Recent Developments in Prosodic Phonetics

Lecture 2: The Facts

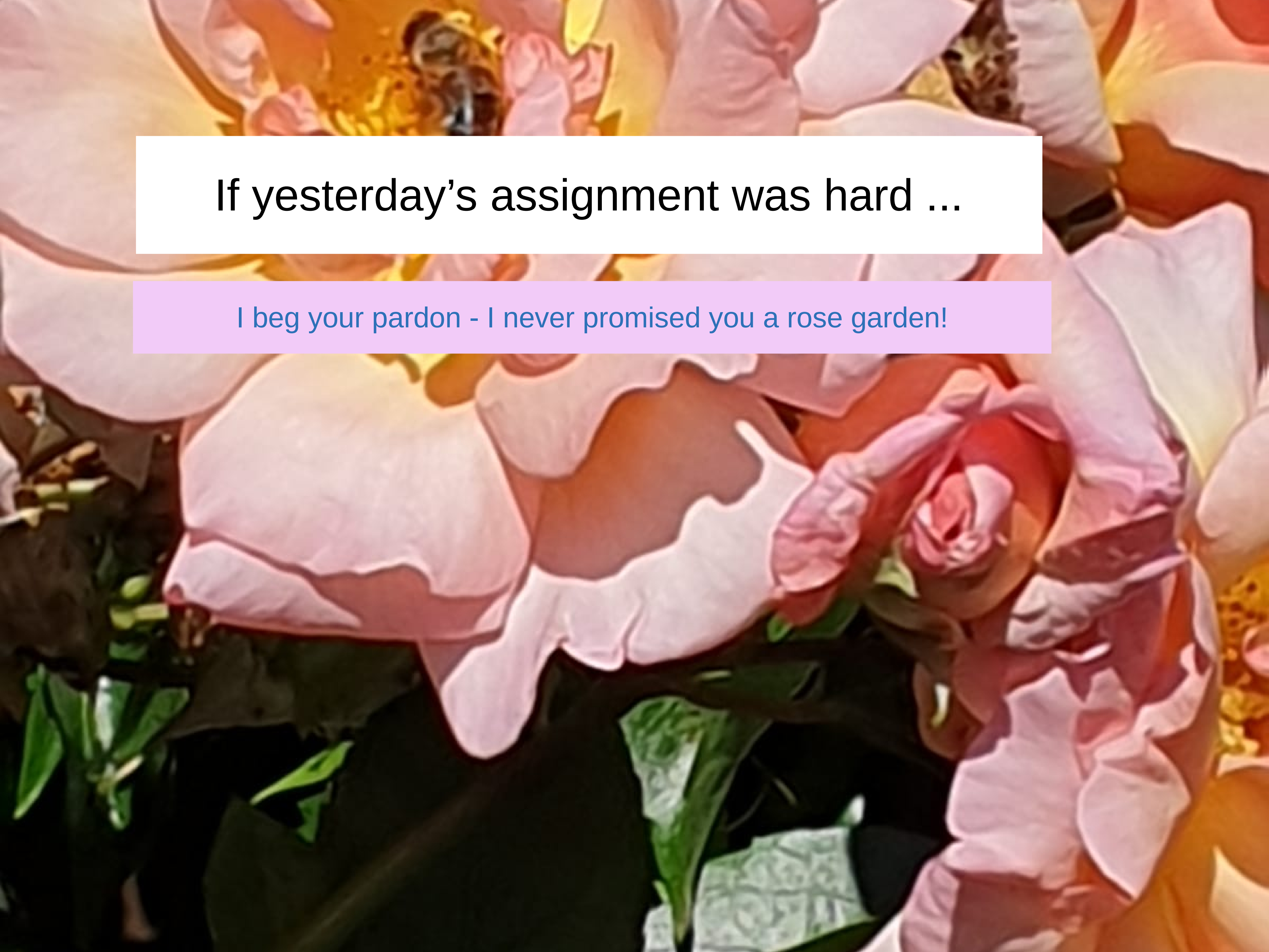
Speech Rhythm
and the role of Speech Melody

Dafydd Gibbon

U Bielefeld, Germany

Online Summer School: Contemporary Phonetics and Phonology, July 2021

If yesterday's assignment was hard ...

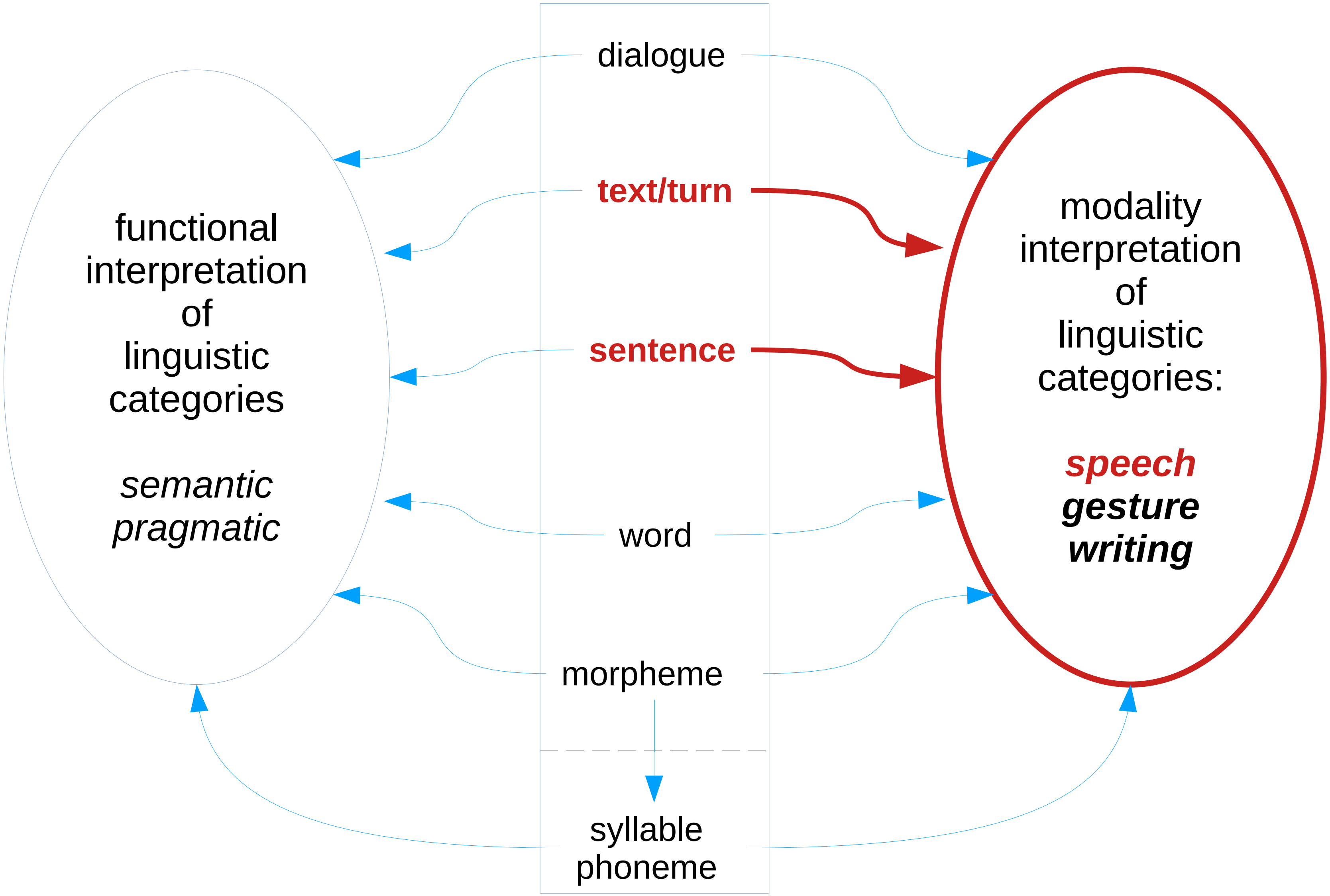


If yesterday's assignment was hard ...

I beg your pardon - I never promised you a rose garden!

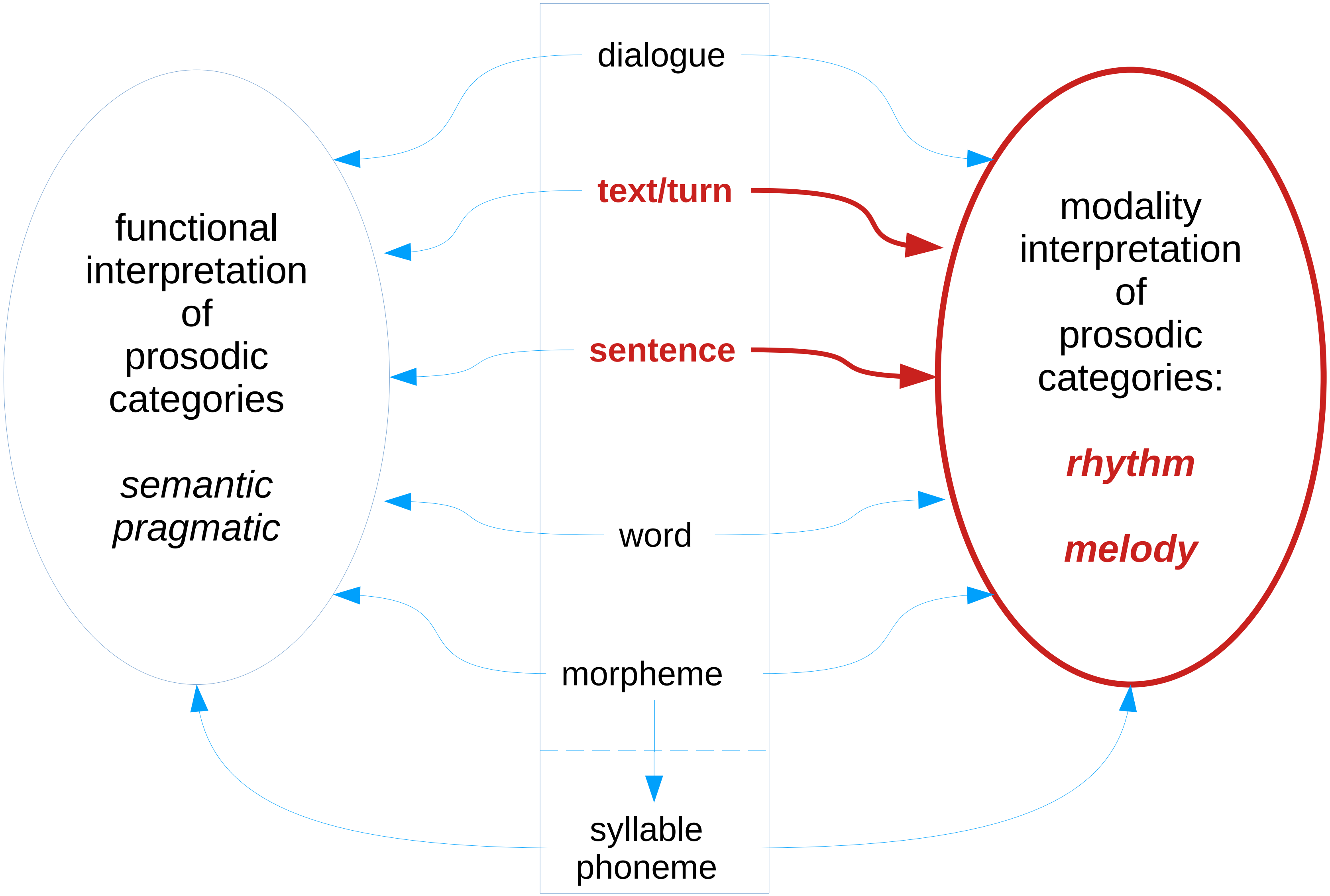
Semiotic Architecture of Speech and Language:

Rank Interpretation Theory

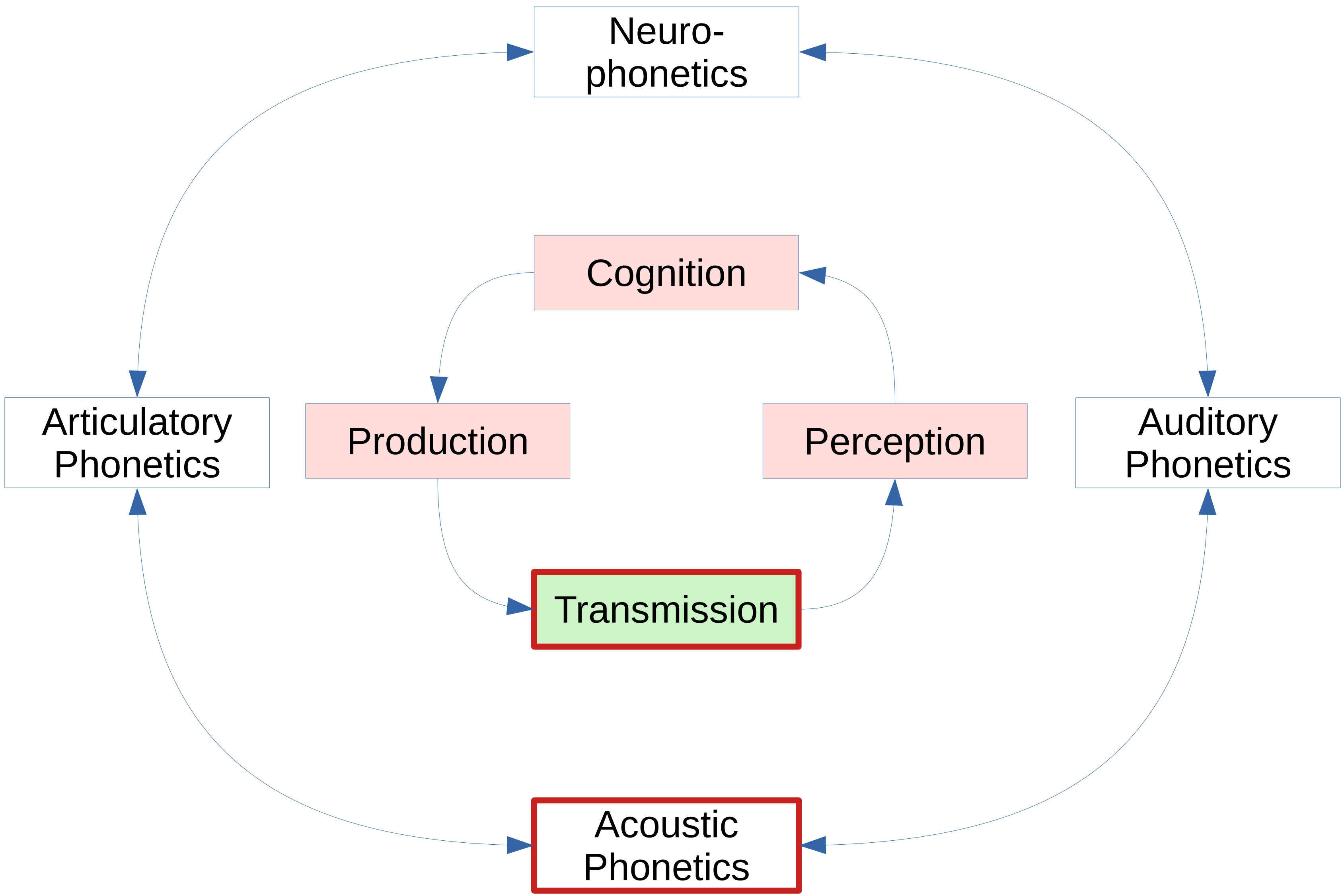


Semiotic Architecture of Speech and Language:

Rank Interpretation Theory

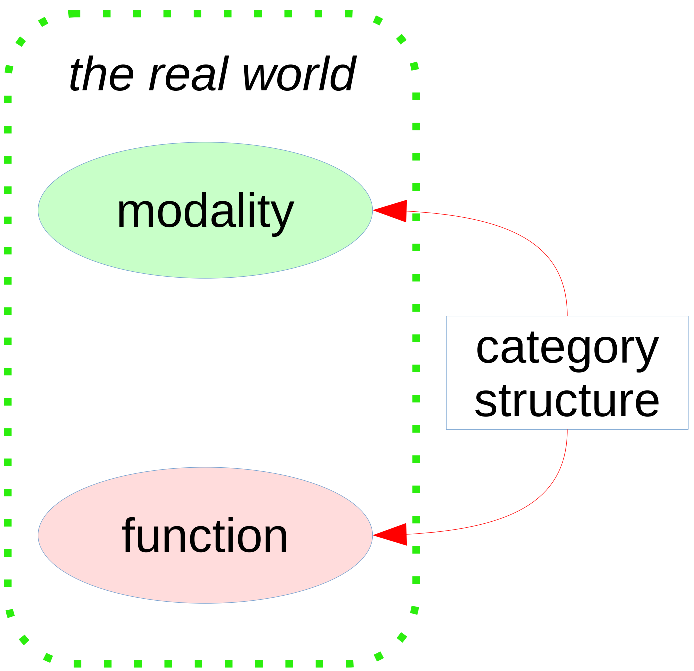


Phonetic domains and methods: the speech cycle



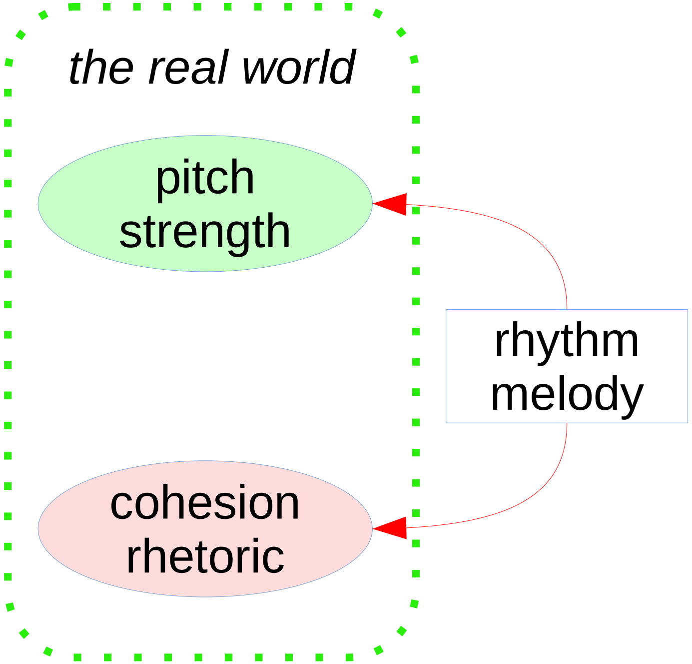
Semiotics of Prosody

- Prosodic categories are SIGNS
  - stress positions, pitch accents, tones
  - intonation in phrase, sentence, text/turn, dialogue
- That is: prosodic categories have
  - physical form with a modality interpretation
    - speech: pitch, rhythm, ...
    - gesture: speech, face, body, ...
    - writing: layout, punctuation, highlights
  - a functional interpretation
    - semantics: metadeictic
    - pragmatics: framing, appraisal

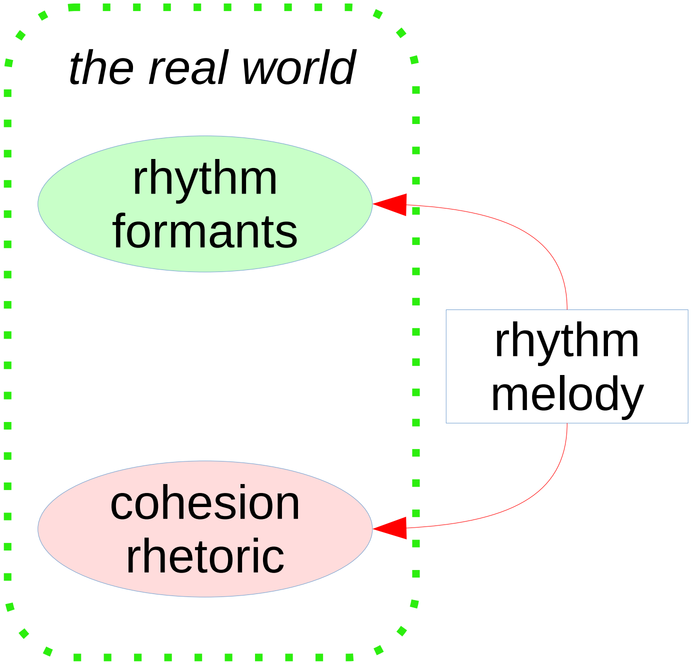


Reminder: Semiotics of Prosody

- Prosodic categories are SIGNS
  - stress positions, pitch accents, tones
  - intonation in phrase, sentence, text/turn, dialogue
- That is: prosodic categories have
  - a modality interpretation as a *Modulation Code*:
    - *Amplitude Modulation*
    - *Frequency Modulation*
  - a functional interpretation
    - semantics: metadeictic
    - pragmatics: framing, appraisal



Rhythm – Form and Meaning



First: basic prosody terminology

Phonetic parameters

In speech production:	phonation rate, glottal closure rate
In speech transmission:	fundamental frequency, F0
In speech perception:	pitch, pitch interval (semitone, minor third, octave, etc.)
Time:	interval duration
Strength:	amplitude, magnitude; intensity, energy; prominence; salience

Linguistic categories

Lexical tone:	correlates with frequency pattern (as in Chinese)
Lexical pitch accent:	correlates with frequency pattern (as in Japanese)
Lexical stress:	designated ‘strong’ position in word (as in English) correlates with stress-pitch accent, duration, intensity patterns
Intonation:	correlates with frequency pattern relative to the phrase (or larger sentence, text or discourse unit) or to the focussed word
Sentence stress:	designated ‘strong’ position in phrase or sentence correlates with frequency contour on focussed word

First: basic prosody terminology

Phonetic parameters

In speech production:	phonation rate, glottal closure rate
In speech transmission:	fundamental frequency, F0
In speech perception:	pitch, pitch interval (semitone, minor third, octave, etc.)
Time:	interval duration
Strength:	amplitude, magnitude; intensity, energy; prominence; salience

Linguistic categories

Lexical tone:	correlates with frequency pattern (as in Chinese)
Lexical pitch accent:	correlates with frequency pattern (as in Japanese)
Lexical stress:	designated ‘strong’ position in word (as in English) correlates with stress-pitch accent, duration, intensity patterns
Intonation:	correlates with frequency pattern relative to the phrase (or larger sentence, text or discourse unit) or to the focussed word
Sentence stress:	designated ‘strong’ position in phrase or sentence correlates with frequency contour on focussed word

Some inconsistencies to look out for:

- in phonetics: ‘pitch’, ‘pitch tracking’ are sometimes used instead of ‘F0’, ‘F0 estimation’ for the acoustic correlate in speech transmission (cf. ‘pitch’ is used for ‘F0’ in Praat, for example)
- in phonology: sometimes ‘stress’ is used in both a phonological and a phonetic sense
- the term ‘pitch accent’ is sometimes used for both Japanese distinctive pitch accent and the English correlate of stress, which has a variable relation to the pitch pattern, and which I therefore refer to as ‘stress-pitch accent’

Characterisation: three levels of rhythm abstraction

METRE

I **wan**dered **lone**ly **as** a **cloud**
That **floats** on **high** o'er **vales** and **hills**,
When **all** at **once** I **saw** a **crowd**,
A **host**, of **gold**en **daf**fo**dils**;
Be**side** the **lake**, be**neath** the **trees**,
Fluttering and **danc**ing **in** the **breeze**.

POETIC CONVENTION

Metrical conventions

iamb, trochee, pyrrhus, spondee
anapaest, dactyl
ambphibrach

STRUCTURAL RHYTHM

I **wan**dered **lone**ly as a **cloud**
That **floats** on **high** o'er **vales** and **hills**,
When **all** at **once** I **saw** a **crowd**,
A **host**, of **gold**en **daf**fodils;
Be**side** the **lake**, be**neath** the **trees**,
Fluttering and **danc**ing in the **breeze**.

CATEGORIES

Lexical and phrasal stress

nouns
verbs
adjectives
adverbs

PERFORMED RHYTHM (example)

I **wan**dered – **lone**ly **as** a **cloud**
That **floats** on **high** o'er **vales** and **hills**,
When – **all** at **once** – I saw a **crowd**, –
A **host**, – of **gold**en **daf**fo**dils**;
Be**side** the **lake**, – be**neath** the **trees**,
Fluttering – and **danc**ing in the **breeze**.

PHONETIC INTERPRETATION

Timing, Amplitude, Frequency

co-determined by
contrast, emphasis
pause, hesitation,
choice of global & local pitch contours

Rhythm performance: the Modulation Code

Methods of rhythm description

There are not so many explicit approaches to the modelling of rhythm:

1. Examples of traditional observational approaches
   1. Stress-pitch accent sequences (numbers; tonetic transcription)
   2. High-Low tone sequences (ToBi)
   3. Informal interpretative analysis (Couper-Kuhlen)
2. Examples of model, measurement, visualisation approaches:
   1. Deductive and annotation based:
      1. duration regularity and irregularity
      2. dispersion measures: e.g. standard deviation, pairwise variability
   2. Inductive and signal based:
      1. Spectral analysis (Todd, Tilsen)
      2. Spectrogram analysis (Gibbon)
      3. Empirical mode decomposition (Tilsen & Arvaniti)

Modulation with Rhythm Formants

1. **Rhythm Formant Theory (RFT):**
   1. A rhythm formant is a frequency zone of higher magnitude values in the normalised low frequency (LF) spectrum.
   2. Rhythm formants are detected in the LF AM spectrum and in the LF FM spectrum.
2. **Rhythm Formant Analysis (RFA):**
   1. The spectrum magnitudes are obtained by FFT and normalised to the range 0,…,1.
   2. A minimum magnitude (e.g. about 0.2) is defined as a cutoff level, below which values are clipped to zero.
   3. The clipped spectra of different recordings are compared using standard distance metrics and represented as distance maps, and hierarchically clustered using standard clustering criteria and represented as dendrograms.

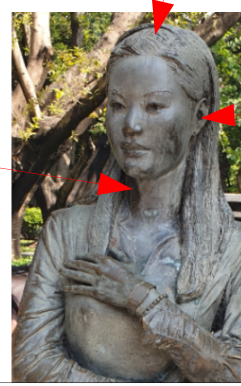
Modulation with rhythm

Amplitude modulation (AM) and Frequency Modulation (FM)

STRUCTURED INFORMATION

FM generator:

tone, pitch accent intonation



AM filter:

phones, syllables, words, phrases, ...

COMPLEX SPEECH SIGNS
gestures
waves
percepts

frequency modulated X amplitude modulated sound

Sound generator:

carrier frequency + harmonics

frequency modulated sound

Characterisation of rhythm modulation

Rhythm modulation consists of

fairly regular **oscillations in the speech signal below about 10 Hz**

- which **modulate the speech source carrier signal** as rhythm formants

  and are detectable **in spectral analysis as magnitude peaks**

  in the **low frequency (LF) spectrum** of both

  - the amplitude modulation (**AM**) envelope of the speech signal

    related to the syllable sonority outline of the waveform, and

  - the frequency modulation (**FM**) envelope of the speech signal,

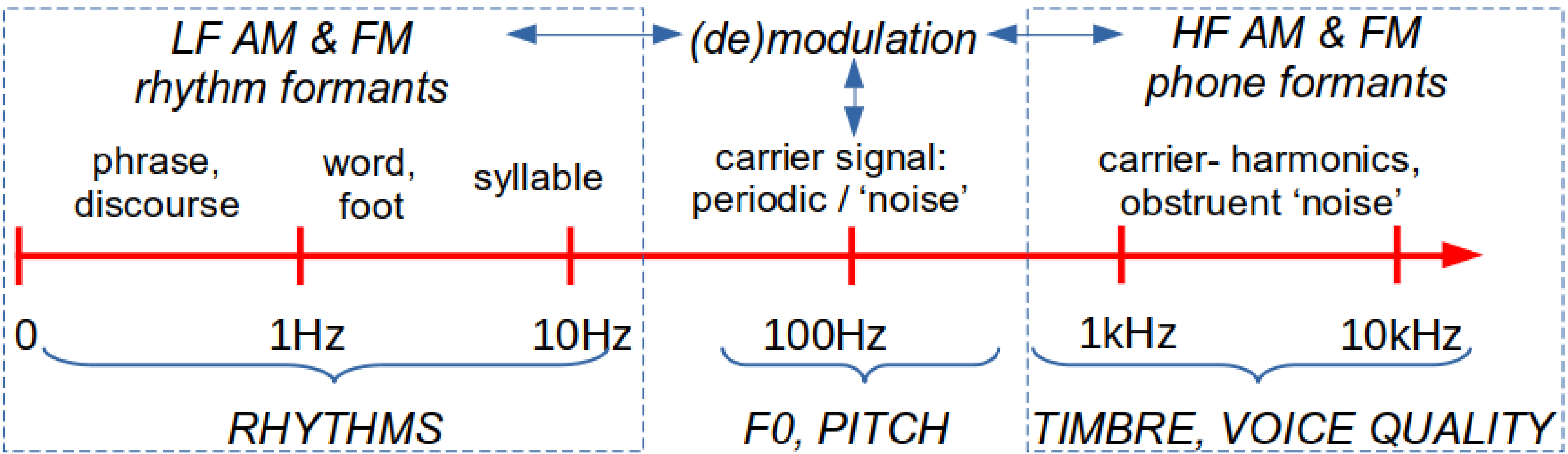
    related to perceived pitch contours

Modulation Code: the Frequency Scale

Low frequencies:
rhythm

Mid frequencies:
rhythm

High frequencies:
consonants and vowels



Low Frequency
AM and FM modulations

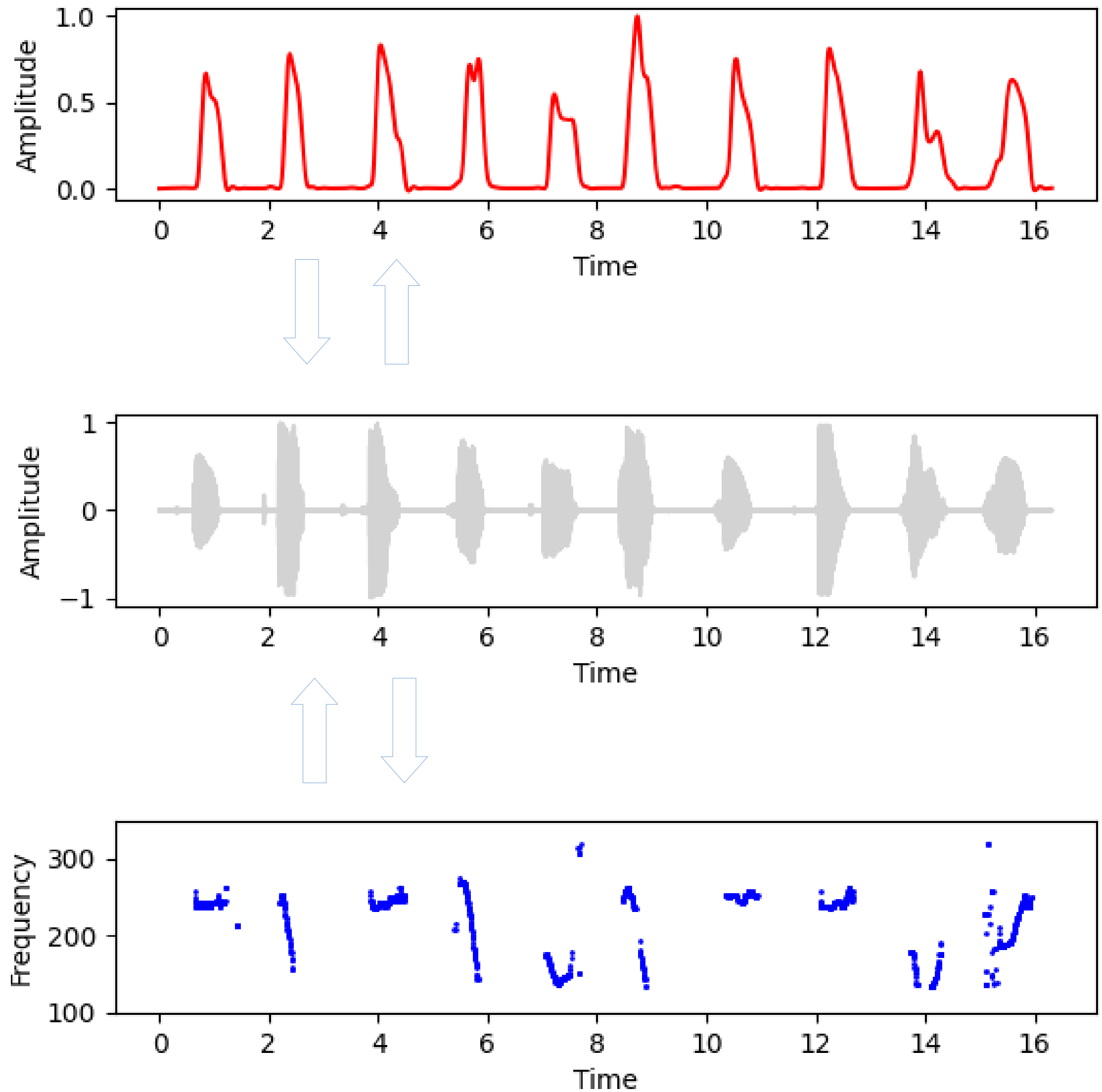
High Frequency
AM and FM modulations

Modulation code: modulation and demodulation

AM envelope modulation signal:
- phonetics:
  amplitude curve, syllable, stress-accent
- phonology:
  sonority curve, syllables, stress

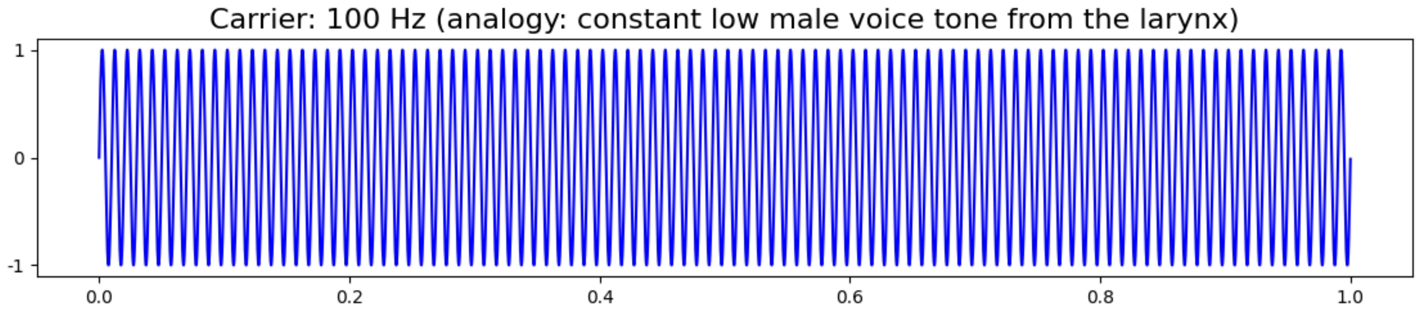
Carrier signal:
source:
- larynx: harmonic sounds
- constriction: noise sounds

FM envelope modulation signal:
- phonetics:
  F0, pitch track
- phonology:
  tones, pitch accents, intonation



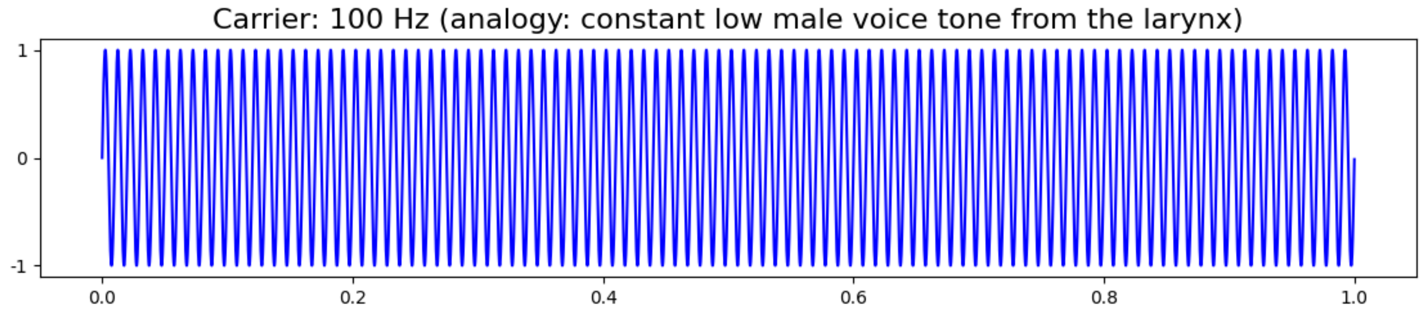
AM and FM: modulation and demodulation

Carrier signal:

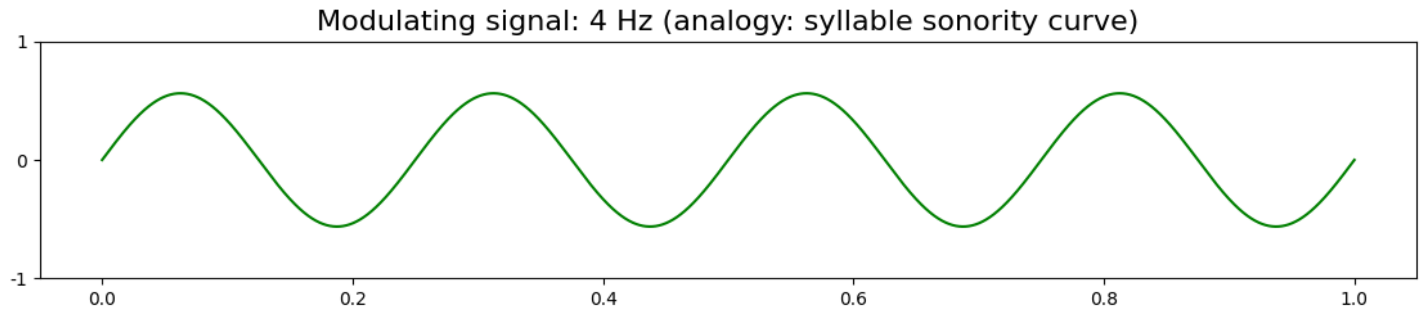


AM and FM: modulation and demodulation

Carrier signal:



Modulation signal (information signal):

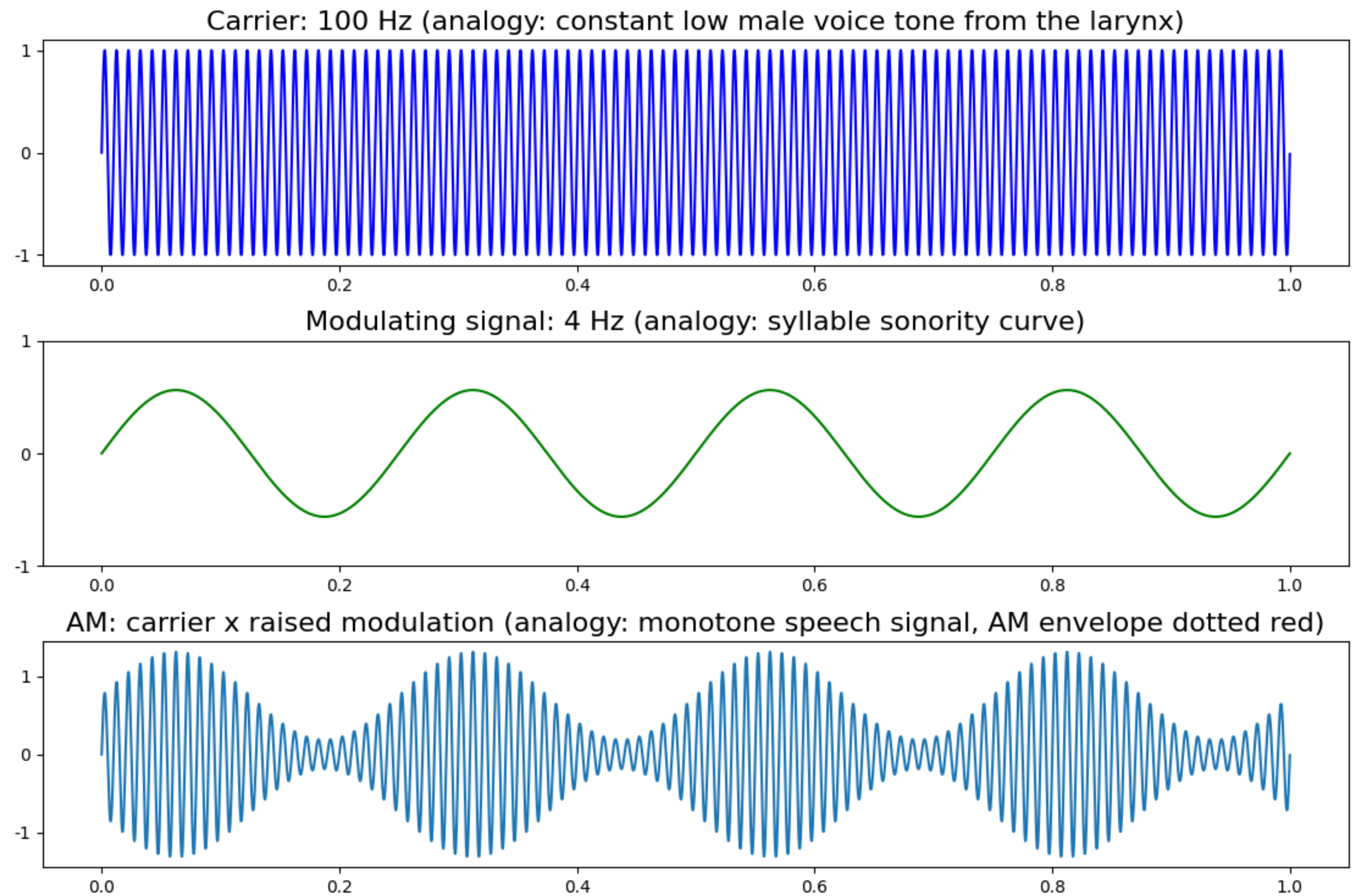


AM and FM: modulation and demodulation

Carrier signal:

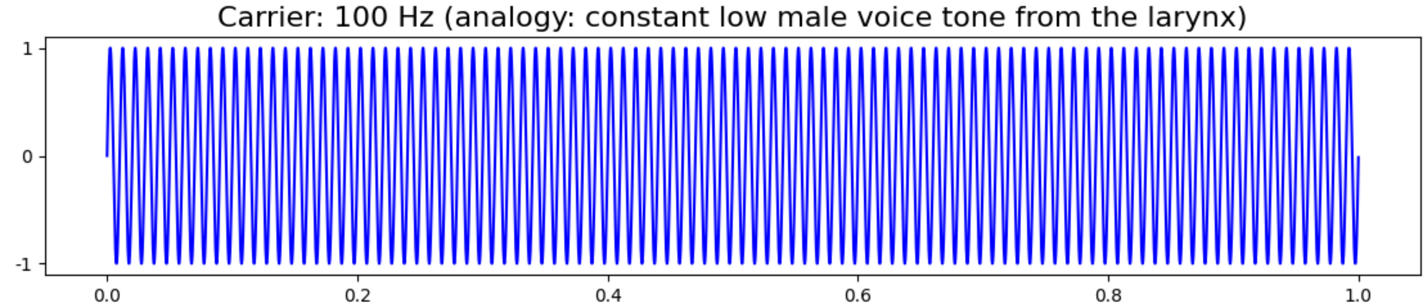
Modulation signal (information signal):

Amplitude modulated signal:

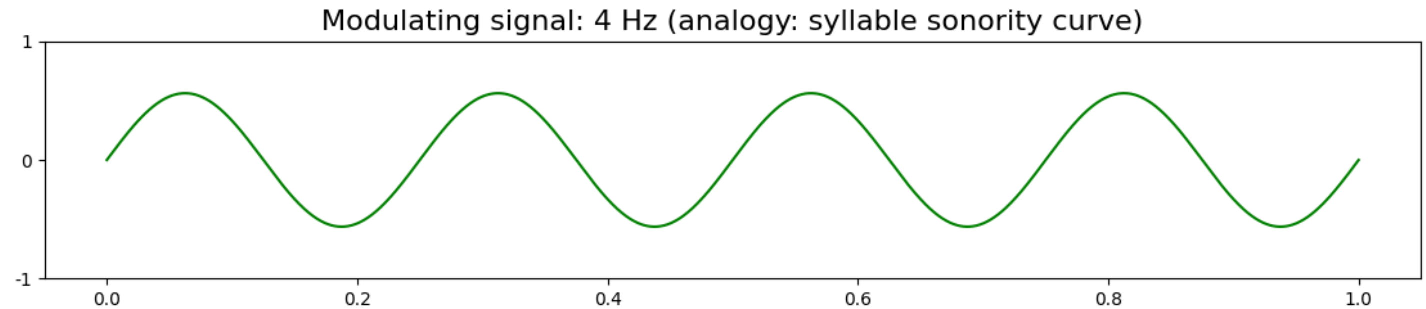


AM and FM: modulation and demodulation

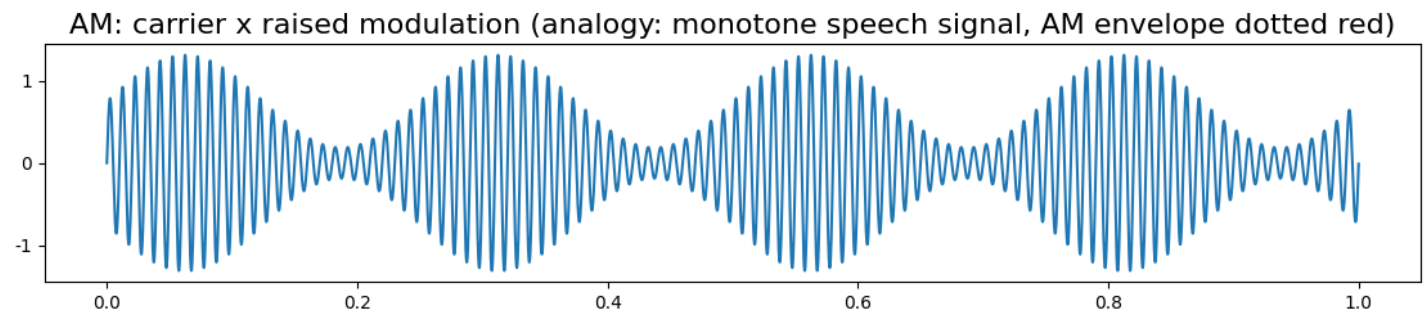
Carrier signal:



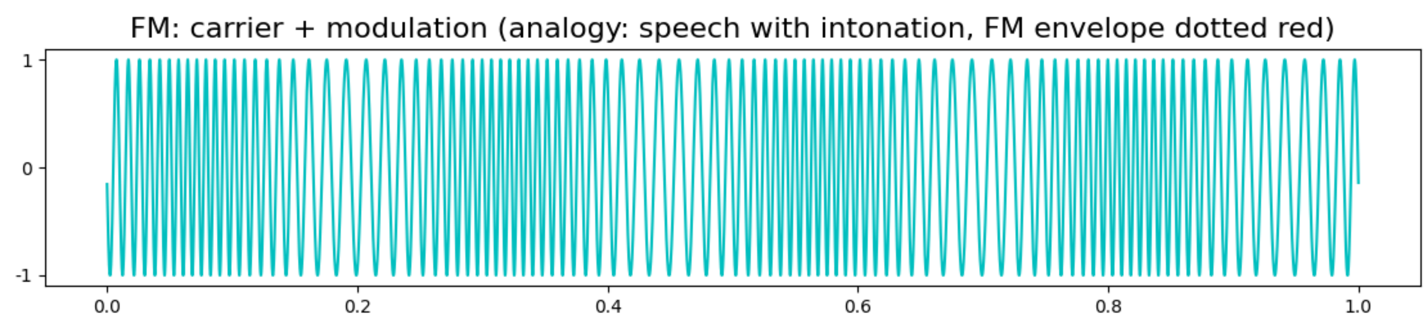
Modulation signal:



Amplitude modulated signal:



Frequency modulated signal:

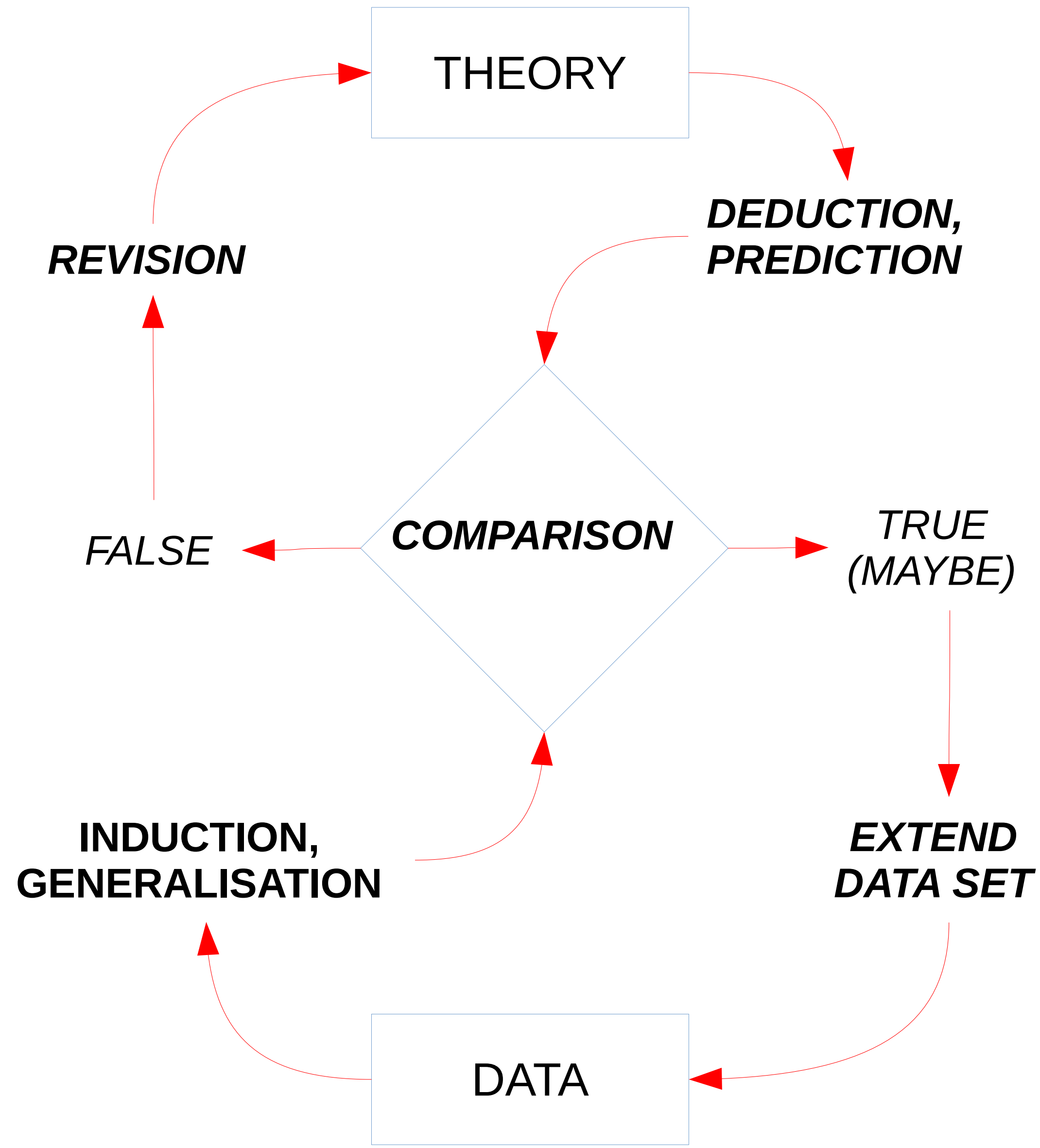


Critical Rationalist methods in linguistics and phonetics

The Logic

From a logical point of view, it is impossible to confirm a theory – you may find a counter-example any time!

So we try hard to falsify the theory. If we succeed, we revise predictions in our research space. If we fail to falsify in many experiments, we call it confirmation.



Induction plus deduction

direct observation, transcription and measurement

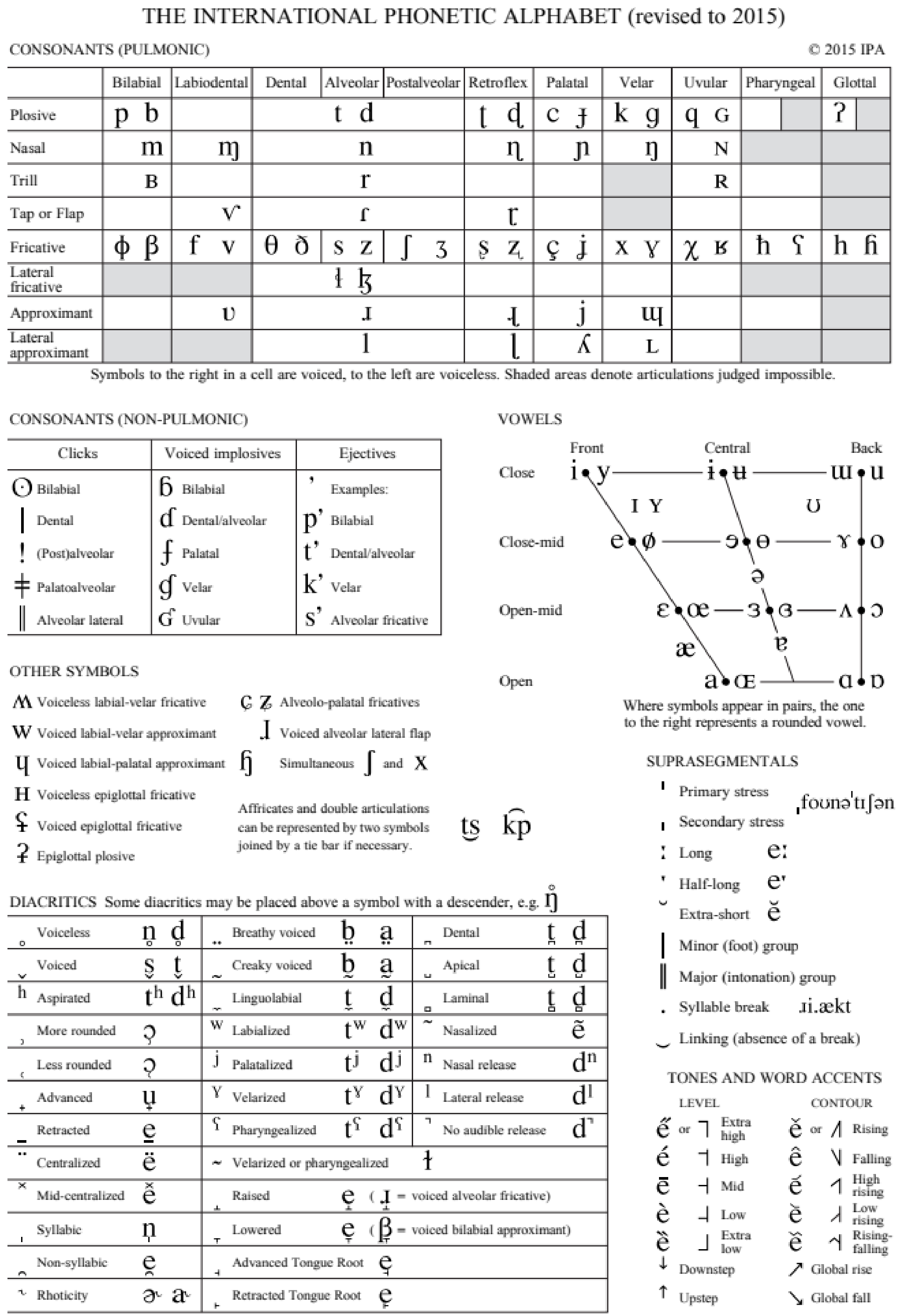
IPA

Seven categories of speech sounds

Clicks
Implosives
Ejectives

Coarticulations
Other sounds

Modifications



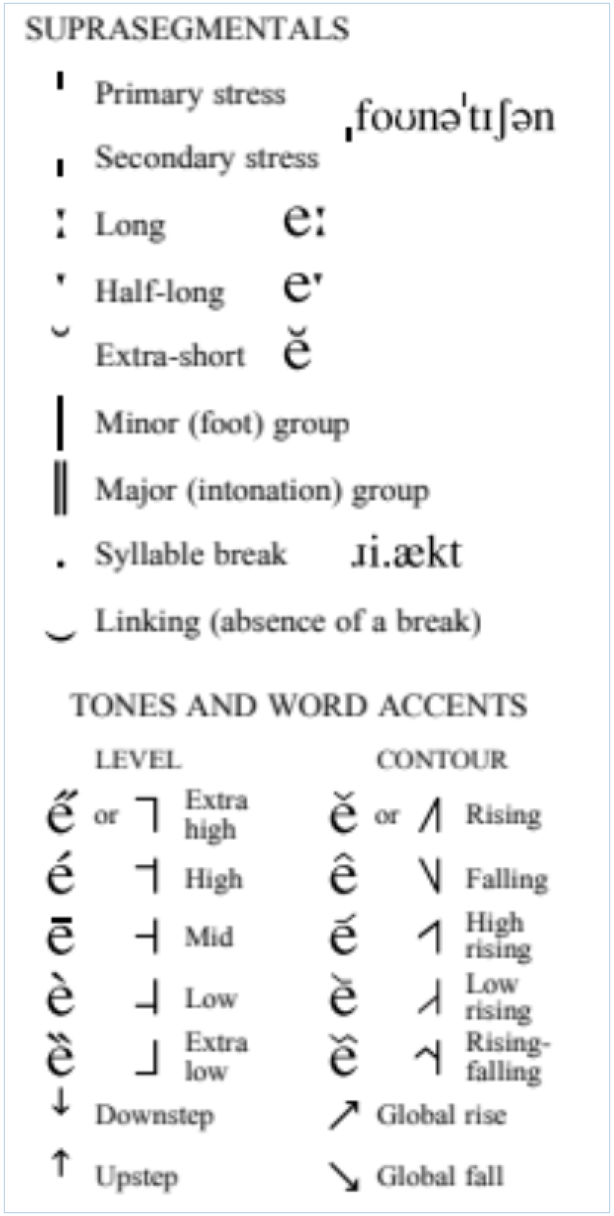
Consonants

Vowels

Stress
Duration
Group breaks
Syllables
Links

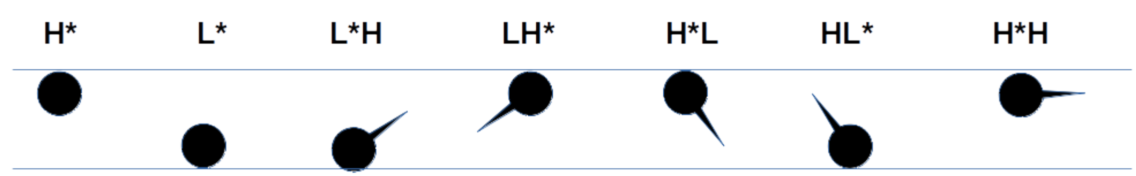
Tones
Accents

Transcription of aspects of of rhythm



There are other conventions

1. Tadpole notation
2. Global contour notation
3. Tonetic stress marks
4. Alphabetic tone marks (ToBI: H, L)
5. Africanist high-low tone convention
6. Sinologist local contour tone convention
7. Numerical marking
   - stress level
   - tone height (Pike; Chao 5-level convention)
   - tone type (Chinese)



F0 modelling

There was a question after the first lecture about different paradigms: here is a selection.

Examples of model, measurement, visualisation approaches:

1. Regression fitting over different domains (my preferred approach)
2. Spline interpolation (Hirst, *Empirical Mode Decomposition*)
3. Pulse sequence with smoothing (Fujisaki Model)
4. Tilt (Black, rise-fall events)

Inductive analysis: from pitch patterns to categories

Phonetic mode (signal analysis):
- Domains:
  - time functions (articulatory, acoustic, auditory)
- Analysis:
  - time domain
  - frequency domain (spectrum)

rhythms
melodies

Tonal tokenisation (e.g. Tobi):

BoundaryTone PitchAccentTone PitchAccentTone* BoundaryTone

Boundary tone: { **H%**, **%L%** }

PitchAccentTone: { **H***, **L***, **L*H**, **LH***, **H*L**, **HL***, **H*H** }

Contour parsing:
prehead head body nucleus tail

Categorial interpretation (prosodic phonologies):
- Configurative: Initial/final boundary; ip, IP boundary
- Contrastive: accents
- Culminative: accent placement

Annotation: segmentation and classification

(‘labelling’)

*and the search for isochrony**

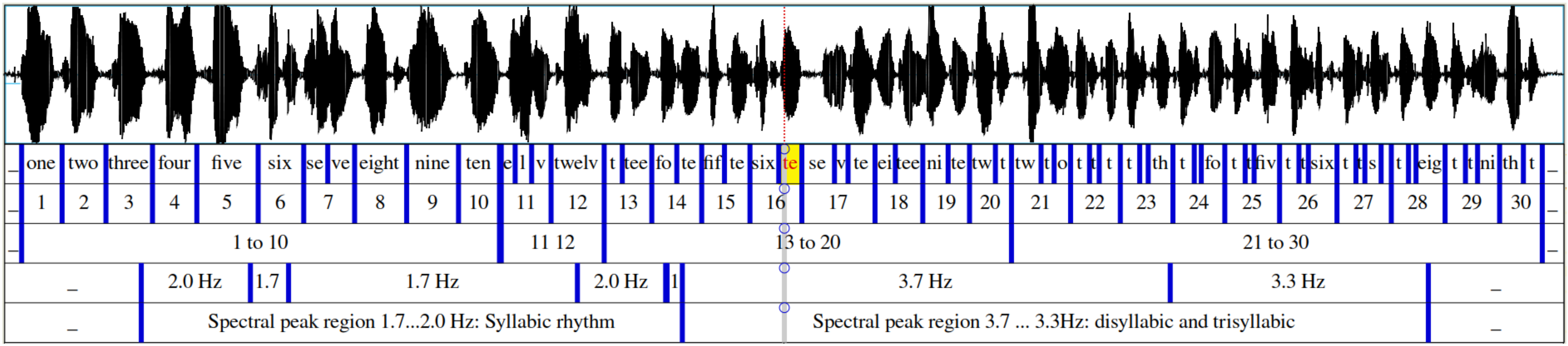
a hybrid deductive-inductive method

isochrony: equal timing, for example as an idealised phonetic interpretation of rhythm units like syllables or stress groups

The deductive annotation-based approach

1. Decide on a set of linguistic categories and inventories
   - phonetic
   - phonological
   - morphological
   - grammatical (part of speech, PoS tags)
   - semantic
   - pragmatic
        1. speech acts
        2. turn-taking
        3. discourse grammar
2. Analyse relevant speech data
   - Search for and record data
   - Listen, transcribe, annotate

Event annotation with ‘Praat’: intervals and labels



Download Praat

- https://www.fon.hum.uva.nl/praat/
- https://www.praat.org

Data

- **Pre-recording**
  - Design systematic filenames
  - Design data scenario
  - Prepare equipment and participants
    - You can record with Praat or Audacity
- **Recording**
  - record with proper distance (1 span)
  - enough to drink
- **Post-recording**
  - save with systematic filename
  - archive systematically

Annotate with Praat

- Read into Praat
- Select “Annotation”
- Annotate.
- Save with systematic filenameAnalysis
- Convert the Praat format to CSV spreadsheet format
  - This can be done easily with a Python script. Contact me.

Analyse the spreadsheet file

- With a spreadsheet.
- With TGA (Time Group Analyser) online tool
  - http://wwwhomes.uni-bielefeld.de/gibbon/TGA/ speaker)

Or ask Dr. Yu Jue!

Event annotation with 'Praat': intervals and labels

What you get is this, the TextGrid format:

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 11.017875
tiers? <exists>
size = 3
item []:
item [1]:
class = "IntervalTier"
name = "Syllables"
xmin = 0
xmax = 11.017875
intervals: size = 62
intervals [1]:
xmin = 0
xmax = 0.48339725121628835
text = "_"
intervals [2]:
xmin = 0.48339725121628835
xmax = 0.6964283269433246
text = "one"
intervals [3]:
xmin = 0.6964283269433246
xmax = 0.9009381596412812
text = "two"
intervals [4]:
xmin = 0.9009381596412812
xmax = 1.155155243342209
text = "three"
intervals [5]:
xmin = 1.155155243342209
xmax = 1.4091692796134065
text = "four"
intervals [6]:
xmin = 1.4091692796134065
xmax = 1.6293013911980108
text = "five"
```

What you need is this, the CSV format:

File	Tier	Label	Start	End	Duration
one-to-thirty-11s_16k	Syllables	_	0.000	0.249	0.249
one-to-thirty-11s_16k	Syllables	_	0.249	0.483	0.234
one-to-thirty-11s_16k	Syllables	one	0.483	0.696	0.213
one-to-thirty-11s_16k	Syllables	two	0.696	0.901	0.205
one-to-thirty-11s_16k	Syllables	three	0.901	1.155	0.254
one-to-thirty-11s_16k	Syllables	four	1.155	1.409	0.254
one-to-thirty-11s_16k	Syllables	five	1.409	1.629	0.220
one-to-thirty-11s_16k	Syllables	six	1.629	1.883	0.254
one-to-thirty-11s_16k	Syllables	se	1.883	2.020	0.137
one-to-thirty-11s_16k	Syllables	ven	2.020	2.148	0.128
one-to-thirty-11s_16k	Syllables	eight	2.148	2.328	0.180
one-to-thirty-11s_16k	Syllables	nine	2.328	2.551	0.223
one-to-thirty-11s_16k	Syllables	ten	2.551	2.751	0.200
one-to-thirty-11s_16k	Syllables	e	2.751	2.821	0.070
one-to-thirty-11s_16k	Syllables	le	2.821	2.936	0.115
one-to-thirty-11s_16k	Syllables	ven	2.936	3.020	0.084
one-to-thirty-11s_16k	Syllables	twelve	3.020	3.296	0.276
one-to-thirty-11s_16k	Syllables	thir	3.296	3.461	0.165
one-to-thirty-11s_16k	Syllables	teen	3.461	3.615	0.154
one-to-thirty-11s_16k	Syllables	four	3.615	3.764	0.149
one-to-thirty-11s_16k	Syllables	teen	3.764	3.921	0.157
one-to-thirty-11s_16k	Syllables	fif	3.921	4.056	0.135
one-to-thirty-11s_16k	Syllables	teen	4.056	4.222	0.166
one-to-thirty-11s_16k	Syllables	six	4.222	4.449	0.227
one-to-thirty-11s_16k	Syllables	teen	4.449	4.547	0.098
one-to-thirty-11s_16k	Syllables	se	4.547	4.680	0.133
one-to-thirty-11s_16k	Syllables	ven	4.680	4.748	0.068
one-to-thirty-11s_16k	Syllables	teen	4.748	4.920	0.172
one-to-thirty-11s_16k	Syllables	eigh	4.920	5.025	0.105
one-to-thirty-11s_16k	Syllables	teen	5.025	5.208	0.183
one-to-thirty-11s_16k	Syllables	nine	5.208	5.356	0.148
one-to-thirty-11s_16k	Syllables	teen	5.356	5.506	0.150
one-to-thirty-11s_16k	Syllables	twen	5.506	5.734	0.228
one-to-thirty-11s_16k	Syllables	ty	5.734	5.863	0.129
one-to-thirty-11s_16k	Syllables	twen	5.863	6.036	0.173
one-to-thirty-11s_16k	Syllables	ny	6.036	6.100	0.064
one-to-thirty-11s_16k	Syllables	one	6.100	6.230	0.130
one-to-thirty-11s_16k	Syllables	twen	6.230	6.432	0.202
one-to-thirty-11s_16k	Syllables	ty	6.432	6.550	0.118
one-to-thirty-11s_16k	Syllables	two	6.550	6.703	0.153
one-to-thirty-11s_16k	Syllables	twen	6.703	6.896	0.193
one-to-thirty-11s_16k	Syllables	ty	6.896	6.959	0.063
one-to-thirty-11s_16k	Syllables	three	6.959	7.132	0.173
one-to-thirty-11s_16k	Syllables	twen	7.132	7.321	0.189
one-to-thirty-11s_16k	Syllables	ty	7.321	7.407	0.086
one-to-thirty-11s_16k	Syllables	four	7.407	7.561	0.154
one-to-thirty-11s_16k	Syllables	twen	7.561	7.741	0.180
one-to-thirty-11s_16k	Syllables	ty	7.741	7.793	0.052
one-to-thirty-11s_16k	Syllables	five	7.793	8.003	0.210
one-to-thirty-11s_16k	Syllables	twen	8.003	8.192	0.189
one-to-thirty-11s_16k	Syllables	ty	8.192	8.239	0.047
one-to-thirty-11s_16k	Syllables	six	8.239	8.477	0.238
one-to-thirty-11s_16k	Syllables	twen	8.477	8.674	0.197
one-to-thirty-11s_16k	Syllables	sen	8.674	8.903	0.229
one-to-thirty-11s_16k	Syllables	twen	8.903	9.071	0.168
one-to-thirty-11s_16k	Syllables	ny	9.071	9.174	0.103
one-to-thirty-11s_16k	Syllables	eight	9.174	9.302	0.128
one-to-thirty-11s_16k	Syllables	twen	9.302	9.462	0.160
one-to-thirty-11s_16k	Syllables	ny	9.462	9.559	0.097
one-to-thirty-11s_16k	Syllables	nine	9.559	9.745	0.186
one-to-thirty-11s_16k	Syllables	thir	9.745	9.996	0.251
one-to-thirty-11s_16k	Syllables	ty	9.996	10.151	0.155
one-to-thirty-11s_16k	Syllables	_	10.151	11.018	0.867

textgridtier2csv.py

```
#!/usr/bin/python
# textgridtier2csv.py D. Gibbon 2015.02.12

# Convert a Praat TextGrid tier to CSV  format

#-----------------------------------------------------------
# Import standard modules

import os, re, sys

#-----------------------------------------------------------
# Input TextGrid from CLII

if len(sys.argv) < 3:
    print("Usage:",sys.argv[0],'<filename> <tiername>')
    exit()
fname = sys.argv[1]
tname = sys.argv[2]
if not os.path.isfile(fname):
    print("File",fname,"does not exist.")
    exit()

textgrid = open(fname,'r').read().split('\n')
fname = sys.argv[1].split('.')[0]

#-----------------------------------------------------------
# Remove initial and final spaces

nugrid = []
for l in textgrid:
    a = "
    l = re.sub(' *$',",l)
    l = re.sub('^ *',",l)
    l = re.sub('\"',",l)
    if l != ":
        nugrid += [l]
```

```
def extracttiers(nugrid,outflag):
    tierkey = "
    returnstring = "
    output = "
    val = "
    start = 0
    if not outflag in ['file','string']:
        tierkey = outflag
    for i in range(len(nugrid)):
        l1 = nugrid[i].split(' = ')
        if len(l1) > 1:
            val = l1[1]
        if val == 'IntervalTier':
                if start > 0:
                    if tierkey == tiername:
                        return output
                    if outflag == 'file':
                        open(fname+'-'+tiername+'.csv','w').write(output)
                    returnstring += output
                output = "
                tiername = nugrid[i+1].split(' = ')[1]
                start = 1
        l2 = nugrid[i].split(' ')
        if l2[0] == 'intervals':
            xmin = "%.3f"%float(nugrid[i+1].split(' = ')[1])
            xmax = "%.3f"%float(nugrid[i+2].split(' = ')[1])
            text = nugrid[i+3].split(' = ')[1]
            dur = "%.3f"%(float(xmax)-float(xmin))
            interval = "\t".join([fname,tiername,text,xmin,xmax,dur])+"\n"

#           interval = fname+'\t'+tiername+'\t'+text+'\t'+"%.3f"%xmin+'\
t'+"%.3f"%xmax+'\t'+"%.3f"%dur+'\n'
            output += interval
    if outflag == 'file':
        open(fname+'-'+tiername+'.csv','w').write(output)
    if outflag == 'string':
        returnstring += output
        return returnstring
    else:
        return "
print(extracttiers(nugrid,tname))
```

Interval analysis and PVI – the search for <u>isochrony</u>

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

raw Pairwise Variability Index

The measure defines an overall ‘next-door neighbour distance’.

Interval analysis and PVI – the search for isochrony

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

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

raw Pairwise Variability Index

normalised Pairwise Variability Index

Interval analysis and PVI – the search for isochrony

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

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

raw Pairwise Variability Index

normalised Pairwise Variability Index

Similarity to Manhattan Distance

Similarity to Canberra Distance (Normalised Manhattan Distance)

$$MD(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

$$NormMD(x, y) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

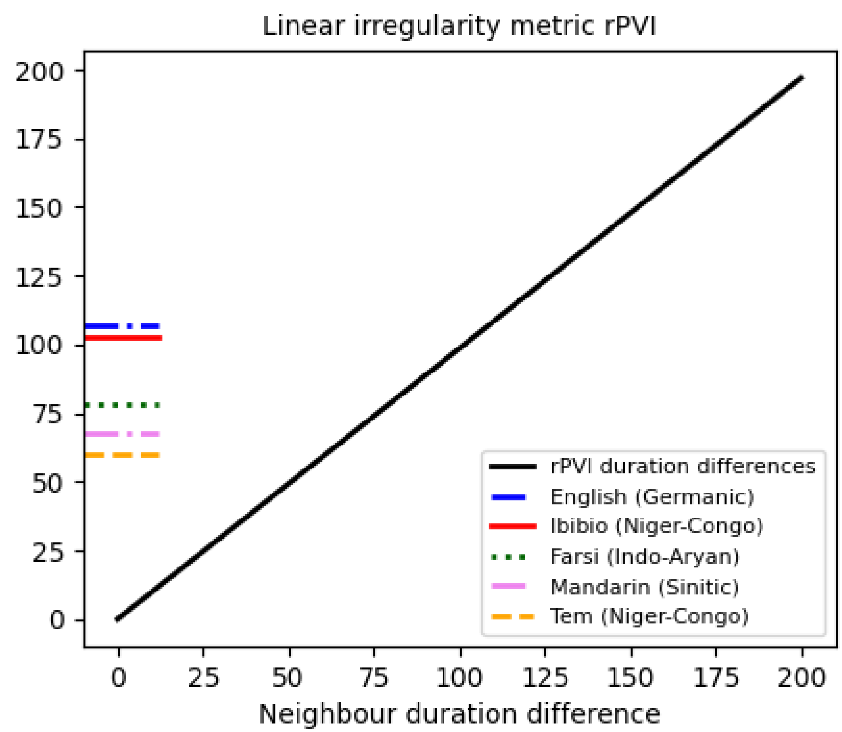
Interval analysis and PVI – the search for isochrony

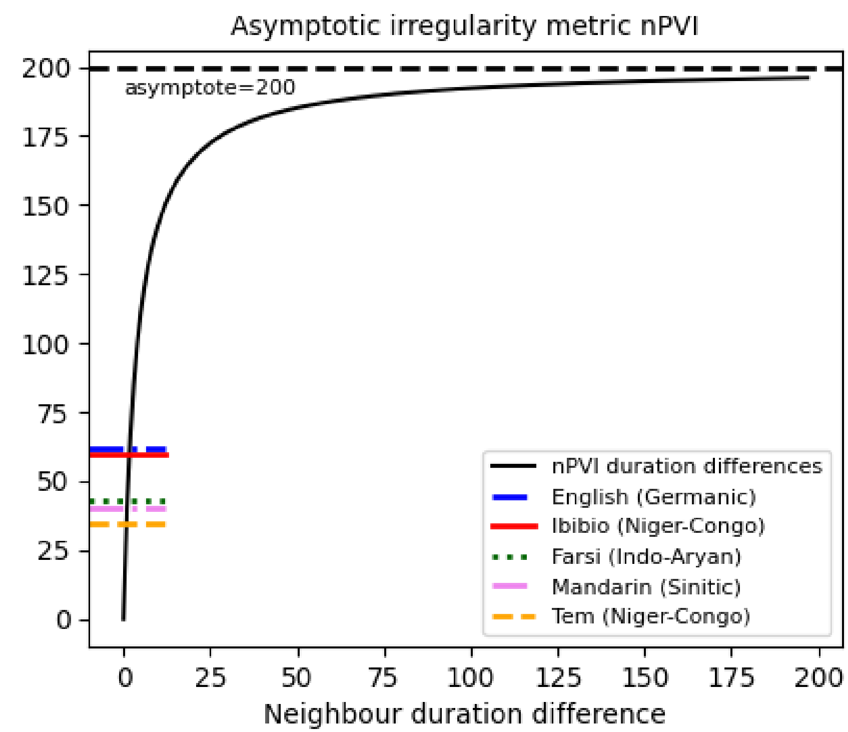
$$rPVI(D) = \sum \left| d_k - d_{k+1} \right| / (n-1)$$

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

raw Pairwise Variability Index

normalised Pairwise Variability Index





Interval analysis and PVI – the search for isochrony

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

Similarity to Manhattan Distance

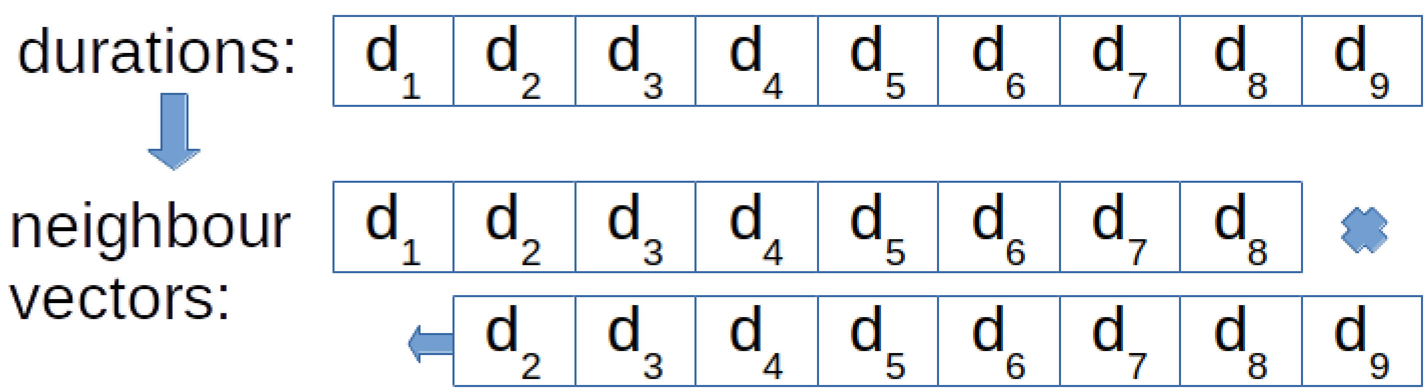
$$MD(x,y) = \sum_{i=1}^{n} \left|x_i - y_i\right|$$

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

Similarity to Canberra Distance
(Normalised Manhattan Distance)

$$NormMD(x,y) = \sum_{i=1}^{n} \frac{\left|x_i - y_i\right|}{\left|x_i\right| + \left|y_i\right|}$$

These measures show an overall 'next-door neighbour distance':



Assessment of interval duration measures

The interval duration measures can be useful heuristic measures.

They have the following properties:

1. the interval duration measures ignore the alternation property of rhythm by using absolute values, and yielding the same values for positive and negative differences;
2. they are often called ‘rhythm metrics’, but the interval duration measures calculate irregularity, not rhythmicality;
3. the ‘irregularity measures’ do not model rhythm and do not constitute either a model or a theory of rhythm;
4. through the annotation procedure the signal is filtered through the perceptual skills of an annotator and the signal is not analysed directly.

Back to Rhythm Formant Theory and Rhythm Formant Analysis

A signal processing approach

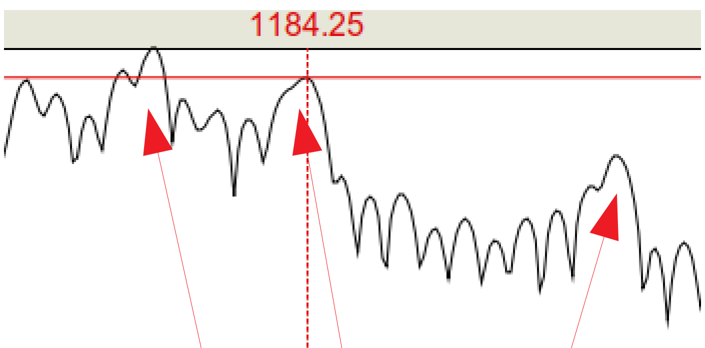
Rhythm Formants

1. Rhythm Formant Theory (RFT):
   1. A rhythm formant is a frequency zone of higher magnitude values in the normalised low frequency (LF) spectrum.
   2. Rhythm formants are detected in the LF AM spectrum and in the LF FM spectrum.
2. Rhythm Formant Analysis (RFA):
   1. The spectrum magnitudes are obtained by FFT and normalised to the range 0,…,1.
   2. A minimum magnitude (e.g. about 0.2) is defined as a cutoff level, below which values are clipped to zero.
   3. The clipped spectra of different recordings are compared using standard distance metrics and represented as distance maps, and hierarchically clustered using standard clustering criteria and represented as dendrograms.

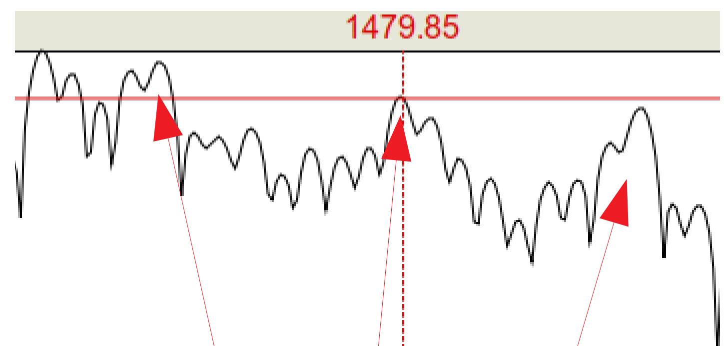
High Frequency *Formants (HF Formants)*

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

HF formant structures, f>300Hz signify vocal tract configurations.



[a] in “five”: 1st, 2nd, 3rd formants

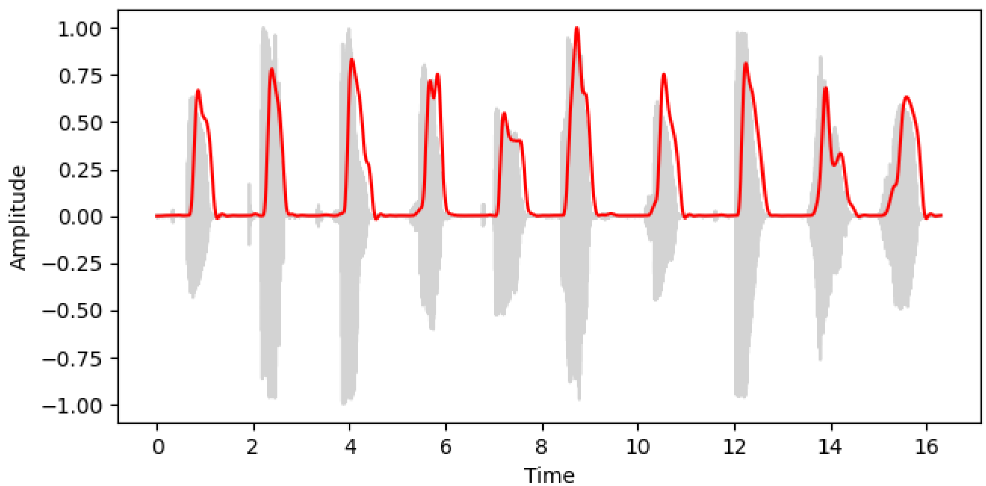


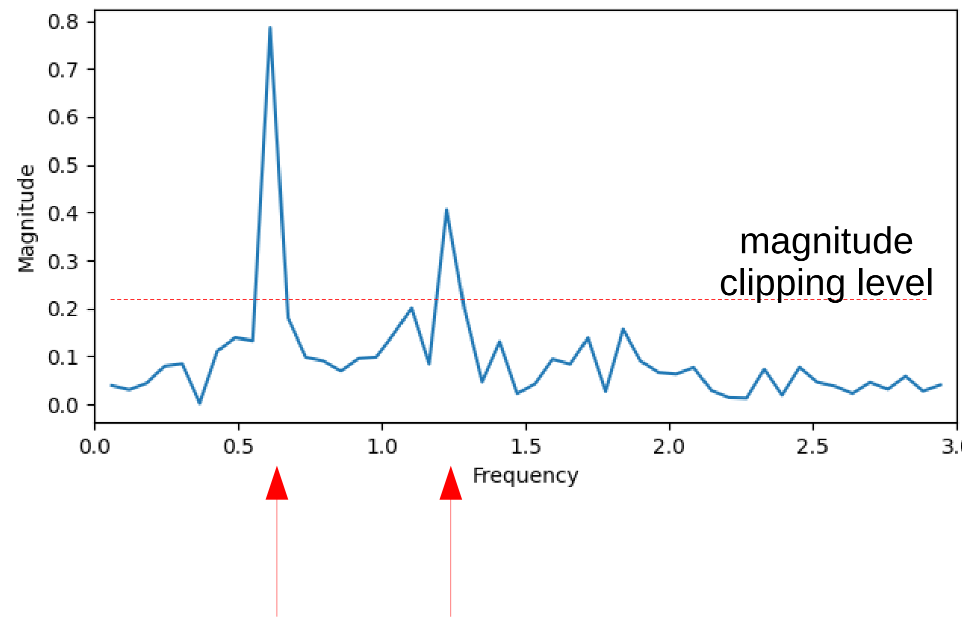
[i] in “five”: 1st, 2nd, 3rd formants

Low Frequency ***Formants (LF Formants)***

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

DATA/one-to-ten-Putonghua-Lara-16k-mono.wav, 16000



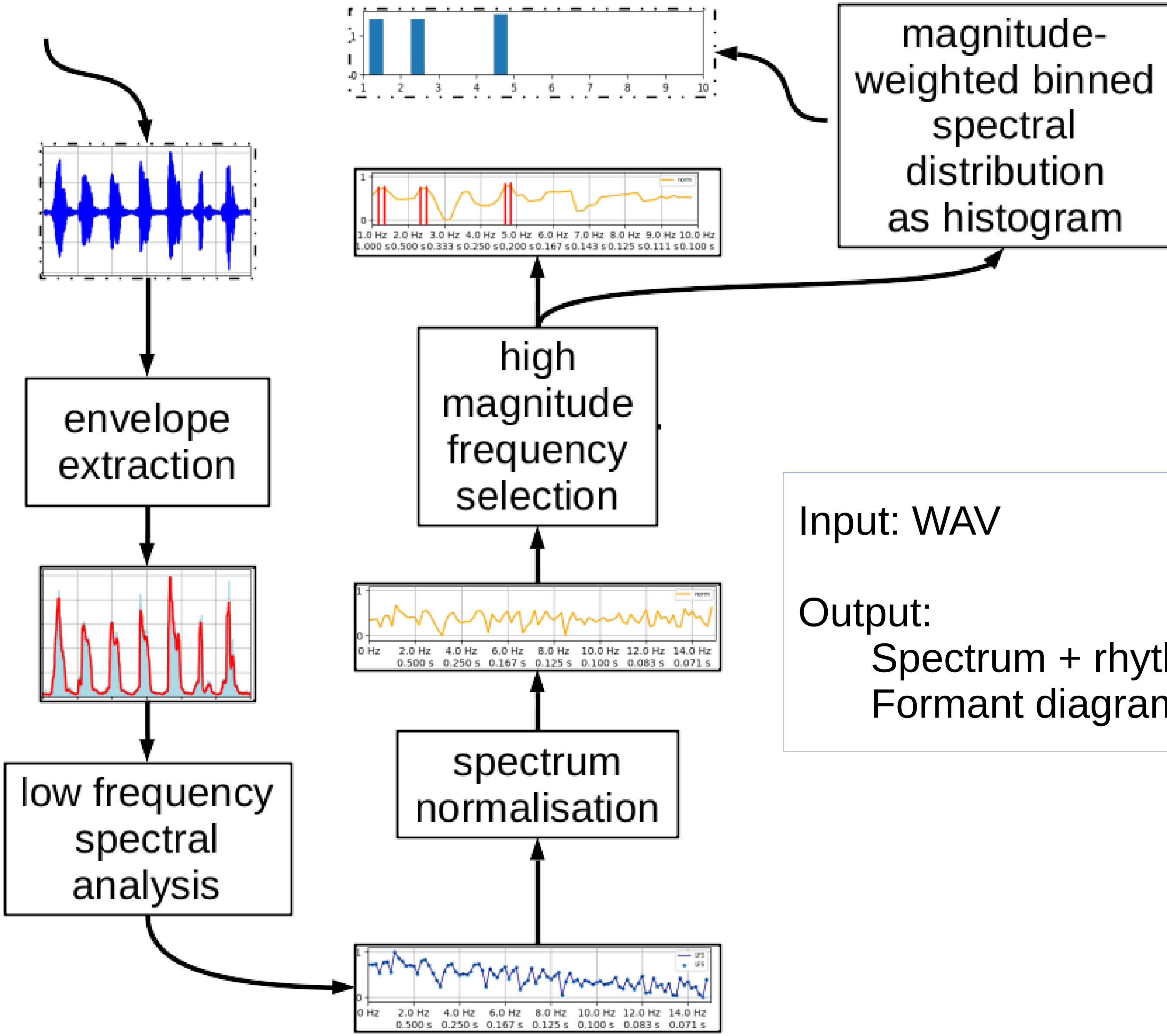


Frequency zones interpreted as rhythm formants

Queries:

1. What is the average length *T* of a syllable plus pause?
2. What kind of repetitive event has a rhythm formant at 0.6 Hz?
3. What kind of repetitive event has a rhythm formant at 1.3 Hz?
4. What is the syllable rate (frequency)?
5. What is the average syllable duration?

Overview of Rhythm Formant Analysis Dataflow



Input: WAV

Output:
- Spectrum + rhythm bars
- Formant diagram

The Modulation Code – practical research questions

How do we get information from the speech signal which is relevant for prosody?

- Amplitude Modulation (AM): demodulation procedure
- Frequency Modulation (FM): demodulation procedure (‘pitch tracking’)

How do we get rhythms out of the demodulated AM and FM?

- Low frequency spectral analysis
- Low frequency rhythm formant identification

How do we induce interesting information from the low frequency formants?

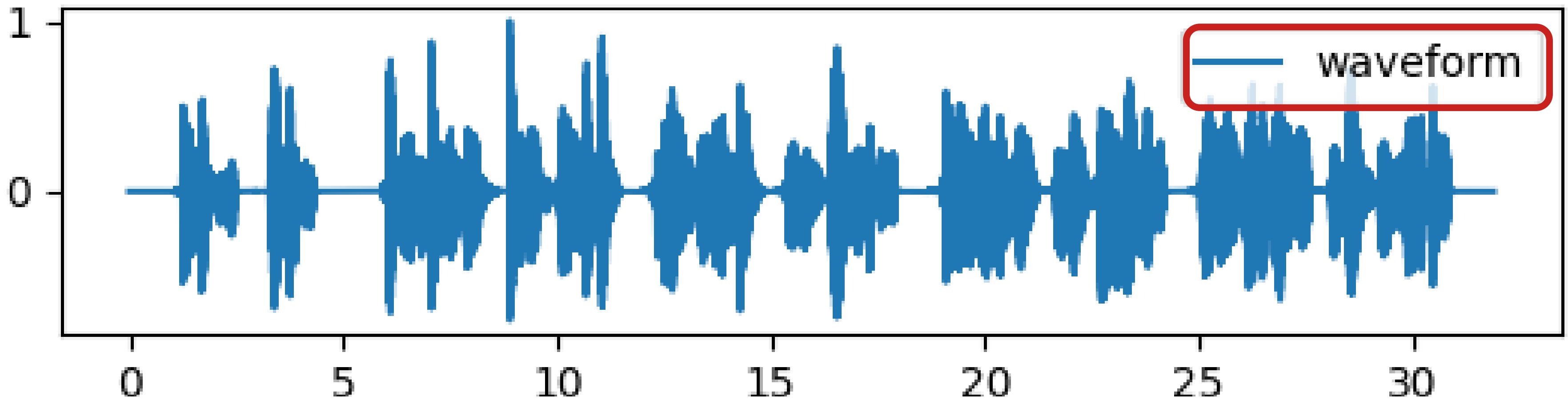
- Measure spectral differences between utterances, display as distance maps.
- Group different utterances into clusters, display as dendrograms (tree graphs).

Which tools do we use for this research schedule?

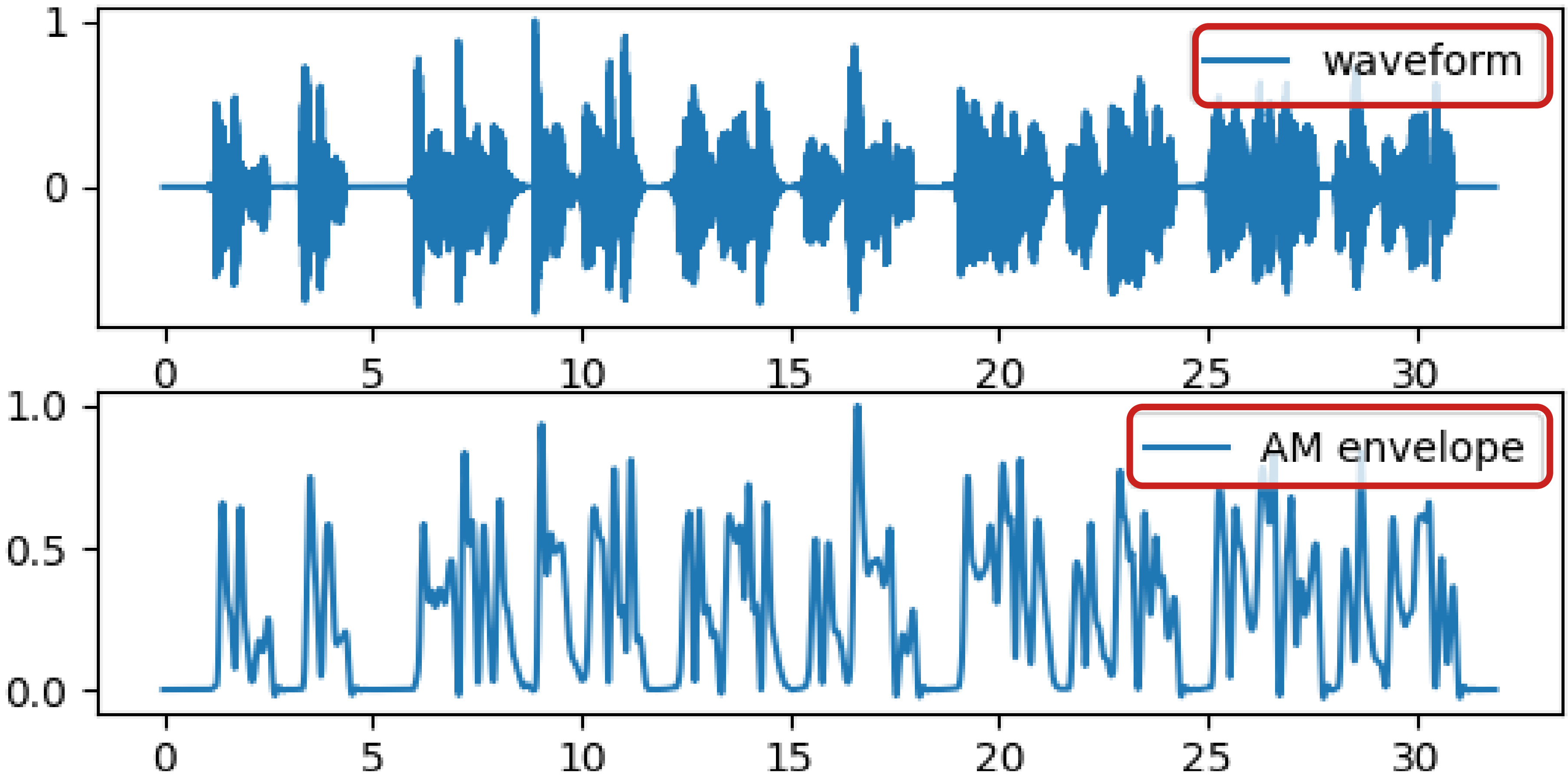
- The answer is the subject of the next lecture

Modelling rhythm as low frequencies (< 10 Hz)

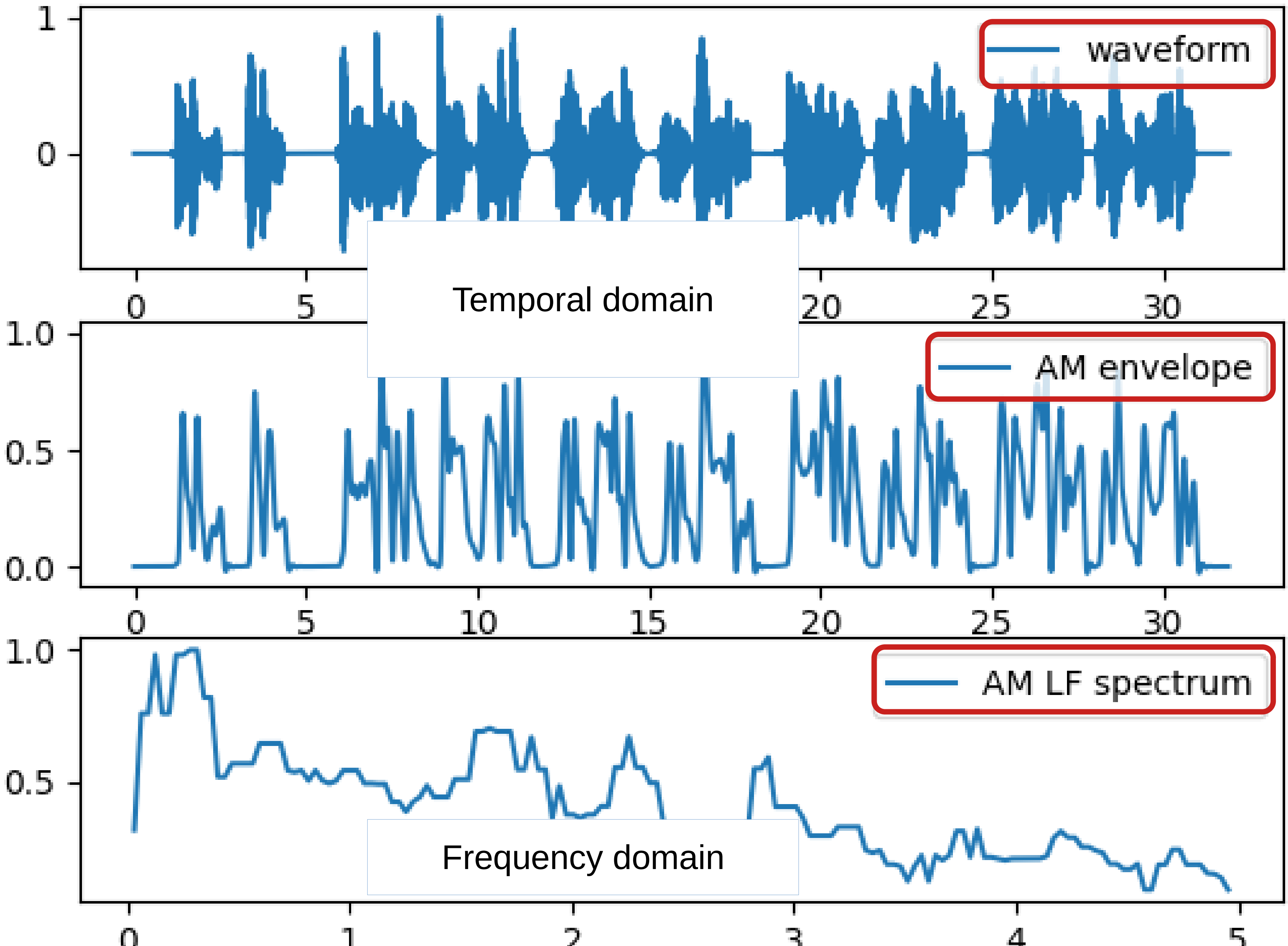
Case study 1: Tang dynasty poetry



Case study 1: Tang dynasty poetry



Tang dynasty poetry



Implementation

In case you think it’s hard, this is basically all it takes...

In case you think it's hard, this is all it takes...

Metadata, signal input, properties

```
# simpleacousticphonetics.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as np
import matplotlib.pyplot as plt
import scipy.io.wavfile as wave
from scipy.signal import butter, lfilter, medfilt

wavfilename = sys.argv[1]
fs, signal = wave.read(wavfilename)
siglen = len(signal)
sigsecs = siglen / fs
```

In case you think it's hard, this is all it takes...

Metadata, signal input, properties

```
# simpleacousticphonetics.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as np
import matplotlib.pyplot as plt
import scipy.io.wavfile as wave
from scipy.signal import butter, lfilter, medfilt

wavfilename = sys.argv[1]
fs, signal = wave.read(wavfilename)
siglen = len(signal)
sigsecs = siglen / fs
```

AM envelope extraction

```
b, a = butter(5, 5 / (0.5 * fs), btype="low")
envelope = lfilter(b, a, abs(signal))
```

In case you think it's hard, this is all it takes...

Metadata, signal input, properties

```
# simpleacousticphonetics.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as np
import matplotlib.pyplot as plt
import scipy.io.wavfile as wave
from scipy.signal import butter, lfilter, medfilt

wavfilename = sys.argv[1]
fs, signal = wave.read(wavfilename)
siglen = len(signal)
sigsecs = siglen / fs
```

AM envelope extraction

```
b, a = butter(5, 5 / (0.5 * fs), btype="low")
envelope = lfilter(b, a, abs(signal))
```

AM spectral transformation

```
magn = np.abs(np.fft.rfft(envelope))
freq = np.linspace(0,fs/2,len(magn))
lfspectrumlen = int(len(magn) * 5 / freq[-1])
magn = magn[1:lfspectrumlen]
freq = freq[1:lfspectrumlen]
magn = medfilt(magn, 7)
```

In case you think it's hard, this is all it takes...

Metadata, signal input, properties

```
# simpleacousticphonetics.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as np
import matplotlib.pyplot as plt
import scipy.io.wavfile as wave
from scipy.signal import butter, lfilter, medfilt

wavfilename = sys.argv[1]
fs, signal = wave.read(wavfilename)
siglen = len(signal)
sigsecs = siglen / fs
```

AM envelope extraction

```
b, a = butter(5, 5 / (0.5 * fs), btype="low")
envelope = lfilter(b, a, abs(signal))
```

AM spectral transformation

```
magn = np.abs(np.fft.rfft(envelope))
freq = np.linspace(0,fs/2,len(magn))
lfspectrumlen = int(len(magn) * 5 / freq[-1])
magn = magn[1:lfspectrumlen]
freq = freq[1:lfspectrumlen]
magn = medfilt(magn, 7)
```

Normalisation for display

```
signal = signal / max(signal)
envelope = envelope / max(envelope)
magn = magn / max(magn)
```

In case you think it’s hard, this is all it takes...

Metadata, signal input, properties

```
# simpleacousticphonetics.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as np
import matplotlib.pyplot as plt
import scipy.io.wavfile as wave
from scipy.signal import butter, lfilter, medfilt

wavfilename = sys.argv[1]
fs, signal = wave.read(wavfilename)
siglen = len(signal)
sigsecs = siglen / fs
```

AM envelope extraction

```
b, a = butter(5, 5 / (0.5 * fs), btype="low")
envelope = lfilter(b, a, abs(signal))
```

AM spectral transformation

```
magn = np.abs(np.fft.rfft(envelope))
freq = np.linspace(0,fs/2,len(magn))
lfspectrumlen = int(len(magn) * 5 / freq[-1])
magn = magn[1:lfspectrumlen]
freq = freq[1:lfspectrumlen]
magn = medfilt(magn, 7)
```

Normalisation for display

```
signal = signal / max(signal)
envelope = envelope / max(envelope)
magn = magn / max(magn)
```

Figure creation

```
x = np.linspace(0, sigsecs, siglen)

plt.subplot(3,1,1)
plt.plot(x, signal, label="waveform")
plt.legend(loc="upper right")

plt.subplot(3,1,2)
plt.plot(x, envelope, label="AM envelope")
plt.legend(loc="upper right")

plt.subplot(3,1,3)     # spectrum
plt.plot(freq, magn, label="AM LF
spectrum")
plt.legend(loc="upper right")
```

In case you think it's hard, this is all it takes...

Metadata, signal input, properties

```
# simpleacousticphonetics.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as np
import matplotlib.pyplot as plt
import scipy.io.wavfile as wave
from scipy.signal import butter, lfilter, medfilt

wavfilename = sys.argv[1]
fs, signal = wave.read(wavfilename)
siglen = len(signal)
sigsecs = siglen / fs
```

AM envelope extraction

```
b, a = butter(5, 5 / (0.5 * fs), btype="low")
envelope = lfilter(b, a, abs(signal))
```

AM spectral transformation

```
magn = np.abs(np.fft.rfft(envelope))
freq = np.linspace(0,fs/2,len(magn))
lfspectrumlen = int(len(magn) * 5 / freq[-1])
magn = magn[1:lfspectrumlen]
freq = freq[1:lfspectrumlen]
magn = medfilt(magn, 7)
```

Normalisation for display

```
signal = signal / max(signal)
envelope = envelope / max(envelope)
magn = magn / max(magn)
```

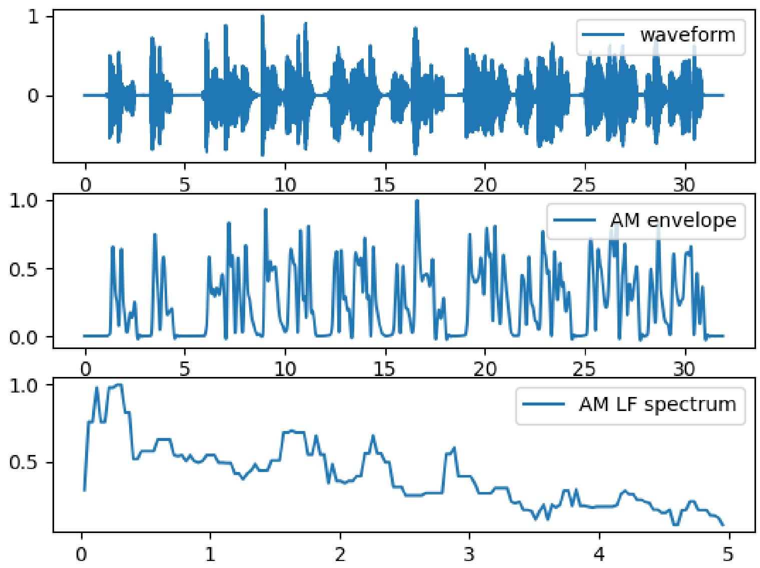
Figure creation

```
x = np.linspace(0, sigsecs, siglen)

plt.subplot(3,1,1)
plt.plot(x, signal, label="waveform")
plt.legend(loc="upper right")

plt.subplot(3,1,2)
plt.plot(x, envelope, label="AM envelope")
plt.legend(loc="upper right")

plt.subplot(3,1,3)     # spectrum
plt.plot(freq, magn, label="AM LF
spectrum")
plt.legend(loc="upper right")
```

Figure storage and display

```
plt.savefig(wavfilename+".png")
plt.show()
```

Amplitude and Frequency demodulation



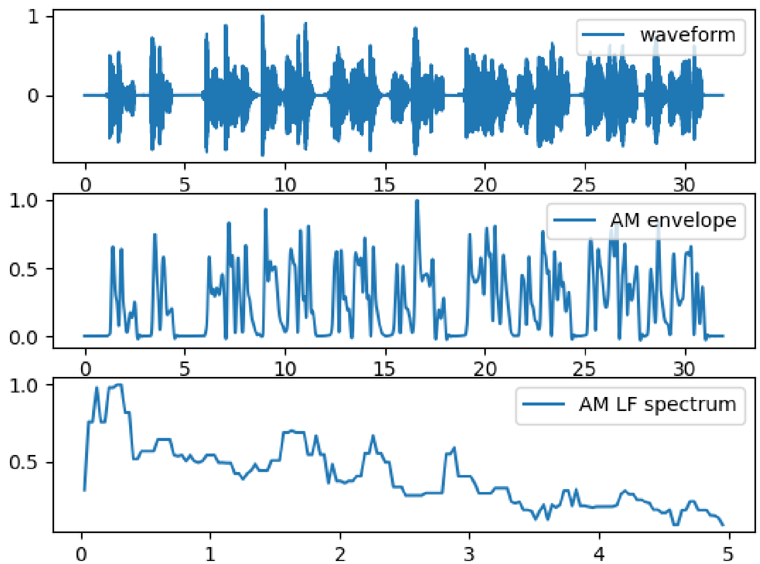
Modulated signal: AM and FM

Demodulated AM signal

Low frequency spectrum of demodulated AM signal

Frequency demodulation (fundamental frequency tracking, F0 estimation) and analysis

Amplitude and Frequency demodulation

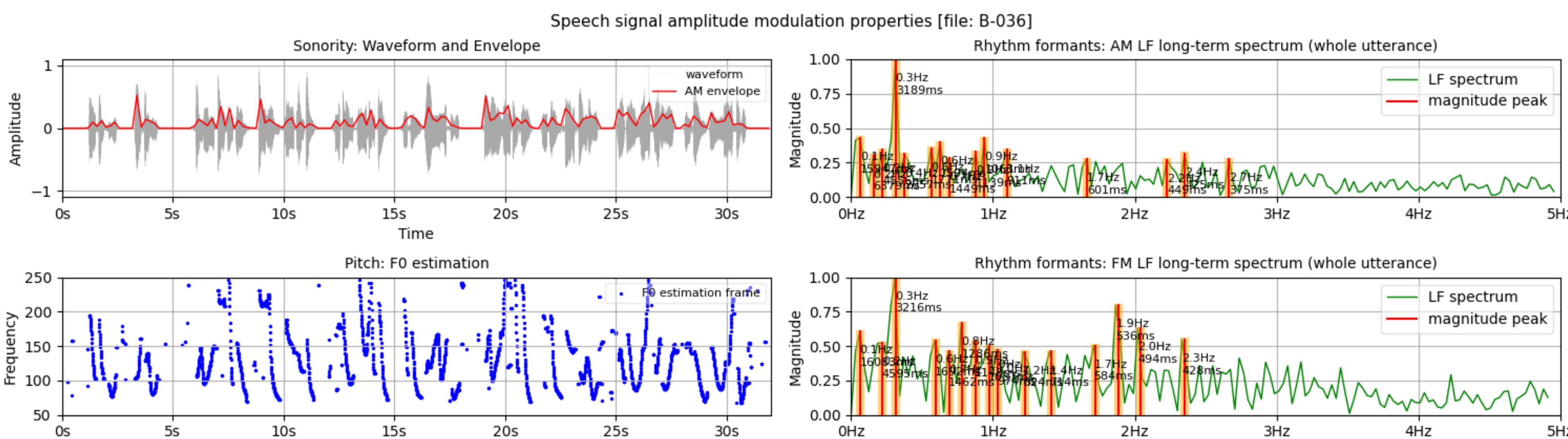


Modulated signal: AM and FM

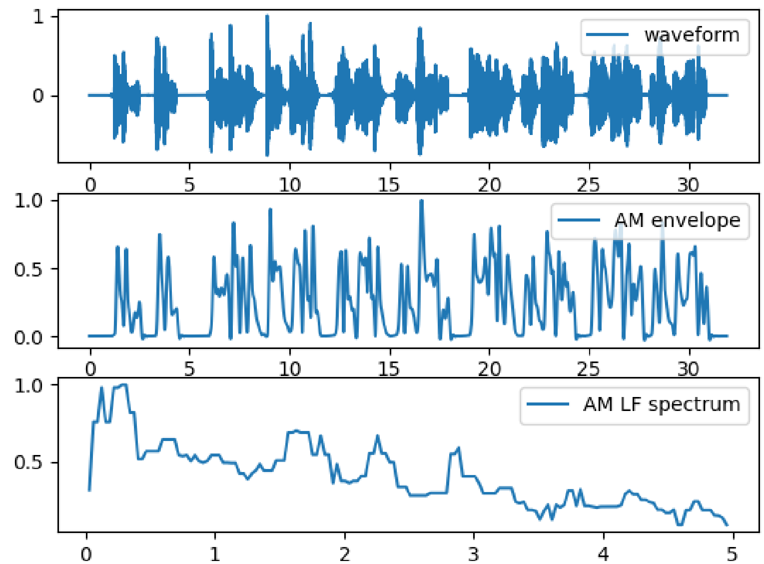
Demodulated AM signal

Analysis: low frequency spectrum of demodulated AM signal

Frequency demodulation (fundamental frequency tracking, F0 estimation) and analysis



Amplitude and Frequency demodulation

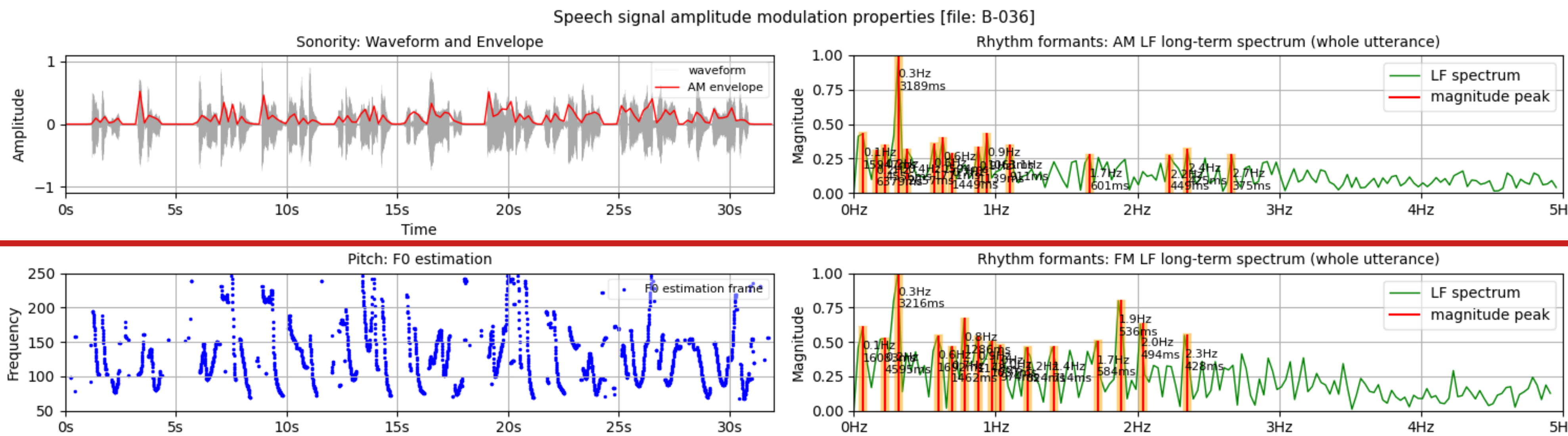


Modulated signal: AM and FM

Demodulated AM signal

Low frequency spectrum of demodulated AM signal

Frequency demodulation (fundamental frequency tracking, F0 estimation) and analysis



Moving from the frequency domain to the temporal domain

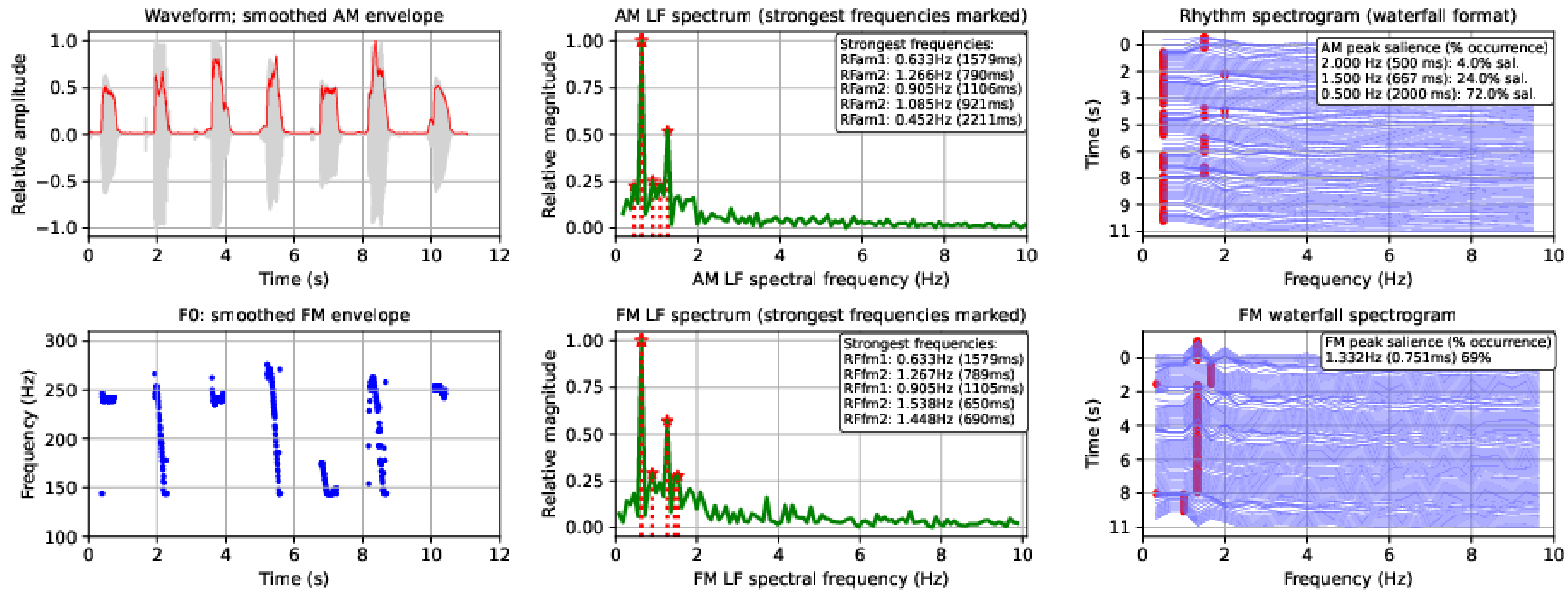
Temporal variability in the rhythm spectrum

The rhythm spectrogram and ‘rhythms of rhythm’:

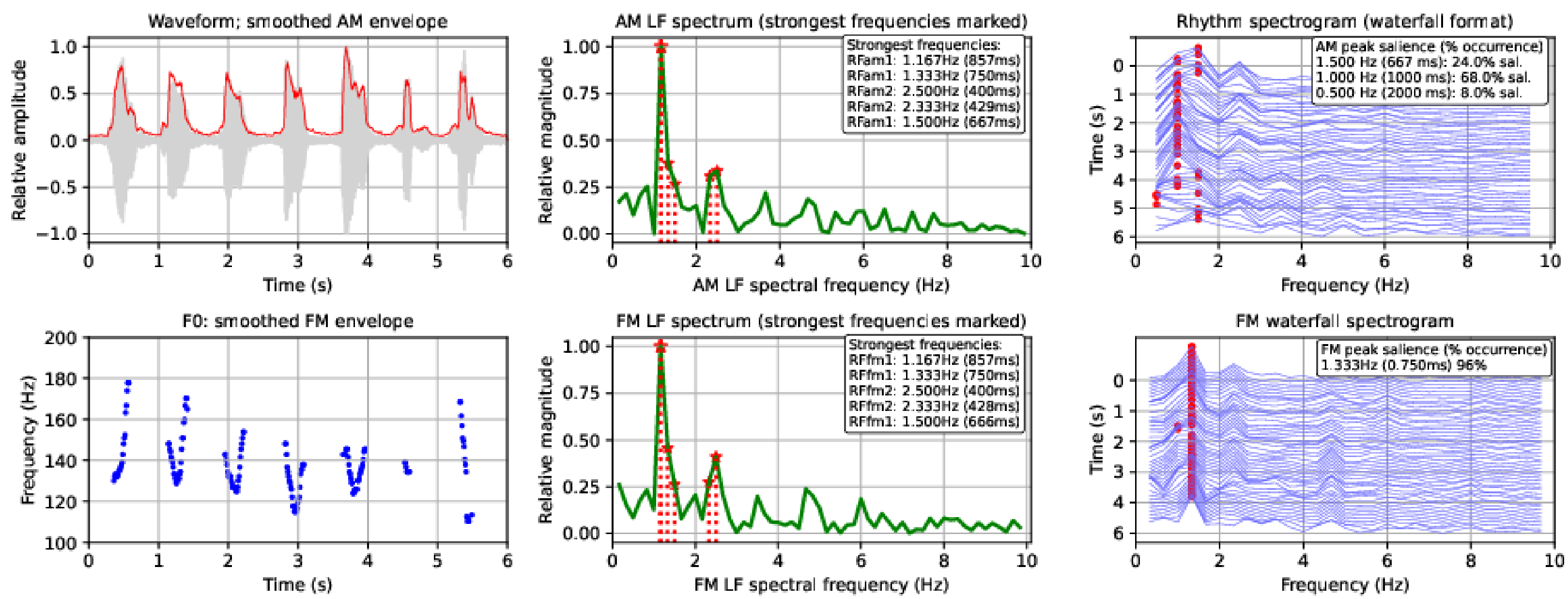
waterfall format
heatmap format

Case Study 2: Counting one to seven

Pǔtōnghuà:



English:



Rhythm and the morphophonology of counting

Case Study 2: Counting one to thirty (English)

Null hypothesis:

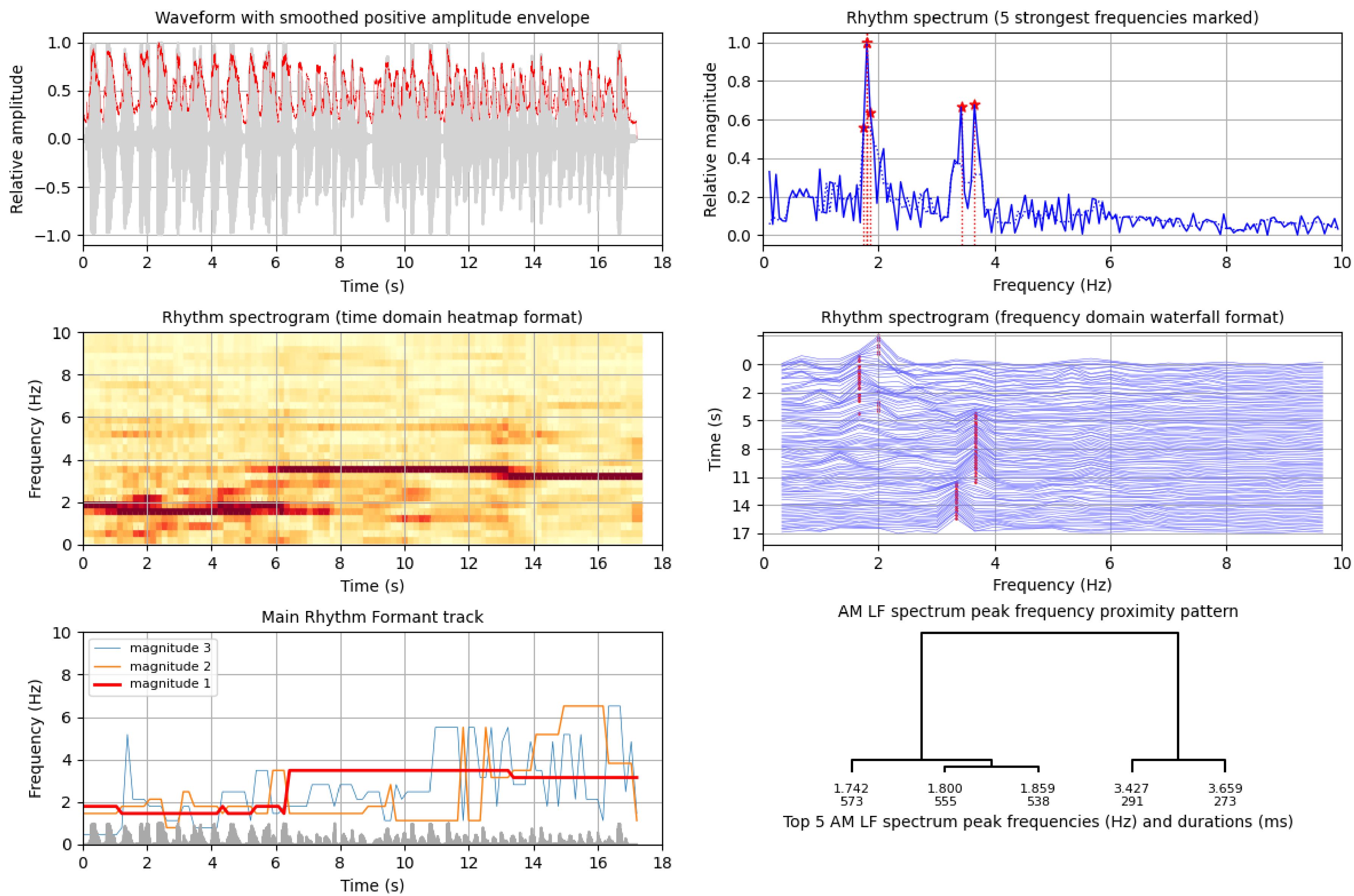
no rhythmic differences in counting from one to thirty

Alternative hypothesis:

rhythmic differences, depending on

syllable count in words due to morphological structure

Case Study 2: Counting one to thirty (English)



Case Study 2: Counting one to thirty (English)

Null hypothesis:

- no rhythmic differences in counting from one to thirty

Alternative hypothesis:

- rhythmic differences
  - depending on syllable count in words
  - due to morphological structure

Result:

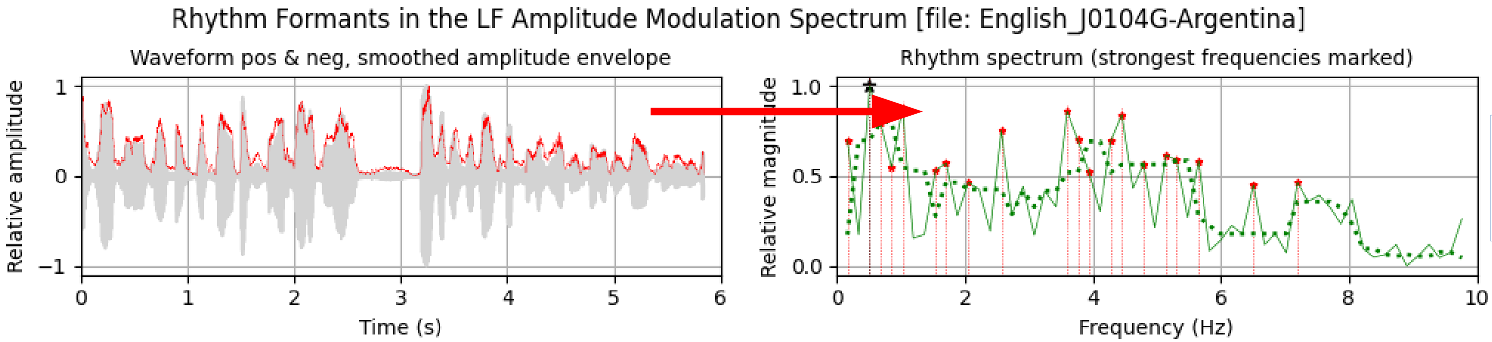
- 1 ... 10, mainly monosyllables, lower frequency rhythm formant
  - from 0.187s to 5.482s
  - frequencies 1.742 Hz, 1.8 Hz, 1.859 Hz;
- 11 ... 30, mainly polysyllables, higher frequency rhythm formant
  - from 5.482s to 16.978s, frequency 3.659 Hz
  - at 12.5 s (in “twenty-three”), frequency 3.427 Hz.

Case study 3:

Rhythms in dialogue

Case study 3: variable rhythms in dialogue

Rhythms in dialogue – a radio interview about football

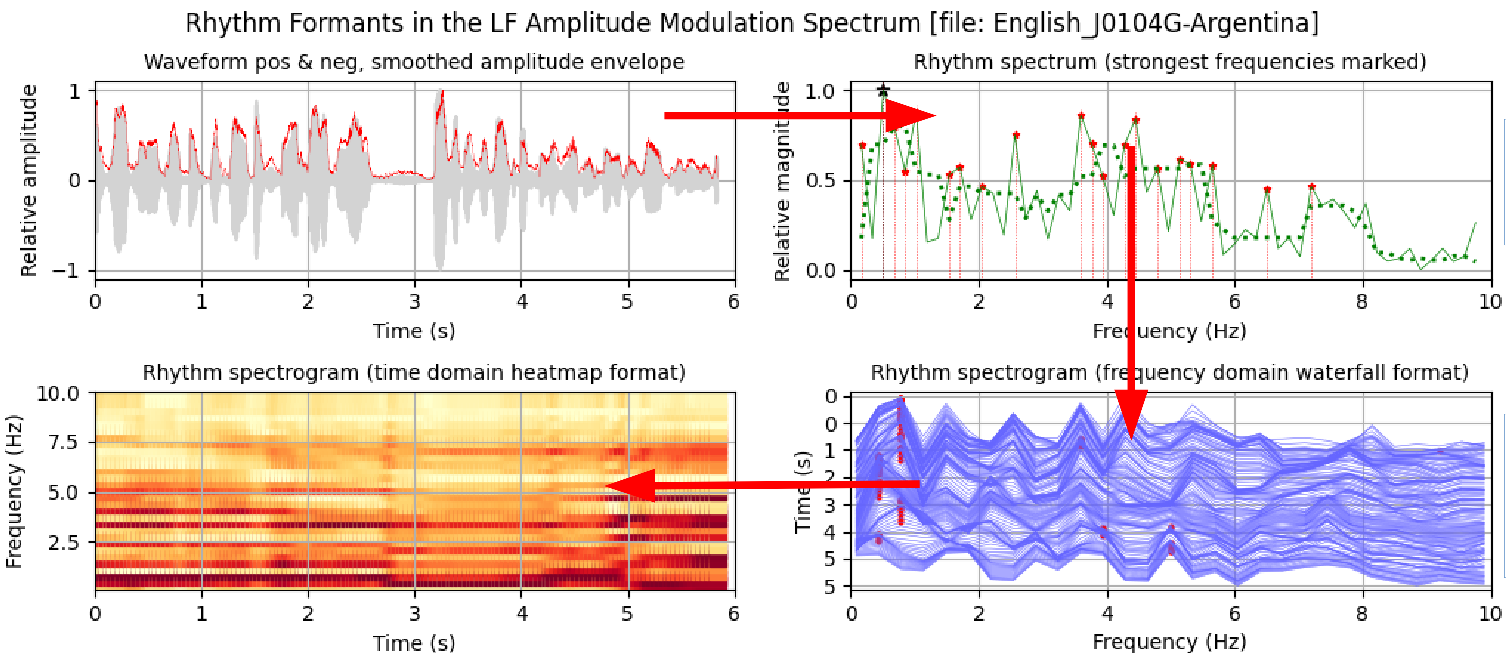


L: *waveform*
R: *long-term spectrum*

Method: Low Frequency Spectral Analysis to identify syllable, word, phrase rhythms

Case study 3: variable rhythms in dialogue

A radio interview about football: rhythms



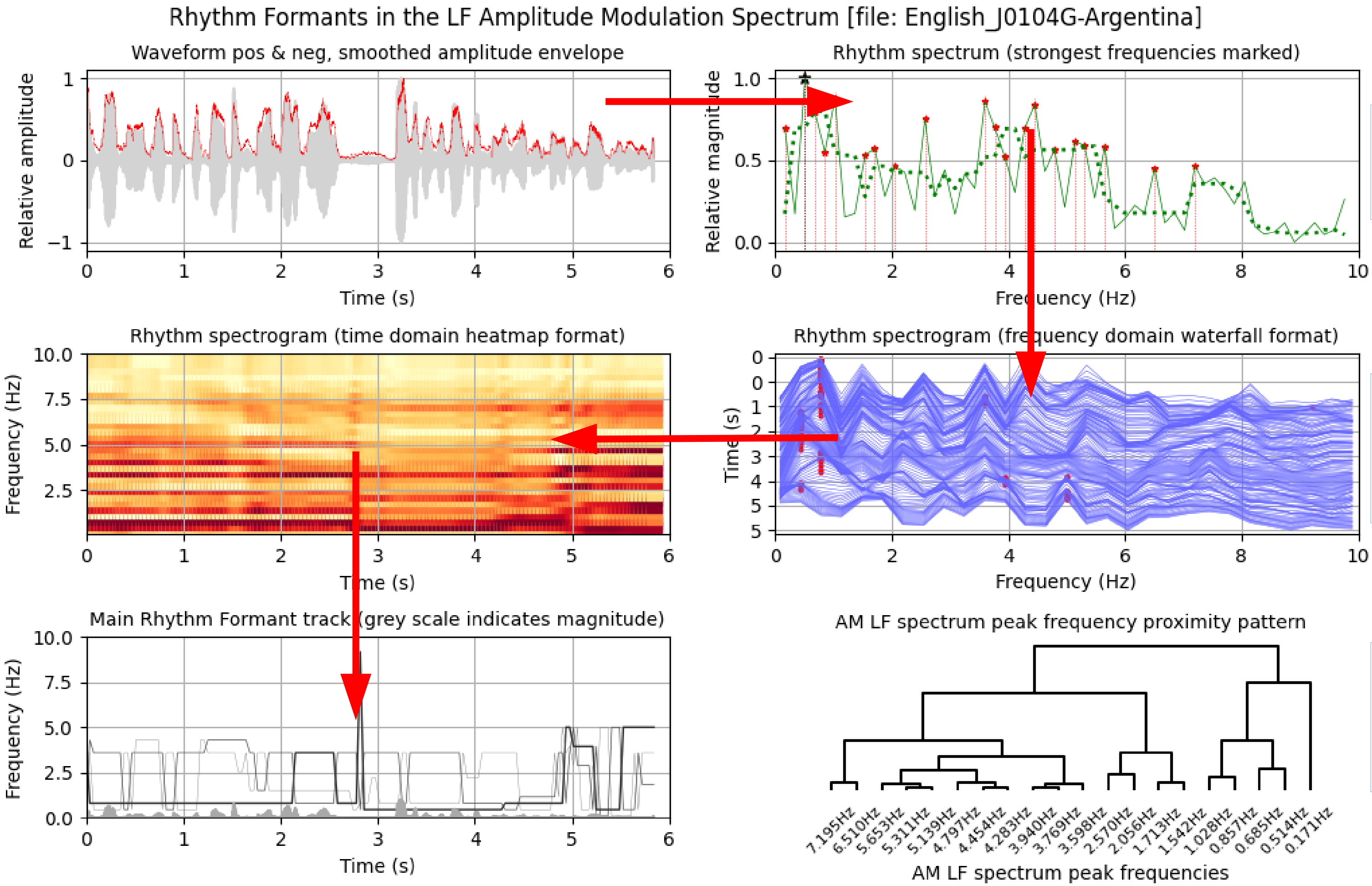
L: *waveform*
R: *long-term spectrum*

Long-term spectrogram:
L: *heatmap format*
R: *waterfall format*

Method: Low Frequency Spectral Analysis to identify syllable, word, phrase rhythms

Case study 3: rhythms in dialogue

Rhythms in dialogue – a radio interview about football



L: *waveform*
R: *long-term spectrum*

Long-term spectrogram:
L: *heatmap format*
R: *waterfall format*

Rhythms:
L: *main rhythms*
R: *rhythm clusters*

Method: Low Frequency Spectral Analysis to identify syllable, word, phrase rhythms

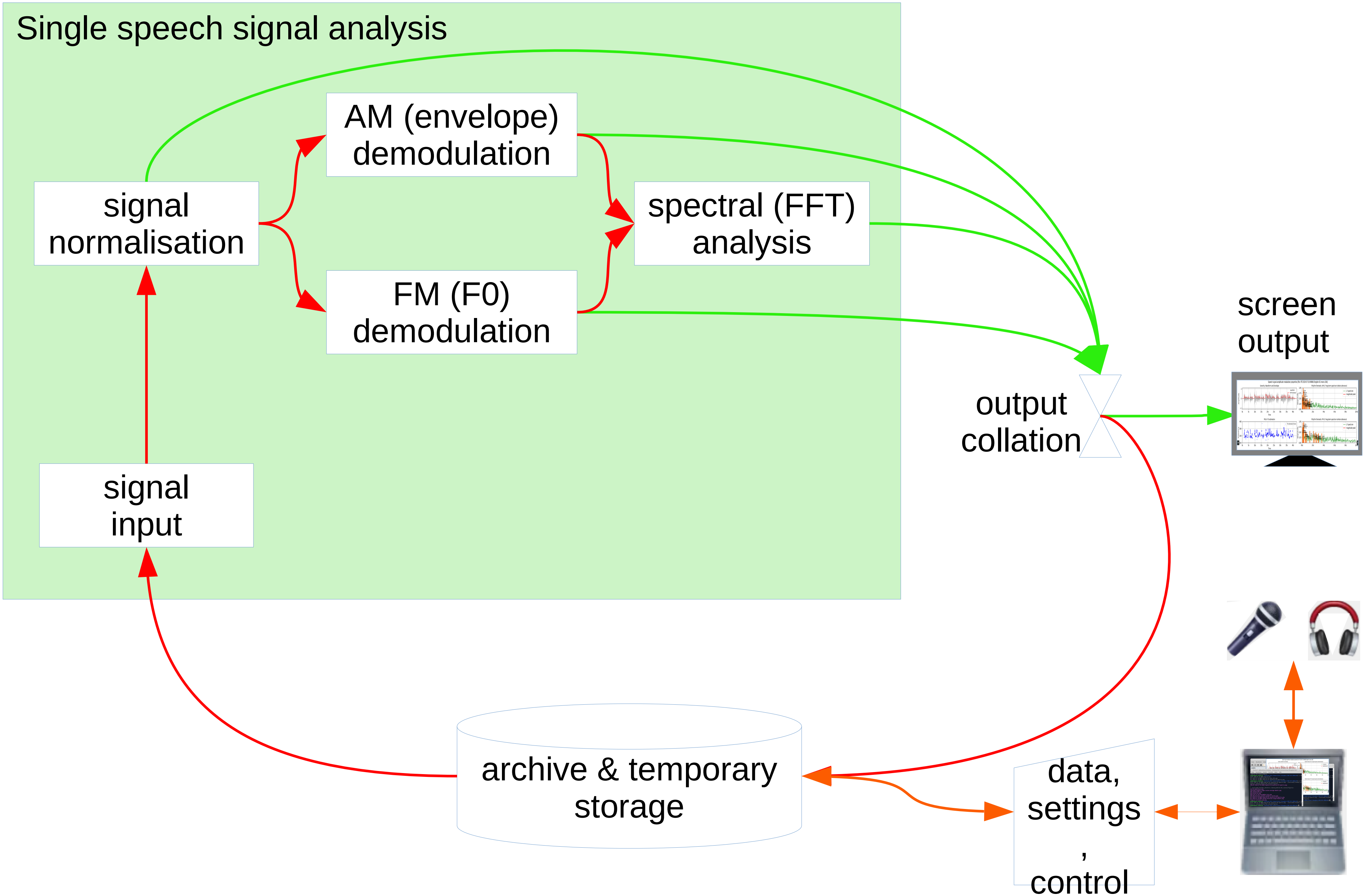
Case Study 4:

Rhythms in story-reading

An English example:
The North Wind and the Sun

A German example:
Nordwind und Sonne

Data flow – single files



Case study 4: rhythms in story-telling

1. Bilingual reader, English and German
2. Story-reading

    “The North Wind and the Sun”

    Nordwind und Sonne

3. Three readings in each language
4. Instruction:

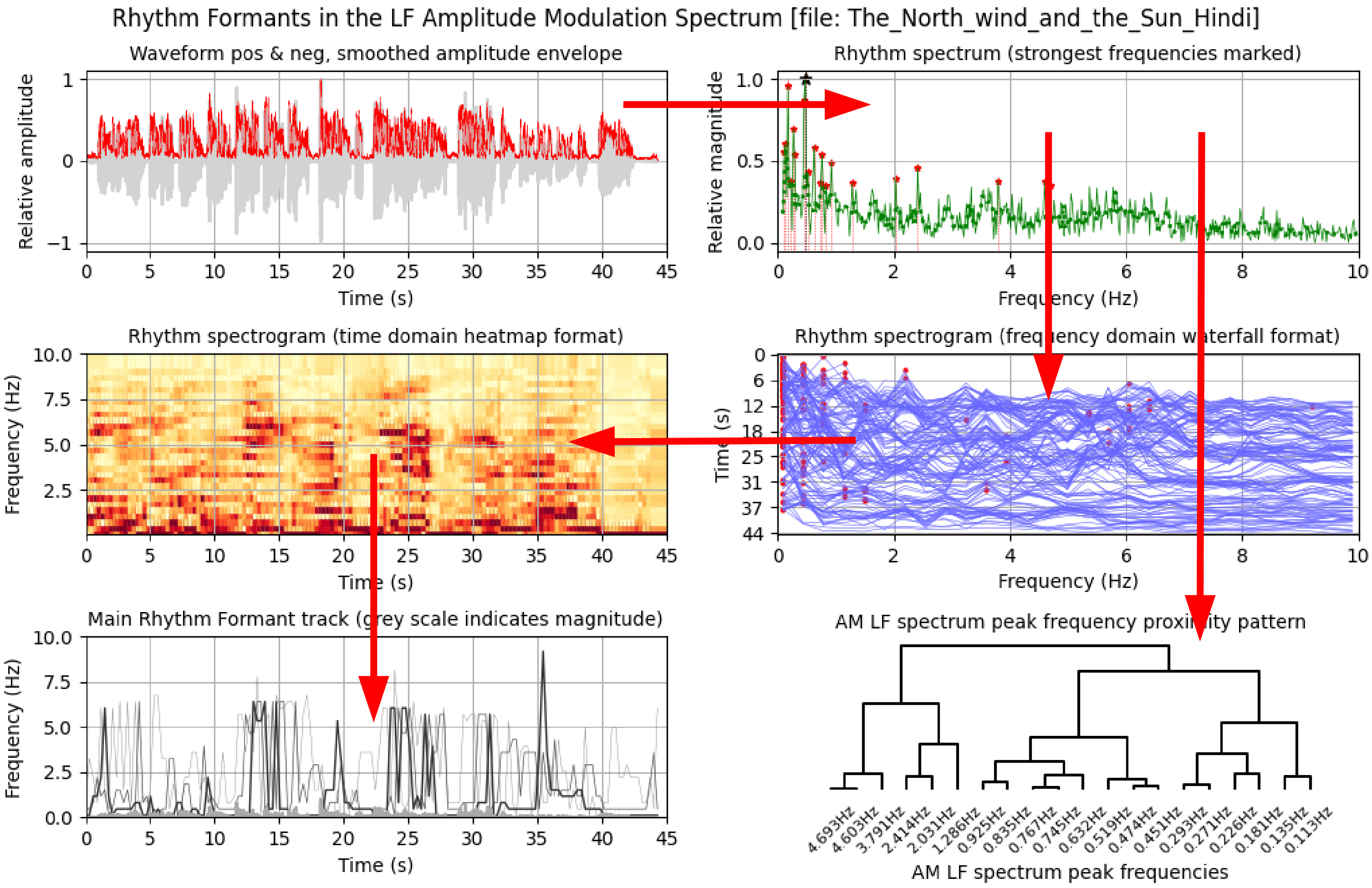
    As if reading a story to a child.

Null hypothesis:

1. No difference between readings
2. Alternative hypothesis:
    1. No difference between readings in one language
    2. Difference between readings in different languages

Case study 3: rhythms in story-telling

Rhythms in a narrative – the North Wind and the Sun (German translation)



L: *waveform*
R: *long-term spectrum*

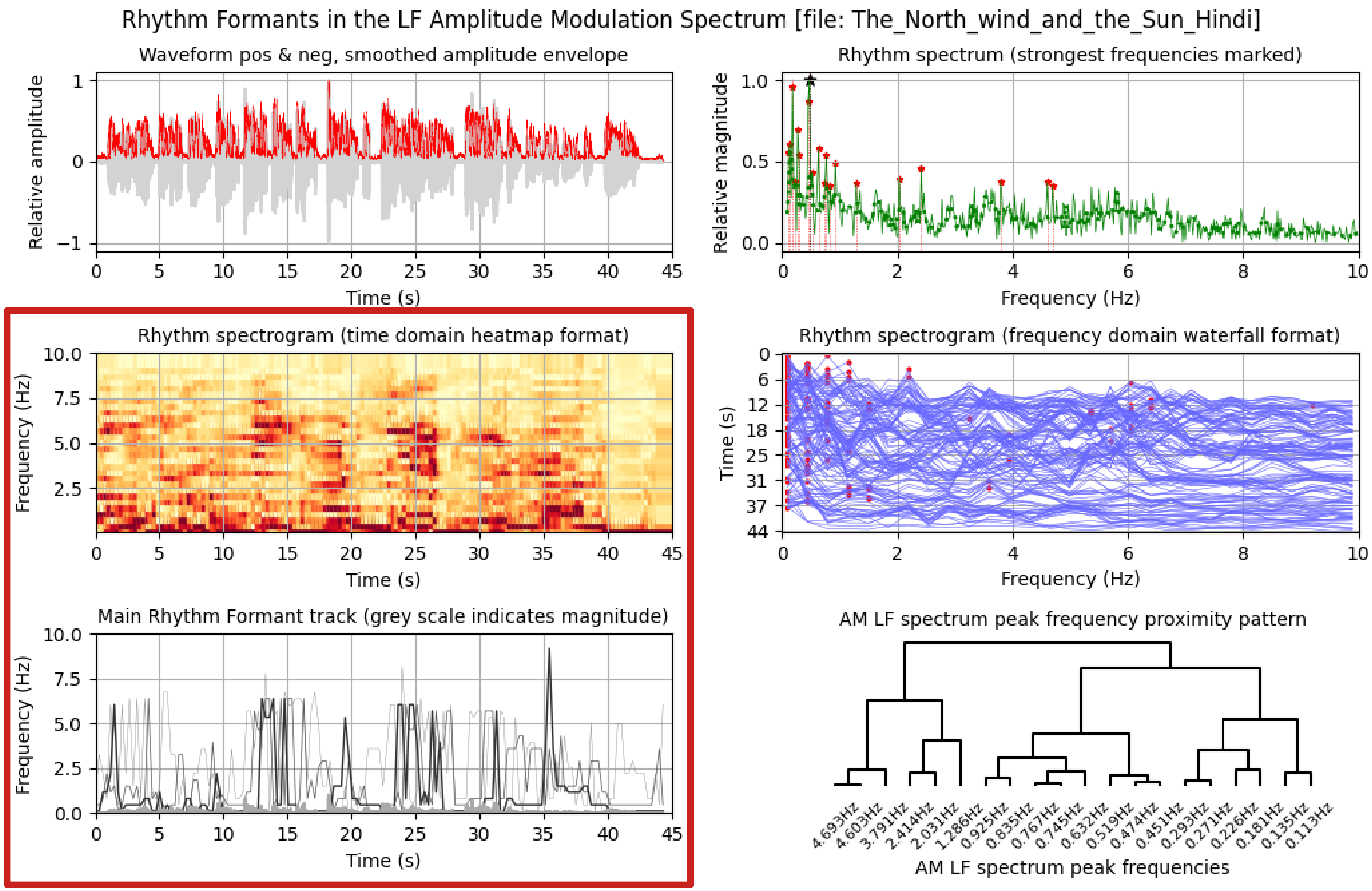
Long-term spectrogram:
L: *heatmap format*
R: *waterfall format*

Rhythms:
L: *main rhythms*
R: *rhythm clusters*

Method: Low Frequency Spectral Analysis to identify syllable, word, phrase rhythms

Case study 3: rhythms in story-telling

Rhythms in a narrative – the North Wind and the Sun (German translation)



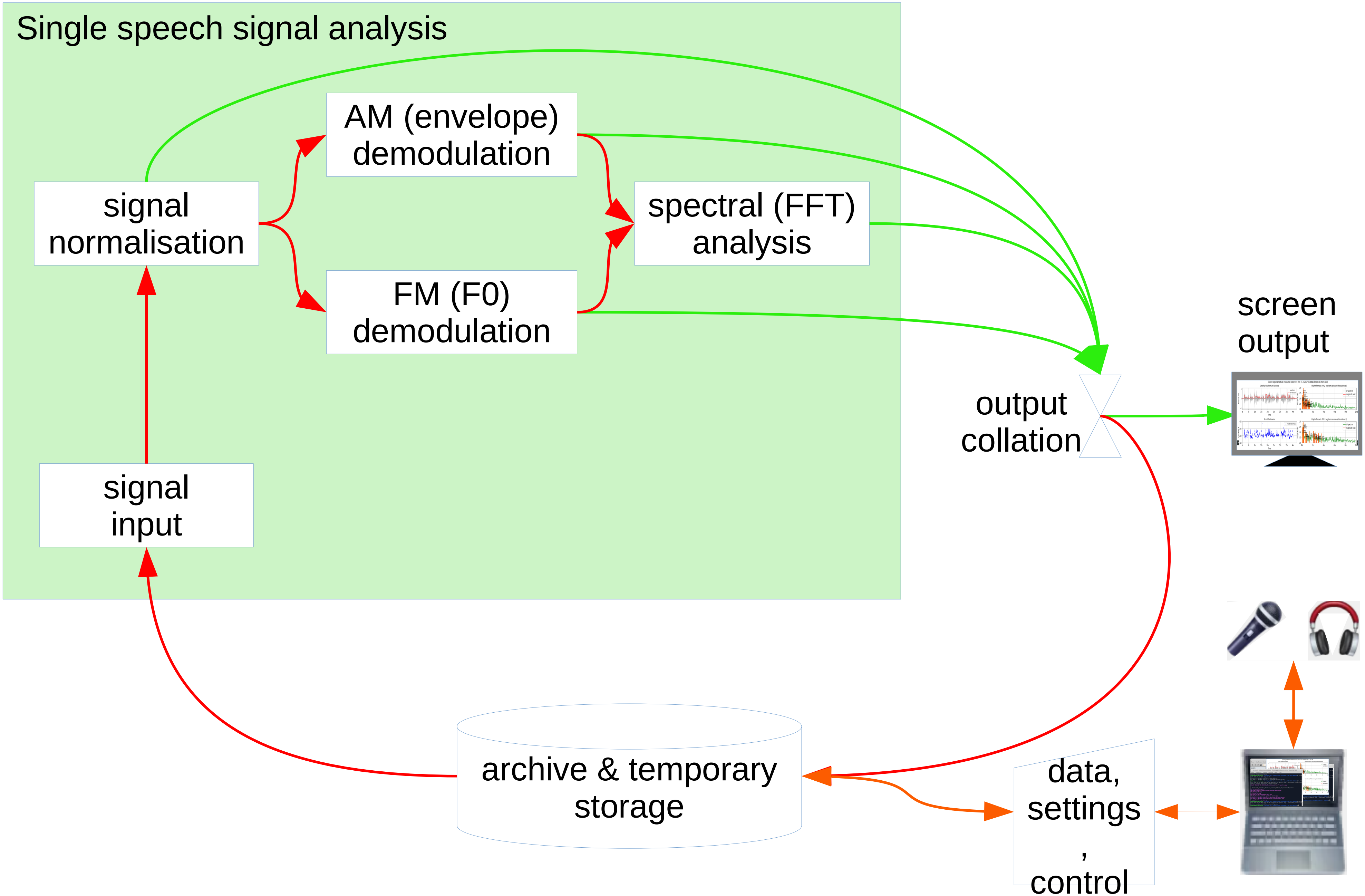
L: *waveform*
R: *long-term spectrum*

Long-term spectrogram:
L: *heatmap format*
R: *waterfall format*

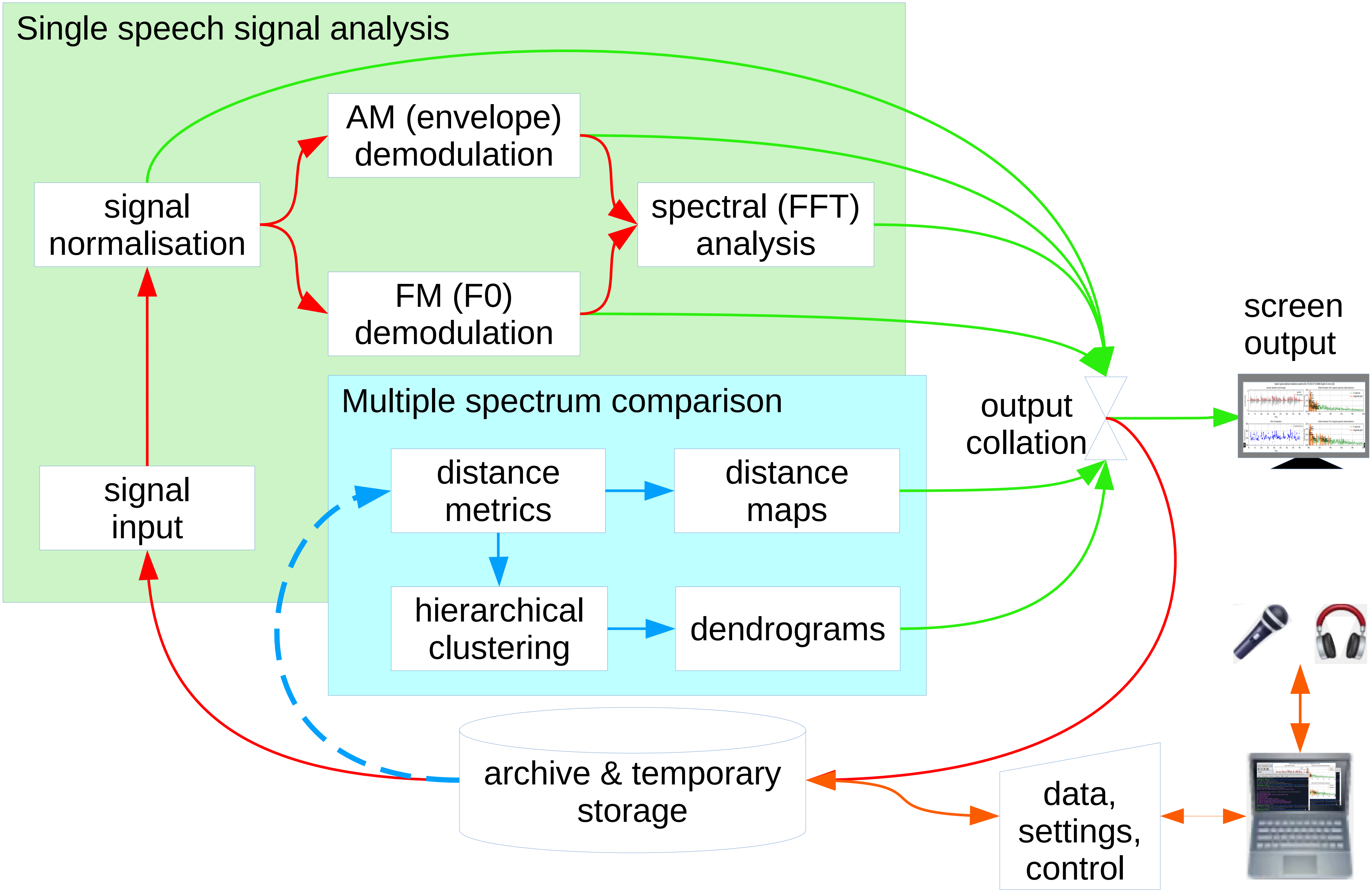
Rhythms:
L: *main rhythms*
R: *rhythm clusters*

Method: Low Frequency Spectral Analysis to identify syllable, word, phrase rhythms

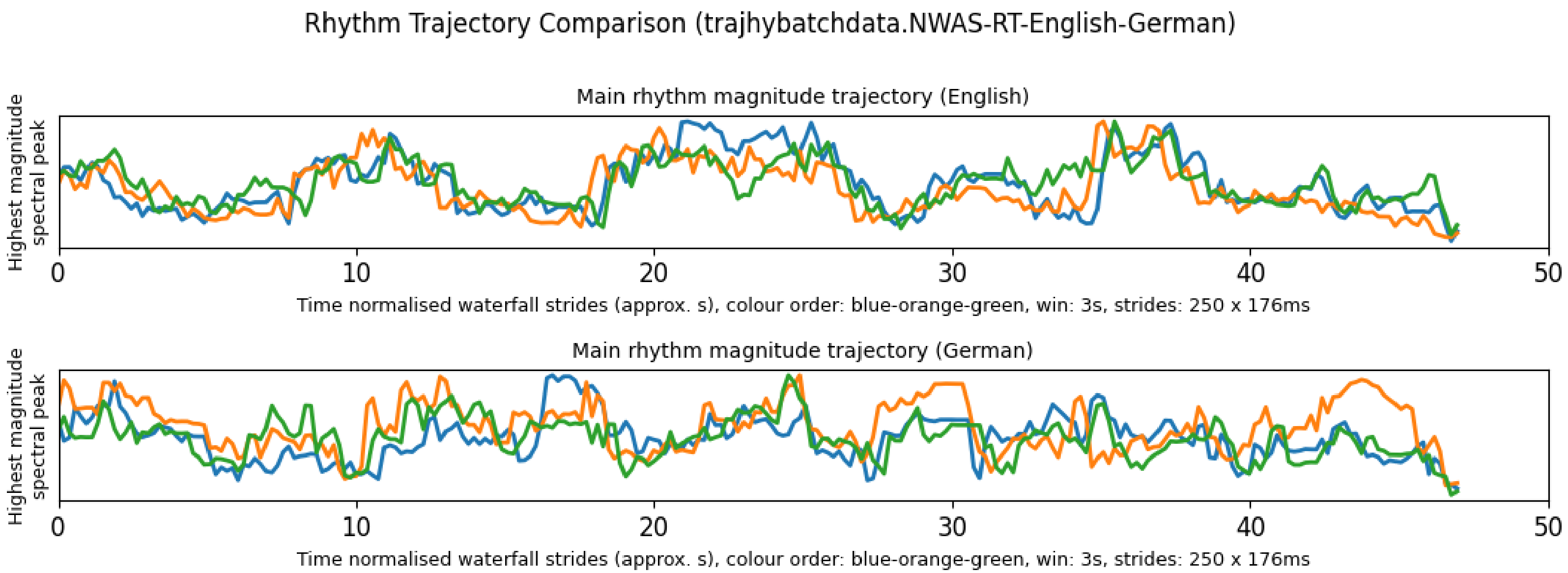
Data flow – single files



Data flow – comparison of multiple files



Case study 3: rhythms in story-telling



Results:

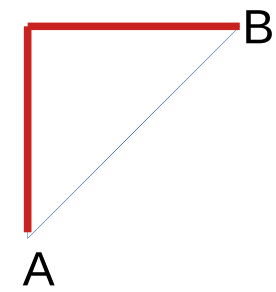
Top: English
Lower: German

All readings show some differences.
Readings in the same language correlate strongly
Readings in different languages do not correlate so strongly

Distance metrics

Manhattan Distance
(Cityblock distance, Taxicab Distance)
'around the corner'

$$\sum_{i=1}^{n} |x_i - y_i|$$

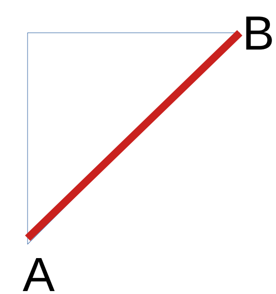


Canberra Distance
(Normalised Manhattan Distance)

$$\sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

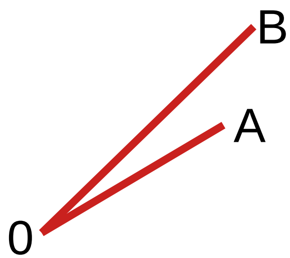
Euclidean Distance
direct distance
'as the crow flies'

$$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$



Cosine Distance
angle, direction, not magnitude
so not distance itself
'hiker's orientation'

$$\frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}$$



Spectrum Comparison: Distance Table

	Eng 01	Eng 02	Eng 03	Ger 01	Ger 02	Ger 03
Eng 01	0.	0.67477731	1.	0.74745837	0.93762055	0.85622088
Eng 02	0.67477731	0.	0.5184008	0.76221046	0.87568858	0.7706713
Eng 03	1.	0.5184008	0.	0.78197106	0.85094568	0.82617612
Ger 01	0.74745837	0.76221046	0.78197106	0.	0.42298678	0.56668163
Ger 02	0.93762055	0.87568858	0.85094568	0.42298678	0.	0.44727788
Ger 03	0.85622088	0.7706713	0.82617612	0.56668163	0.44727788	0.

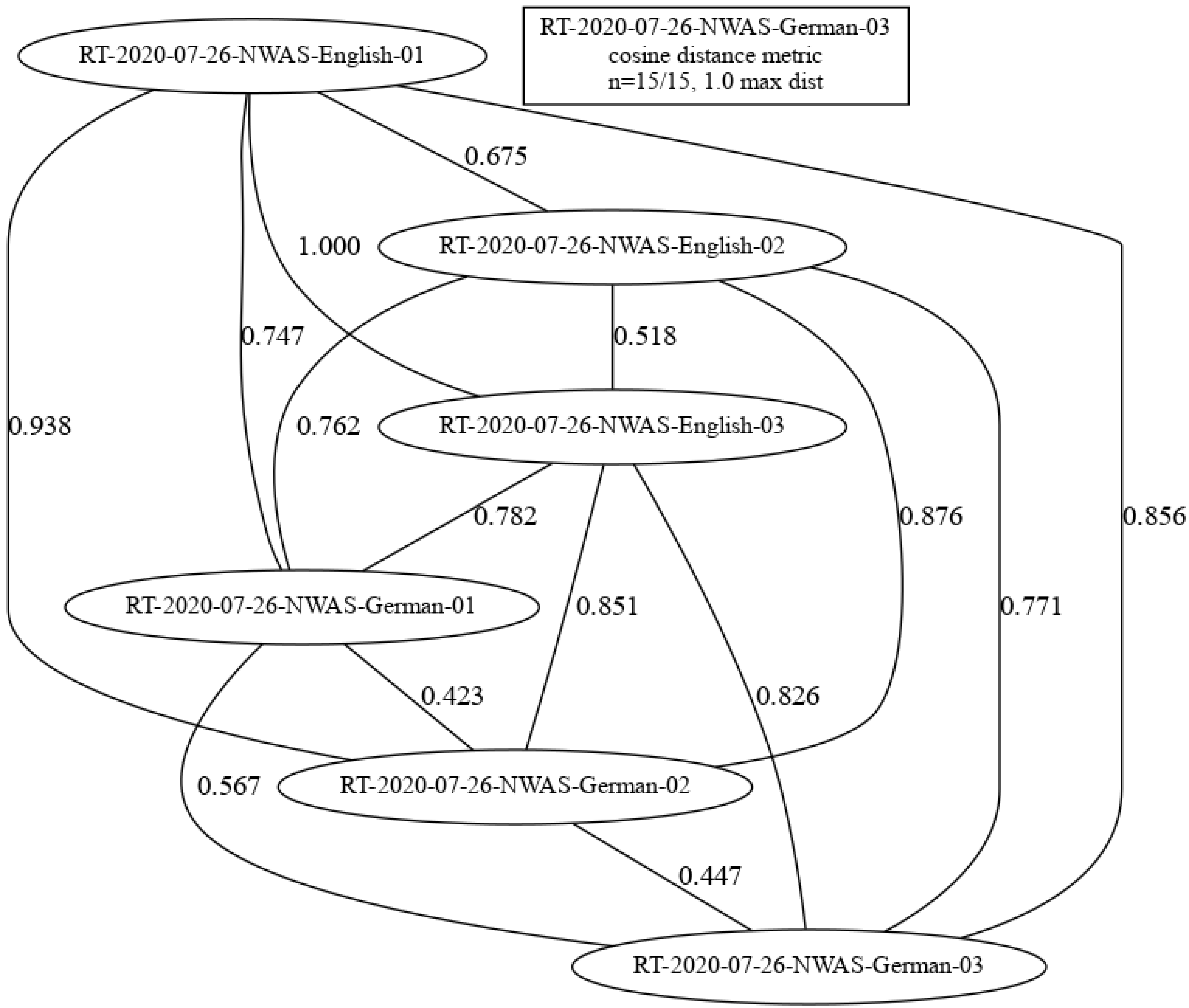
Adult Female English-German bilingual reading
The North Wind and the Sun,
3 English, 3 German, in order of production.

Spectrum Comparison: Distance Table

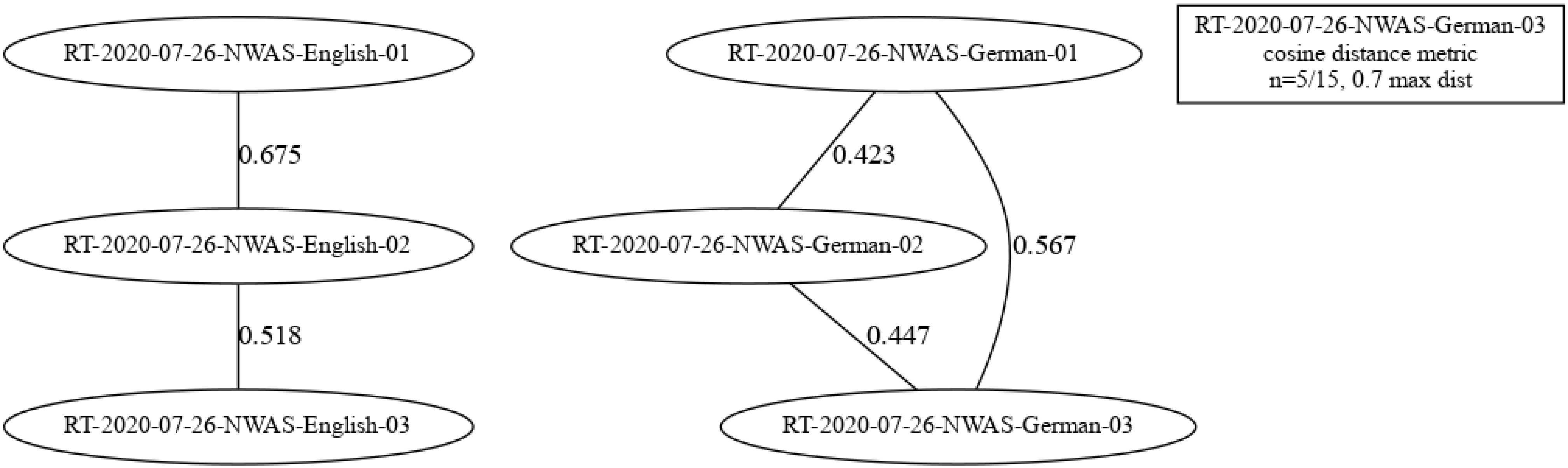
	Eng 01	Eng 02	Eng 03	Ger 01	Ger 02	Ger 03
		0.67477731	1.	0.74745837	0.93762055	0.85622088
			0.5184008	0.76221046	0.87568858	0.7706713
				0.78197106	0.85094568	0.82617612
					0.42298678	0.56668163
						0.44727788

Adult Female English-German bilingual reading
The North Wind and the Sun,
3 English, 3 German, in order of production.

Distance map with normalised distances



Distance map with normalised distances < 0.7



An English example:
The North Wind and the Sun

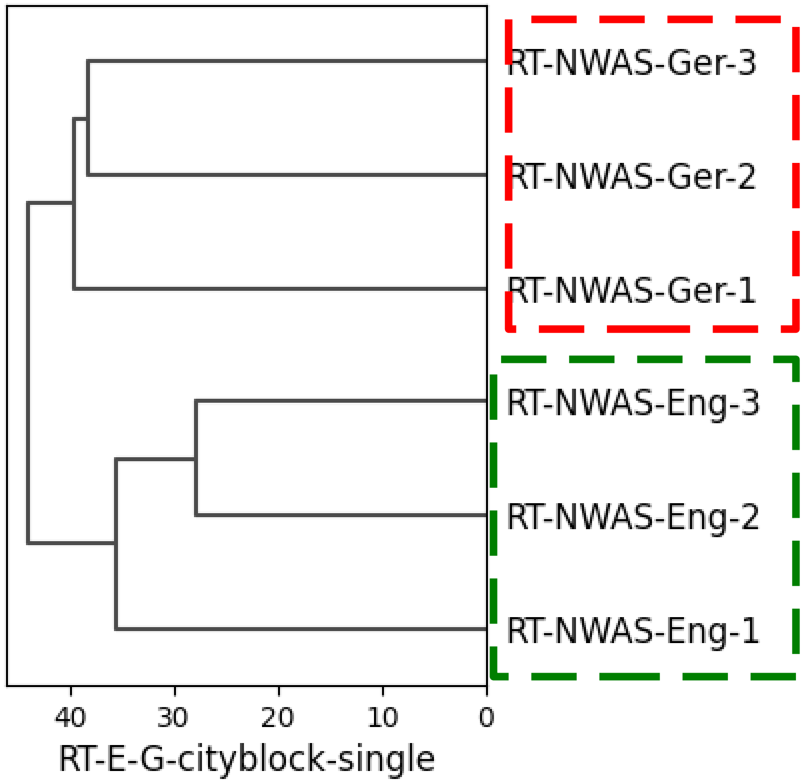
A German example:
Nordwind und Sonne

What can we do with these analyses?

Successful clustering.

Note:
1. The degree of similarity is reflected in the length of the horizontal lines.
2. So English #1 is quite close to the German readings, and German #1 is quite close to the English readings.
3. In fact, some distance measures group English #1 and German #1 together.

Since the analysis was successful, the data set is extended ...



RT-E-G-cityblock-single-dendro.tif

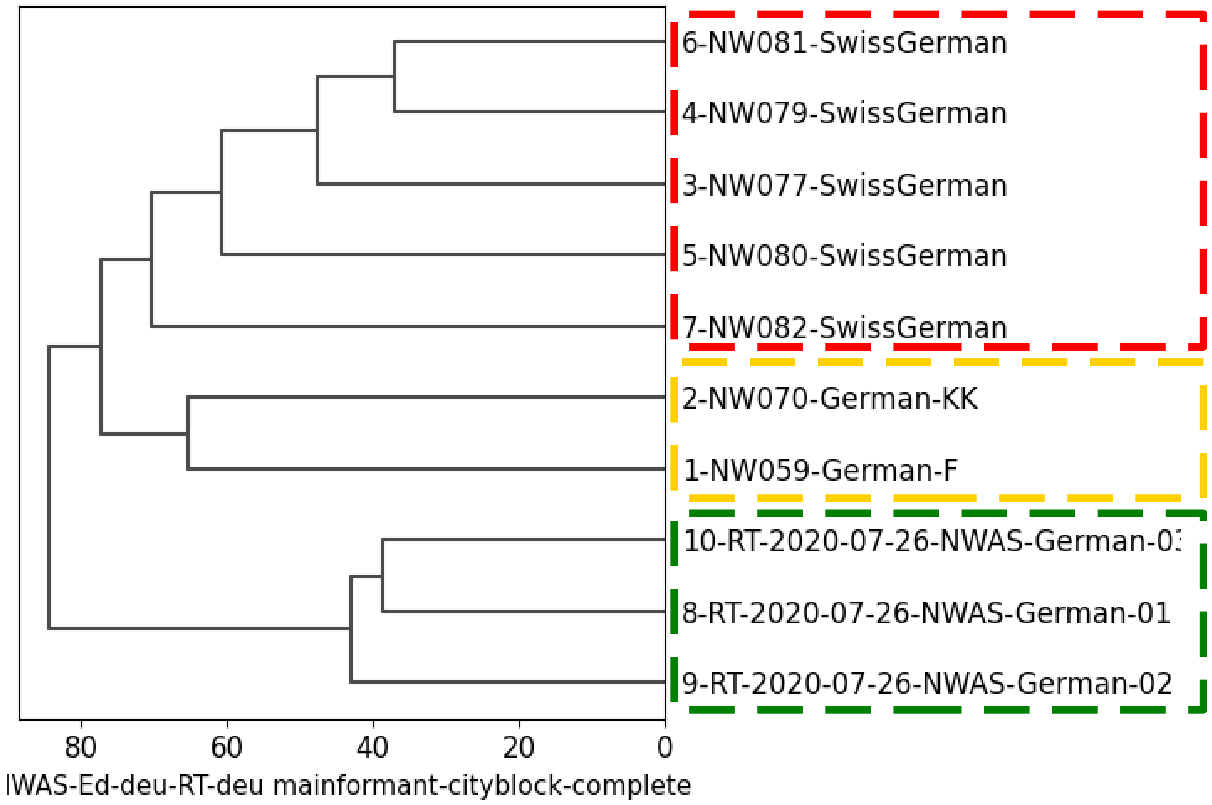
Red:	Female bilingual reading in her primary language
Green:	Female bilingual reading in her secondary language

What can we do with these analyses?

Successful clustering with the extended data set.

Since the analysis was successful, the data set is extended again, but that's for another day …

See my article *The Rhythms of Rhythm*, which is due to appear in the *Journal of the International Phonetic Association* (JIPA) later this year.



NWAS-Ed-deu-RT-deu mainformant-cityblock-complete.png

Red:	Swiss German (male and female speakers)
Yellow:	North German (male speakers)
Green:	Southern Rhineland (3 readings, same female speaker)

So what is the practical value of analyses like these?

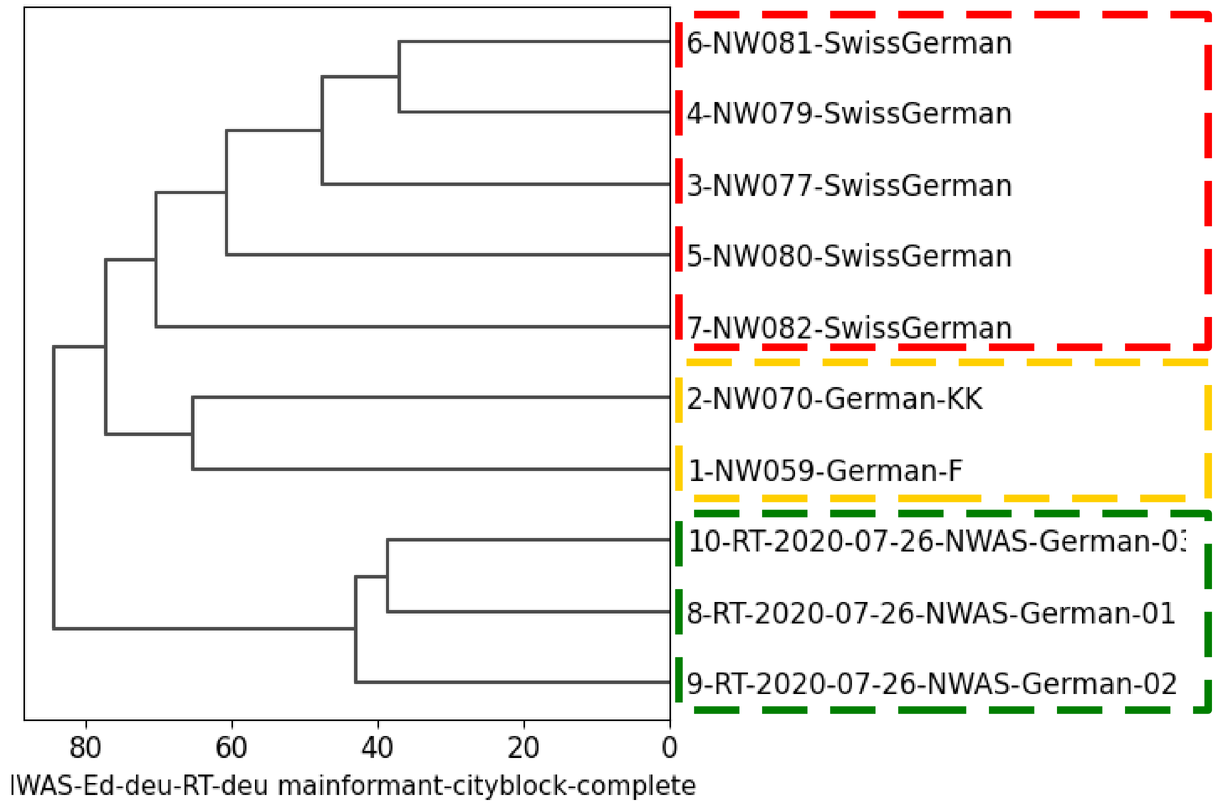
1. Second language assessment:
1. Phrasing
2. Fluency

2. Clinical applications:
1. Diagnosis (e.g. Parkinson)
2. Therapy monitoring

3. Language typology:
1. Classification
2. Identification

4. Discourse analysis:
1. Speech style
2. Variability in episodes

5. Speech engineering:
1. Style recognition
2. Language recognition
3. Speaker recognition
4. Natural speech synthesis



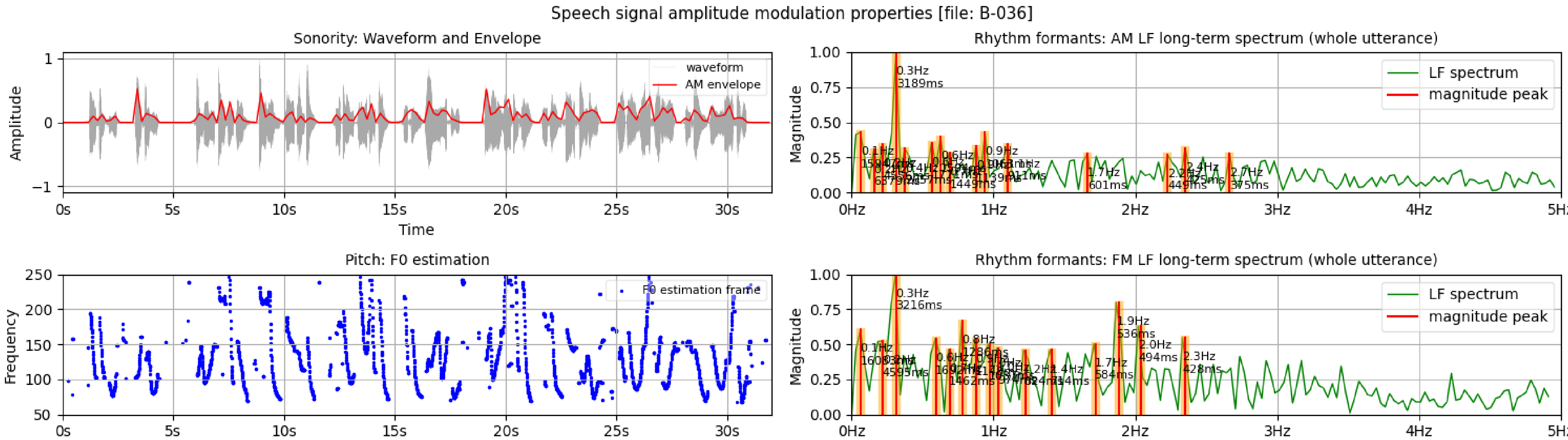
NWAS-Ed-deu-RT-deu mainformant-cityblock-complete.png

Red:	Swiss German (male and female speakers)
Yellow:	North German (male speakers)
Green:	Southern Rhineland (3 readings, same female speaker)

The Tang dynasty revisited: a closer look

In cooperation with Dr. Lin Xuewei, JNU, Guangzhou

Classification of Tang dynasty poetry recitation by rhythm



Time domain
- amplitude modulation (AM)
- frequency modulation (FM)

Spectral frequency domain
- AM
-FM

Classification of Tang dynasty poetry recitation by rhythm

Procedure:

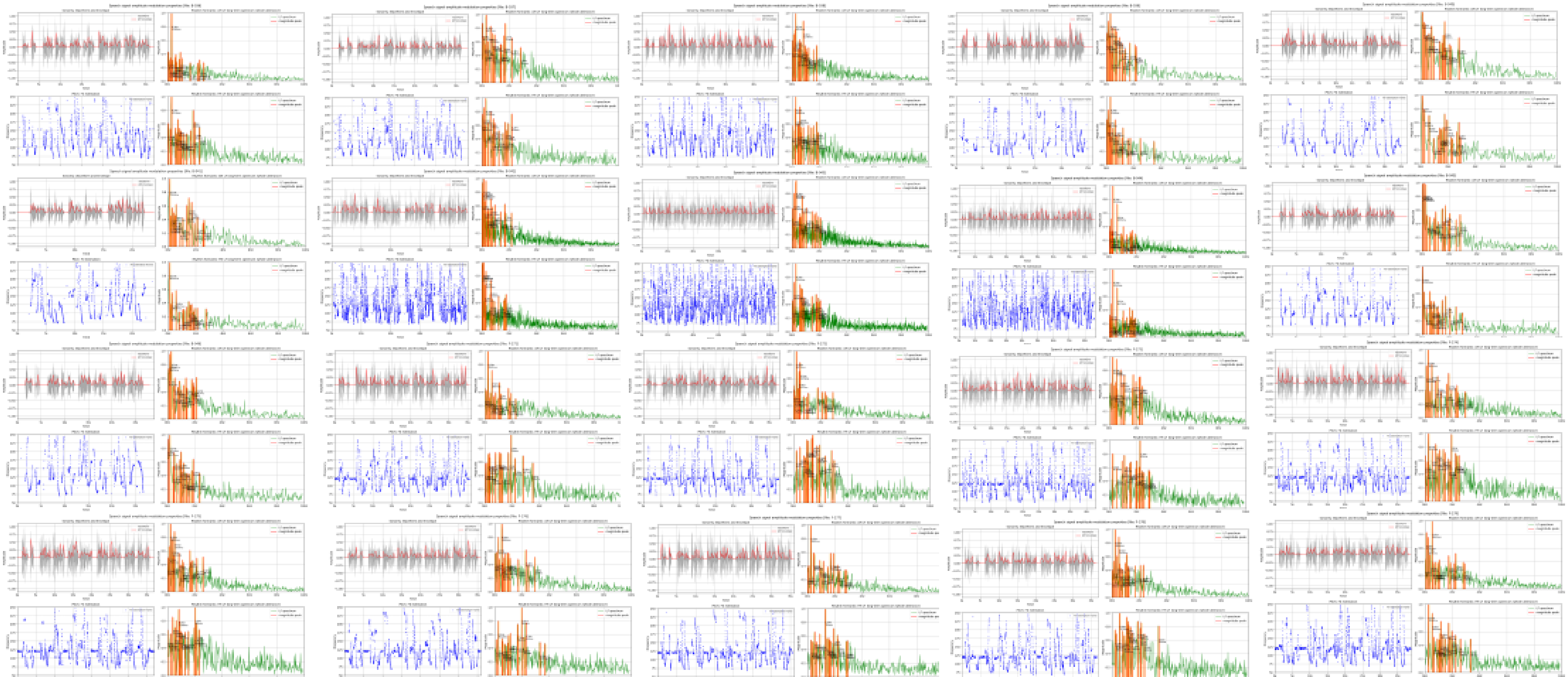
For the AM spectra of all 22 recordings,

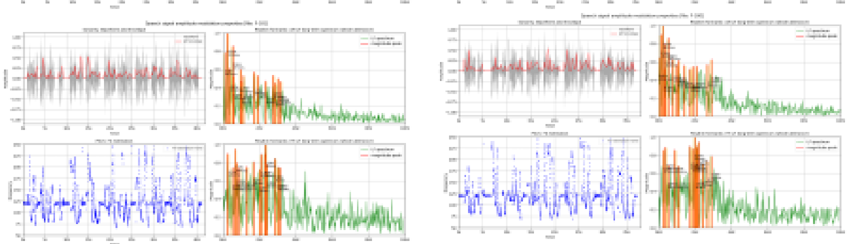
Collect the highest magnitude spectral values above a selected level

Compare these values pairwise:

- Select the relevant distance metrics
  - from, for example, Manhattan Distance, Normalised Manhattan Distance, Chebyshev Distance, Cosine Distance, Euclidean distance, ...
- Create a ‘distance matrix’
  - Display the contents of the distance matrix as a ‘distance map’
- Inductively create a hierarchy of pairs of recordings, and of pairs of pairs of recordings (hierarchical clustering)
  - Display the hierarchy as a dendrogram

Classification of Tang dynasty poetry recitation by rhythm





Data:
22 recordings (11 each of genres B and F)
Method:
Compare the orange patterns pairwise to find similarities and distances.
This is what the distance analysis does, for the 55 possible combinations.

Classification of Tang dynasty poetry recitation by rhythm

0.000	0.129	0.143	0.140	0.151	0.172	0.172	0.124	0.159	0.136	0.143	0.190	0.174	0.162	0.227	0.197	0.169	0.220	0.179	0.193	0.164	0.209
0.129	0.000	0.156	0.105	0.114	0.148	0.138	0.146	0.174	0.126	0.126	0.178	0.154	0.165	0.208	0.182	0.159	0.197	0.182	0.192	0.141	0.218
0.143	0.156	0.000	0.156	0.166	0.187	0.179	0.150	0.154	0.180	0.146	0.159	0.170	0.173	0.239	0.185	0.144	0.204	0.164	0.204	0.172	0.217
0.140	0.105	0.156	0.000	0.087	0.130	0.102	0.158	0.190	0.133	0.106	0.208	0.134	0.196	0.240	0.173	0.182	0.199	0.175	0.211	0.148	0.198
0.151	0.114	0.166	0.087	0.000	0.120	0.107	0.145	0.189	0.151	0.100	0.177	0.143	0.160	0.203	0.150	0.163	0.175	0.175	0.179	0.125	0.179
0.172	0.148	0.187	0.130	0.120	0.000	0.118	0.157	0.171	0.133	0.145	0.211	0.189	0.171	0.221	0.184	0.183	0.194	0.186	0.182	0.174	0.185
0.172	0.138	0.179	0.102	0.107	0.118	0.000	0.167	0.209	0.157	0.143	0.214	0.166	0.203	0.228	0.187	0.194	0.214	0.183	0.210	0.166	0.211
0.124	0.146	0.150	0.158	0.145	0.157	0.167	0.000	0.139	0.131	0.137	0.198	0.192	0.177	0.234	0.213	0.171	0.216	0.204	0.206	0.188	0.228
0.159	0.174	0.154	0.190	0.189	0.171	0.209	0.139	0.000	0.149	0.164	0.212	0.205	0.185	0.245	0.212	0.167	0.211	0.202	0.212	0.211	0.233
0.136	0.126	0.180	0.133	0.151	0.133	0.157	0.131	0.149	0.000	0.122	0.190	0.159	0.164	0.215	0.188	0.156	0.180	0.168	0.172	0.156	0.198
0.143	0.126	0.146	0.106	0.100	0.145	0.143	0.137	0.164	0.122	0.000	0.161	0.131	0.163	0.184	0.150	0.135	0.164	0.167	0.159	0.128	0.183
0.190	0.178	0.159	0.208	0.177	0.211	0.214	0.198	0.212	0.190	0.161	0.000	0.127	0.131	0.136	0.111	0.112	0.142	0.117	0.115	0.099	0.144
0.174	0.154	0.170	0.134	0.143	0.189	0.166	0.192	0.205	0.159	0.131	0.127	0.000	0.149	0.177	0.116	0.143	0.128	0.119	0.156	0.118	0.136
0.162	0.165	0.173	0.196	0.160	0.171	0.203	0.177	0.185	0.164	0.163	0.131	0.149	0.000	0.150	0.150	0.118	0.161	0.149	0.121	0.138	0.135
0.227	0.208	0.239	0.240	0.203	0.221	0.228	0.234	0.245	0.215	0.184	0.136	0.177	0.150	0.000	0.156	0.135	0.158	0.155	0.094	0.123	0.136
0.197	0.182	0.185	0.173	0.150	0.184	0.187	0.213	0.212	0.188	0.150	0.111	0.116	0.150	0.156	0.000	0.117	0.148	0.092	0.113	0.104	0.135
0.169	0.159	0.144	0.182	0.163	0.183	0.194	0.171	0.167	0.156	0.135	0.112	0.143	0.118	0.135	0.117	0.000	0.165	0.118	0.099	0.117	0.125
0.220	0.197	0.204	0.199	0.175	0.194	0.214	0.216	0.211	0.180	0.164	0.142	0.128	0.161	0.158	0.148	0.165	0.000	0.143	0.152	0.135	0.128
0.179	0.182	0.164	0.175	0.175	0.186	0.183	0.204	0.202	0.168	0.167	0.117	0.119	0.149	0.155	0.092	0.118	0.143	0.000	0.119	0.119	0.125
0.193	0.192	0.204	0.211	0.179	0.182	0.210	0.206	0.212	0.172	0.159	0.115	0.156	0.121	0.094	0.113	0.099	0.152	0.119	0.000	0.109	0.111
0.164	0.141	0.172	0.148	0.125	0.174	0.166	0.188	0.211	0.156	0.128	0.099	0.118	0.138	0.123	0.104	0.117	0.135	0.119	0.109	0.000	0.117
0.209	0.218	0.217	0.198	0.179	0.185	0.211	0.228	0.233	0.198	0.183	0.144	0.136	0.135	0.136	0.135	0.125	0.128	0.125	0.111	0.117	0.000

Distance map

1. The diagonal from top-left to bottom-right contains only zeroes.
2. The bottom left triangle is a mirror image of the top right triangle.

Classification of Tang dynasty poetry recitation by rhythm

0.000	0.129	0.143	0.140	0.151	0.172	0.172	0.124	0.159	0.136	0.143	0.190	0.174	0.162	0.227	0.197	0.169	0.220	0.179	0.193	0.164	0.209
0.129	0.000	0.156	0.105	0.114	0.148	0.138	0.146	0.174	0.126	0.126	0.178	0.154	0.165	0.208	0.182	0.159	0.197	0.182	0.192	0.141	0.218
0.143	0.156	0.000	0.156	0.166	0.187	0.179	0.150	0.154	0.180	0.146	0.159	0.170	0.173	0.239	0.185	0.144	0.204	0.164	0.204	0.172	0.217
0.140	0.105	0.156	0.000	0.087	0.130	0.102	0.158	0.190	0.133	0.106	0.208	0.134	0.196	0.240	0.173	0.182	0.199	0.175	0.211	0.148	0.198
0.151	0.114	0.166	0.087	0.000	0.120	0.107	0.145	0.189	0.151	0.100	0.177	0.143	0.160	0.203	0.150	0.163	0.175	0.175	0.179	0.125	0.179
0.172	0.148	0.187	0.130	0.120	0.000	0.118	0.157	0.171	0.133	0.145	0.211	0.189	0.171	0.221	0.184	0.183	0.194	0.186	0.182	0.174	0.185
0.172	0.138	0.179	0.102	0.107	0.118	0.000	0.167	0.209	0.157	0.143	0.214	0.166	0.203	0.228	0.187	0.194	0.214	0.183	0.210	0.166	0.211
0.124	0.146	0.150	0.158	0.145	0.157	0.167	0.000	0.139	0.131	0.137	0.198	0.192	0.177	0.234	0.213	0.171	0.216	0.204	0.206	0.188	0.228
0.159	0.174	0.154	0.190	0.189	0.171	0.209	0.139	0.000	0.149	0.164	0.212	0.205	0.185	0.245	0.212	0.167	0.211	0.202	0.212	0.211	0.233
0.136	0.126	0.180	0.133	0.151	0.133	0.157	0.131	0.149	0.000	0.122	0.190	0.159	0.164	0.215	0.188	0.156	0.180	0.168	0.172	0.156	0.198
0.143	0.126	0.146	0.106	0.100	0.145	0.143	0.137	0.164	0.122	0.000	0.161	0.131	0.163	0.184	0.150	0.135	0.164	0.167	0.159	0.128	0.183
0.190	0.178	0.159	0.208	0.177	0.211	0.214	0.198	0.212	0.190	0.161	0.000	0.127	0.131	0.136	0.111	0.112	0.142	0.117	0.115	0.099	0.144
0.174	0.154	0.170	0.134	0.143	0.189	0.166	0.192	0.205	0.159	0.131	0.127	0.000	0.149	0.177	0.116	0.143	0.128	0.119	0.156	0.118	0.136
0.162	0.165	0.173	0.196	0.160	0.171	0.203	0.177	0.185	0.164	0.163	0.131	0.149	0.000	0.150	0.150	0.118	0.161	0.149	0.121	0.138	0.135
0.227	0.208	0.239	0.240	0.203	0.221	0.228	0.234	0.245	0.215	0.184	0.136	0.177	0.150	0.000	0.156	0.135	0.158	0.155	0.094	0.123	0.136
0.197	0.182	0.185	0.173	0.150	0.184	0.187	0.213	0.212	0.188	0.150	0.111	0.116	0.150	0.156	0.000	0.117	0.148	0.092	0.113	0.104	0.135
0.169	0.159	0.144	0.182	0.163	0.183	0.194	0.171	0.167	0.156	0.135	0.112	0.143	0.118	0.135	0.117	0.000	0.165	0.118	0.099	0.117	0.125
0.220	0.197	0.204	0.199	0.175	0.194	0.214	0.216	0.211	0.180	0.164	0.142	0.128	0.161	0.158	0.148	0.165	0.000	0.143	0.152	0.135	0.128
0.179	0.182	0.164	0.175	0.175	0.186	0.183	0.204	0.202	0.168	0.167	0.117	0.119	0.149	0.155	0.092	0.118	0.143	0.000	0.119	0.119	0.125
0.193	0.192	0.204	0.211	0.179	0.182	0.210	0.206	0.212	0.172	0.159	0.115	0.156	0.121	0.094	0.113	0.099	0.152	0.119	0.000	0.109	0.111
0.164	0.141	0.172	0.148	0.125	0.174	0.166	0.188	0.211	0.156	0.128	0.099	0.118	0.138	0.123	0.104	0.117	0.135	0.119	0.109	0.000	0.117
0.209	0.218	0.217	0.198	0.179	0.185	0.211	0.228	0.233	0.198	0.183	0.144	0.136	0.135	0.136	0.135	0.125	0.128	0.125	0.111	0.117	0.000

Distance map

1. The diagonal from top-left to bottom-right contains only zeroes.

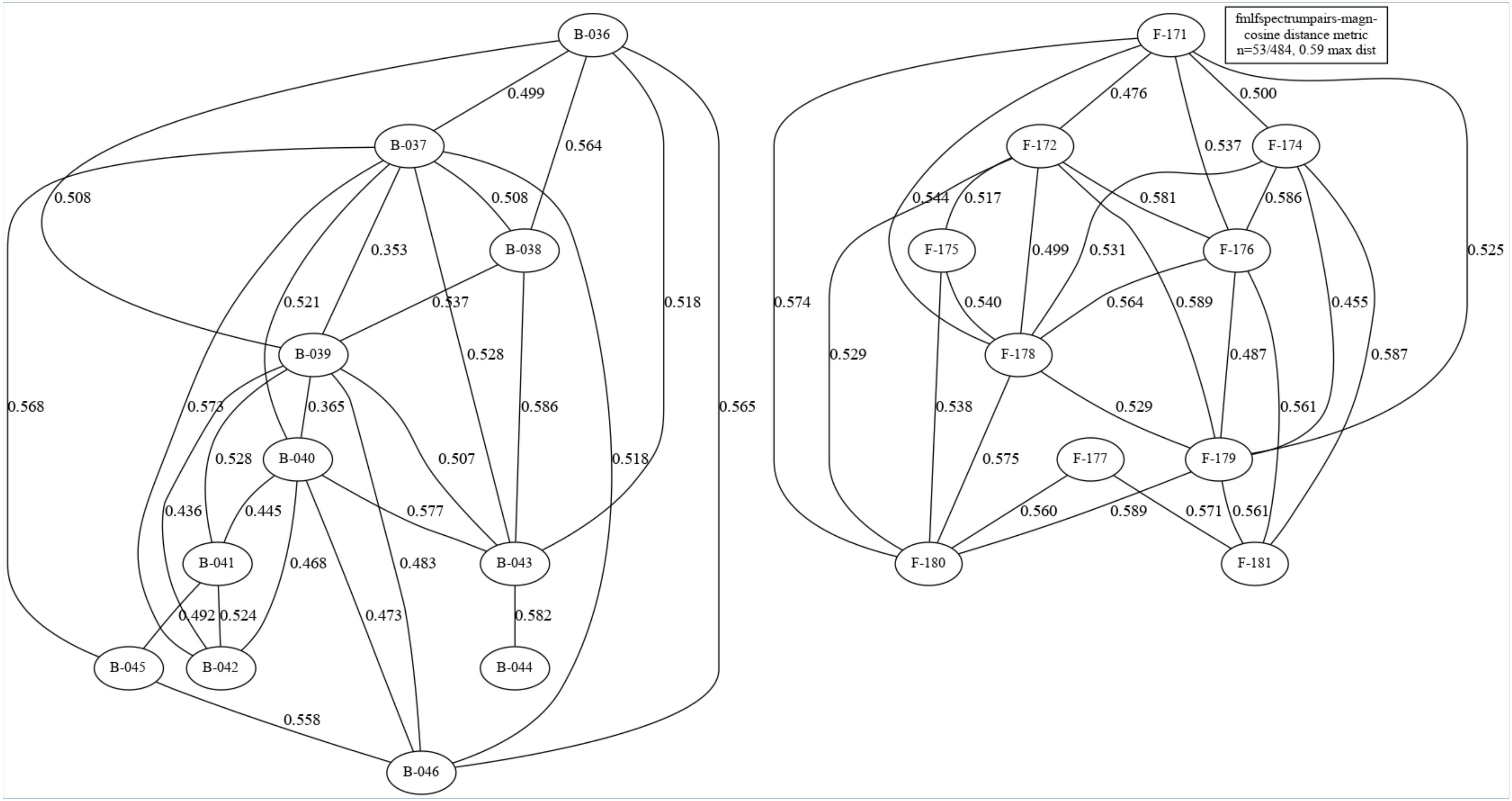
Classification of Tang dynasty poetry recitation by rhythm

0.000	0.129	0.143	0.140	0.151	0.172	0.172	0.124	0.159	0.136	0.143	0.190	0.174	0.162	0.227	0.197	0.169	0.220	0.179	0.193	0.164	0.209
	0.000	0.156	0.105	0.114	0.148	0.138	0.146	0.174	0.126	0.126	0.178	0.154	0.165	0.208	0.182	0.159	0.197	0.182	0.192	0.141	0.218
		0.000	0.156	0.166	0.187	0.179	0.150	0.154	0.180	0.146	0.159	0.170	0.173	0.239	0.185	0.144	0.204	0.164	0.204	0.172	0.217
			0.000	0.087	0.130	0.102	0.158	0.190	0.133	0.106	0.208	0.134	0.196	0.240	0.173	0.182	0.199	0.175	0.211	0.148	0.198
				0.000	0.120	0.107	0.145	0.189	0.151	0.100	0.177	0.143	0.160	0.203	0.150	0.163	0.175	0.175	0.179	0.125	0.179
					0.000	0.118	0.157	0.171	0.133	0.145	0.211	0.189	0.171	0.221	0.184	0.183	0.194	0.186	0.182	0.174	0.185
						0.000	0.167	0.209	0.157	0.143	0.214	0.166	0.203	0.228	0.187	0.194	0.214	0.183	0.210	0.166	0.211
							0.000	0.139	0.131	0.137	0.198	0.192	0.177	0.234	0.213	0.171	0.216	0.204	0.206	0.188	0.228
								0.000	0.149	0.164	0.212	0.205	0.185	0.245	0.212	0.167	0.211	0.202	0.212	0.211	0.233
									0.000	0.122	0.190	0.159	0.164	0.215	0.188	0.156	0.180	0.168	0.172	0.156	0.198
										0.000	0.161	0.131	0.163	0.184	0.150	0.135	0.164	0.167	0.159	0.128	0.183
											0.000	0.127	0.131	0.136	0.111	0.112	0.142	0.117	0.115	0.099	0.144
												0.000	0.149	0.177	0.116	0.143	0.128	0.119	0.156	0.118	0.136
													0.000	0.150	0.150	0.118	0.161	0.149	0.121	0.138	0.135
														0.000	0.156	0.135	0.158	0.155	0.094	0.123	0.136
															0.000	0.117	0.148	0.092	0.113	0.104	0.135
																0.000	0.165	0.118	0.099	0.117	0.125
																	0.000	0.143	0.152	0.135	0.128
																		0.000	0.119	0.119	0.125
																			0.000	0.109	0.111
																				0.000	0.117
																					0.000

Distance map

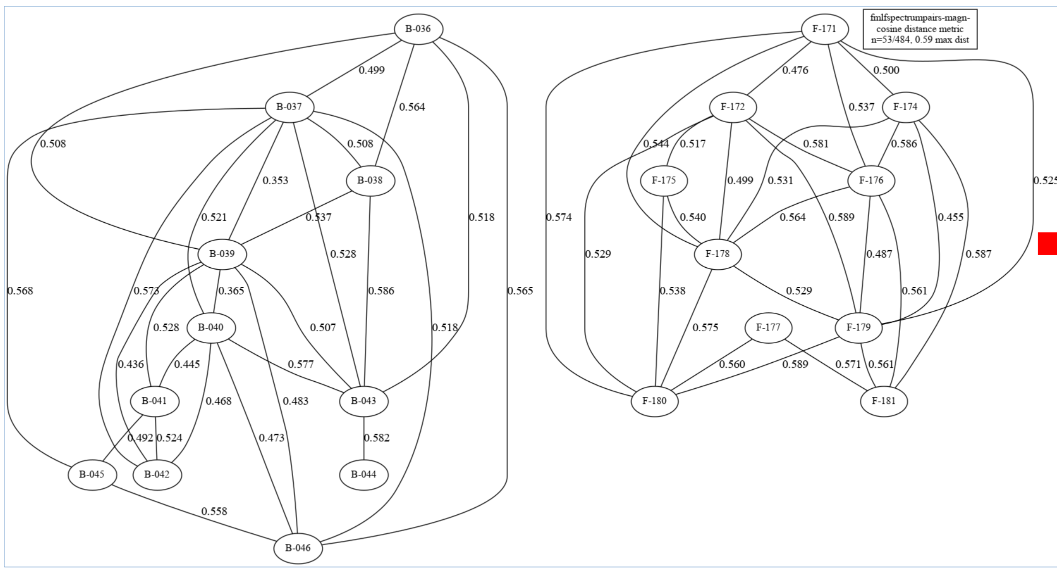
1. The diagonal from top-left to bottom-right contains only zeroes.
2. The bottom left triangle is a mirror image of the top right triangle.

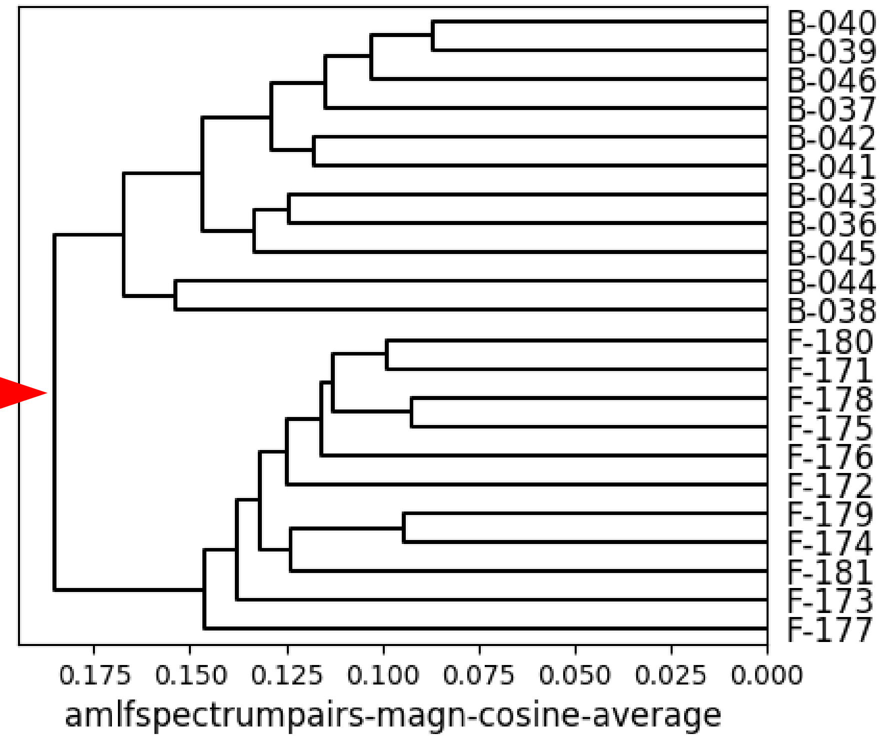
Classification of Tang dynasty poetry recitation by rhythm



partitioned distance map

Classification of Tang dynasty poetry recitation by rhythm





partitioned distance map

partitioned dendrogram

Acknowledgment: recordings were kindly provided by Dr. Xuewei Lin, Jinan University, Guangzhou.

Summary and Conclusion

RFT and RFA

1. Rhythm Formant Theory (RFT):
   1. A rhythm formant is a frequency zone of higher magnitude values in the normalised low frequency (LF) spectrum.
   2. Rhythm formants are detected in the LF AM spectrum and in the LF FM spectrum.
2. Rhythm Formant Analysis (RFA):
   1. The spectrum magnitudes are obtained by FFT and normalised to the range 0,…,1.
   2. A minimum magnitude (e.g. about 0.2) is defined as a cutoff level, below which values are clipped to zero.
   3. The clipped spectra of different recordings are compared using standard distance metrics and represented as distance maps, and hierarchically clustered using standard clustering criteria and represented as dendrograms.

So what is the practical value of analyses like these?

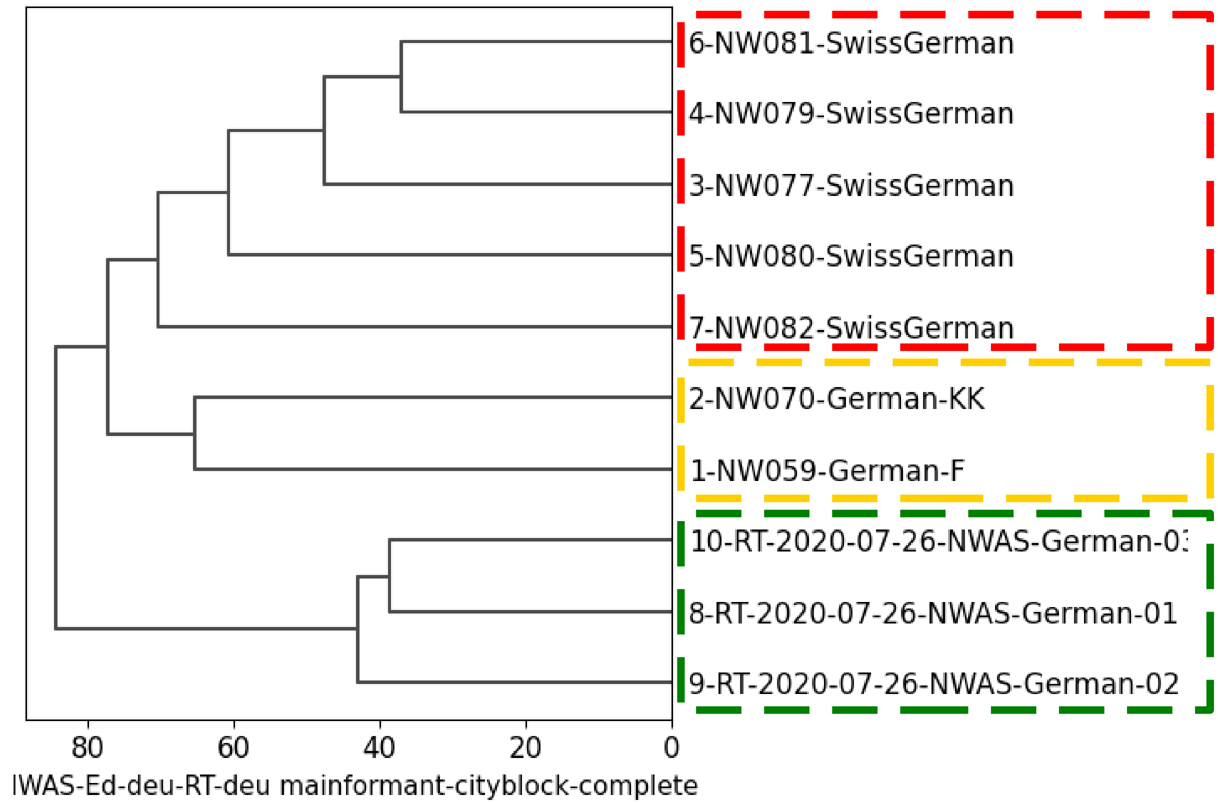
1. **Second language assessment:**
   1. Phrasing
   2. Fluency

2. **Clinical applications:**
   1. Diagnosis (e.g. Parkinson)
   2. Therapy monitoring

3. **Language typology:**
   1. Classification
   2. Identification

4. **Discourse analysis:**
   1. Speech style
   2. Variability in episodes

5. **Speech engineering:**
   1. Style recognition
   2. Language recognition
   3. Speaker recognition
   4. Natural speech synthesis



NWAS-Ed-deu-RT-deu mainformant-cityblock-complete.png

Red:	Swiss German (male and female speakers)
Yellow:	North German (male speakers)
Green:	Southern Rhineland (3 readings, same female speaker)

谢谢

Many thanks for participating!

Assignment: Same as yesterday ...