Lecture 1: Basic Prosody

1A: Background

Dafydd Gibbon
Bielefeld University, Germany
2022-04-25

II Brazilian Congress of Prosody
Minicourse 9, 25, 27, 39 April 2022
(09:00-11:30 Brazilian Standard Time)

Background and Objective

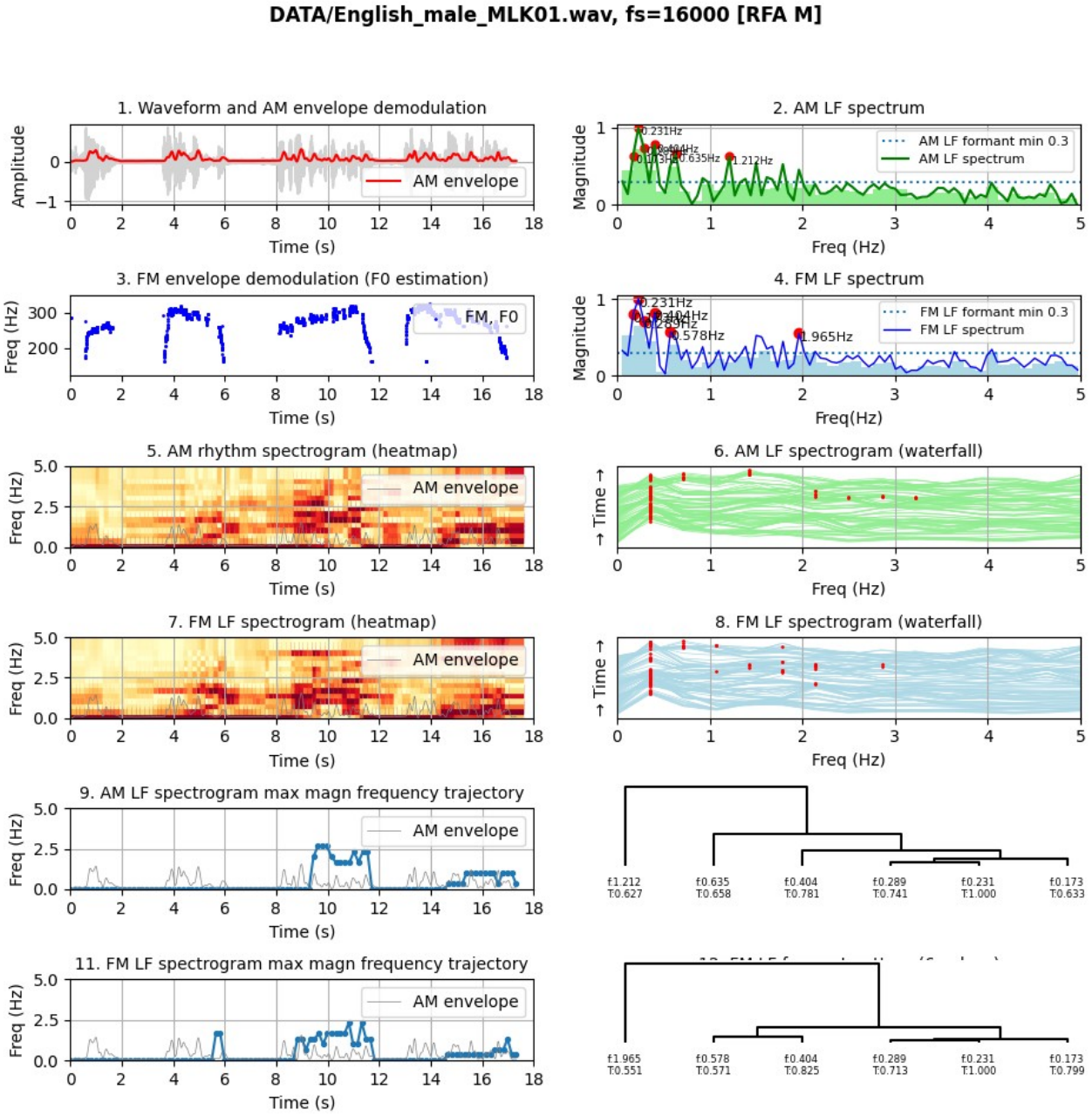
Rhythm Formant Theory:

- Speech rhythms can be measured and related to language and speech style in the Low Frequency spectrum of the speech signal.
- speech rhythm variation can be measured and related to language and speech style in the Low Frequency spectrogram of the speech signal.
- Method:
  - Envelope extraction
  - F0 estimation
  - FFT (spectrum)
  - FFT windowing (spectrogram)
  - Data clustering

Background and Objective

Rhythm Formant Theory:

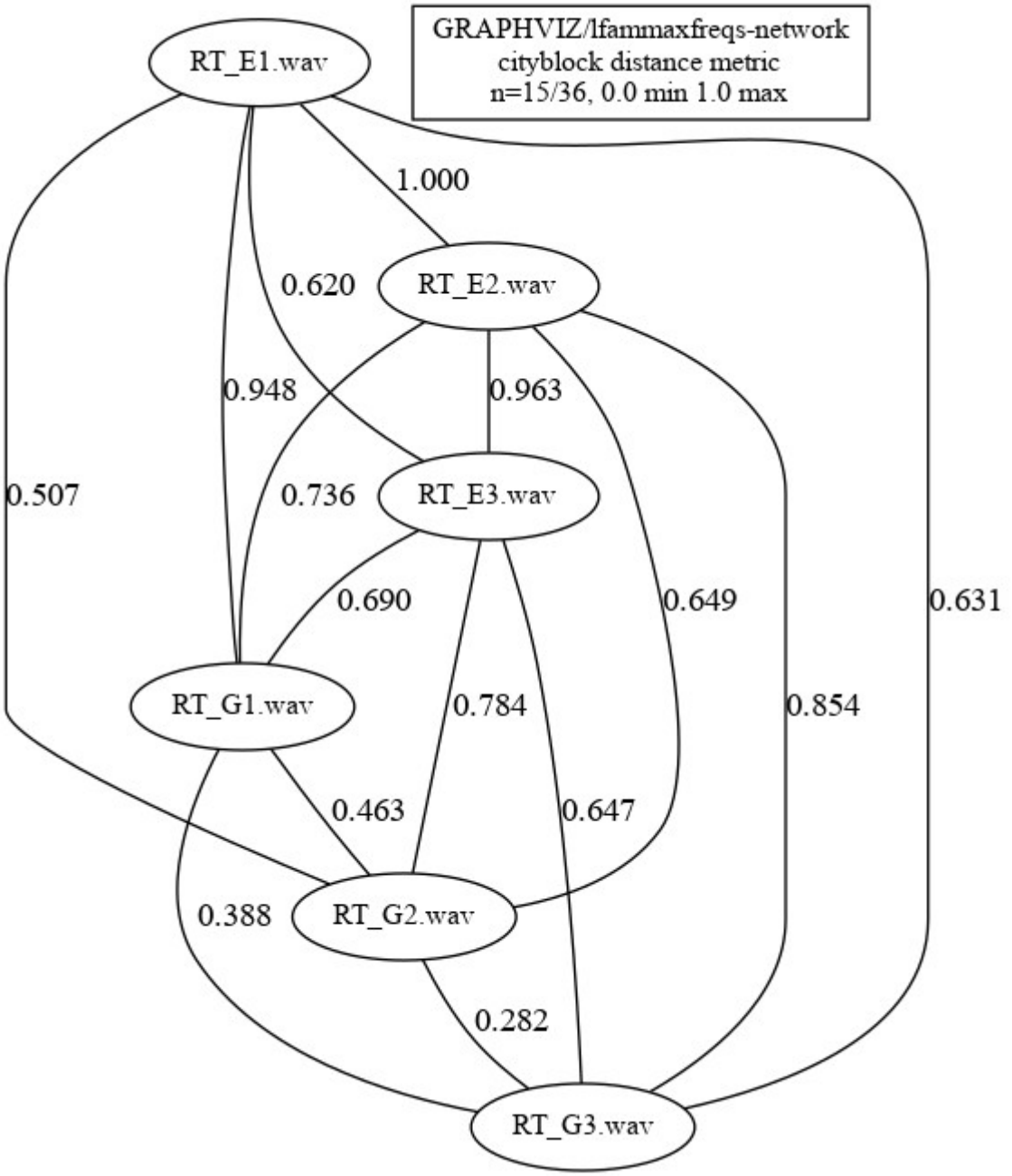
- Speech rhythms can be measured and related to language and speech style in the Low Frequency spectrum of the speech signal.
- speech rhythm variation can be measured and related to language and speech style in the Low Frequency spectrogram of the speech signal.
- Method:
  - **Envelope extraction**
  - **F0 estimation**
  - **FFT (spectrum)**
  - **FFT windowing (spectrogram)**
  - Data clustering



Background and Objective

Rhythm Formant Theory:

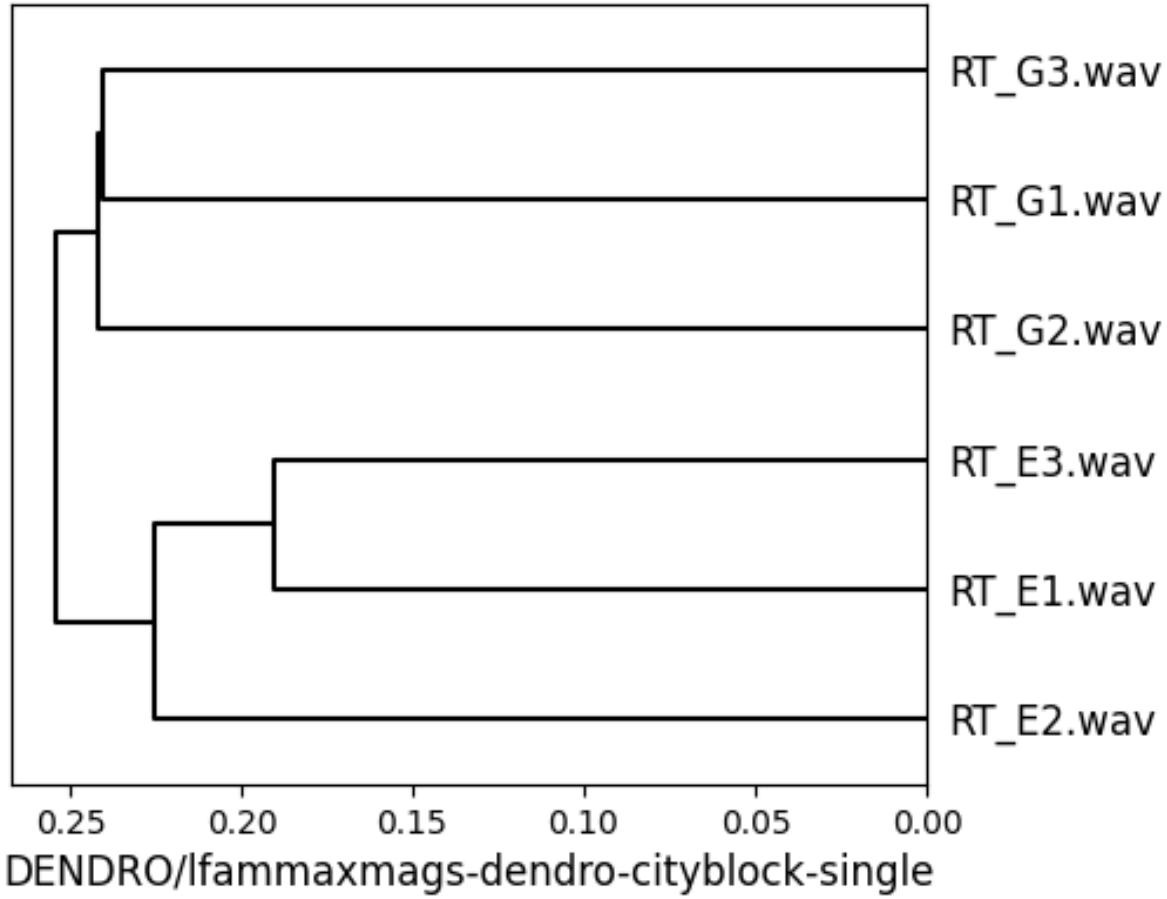
- Speech rhythms can be measured and related to language and speech style in the Low Frequency spectrum of the speech signal.
- speech rhythm variation can be measured and related to language and speech style in the Low Frequency spectrogram of the speech signal.
- Method:
  - Envelope extraction
  - F0 estimation
  - FFT (spectrum)
  - FFT windowing (spectrogram)
  - **Data clustering**



Background and Objective

Rhythm Formant Theory:

- Speech rhythms can be measured and related to language and speech style in the Low Frequency spectrum of the speech signal.
- speech rhythm variation can be measured and related to language and speech style in the Low Frequency spectrogram of the speech signal.
- Method:
  - Envelope extraction
  - F0 estimation
  - FFT (spectrum)
  - FFT windowing (spectrogram)
  - **Data clustering**



RFA Software Repository

Rhythm Formant Analysis (RFA) repository:

https://github.com/dafyddg/RFA/

This repository is for the code which implements the Rhythm Formant Analysis methodology for Rhythm Formant Theory, as described in the following publication:
Gibbon, Dafydd. 2021. The rhythms of rhythm. *Journal of the International Phonetic Association*, 16 August 2021, 1-33. *First View*. (online, open access)
https://www.doi.org/10.1017/S002510032100008 6

Repository contents:

README.1st

Articles

LittleHelpers
LittleHelpers.zip

RFA_single_signal_processing
RFA_single_signal_processing.zip

RFA_multiple_signal_processing
RFA_multiple_signal_processing.zip

Sesson 1: Overview

- Prosody
  - poetic prosody
  - speech prosody
- Qualitative methods phonetics
  - epistemological basis of phonetics and linguistics
  - (computational) phonology
  - transcription and manual annotation
- Quantitative phonetics
  - (semi)automatic annotation with statistical training
  - analysis of annotations
  - signal processing
  - (un)supervised machine learning

Monday (basics)

Wednesday (melody)

Friday (rhythm)

“Prosody”

Practical prosody work
depends on theoretical understanding.

One aspect is consistent terminology.

Wikipedia

Prosody: may refer to:

Prosody (Sanskrit), the study of poetic meters and verse in Sanskrit and one of the six Vedangas, or limbs of Vedic studies

Prosody (Greek), the theory and practice of Greek versification

Prosody (Latin), the study of Latin versification and its laws of meter

Prosody (linguistics), the suprasegmental characteristics of speech

Prosody (music), the manner of setting words to music

Prosody (software), a cross-platform XMPP server written in Lua

Metre (poetry), the rhythmic structure of versed text

See also[edit]

Arabic prosody, study of poetic meters in Arabic; sometimes called the Science of Poetry

Semantic prosody, the way neutral words can be perceived as positive or negative

Emotional prosody, perception of emotion in speech

Poetic prosody

- has been a fruitful source of metaphors in phonology:
  - ‘iambic’ (right-headed), ‘trochaic’ (left-headed)
  - ‘foot’, ‘ictus’ (stressed), ‘remiss’ (unstressed) **– Abercrombie**
  - anacrusis (initial unstressed) **– Jassem**
  - ‘metre’ **– Metrical Phonology**
- but it is useful to know the original meanings:
  - poetic prosody:
    - Greek: προσῳδία (prosōidía), a song sung to music; pronunciation of the syllable
    - Greek and Latin versification: metre based on *length*
    - English versification: metre based on *stress*
    - Chinese versification (traditional):
      - no metre, based on syllable-dependent ‘yin’ and ‘yang’ (flat and sharp) tone types
  - literary studies:
    - structuralist approaches, ‘coupling’ (Jakobson, Levin):
      - patterns of rhyme, alliteration, assonance, rhythm

Poetic prosody – for example: metre

FOOT	left-headed	right-headed	mid-headed	multi-headed	nil-headed
disyllabic:	trochaic (trochee)	iambic (iamb)		spondaic (spondee)	pyrrhic (pyrrhus)
trisyllabic:	dactylic (dactyl)	anapaestic (anapaest)	amphibrachic (amphibrach)		
undefined:					anacrusis

LINE

monometer:	1 foot
dimeter:	2 feet
trimeter:	3 feet
tetrameter:	4 feet
pentameter:	5 feet
hexameter:	6 feet
heptameter:	7 feet
octameter:	8 feet

Fleas

Strickland Gillilan
Adam
Had ‘em.

iambic pentameter (5 iambs, 10 syllables)
That time | *of year* | *thou mayst* | *in me* | *behold*
trochaic tetrameter (4 trochees, 8 syllables)
Tell me | *not in* | *mournful* | *numbers*
anapestic trimeter (3 anapests, 9 syllables)
And the sound | *of a voice* | *that is still*

It's four in / the morning, / the end of / December
I'm writing / you now just / to see if / you're better
New York / is cold, but / I like where / I'm living
There's music / on Clinton / Street all through / the evening.

Speech Prosody: Rhythms and Melodies of Speech

Speech Prosody: Rhythms and Melodies of Speech

- Prosodic units:

  Units of speech which are longer than speech sounds.

  Units of speech which are not speech sounds.

    Speech sounds (phones, phonemes) are traditionally referred to as ‘segments’

      strange: ‘segment’ just means part of a disc or line, therefore in general terms syllables, words, etc. are also segments...

  So prosodic units are often called ‘suprasegmentals’.
- Firth: ‘prosodies’
- Goldsmith: ‘autosegments’
- Most people: ‘intonation’, ‘rhythm’, ‘(pitch) accent’, ‘(nuclear, lexical, morphological) tone’ (also ‘stress’, though this is a tricky term)

Qualitative methods

including transcription, annotation, choice of methods and choice of evaluation criteria

IPA

Seven categories of speech sounds

Clicks
Implosives
Ejectives

Coarticulations
Other sounds

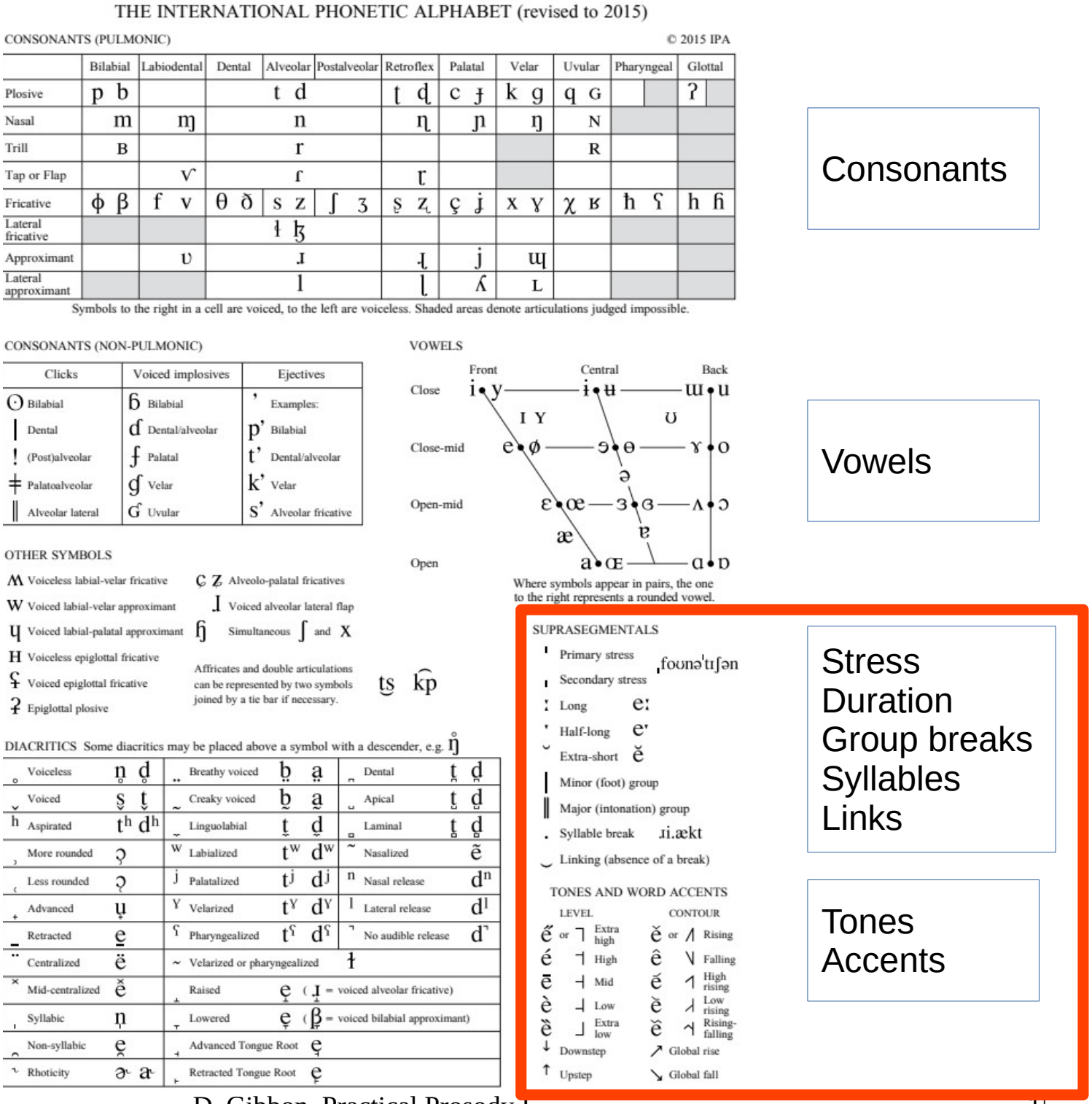
Modifications

Consonants

Vowels

Stress
Duration
Group breaks
Syllables
Links

Tones
Accents



Transcription of aspects of of rhythm and intonation

SUPRASEGMENTALS

ˈ	Primary stress	ˌfoʊnəˈtɪʃən
ˌ	Secondary stress	
ː	Long	eː
ˑ	Half-long	eˑ
˘	Extra-short	ĕ
|	Minor (foot) group	
‖	Major (intonation) group	
.	Syllable break	ɹi.ækt
‿	Linking (absence of a break)	

TONES AND WORD ACCENTS

LEVEL			CONTOUR		
e̋ or e᷄	˥	Extra high	ě or e᷅	˩˥	Rising
é	˦	High	ê	˥˩	Falling
ē	˧	Mid	e᷈	˦˥	High rising
è	˨	Low	e᷄	˩˧	Low rising
e̗	˩	Extra low	e᷅	˧˩	Rising-falling
↓		Downstep	↗		Global rise
↑		Upstep	↘		Global fall

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)

SAMPROSA

Register	H	72	High pitch
	L	76	Low pitch
	T	84	Top pitch (extreme H)
	B	66	Bottom pitch (extreme L)
	M	77	Mid pitch
	+	43	Higher pitch
	++	43,43	Much higher pitch
	+-	43,45	Peak (upward-downward)
	-	45	Lower pitch
	--	45,45	Much lower pitch
	-+	45,43	Trough (downward-upward)
	^	94	Upstep
	^^	94,94	Wide upstep
	!	33	Downstep
	!!	33,33	Wide downstep
	= or > or S	61 62 or 83	Level or same tone

Global tone: from Registerl and Tone repertoires

Terminal tone: from Register and Tone repertoires

Tone	-	45	Level tone (before tone group boundary)
	' or / or R	39 47 or 82	Rising tone
	` or \ or F	96 92 or 70	Falling tone
	`' (etc.)	96,39 (etc.)	Fall-rise
	'` (etc.)	39,96 (etc.)	Rise-fall
Length	:	58	Segment length mark
Stress	"	34	Primary stress
	%	37	Secondary stress
Pause	...	46,46,46	Silence
Boundary	$	36	Syllable boundary
	#	35	Word boundary
	|	124	Tone group boundary (non-directional)
	[	91	Tone group boundary (left)
	]	93	Tone group boundary (right)
Metasigns	-	45	Separator (the underscore, _, ASCII 95, may replace this owing to ambiguity with level tone)
	*	42	Conjunctor

Intonation: ToBI (Tones and Break Indices)

Pitch accents:

H* or L* L*+H, L+H*, H*+L, H+L*, ...: on words with most information in a sentence

Boundary tones:

H% and L%: at phrase edges (‘nuclear/final tone’, incomplete vs. complete meaning

Phrase accents:

H- or L-: tones between pitch accent and boundary tone, modification of boundary tone,

H* L* H%: fall-rise nuclear tone

L* H- L%: rise-fall nuclear tone

L* H- H%: rise nuclear tone

Prosodic hierarchy markers (strength of breaks between words)

0 = **clitic** boundary, e.g. who's
1 = normal **word** boundary
2 = perceived **juncture with no intonation effect**, or apparent intonational boundary without a pause or any other clues
3 = **intermediate phrase**, marked with H- or L-.
4 = **full intonation phrase**, marked L% or H%, at the end of a phrase or sentence

Intonation: ToBI (Tones and Break Indices)

Pitch accents:

H* or L* L*+H, L+H*, H*+L, H+L*, ...: on words with most information in a sentence

Boundary tones:

H% and L%: at phrase edges (‘nuclear/final tone’, incomplete vs. complete meaning

Phrase accents:

H- or L-: tones between pitch accent and boundary tone, modification of boundary tone,

H* L* H%: fall-rise nuclear tone

L* H- L%: rise-fall nuclear tone

L* H- H%: rise nuclear tone

Note that the boundary markers are essentially phonological and language-specific, based on linguistically interpretable categories.

Prosodic hierarchy markers (strength of breaks between words)

0 = **clitic** boundary, e.g. who's
1 = normal **word** boundary
2 = perceived **juncture with no intonation effect**, or apparent intonational boundary without a pause or any other clues
3 = **intermediate phrase**, marked with H- or L-.
4 = **full intonation phrase**, marked L% or H%, at the end of a phrase or sentence

Intonation: IntSint

Speaker-dependent or Utterance-dependent parameters:

key: like a musical key,establishes reference as F0 value (Hz)

range: interval between highest and lowest F0 (Hz)

Absolute:

T (Top) := key + range/2
M (mid) := key
B (Bottom) := key - range/2

Relative:

H (Higher) $F_i := (F_{i-1} + T) \times 0.5$
U (Upstepped) $F_i := (3 \times F_{i-1} + T) \times 0.25$
S (Same) $F_i := F_{i-1}$
D (Downstepped) $F_i := (3 \times F_{i-1} + B) \times 0.25$
L (Lower) $F_i := (F_{i-1} + B) \times 0.5$

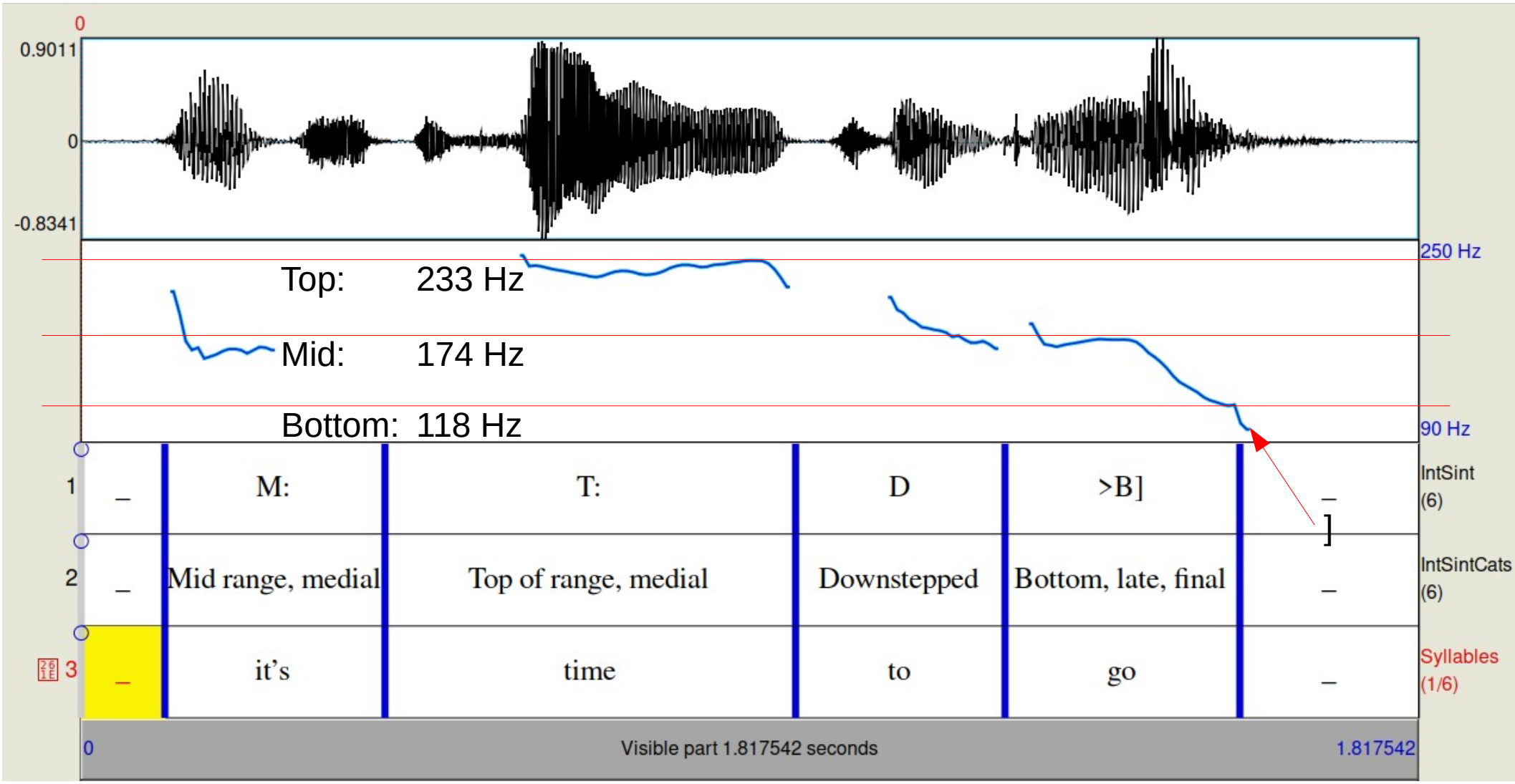
Note that this transcription is basically oriented towards speech synthesis.

[(initial)
< (early)
: (medial)
> (late)
] (final)

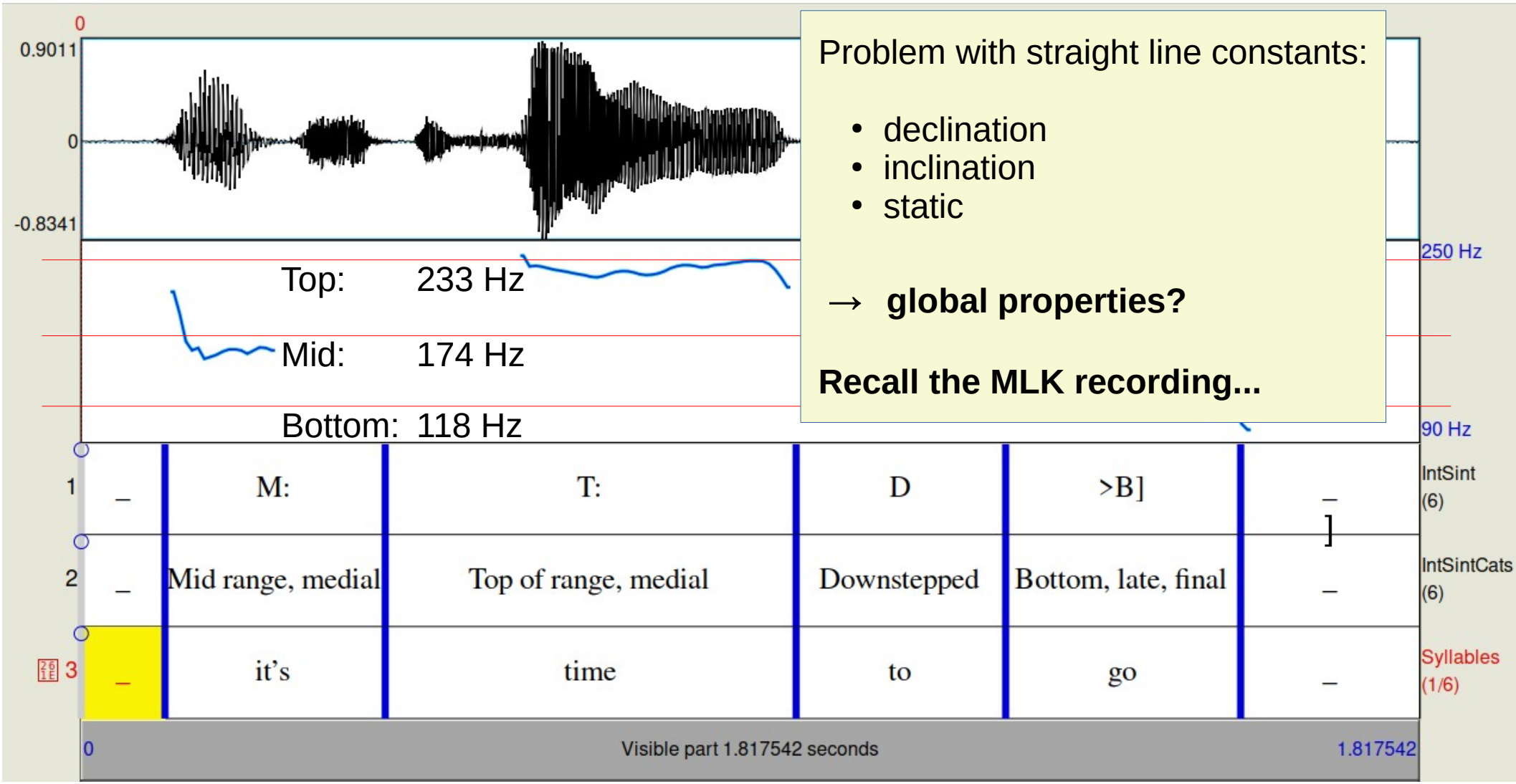
Example:

”It’s time to go!”
M: /ɪts/ T: /taɪmtə/ D <B] /goʊ/

Intonation: IntSint



Intonation: IntSint



Problem with straight line constants:

- declination
- inclination
- static

→ **global properties?**

Recall the MLK recording...

Before we can sensibly do anything practical with prosody, we need ...

Before we can sensibly do anything practical with prosody,
we need ...

CLARIFICATION OF THE DOMAIN

TERMINOLOGY

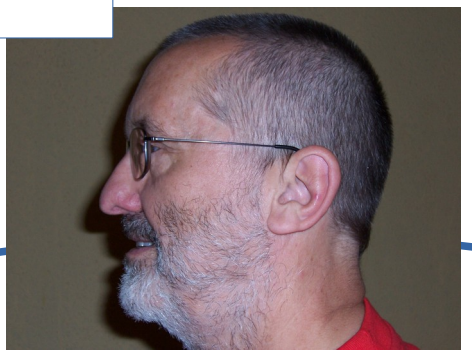
HIERARCHIES

FUNCTIONALITIES

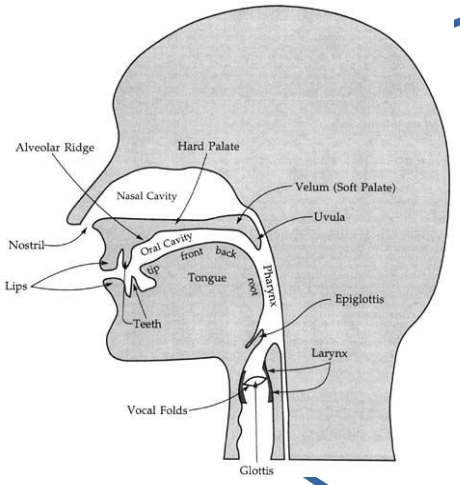
→ *Rank Interpretation Model*

TERMINOLOGY: four domains

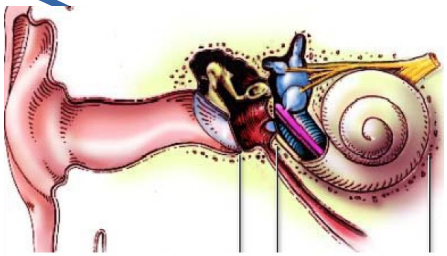
Phonology



Articulatory Phonetics
production: phonation

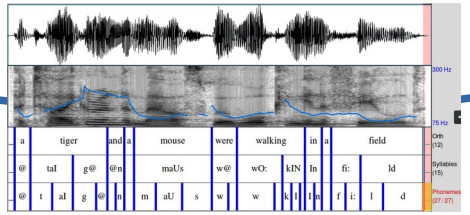


Auditory Phonetics
pitch prominence



outer ear middle ear inner ear

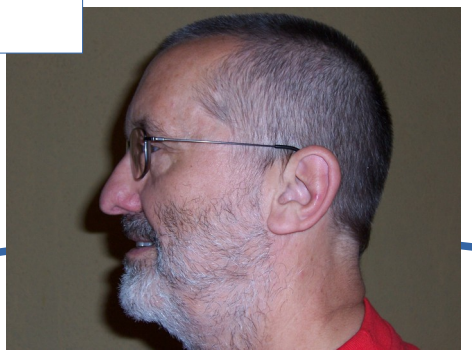
"prosody"



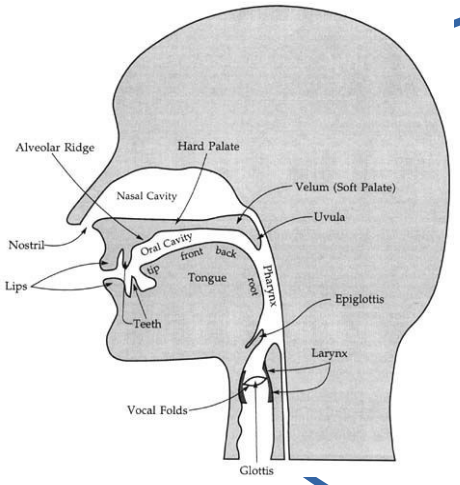
Acoustic Phonetics
transmission:
F0, timing, intensity

Terminology in four domains: *stress*

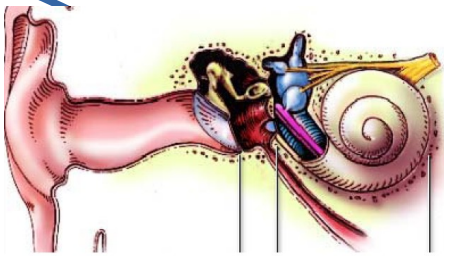
stress position
(phonology)



effort

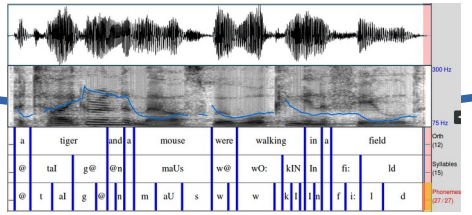


prominence
salience
pitch accent



outer ear
middle ear
inner ear

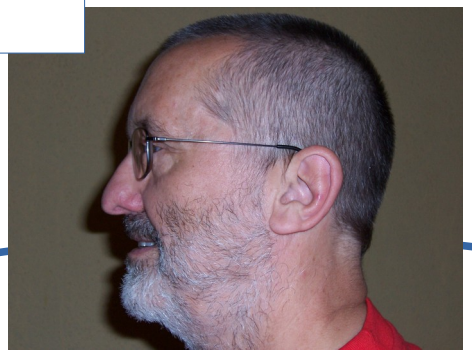
“stress”



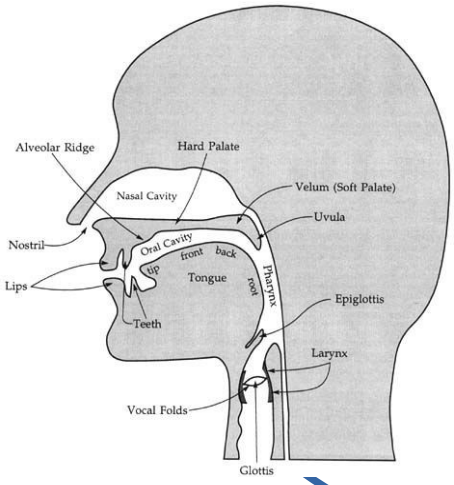
F0 height, F0 contour
syllable duration
syllable intensity

Terminology in four domains: *melody*

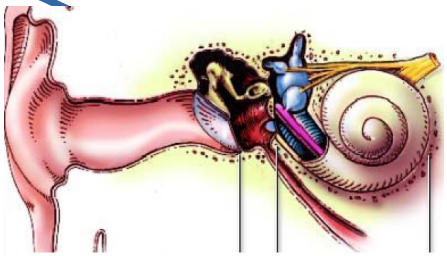
tone, intonation (phonology)



phonation rate

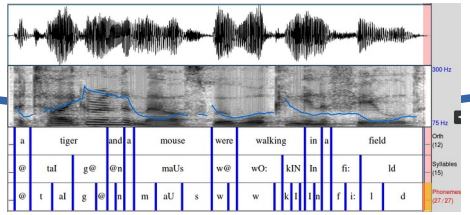


pitch pattern



outer ear middle ear inner ear

"intonation"



F0 height, F0 contour

Terminological points

Phonetic parameters

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

Linguistic categories

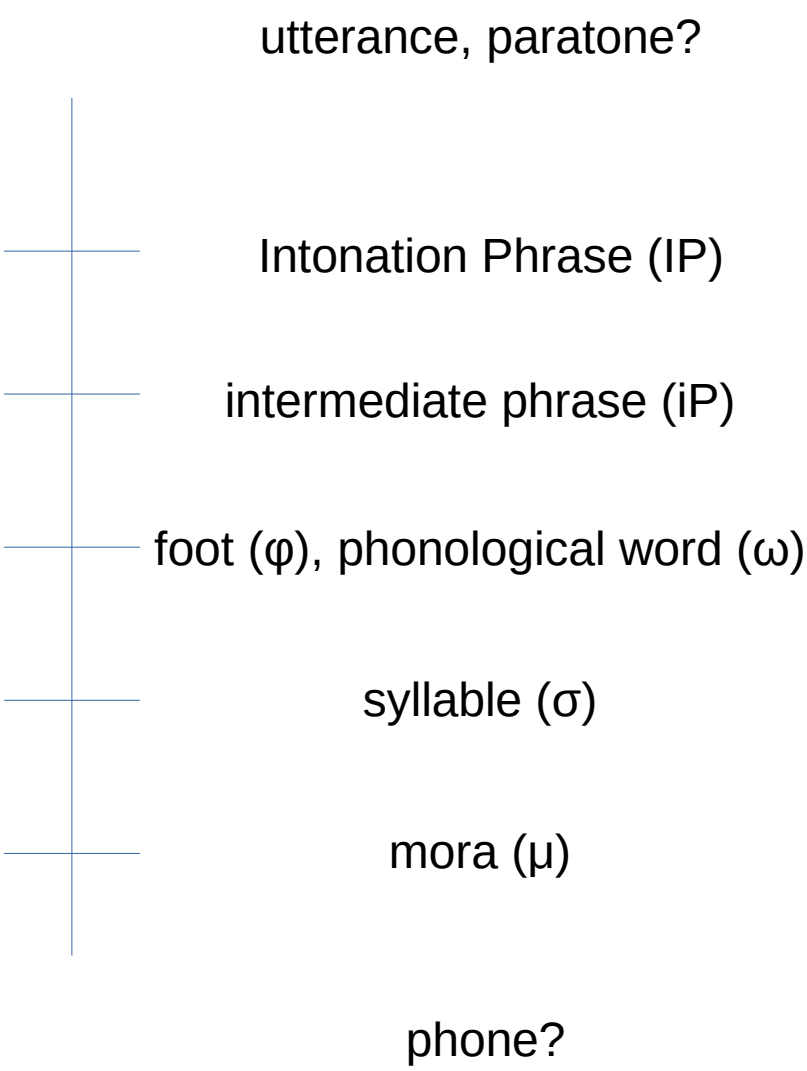
Lexical tone (phon / morph): Lexical pitch accent:	correlates with frequency pattern (as in Chinese) correlates with frequency pattern (as in Japanese)
Lexical stress:	designated ‘strong’ position in word (as in English) correlates with stress-pitch accent, duration, intensity patterns
Phrasal pitch accent:	correlates with frequency pattern (as in English
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:

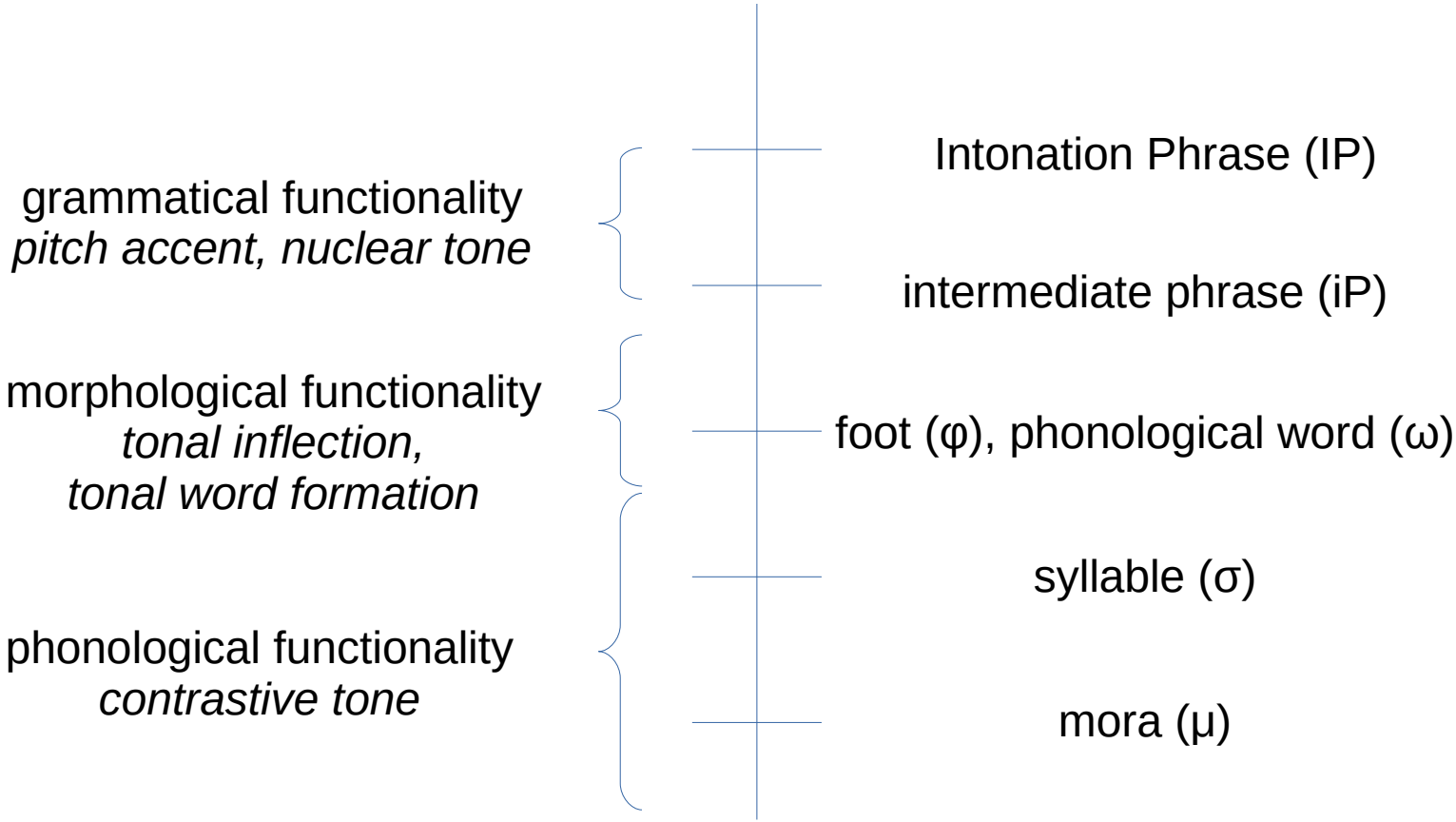
- in phonetics: ‘pitch’, ‘pitch tracking’ are sometimes used instead of ‘F0’, ‘F0 estimation’ for the acoustic correlate in speech transmission (cf. ‘pitch’ is used for ‘F0’ in Praat, for example)
- in phonology: sometimes ‘stress’ is used in both a phonological and a phonetic sense
- the term ‘pitch 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’

Prosodic hierarchies

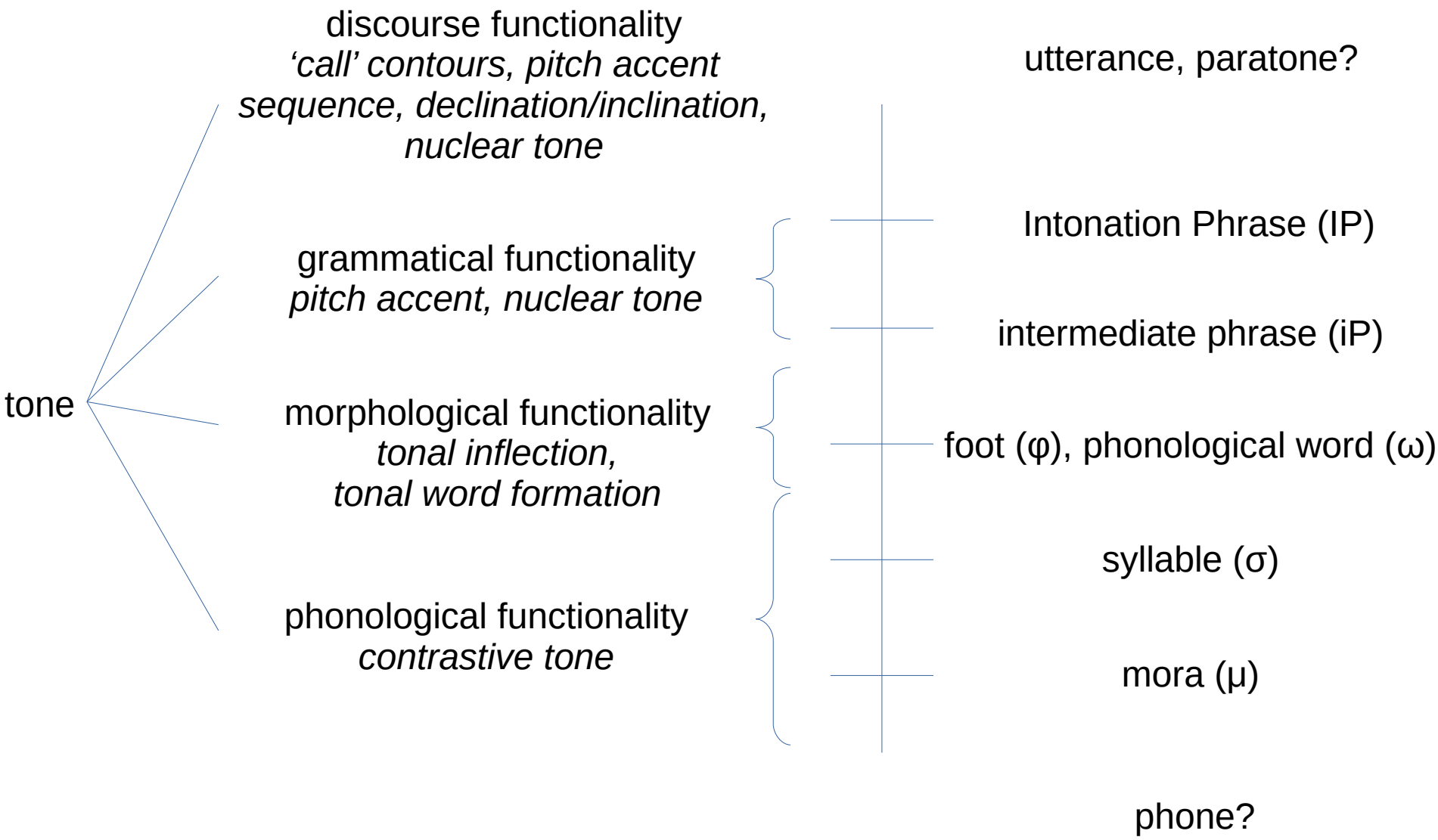
Prosodic hierarchies: a standard version



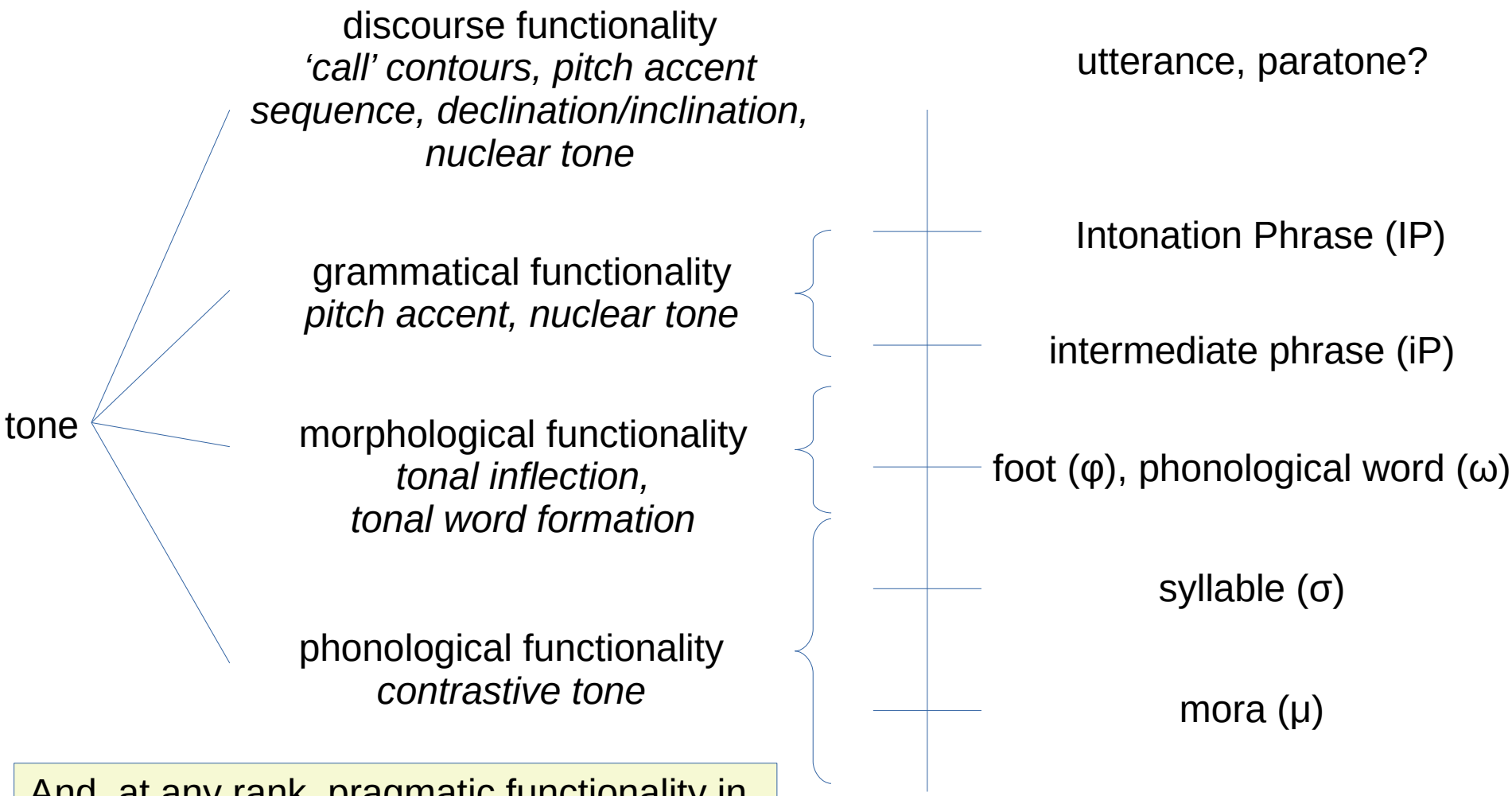
Prosodic hierarchies: a standard version



Prosodic hierarchies: a standard version



Prosodic hierarchies: a standard version



And, at any rank, pragmatic functionality in terms of *focus*, *contrast*, *emphasis*:

- pitch height, range
- intensity
- duration

RANK INTERPRETATION SEMIOTIC MODEL

OF LANGUAGE AND SPEECH

Rank Hierarchy:

Units, Categories, Functions

Interpretations:

Functional Interpretation

Modality Interpretation

- Auditory
- Visual

Semiotics and Prosody

Prosody is

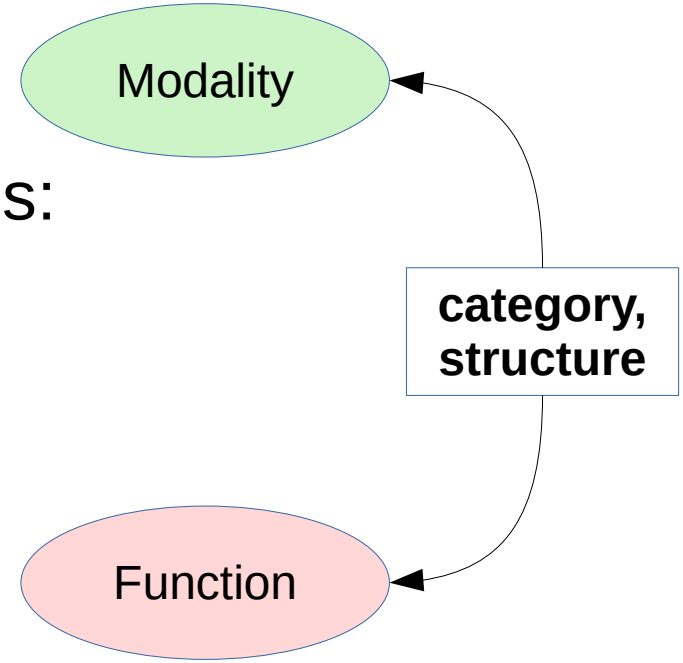
an independent sign system

with two main subsystems / channels:

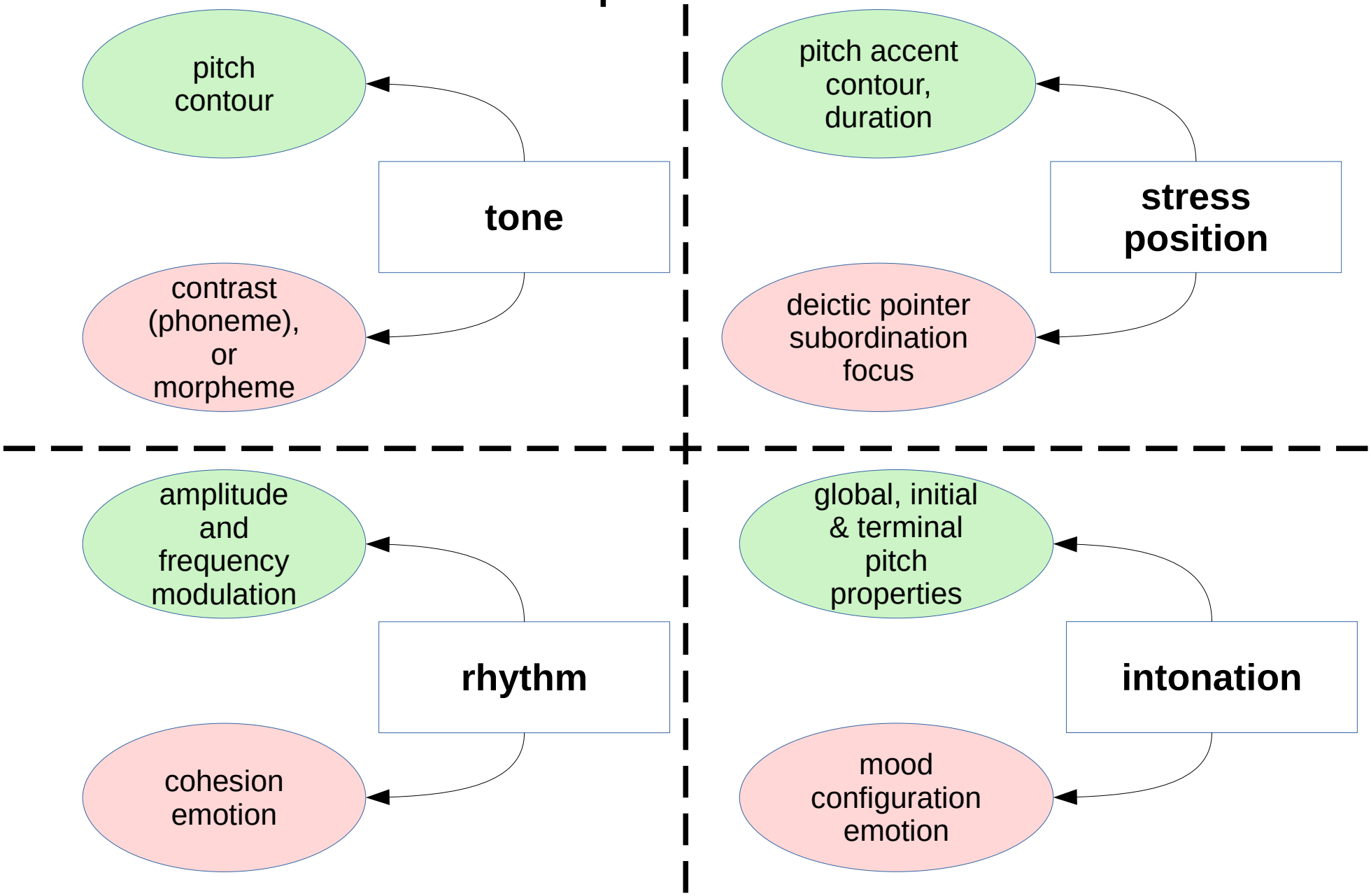
- rhythms
- melodies

with its own

- syntax:
  - linear and hierarchical patterns
- modality:
  - low frequency amplitude and frequency modulation of speech
  - layout, punctuation and highlighting hierarchy in writing
- functionality:
  - semantics: deictic pointing to associated words, phrases
  - pragmatics: attitudinal and emotional meanings

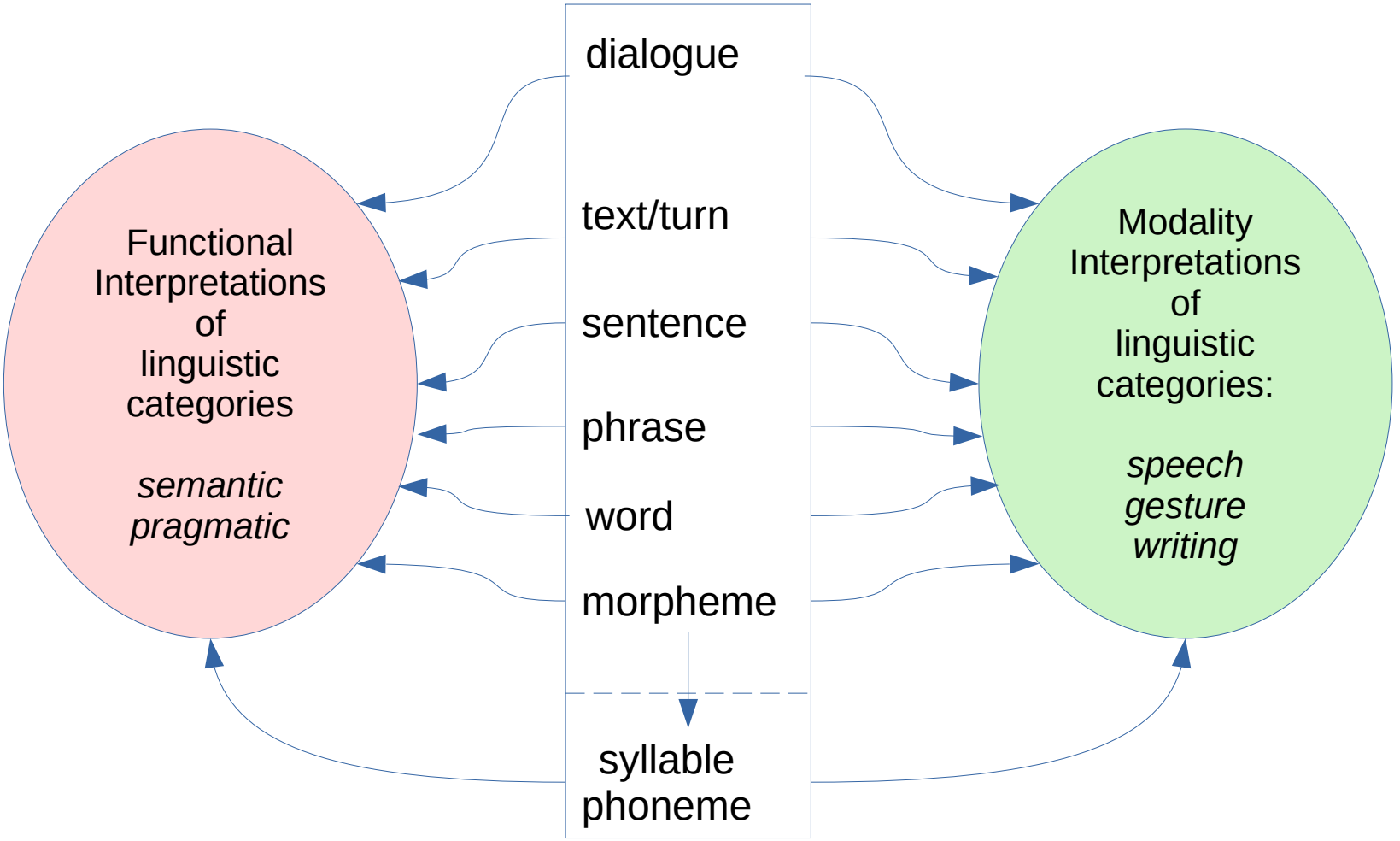


Semiotics of Prosody – four categories and their interpretations

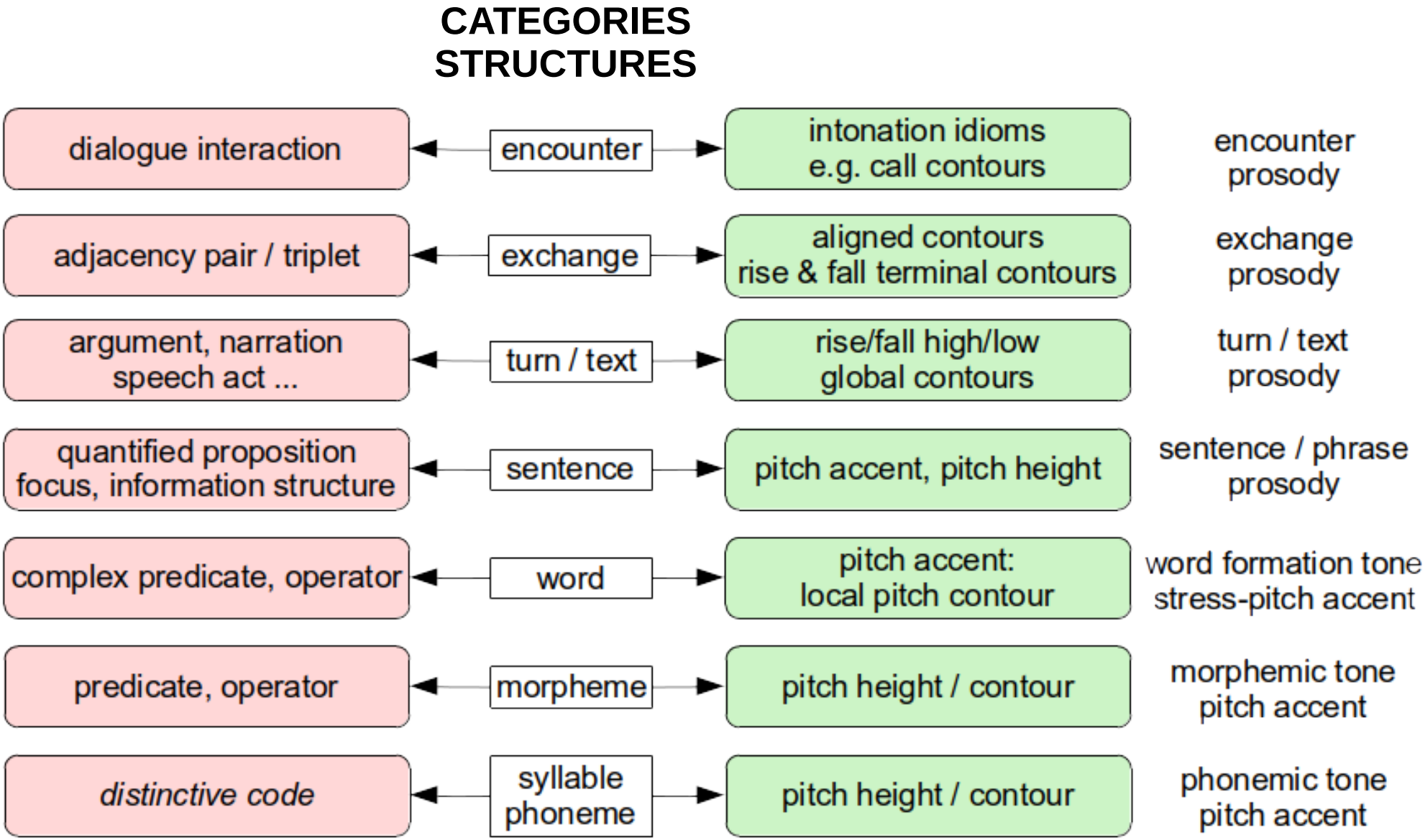


Rank Interpretation Architecture

1. Hierarchical ranks of signs
2. For each rank, its interpretations



Rank-Interpretation Architecture of Prosody



Quantitative methods

- Qualitative methods phonetics
  - epistemological basis of phonetics and linguistics
  - phonology
  - transcription and manual annotation
  - statistical analysis of annotation
  - signal processing and analysis

Quantitative methods

```
#!/usr/bin/python3
# basic_am_demod.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 medfilt
```

Quantitative methods

```
#!/usr/bin/python3
# basic_am_demod.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 medfilt
```

```
# Input WAV file
wavfilename = sys.argv[1]
pngfilestem = re.sub(".wav","",wavfilename)
appfilestem = re.sub(".py","",sys.argv[0])
samplerate, signal = wave.read(wavfilename)
period = 1/samplerate
siglen = len(signal)
sigsecs = siglen / samplerate
```

Quantitative methods

```
#!/usr/bin/python3
# basic_am_demod.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 medfilt
```

```
# Input WAV file
wavfilename = sys.argv[1]
pngfilestem = re.sub(".wav","",wavfilename)
appfilestem = re.sub(".py","",sys.argv[0])
samplerate, signal = wave.read(wavfilename)
period = 1/samplerate
siglen = len(signal)
sigsecs = siglen / samplerate
```

```
# Extract AM envelope, apply FFT
env = np.abs(signal)
mags = np.abs(np.fft.rfft(env))
freqs = np.fft.rfftfreq(env.size, period)

# Normalise data values for graphics
signal = signal / max(signal)
env = medfilt(env / max(env), 301)
mags = mags / max(mags)
```

Quantitative methods

```
#!/usr/bin/python3
# basic_am_demod.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 medfilt
```

```
# Input WAV file
wavfilename = sys.argv[1]
pngfilestem = re.sub(".wav","",wavfilename)
appfilestem = re.sub(".py","",sys.argv[0])
samplerate, signal = wave.read(wavfilename)
period = 1/samplerate
siglen = len(signal)
sigsecs = siglen / samplerate
```

```
# Extract AM envelope, apply FFT
env = np.abs(signal)
mags = np.abs(np.fft.rfft(env))
freqs = np.fft.rfftfreq(env.size, period)

# Normalise data values for graphics
signal = signal / max(signal)
env = medfilt(env / max(env), 301)
mags = mags / max(mags)
```

```
# Select LF spectrum
maxfreq = 6
maxsample = int(round(maxfreq*sigsecs))
mags = mags[int(round(sigsecs/2)):maxsample]
freqs = freqs[int(round(sigsecs/2)):maxsample]

# Generate graphics
x = np.linspace(0, sigsecs, siglen)
plt.subplot(3,1,1)     # waveform
plt.plot(x, signal)
plt.subplot(3,1,2)     # env
plt.plot(x, env)
plt.subplot(3,1,3)     # env
plt.plot(freqs, mags)

# Save and show graphics
plt.savefig("%s_%s.png"%(appfilestem, pngfilestem))
plt.show()
```

https://github.com/dafyddg/RFA/tree/main/LittleHelpers

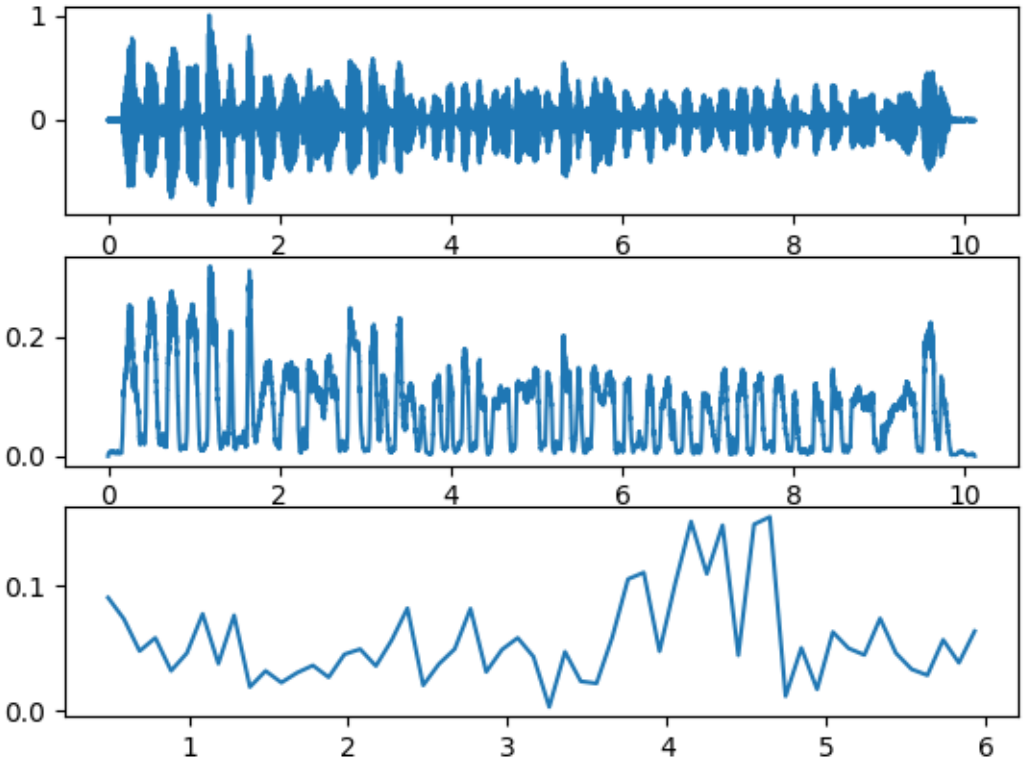
Quantitative methods

```
#!/usr/bin/python3
# basic_am_demod.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as
import matplotlib
import scipy.io.w
from scipy.signa
```

```
# Input WAV file
wavfilename = s
pngfilestem = re
appfilestem = re
samplerate, sign
period = 1/samp
siglen = len(sign
sigsecs = siglen
```

```
# Extract AM en
env = np.abs(sig
mags = np.abs(r
freqs = np.fft.rfftf

# Normalise data values for graphics
signal = signal / max(signal)
env = medfilt(env / max(env), 301)
mags = mags / max(mags)
```

```
# Select LF spectrum
maxfreq = 6
maxsample = int(round(maxfreq*sigsecs))
mags = mags[int(round(sigsecs/2)):maxsample]
freqs = freqs[int(round(sigsecs/2)):maxsample]
```



https://github.com/dafyddg/RFA/tree/main/LittleHelpers

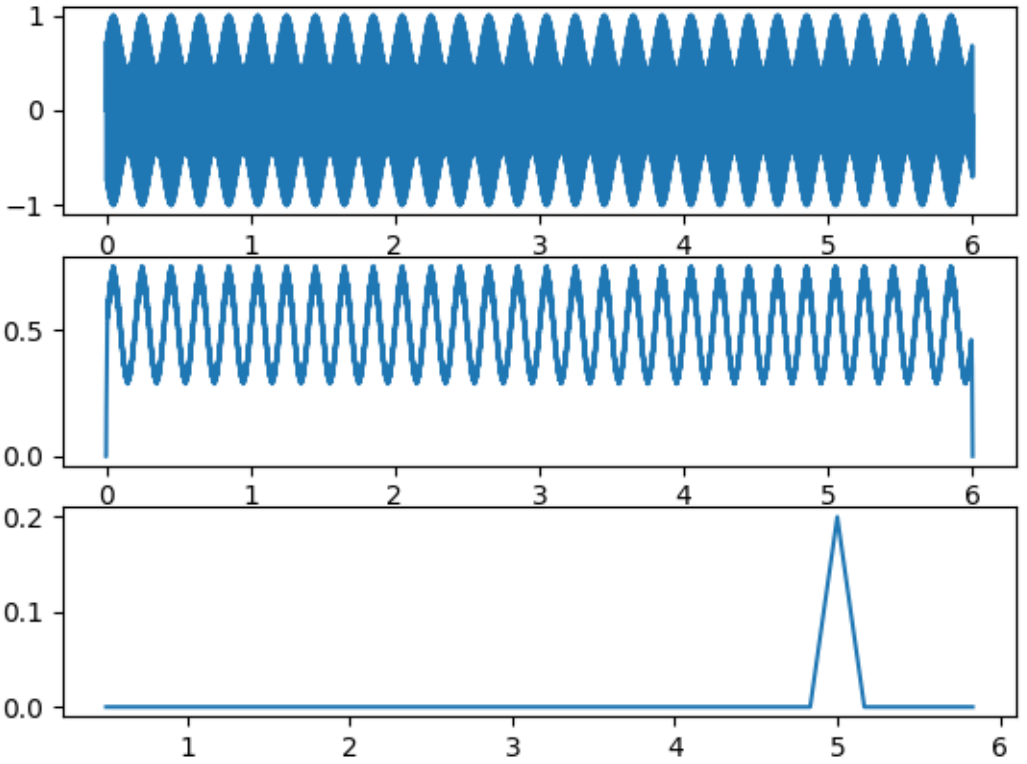
Quantitative methods

```
#!/usr/bin/python3
# basic_am_demod.py
# D. Gibbon, 2021-03-16

import sys, re
import numpy as
import matplotlib
import scipy.io.w
from scipy.signa
```

```
# Input WAV file
wavfilename = s
pngfilestem = re
appfilestem = re
samplerate, sign
period = 1/samp
siglen = len(sign
sigsecs = siglen
```

```
# Extract AM en
env = np.abs(sig
mags = np.abs(r
freqs = np.fft.rfftf

# Normalise data values for graphics
signal = signal / max(signal)
env = medfilt(env / max(env), 301)
mags = mags / max(mags)
```

```
# Select LF spectrum
maxfreq = 6
maxsample = int(round(maxfreq*sigsecs))
mags = mags[int(round(sigsecs/2)):maxsample]
freqs = freqs[int(round(sigsecs/2)):maxsample]
```



https://github.com/dafyddg/RFA/tree/main/LittleHelpers

BREAK, 30 minutes

Tasks:

1) get Praat ready,
2) locate http://wwwhomes.uni-bielefeld.de/gibbon/TGA/