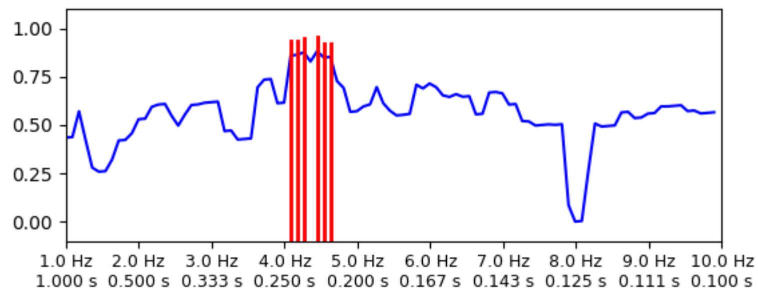
Quantifying and Correlating Rhythm Formants in Speech



Dafydd Gibbon

Andrea Lee

Bielefeld University, Germany
Jinan University, Guangzhou, China

Guangdong University of Finance,
Guangzhou, China

Overview

Part One: Problem and Proposal

Part Two: Frameworks for describing Speech Rhythm

Part Three: A Generalised Theory of Formants

Part Four: Rhythm Formants in Public Discourse

Summary, Conclusion and Outlook

Part One: Problem and Proposal

The Rhythm Challenge

1) Rhythms are directly observable events
2) Definition:
   1) Alternating pattern
   2) specific duration
   3) repeated (typically > 3 times)
3) Corollaries – can be described as:
   1) Iteration model (cf. finite state models)
   2) Alternating hierarchy (cf. generative and metrical models)
   3) Equal durations (cf. isochrony metrics)
   4) Oscillation (cf. coupled oscillator and entrainment approaches)
4) Issues with current approaches:
   1) Phonetics: isochrony, no oscillation, no general theory, annotation needed
   2) Linguistics: general theory, but controversy about physical correlates
   3) Acoustics: mainly clinical diagnosis and language identification
   4) All approaches: no account of slower discourse rhythms

The Rhythm Challenge

1) Rhythms are directly observable events
2) Definition:
   1) Alternating pattern
   2) specific duration
   3) repeated (typ
3) Corollaries –
   1) Iteration moc
   2) Alternating hi
   3) Equal duratio
   4) Oscillation (c ches)

So here is the challenge:

- **account for rhythm as oscillation**
- **account for slower discourse rhythms**
- **account for rhythm variation**
- **embed in a general theory**
- **implement automatic rhythm analysis**

4) Issues with current approaches:
   1) Phonetics: isochrony, no oscillation, no general theory, annotation needed
   2) Linguistics: general theory, but controversy about physical correlates
   3) Acoustics: mainly clinical diagnosis and language identification
   4) All approaches: no account of slower discourse rhythms

A Proposal: Rhythm Formant Theory, Rhythm Formant Analysis

A theory of rhythm which

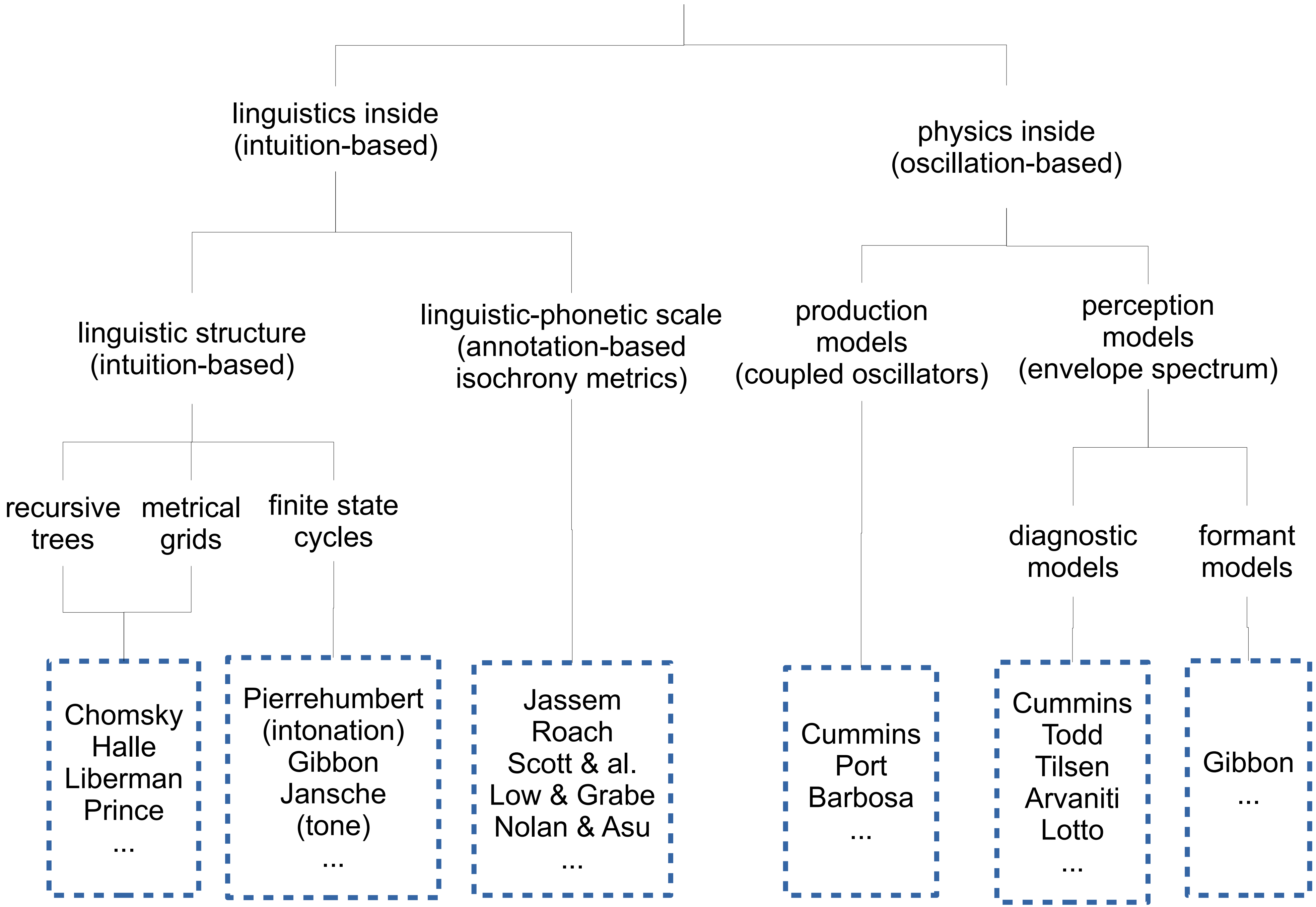
- is language-independent
- takes rhythm as oscillation into account
  - and therefore *a fortiori* isochrony
- relates to a range of low frequency rhythms:
  - syllable rhythms, 3...12 Hz
  - slower word/foot rhythms, 1...3 Hz
  - slower phrase rhythms, 0.5...1 Hz
  - slower discourse rhythms, < 0.2 Hz
- has a straightforward implementation

Part Two: Frameworks for describing speech rhythm

1) Typology of frameworks

2) A specific case: selected isochrony metrics

Typology of Rhythm Description Frameworks



A popular Isochrony Metric: Pairwise Variability Index

For a vector D = (d_1, …, d_n) of annotated durations:

$$rPVI(D) = \left(\sum_{k=1}^{n-1} |d_k - d_{k+1}|\right) / (n-1)$$

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

A popular Isochrony Metric: Pairwise Variability Index

Strangely, the formal and empirical foundations of the PVI are not questioned by its practitioners. So let’s take a quick look...

For a vector D = (d_1, …, d_n) of annotated durations:

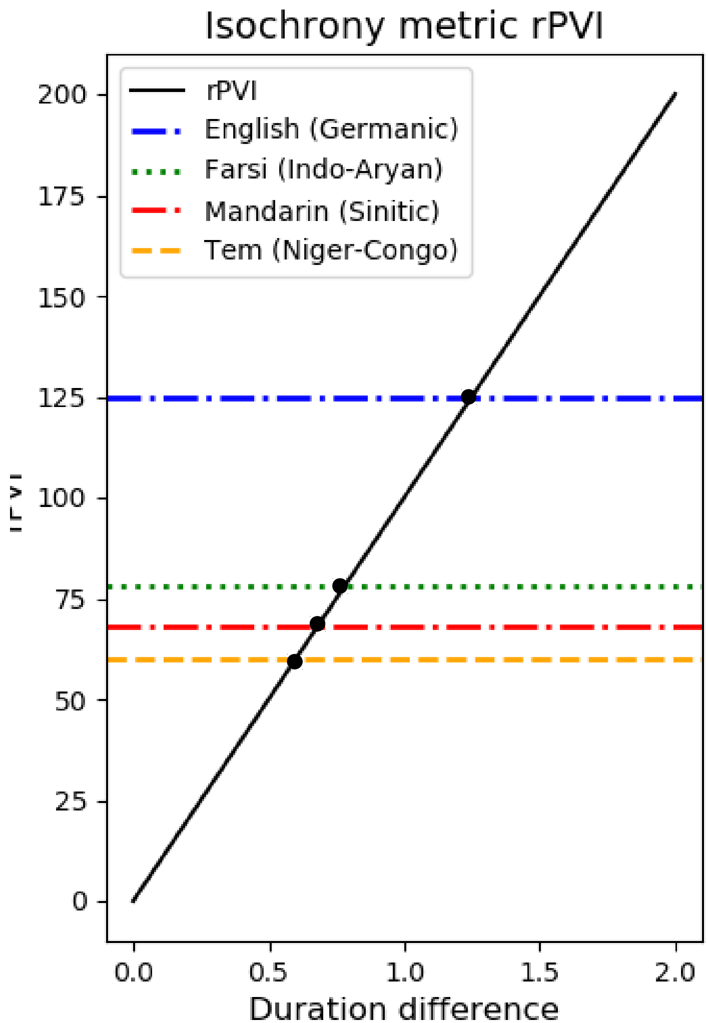
$$rPVI(D) = \sum_{k=1}^{n-1} |d_k - d_{k+1}| / (n-1)$$

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

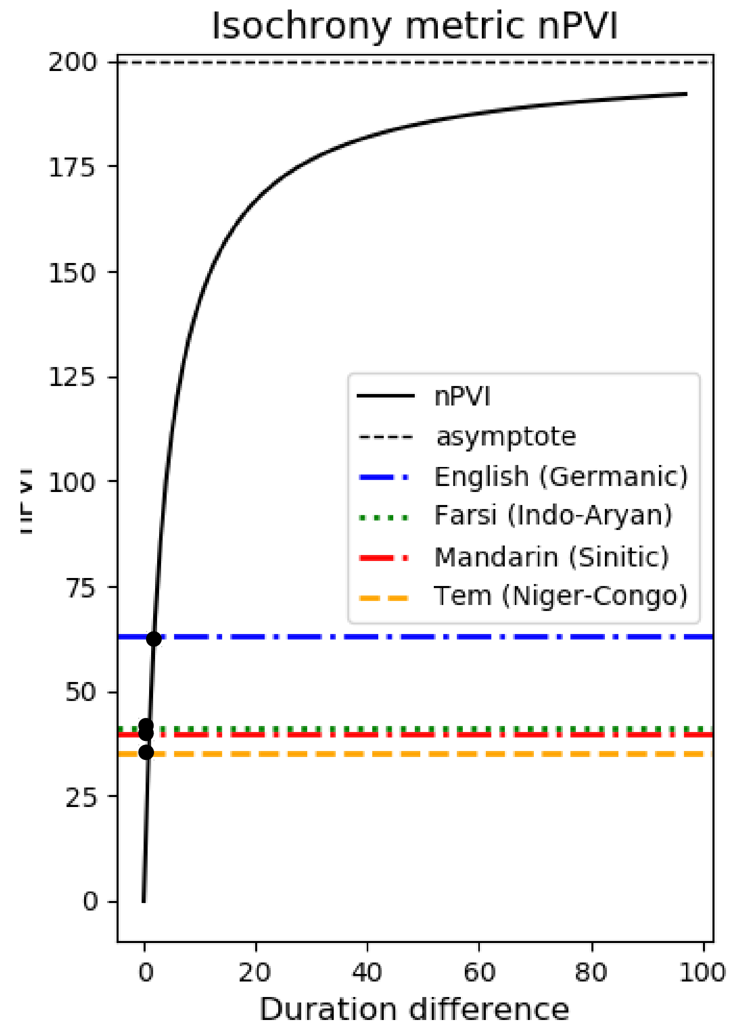
Modifications of standard distance measures:
- Manhattan Distance (*rPVI*)
- Canberra Distance (*nPVI*)

A popular Isochrony Metric: Pairwise Variability Index

rPVI: linear
scale, syllables



nPVI: non-linear
scale, syllables



A popular Isochrony Metric: Pairwise Variability Index

absolute value: ambiguous index, same for alternating and non-alternating sequences
Therefore:

NOT A RHYTHM METRIC ☺

subtraction restricts the metric to a binary relation

For a vector $D = (d_1, \ldots, d_n)$ of annotated durations:

$$rPVI(D) = \sum_{k=1}^{n-1} |d_k - d_{k+1}| / (n-1)$$

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

Language-dependent
Filtered by the annotation procedure.

The distance measures are binary:
- Manhattan Distance (*rPVI*)
- Canberra Distance (*nPVI*)

2-dimensional isochrony models

Asu & Nolan:

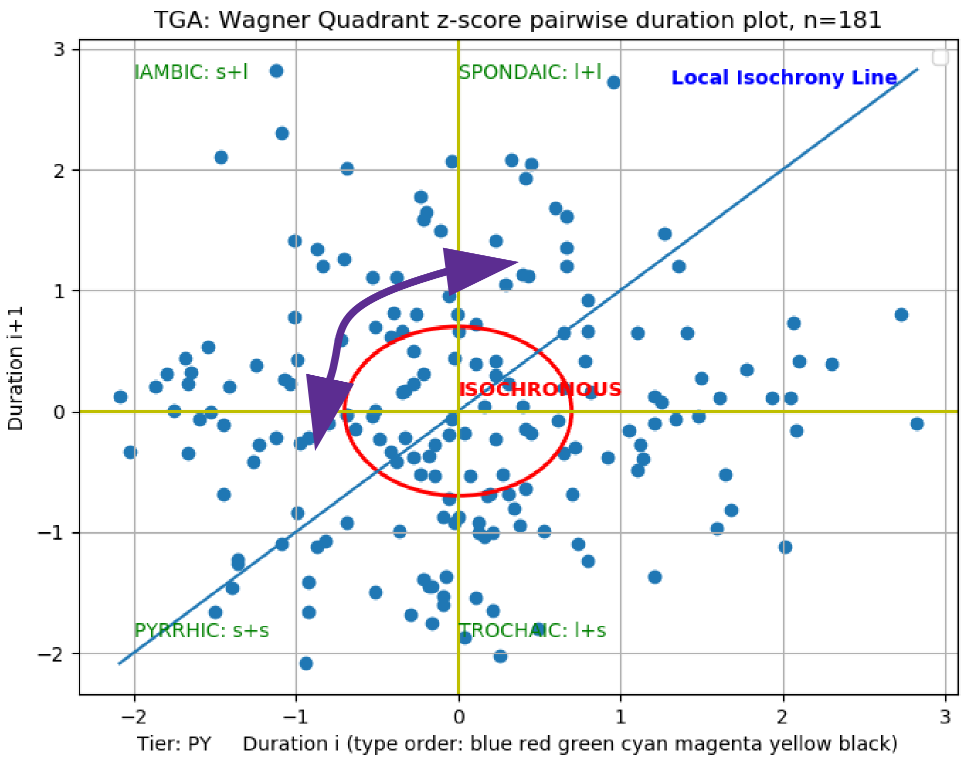
comparison of PVI for foot X syllable in Estonian X English
foot results are similar
syllable results are different

Wagner:

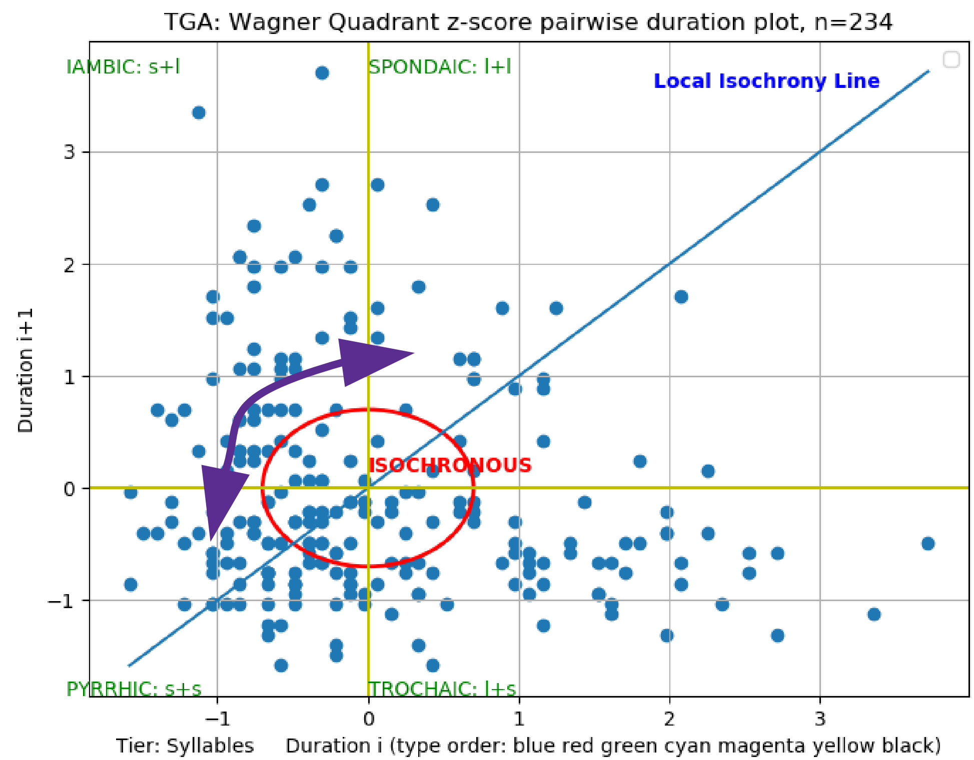
from the sequence of durations $D = (d_1, \ldots, d_n)$
plot z-scored scatter plot quadrants subsequences
$(d_1, \ldots, d_{n-1}) \quad X \quad (d_2, \ldots, d_n)$

2-dimensional isochrony models: Wagner

Mandarin
Note the even distribution around the mean.



English
Note the skewed distribution with many shorter than average syllables.



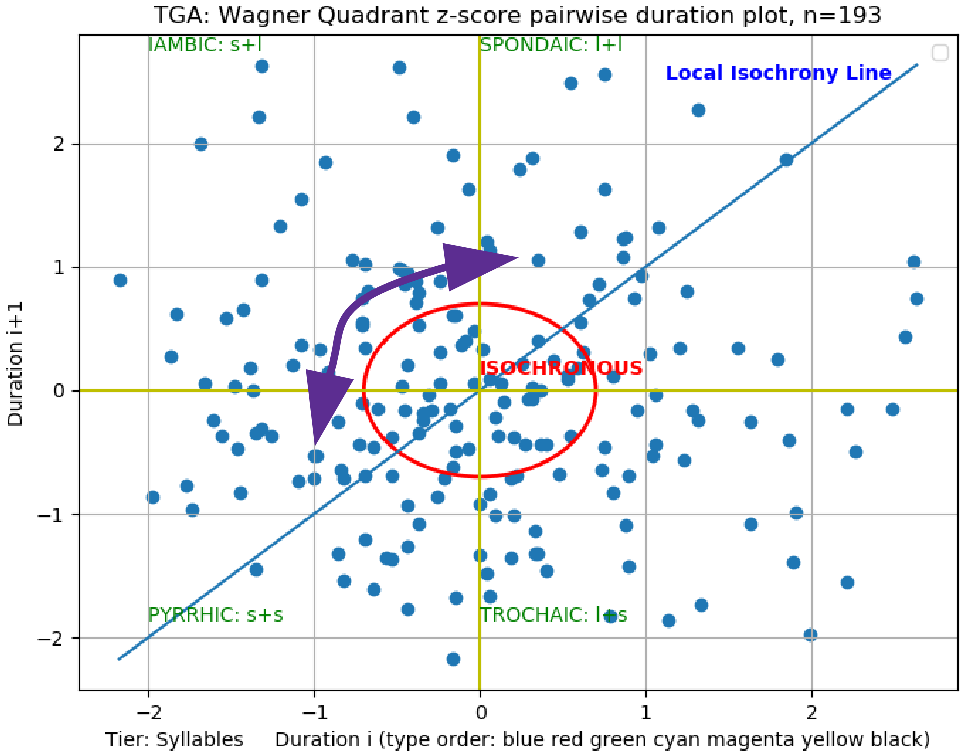
Pyrrhic (short-short) and Spondaic (long-long) counts:

Mandarin: ratio approximately 1:1
English: ratio approaches 2:1

<u>2-dimensional</u> isochrony models: Wagner

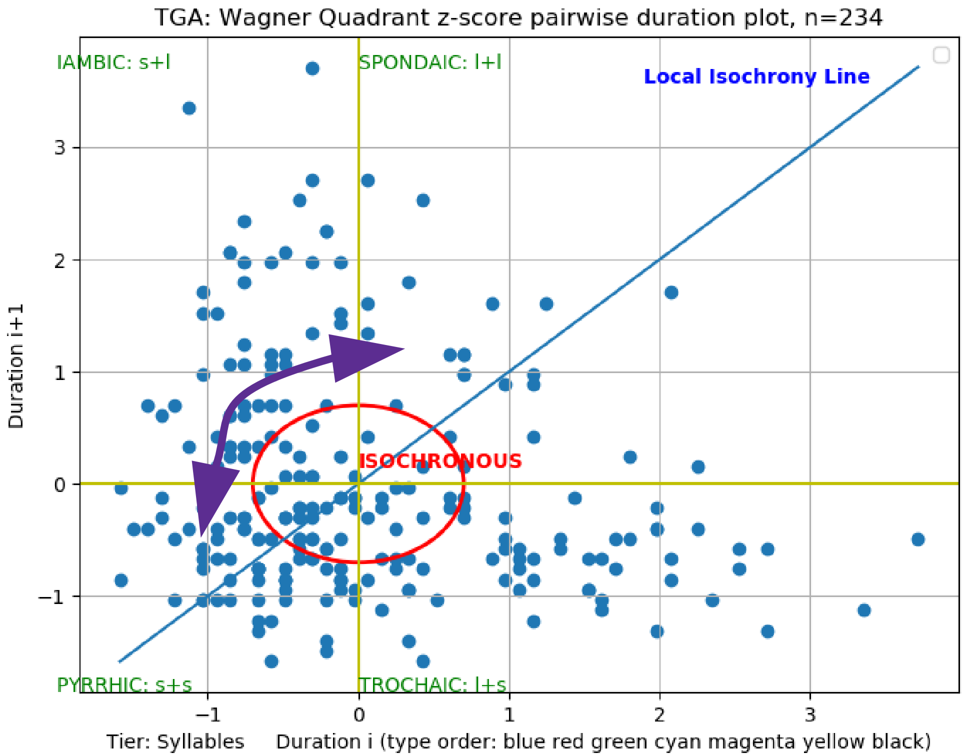
Farsi

Note the relatively even distribution around the mean.



English

Note the skewed distribution with many shorter than average syllables.



<u>Pyrrhic</u> (short-short) and <u>Spondaic</u> (long-long) counts:

Farsi: ratio approaches 1:1
English: ratio approaches 2:1

Summary of issues with isochrony metrics

Isochrony metrics are popular, but ...

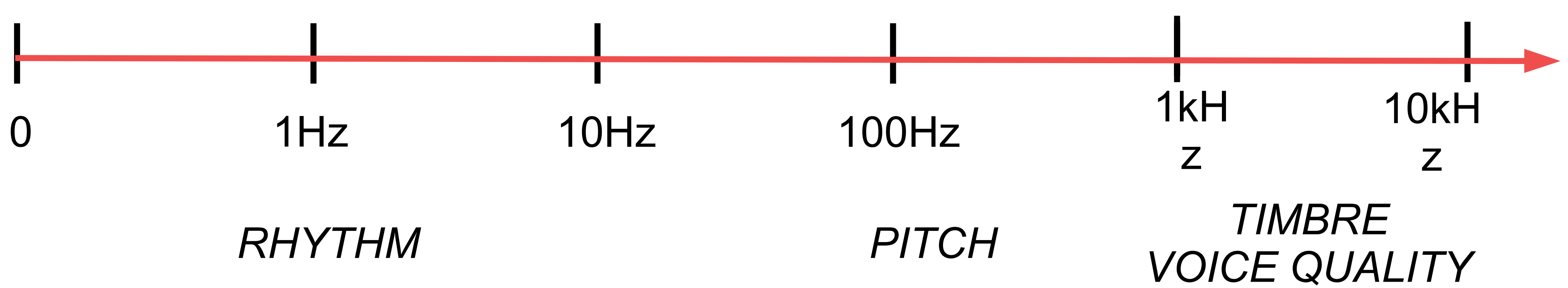
- no adequate explanation for
  - rhythm
  - rhythm variation for the same speaker / dialect / language
- too little:
  - *isochrony* but not *oscillation*
  - only binary patterns

    but rhythms can be ternary, quaternary, etc., or even unary
- too much:
  - indices can be ambiguous for alternating and non-alternating values (because absolute not actual differences)
- dependent on human annotation decisions
- one-dimensional metrics with single value
- neither a descriptive model nor a predictive theory

Part Three: From Formants to Rhythm Formants

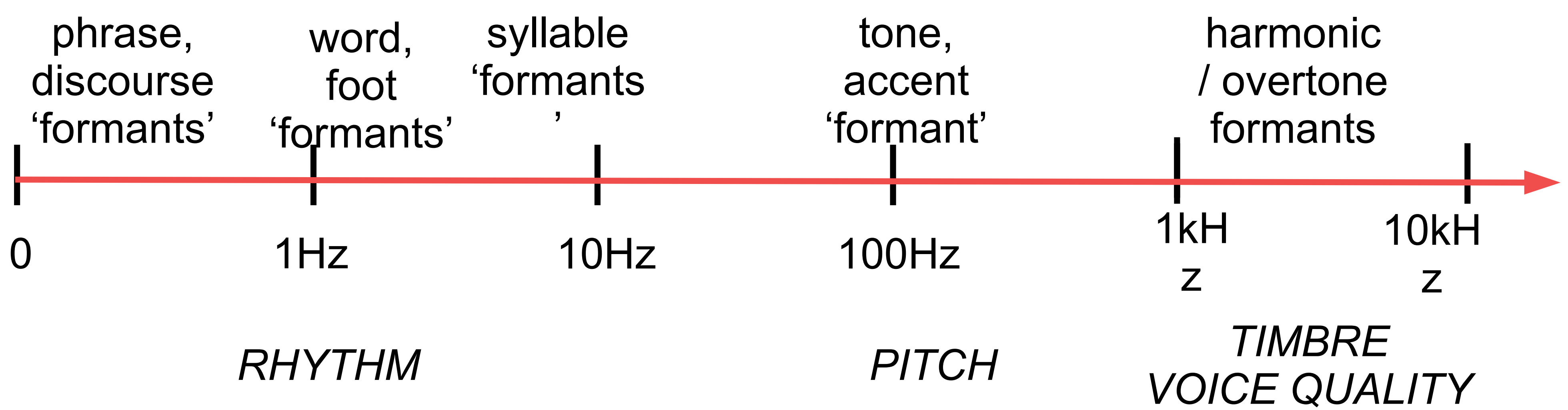
language-independent
automatic identification of speech rhythms
in syllables, words, discourse
embedded in a general formant theory

Rhythms as Oscillations – Oscillations as Rhythms
Frequency Zones and Rhythm Formants

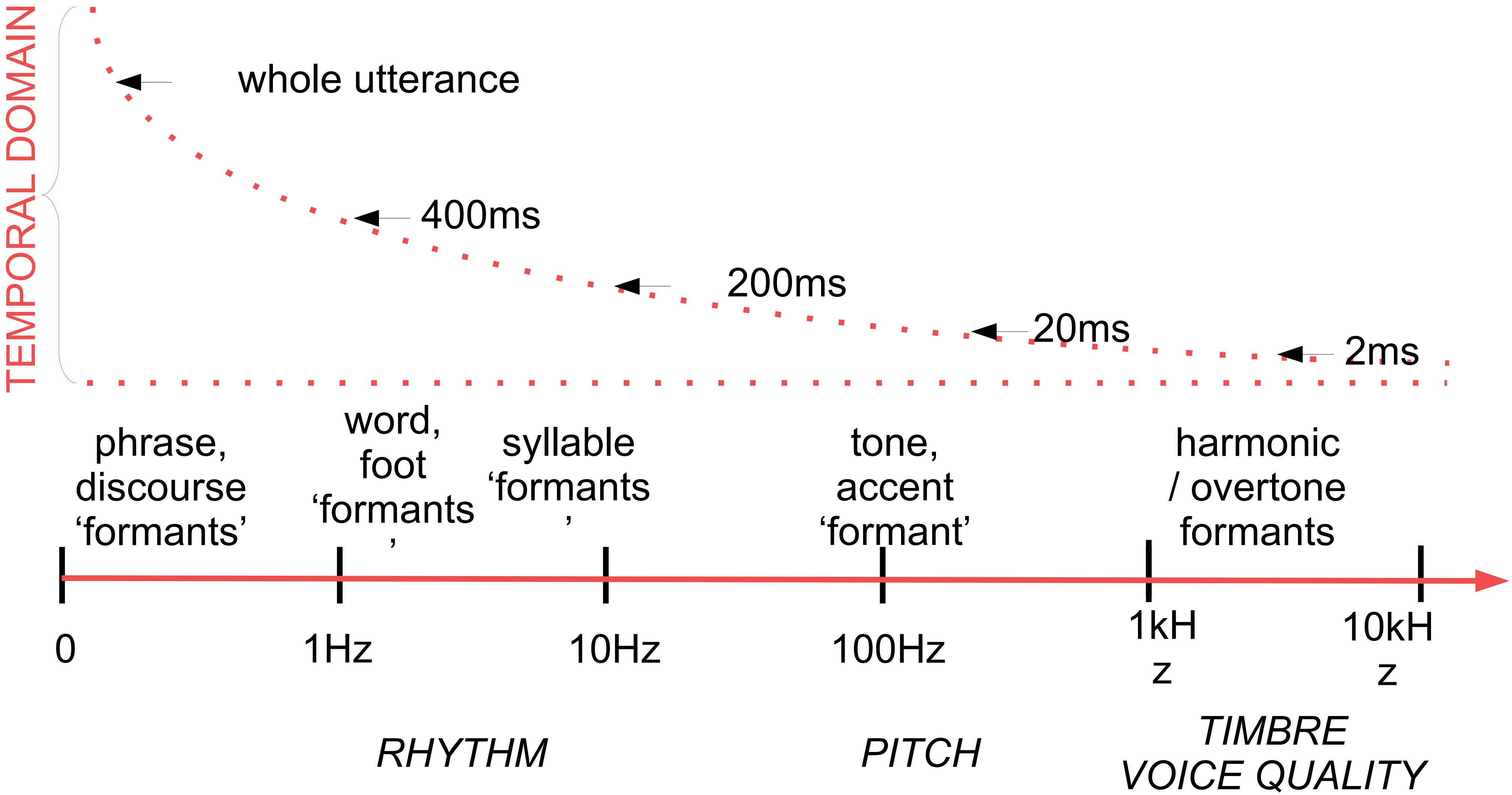


Cf. the classic of Musical Relativity Theory / Overtone Theory in musicology:
Cowell, Henry. 1930. *New Musical Resources*. New York: Alfred A. Knopf Inc.

Rhythms as Oscillations – Oscillations as Rhythms
Frequency Zones and Rhythm Formants



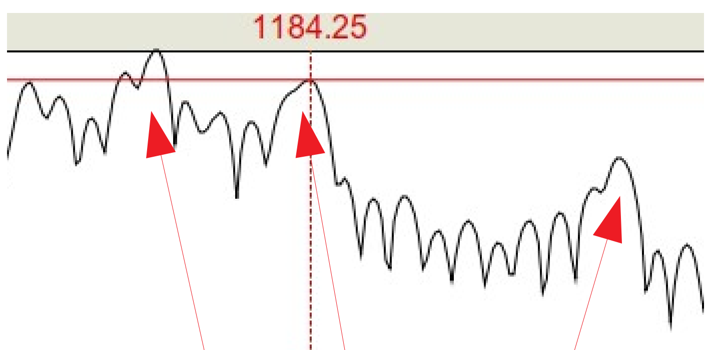
Rhythms as Oscillations – Oscillations as Rhythms
Frequency Zones and Rhythm Formants



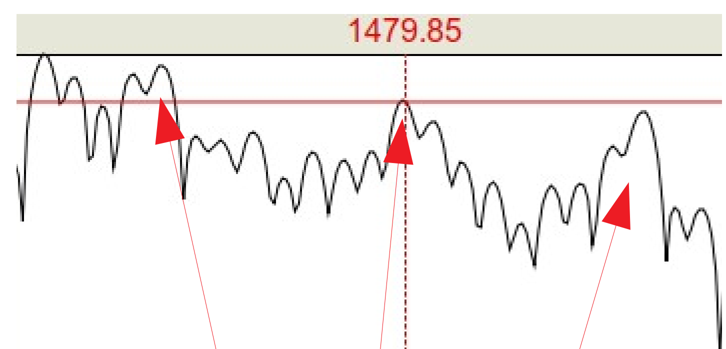
<u>High Frequency</u> Formants (HF Formants)

1. Formants are the resonant frequencies of the vocal tract.
2. Formants are distinctive frequency components of speech.

<u>HF formant structures</u>, f>600Hz signify vocal tract configurations.



[a] in “five”: 1st, 2nd, 3rd formants

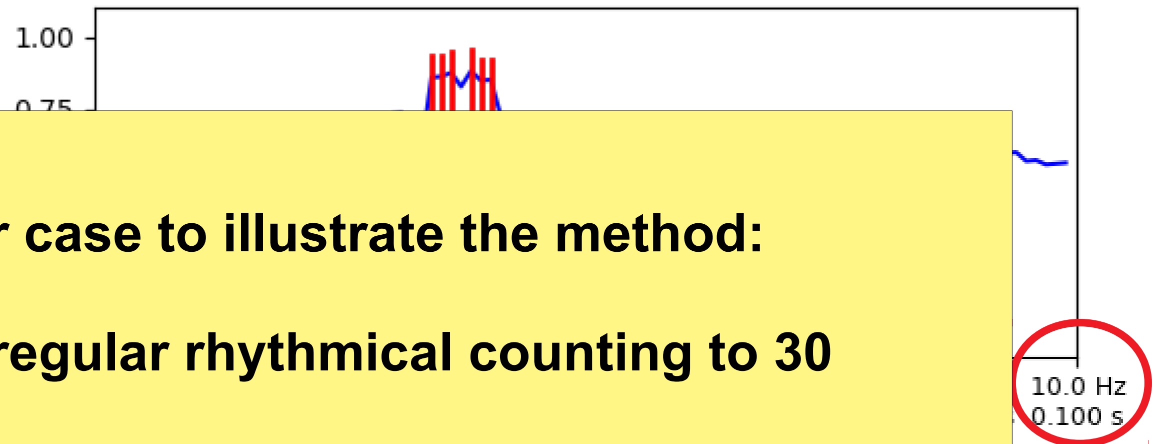


[i] in “five”: 1st, 2nd, 3rd formants

<u>Low Frequency</u> Formants (LF Formants)

1. ~~Formants are the resonant frequencies of the vocal tract.~~
2. Formants are distinctive frequency components of speech.

<u>LF formant structures</u>, f<20Hz, signify rhythms,
e.g. a 4.3Hz LF formant may signify a syllable sequence of mean duration 235ms.



LF spectr

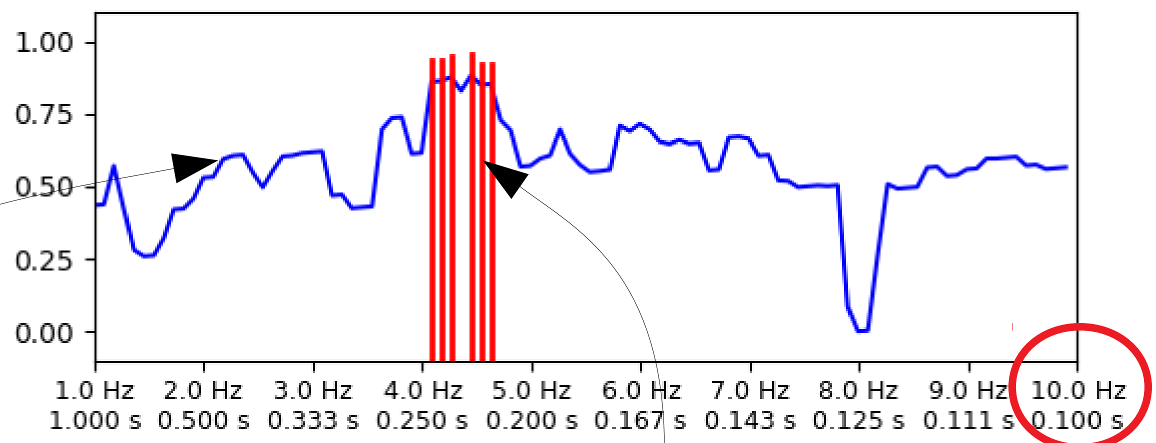
A clear case to illustrate the method:

- **fast regular rhythmical counting to 30**

<u>Low Frequency</u> Formants (LF Formants)

1. ~~Formants are the resonant frequencies of the vocal tract.~~
2. Formants are distinctive frequency components of speech.

<u>LF formant structures</u>, f<20Hz, signify rhythms
e.g. a 4.3Hz LF formant may be a syllable sequence of mean duration 235ms.

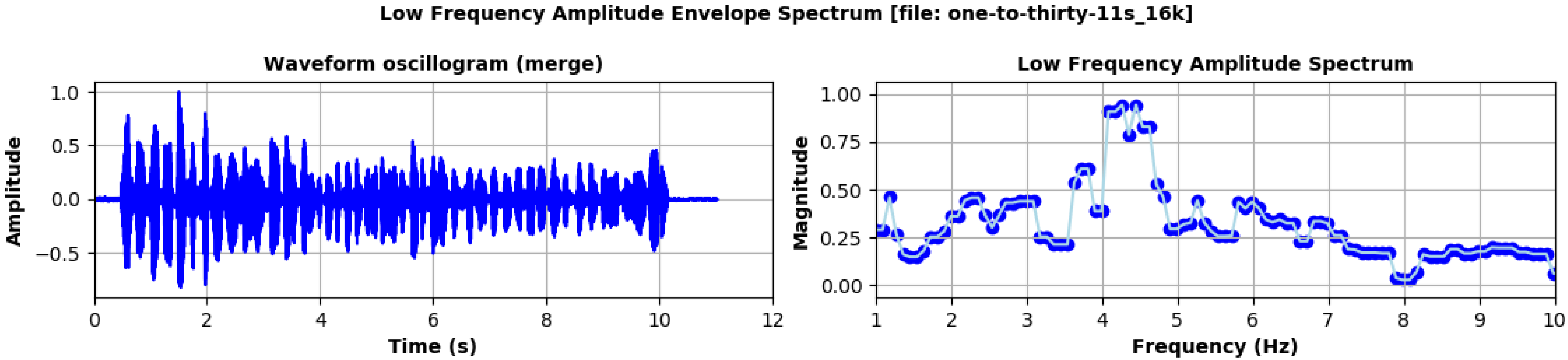


LF spectrum

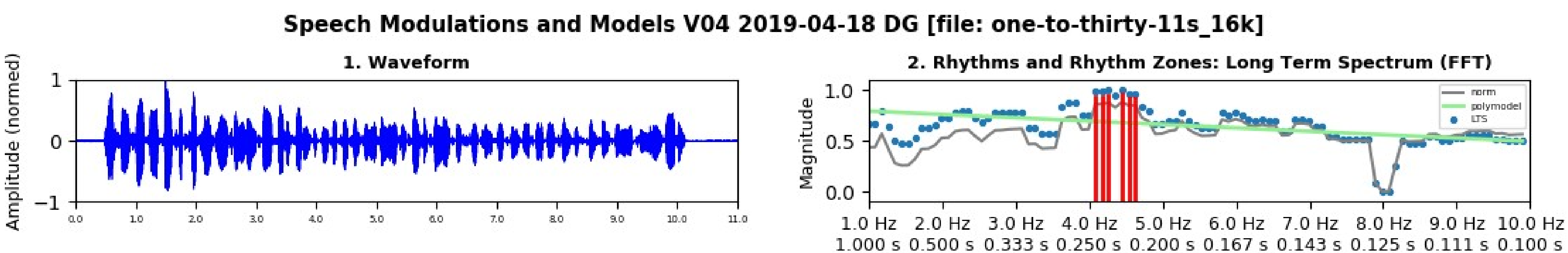
highest magnitude frequencies, ‘rhythm bars’

<u>Low Frequency</u> Formants (LF Formants)

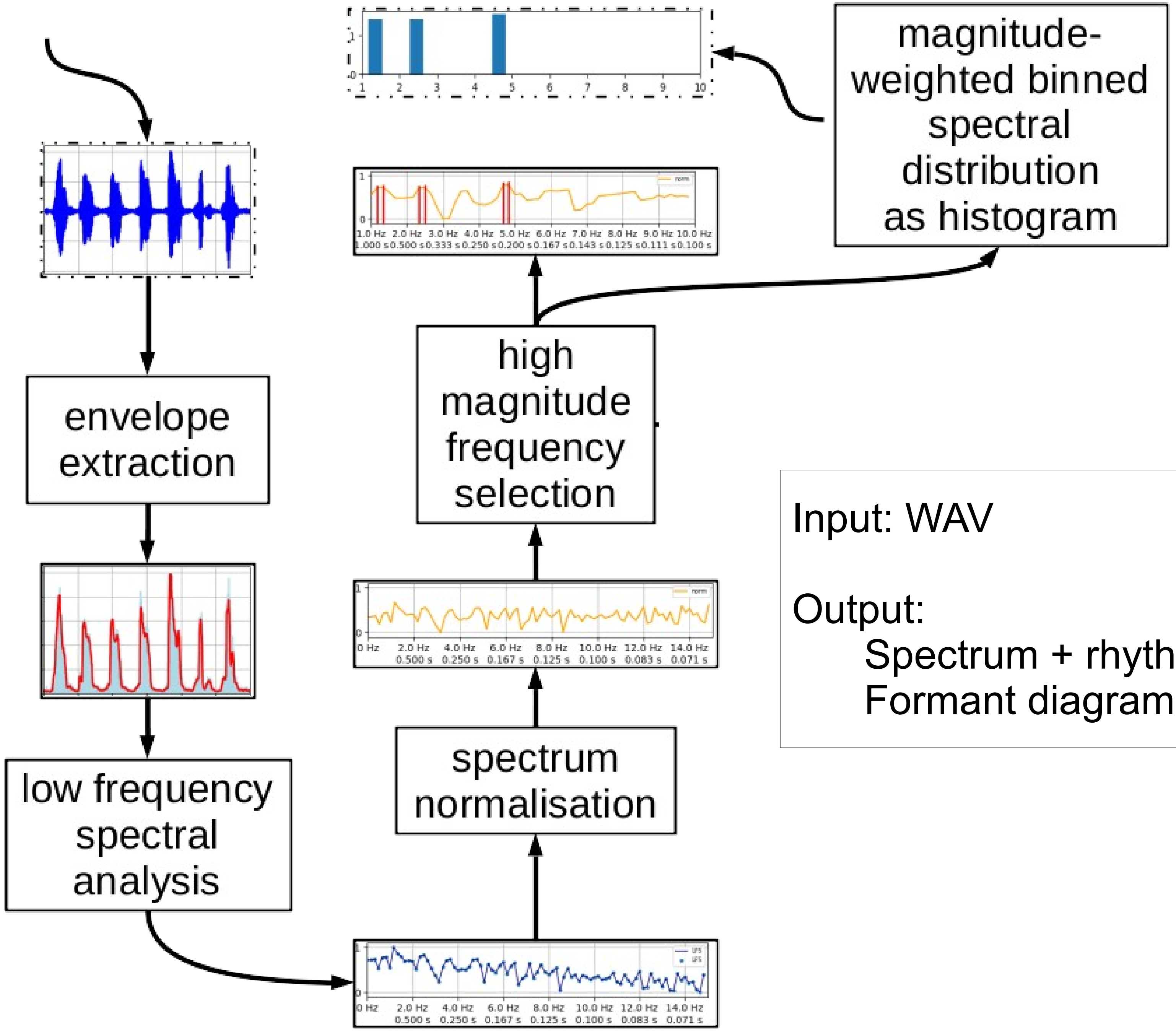
Non-normalised LF spectrum



Normalised LF spectrum with ‘rhythm bars’



Overview of Rhythm Formant Analysis Dataflow



Input: WAV

Output:
- Spectrum + rhythm bars
- Formant diagram

Part Four: Discourse Rhythms in Public Speaking

Campaign Speeches of Donald Trump (2016)
for a study of impoliteness (Li 2017)

An exploratory pilot study

Case Study on Impoliteness

- Problem:
  - Which method of analysis to use?
  - Experimental elicitation of impoliteness is problematic
  - Individual judgments of politeness are problematic
- Solution:
  - Phonetic corpus analysis
  - Opinion survey, classification of results

- Problem:
  - Where to find real impoliteness ‘in the wild’?
- Solution:
  - –

Case Study on Impoliteness

- Problem:
  - Which method of analysis to use?
  - Experimental elicitation of impoliteness is problematic
  - Individual judgments of politeness are problematic
- Solution:
  - Phonetic corpus analysis
  - Opinion survey, classification of results

- Problem:
  - Where to find real impoliteness ‘in the wild’?
- Solution:
  - Election campaign speeches by Donald Trump

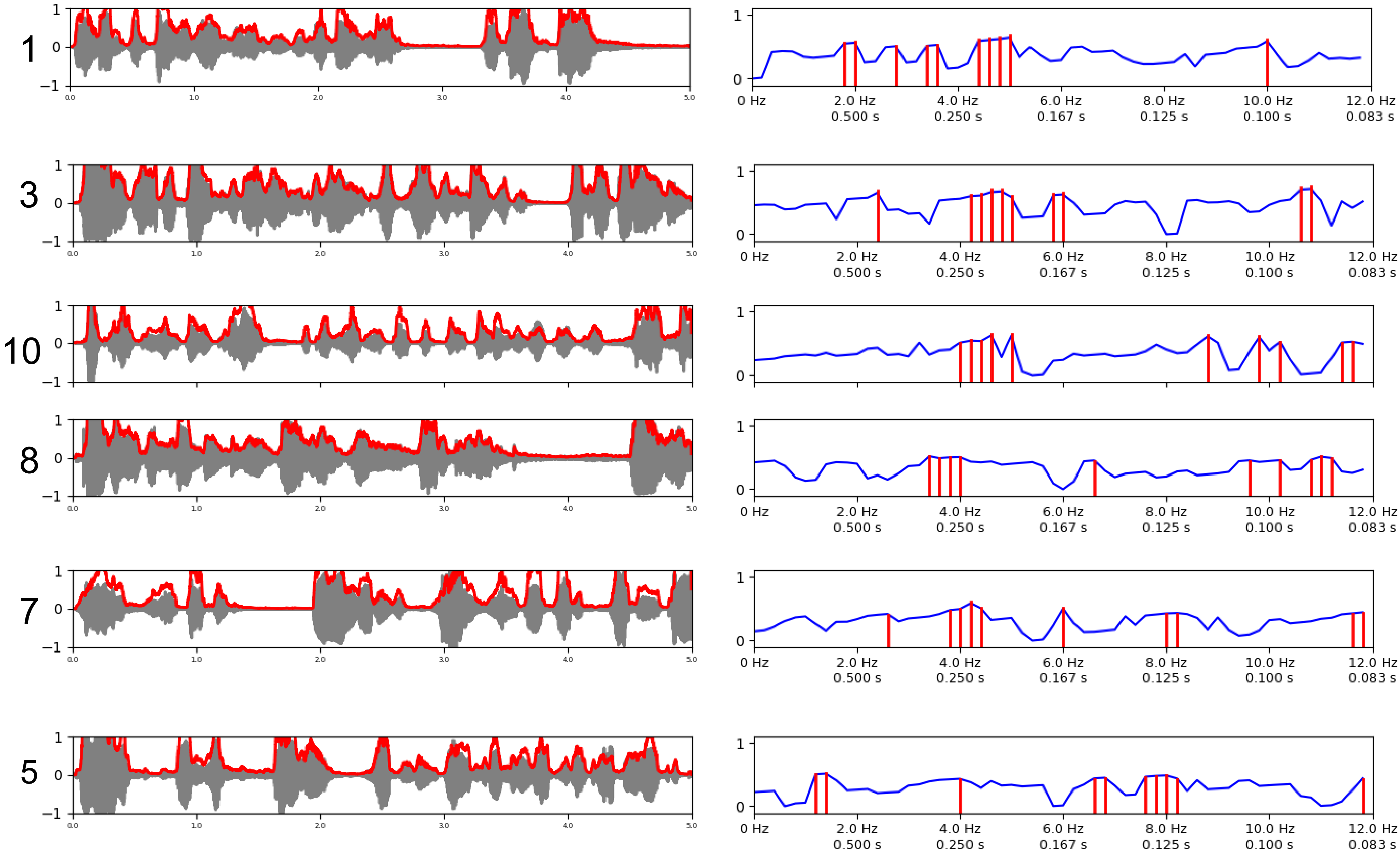
Rhythm Formant Analysis (RFA)

1. Categorise each of 10 utterances linguistically

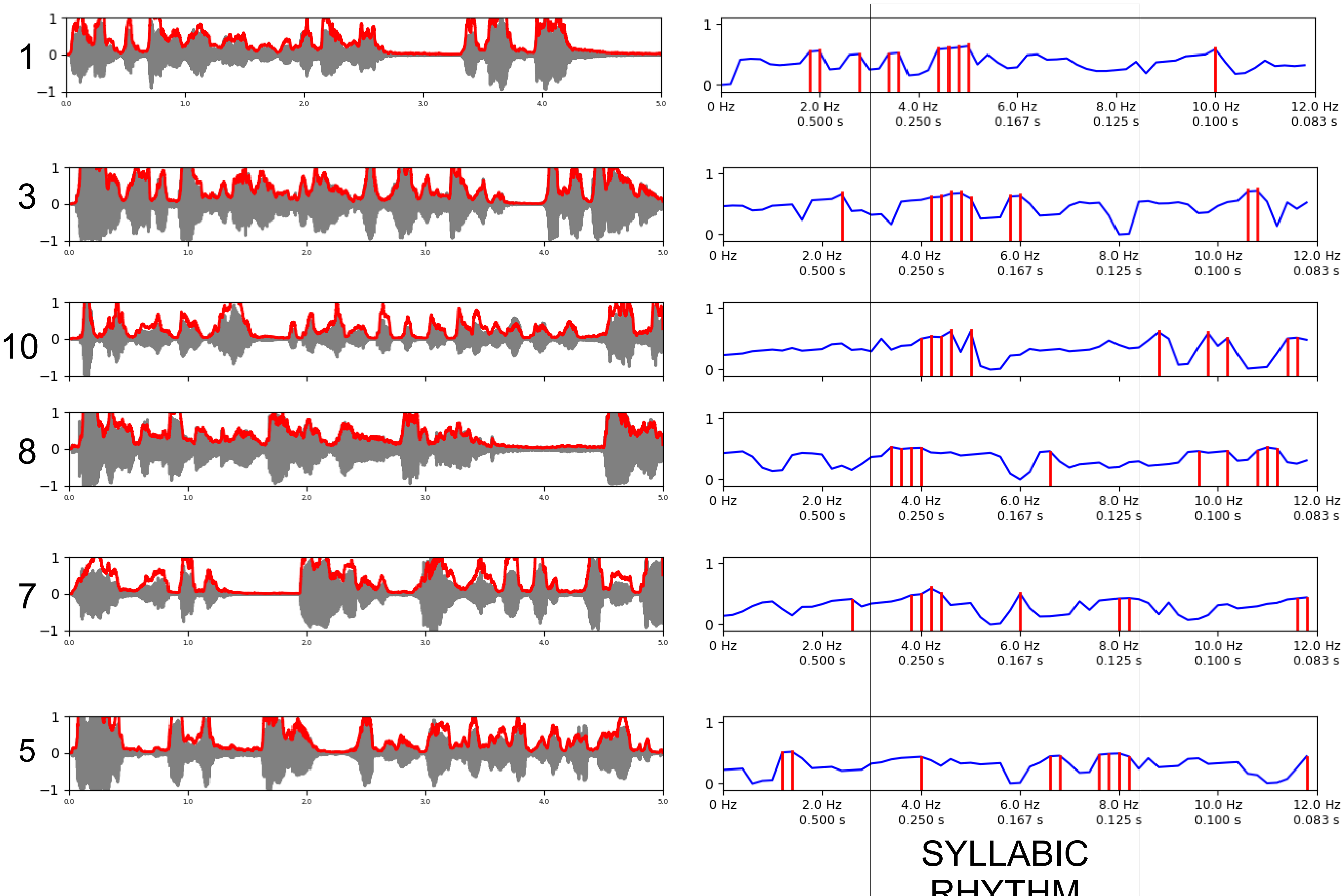
   e.g. genre categories *narrative* or *non-narrative*

2. Apply Rhythm Formant Analysis to each utterance.
3. Calculate pairwise distances (Cosine, Manhattan, ...)
   - of low frequency spectrum
   - based on the distance measures
   - display as a dendrogram
4. Generate a hierarchical classification
   - based on the distance measures
   - display as a dendrogram
5. Assign linguistic categories to dendrogram end nodes
6. Agreement → reasonable agreement

Narrative style: <u>regular rhythmical syllabic timing</u>

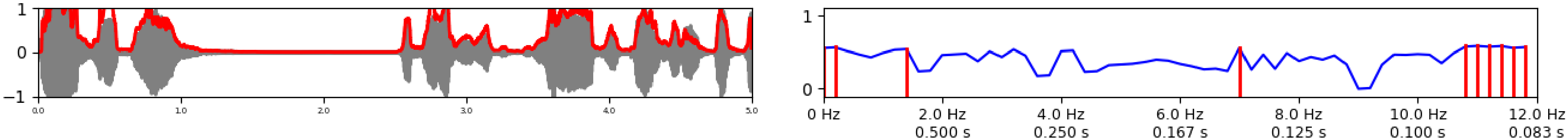


Narrative style: regular rhythmical syllabic timing

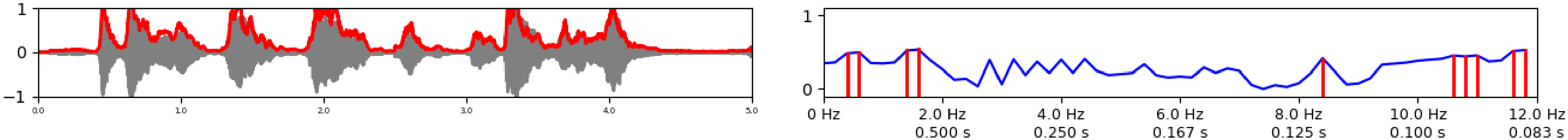


Face-threatening style: short syllables, regular pauses

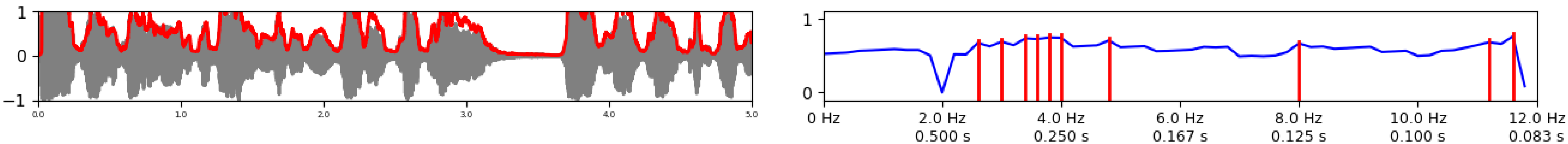
2



9

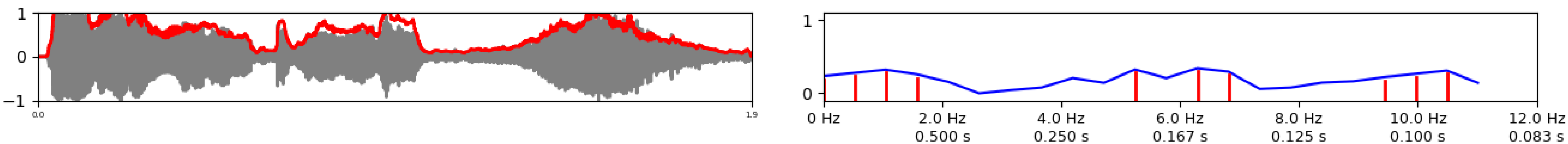


6

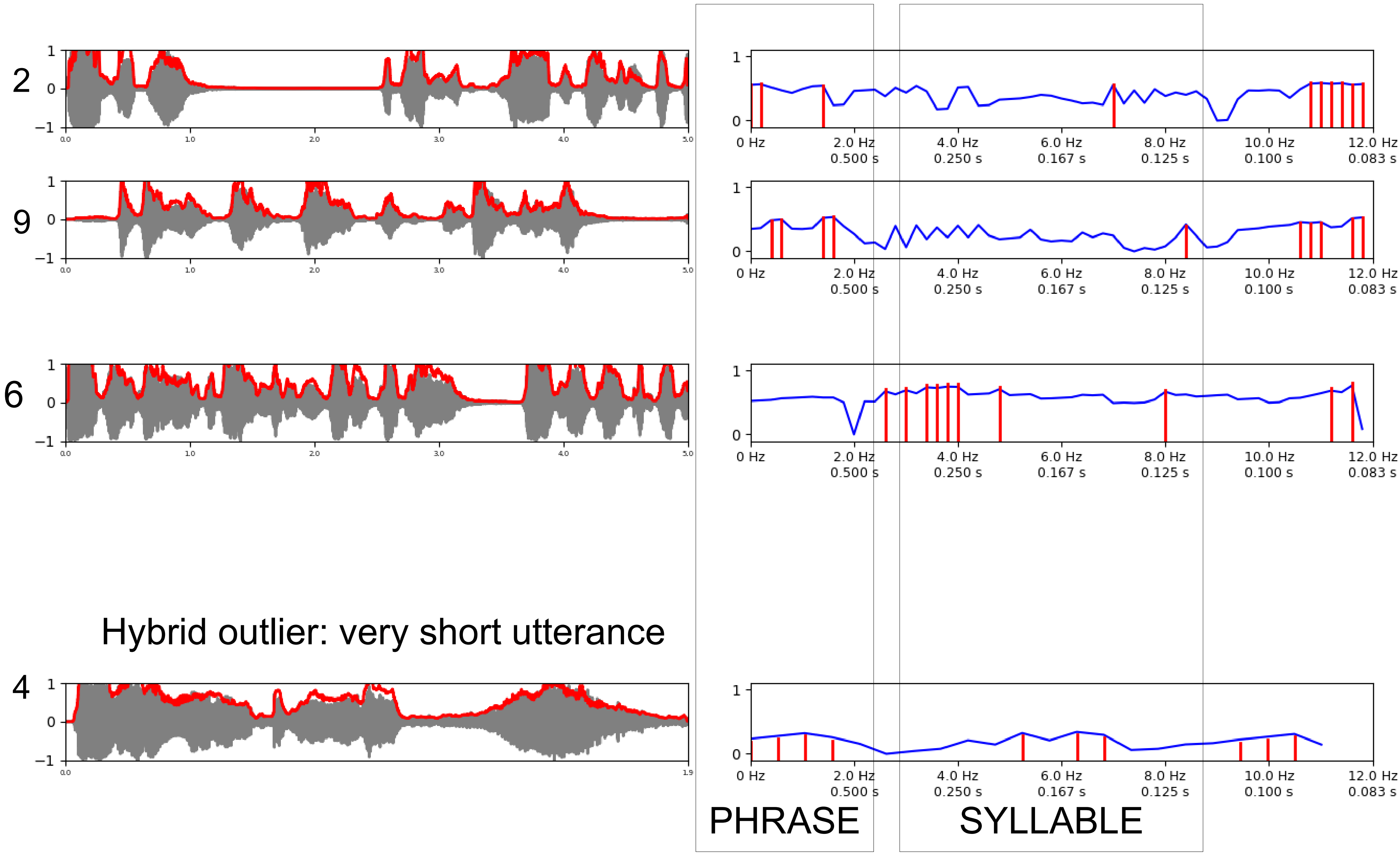


Hybrid outlier: very short utterance

4



Non-narrative style: phrase rhythms with pauses



Exploratory results for pilot case study

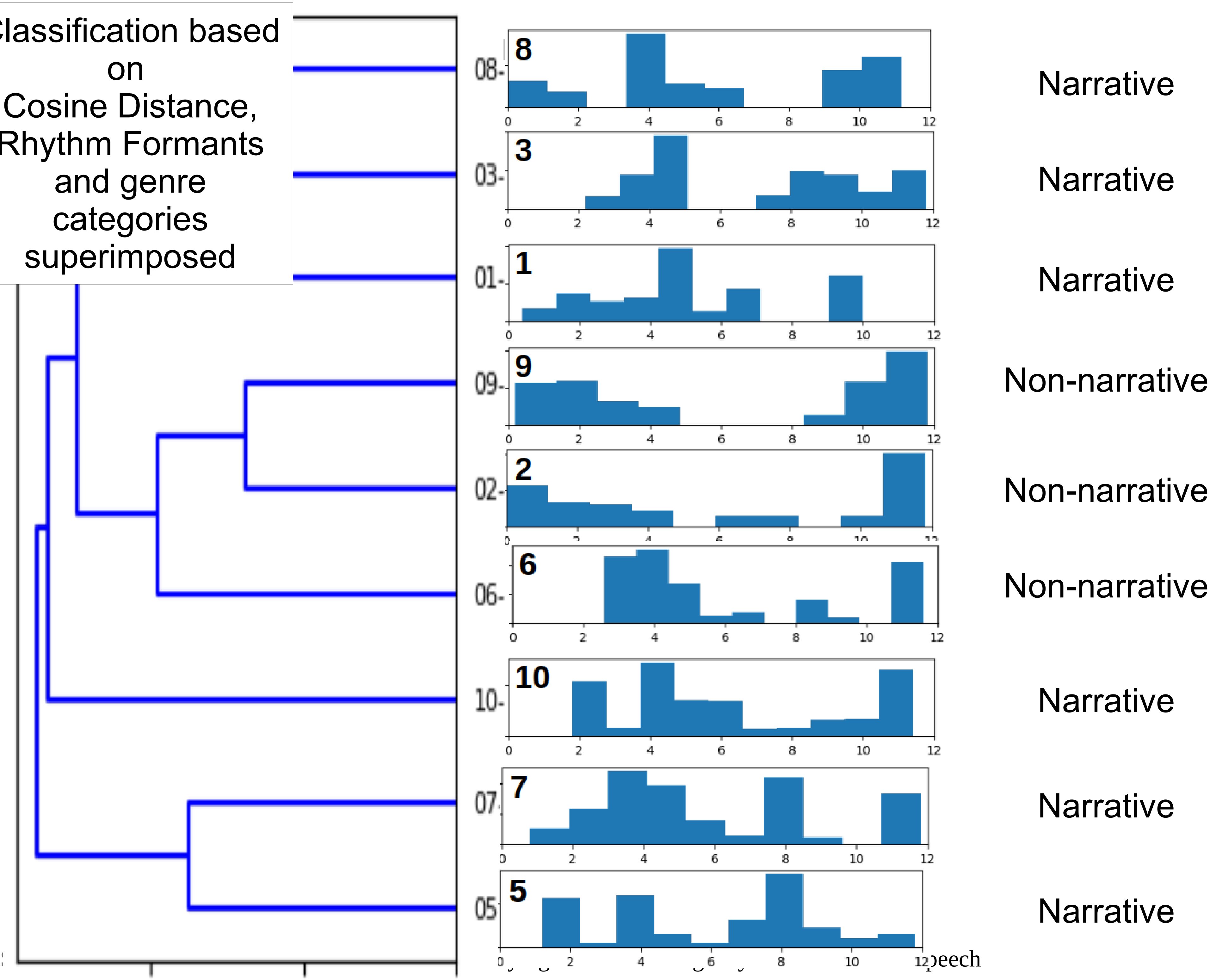
Approximate language unit correspondence	Narrative (1, 3, 5, 7, 8, 10)	Non-narrative (2, 4, 9)
weak syllables	approx. 11 Hz	approx. 11 Hz
strong syllables	approx. 4.5 Hz	
words/feet	approx. 2 Hz	
pause units		< 2Hz

Approximate language unit correspondence determined by comparison with annotations and automatic TGA (Time Group Analyser) analysis.

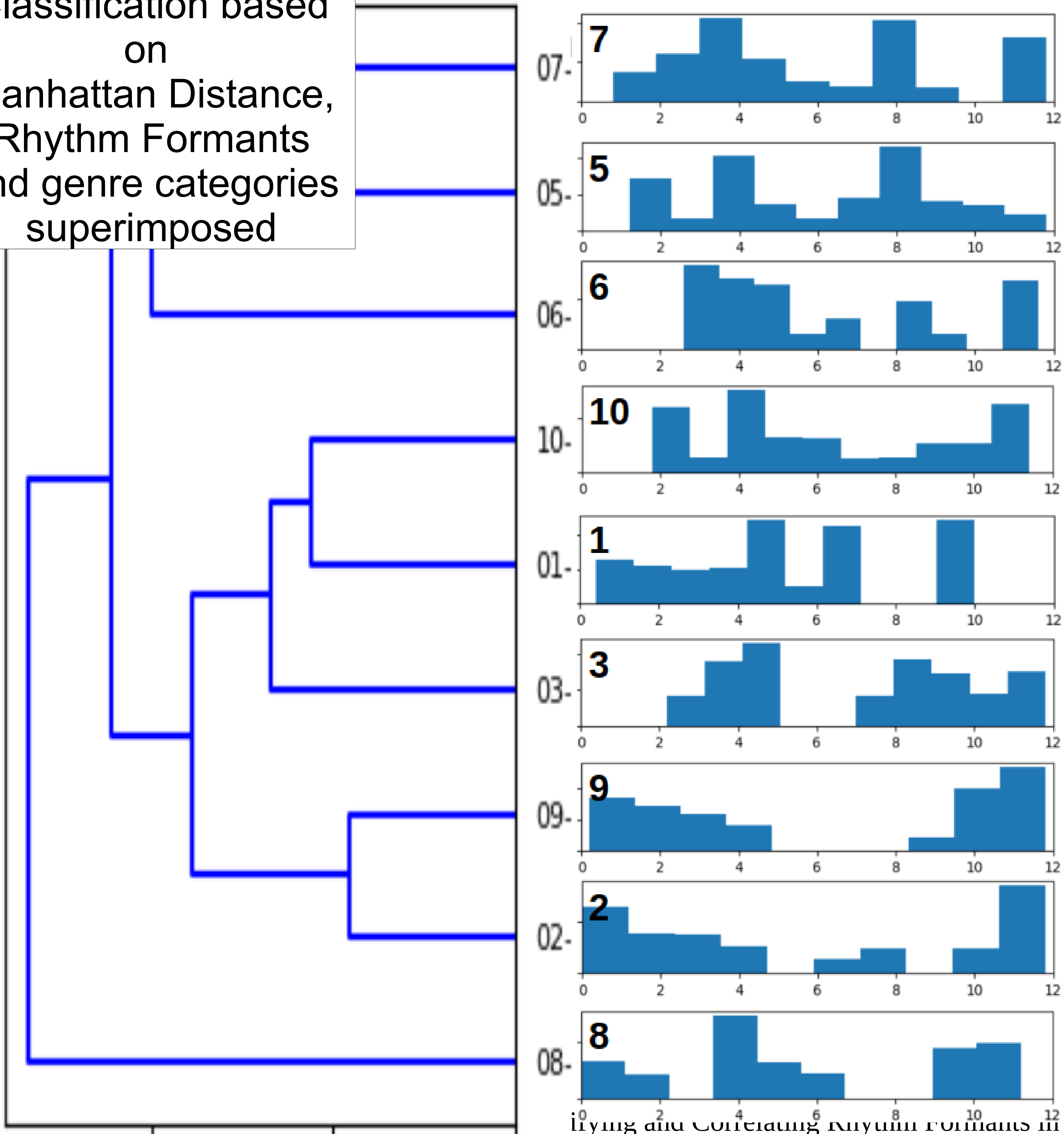
Test

Does automatic classification correspond to intuitive categories?

Classification based on
Cosine Distance,
Rhythm Formants
and genre
categories
superimposed



Classification based on
Manhattan Distance,
Rhythm Formants
and genre categories
superimposed



Narrative

Narrative

Non-narrative

Narrative

Narrative

Narrative

Non-narrative

Non-narrative

Narrative

Summary, Conclusion and Outlook

Summary

- Isochrony metric approaches
  - issues with isochrony metrics
  - *rPVI* and *nPVI* as modified distance metrics
  - Wagner’s 2-dimensional z-scored scatter plot quadrants
- Generalisation of formants to Rhythm Formant Theory
  - high frequency formants (voiced segments)
  - low frequency formants (rhythms)
- Rhythm Formant Analysis, case study: public speaking

- More specific issues are discussed in more detail in the paper, including:
  - the role of F0 / ‘pitch’ in rhythm patterning
  - other interpretations of the functionality of rhythms

Conclusion

Rhythm Formant Theory is ...

- language independent but linguistically interpretable
- oscillation-based
- perception-oriented
- explanatory and predictive *RHYTHM* theory, accounts for
  - relations between acoustic frequency ranges and language units
  - rhythmic variation in speech styles, genres, dialects, languages

Rhythm Formant Analysis …

- has a straightforward implementation
- permits fast analyses of case studies or large databases

Claim:

- potentially a versatile and future-oriented new paradigm

Outlook

- Research programme
  - Moving window for rhythm variation
  - Association with linguistic annotations
  - Validation with larger ‘clear case’ data sets
  - Application to data from different varieties:
    - genre: reading, public speaking, conversation, …
    - gender
    - age
    - dialects
  - Application to language typology data

Many thanks for your time and attention!