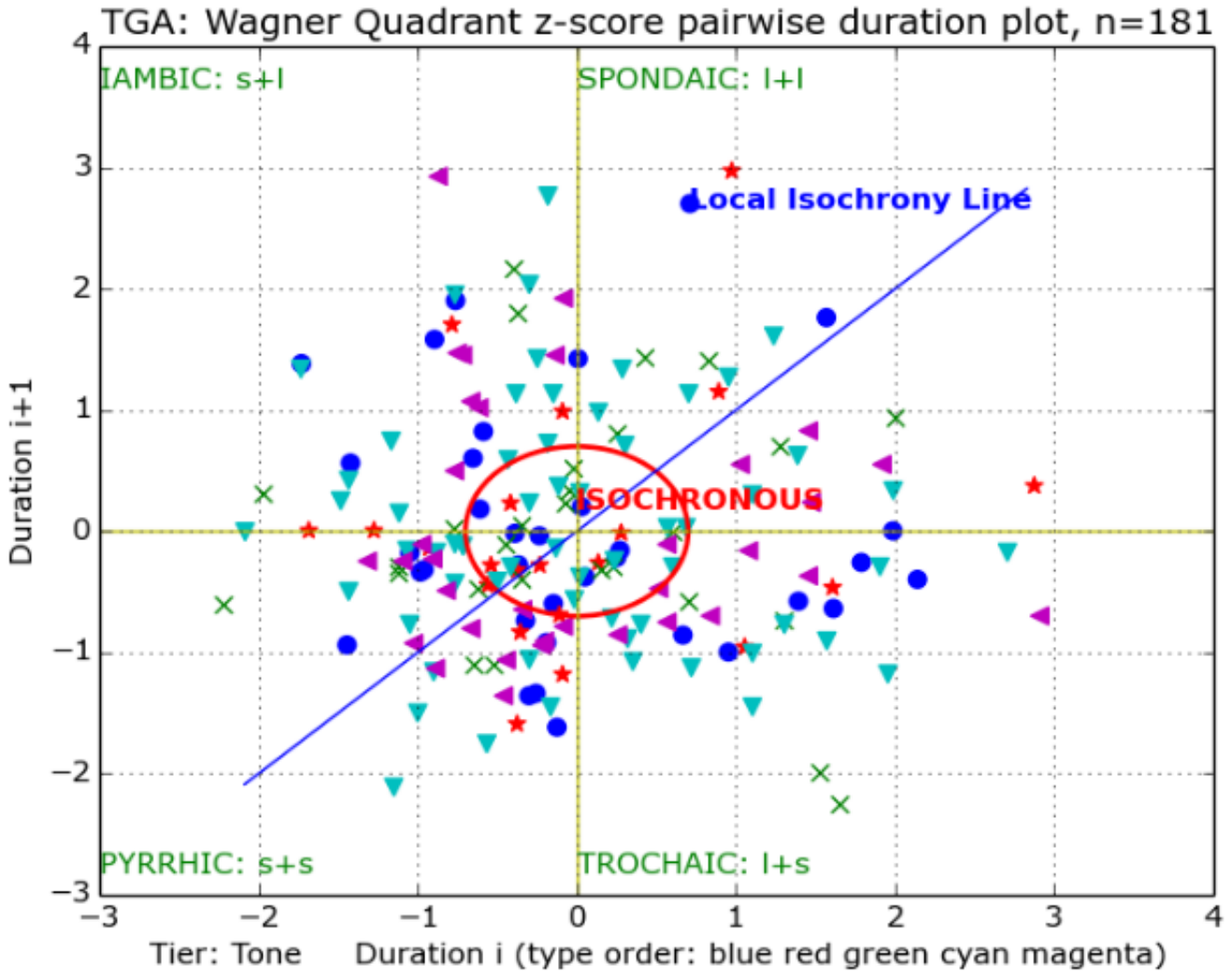
The 5 tones of Mandarin (1, 2, 3, 4 and neutral)



Result:

- Boxes overlap strongly, thus no significant difference.
- So the null hypothesis is not refuted:
  H and L syllables tend to have the same length.

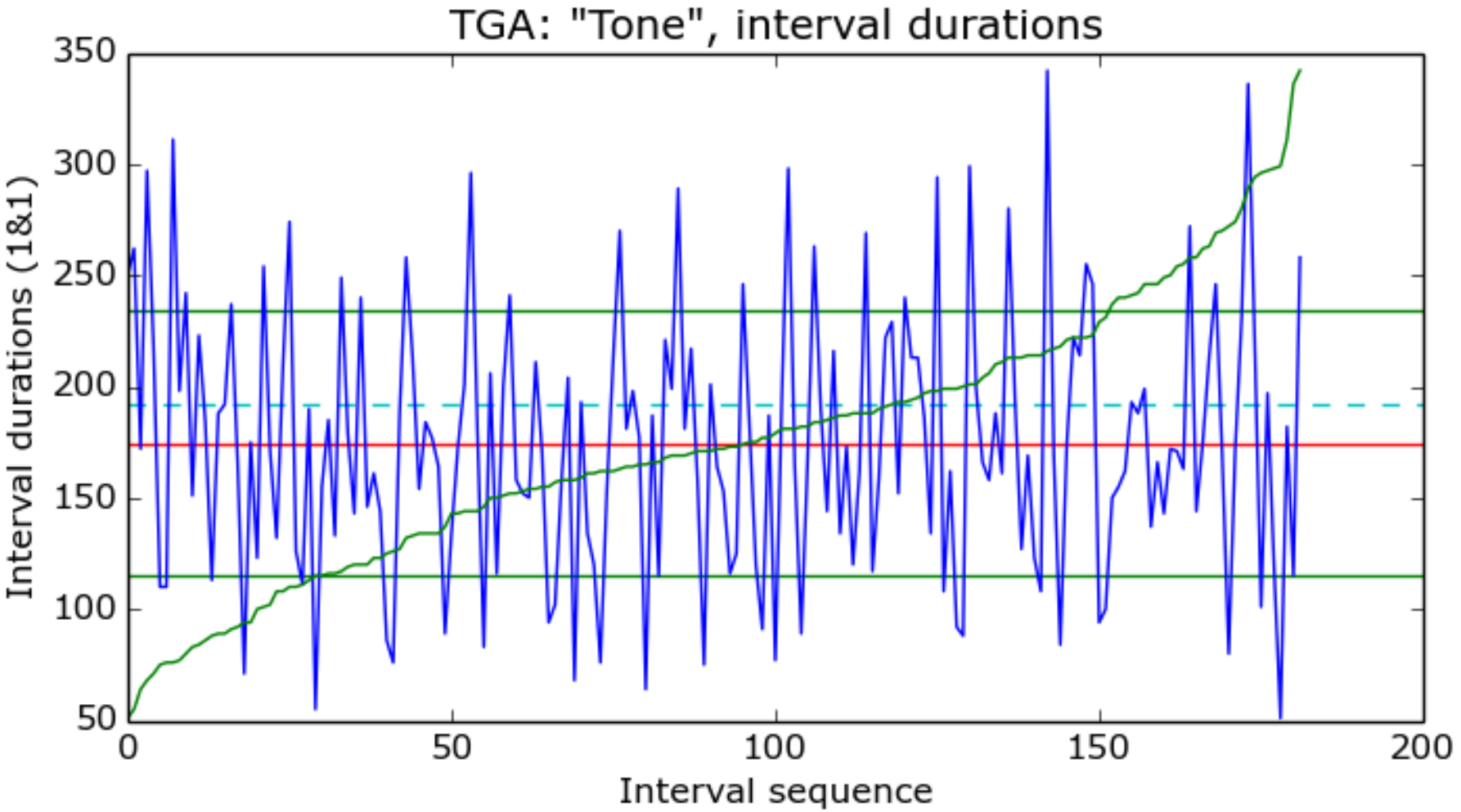
The 5 tones of Mandarin (1, 2, 3, 4 and neutral)



Quasi-isochrony:

- L tone: blue, H tone: red
- Durations of neighbouring syllable pairs tend to be scattered randomly close to zero in the normalised z-scores (zero is the mean duration)
- For details see the box plot.

The 5 tones of Mandarin (1, 2, 3, 4 and neutral)



Durations:

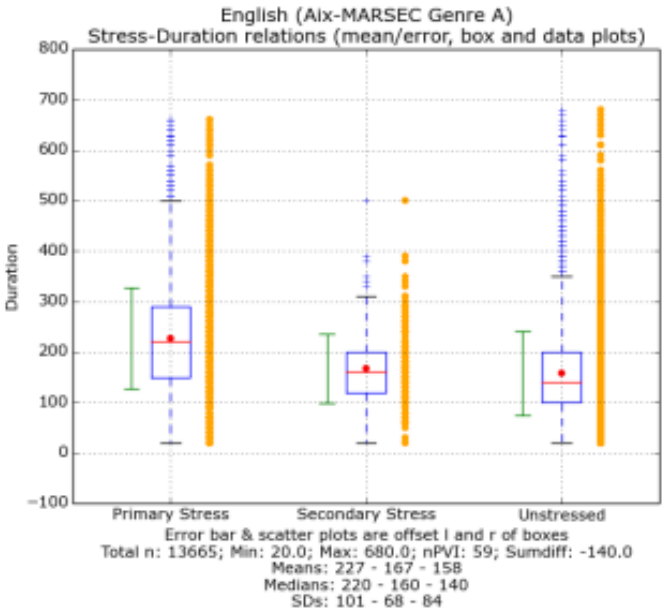
- Durations are similar and vary randomly around the mean.
- As the speech session continues, syllables get slightly shorter (i.e. the syllable rate gets faster, accelerates)

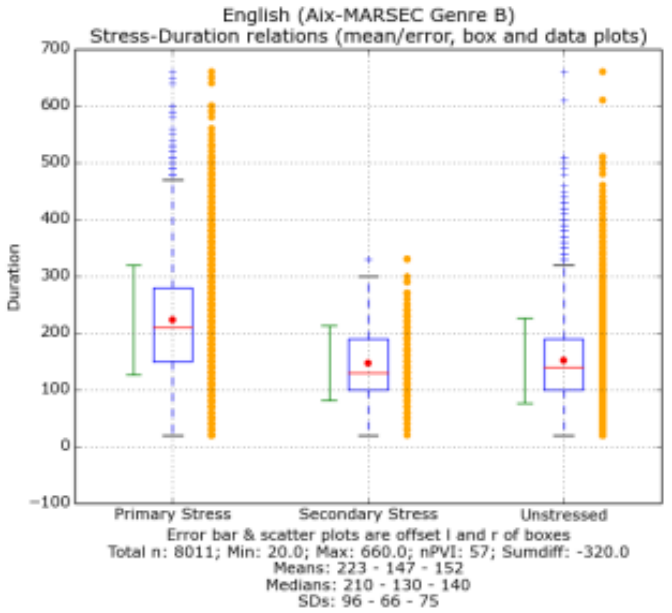
1.3. Three stress levels in English

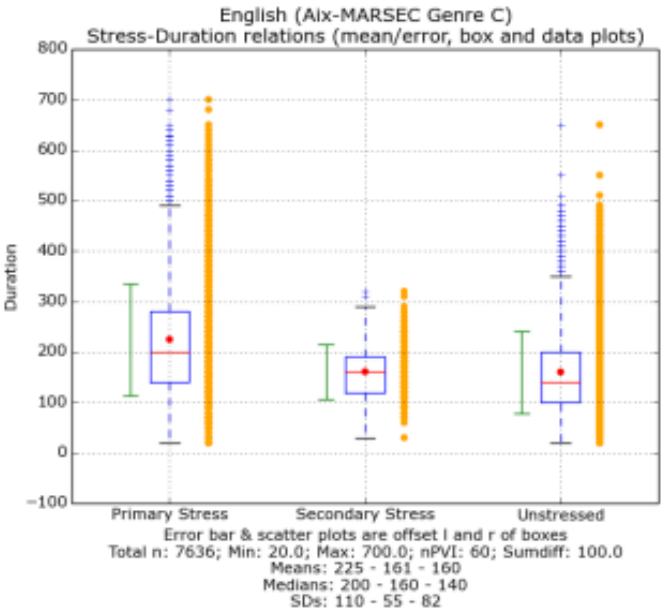
Corpus description

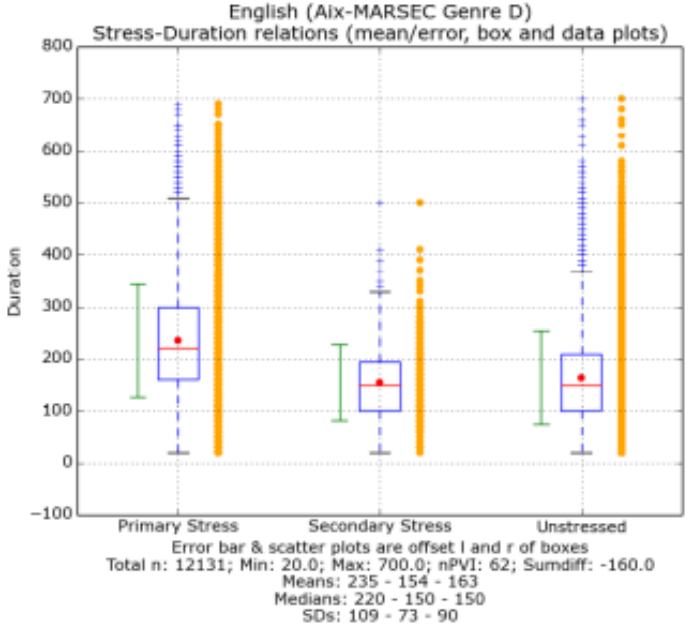
Auran, Cyril, Caroline Bouzon and Daniel Hirst. 2004. The Aix-MARSEC Project: An Evolutive Database of Spoken British English. Proceedings of Speech Prosody 2.

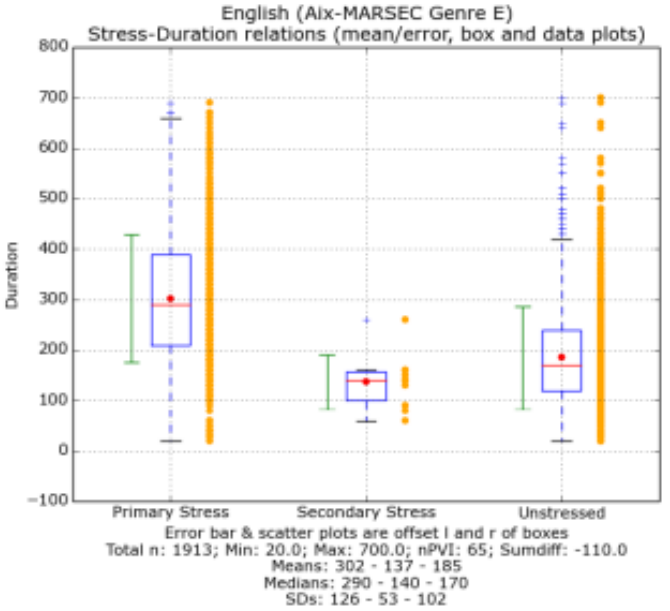
Three stress levels in English

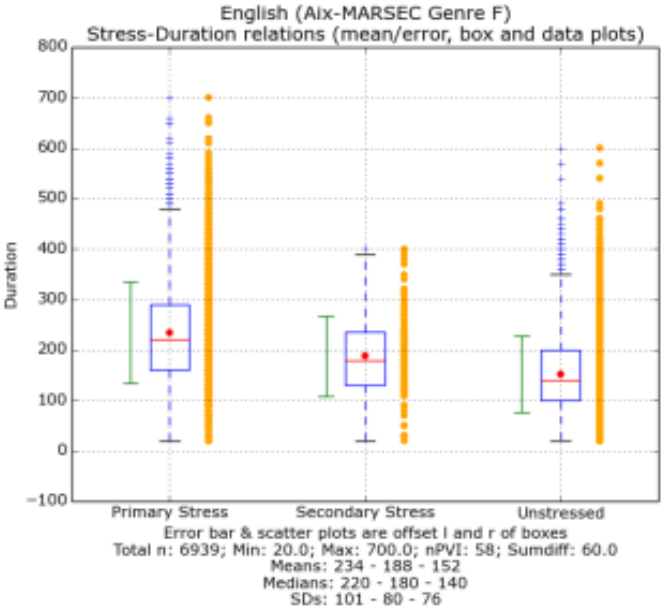




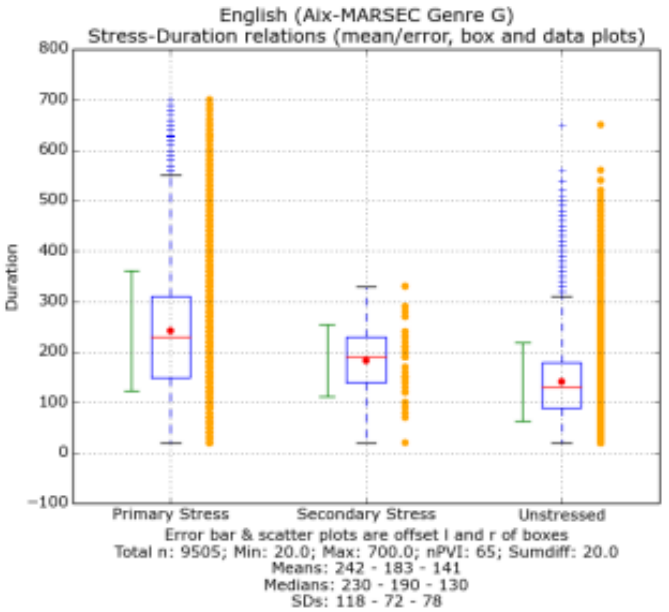


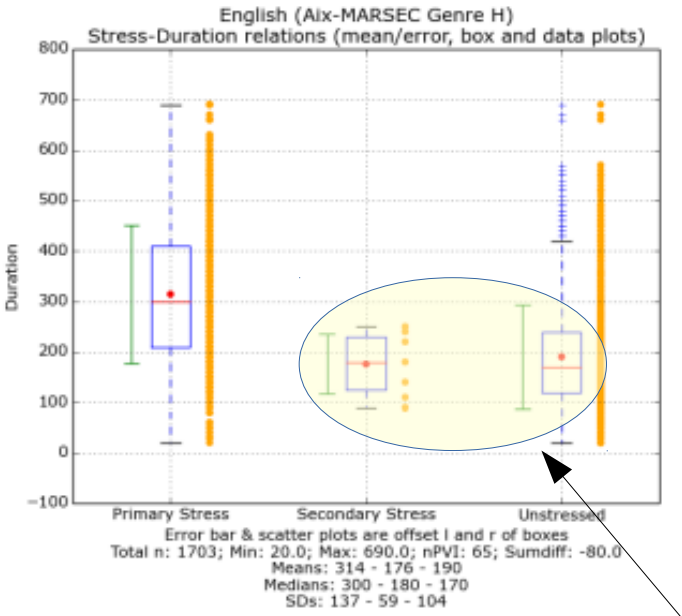


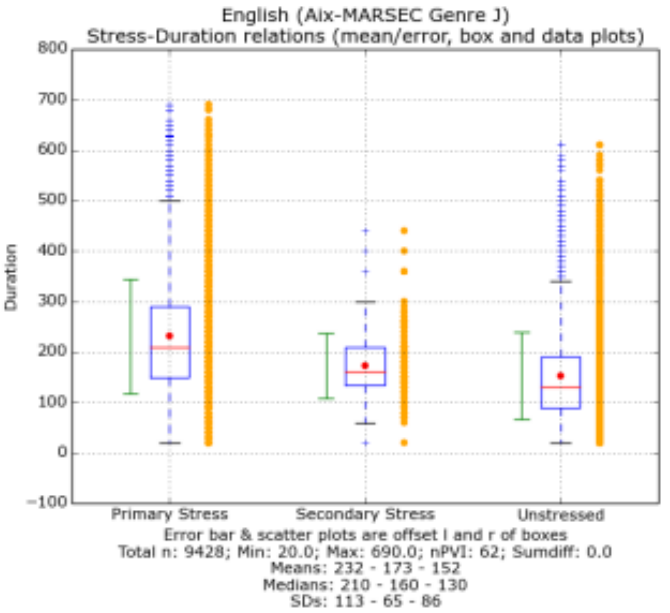


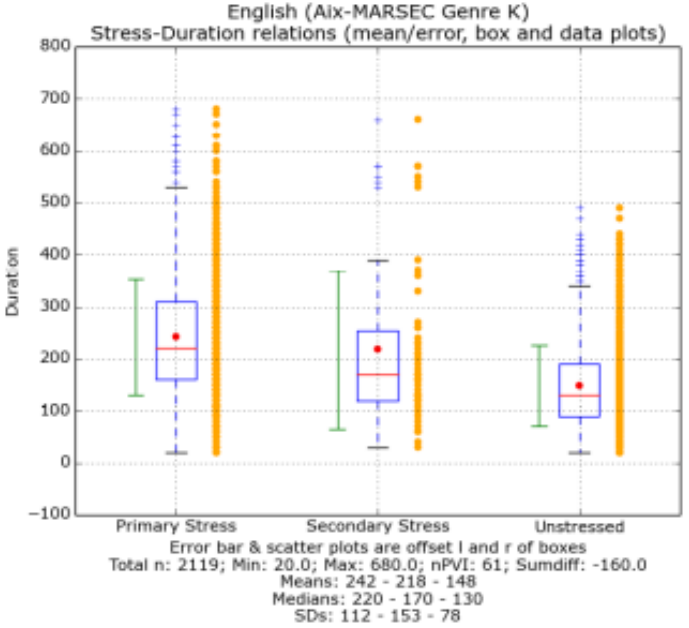


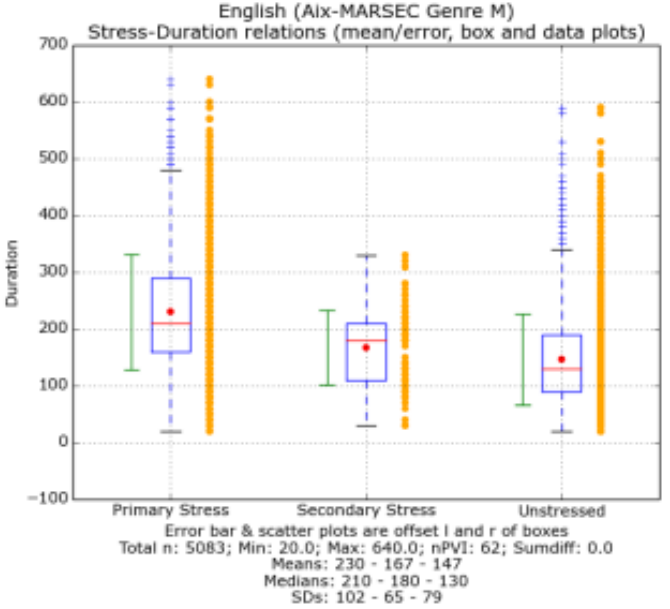
Three stress levels in English





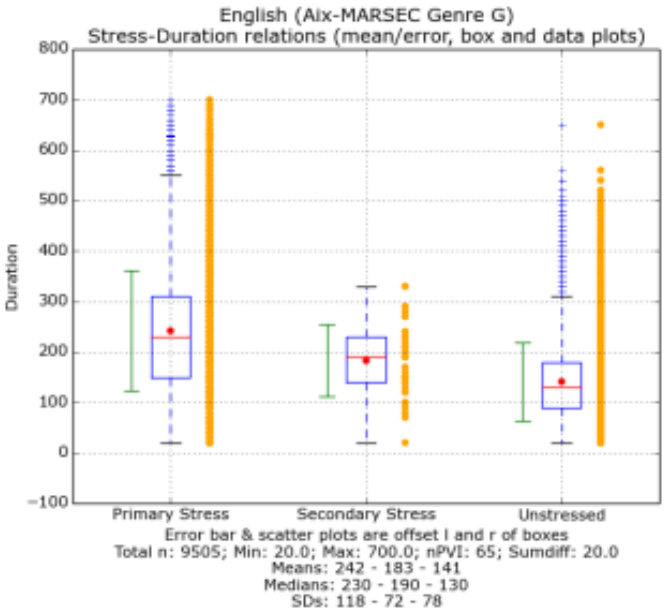


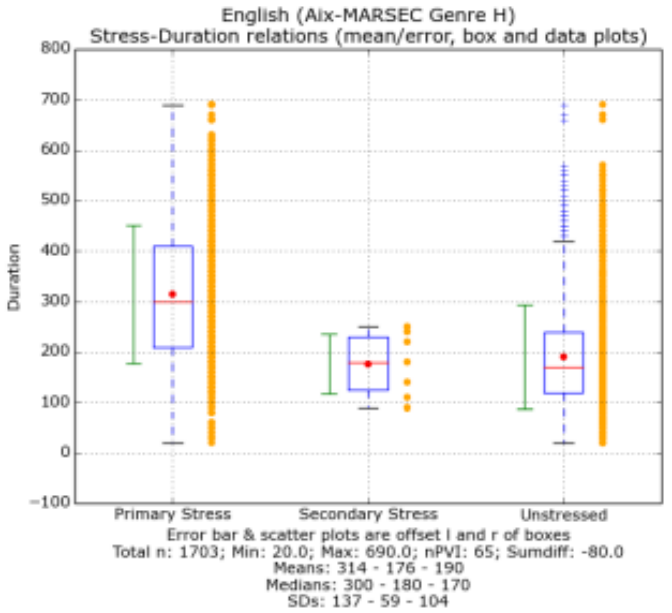


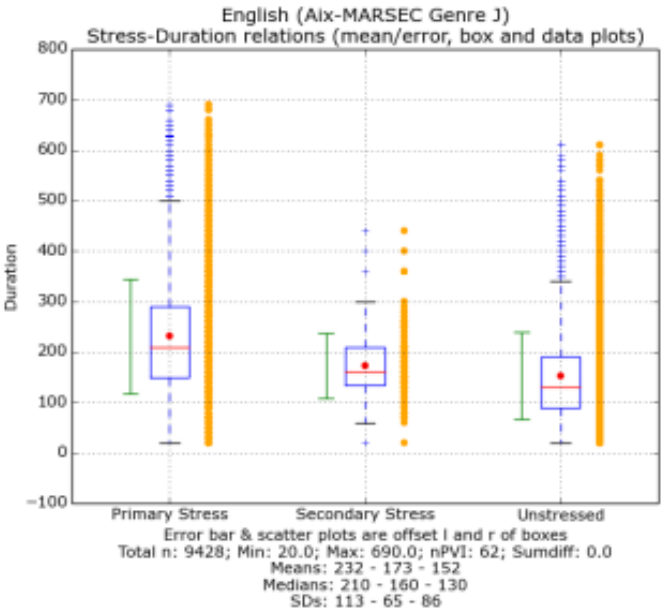


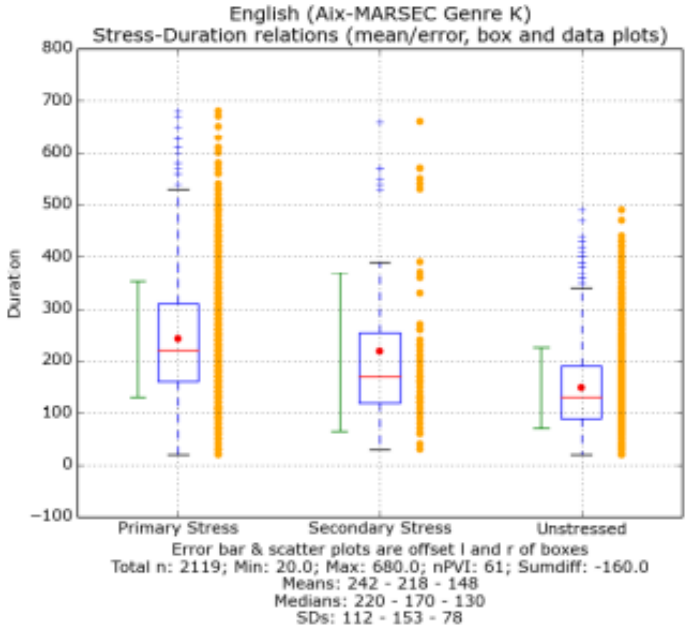
Durations of primary, secondary, unstressed syllables in English relate quite consistently (<u>but not always</u>) to longer, medium and shorter syllable lengths.

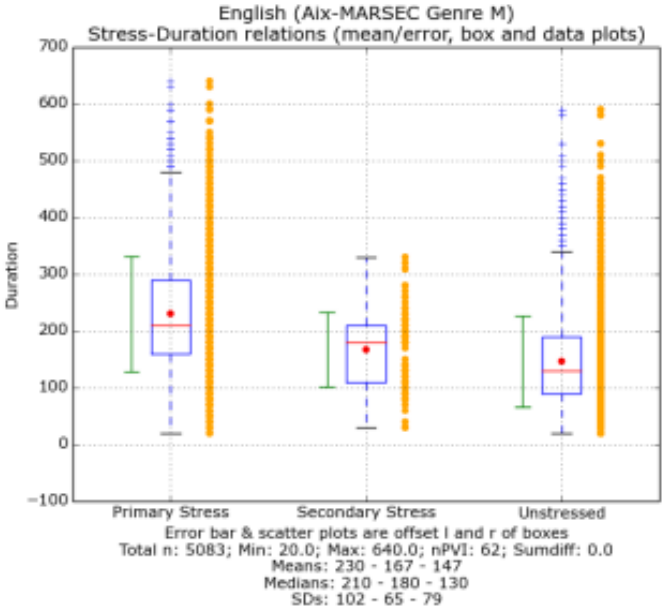
Three stress levels in English











Corpus description

Auran, Cyril, Caroline Bouzon and Daniel Hirst. 2004. The Aix-MARSEC Project: An Evolutive Database of Spoken British English. *Proceedings of Speech Prosody 2*.

https://en.wikipedia.org/wiki/Spoken_English_Corpus

PHONETIC APPROACHES: STATIC TIMING MODELS

3. Rhythm and isochrony (duration similarity) or anisochrony (irregularity, smoothness, dissimilarity): the PIM, PFD, rPVI, nPVI measures (one could also add standard deviation)

These measures are sometimes referred to as ‘rhythm metrics’.

This term is unsuitable, because they only measure interval length averages, not the alternation ‘oscillation’ property.

Definitions of 'rhythm'

- "An ordered recurrent alternation of strong and weak elements in the flow of sound and silence in speech." (Webster web version)
- "Rhythm is the directional periodic iteration of a possibly hierarchical temporal pattern with constant duration and alternating strongly marked (focal, foreground) and weakly marked (non-focal, background) values of some observable parameter." (Gibbon & Gut 2001)
- "Rhythm is viewed here as the hierarchical organisation of temporally coordinated prosodic units ... certain salient events (beats) are constrained to occur at particular phases of an established period." ... "Rhythm is manifested as the temporal binding of events to specific and predictable phases of a superordinate cycle." (Cummins & Port 1998)

Emergent & Physical Rhythm Theories

- Are we talking about RHYTHM
  - or more generally TIMING?
  - In recent work, mainly phonetic, phonological and signal processing perspectives, leading to a wide range of non—comparable methodologies.
- But: speech rhythm is a function of many `hidden' physiological, cognitive and linguistic factors
  - So far, different selections from these factors, leading to incomplete models
  - Emergent Rhythm Theory necessary:
    - Coordination at different ranks leads to rhythm as an emergent function of all ranks
    - ERT is still too complex and inexplicit to be falsifiable.
  - So:
    - start wih Physical Rhythm Theory (PRT) approaches, using speech signal parameters
- But check: Which models really model rhythm?

Rhythm Periodicity Condition

Basic conditions on rhythm:

- Observable parameter (simple or complex) in the acoustic, visual or tactile modalities.
- Alternating (often binary) pattern (with simple or complex components) of one strong and possibly several weak values of this parameter.
- Iteration of the alternating pattern (Rhythm Unit).
- Isochrony (equal timing) of the iterations of the alternating pattern.
- Absolute durations of rhythm units vary: 0.3 ... 1.0 sec.

- *Steele, 18th century British actor*:
  - *SPEECH RHYTHM ≈ HEARTBEAT-PACED PERIODICITY*
  - *Also notice that English stress approximately corresponds to moderate walking speed!*

- ☺

Temporal relations – isochrony metrics

$$PIM(I_{1,\ldots n}) = \sum_{i \neq j} |\log \frac{I_i}{I_j}|$$

- ‘Rhythm metrics’ of relative ('fuzzy', 'sloppy') isochrony:
  - measures of regularity...irregularity of timing units
    - *PIM*: Pairwise Irregularity Measure
    - *PFD*: Pairwise Foot Difference
    - *rPVI*, *nPVI*: raw and normalised Pairwise Variability Index
  - not rhythm, though: they ignore rhythmic alternation

Temporal relations – isochrony metrics

$$PFD(foot_{1...n}) = \frac{100 \times \sum |MFL - len(foot_i)|}{len(foot_{1...n})}$$

where MFL = 'mean foot length'

- 'Rhythm metrics' of relative ('fuzzy', 'sloppy') isochrony:
  - measures of regularity...irregularity of timing units
    - *PIM*: Pairwise Irregularity Measure
    - *PFD*: Pairwise Foot Difference
    - *rPVI*, *nPVI*: raw and normalised Pairwise Variability Index
  - not rhythm, though: they ignore rhythmic alternation

Temporal relations – isochrony metrics

$$nPVI(d_{1...m}) = 100 \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1)$$

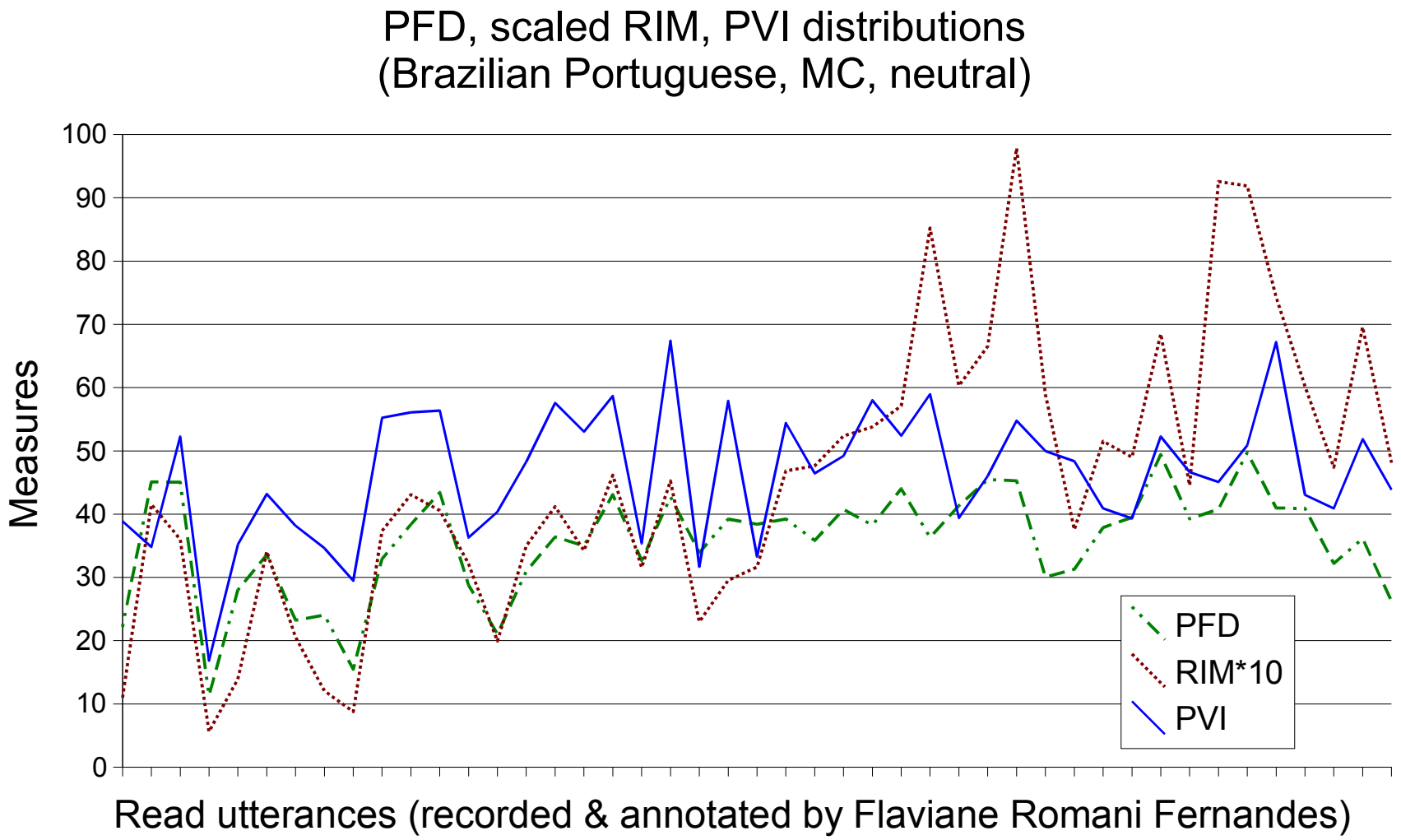
- ‘Rhythm metrics’ of relative ('fuzzy', 'sloppy') isochrony:
  - measures of regularity...irregularity of timing units
    - *PIM*: Pairwise Irregularity Measure
    - *PFD*: Pairwise Foot Difference
    - *rPVI*, *nPVI*: raw and normalised Pairwise Variability Index
  - not rhythm, though: they ignore rhythmic alternation

Temporal relations – isochrony metrics

$PIM(I_{1,\ldots n}) =$	$\sum_{i \neq j} \lvert \log \frac{I_i}{I_j} \rvert$
$PFD(foot_{1\ldots n}) =$	$\frac{100 \times \sum \lvert MFL - len(foot_i) \rvert}{len(foot_{1\ldots n})}$ where MFL = 'mean foot length'
$rPVI(d_{1\ldots m}) =$	$\sum_{k=1}^{m-1} \lvert d_k - d_{k+1} \rvert / (m-1)$
$nPVI(d_{1\ldots m}) =$	$100 \times \sum_{k=1}^{m-1} \lvert \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \rvert / (m-1)$

- 'Rhythm metrics' of relative ('fuzzy', 'sloppy') isochrony:
  - measures of regularity...irregularity of timing units
    - *PIM*: Pairwise Irregularity Measure
    - *PFD*: Pairwise Foot Difference
    - *rPVI*, *nPVI*: raw and normalised Pairwise Variability Index
  - not rhythm, though: they ignore rhythmic alternation

Empirical comparison of RIM, PFD and PVI measures



Critique of 'isochrony-only' approaches

- Isochrony-only models
  - fail to recognise
    - the structure of rhythm: alternation + isochrony
    - directionality of alternation between neighbouring syllables, by using absolute values of differences
  - PIM and PFD
    - global measures of difference, like standard deviation: no directionality
  - rPVI and nPVI
    - measure of local differences: no directionalityconvenient abstraction away from change of articulation rate (acceleration, deceleration)
    - uses a *binary* rhythm model by comparing neighbours, but:
      - *Big John swam fast past Jane's boat (unary, syllable timed)*
      - *Johnny walked along the pathway (binary, foot-timed)*
      - *Jonathan Appleby hurried away on his bicycle (ternary, foot-timed)*

Critique of 'isochrony-only' approaches

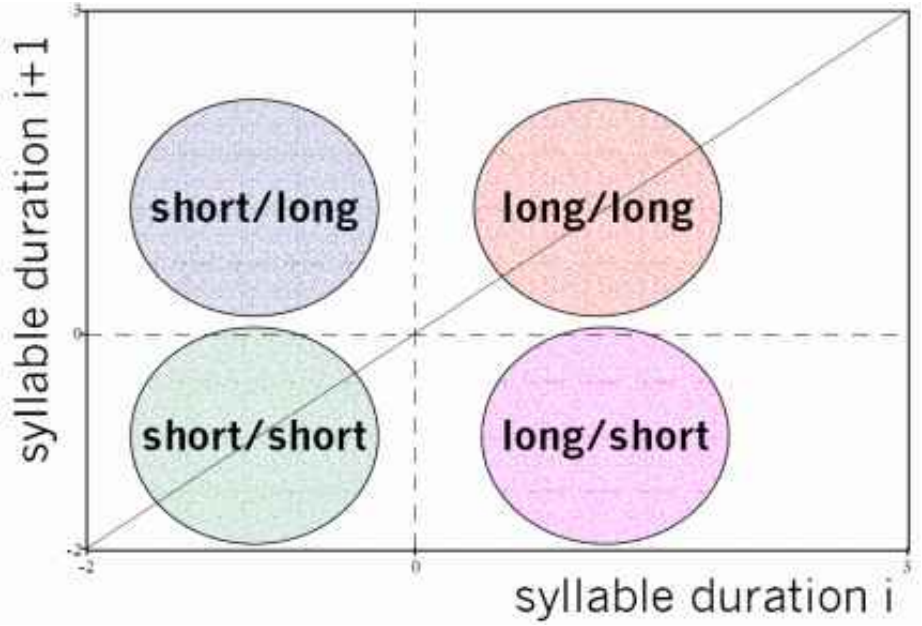
- There are many other useful isochrony / irregularity / smoothness measures

  perhaps most prominently in the past 5 years the 2-dimensional Ramus model: ΔC X %V

- But beware:
  - Isochrony/irregularity/smoothness:
    - a necessary but not a sufficient condition for rhythm:
      - cf. Cummins (2002) on Ramus:
      - *Where is the bom-di-bom-bom in %V?*
  - Cummins' critique applies to all isochrony-only models:
    - the necessary condition of alternation must combine with the necessary condition of isochrony to create a more complex sufficient condition for rhythm

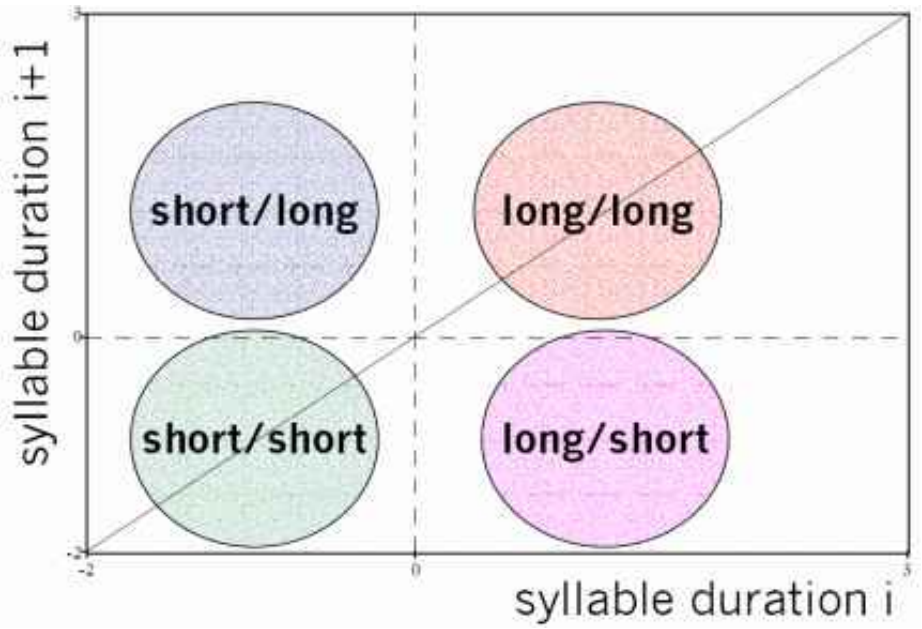
Partial recovery of periodicity: Wagner Quadrants

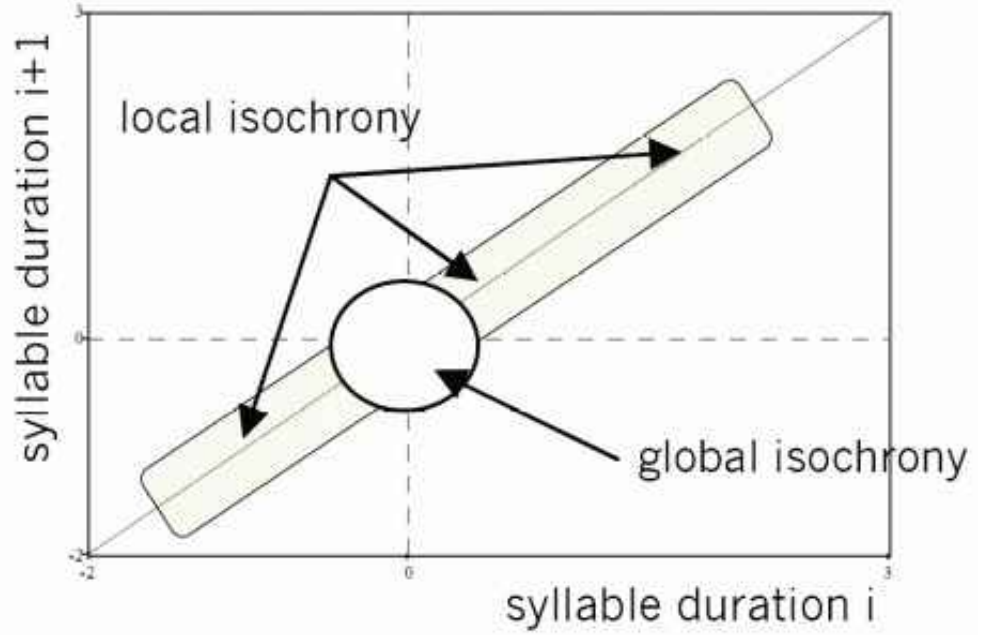
Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting DUR(i) x DUR(i+1):



Partial recovery of periodicity: Wagner Quadrants

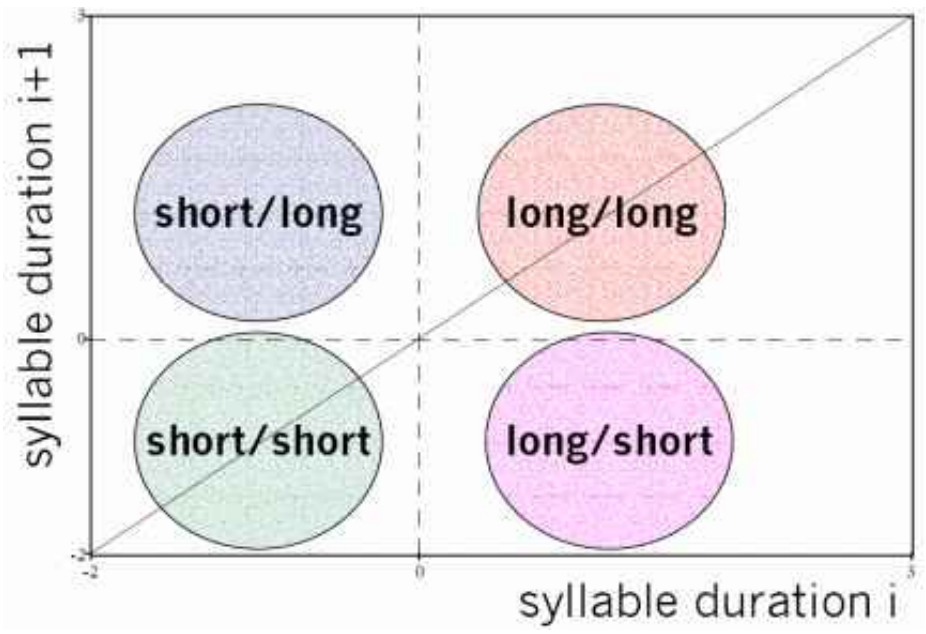
Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting DUR(i) x DUR(i+1):

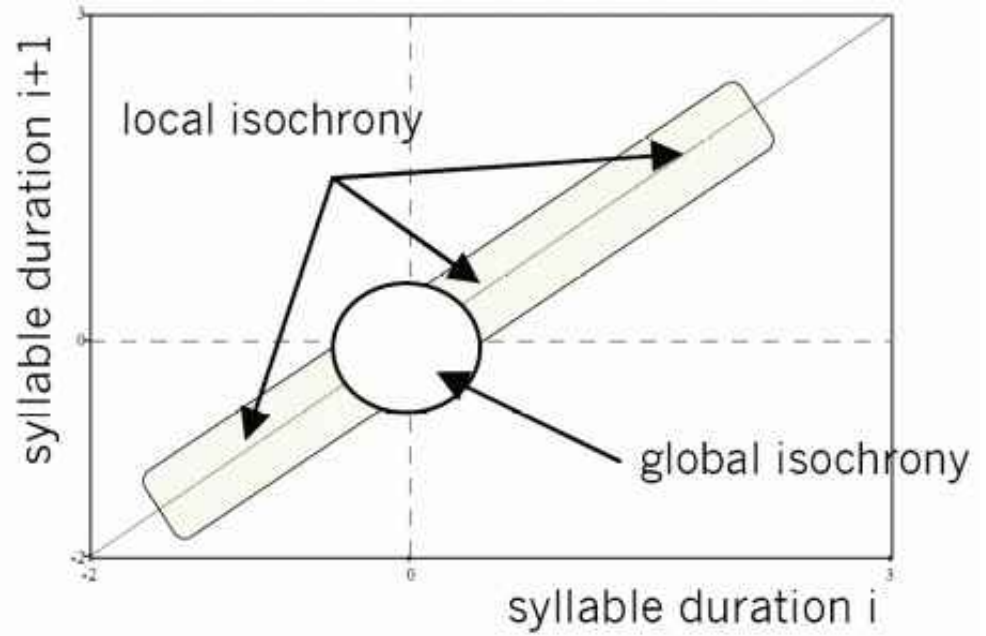




Partial recovery of periodicity: Wagner Quadrants

Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting DUR(i) x DUR(i+1):



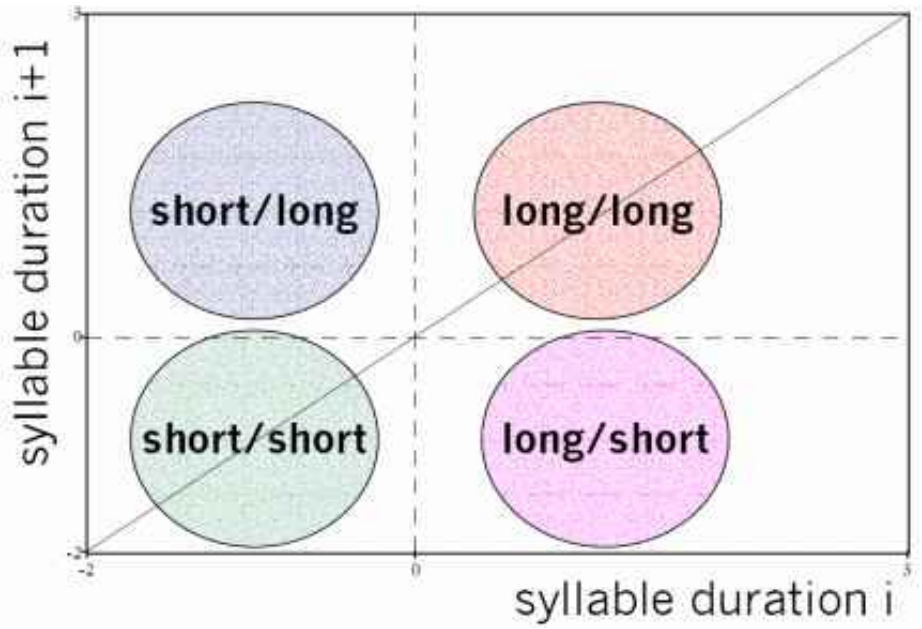


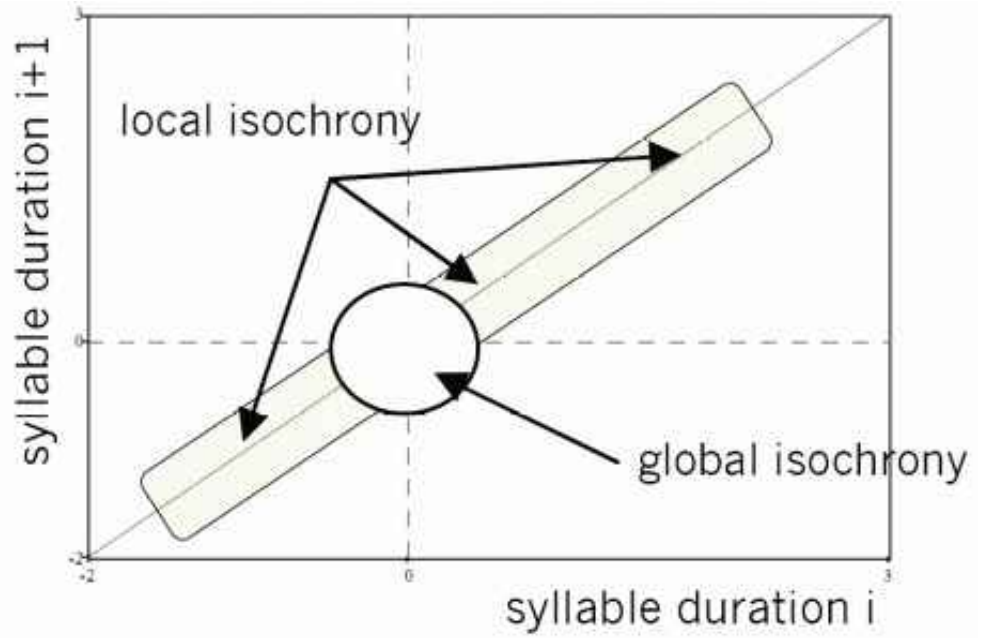
Heuristic idea

In isochonous sequences, items will lie within the global isochrony circle.
In binary alternating long-short/short-longsequences, differences will lie in the top left and bottom right quadrants.
In non-binary sequences, similar longer than average (e.g. unary) sequences will be top right, similar shorter than average sequences will be bottom left.

Partial recovery of periodicity: Wagner Quadrants

Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting DUR(i) x DUR(i+1) in 4 quadrants:

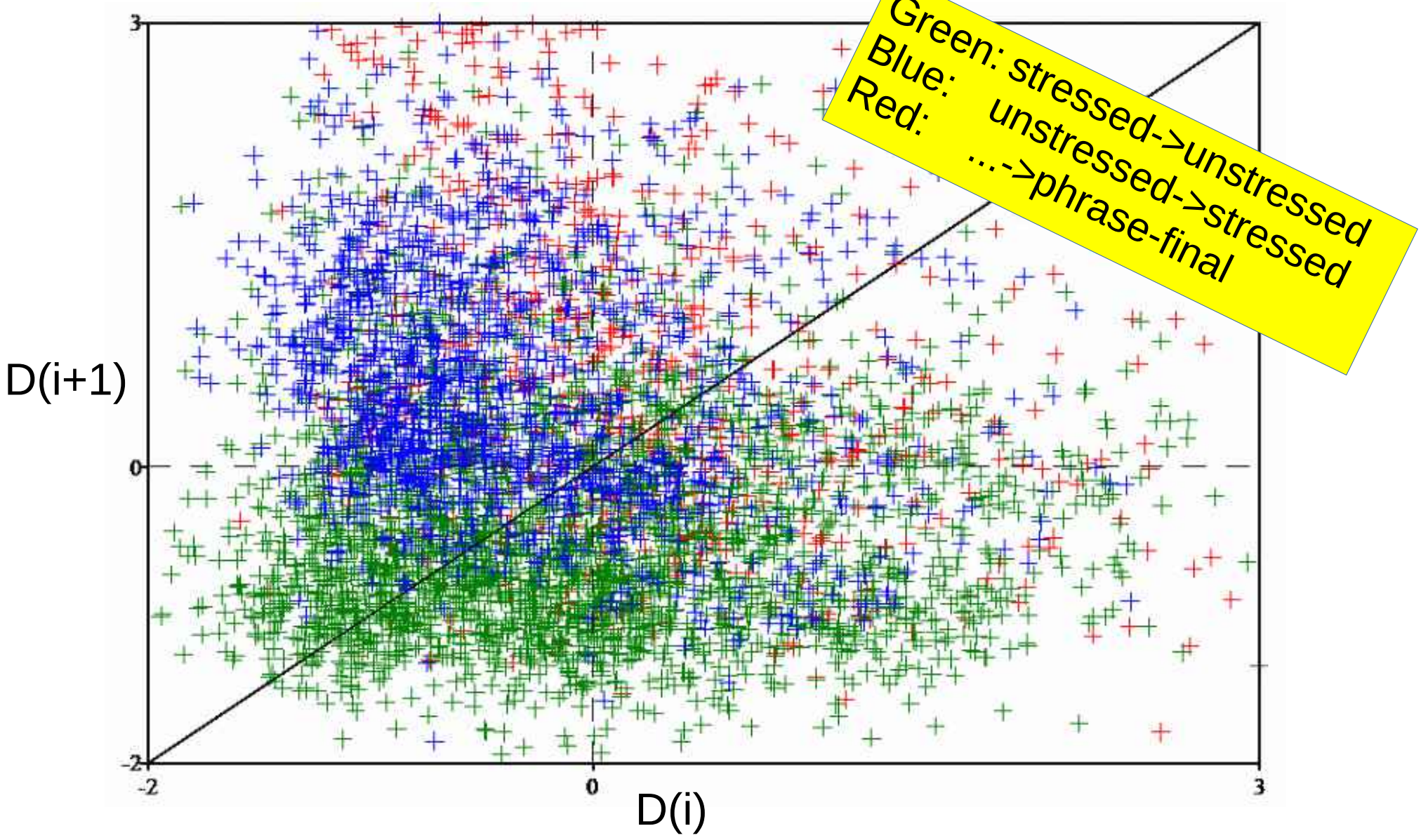




Note:

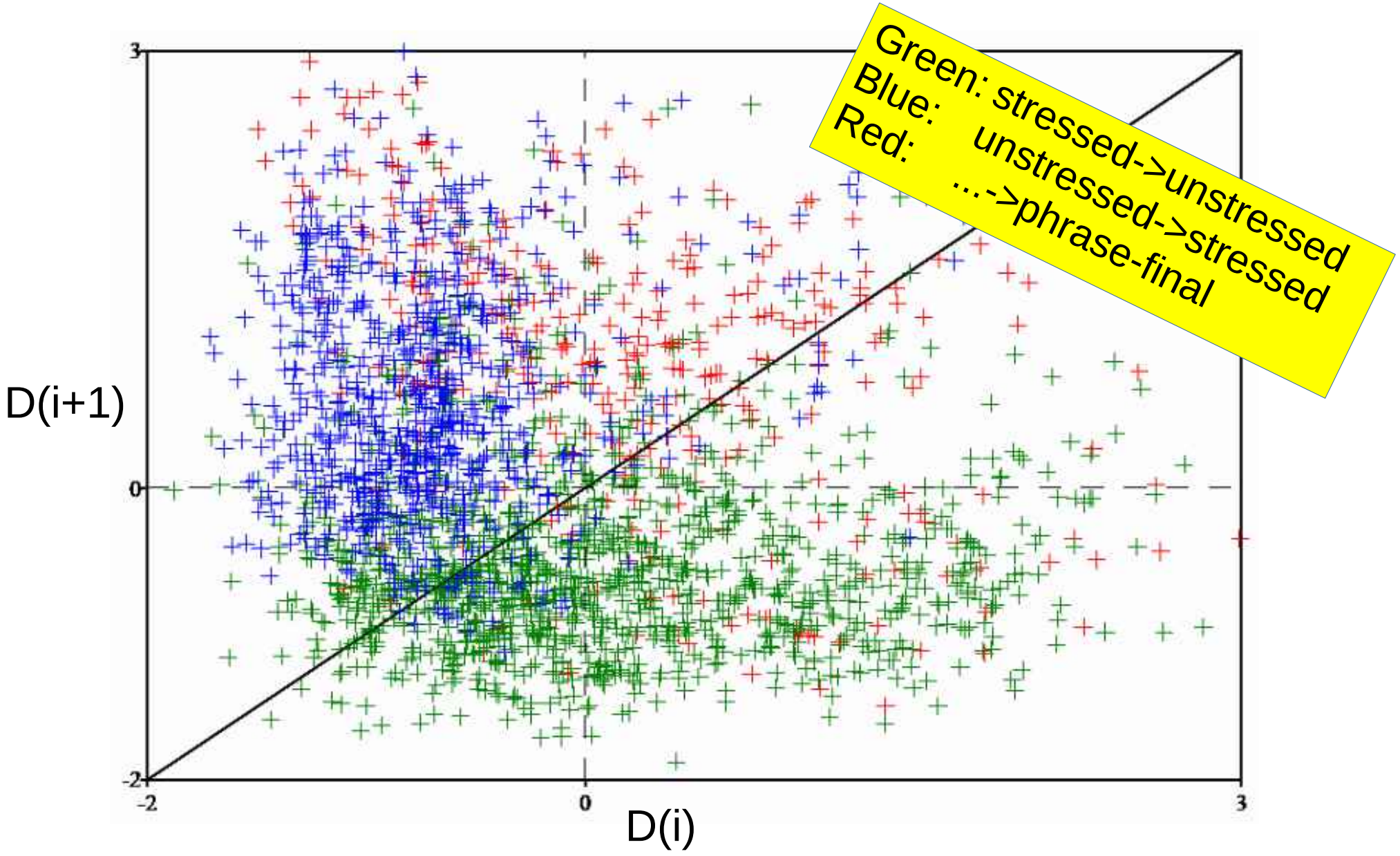
- the relations are still binary only
- the method distinguishes both directionality and difference

Binary duration relations: German



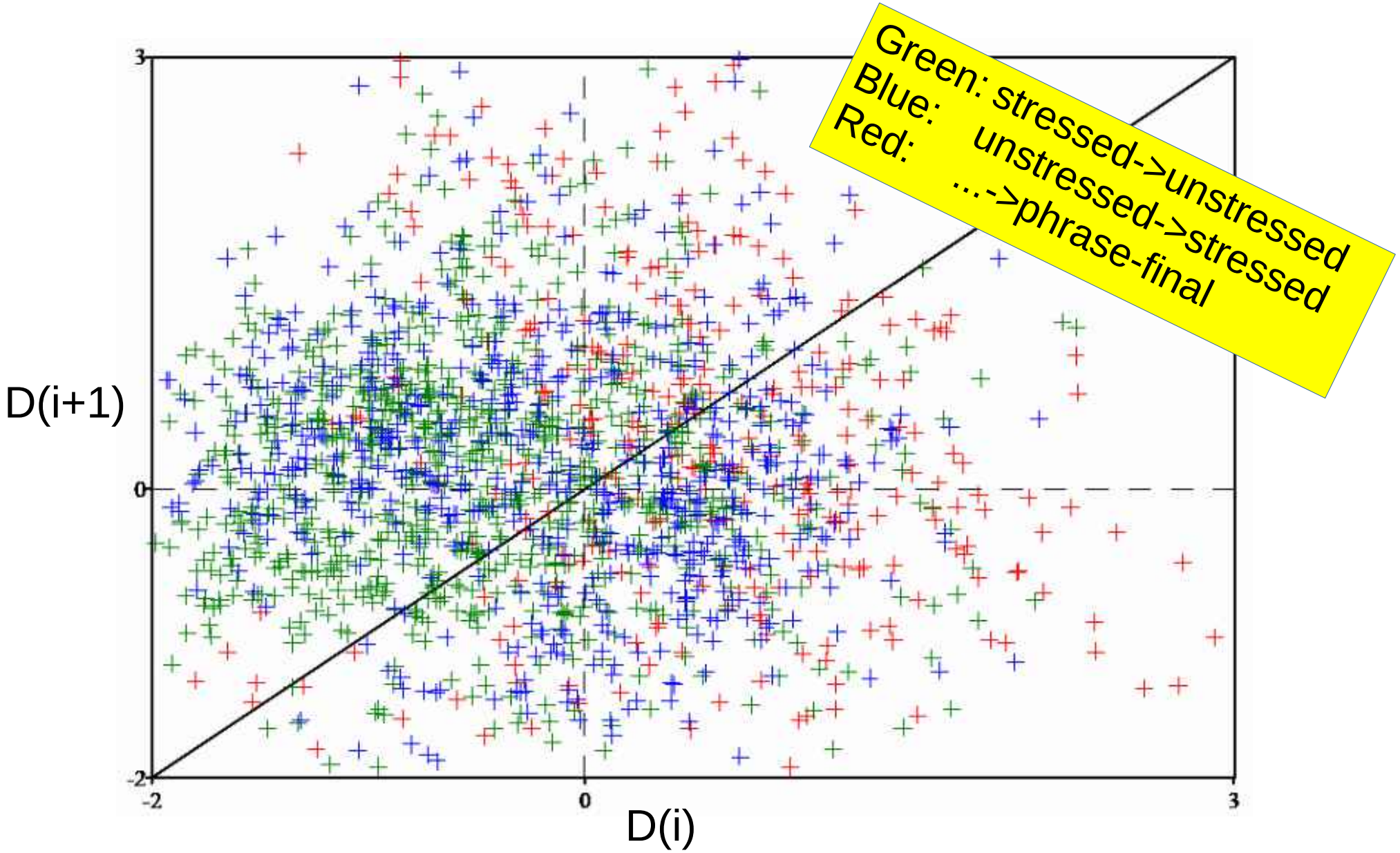
Comment: stress timed - green & blue disjoint

Wagner Quadrants: English



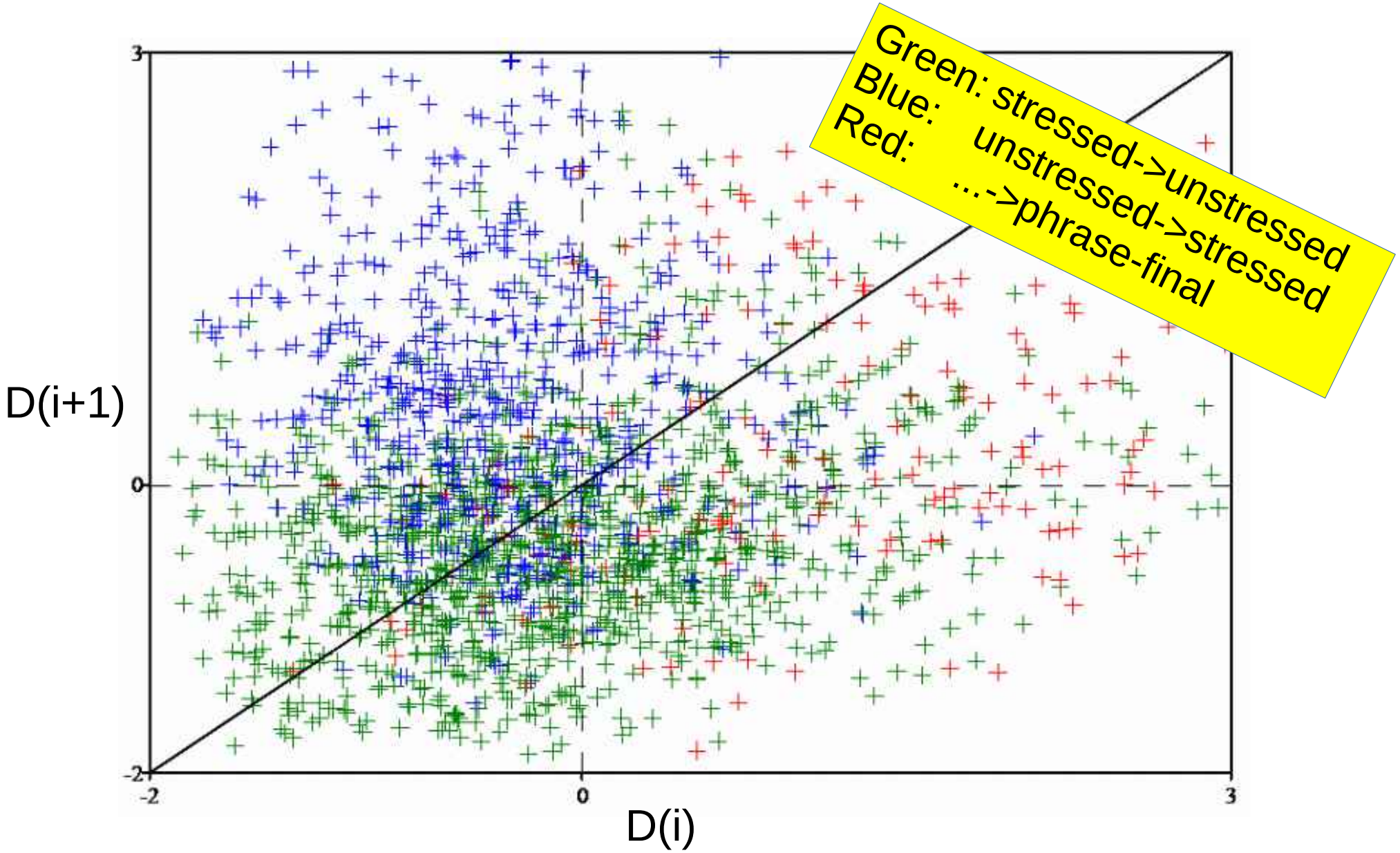
Comment: stress timed - green & blue disjoint

Wagner Quadrants: French



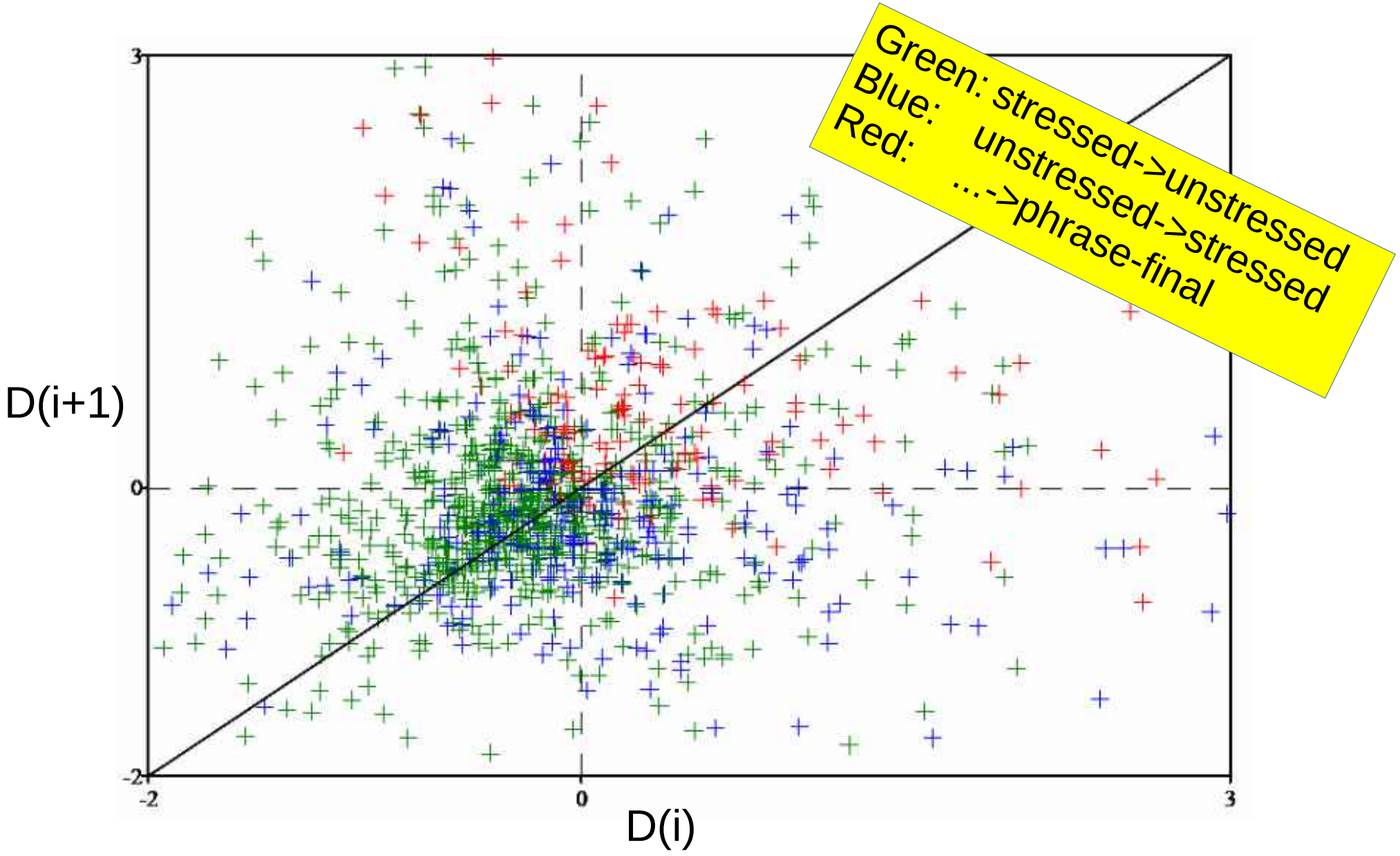
Comment: syllable timed - green & blue overlap

Wagner Quadrants: Italian



Comment: stress timed - green & blue disjoint

Wagner Quadrants: Polish



Comment: highly syllable timed - green & blue overlap

PHONETIC APPROACHES: STRUCTURED TIMING MODELS

The Phonology of Timing

Need for structure in timing models

- The necessary condition of alternation must combine with the necessary condition of isochrony to create a more complex sufficient condition for rhythm
- Consequently more structure is needed, e.g.:
  - Abercrombie's depth 1 hierarchical model (English):
    - Foot = Ictus Remiss (Strong Weak)
  - Jassem's depth 2 hierarchical model (English):
    - TRU = ANA NRU

      NRU = Stress Unstressed* #
      - TRU: Total Rhythm Unit; ANA: Anacrusis; NRU: Narrow Rhythm Unit
      - *: sequence of any length;
      - #: all unstressed syllables until the nearest grammatical boundary

Phonological models

There are numerous phonological models of accentuation:

- descriptive phonetic models
  - Abercrombie
  - Generative Phonology, Metrical Phonology
- phonetically grounded quantitative models
  - Jassem
  - Campbell
  - Wagner

Adding structure

- Sequential / overlap temporal structure
  - Wagner:
    - Nouns, Numerals, Proper Names
    - Adverbs, Adjectives
    - Verbs, Demonstrative Pronouns, WH-Pronoun
    - Modal & Auxiliary Verbs, Affirmative & Negation Particles
    - Determiners, Conjunctions, Subjunctions, Prepositions

Remember the Event Logic relations?

- Hierarchical temporal structure:
  - Jassem: TotalRhythmUnit = ANAcrusis NarrowRhythmUnit
    - where ANA is anisochronous, NRU is isochronous
- Barbosa, Cummins:
  - Oscillation hierarchy
- Wagner:
  - Explication of alternation and iteration with Finite State Automaton (FSA)

PHONETIC APPROACHES: DYNAMIC TIMING MODELS

(Selection from signal-theoretic models of timing)

Barbosa's dynamic timing model of rhythm

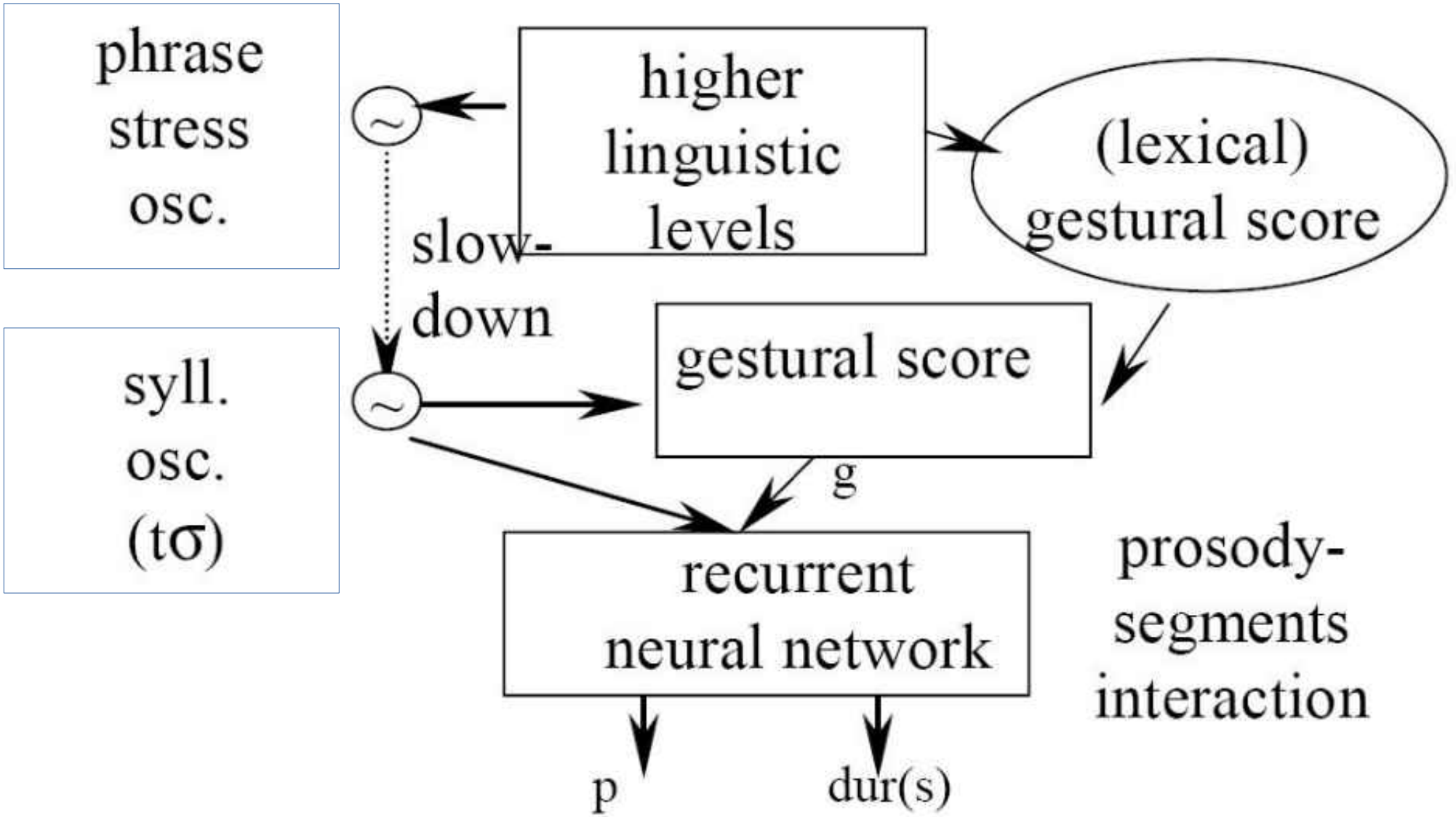
Perceptual definition of rhythm:

- speech rhythm is understood as the consequence of the variation of perceived duration along the entire utterance.

Model:

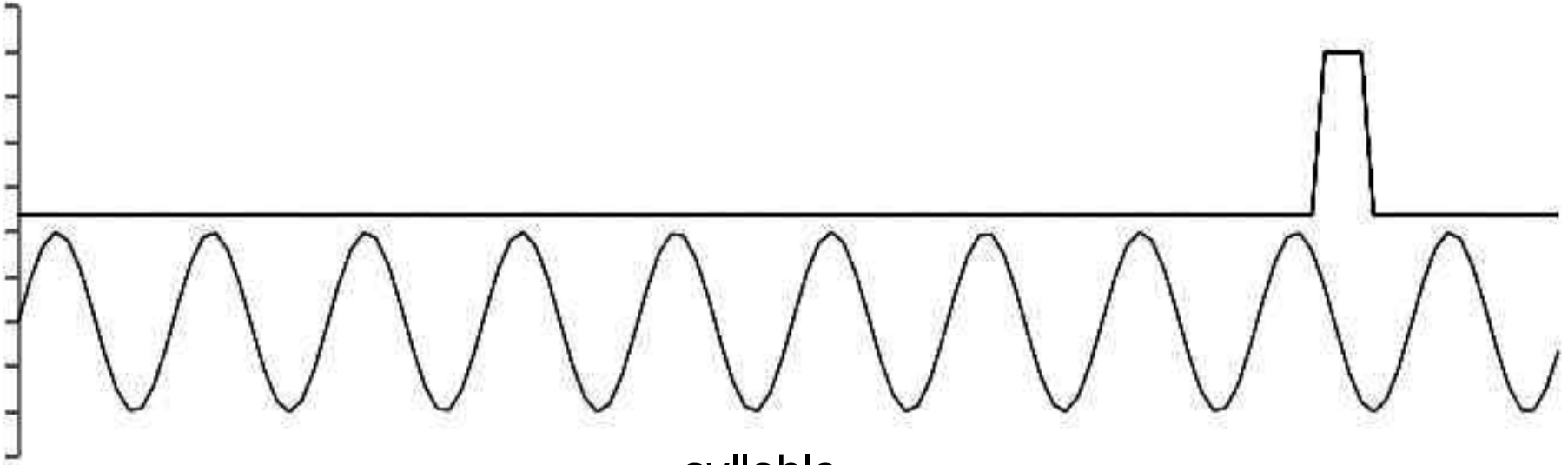
- Two levels of duration encoding / control / specification, coupling between 2 oscillators:
  - syllabic: intrinsic lexical level
  - phrasal: extrinsic, properly rhythmic level
- Entrainment (coupling) of the oscillators
- Emulation of results of other studies:
  - more like stress-timing?
  - more like syllable-timing?

Barbosa's dynamic timing model of rhythm



Barbosa's dynamic timing model of rhythm

phrase pulse



syllable
oscillations

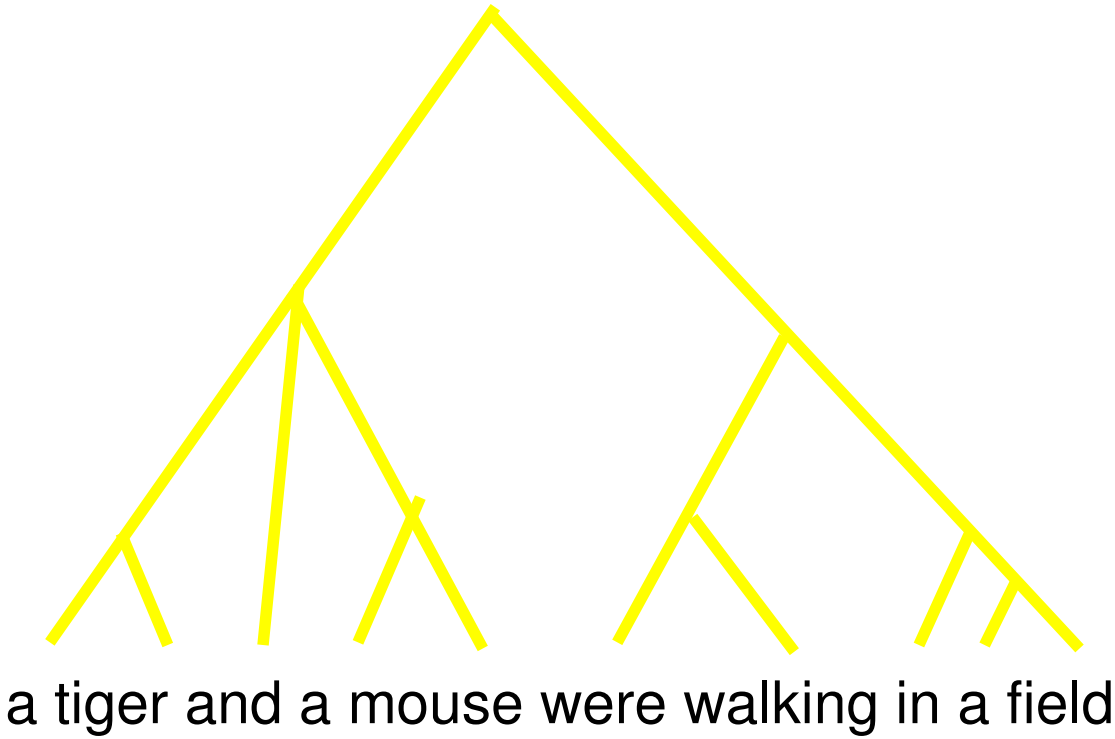
(for English these could be stress oscillations)

Note also work by Cummins, Port, Wagner, Windman and others on oscillator models of rhythm.

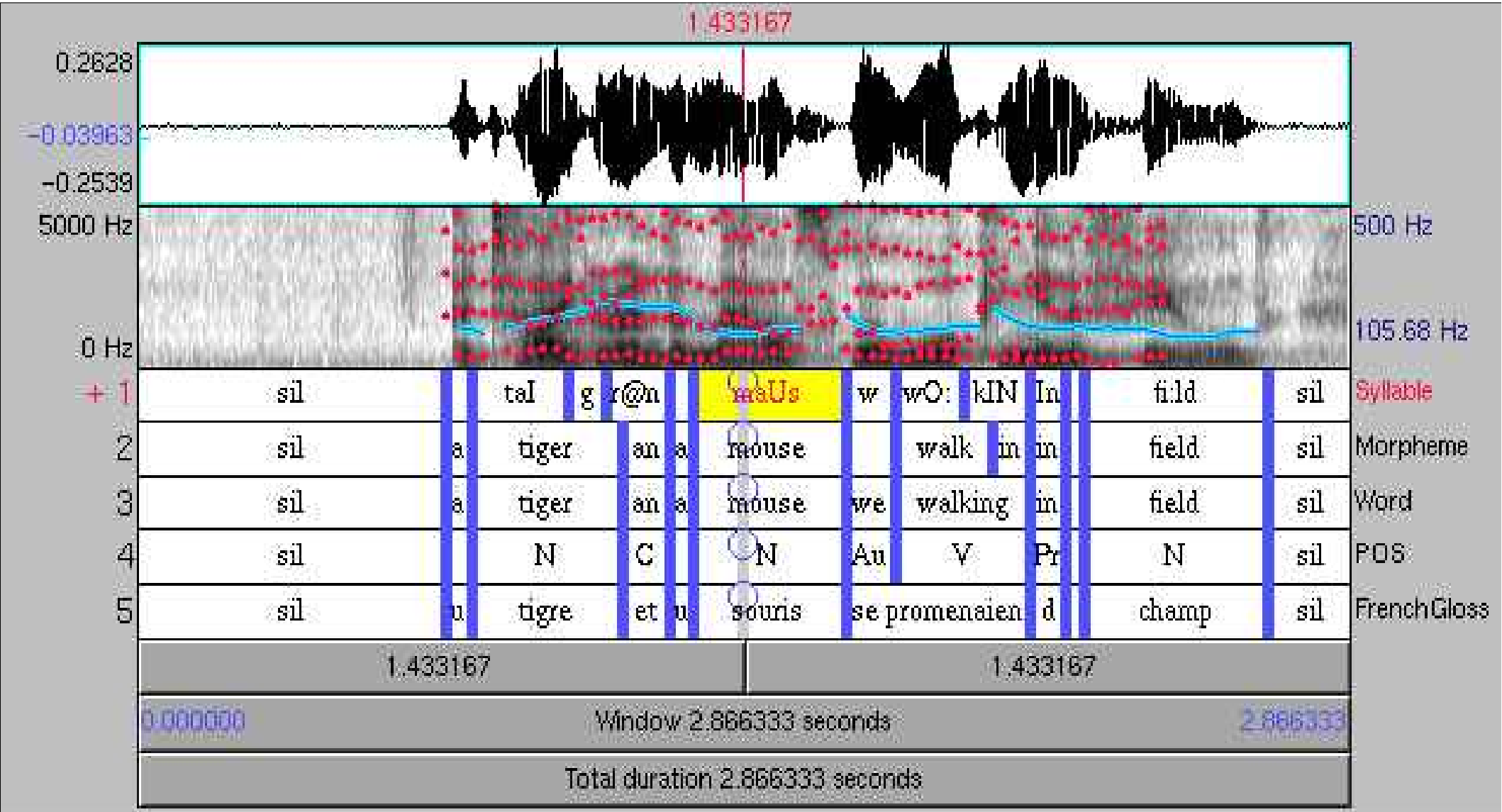
How do we analyse timing?

ANNOTATION MINING

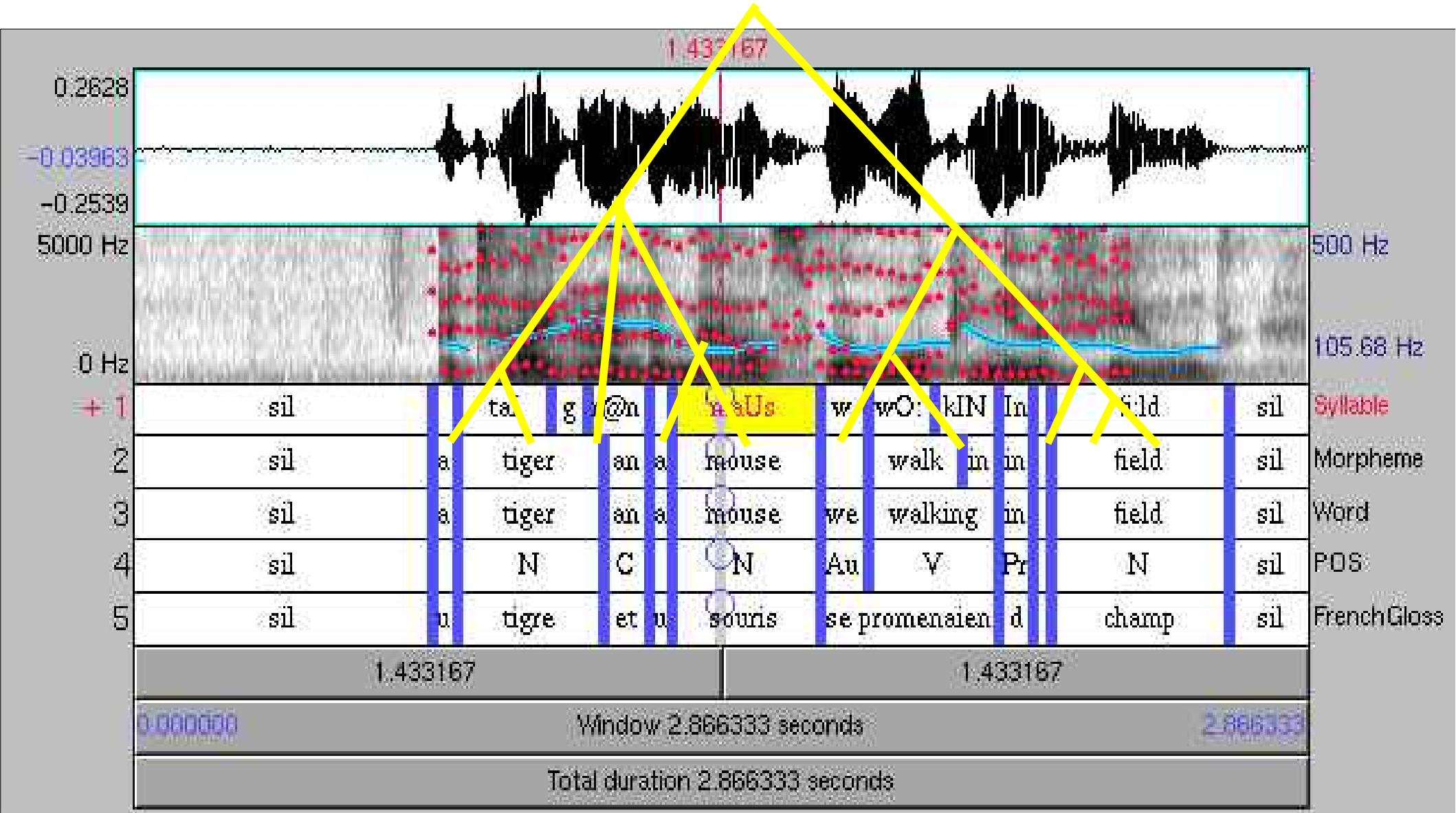
Parsing



Annotation mining



Annotation mining plus parsing

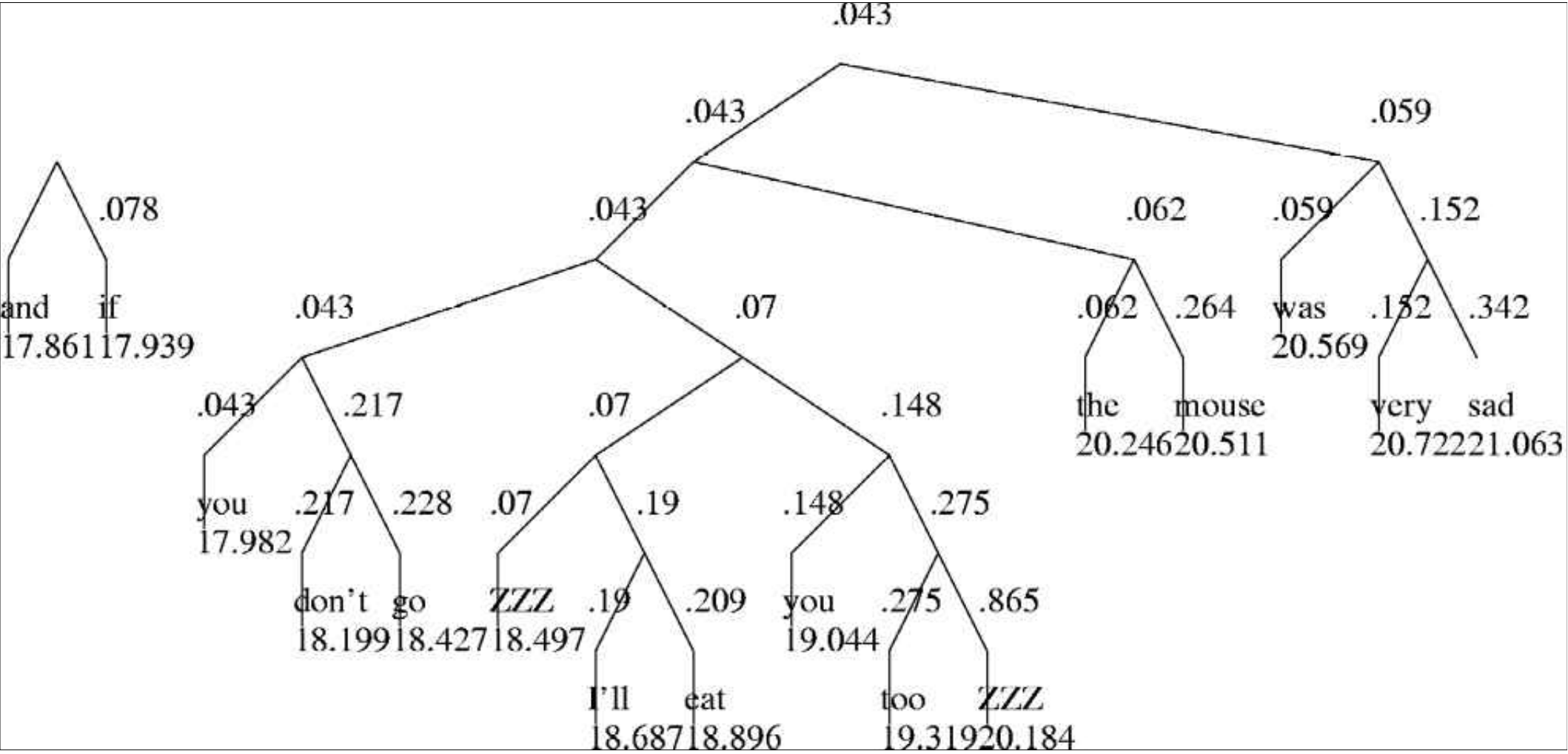


A tiger and a mouse

A tiger and a mouse were walking in a field when they saw a big lump of cheese lying on the ground. The mouse said: "Please, tiger, let me have it. You don't even like cheese. Be kind and find something else to eat." But the tiger put his paw on the cheese and said: "It's mine! And if you don't go I'll eat you too." The mouse was very sad and went away. The tiger tried to swallow all of the cheese at once but it got stuck in his throat and whatever he tried to do he could not move it. After a while, a dog came along and the tiger asked it for help. "There is nothing I can do." said the dog and continued on his way. Then, a frog hopped along and the tiger asked it for help. "There is nothing I can do." said the frog and hopped away. Finally, the tiger went to where the mouse lived. She lay in her bed in a hole which she had dug in the ground. "Please help me," said the tiger. "The cheese is stuck in my throat and I cannot remove it." "You are a very bad animal," said the mouse. "You wouldn't let me have the cheese, but I'll help you nonetheless. Open your mouth and let me jump in. I'll nibble at the cheese until it is small enough to fall down your throat." The tiger opened his mouth, the mouse jumped in and began nibbling at the cheese. The tiger thought: "I really am very hungry.."

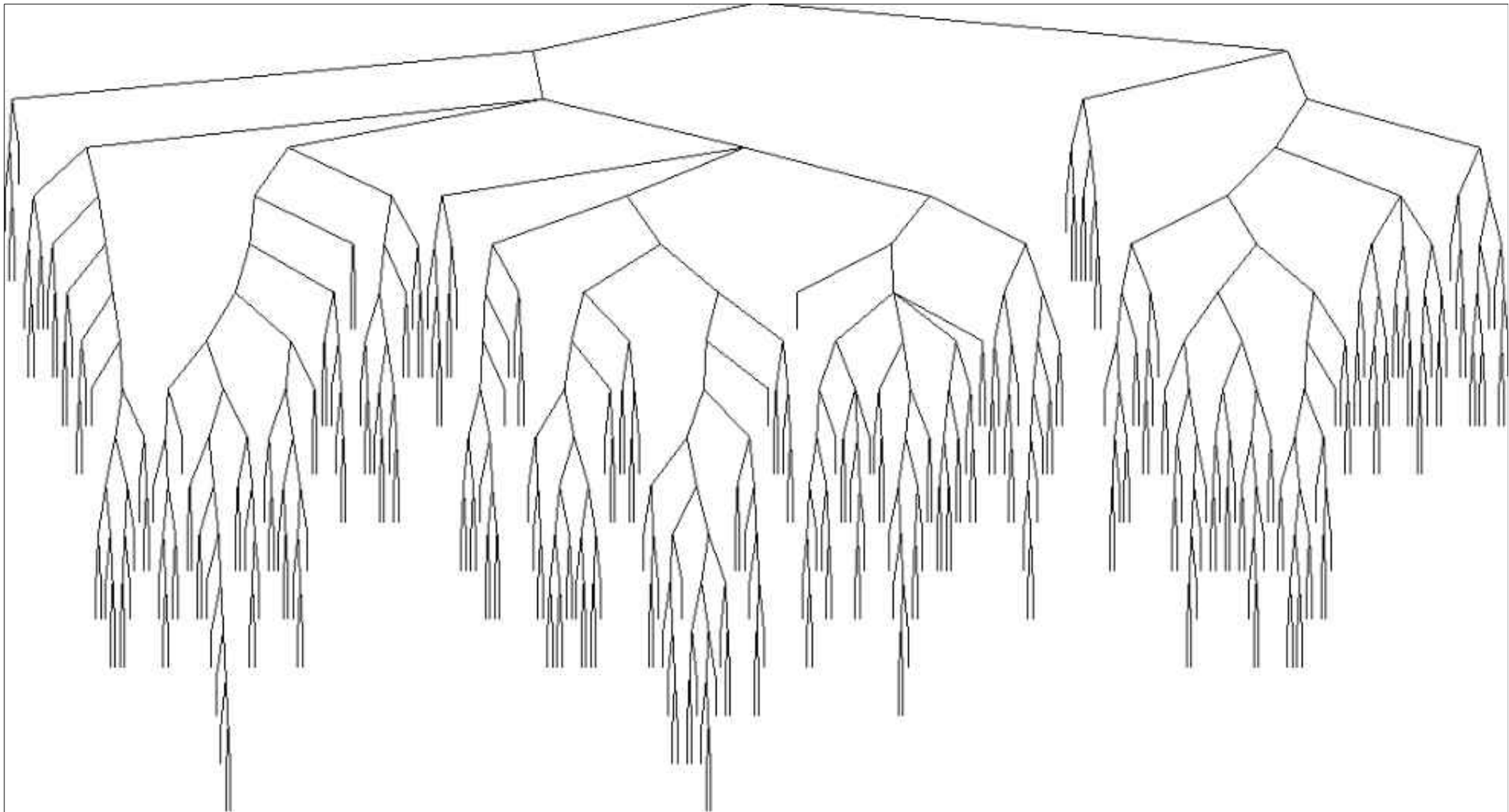
Annotation mining and timing hierarchies: tree induction

Part of a narrative:



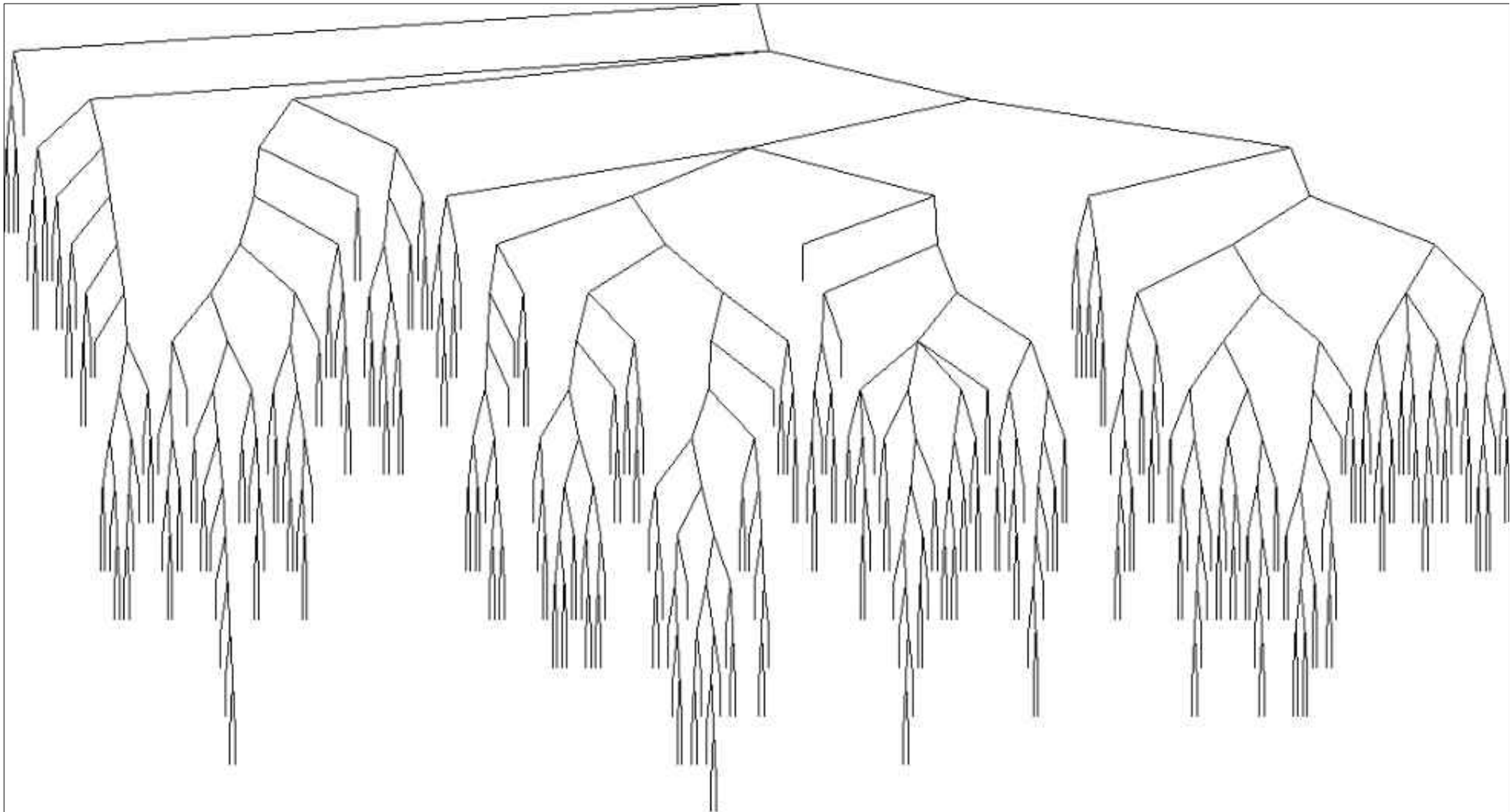
Annotation mining and timing hierarchies: tree induction

A complete narrative - parametrisation A:



Annotation mining and timing hierarchies: tree induction

A complete narrative - parametrisation B:



CONCLUSION, SUMMARY, OUTLOOK

Towards an Emergent Rhythm Theory

- Toward an *Emergent Rhythm Theory*
  - Recall Dauer (1983): different rhythms as conditioned by many structural factors – phonotactics, grammar, ...
- Structural criteria:
  - relevant units (syllable, …)
  - alternation pattern
  - iteration
  - isochrony
- Process criteria:
  - coordinative entrainment of production processes by superordinate oscillator (Cummins)
  - relating linguistic information with interacting phrase and syllable (maybe also other) oscillators (Barbosa)

Summary and Outlook – Applications of Timing Analysis

Phonetics:

- direct ‘bottom-up’ phonetic analysis of timing

- Phonology:
  - timing domains in prosodic typology ofe.g. mora, syllable, foot timing (depending on annotation)
- Language teaching
- Musicology
  - e.g. annotated music performances
- Speech technology
  - measuring foreign language phonetic proficiency
  - diagnosis and therapy in speech pathology
  - creation and benchmarking of
    - duration models in natural speech synthesis
    - designing disambiguation models in speech recognition

Towards an Emergent Rhythm Theory: the Context

Multilinear Grammar – Ranks and Interpretations (MLG-RI) Architecture