Prosody: speech rhythms and melodies

7. Speech timing

Dafydd Gibbon

Summer School
Contemporary Phonology and Phonetics
Tongji University 9-15 July 2016

Top-down ('grammar first') approaches:

linguistic domains of timing analysis

The finite depth grammar of the Prosodic Hierarchy

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): phonotactic patterns, stress-bearing unit, (phonetically: local sonority peak)

Mora (μ): tone placement, phonotactic patterns

Segment: smallest 'leaf' element in prosodic hierarchy

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

The finite depth grammar of the Prosodic Hierarchy

Prosodic Category (PC) inventory:

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

Prosodic Hierarchy (PH) ordering:

L = < ***Utt***, ***IP***, ***PhP***, ***PrWd***, ***Ft***, ***syll***, ***mora***, ***seg*** >

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

Structural constraints on PH:

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

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

<u>Headedness</u>:

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

Domains of speech timing

- Utterance (Utt): rhythm
- Intonational Phrase (IP): boundary effects, grammatical structure, information structure
- Phonological phrase (PhP): grammatical structure, phrase accent
- Phonological word: word stress, morphological structure
- Foot (φ): ‘stress timing’ hypothesis, stressed-unstressed / strong-weak syllable sequences, SD of durations
- Syllable (σ): ‘syllable timing hypothesis, relative lengths of syllable constituents, SD of durations
- Mora (μ): relative length of mora constituents
- Segment: relative length of C and V, %V, %C, δV, δC
- Subsegment: relative lengths of components of affricates, diphthongs, segments (e.g. stop closure-pause-release, voice onset time)

Domains of speech timing

- Utterance (Utt): rhythm
- Intonational Phrase (IP): boundary effects, grammatical structure, information structure
- Phonologi ture, phrase accent
- Phonolog l structure
- Foot (φ): ' unstressed / strong-we ons
- Syllable (ve lengths of syllable co
- Mora (μ): relative length of mora constituents
- Segment: relative length of C and V, %V, %C, δV, δC
- Subsegment: relative lengths of components of affricates, diphthongs, segments (e.g. stop closure-pause-release, voice onset time)

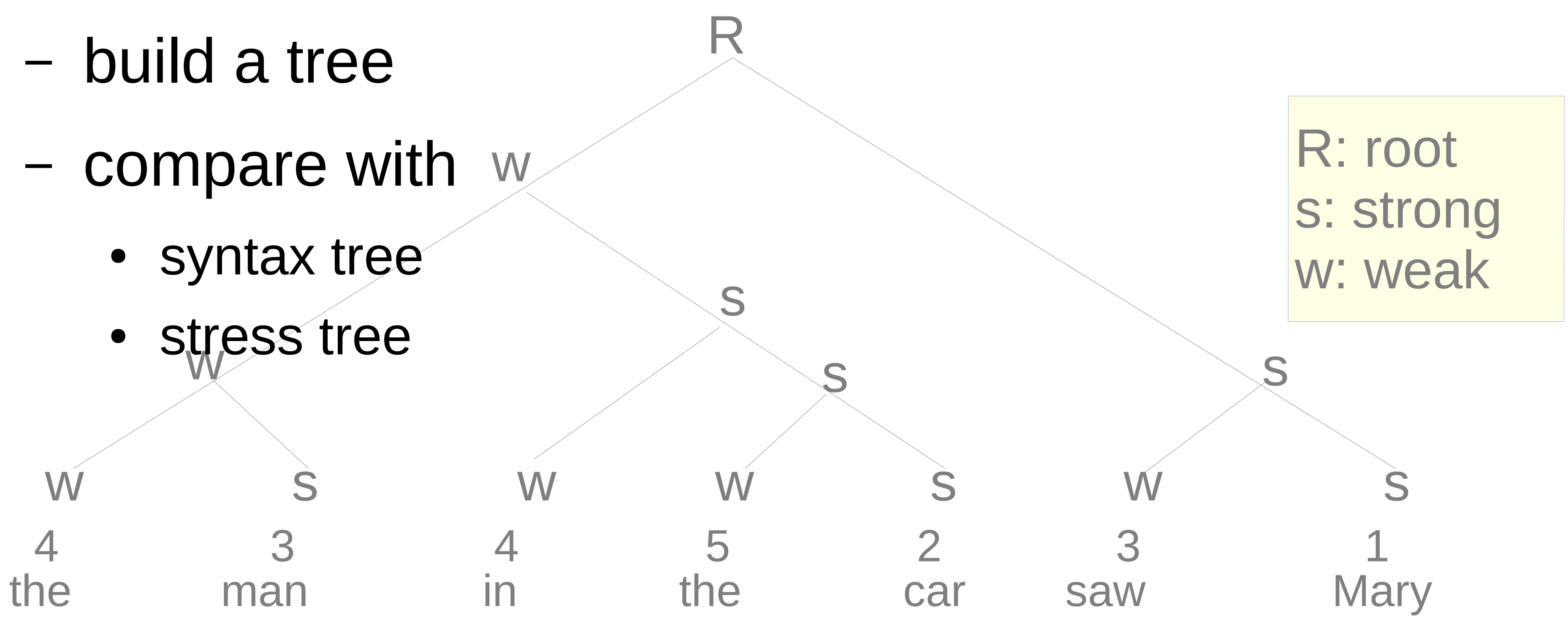
So if you want a complete description of timing for a language / dialect / speaker …

... that's quite a lot of work!

Selected top-down domains: sentence

Timing and stress hierarchies

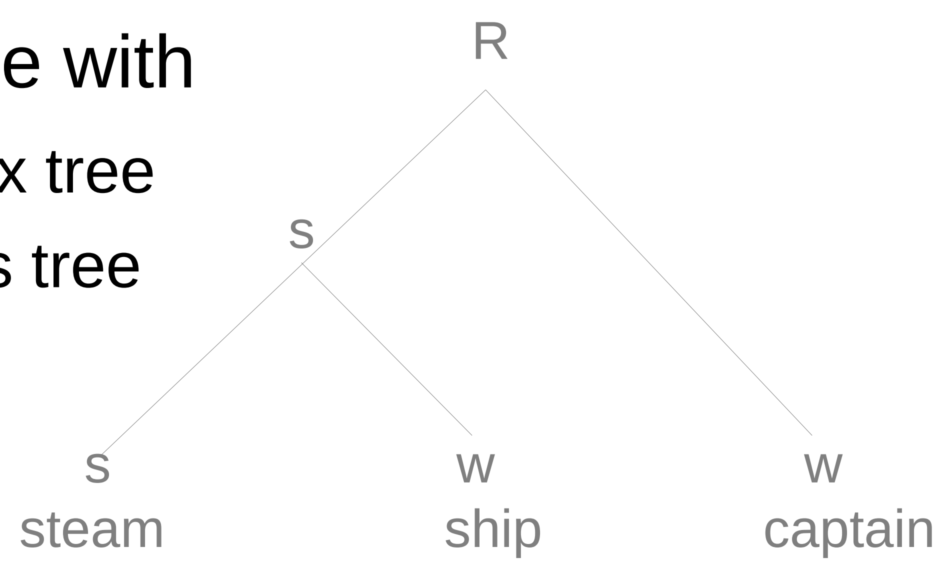
- Syllable timing tends to follow the hierarchical structure of the stress pattern
- Time Tree analysis:
  - compare durations of neighbouring syllables
  - select criterion (longest/shortest item last/first)
  - build a tree
  - compare with
    - syntax tree
    - stress tree



Selected top-down domains: word

Timing and stress hierarchies

- Syllable timing tends to follow the hierarchical structure of the stress pattern
- 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

A bottom up (‘phonetics first’) approach:

pauses, syllables and interpausal units

Question:

How are tones and syllable durations related?

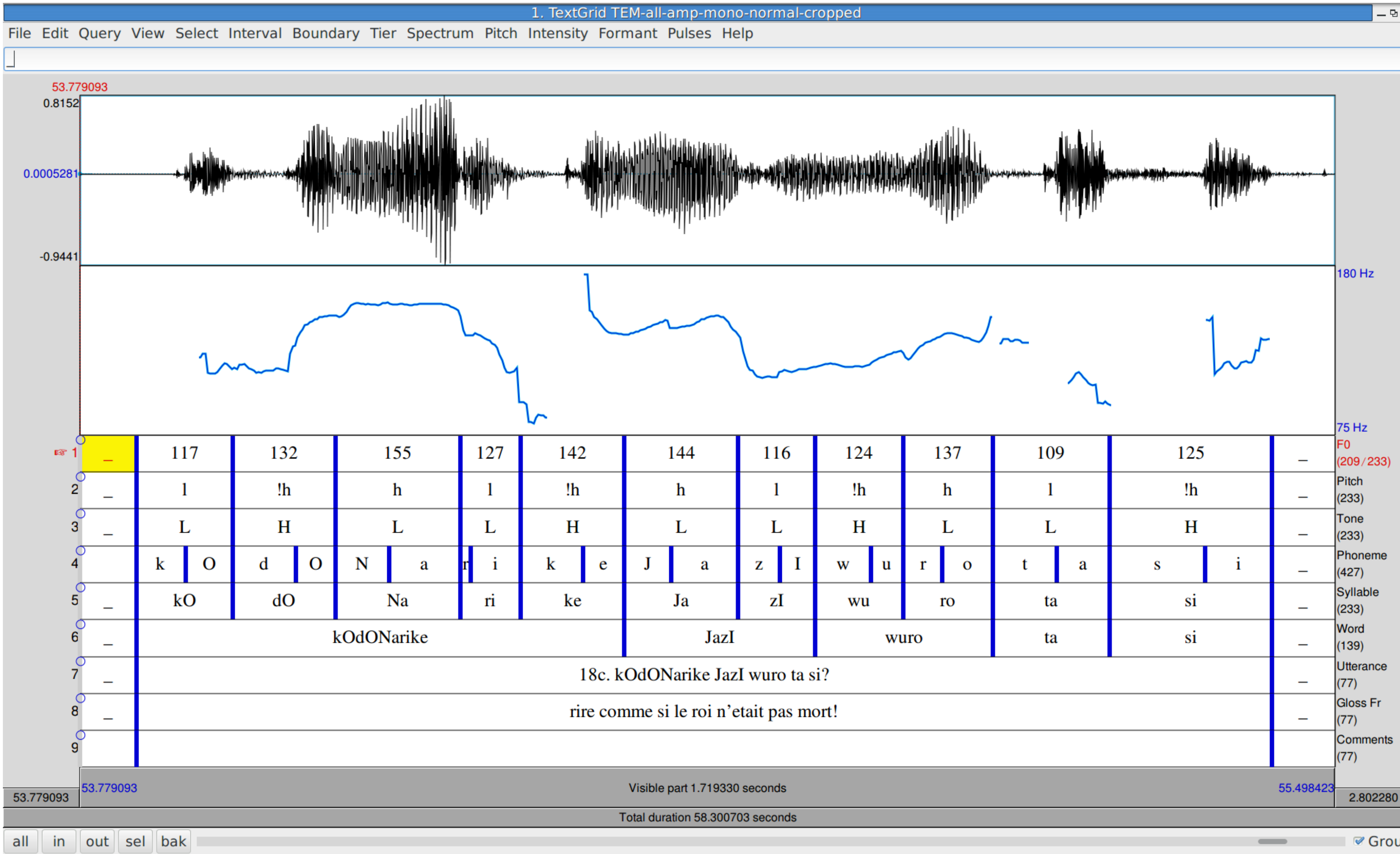
Question:

How are tones and syllable durations related?

What are we looking for?

- Null hypothesis:
  - the syllable is the tone-bearing unit,
  - all syllables have the same duration (syllable timing),
  - therefore the tones have the same duration.

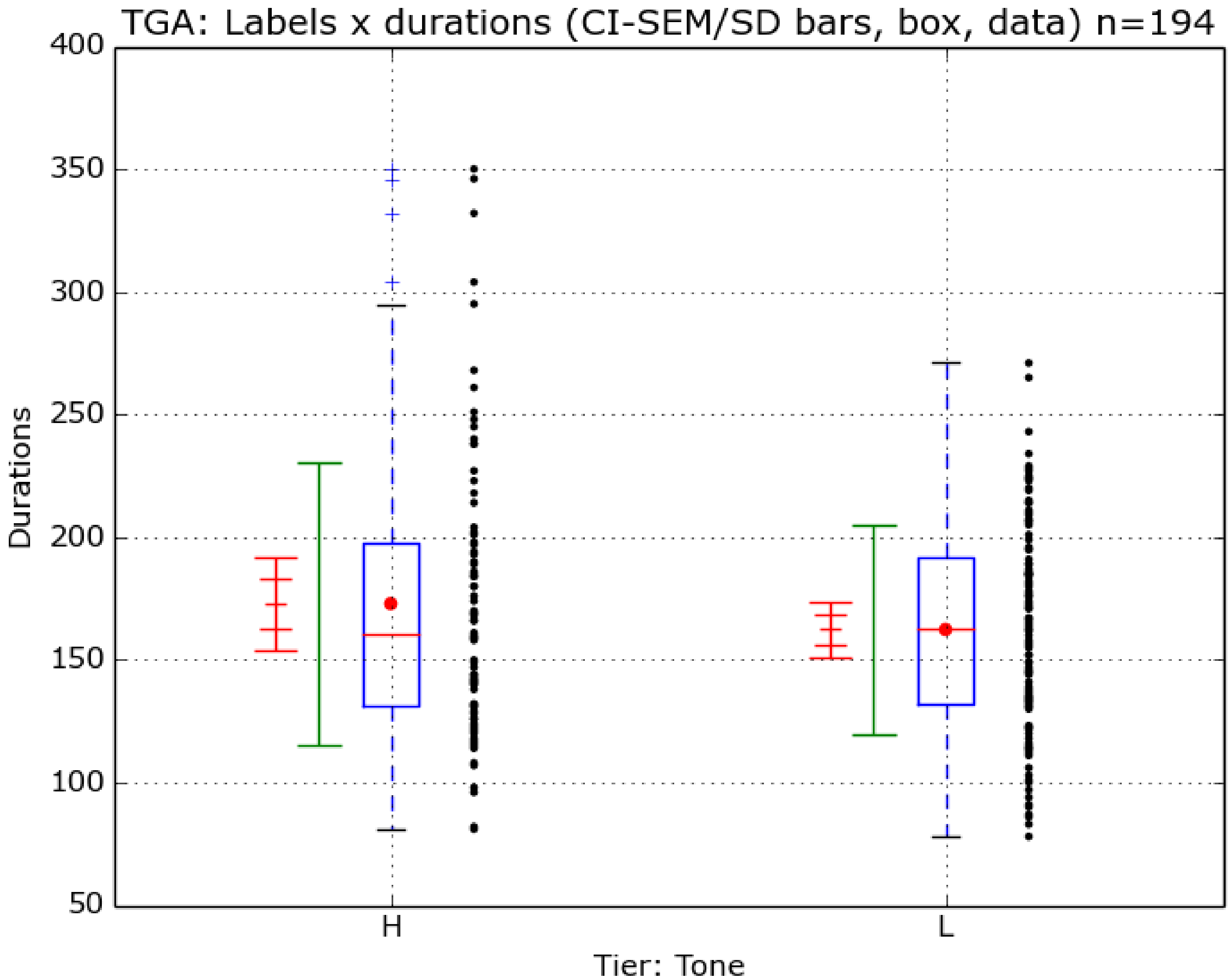
Tem (Niger-Congo)



TEM kodoNa

Tem (tone)

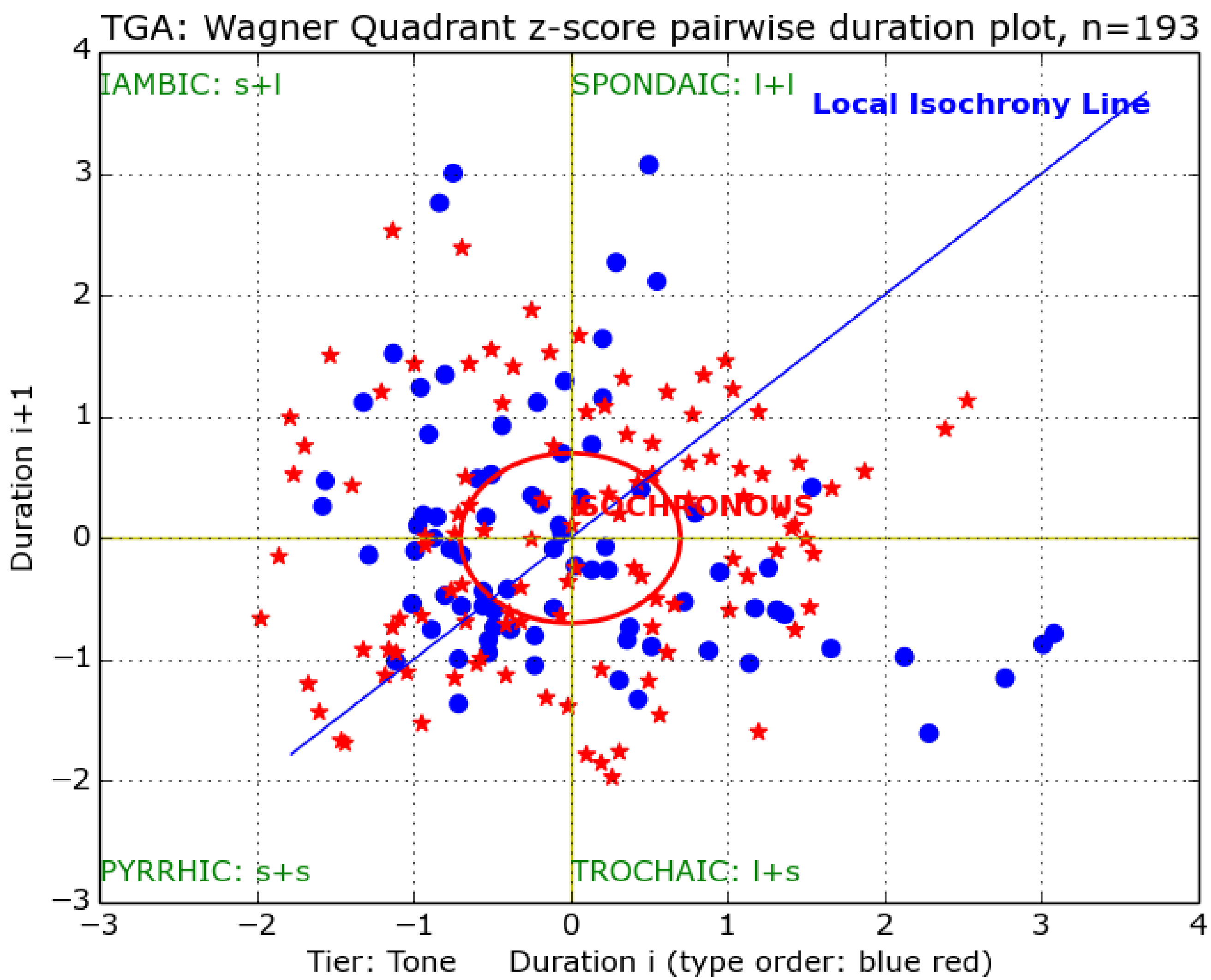
Tem (Niger-Congo)



Result:

- Null hypothesis is not refuted.
- H and L syllables tend to have the same length.

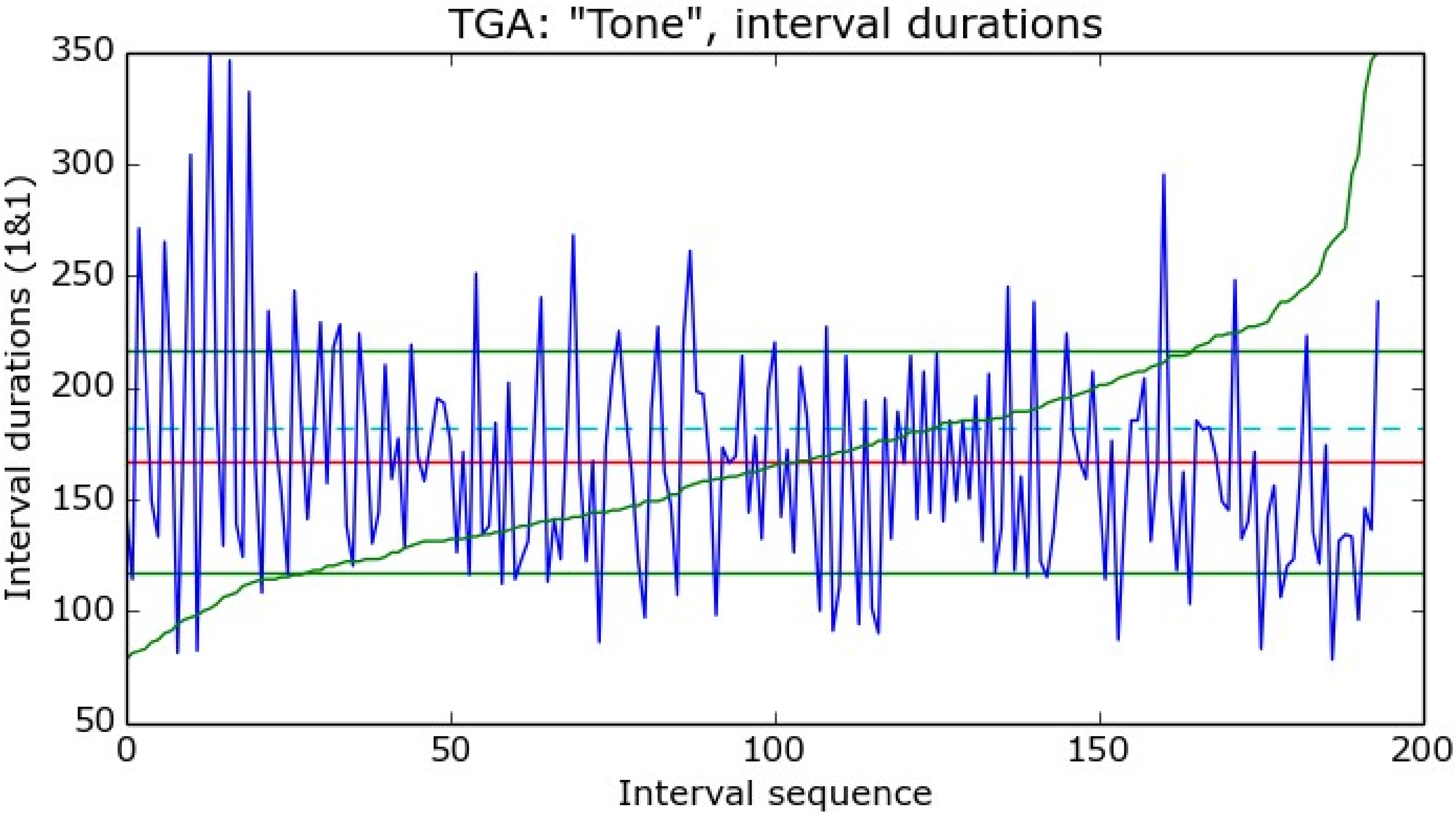
Tem (Niger-Congo)



Quasi-isochrony:

- Durations of neighbouring syllable pairs tend to be scattered randomly around the zero in the normalised z-scores (zero is the mean duration).

Tem (Niger-Congo)

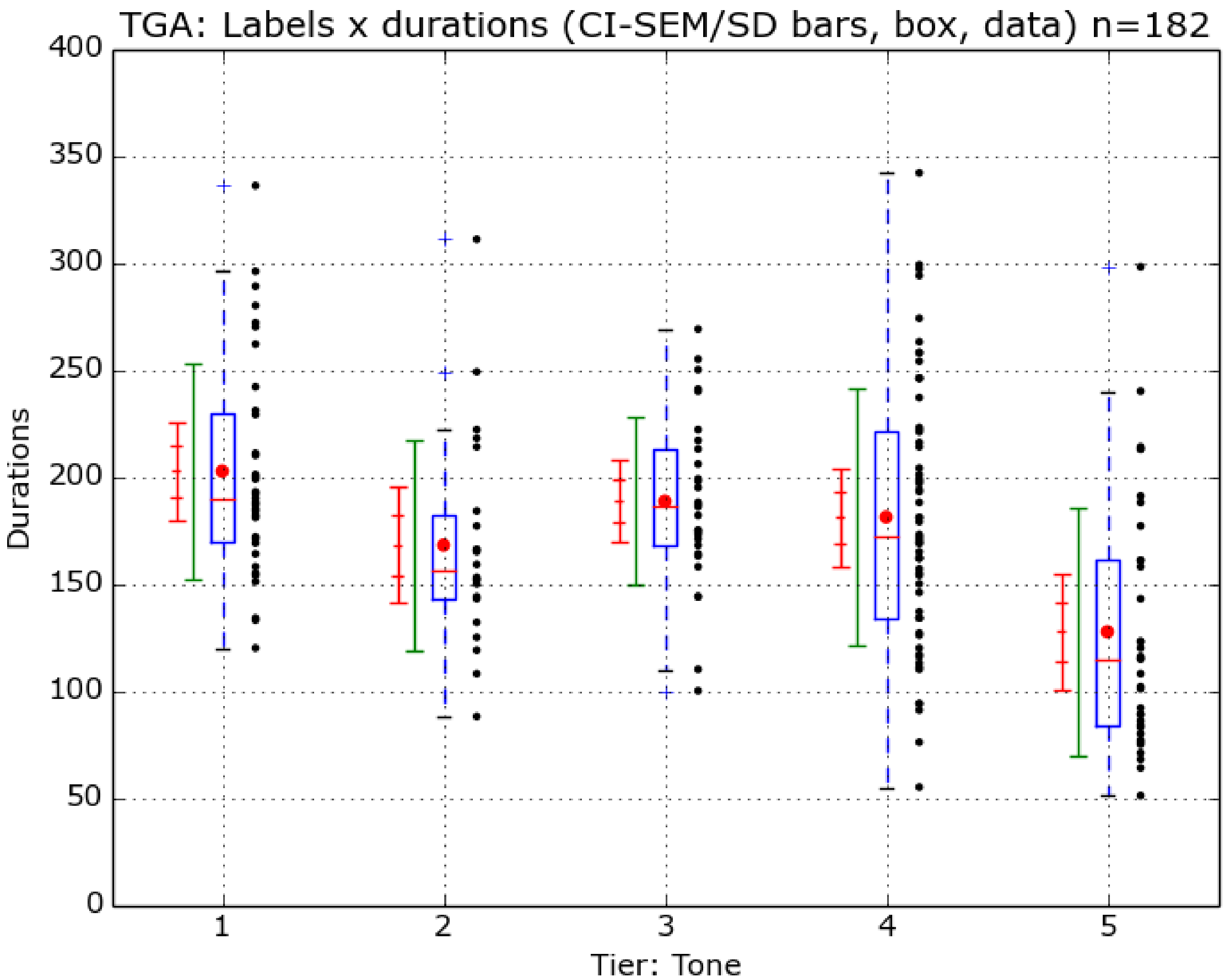


Durations:

- 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)

Mandarin (tone)

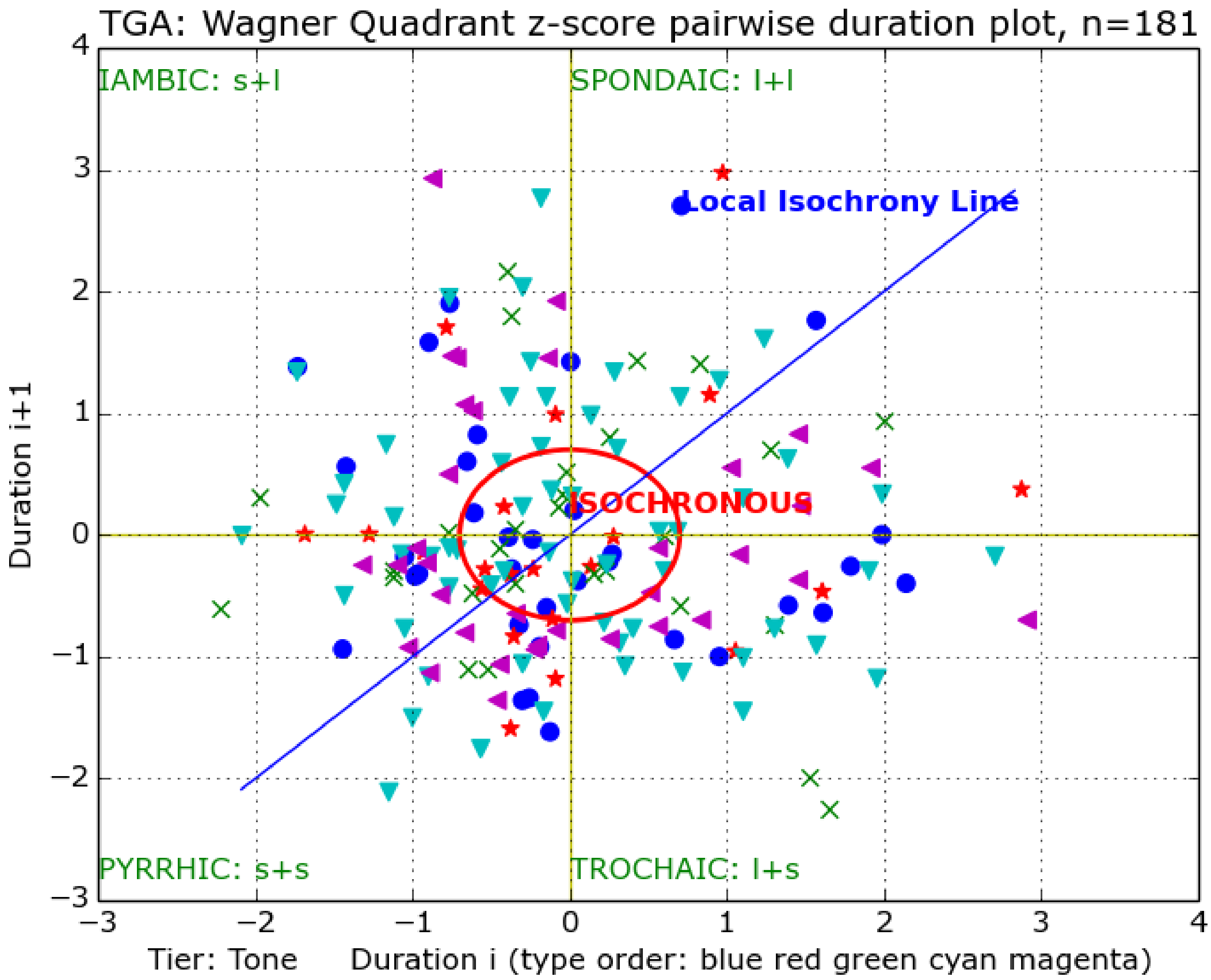
Mandarin (tone)



Result:

- Null hypothesis is not refuted.
- H and L syllables tend to have the same length.

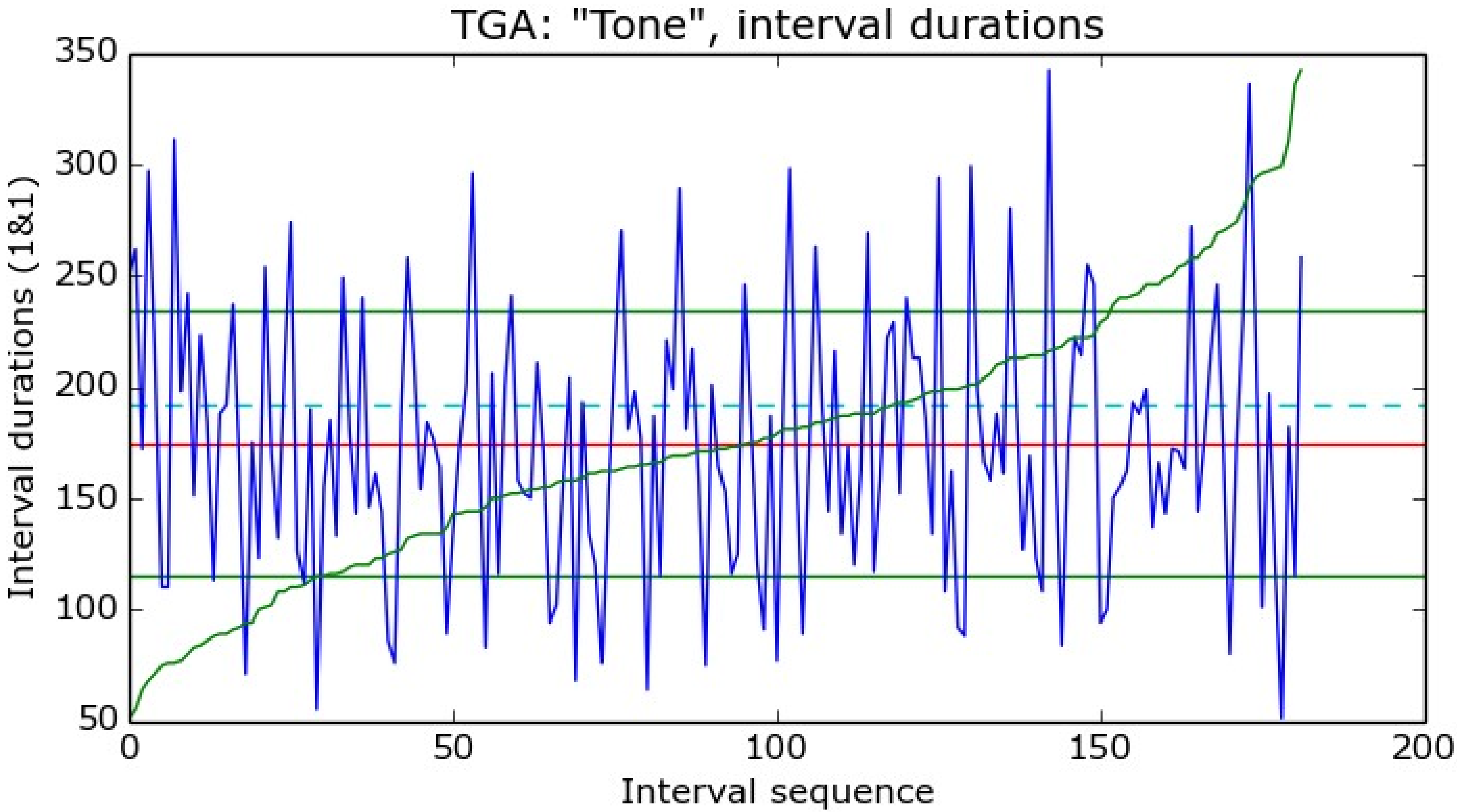
Mandarin (tone)



Quasi-isochrony:

- Durations of neighbouring syllable pairs tend to be scattered randomly around the zero in the normalised z-scores (zero is the mean duration).

Mandarin (tone)

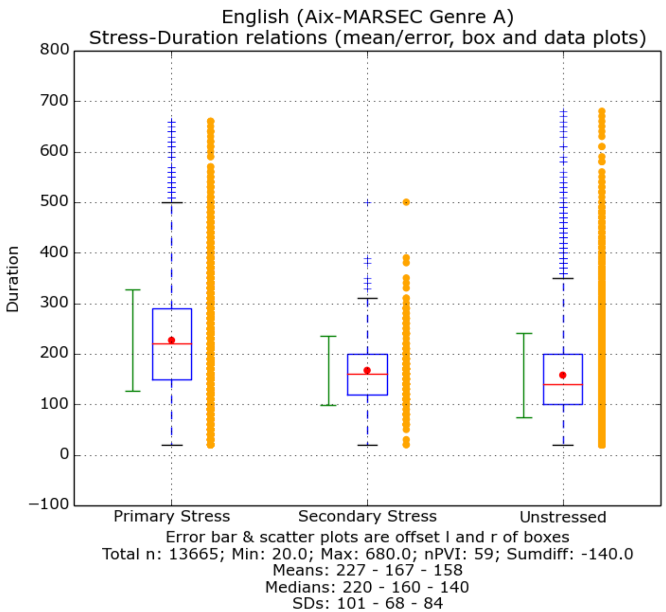


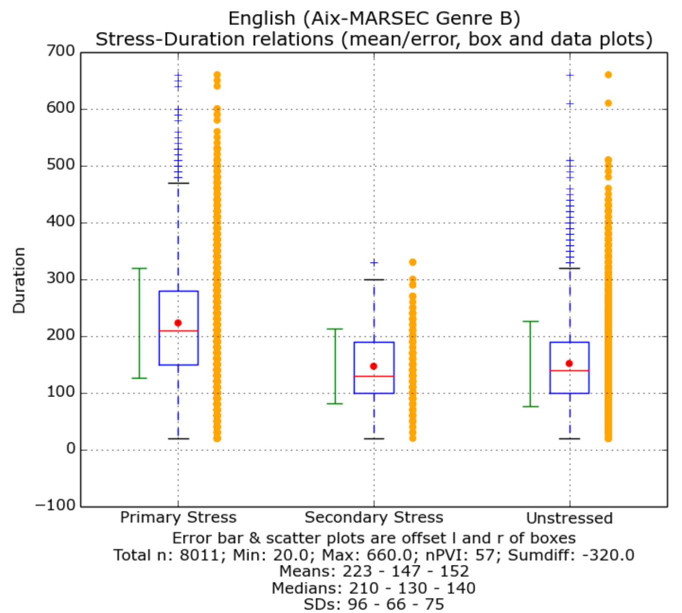
Durations:

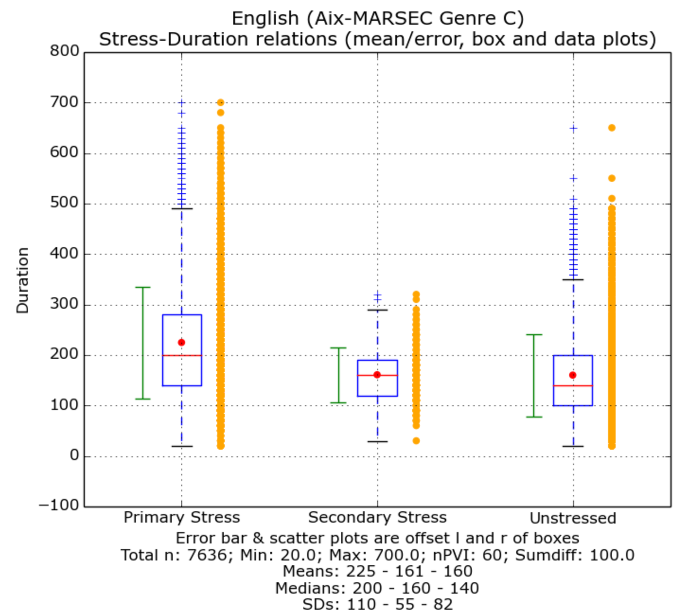
- Durations are similar and vary randomly around the mean.
- As the speech session continues, syllables get shorter and shorter (i.e. the syllable rate gets faster)

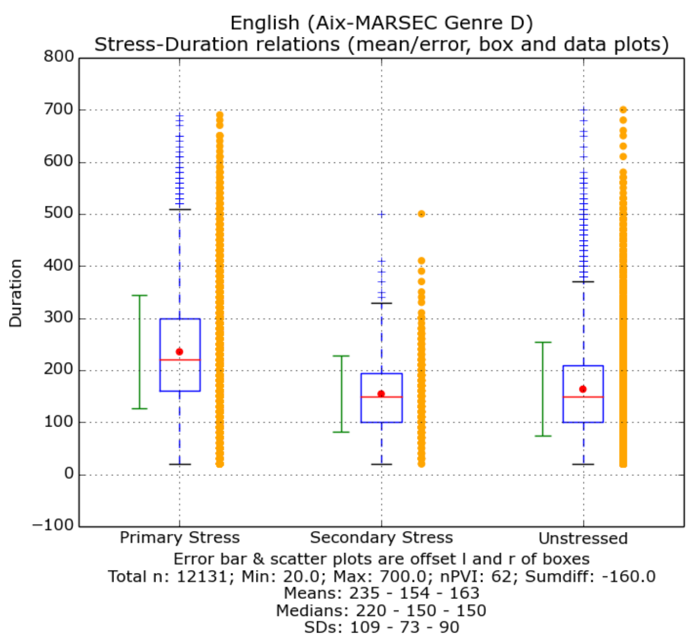
English (stress)

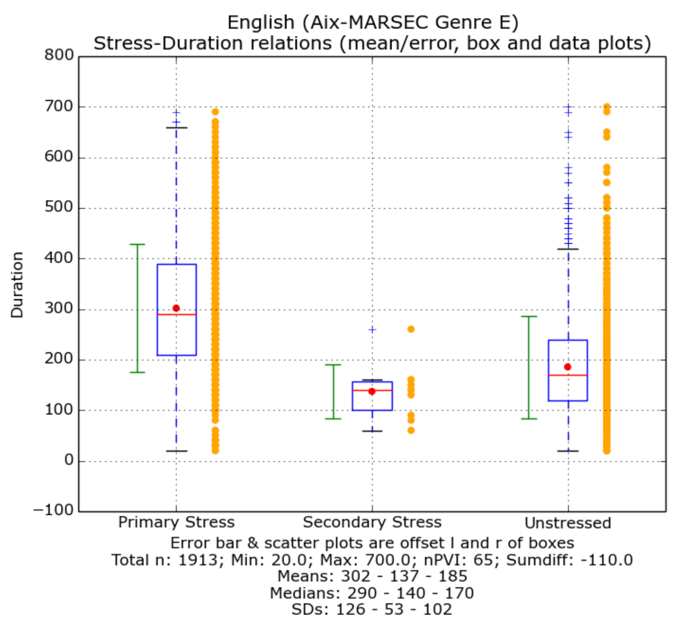
English (stress)

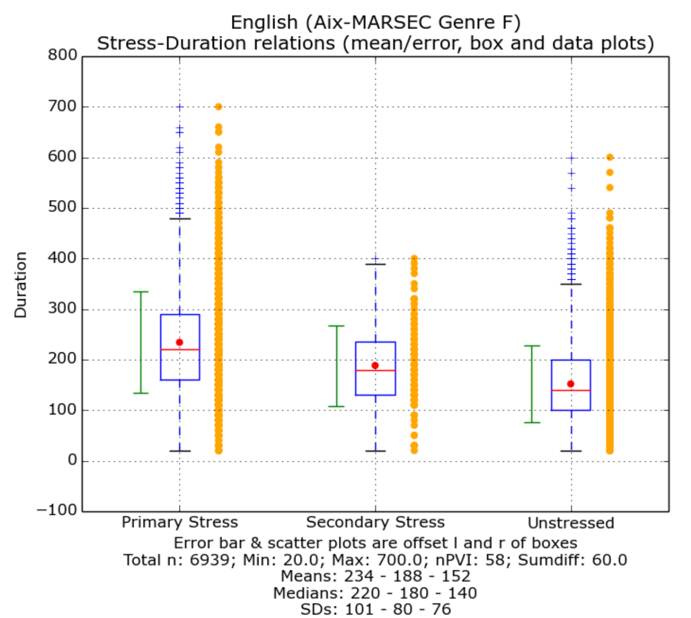




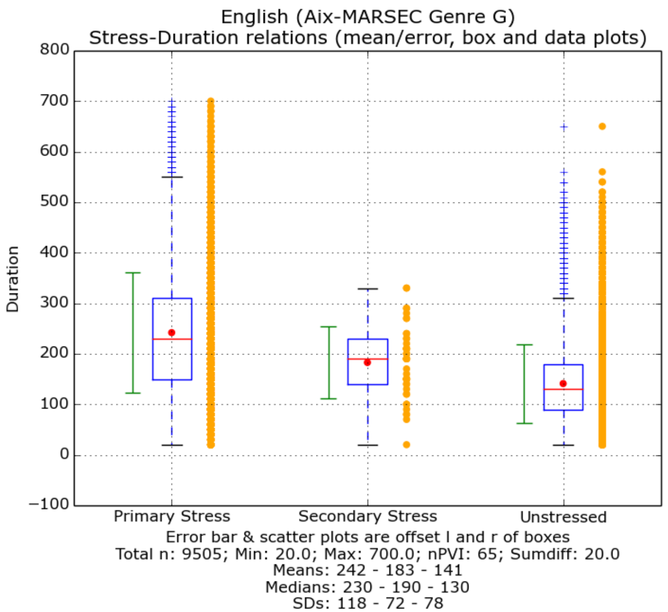


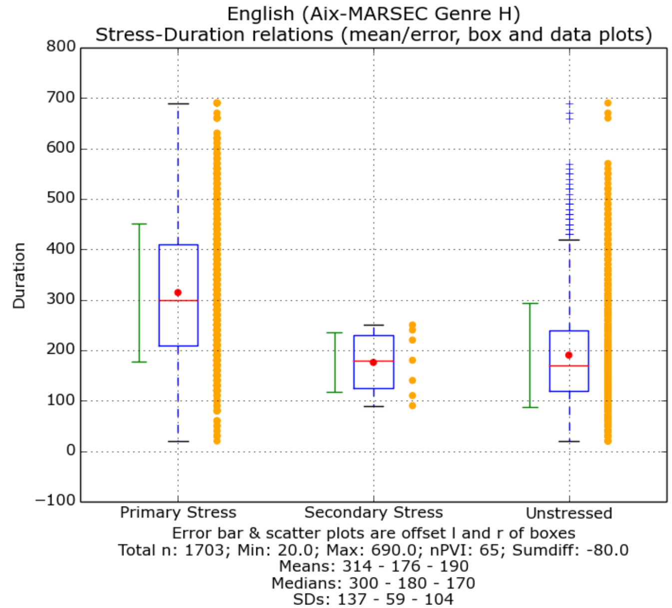


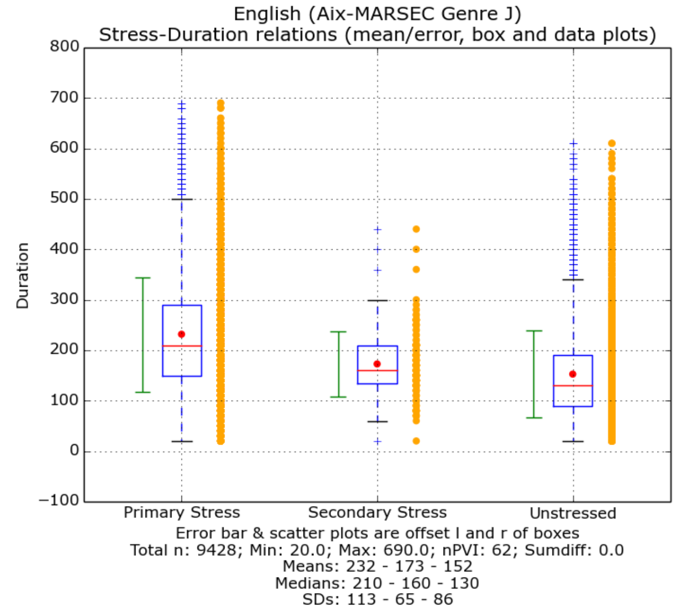


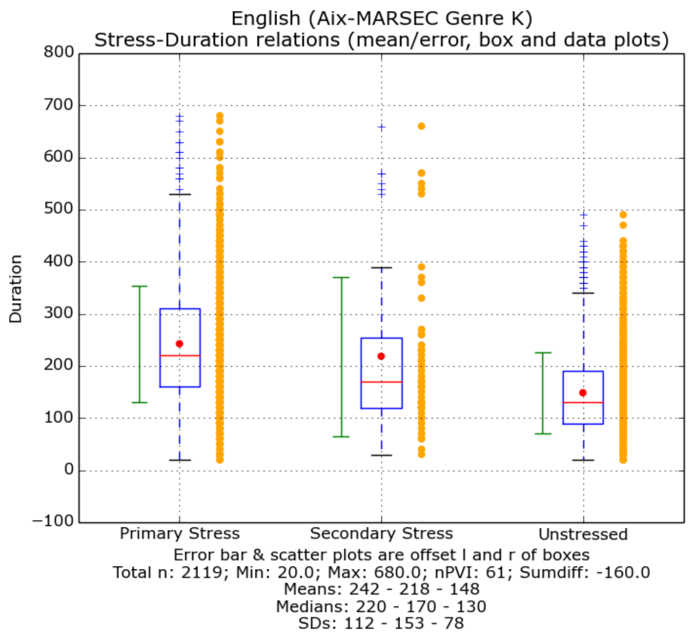


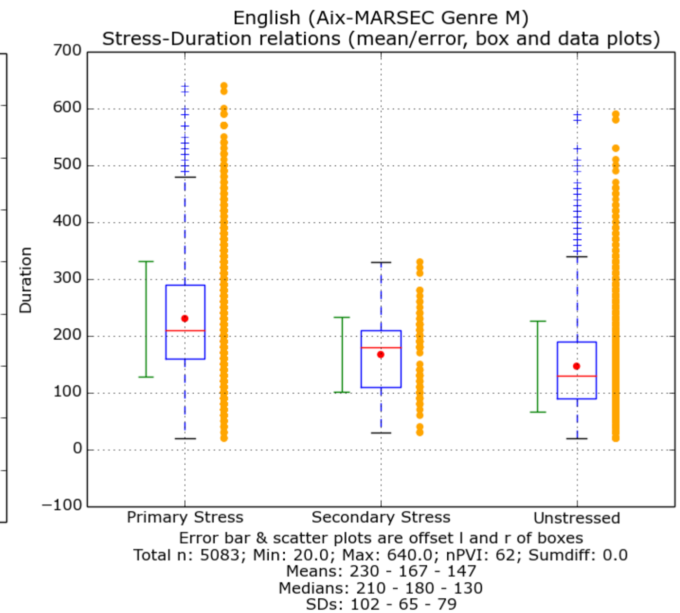
English (stress)



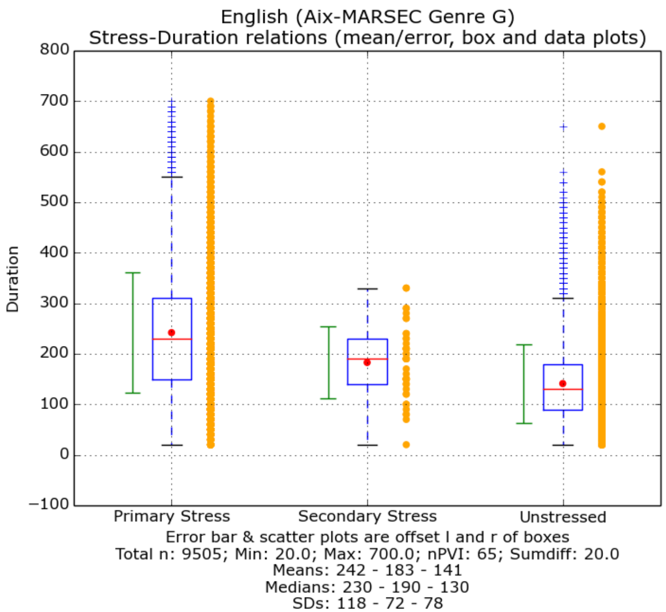


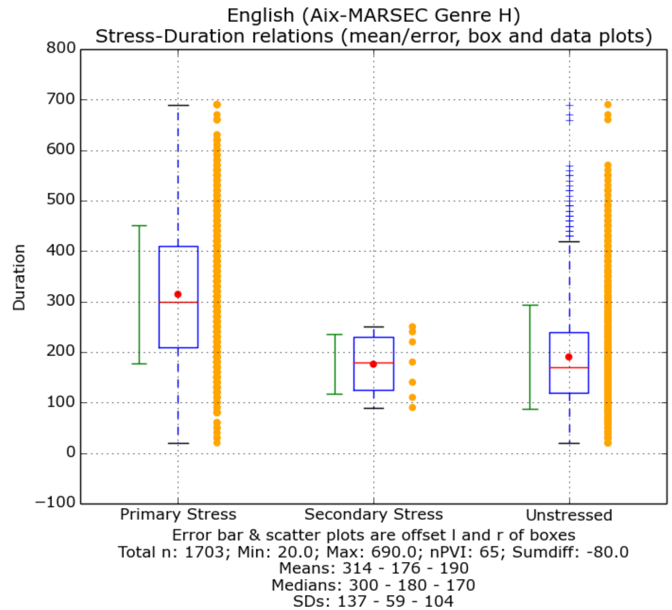


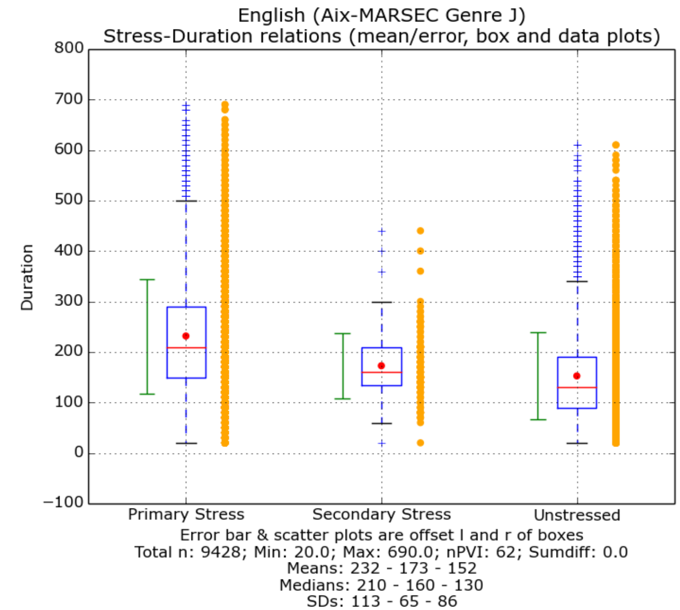


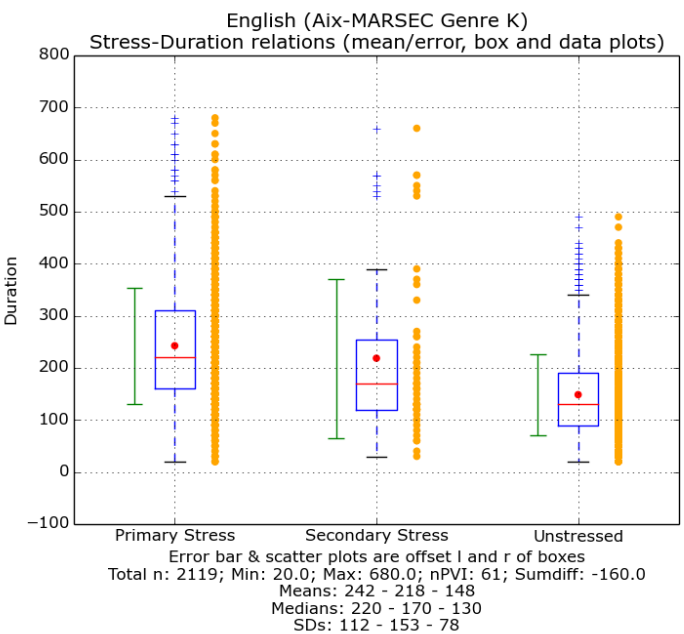


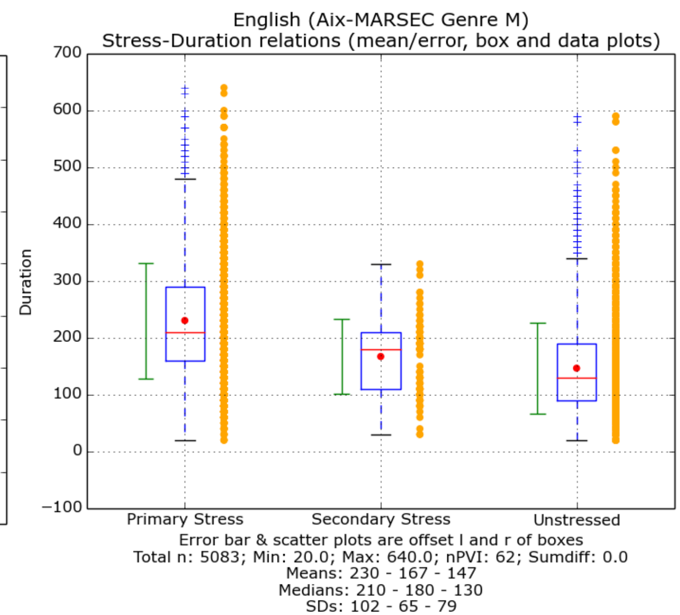
English (stress)











Daniel Hirst & colleagues:
Aix-MARSEC Corpus of Spoken English

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

So-called "rhythm metrics" of duration differences,

which are useful but actually

"relative isochrony metrics"
"regularity / irregularity metrics"
"smoothness metrics"

(rather like standard deviation)

and only part of a full rhythm metric!

Temporal relations – isochrony metrics

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

- ‘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

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

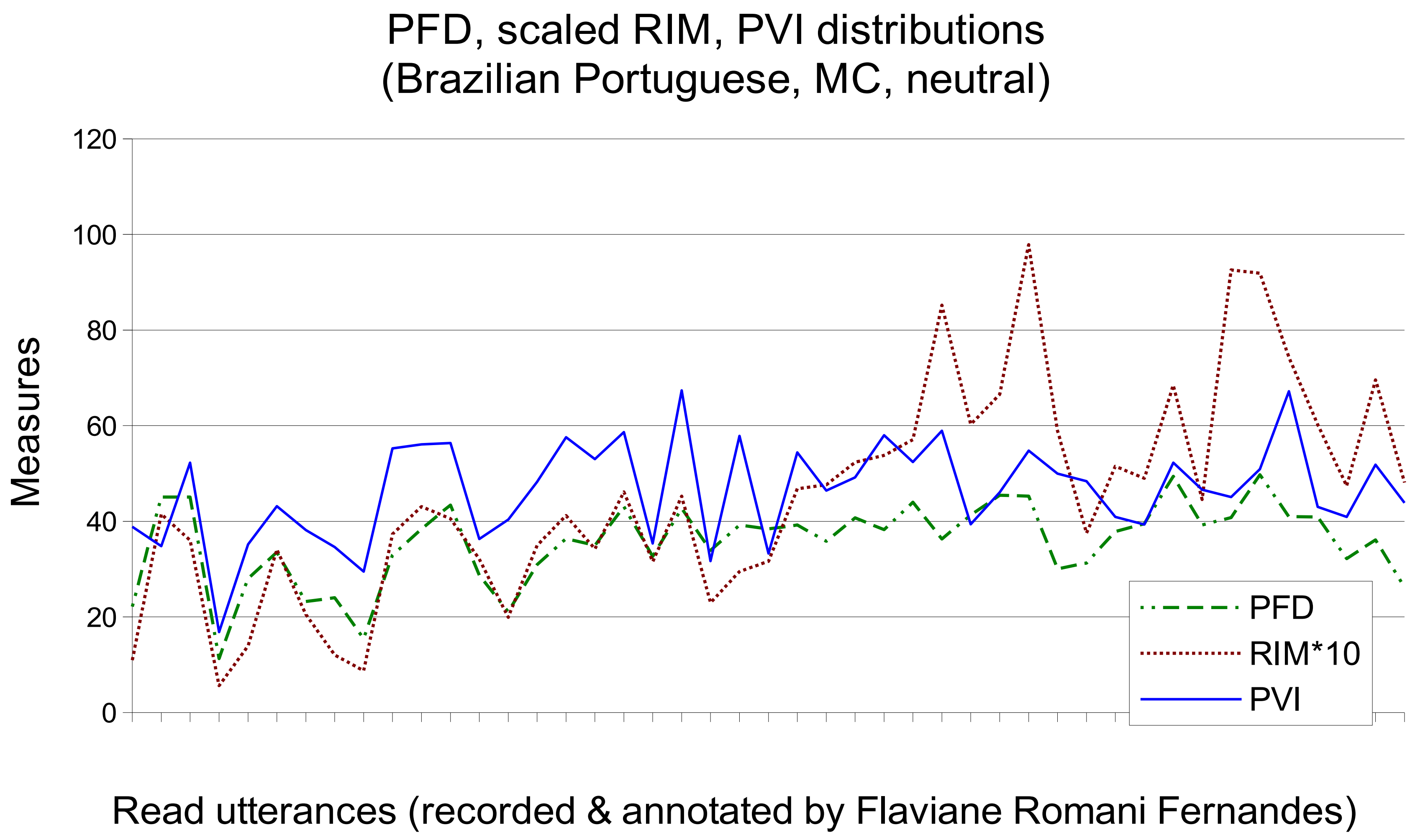
- ‘Rhythm metrics’ of relative ('fuzzy', 'sloppy') isochrony:
  - measures of regularity...irregularity of timing units
    - *PIM*: Pairwise Irregularity Measure
    - *PFD*: Pairwise Foot Difference
    - *rPVI*, *nPVI*: raw and normalised Pairwise Variability Index
  - not rhythm, though: they ignore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



Critique of smoothness approaches

There are many other isochrony / irregularity measures, perhaps most prominently in the past 5 years the 2-dimensional Ramus model: ΔC X %V

However, isochrony/irregularity is not a sufficient condition (cf. the Rhythm Periodicity Model):

cf. Cummins (2002) on Ramus:

Where is the bom-di-bom-bom in %V?

In other words:

The isochrony approaches ignore the *ordering and directionality*, of rhythm, *alternation* within Rhythm Units and *iteration* of Rhythm Units.

Consequently, as a first step, *structure* is needed:

Cf. Jassem's analysis:

Structure is missing...

Sequential temporal structure

Overlap relation to locutionary structure:

- Generative Phonology:
  - phrases; POS
- Wagner:
  - Nouns, Numerals, Proper Names
  - Adverbs, Adjectives
  - Verbs, Demonstrative Pronouns, WH-Pronouns
  - 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

Cummins: Hierarchy

Process is also missing...

- Processes:
  - At the level of phonetic/prosodic periodicity
  - Coordinating different levels of activity
  - Resulting in rhythm as an emergent property of different levels
- Cummins & Port (functional coordination & entrainment):
  - “Rhythm is viewed here as the hierarchical organization of temporally coordinated prosodic units ... certain salient events (beats) are constrained to occur at particular phases of an established period ... the establishment of this period serves a coordinative function.”
  - “Rhythm is manifested as the temporal binding of events to specific and predictable phases of a superordinate cycle.”

Timing: a case study of Tem

Timing: a case study of Tem (Gur; ISO 639-2 kth)

N	dur	rate	mean	median	stdev	npvi	median npvi	inter-cept	slope
11	1795	6.13	163.18	162.00	36.57	24	19	132.27	6.18

Above: example of TGA information output for one interpausal unit.

Below: reformatted information.

N:		17
duration (ms):		95
syllable rate:		6.13
mean syllable duration:		163.18
median syllable duration:		162.00
standard deviation of syllable duration:		36.57
nPVI:		24
median-nPVI:		19
linear regression	intercept:	132.27
	slope:	6.18

Timing: a case study of Tem (Gur; ISO 639-2 kth)

Example of TGA information output for a sequence of interpausal units

Duration properties (syllables)

Attributes	Values	Attributes	Values
n:	194	intercept:	185.235
min:	78	slope:	-0.19
max:	350	std:	49.979
mean:	166.8	nPVI:	35
median:	162.0	rPVI:	60
total:	32360	100*rPVI/med:	37
range:	272	nPVI*med/100:	57

Timing: a case study of Tem (Gur; ISO 639-2 kth)

TGA output sample: statistics

Overall duration:	32360	Overall raw longer, ms:	5804	Overall raw shorter, ms:	5711
Overall min:	78.00	Overall max:	350.00	Overall range:	272.00
Valid Time Groups:	38	Overall rate/sec:	01.06.00		

Overall mean:	166.80	Overall median:	162.00	Overall SD:	49.98
Overall npvi:	35.00	Overall intercept:	185.23	Overall slope:	-0.18

Mean of means:	171.36	Median of means:	167.60	SD of means:	18.02.14
Mean of medians:	164.57	Median of medians:	168.25	SD of medians:	21.61
Mean of SDs:	45.61	Median of SDs:	41.03	SD of SDs:	20.41

mean::TGdur:	-0.403	median::TGdur:	-0.201	SD::TGdur:	-0.140
nPVI::TGdur:	-0.199	slope::TGdur:	-0.373	intercept::TGdur:	-0.113
nPVI::mean:	0.242	slope::mean:	0.718	intercept::mean:	0.003
nPVI::median:	-0.053	slope::median:	0.358	intercept::median:	0.266
nPVI::SD:	0.798	slope::SD:	0.840	intercept::SD:	-657

Timing: a case study of Tem (Gur; ISO 639-2 kth)

TGA online tool: visualisation of syllable time relations

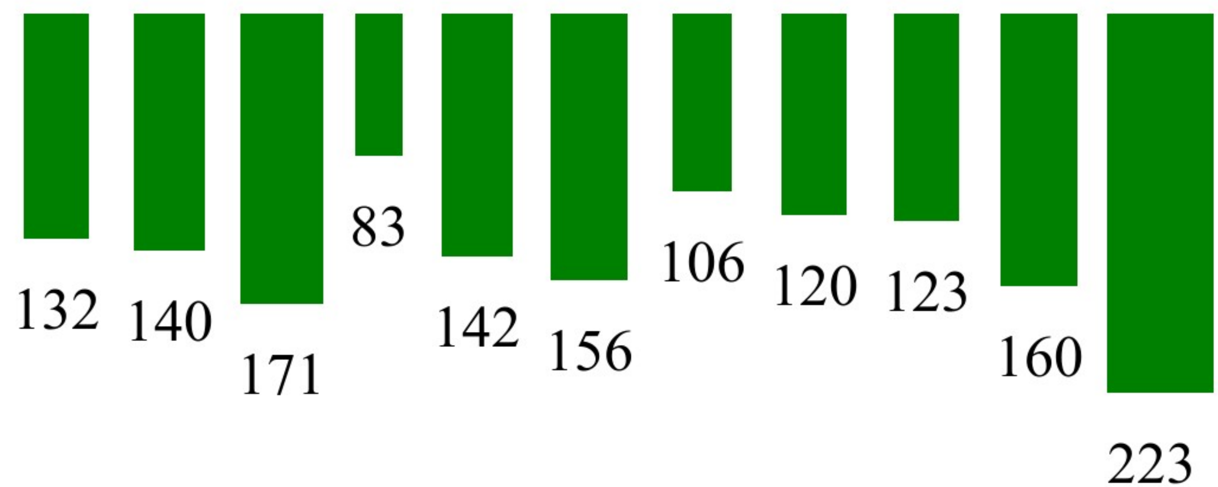
http://wwwhomes.uni-bielefeld.de/gibbon/TGA/

Duration difference tokens: = = / \ = / = = = \

Keyboard friendly transcription: kO dO Na ri ke Ja zI wu ro ta si

2D visualisation of durations:

Syllable durations:



Duration difference tokens for utterance #37:

- pos, neg, equal differences between neighbours: /, \, =
- difference threshold: 40ms.
- clear indication of syllable isochrony

Timing: a case study of Tem (Gur; ISO 639-2 kth)

n-grams	10ms		20ms		30ms		40ms		50ms		60ms		70ms		80ms	
	%	seq	%	seq	%	seq	%	seq	%	seq	%	seq	%	seq	%	seq
Uni 232	35	\	31	\	30	\	27	=	36	=	43	=	49	=	51	=
	24	/	21	/	19	=	25	\	20	\	17	\	16	+	16	+
	16	+	16	+	19	/	16	+	16	+	16	+	16	#	16	#
Di 194	20	/\	15	/\	14	/\	13	=\	19	==	24	==	31	==	35	==
	15	\/	14	\/	12	\/	11	\#	15	+=	16	+=	19	+=	19	+=
	12	\#	12	\#	11	\#	11	==	13	=\	13	=\	13	=\	12	=\
Tri 156	12	+/\	9	+/\	14	/\	8	=\#	10	===	15	===	19	===	24	===
	11	\/\	8	\/\	12	\/	8	+=\	10	+==	11	+==	14	+==	15	+==
	10	/\/	8	/\/	11	\#	6	+/\	8	+=\	8	+=\	10	==#	11	==#
Quad 118	12	+/\/	9	+/\/	9	+/\/	7	+=\/	8	+=\/	8	====	10	====	14	====
	7	/\/#	6	+=\/	6	+=\/	7	+/\/	5	=\/#	8	+===	10	+===	14	+===
	6	\/\\	5	\/\#	4	\/\#	4	=\/#	5	====	7	=\=#	9	===#	11	===#

Duration difference tokens for utterance #37:

- *n*-grams: unigrams, digrams, trigrams, quadgrams
- thresholds 10...80 ms

Timing: a case study of Tem (Gur; ISO 639-2 kth)

Visualisation of duration difference relation hierarchy:

- Iambic grouping:
  - Greater-than:
    - 10ms: (((kO (dO Na)) (((ri ke) Ja) (((zI wu) (ro ta)) si)))
    - 40ms: ((kO (dO (Na ((ri ke) (Ja (zI (wu (ro (ta si)))))))))
  - Greater-than-or-equal:
    - 10ms: (kO dO) Na ri ke Ja zI (wu ro)
    - 40ms: ((kO dO) Na) ri ((ke Ja) (((zI wu) ro) ta))
- Trochaic:
  - 10ms: (kO (dO (Na ri))) (ke (Ja zI)) wu ro ta si
  - 40ms: (kO (dO (Na ri))) ke (Ja zI) wu ro ta si

Duration difference tokens for utterance #37

Note that unidirectional branching is effectively linear.

Timing: a case study of Tem (Gur; ISO 639-2 kth)

TGA output sample:

Visualisation of interpausal units

(screenshot)

31	4	683	5.86	170.75	173.00	31.19	37	40	155.30	10.30	\ / \ kpo no kpo no 150 196 131 206	kpo:150 no:196 kpo:131 no:206 **PAUSE**:766 # **iambicTTgt:** ((kpo no) ((kpo no) **PAUSE**)) **iambicTTgte:** kpo no kpo no **PAUSE** **trochaicTTlt:** kpo (no kpo) no **PAUSE** **trochaicTTlte:** kpo no kpo no **PAUSE**
32	4	617	6.48	154.25	127.50	53.00	47	57	137.60	11.10	= \ / fe u sO m 117 137 245 118	fe:117 u:137 sO:245 m:118 **PAUSE**:782 # **iambicTTgt:** ((fe (u sO)) (m **PAUSE**)) **iambicTTgte:** (fe u) sO m **PAUSE** **trochaicTTlt:** fe u (sO m) **PAUSE** **trochaicTTlte:** fe u sO m **PAUSE**
33	4	635	6.30	158.75	141.00	48.85	56	64	157.40	0.90	/ \ / fe u sO m 160 115 238 122	fe:160 u:115 sO:238 m:122 **PAUSE**:619 # **iambicTTgt:** ((fe (u sO)) (m **PAUSE**)) **iambicTTgte:** fe u sO m **PAUSE** **trochaicTTlt:** (fe u) (sO m) **PAUSE** **trochaicTTlte:** fe u sO m **PAUSE**
34	4	642	6.23	160.50	151.50	41.26	22	22	106.50	36.00	= = \ Ja na wI sI 115 135 168 224	Ja:115 na:135 wI:168 sI:224 **PAUSE**:714 # **iambicTTgt:** ((Ja (na (wI sI))) **PAUSE**) **iambicTTgte:** ((Ja na) wI) sI **PAUSE** **trochaicTTlt:** Ja na wI sI **PAUSE** **trochaicTTlte:** Ja na wI sI **PAUSE**
35	4	713	5.61	178.25	173.50	18.21	13	7	167.30	7.30	= = \ Ja na wI sI 180 167 159 207	Ja:180 na:167 wI:159 sI:207 **PAUSE**:672 # **iambicTTgt:** (Ja (na ((wI sI) **PAUSE**))) **iambicTTgte:** Ja na wI sI **PAUSE** **trochaicTTlt:** Ja na wI sI **PAUSE** **trochaicTTlte:** ((Ja na) wI) sI **PAUSE**
36	11	1839	5.98	167.18	164.00	51.98	35	37	119.86	9.46	/ \ / \ \ = = / = \ kO dO Na ri ke Ja zI wu ro ta si 155 114 176 87 143 185 185 204 131 164 295	kO:155 dO:114 Na:176 ri:87 ke:143 Ja:185 zI:185 wu:204 ro:131 ta:164 si:295 **PAUSE**:476 # **iambicTTgt:** ((kO ((dO Na) (((ri ke) Ja) (zI (wu (ro (ta si))))))) **PAUSE**) **iambicTTgte:** kO dO Na ri ke ((Ja zI) wu) (ro ta) si **PAUSE** **trochaicTTlt:** (kO dO) (Na ri) ke (Ja (zI (wu ro))) ta si **PAUSE** **trochaicTTlte:** kO dO Na ri ke Ja zI wu ro ta si **PAUSE**
37	11	1795	6.13	163.18	162.00	36.57	24	19	132.27	6.18	= \ / \ = = = = = \ kO dO Na ri ke Ja zI wu ro ta si 152 118 162 103 185 181 182 170 149 145 248	kO:152 dO:118 Na:162 ri:103 ke:185 Ja:181 zI:182 wu:170 ro:149 ta:145 si:248 **PAUSE**:549 # **iambicTTgt:** ((kO ((dO Na) ((ri ke) (Ja (zI (wu (ro (ta si)))))))) **PAUSE**) **iambicTTgte:** kO dO Na ri ke (Ja zI) wu ro ta si **PAUSE** **trochaicTTlt:** kO dO (Na ri) ke Ja zI wu ro ta si **PAUSE** **trochaicTTlte:** (kO dO) Na ri ((((ke Ja) (zI wu)) ro) ta) si **PAUSE**

So what is rhythm if not just relative isochrony?

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

ERT is still too complex and inexplicit to be falsifiable.

Start wih Physical Rhythm Theory (PRT)

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” (Cummins & Port 1998)

Rhythm Periodicity Model

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.

Summary (d*ue to Steele...*), maybe:

SPEECH RHYTHM ≈ HEARTBEAT-PACED PERIODICITY

☺

So what is rhythm if not just relative isochrony?

- For some unit, e.g. syllable, foot a function of both
  - isochrony
  - alternation

  And often at different hierarchical levels

- So a measure of alternation is needed, too:
  - models of oscillation:
    - Barbosa, Wagner, Windmann, ...

Rhythm Periodicity Model

Alternating Rhythm Event (Rhythm Unit)

RE=<PE,NE>

Sequential decomposition:

Prominent Event (Foreground Event) PE = <PP, PI>

Nonprominent Event (Background Event) NE = <NP, NI>

Simultaneous decomposition RE=<RP, RI>

Rhythm Pattern RP

Rhythm interval (cf. isochrony condition) RI

Rhythm Periodicity Model

Alternating Rhythm Event (Rhythm Unit)

RE=<PE,NE>

- Sequential decomposition:
  - Prominent Event (Foreground Event) PE = <PP, PI>
    - Simultaneous decomposition:
      - Prominent Pattern (Foreground Pattern) PP
      - Prominent Interval PI
  - Nonprominent Event (Background Event) NE = <NP, NI>
    - Simultaneous decomposition
      - Nonprominent Pattern NP
      - Nonprominent Interval NI
- Simultaneous decomposition RE=<RP, RI>
  - Rhythm Pattern RP
  - Rhythm interval (cf. isochrony condition) RI

Rhythm Periodicity Model

Alternating Rhythm Event (Rhythm Unit) RE=<PE,NE>

Sequential decomposition:

Prominent Event (Foreground Event) PE = <PP, PI>

Simultaneous decomposition:

Prominent Pattern (Foreground Pattern) PP

Prominent Interval PI

Nonprominent Event (Background Event) NE = <NP, NI>

Simultaneous decomposition

Nonprominent Pattern NP

Nonprominent Interval NI

Simultaneous decomposition RE=<RP, RI>

Rhythm Pattern RP

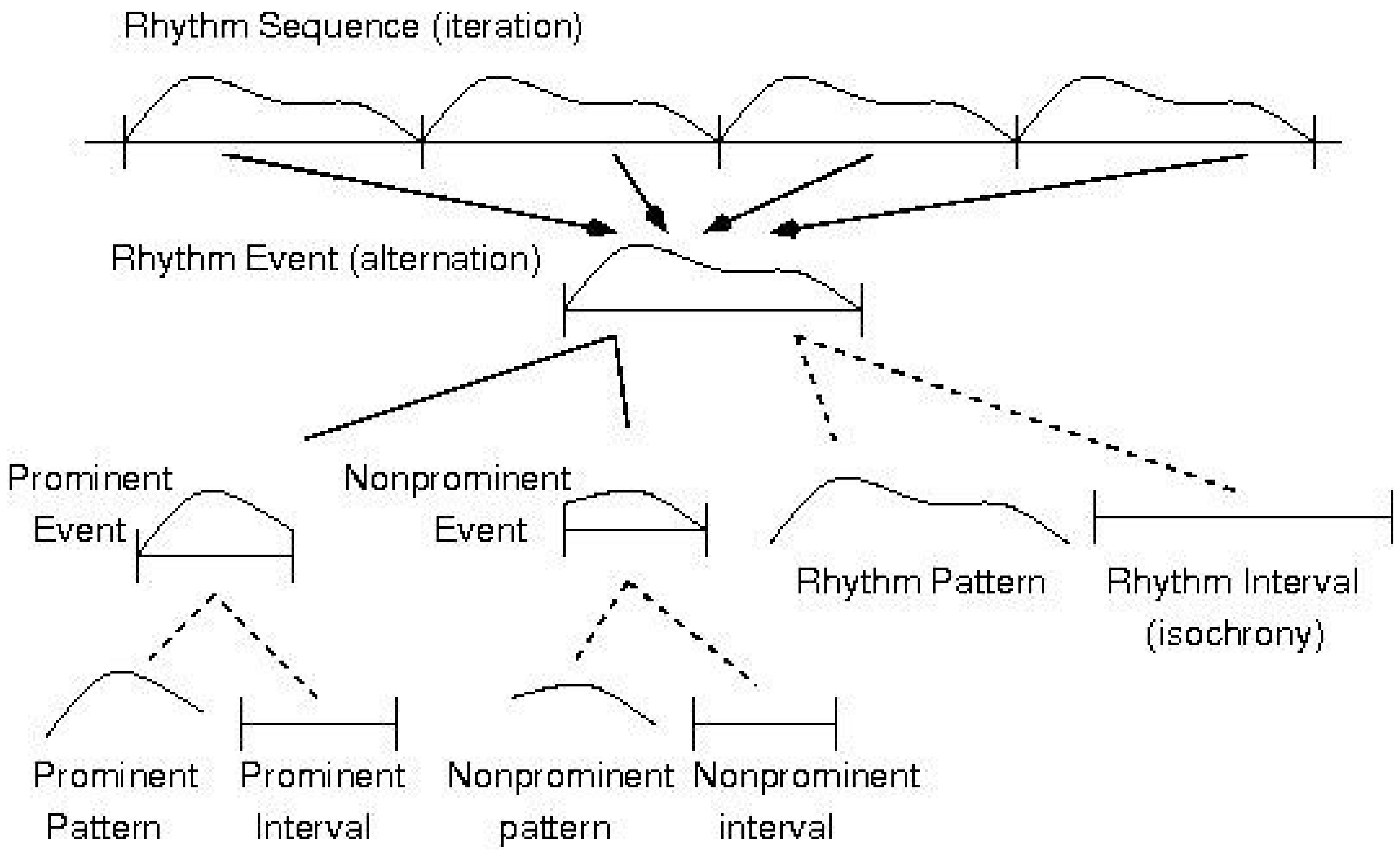
Rhythm interval (cf. isochrony condition) RI

Iteration of Rhythm Units:

Isochrony condition: $\forall$ i, j: $card(RI_i) = card(RI_j)$

Similarity condition: $\forall$ i, j: $RP_i \approx RP_j$

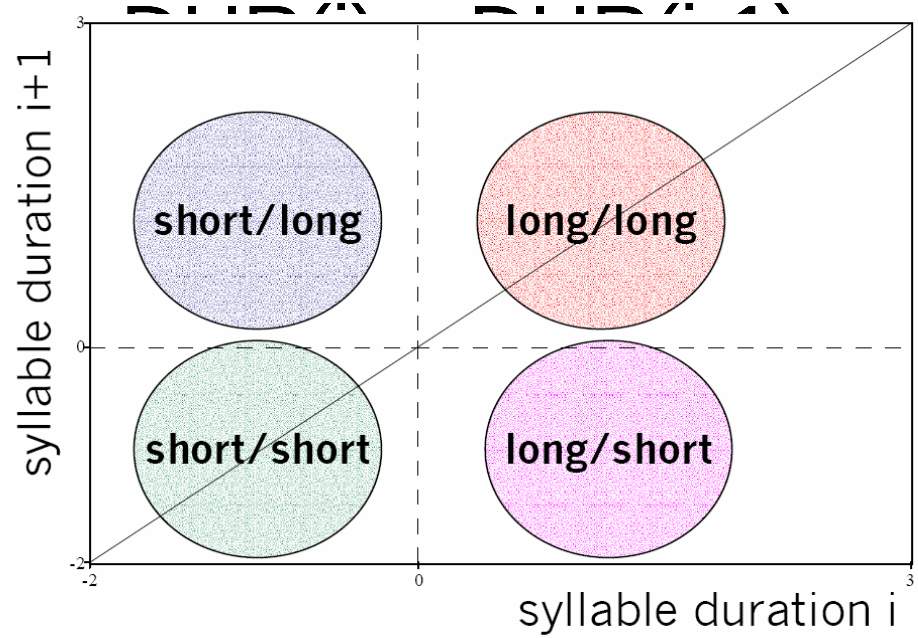
Rhythm Periodicity Model



RECOVERING THE RELATIONS

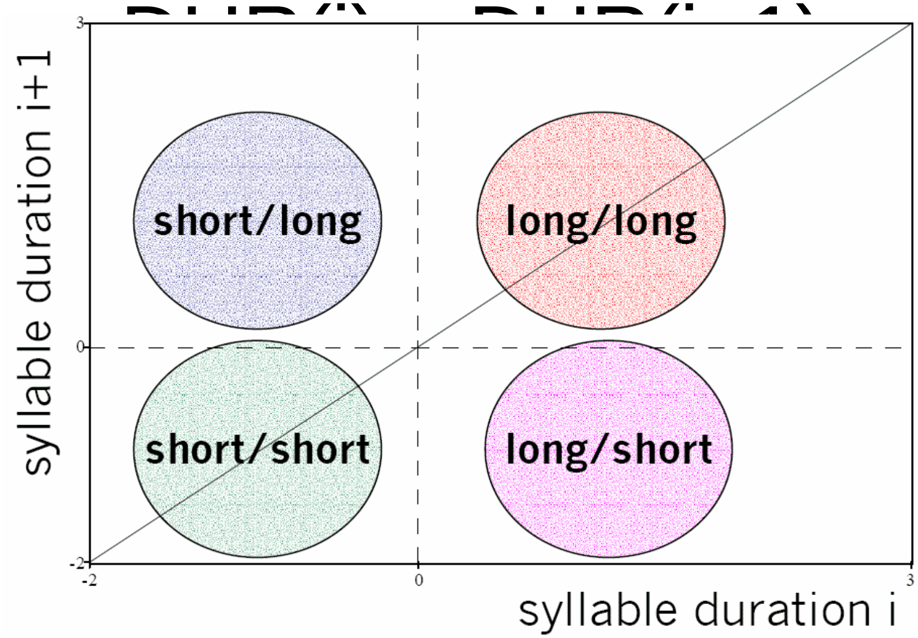
Partial recovery of alternation

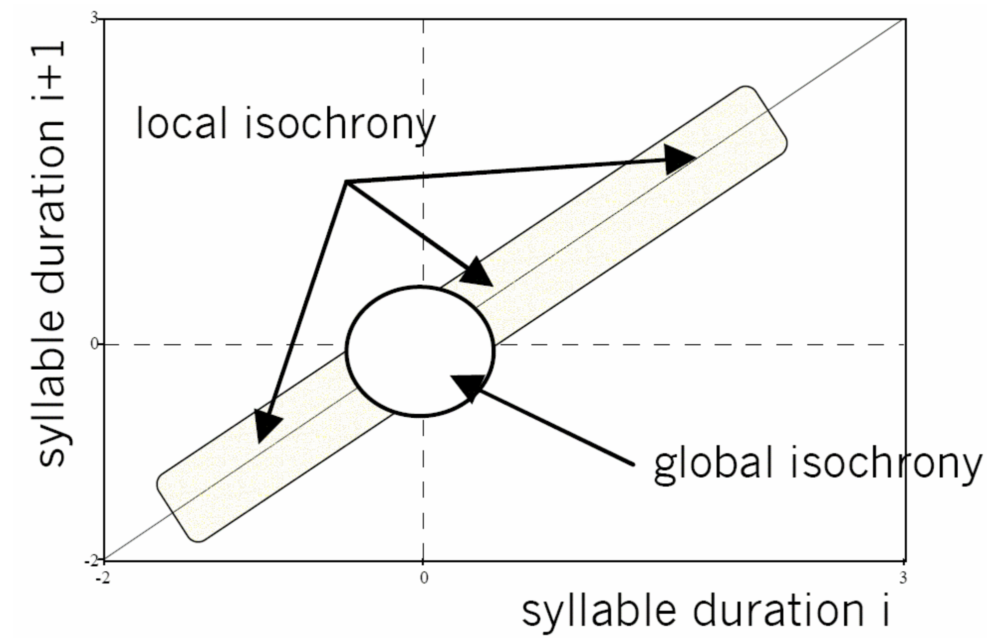
Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting DUR(i) – DUR(i-1)



Partial recovery of alternation

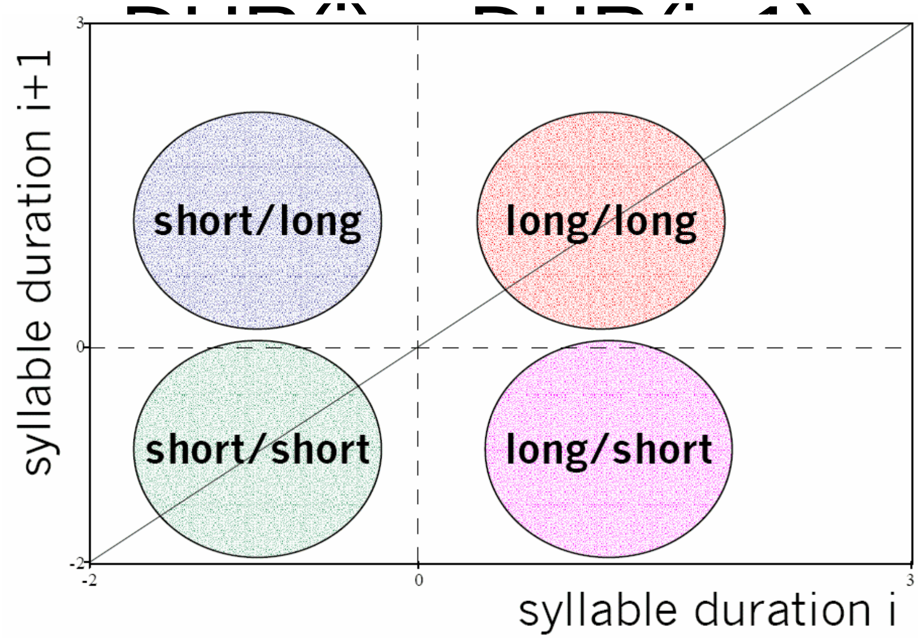
Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting

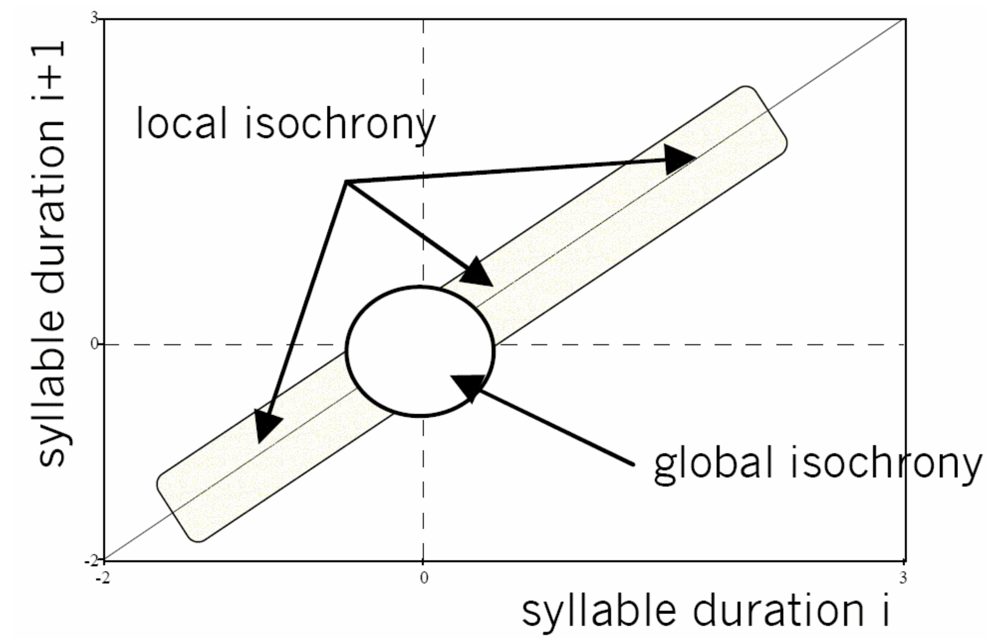




Partial recovery of alternation

Wagner (2006) has a topological procedure for recovering non-absolute differences by plotting

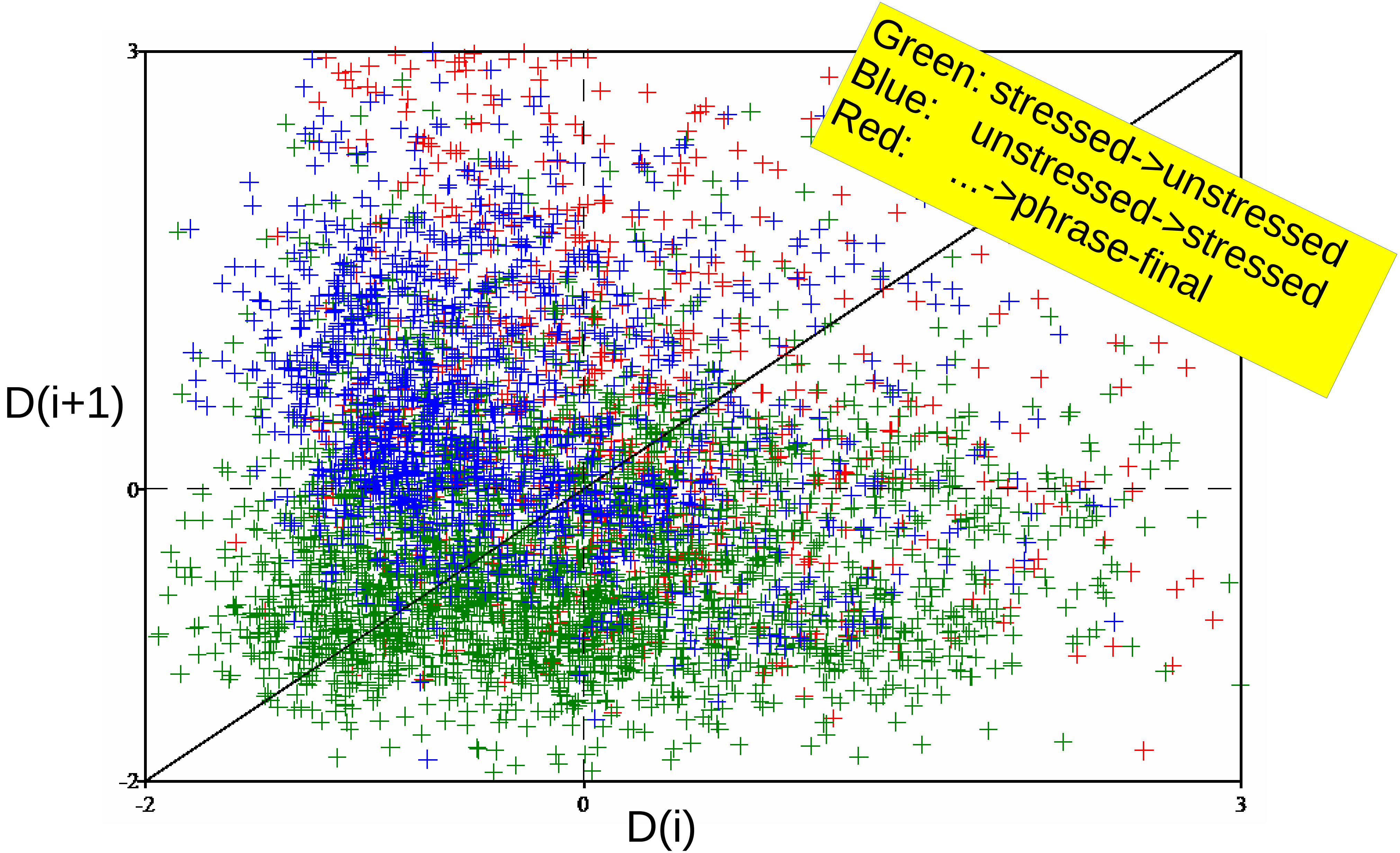




Note: still binary relations only on the surface

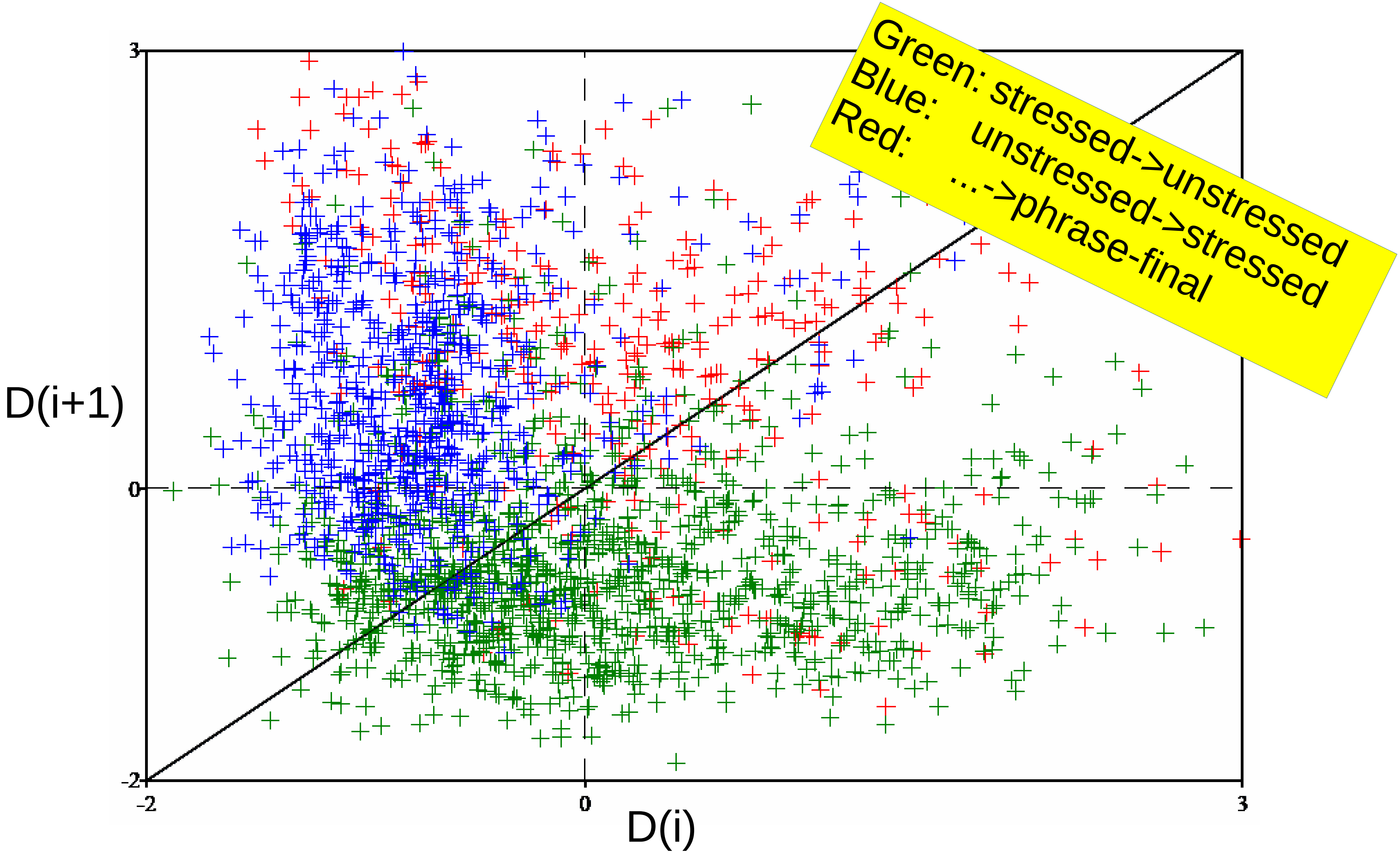
However, 4 quadrants permit distinguishing between long-short & short-long

Binary duration relations: German



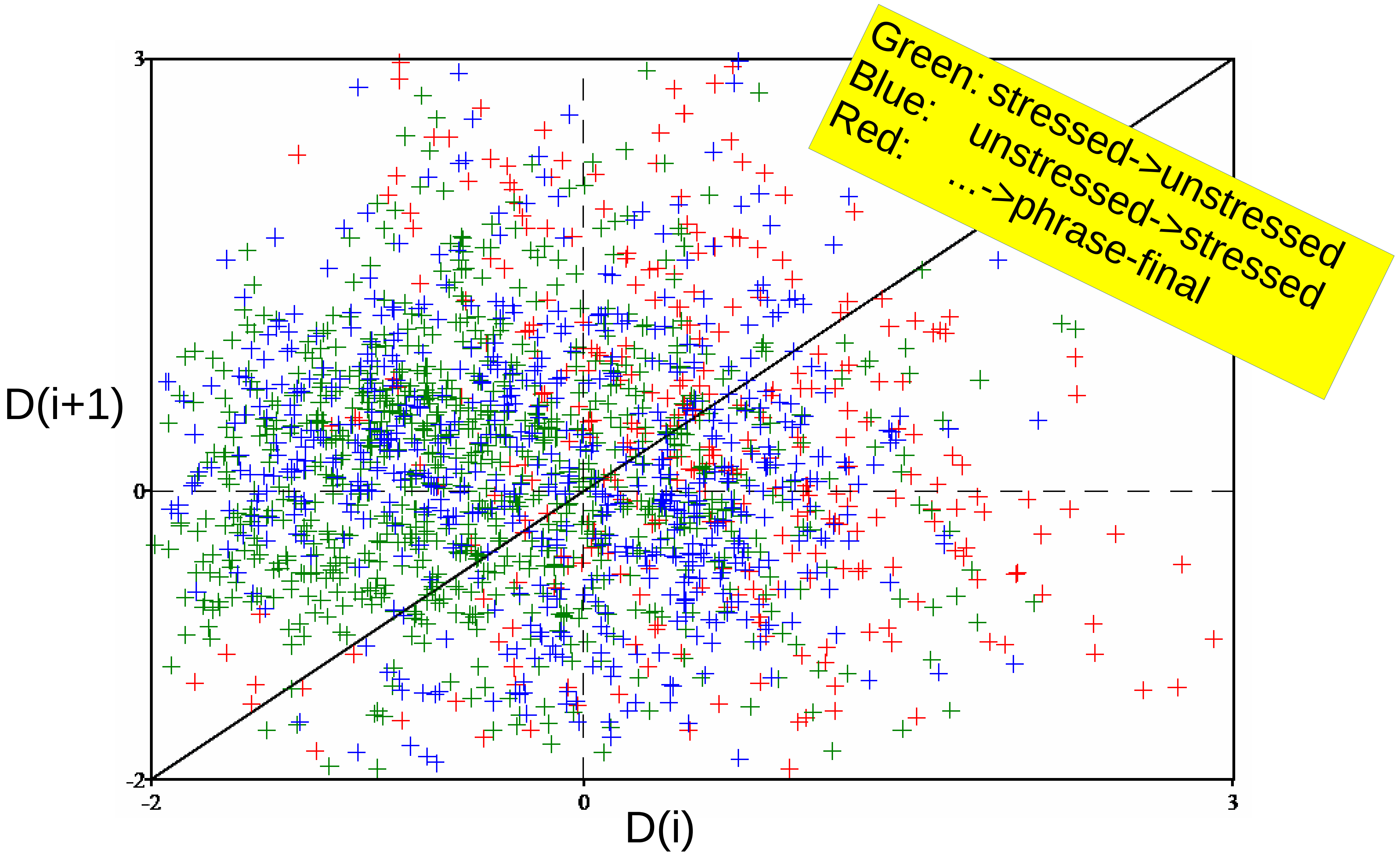
Comment: stress timed - green & blue disjoint

Binary duration relations: English



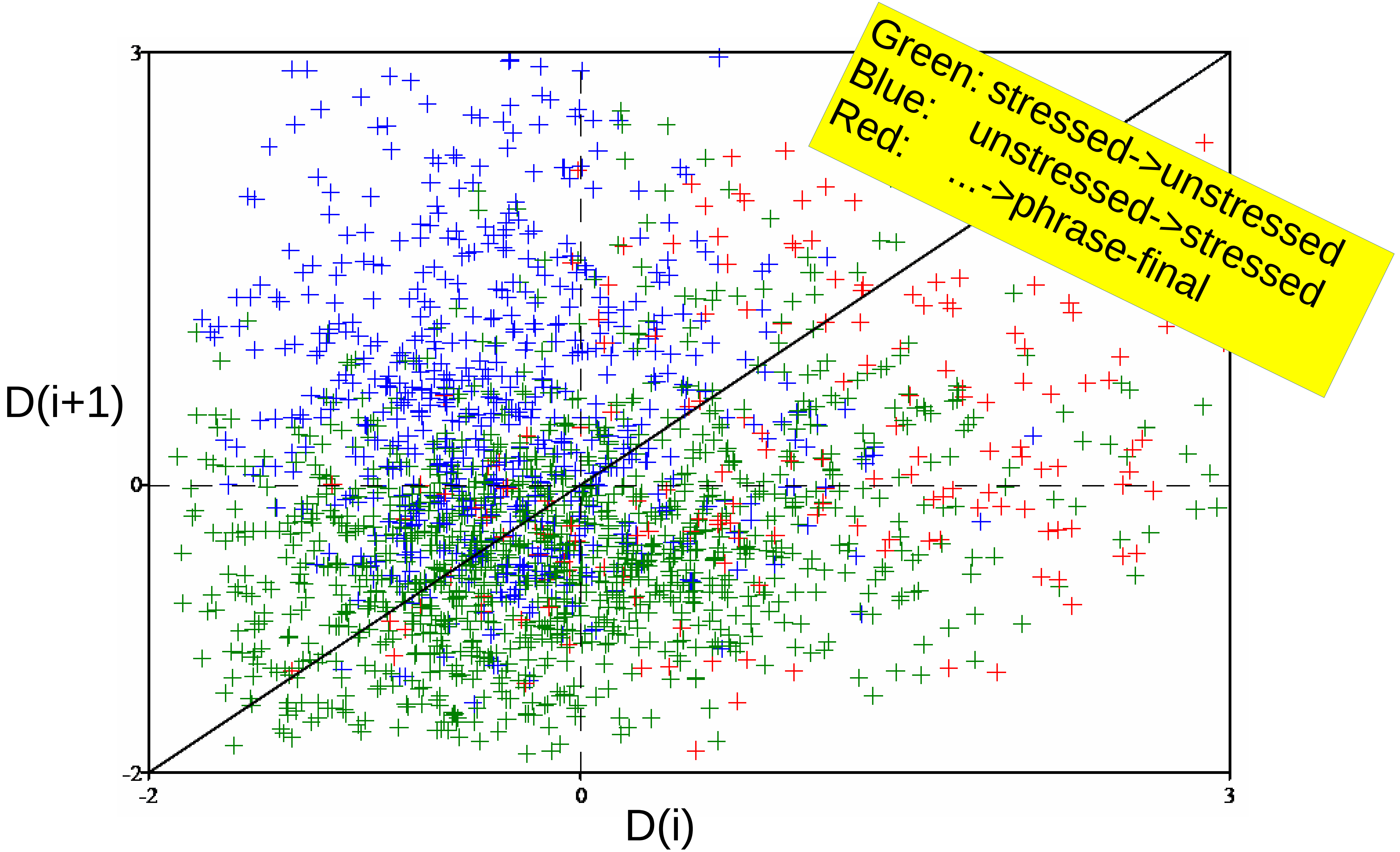
Comment: stress timed - green & blue disjoint

Binary duration relations: French



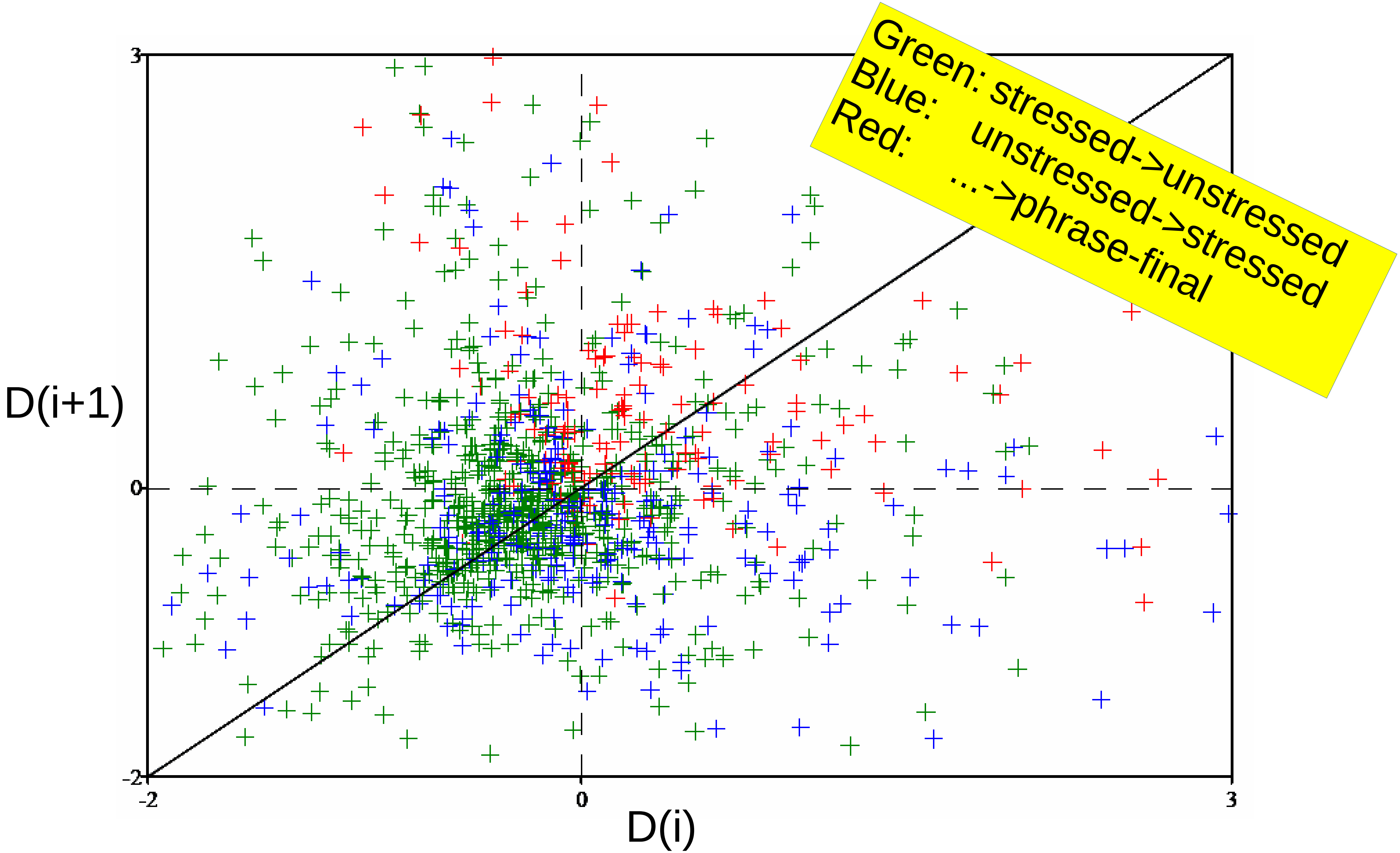
Comment: syllable timed - green & blue overlap

Binary duration relations: Italian



Comment: stress timed - green & blue disjoint

Binary duration relations: Polish



Comment: highly syllable timed - green & blue overlap

DYNAMIC TIMING MODELS

Barbosa's dynamic timing model

Def. "rhythm": speech rhythm is understood as the consequence of the variation of perceived duration along the entire utterance.

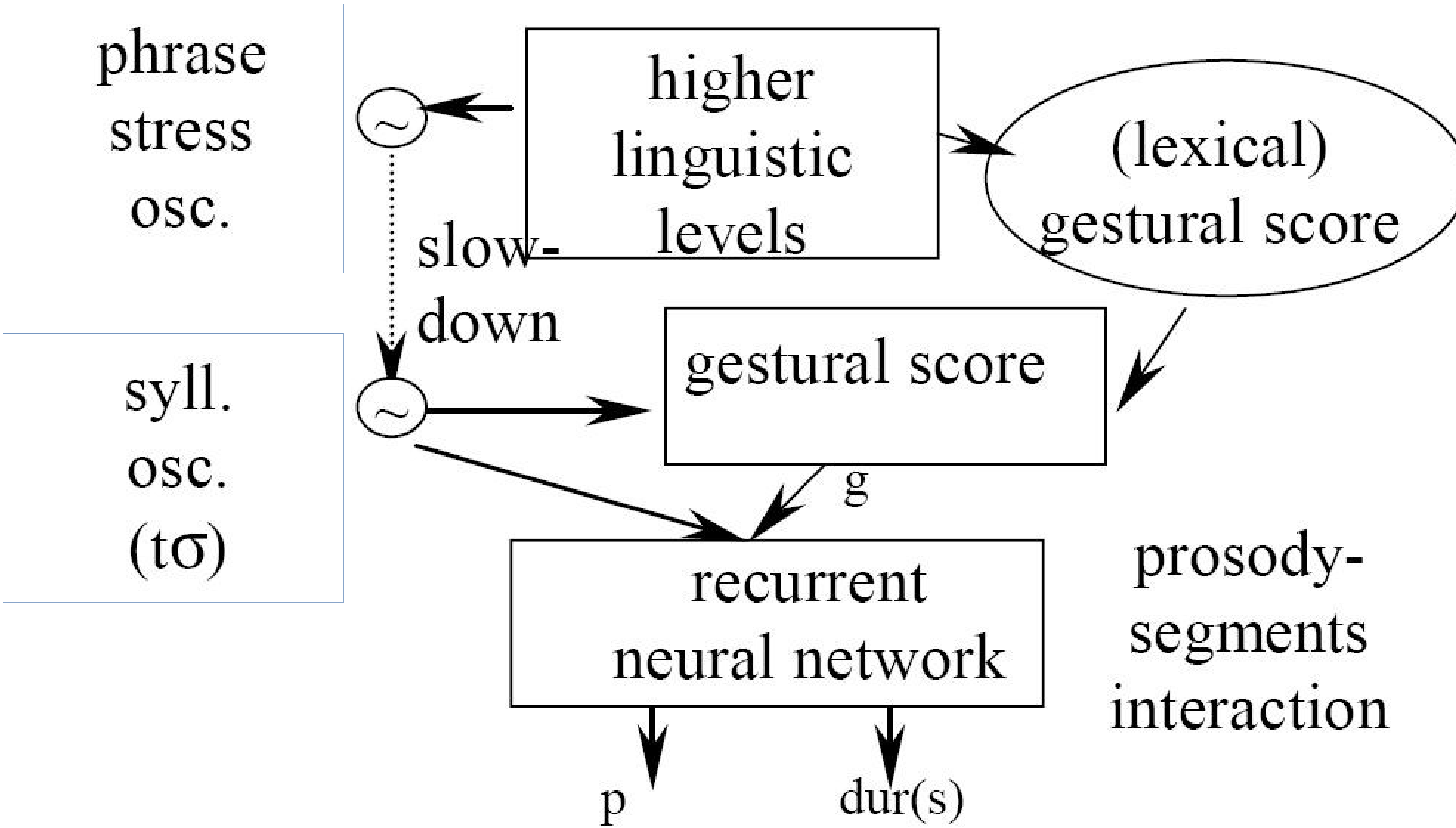
Two levels of duration encoding / control / specification, coupling between 2 oscillators:

- syllabic: intrinsic lexical level
- phrasal: extrinsic, properly rhythmic level
- entrainment (coupling) of the oscillators

Emulation of results of other rhythm studies:

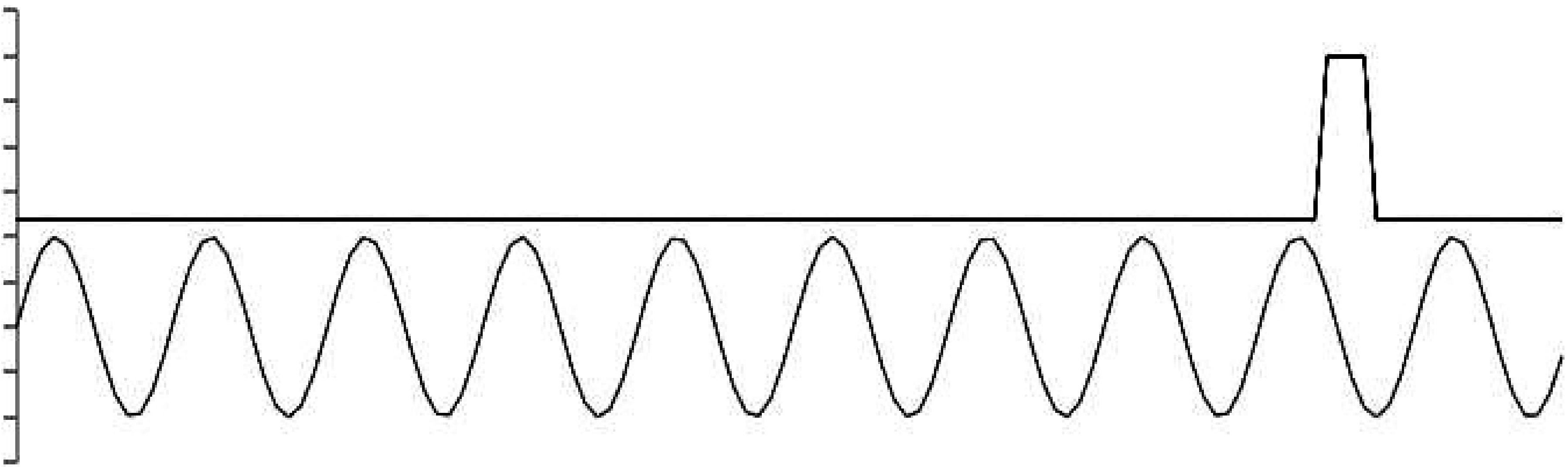
- the greater wo, the more like stress-timing
- the smaller wo, the more like syllable-timing

Barbosa's dynamic rhythm model



Barbosa's dynamic rhythm model

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.

HIERARCHICAL MODELS

Phonological models

- There are numerous phonological models of accentuation
- The most well-known ones used in Speech Synthesis are
  - Generative Phonology
  - Campbell’s model
  - Wagner’s model

Phonological models

There are a number of mildly hierarchical models of prosody in general

in the impressionistic, language teaching field, for 100 years

Selkirk's *Prosodic Hierarchy*, since 1984 (and later variants)

and work in this area is in its infancy

There are numerous hierarchical phonological models of accentuation:

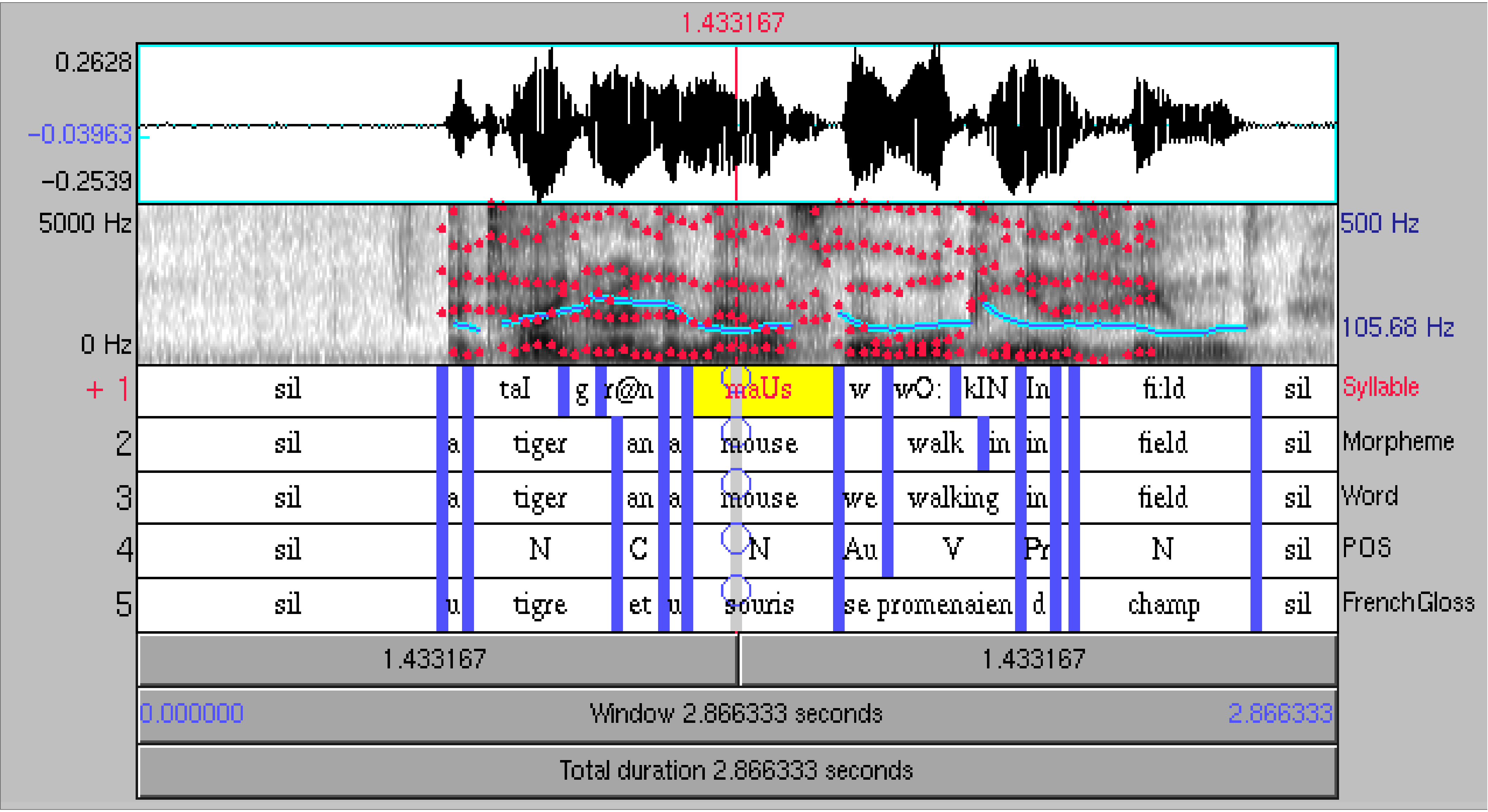
- Generative Phonology, Metrical Phonology
- Campbell's and Wagner's models for Speech Synthesis

For now I will ignore all of these, for lack of time, and move on to speech timing ...

How do we analyse timing?

ANNOTATION MINING

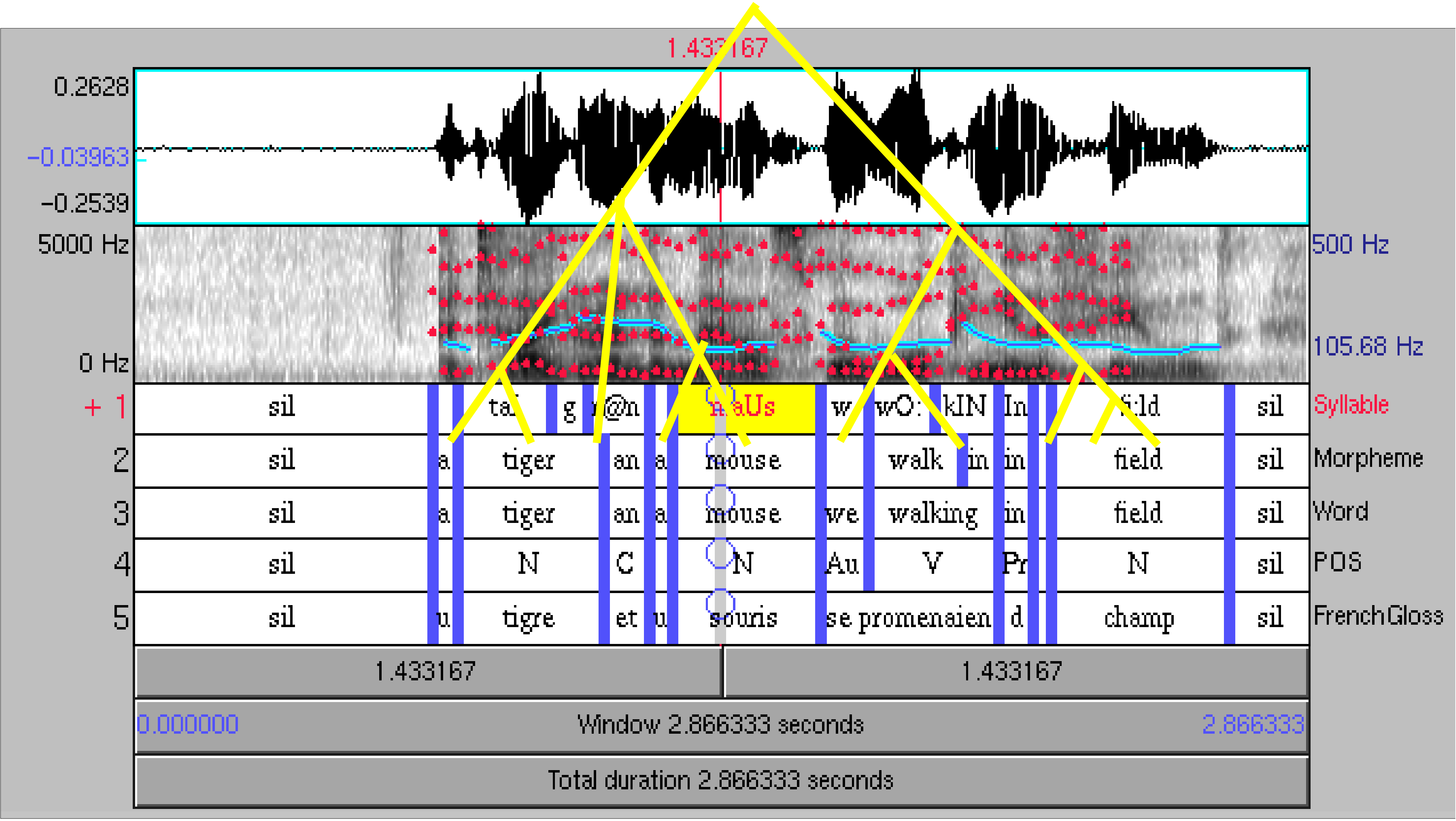
Annotation mining



Parsing

a tiger and a mouse were walking in a field

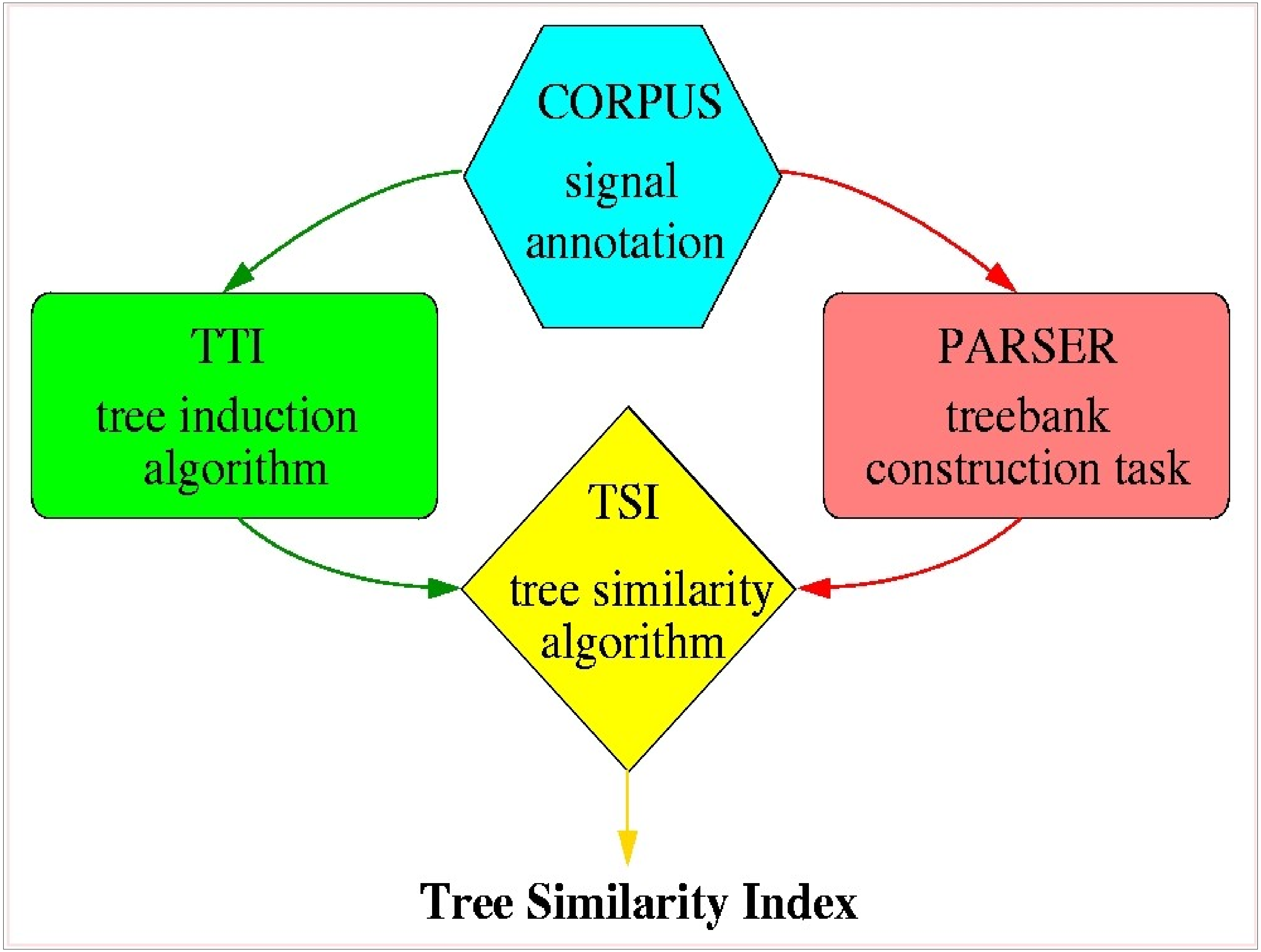
Annotation and parsing



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.."

Comparing parsing & annotation



General strategy

General strategy:

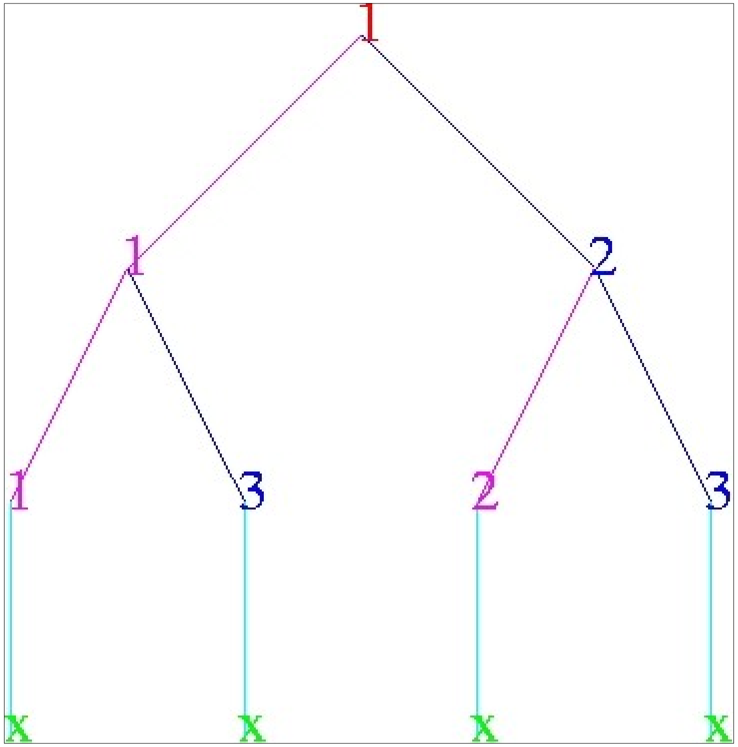
- take the local distance measure from the PVI
- do not throw directionality away by taking absolute values
- but use directionality (polarity) to determine grouping

Specific procedure:

- using annotation time-stamps, recursively build tree structures (Time Trees):
  - iambic parametrisation:
    - if right neighbour is stronger,
      - then group
      - else stack and wait for a stronger right neighbour
  - trochaic parametrisation:
    - if right neighbour is stronger,
      - then group
      - else stack and wait for a weaker right neigbour

Tree induction: algorithm sketch

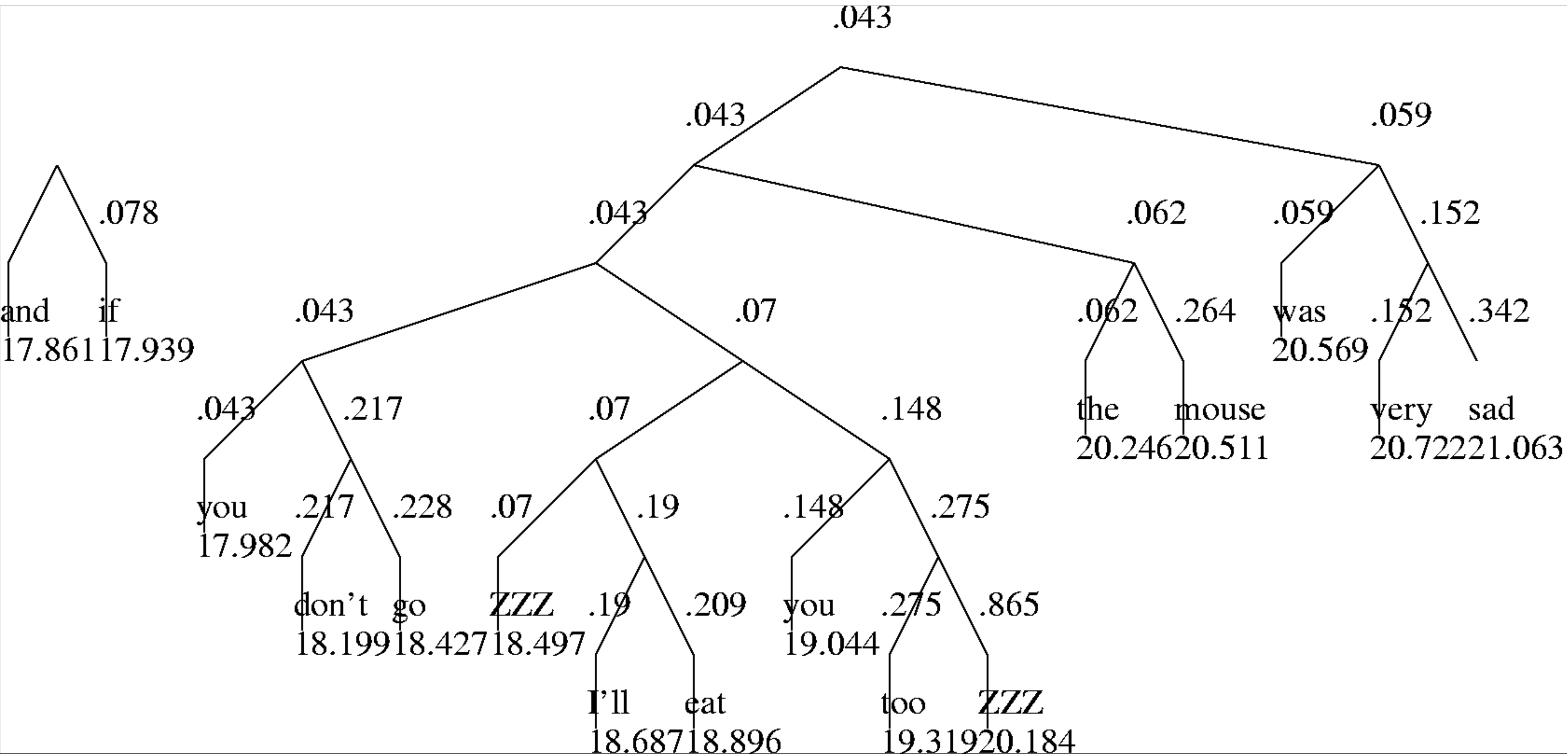
while items still left and stack not empty
- while next weaker[1] than current
  - push current on stack
  - make next current
- push current on stack
- while top stack weaker[1] than second stack
  - pop top and second from stack
  - adjoin top and second into a new node
  - push new node on stack



[1] Depending on parametrisation, comparison is weaker (A) or stronger (B)

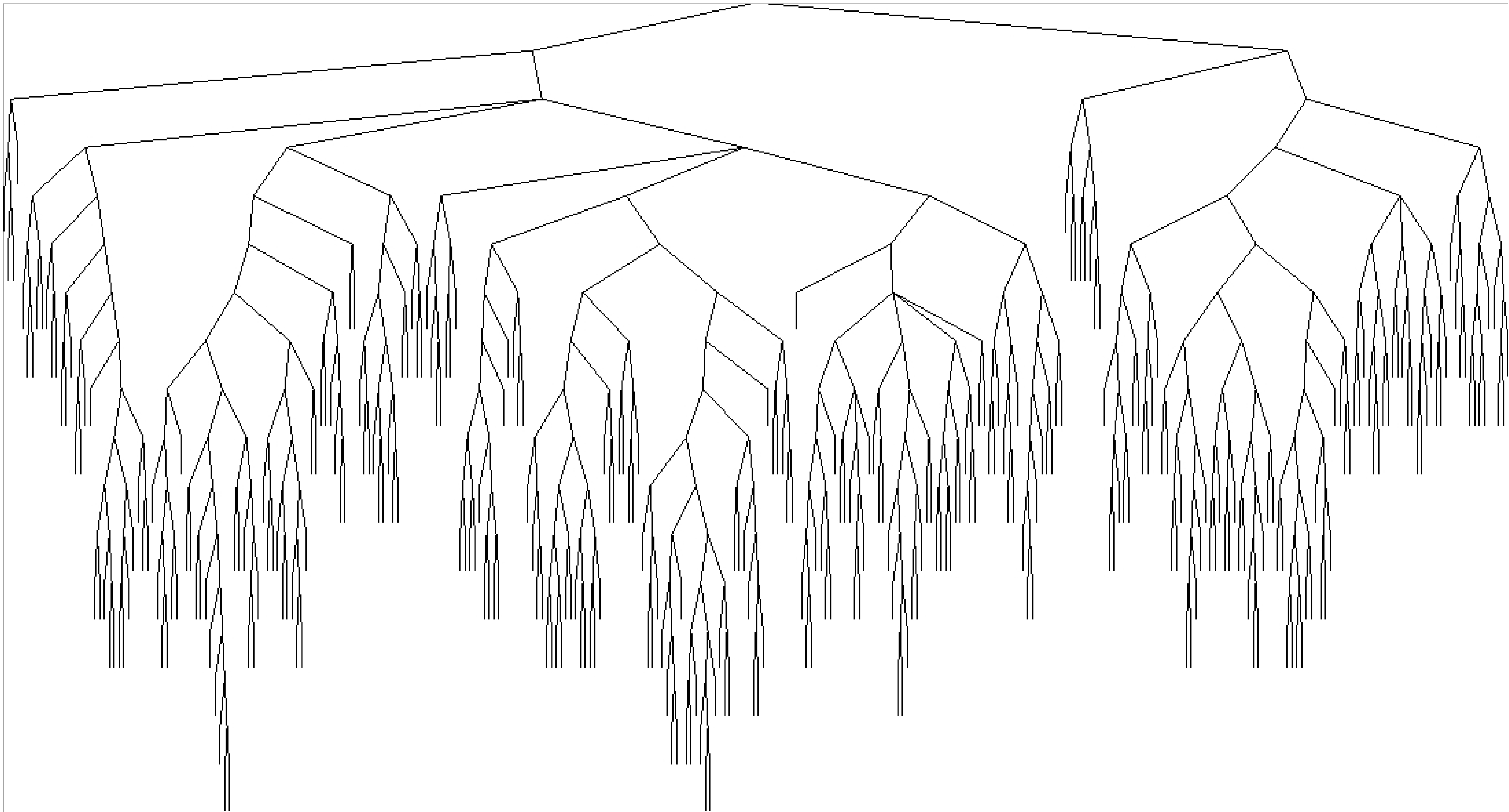
Tree induction: zoom in

Part of a narrative:



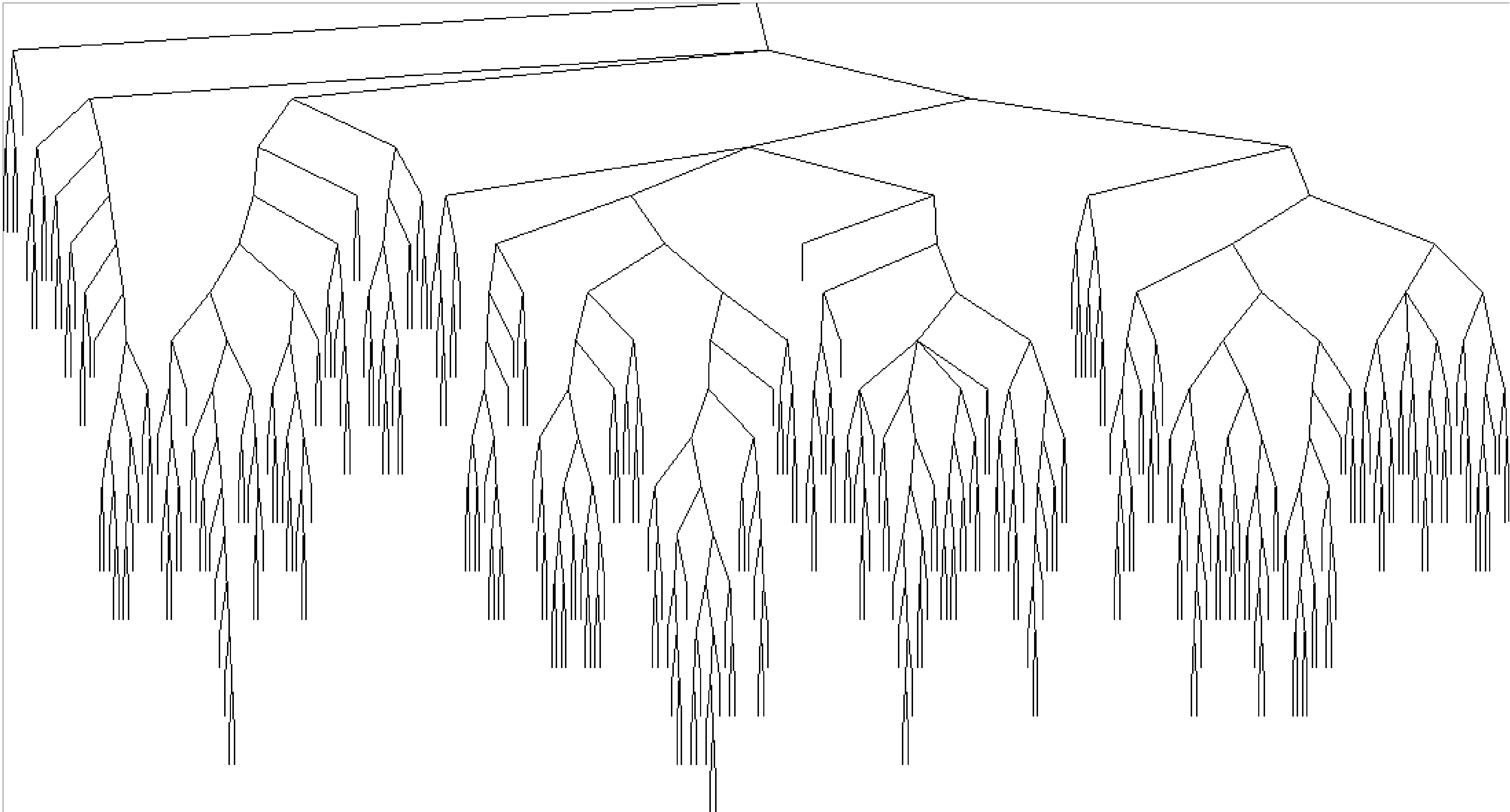
Tree induction: zoom out

A complete narrative - parametrisation A:



Tree induction: zoom out

A complete narrative - parametrisation B:



Syntax: “subjective parsing”

Six linguistically trained subjects were asked to

- bracket separate sentences (tree-equivalent notation)
- without category labels
- to show grammatical grouping
- ill-formed bracketings completed at beginning or end

Example:

- English: (((a tiger) and (a mouse)) ((were walking) (in (a field))))

Rhythm tree ≈ syntax tree?

Treat each sentence in text separately.

Uniquely label terminals (leaves) in string shared by trees.

For each tree in the tree pair

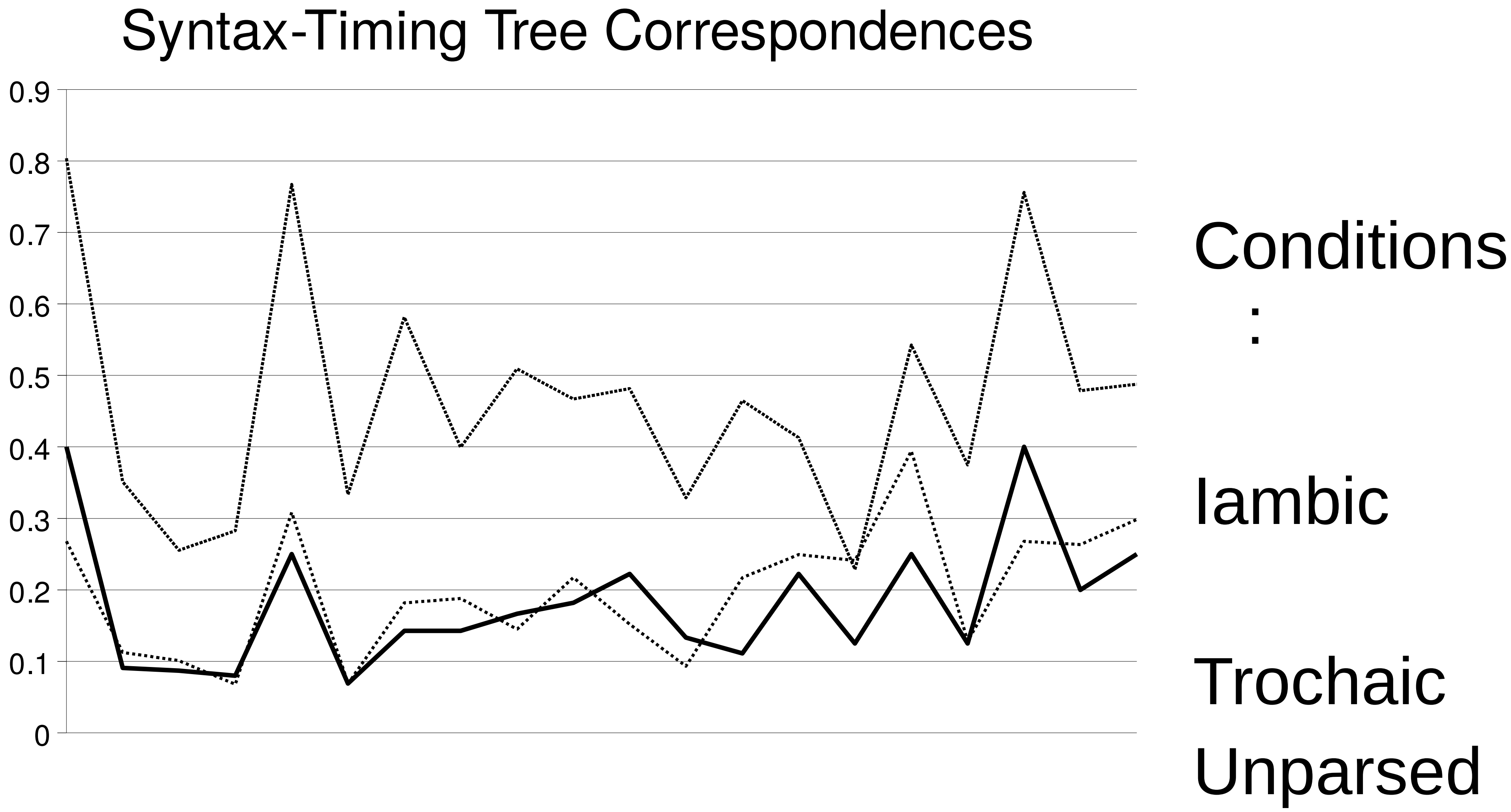
for each node in the tree

collect the substring covered by nodes into a set.

Divide the cardinality of the intersection of the sets by the some function of the cardinality of the two sets (min, mean, max, ...).

This yields the TSI (Tree Similarity Index).

Tree Similarity Index for narrative



Results

Condition	*mean UP*	*mean TSI*
Parsed + TTI-A:	0.85	0.47
Parsed + TTI-B:	0.89	0.2
Unparsed + TTI-A:		0.19
Unparsed + TTI-B:		0.19

Discussion

The thick solid line shows correspondence between timing trees and unparsed (UP) sentences.

For parsed (P) sentences, the higher thin line shows mean TSI for TTI-A short-long (iambic) grouped trees, the lower thin line shows mean TSI for TTI-B long-short (trochaic) grouped trees.

Both TTI-A (0.85) and TTI-B (0.89) TSI sequences correlate highly with the UP sequence, maybe due to shallow bracketing and short sentences. TSI levels differ considerably.

The mean TSI for TTI-A trees (iambic) is much higher than for TTI-B trees (trochaic) or UP strings, which are indistinguishable.

Syntax trees are thus more similar to iambic timing trees than to trochaic timing trees.

Preference for iambic

The results show a preference for

a match between grammatical structures and iambic groups,

with short-long constituent pairs,

indicating that the measure provides substantive and relevant information related to patterns

(such as the iambic Nuclear Stress Rule)

which figure in traditional descriptions of the intonation of West Germanic languages.

Possible extensions ...

Extensions to...

other genres and languages,

other levels, layers, weights,

deeper, non-binary bracketing,

normalisation for length effects,

size of subject set,

use of treebanks,

full statistical treatment, ...

Applications to European and African languages.

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:

- Direct ‘bottom-up’ phonetic analysis of timing
- Timing domains in prosodic typology ofe.g. mora, syllable, foot timing (depending on annotation)
- Studies in musicology

    e.g. annotated music performances

- Software development for
    - measuring foreign language phonetic proficiency
    - diagnosis and therapy in speech pathology
    - benchmarking
        - duration models in natural speech synthesis
        - designing disambiguation models in speech recognition