Prosody: Thinking Outside the Box

Lecture 3

The Phonetics of Prosody 2: Melody

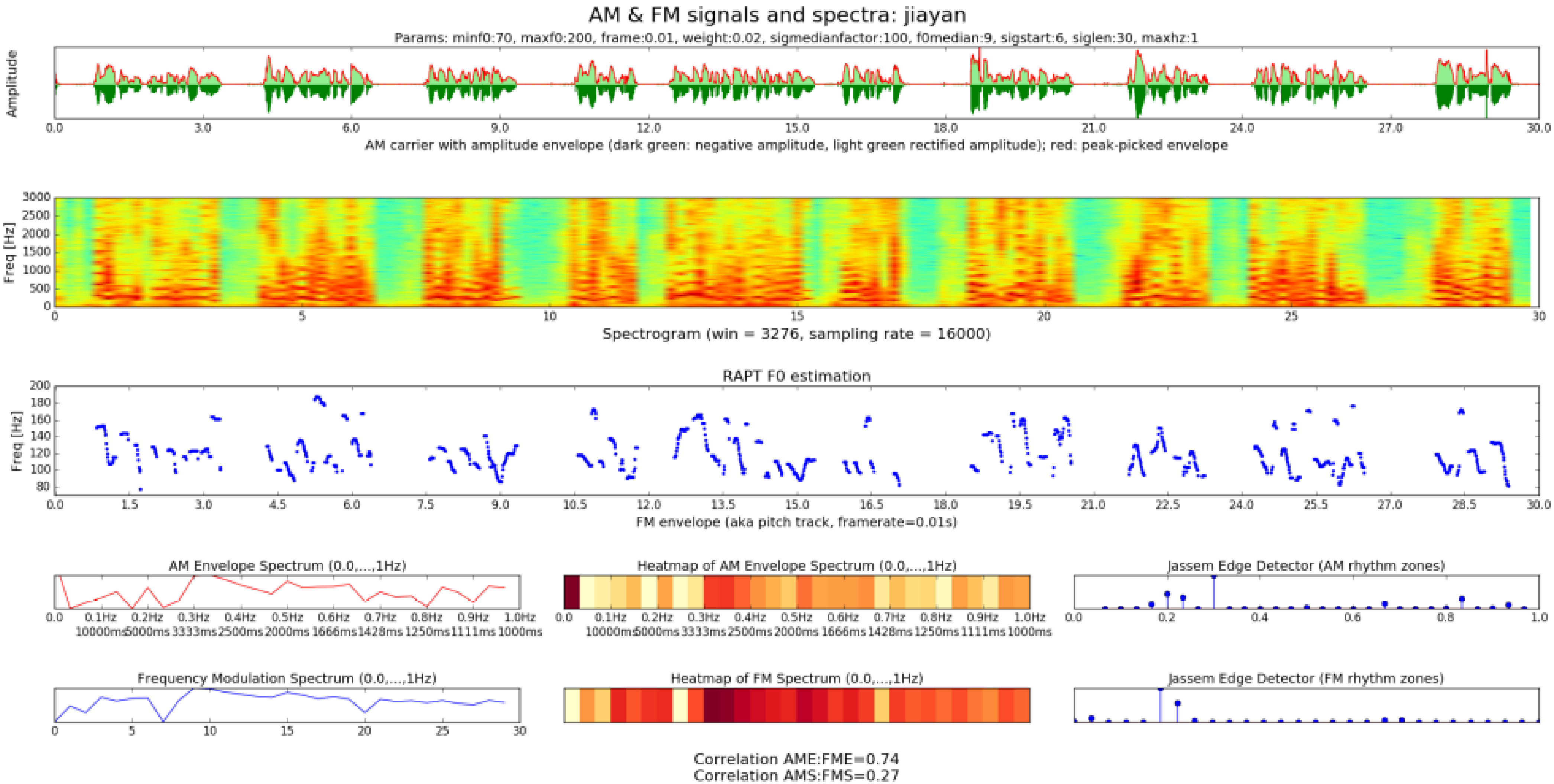
Dafydd Gibbon

Bielefeld University

Fudan University Summer School: Contemporary Phonetics and Phonology
Shanghai, 7–13 July 2018

Observations – Measurements – Models

Models of Prosody

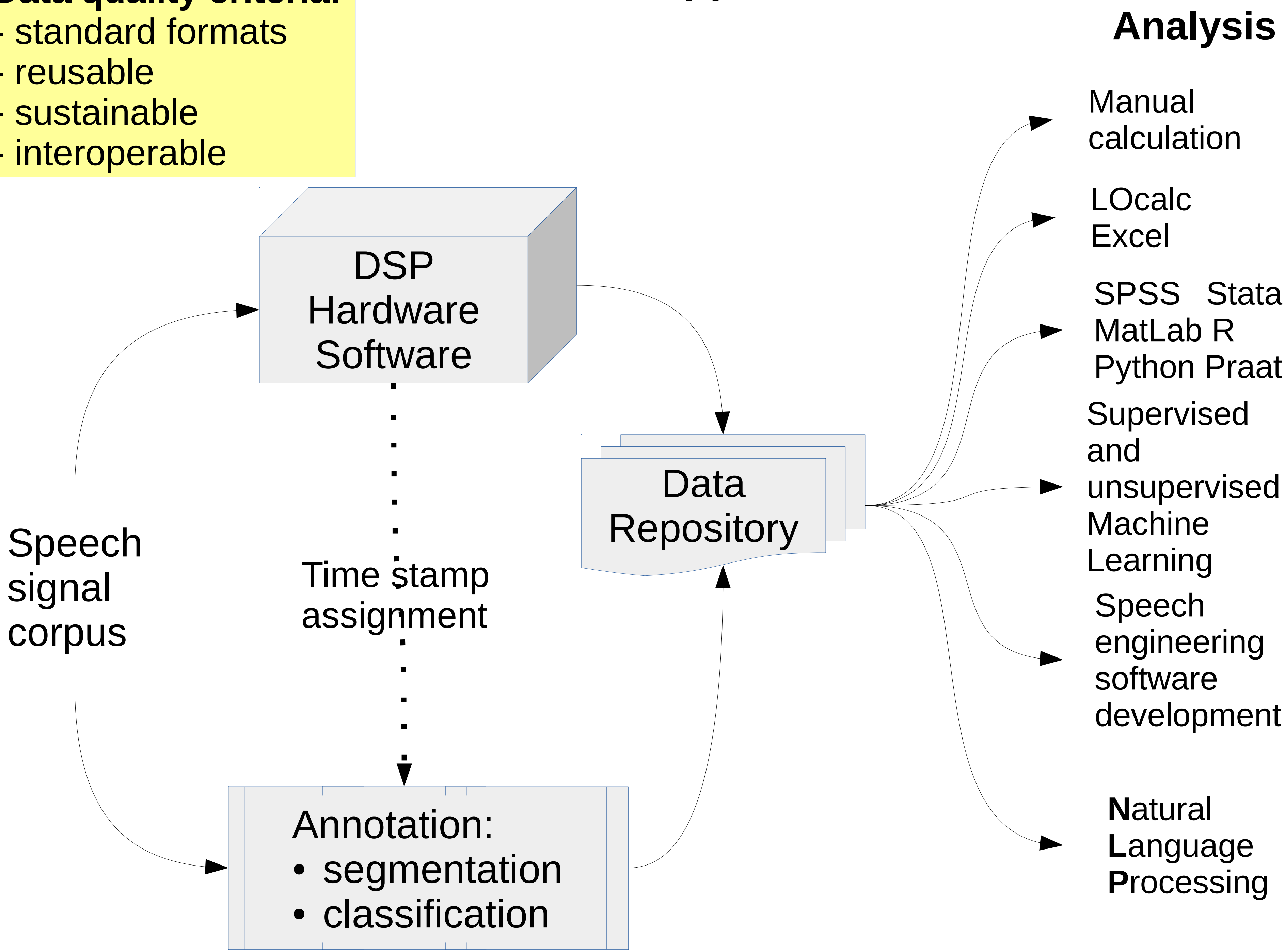


The Phonetics of Prosody: Melody – Overview

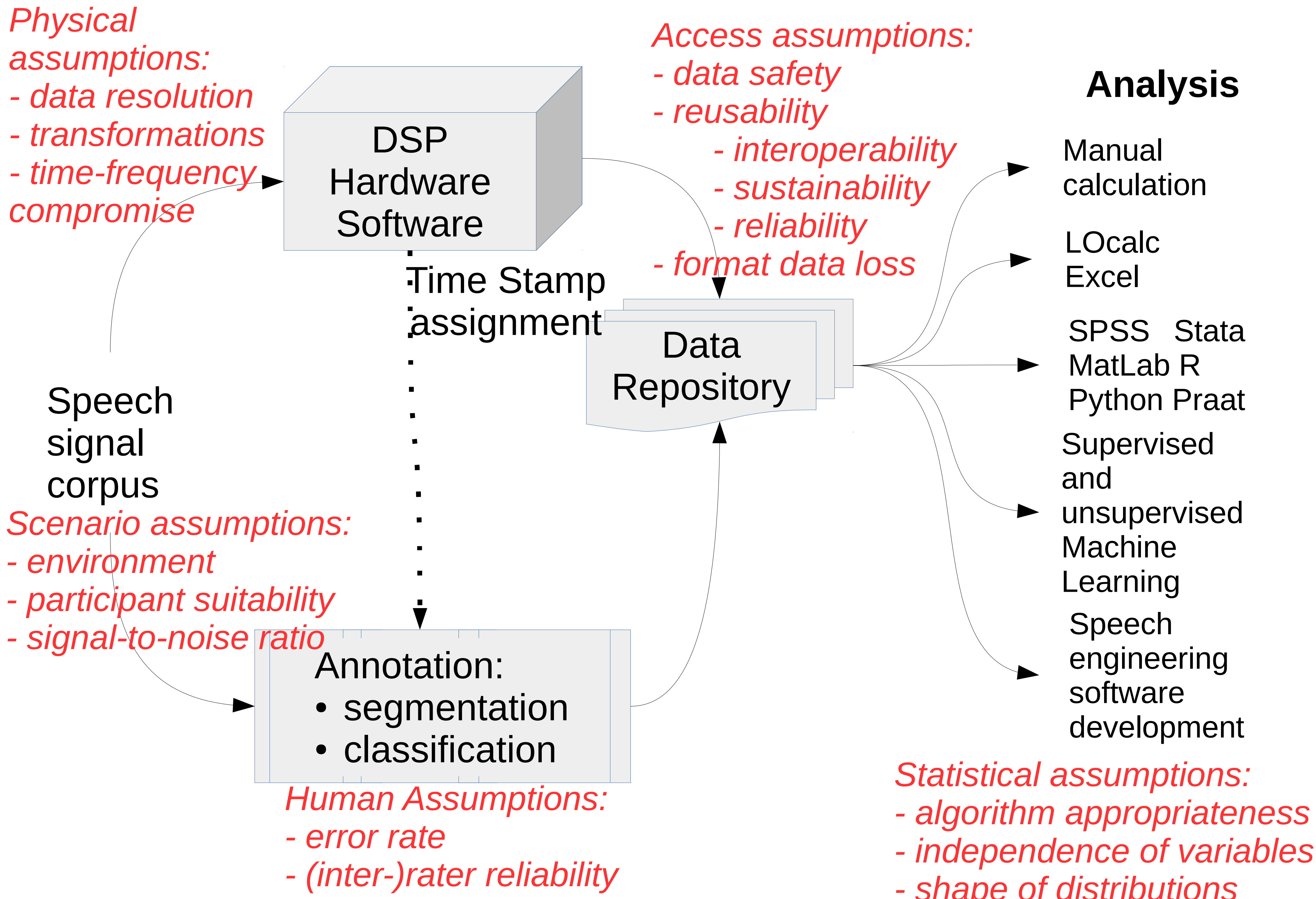
1. Orientation
   - Data as a valuable resource
2. The physiology of melody
   - production: the larynx
   - perception: the cochlea
3. The physics of melody
   - amplitude and frequency modulation
   - frequency processing
     - frequency modulation and demodulation
4. F0 estimation (F0 / pitch; detection / extraction / tracking)
   - time domain: period measurement – peak-picking, zero-crossings
   - frequency domain: harmonic difference measurement
5. Modelling melody: from discourse to phoneme

Data-driven approaches

Data quality criteria:
- standard formats
- reusable
- sustainable
- interoperable



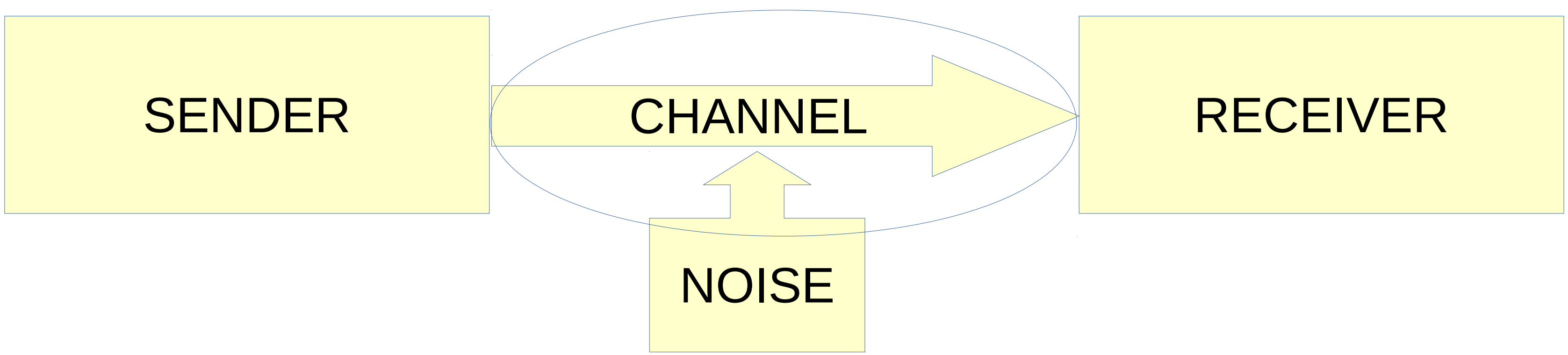
Data-driven approaches



The Domains of Melody

Phonetic Subdomains as Time Phases

1. Speaker: production, articulatory phonetics:
   - articulation rate - effort
2. Channel: acoustic phonetics:
   - fundamental frequency - intensity
3. Hearer: reception, auditory phonetics:
   - pitch - loudness

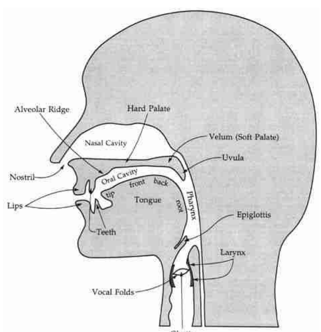


The Domains of Phonetics: the Phonetic Cycle

A tiger and a mouse were walking in a field ...

The Domains of Phonetics: the Phonetic Cycle

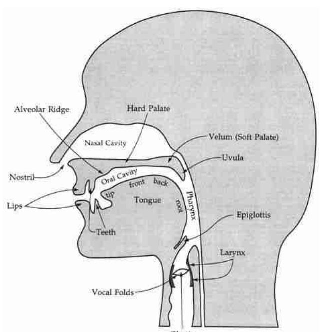
Sender:
Articulatory
Phonetics



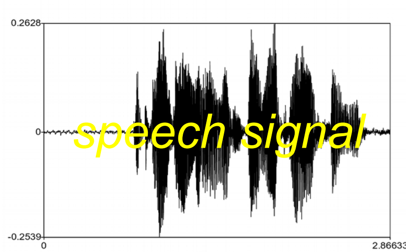
A tiger and a mouse were walking in a field ...

The Domains of Phonetics: the Phonetic Cycle

Sender:
Articulatory
Phonetics

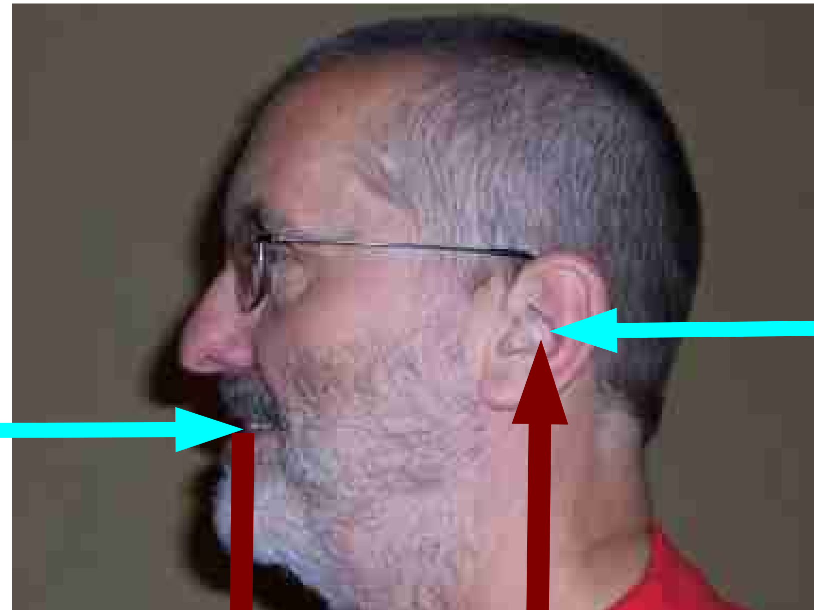


A tiger and a mouse were walking in a field ...

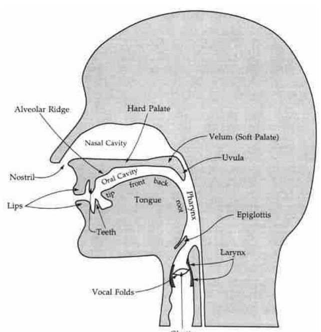


Channel:
Acoustic
Phonetics

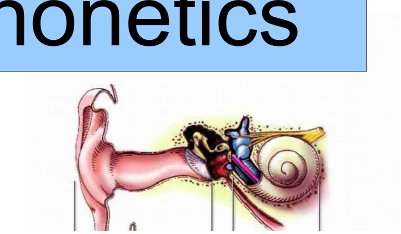
The Domains of Phonetics: the Phonetic Cycle



Sender:
Articulatory
Phonetics

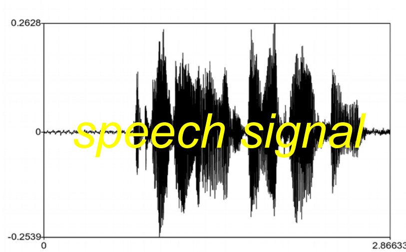


Receiver:
Auditory
Phonetics



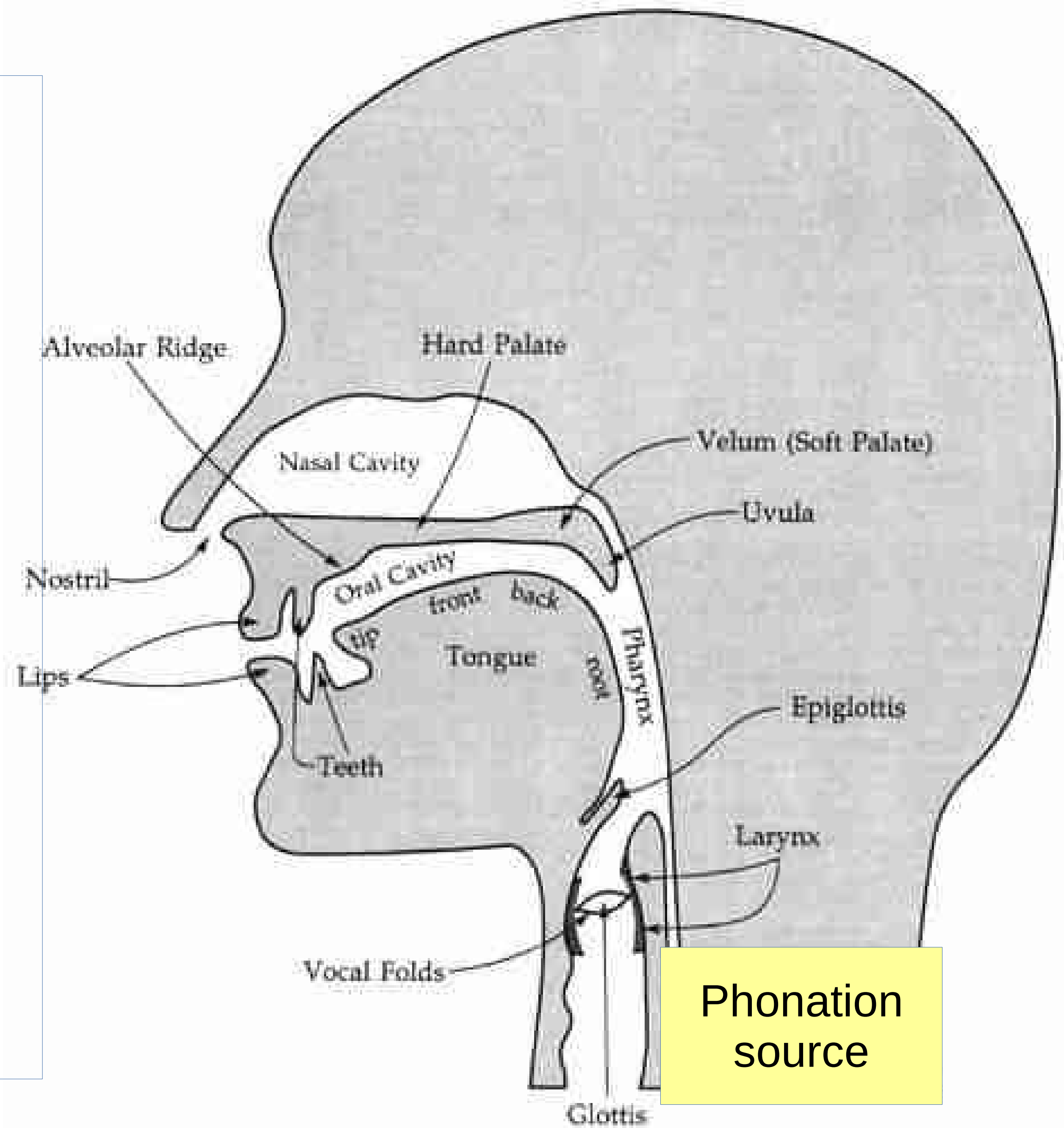
A tiger and a mouse were walking in a field ...

Channel:
Acoustic
Phonetics

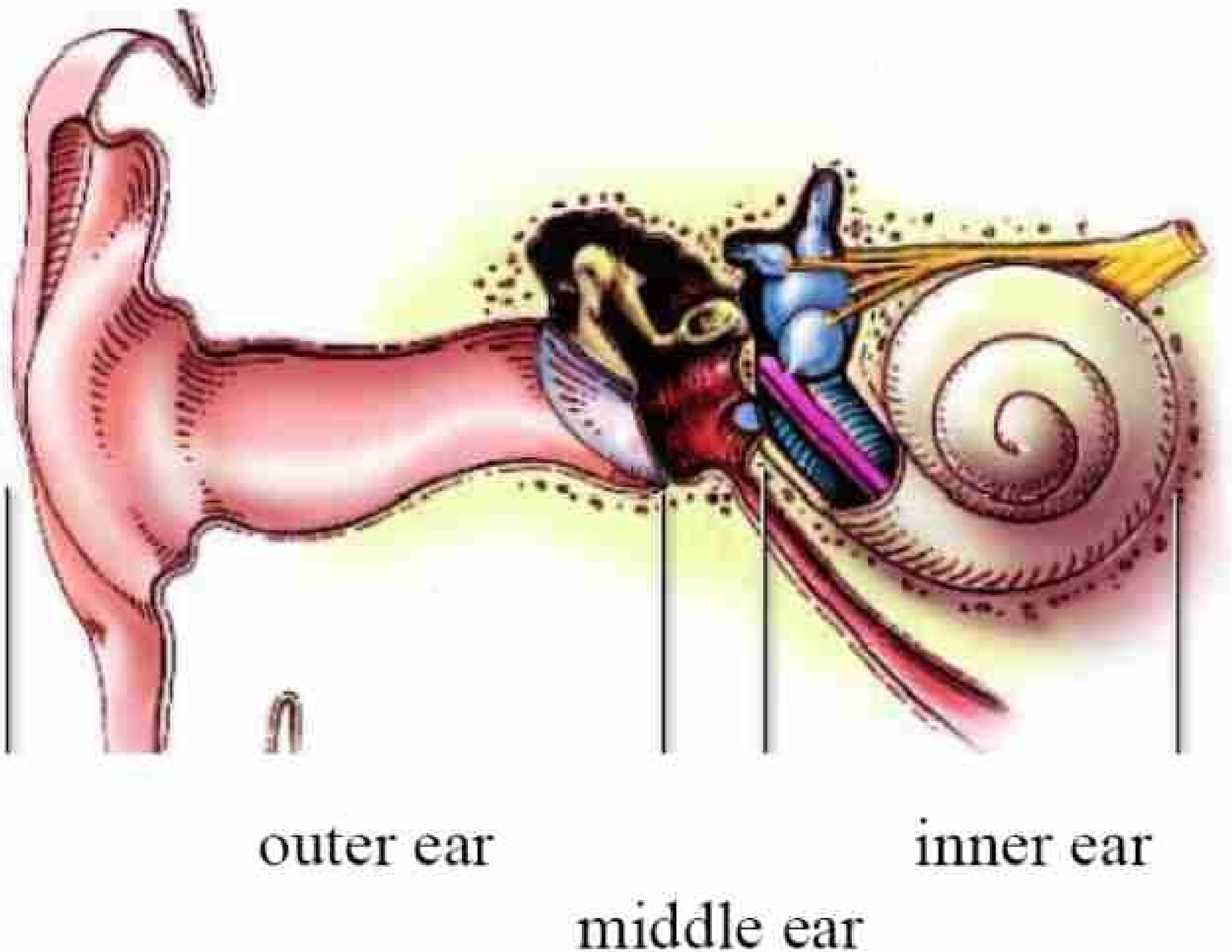


The articulatory domain

1. Domain of speech production
2. Articulatory organs are relatively easily observable
3. Domain of reference for phonetic categories of the IPA
4. Investigated via
   - corpus creation
   - experiment paradigm



The Auditory Domain

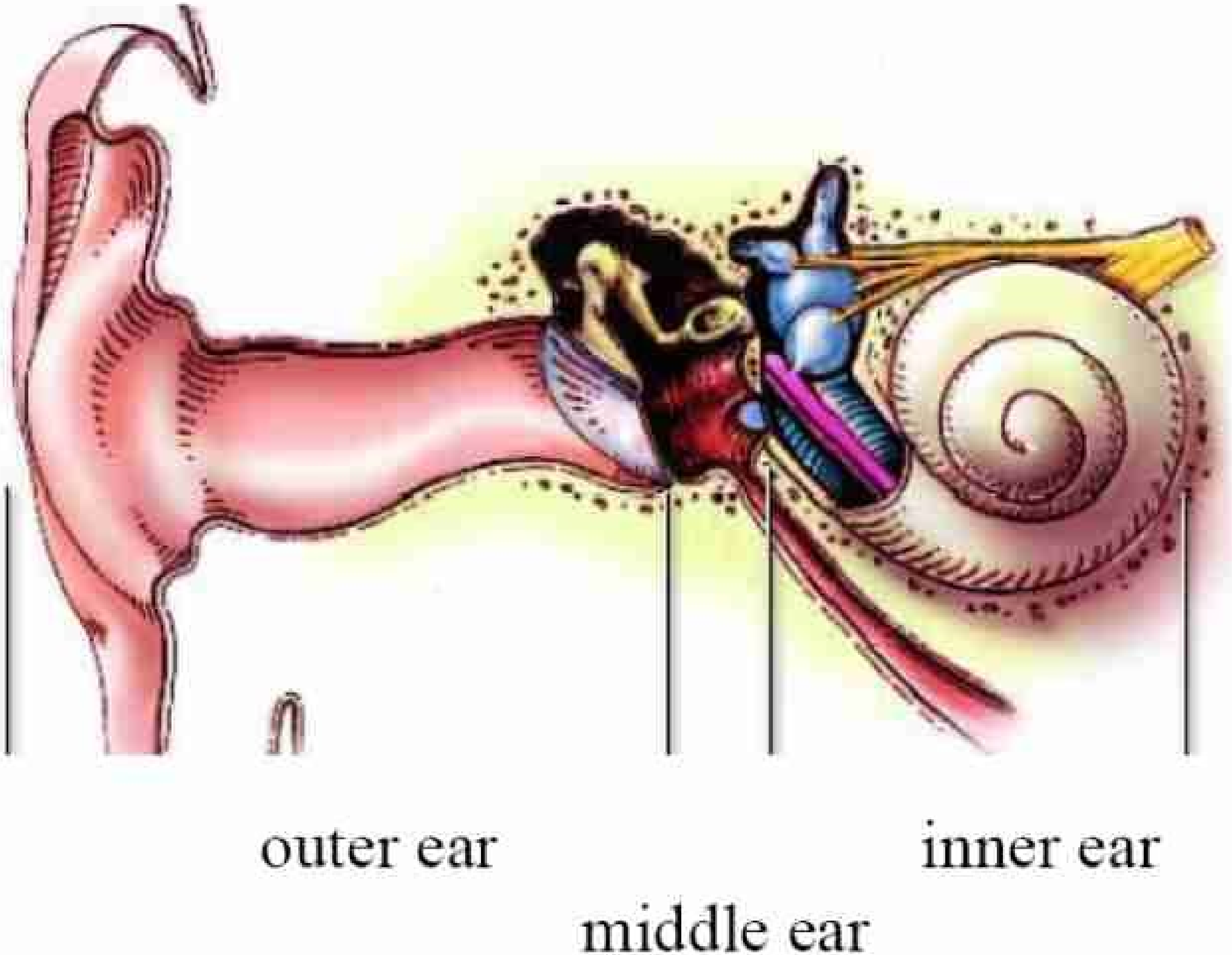


The Auditory Domain: Anatomy of the Ear

microphone

amplifier

Fourier transform



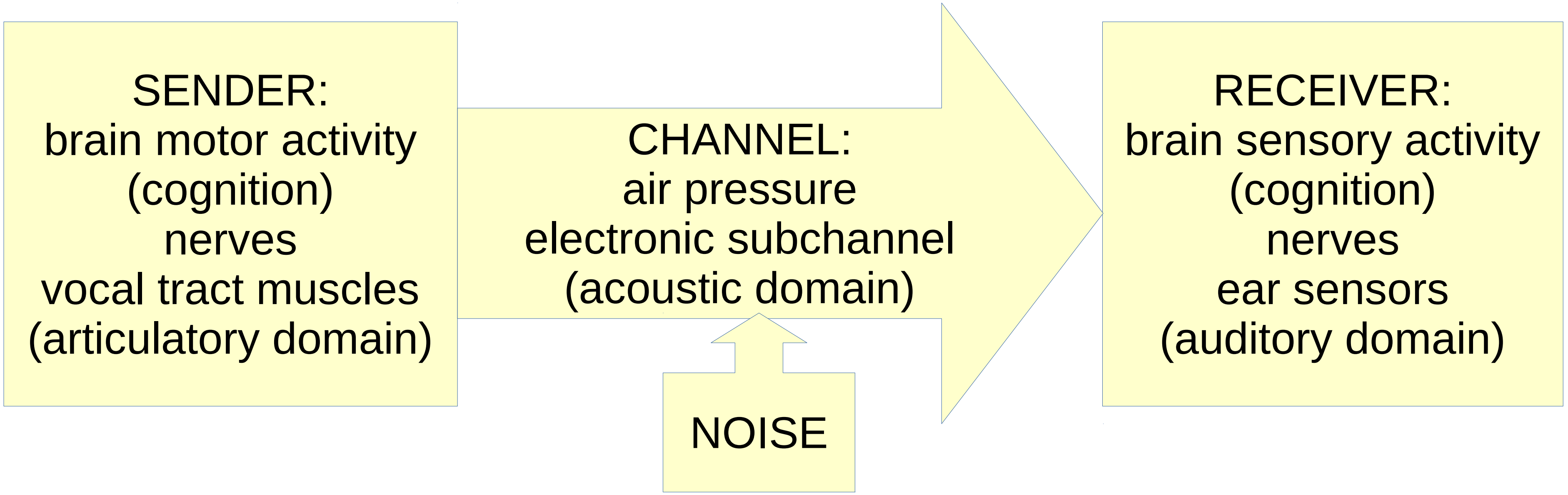
The Phonetic Cycle

Each of the phases has subphases:

→ brain motor activity → nerves → vocal tract muscles

→ air pressure → (electronic channel →) air pressure

→ ear sensors – nerves – brain sensory activity

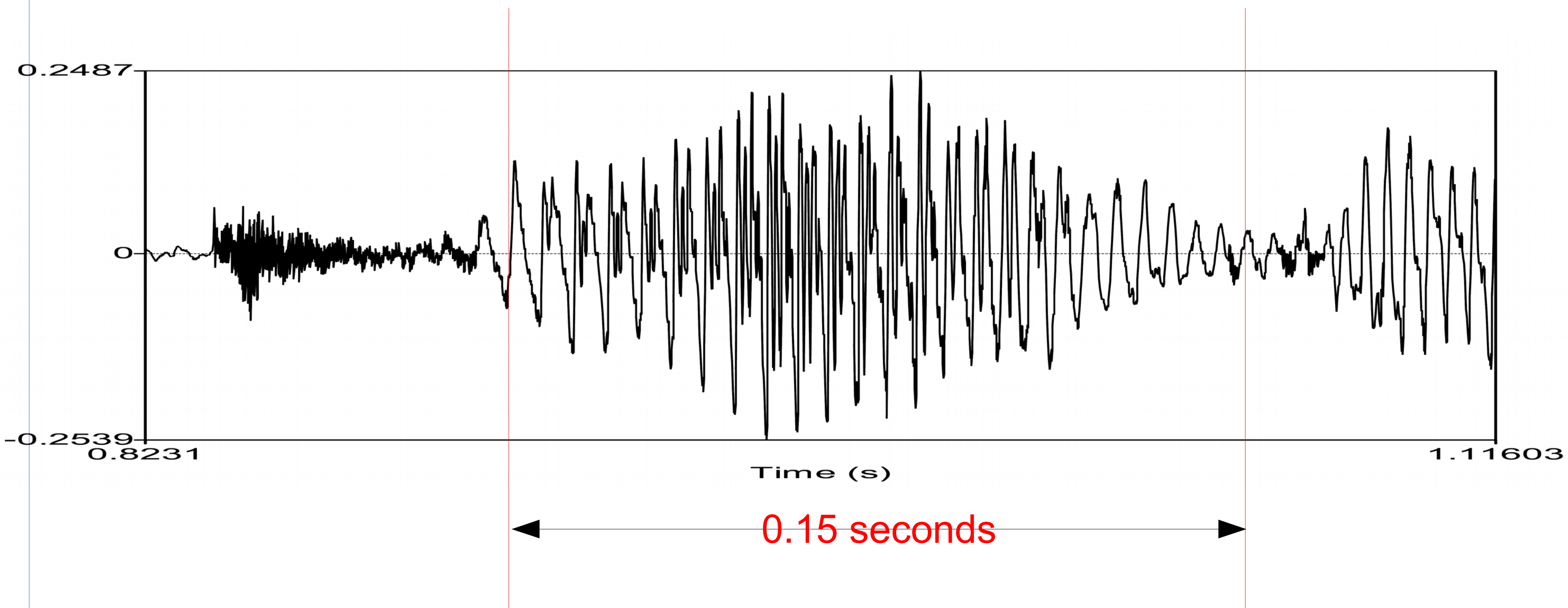


The Acoustic Signal

1. The *period* or *interval* of a single wave in a speech signal is the duration of this single wave.
   - A *resonant signal* is a signal whose periods are regular, i.e. even in duration.
   - A signal is *noisy* if the periods are irregular, i.e. uneven in duration
   - The average period of a speech signal
2. The *wavelength* λ (lambda) in cm of a speech signal is the speed of sound in cm/sec divided by the number of periods per second.
   - *You can forget the definition of wavelength...*
   - *A task for the very interested:*
     - *What is the speed of sound?*
     - *What is the wavelength of a sound with 100 periods per second?*

Time Domain and Frequency Domain

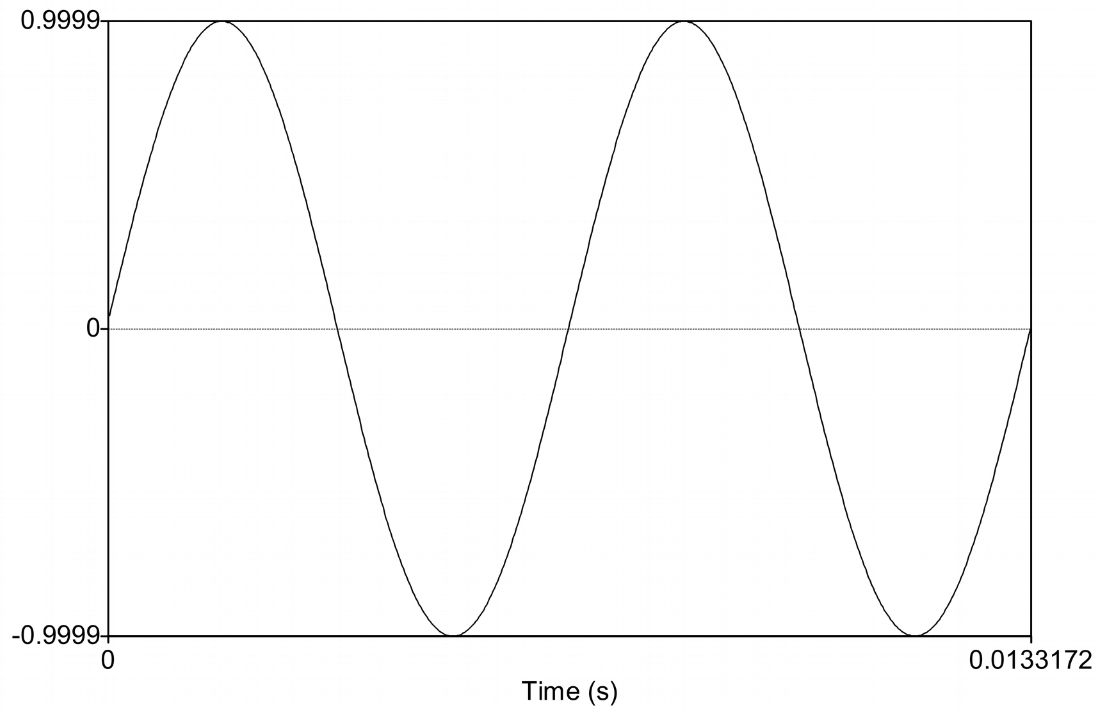
1. The *frequency* of a speech signal is the number of waves (periods) per second in the waveform



1. Question:
   - Ignoring the irregularities in the waveforms: what is the average frequency of the segment between the red lines?

Sine Waves

A sine wave of frequency F is produced by an evenly swinging pendulum (a very slow sine wave, of course).



The speech signal is not a simple sine wave, however, but a complex signal.

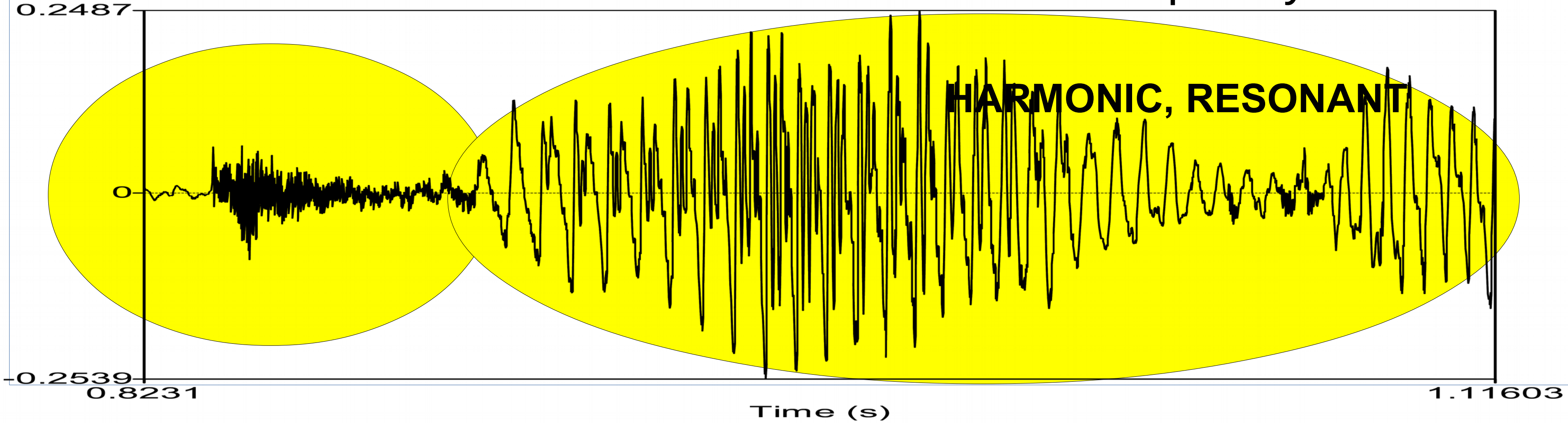
The Frequency Structure of Speech

The SOURCE

- for harmonic, voiced sounds
- is the larynx.
- The larynx produces
- a complex waveform, consisting of
  - a fundamental frequency
    - about 80 Hz - 200 Hz for men
    - about 160 Hz - 300 Hz for women
  - many overtones, which are audible up to about 20 kHz
  - different intensities of the overtones, relative to each other, determines the overall waveform, and therefore the kind of sound which the source produces
  - during voicing, the larynx generates a waveform which is rather like a “sawtooth” sequence

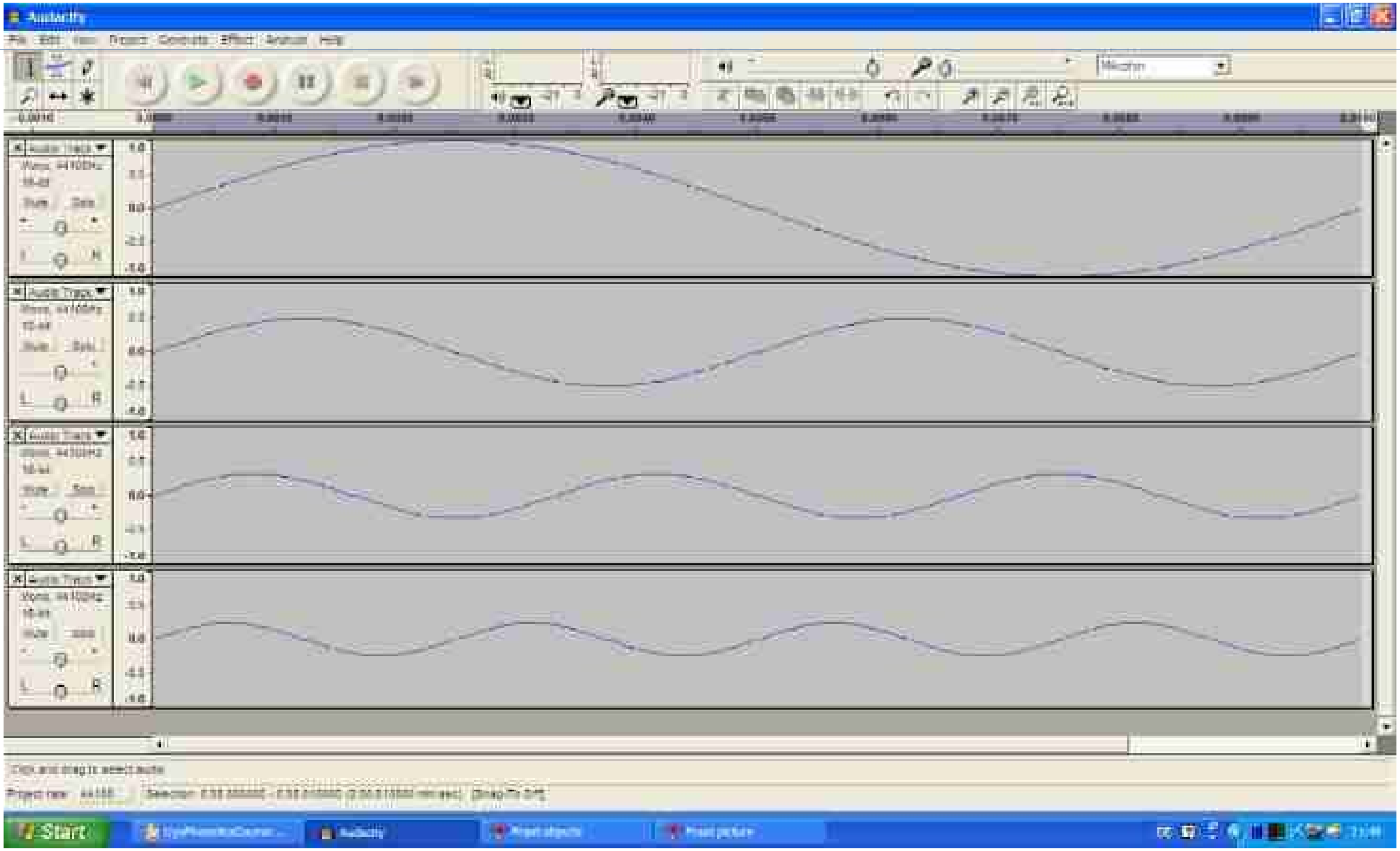
Complex Sources: noisy & harmonic signals

1. If many sine waves of arbitrary frequencies occur together, the result is NOISE.
2. If many sine waves occur together, with each being an integer multiple of some lowest frequency,
   - the resulting overall wave is a HARMONIC wave:
   - the lowest frequency of a harmonic waveform is the *fundamental frequency*, F0 (f-zero, f-nought)
   - the higher frequencies in a harmonic waveform are called the *harmonics* or *overtones* of the fundamental frequency



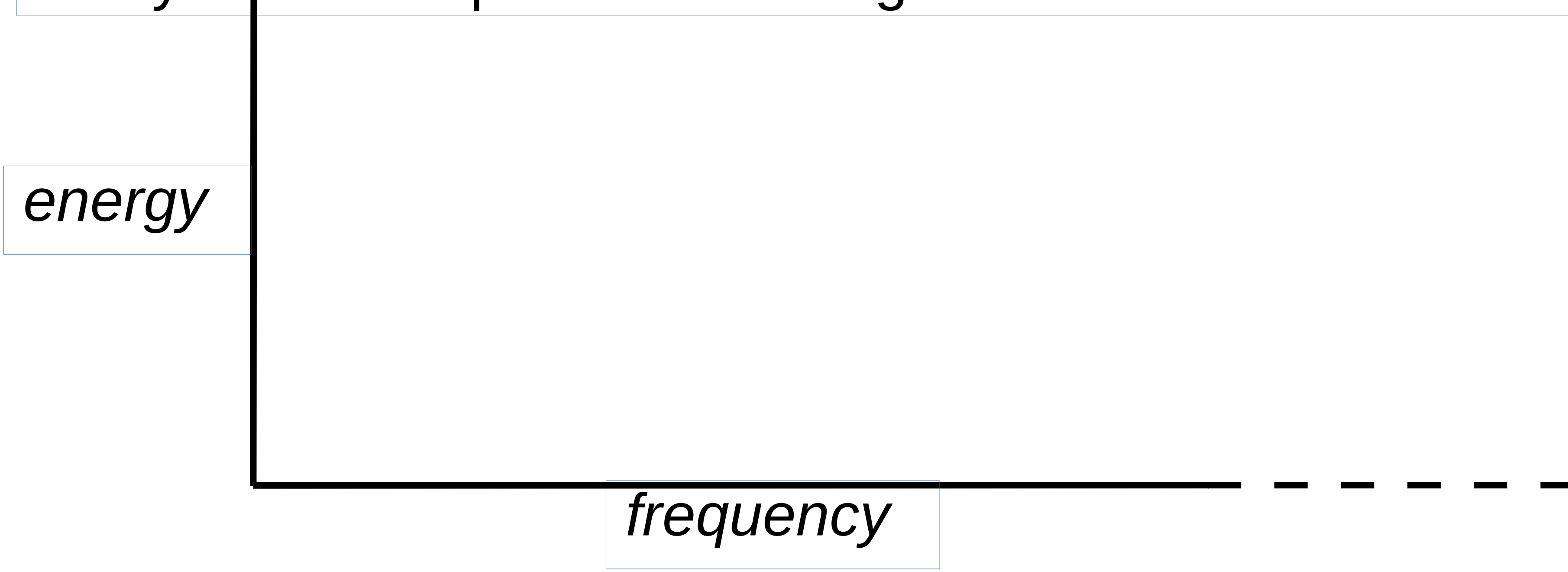
Sources with Integer Multiples of Sine Waves

1. Harmonic, resonant frequencies are created by adding several sine waves together, point by point
2. The larynx sound source is a special case of this



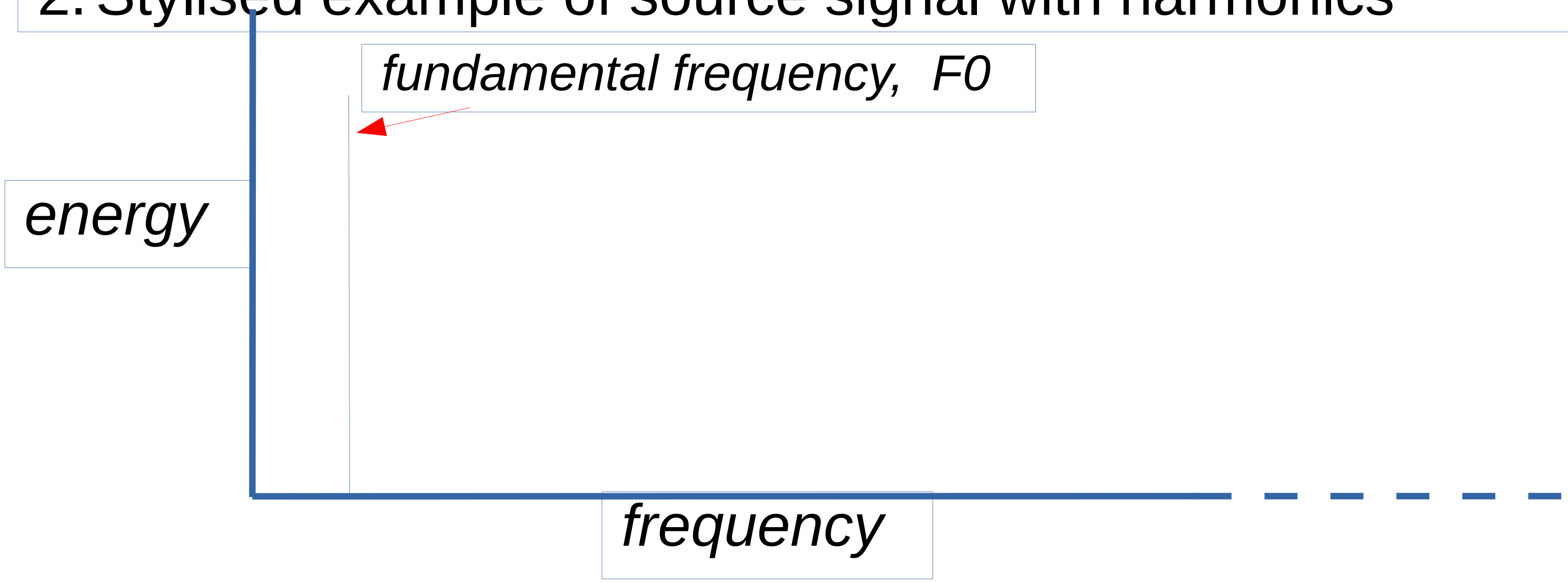
Harmonics / overtones in complex signals

1. If a complex signal consists of
   - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
     - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
   - then the signal is a resonant signal
   - and *f* is the *fundamental frequency* F0
   - while 2*f*, 3*f*, ..., *nf* are harmonics of the fundamental frequency
2. Stylised example of source signal with harmonics



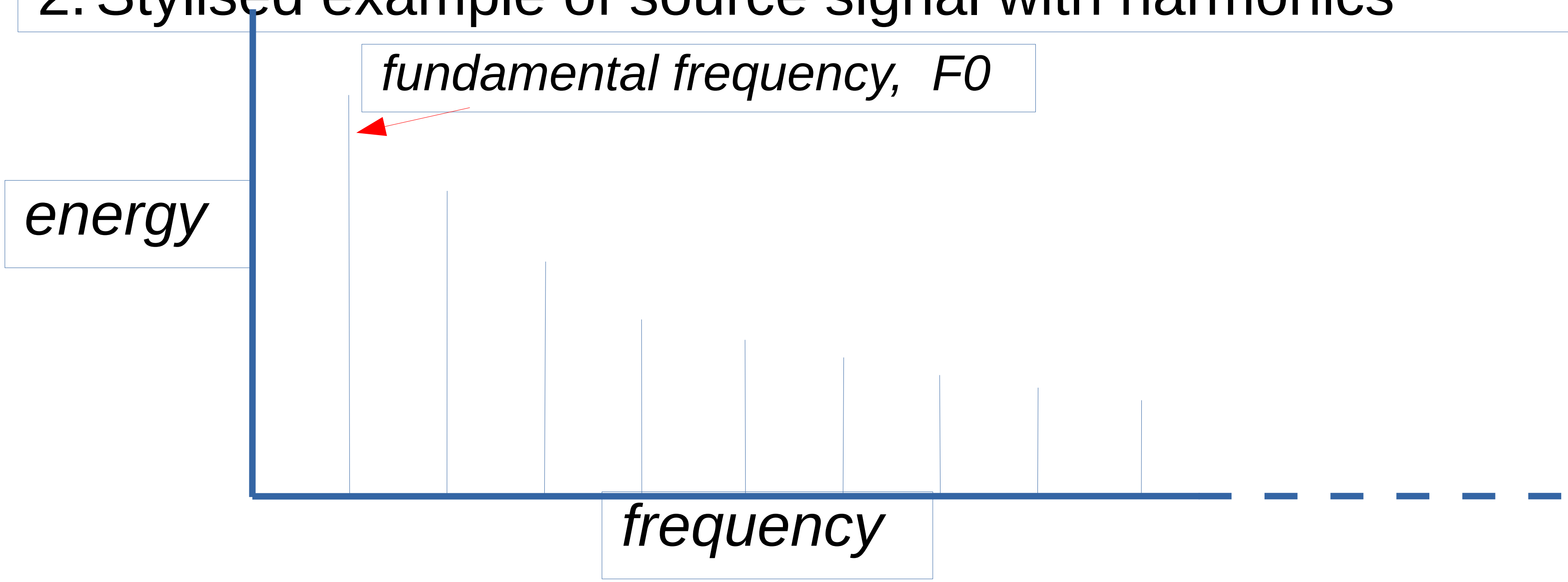
The Spectrum of Complex Signals

1. If a complex signal consists of
   - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
     - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
   - then the signal is a resonant signal
   - and *f* is the *fundamental frequency* F0
   - while 2*f*, 3*f*, ..., *nf* are *harmonics* of the fundamental frequency
2. Stylised example of source signal with harmonics



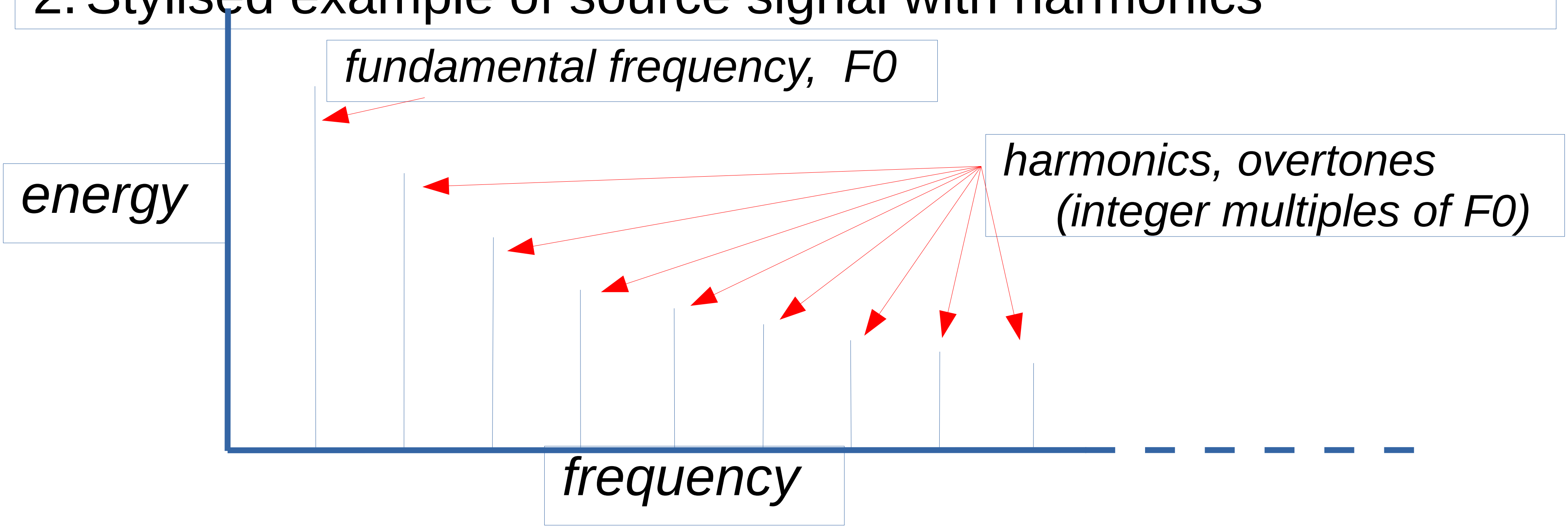
The Spectrum of Complex Signals

1. If a complex signal consists of
   - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
     - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
   - then the signal is a resonant signal
   - and *f* is the *fundamental frequency* F0
   - while 2*f*, 3*f*, ..., *nf* are harmonics of the fundamental frequency
2. Stylised example of source signal with harmonics



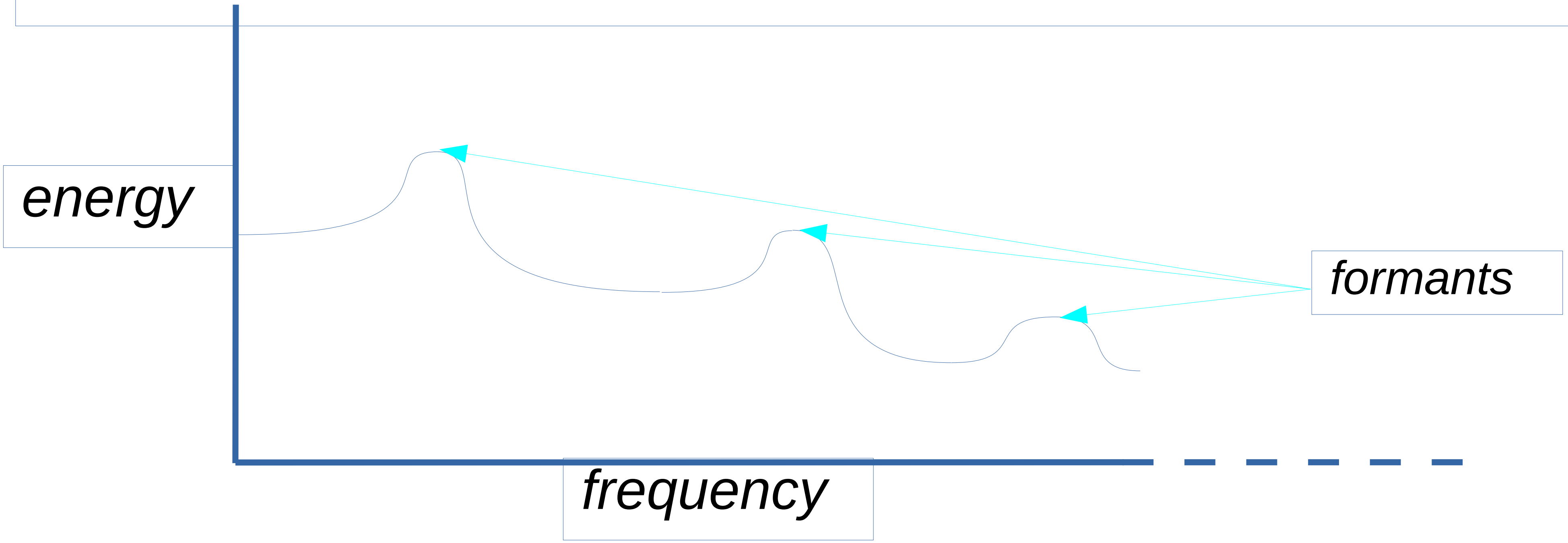
The Spectrum of Complex Signals

1. If a complex signal consists of
   - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
     - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
   - then the signal is a resonant signal
   - and *f* is the *fundamental frequency* F0
   - while 2*f*, 3*f*, ..., *nf* are harmonics of the fundamental frequency
2. Stylised example of source signal with harmonics



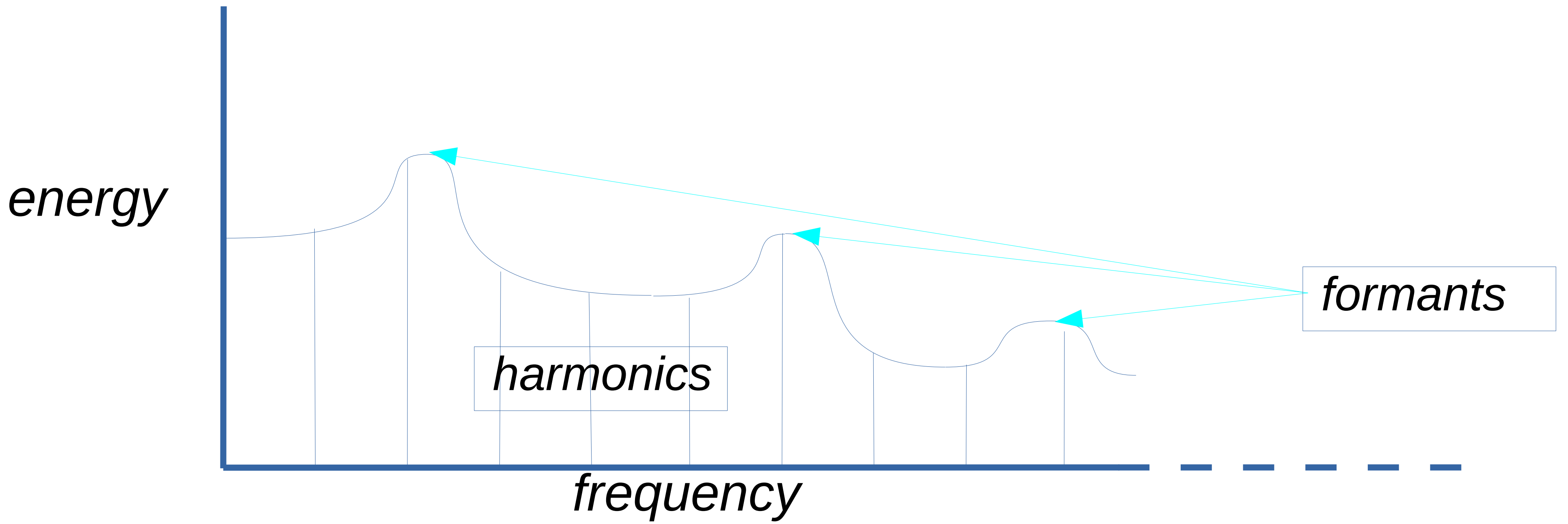
The Spectrum of Complex Signals

1. The filter system consists of the pharyngeal, nasal and oral cavities, which have cavities have specific resonant frequencies
2. These filter frequency bands are called *formants*
3. Formant frequencies of the oral cavity can be modified by the variable filters (articulators *tongue* and *lips*)



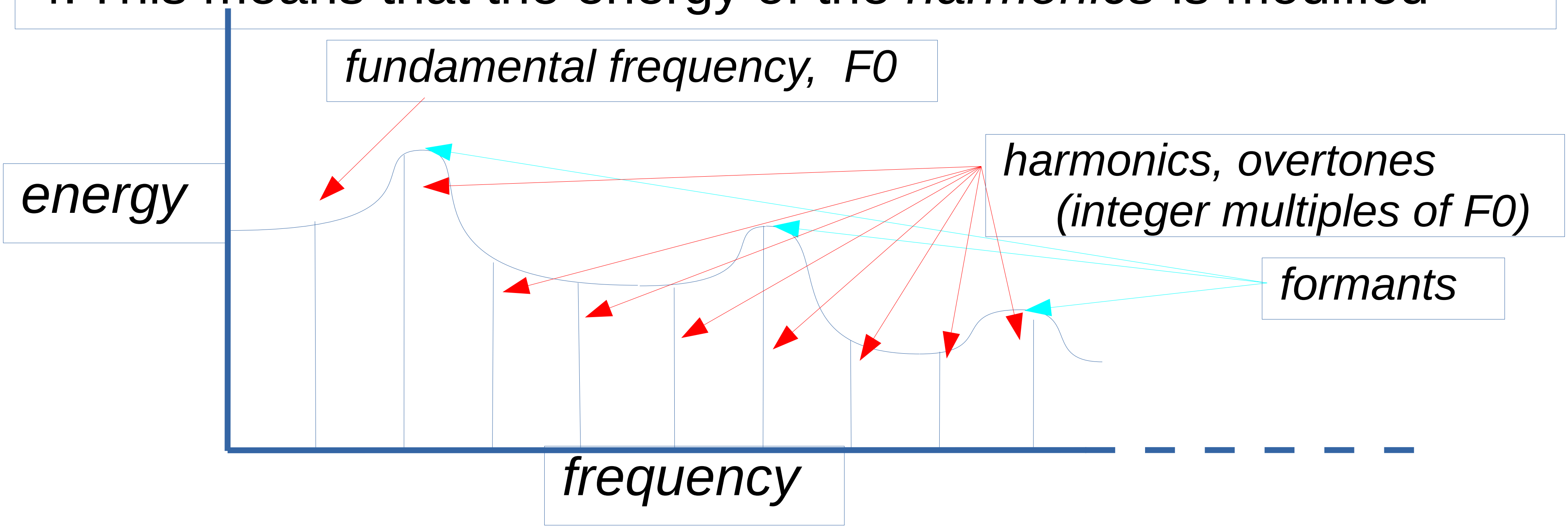
The Spectrum of Complex Signals

1. The filter system consists of the pharyngeal, nasal and oral cavities
2. The cavities have specific resonant frequencies
3. The frequencies of the oral cavity can be modified by the variable filters (the articulators *tongue* and *lips*)
4. This means that the energy of the *harmonics* is modified



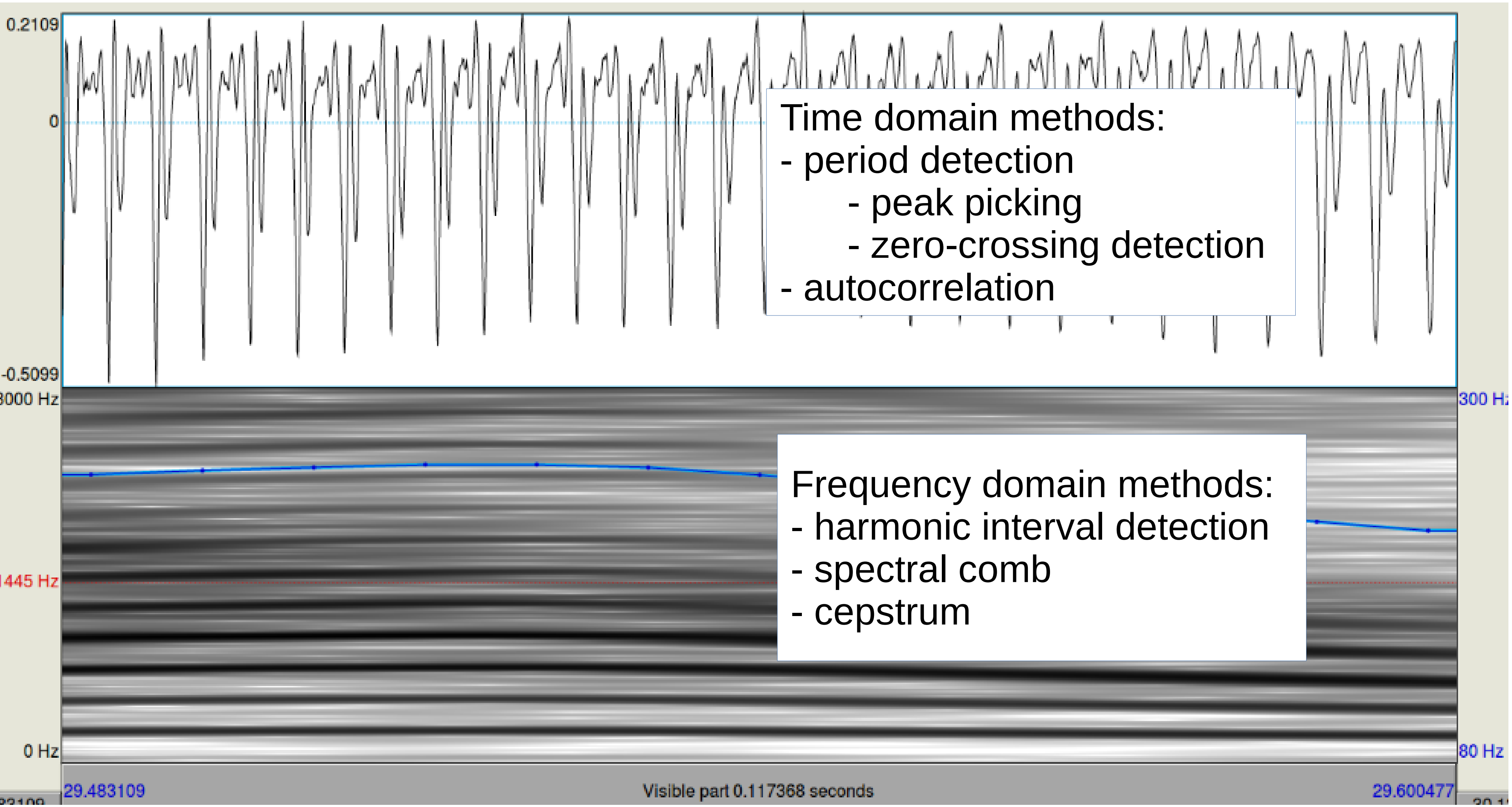
The Spectrum of Complex Signals

1. The filter system consists of the pharyngeal, nasal and oral cavities
2. The cavities have specific resonant frequencies
3. The frequencies of the oral cavity can be modified by the variable filters (the articulators *tongue* and *lips*)
4. This means that the energy of the *harmonics* is modified



F0 estimation (aka ‘pitch tracking’)

From waveform to F0



Time domain methods:
- period detection
    - peak picking
    - zero-crossing detection
- autocorrelation

Frequency domain methods:
- harmonic interval detection
- spectral comb
- cepstrum

From Measurements to Models

Regression smoothing examples

1. Identify voiced intervals
2. Extract F0
3. Interpolate silent intervals

   Simplified in the following examples:

   3rd quartile (75th percentile)

4. Modelling:
   - Calculate overall ‘smoothing’ function

     Linear, quadratic etc. (polynomial) regression over interpolated F0 sequence

   - Calculate residuals (microprosody model):

     Subtract regression values from F0 values

From phonetic measurements to phonetic models

F0 modelling (aka ‘F0 stylisation’)

- the simplification of the F0 trajectory to remove
  - irrelevant properties
  - noise
- the methods can be seen as ‘smoothing operations’
  - moving median window
  - global regression (Huber)
  - local regression (IPO, Instituut voor Perceptie-Onderzoek, Eindhoven)
  - local quadratic spline segment interpolation (Hirst)
  - Fujisaki production model: phrase contour + accent contour sequence
  - Liberman & Pierrehumbert: downtrend + accent sequence
- the modelling functions can be applied to different interval types:
  - voiced sequences, interpausal units
  - whole utterances, whole dialogue exchanges
  - …. (in other words, whatever you think is a useful domain)

F0 smoothing: different approaches

1. Smoothing by median filter:
   - the median of sequences of 3 (or any odd number of) measurements
2. Smoothing by linear regression

   $y = a_0 + a_1x + \varepsilon$

3. Smoothing by polynomial regression:

   $y = a_0 + a_1 \cdot t + a_2 \cdot t^2 + a_3t^3 + \ldots + a_0t^n + \varepsilon$

4. Smoothing by asymptotic descent:

   $F0(t_{t+1}) = m \cdot F0(t_i) + \varepsilon$, for $m < 0$

   $a + F0(t_{i+1}) = a + m \cdot F0(t_i) + \varepsilon$, $m < 0$ non-zero asymptote

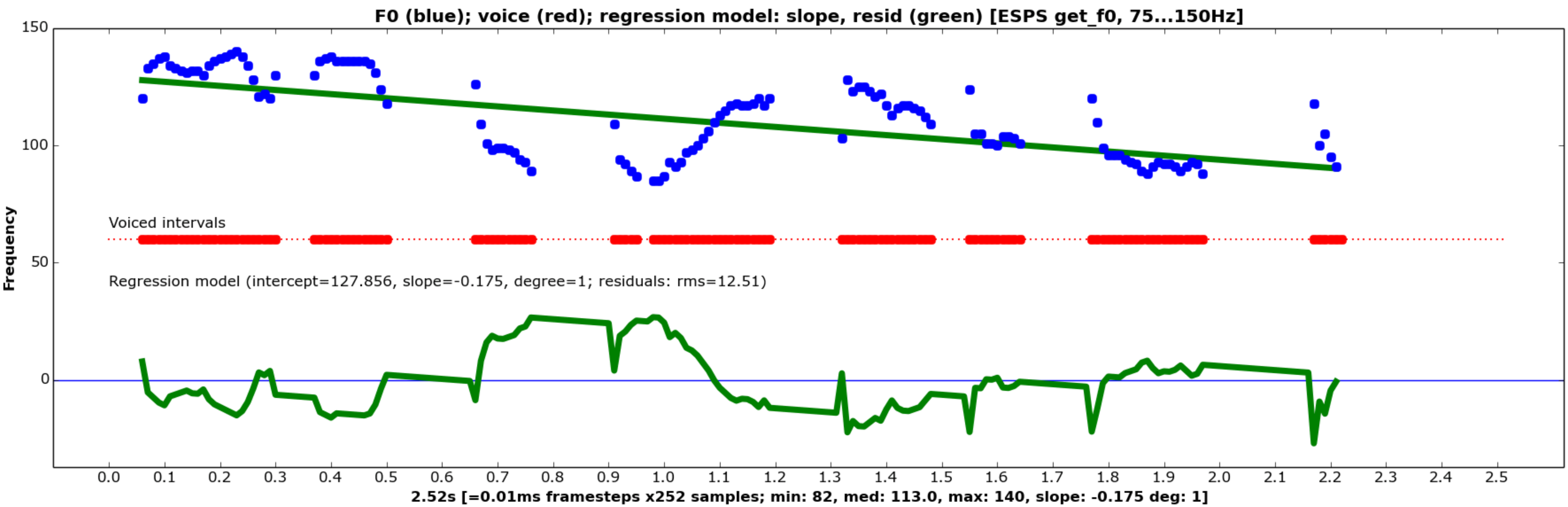
F0 Models: Degrees of Approximation

Smoothing by linear regression (degree 1)

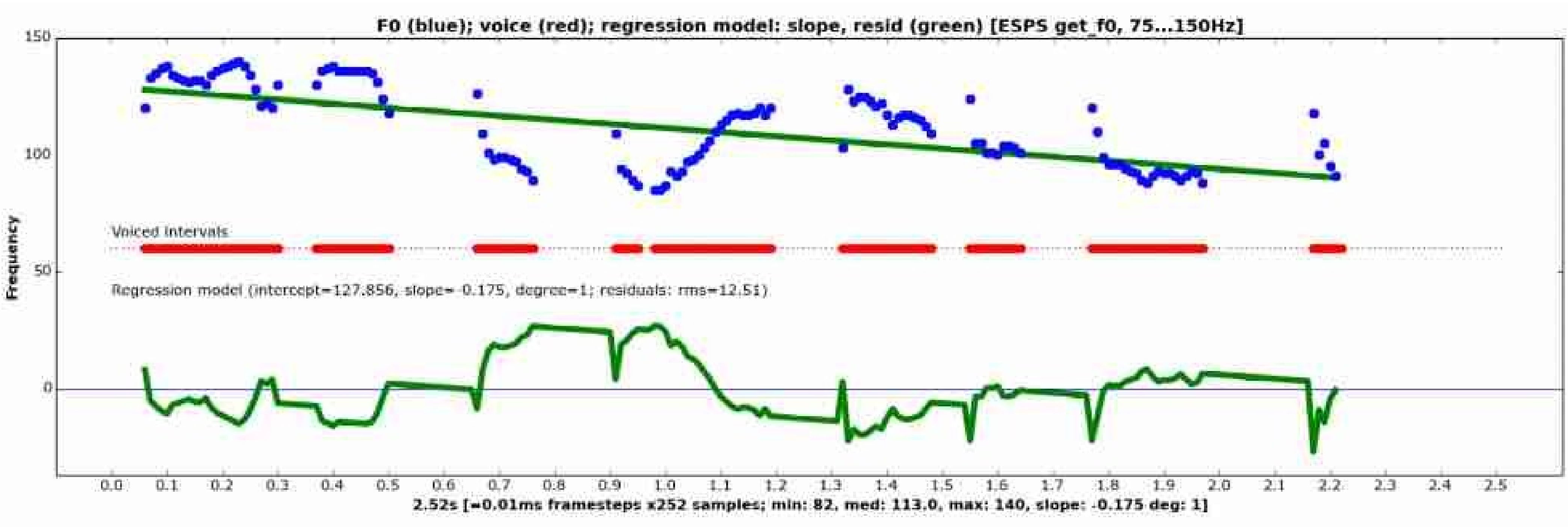
$$y = a_0 + a_1 x + \varepsilon$$

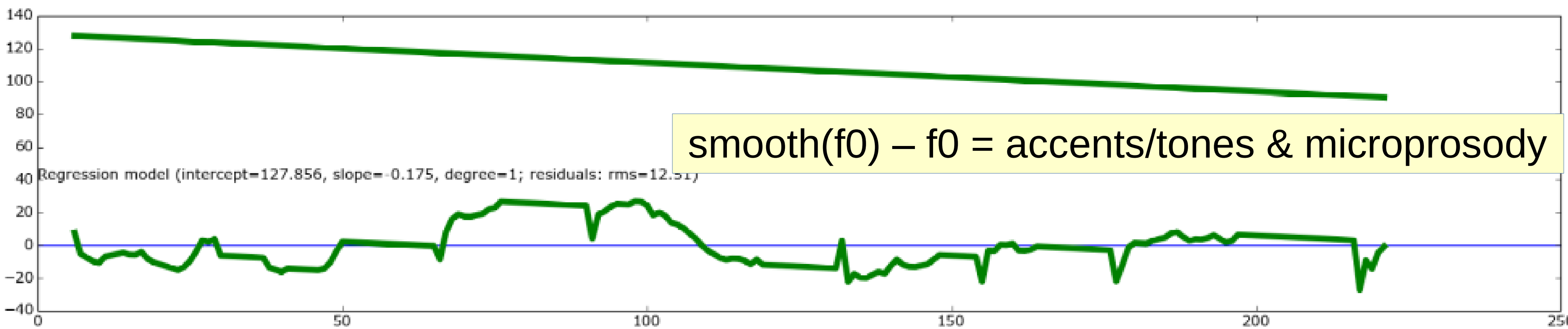
Smoothing by polynomial regression (degree *n*):

$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 x^3 + \ldots + a_0 x^n + \varepsilon$$



Global linear regression contour

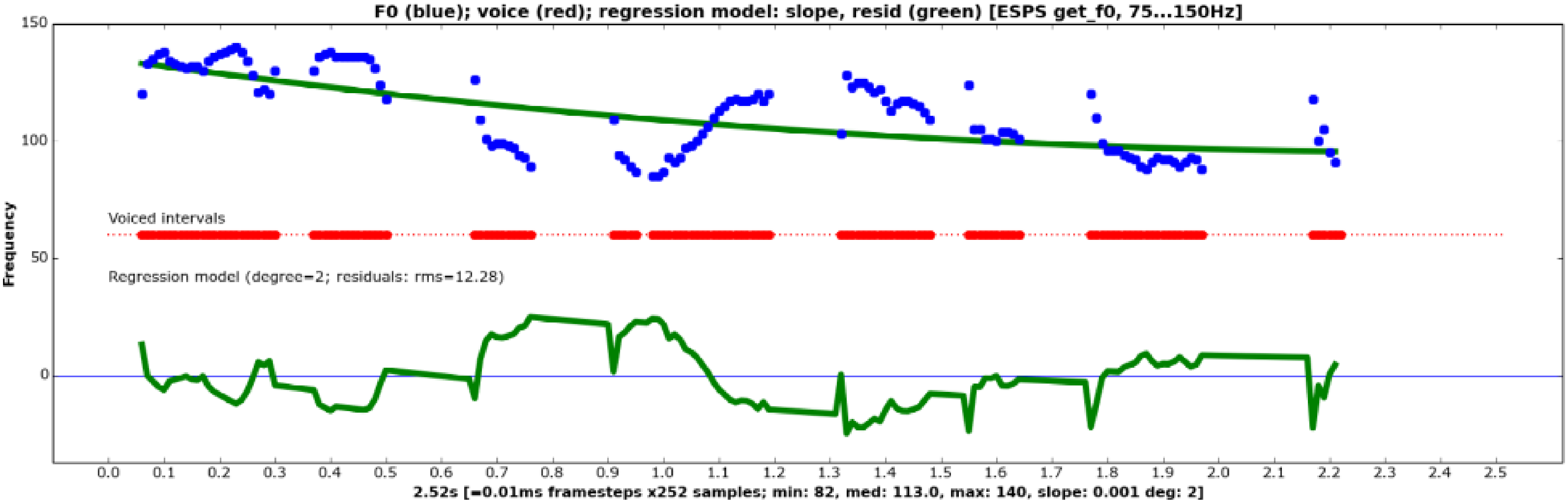


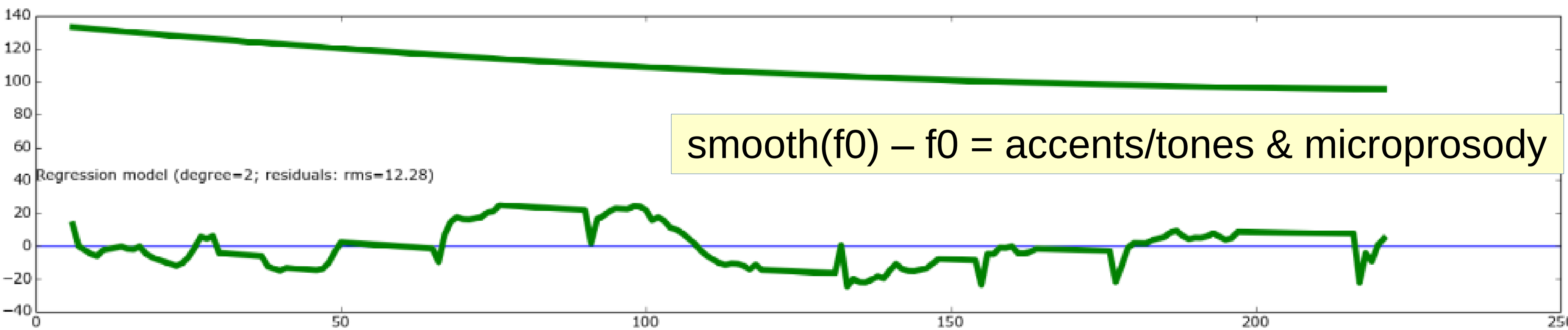


smooth(f0) – f0 = accents/tones & microprosody

Endlich gab der Nordwind den Kampf auf.

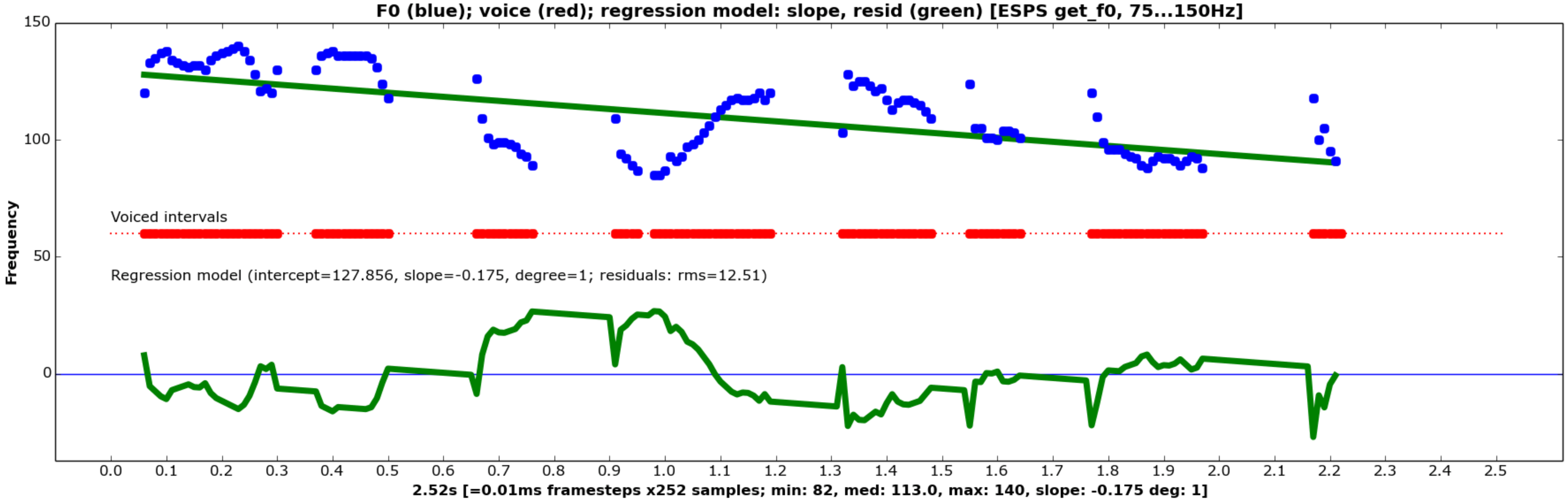
Global quadratic regression contour





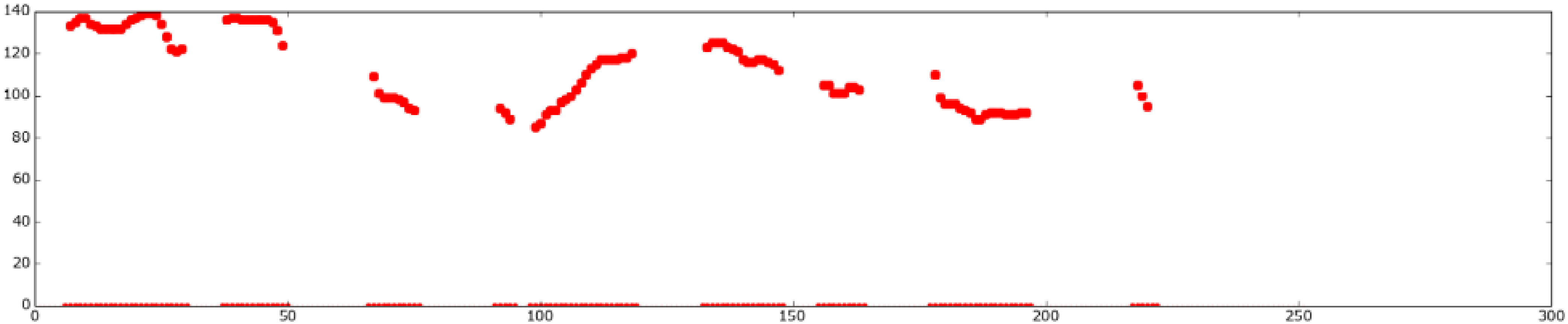
Endlich gab der Nordwind den Kampf auf.

Global regression contours, up to degree 20

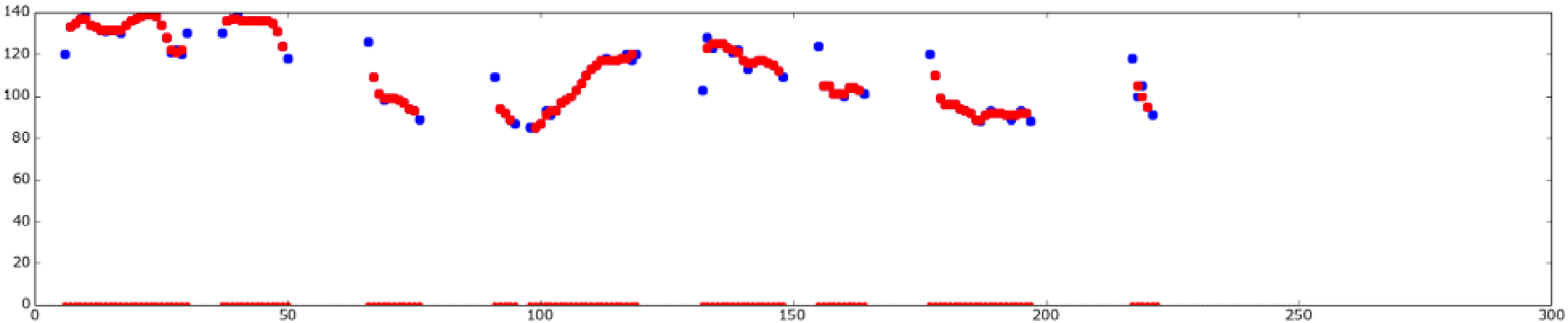


Endlich gab der Nordwind den Kampf auf.

Simple median filter (scope: 3), often used

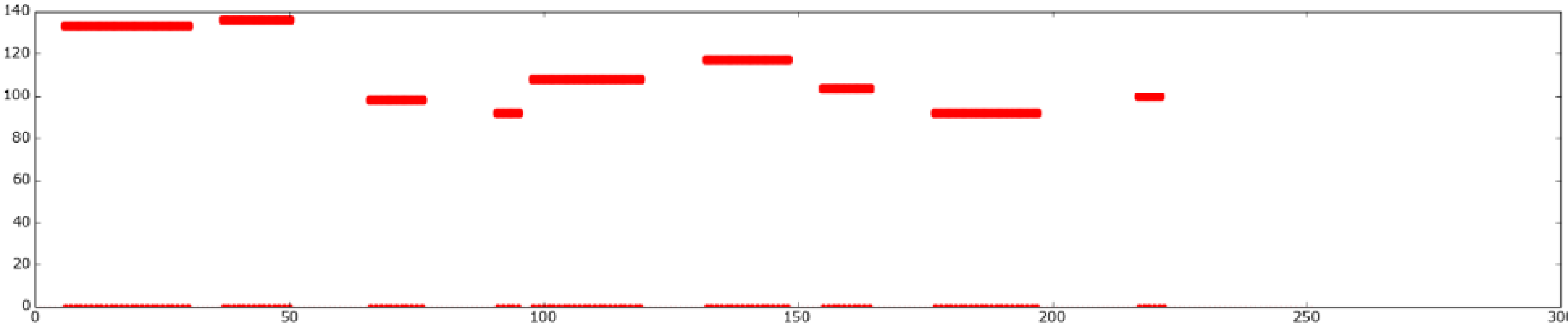


Endlich gab der Nordwind den Kampf auf.

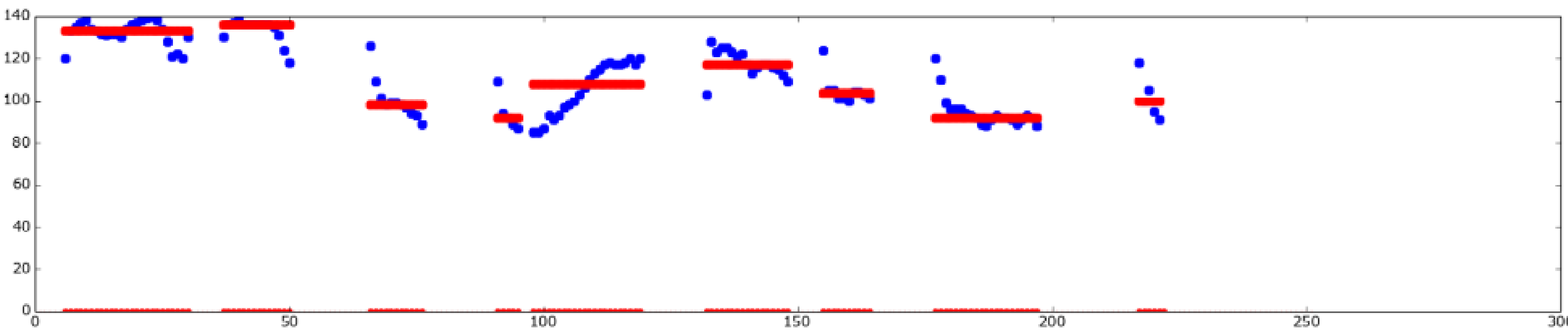


Each F0 value is normalised to the median F0 value of its immediate neighbours

Simple local median levelling filter – robotic!

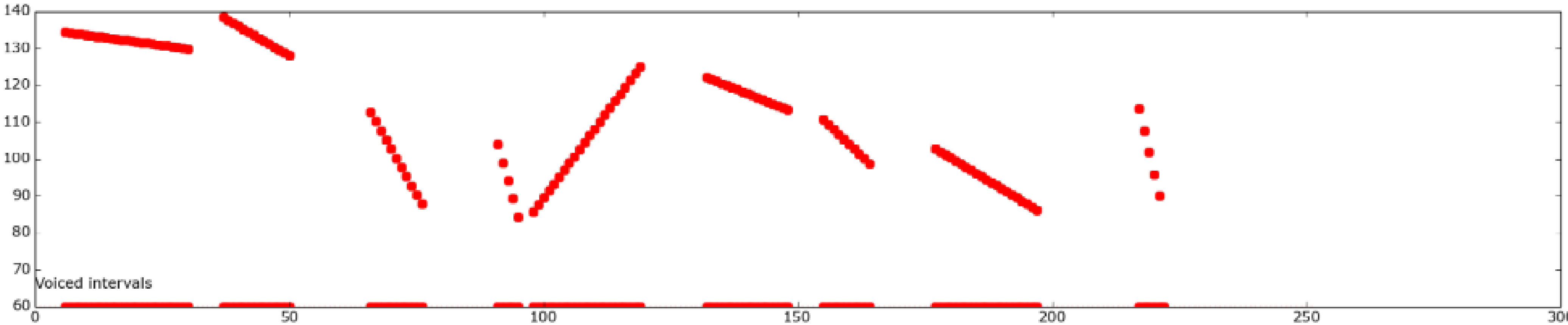


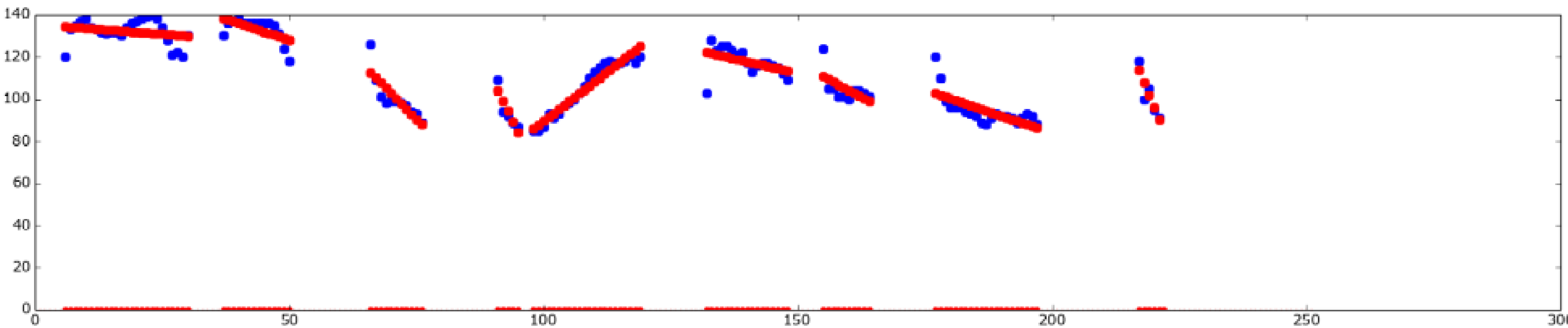
Endlich gab der Nordwind den Kampf auf.



Each F0 value in a sequence is normalised to the median F0 value for the sequence

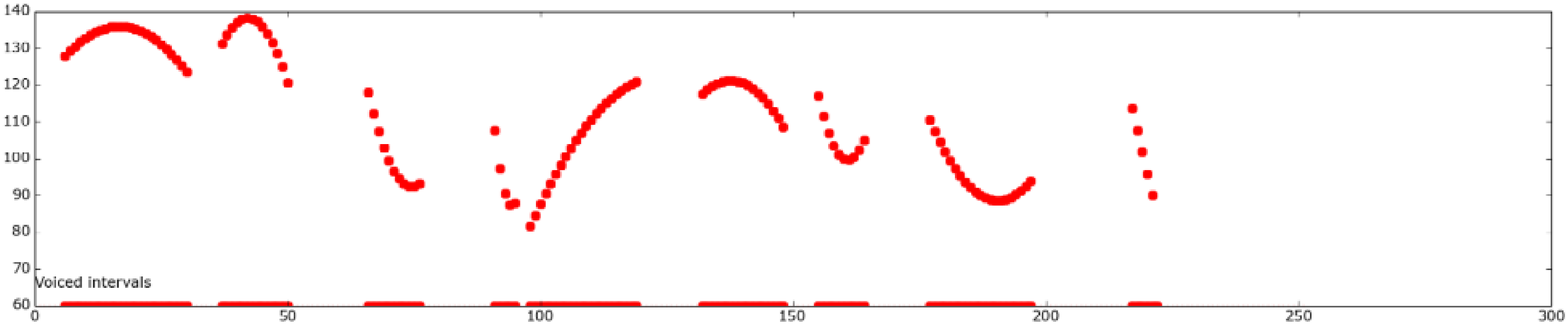
Local voicing regression contours, degree 1

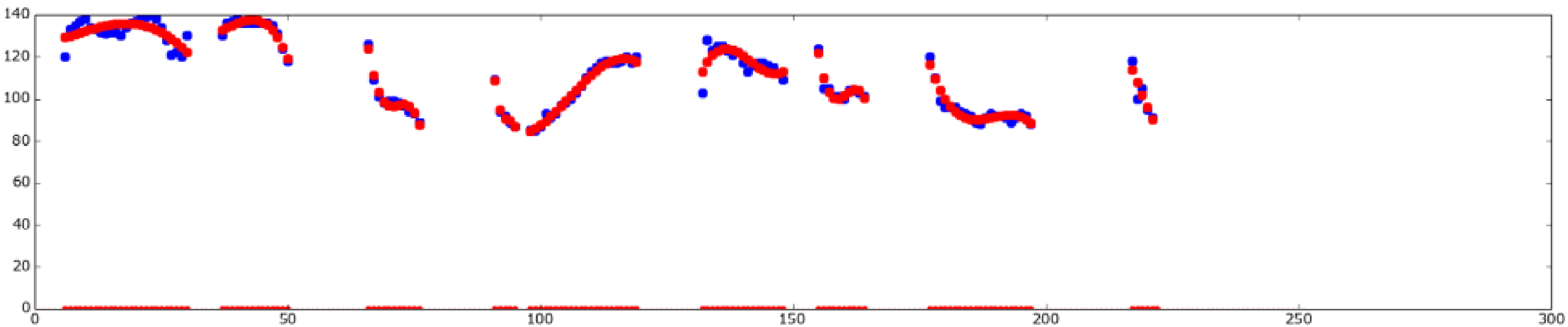




Endlich gab der Nordwind den Kampf auf.

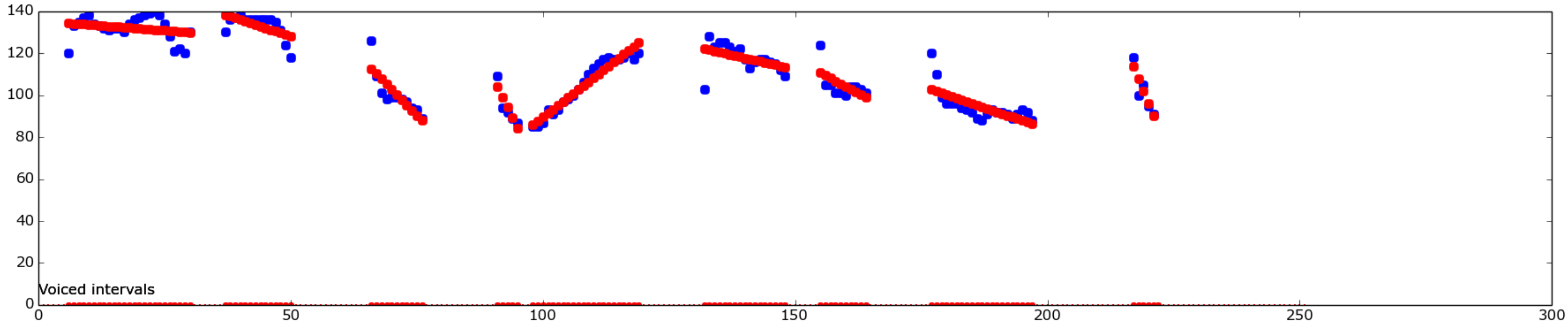
Local voicing regression contours, degree 2





Endlich gab der Nordwind den Kampf auf.

Local voicing regression contours (1...5)



Endlich gab der Nordwind den Kampf auf.

Higher degrees of polynomial regression can be difficult to interpret.

Note the progression:

- from <u>underfitting</u> with linear regression
- to <u>overfitting</u> with higher degrees polynomial regression

Other Types of Model

Hirst: quadratic spline - 'piecewise quadratic function'

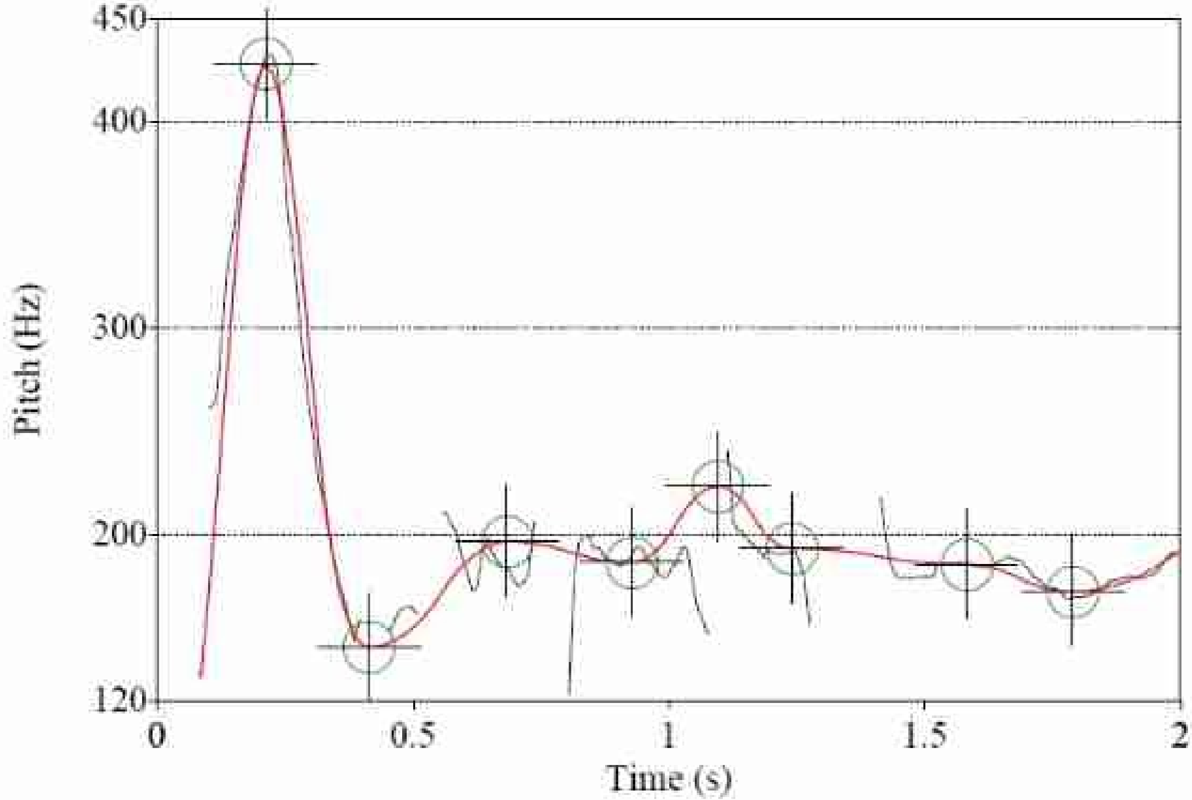


Fig. 6.7 Macromelodic profile (red) for a two-second extract from recording A01, defined as quadratic transitions between anchor points (green).

Hirst: quadratic spline - ‘piecewise quadratic function’

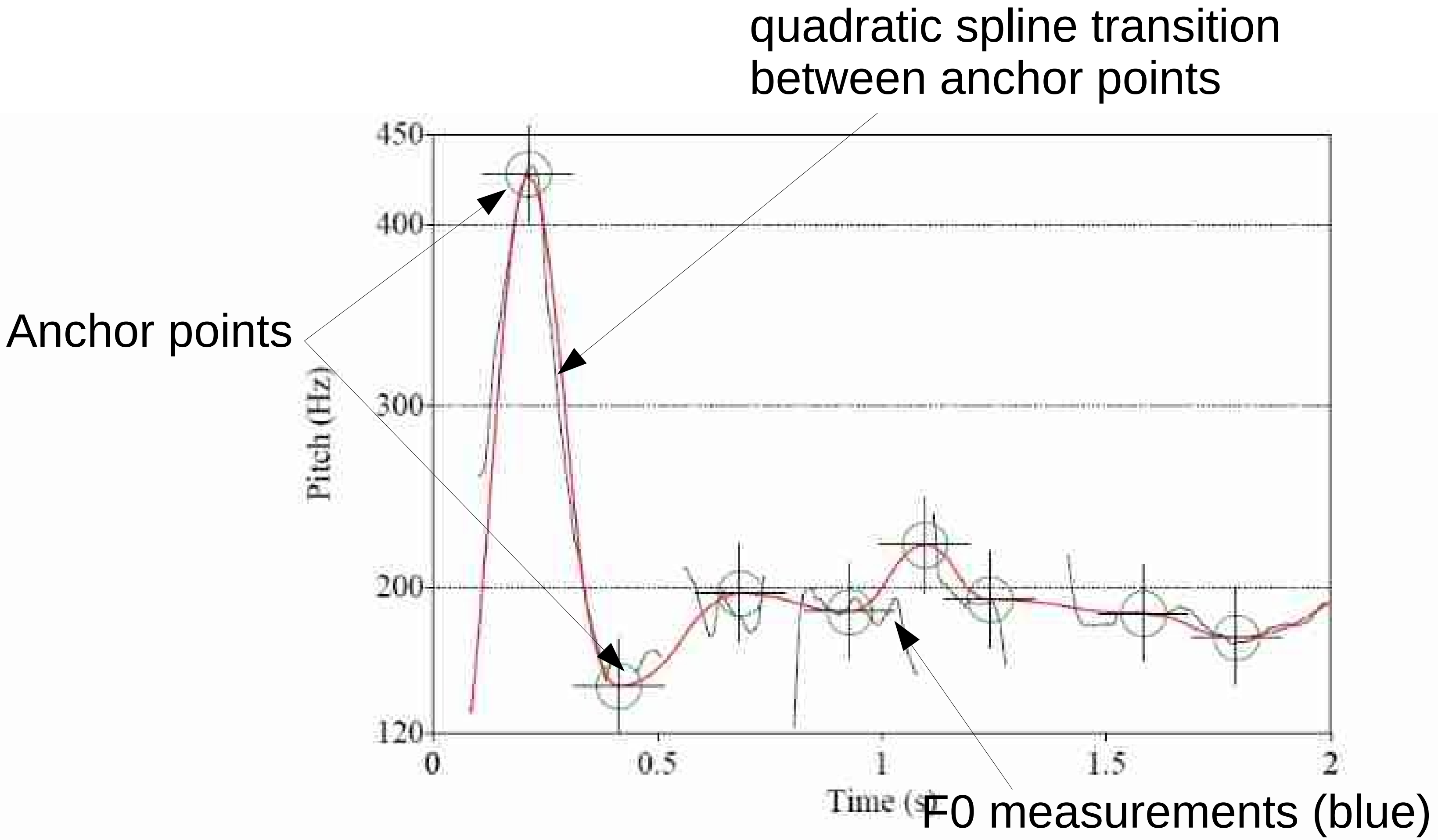


Fig. 6.7 Macromelodic profile (red) for a two-second extract from recording A01, defined as quadratic transitions between anchor points (green).

Models of f0 patterning: Liberman & Pierrehumbert

Subtract the reference line from the F0 trajectory

Define the asymptotic declination line

Define the relation between focus and non-focus accent types

Define the relation between first pitch accent and reference line

Define final lowering

Models of f0 patterning: Liberman & Pierrehumbert

Model 1

a. General F0 transform

$$T(P) = P - r$$

P and r in Hz

Modified transform for model 1

$$T(P) = (1/l) \cdot (P - r)$$

where $l < 1$ in final position, $l = 1$ otherwise

b. Downstep

$$T(P_i) = s \cdot T(P_{i+1})$$

where P_i is the F0 target in Hz of a step accent in position i, downstepped with respect to the previous accent target P_{i-1}

c. Answer-background relation

$$T(P_A) = k \cdot T(P_B)$$

where P_A is the F0 target in Hz of the A accent, and P_B the B accent

d. Relation of r to initial accent target

$$r = f \cdot (P_0 - b)^e + d + b$$

where P_0 is the target in Hz of the first pitch accent, and d, e, f, and b are constants

e. Final Lowering

$$P \rightarrow r + l \cdot (P - r) / \_\_\_\_ \$$$

where $l < 1$

Model 1A

Substitute

$$r = f \cdot (P_0)^e + d$$

for equation (5d) in model 1.

Model 1B

Substitute

$$r = f \cdot P_0 + d$$

for equation (5d) in model 1.

Model 1C

Substitute

$$P \rightarrow l \cdot P / \_\_\_\_ \$$$

for rule (5e) in model 1.

Models of f0 patterning: Liberman & Pierrehumbert

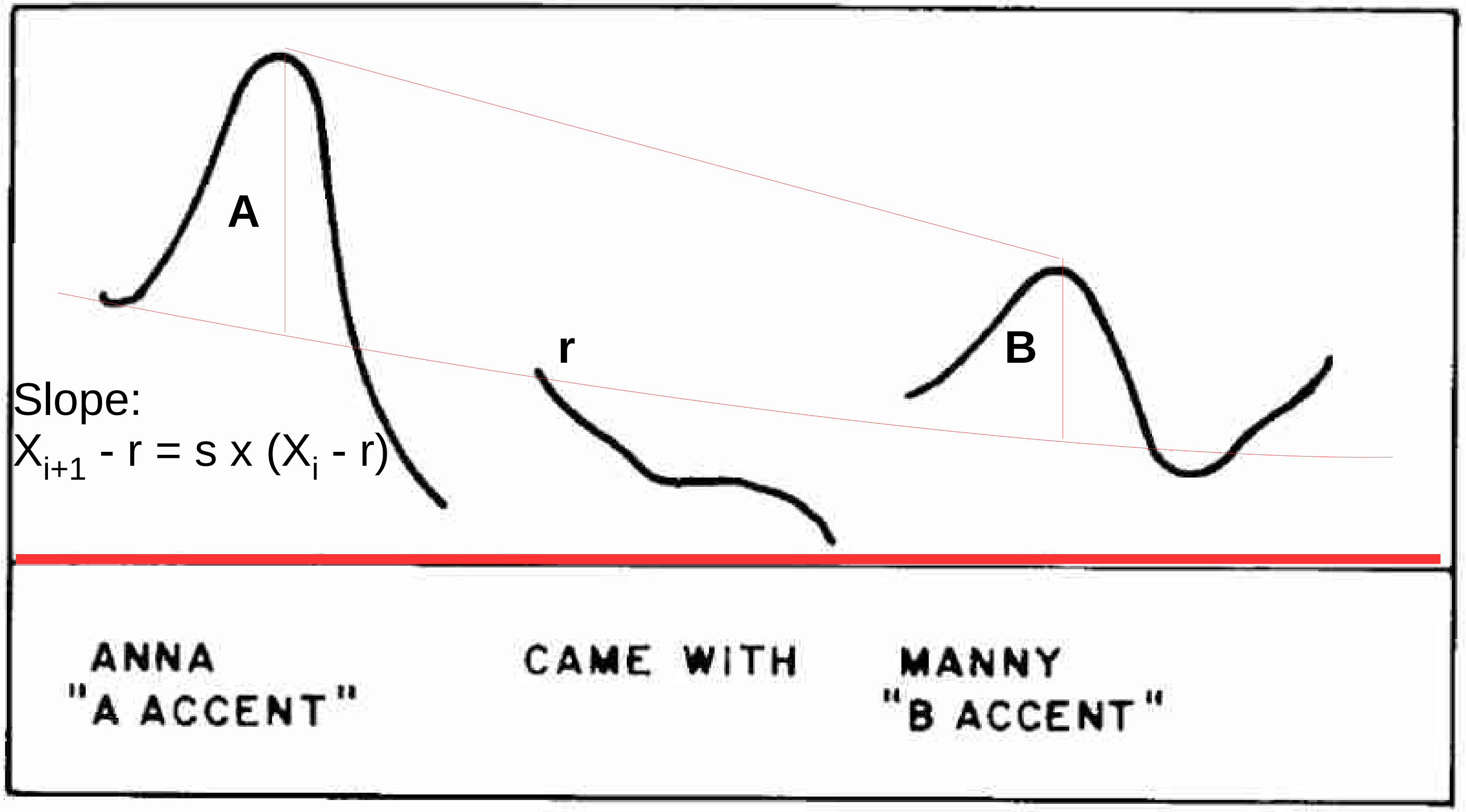


Figure 9
An F0 contour for *Anna came with Manny*, produced as a response to *What about Manny? Who came with him?*

Models of f0 patterning: Liberman & Pierrehumbert

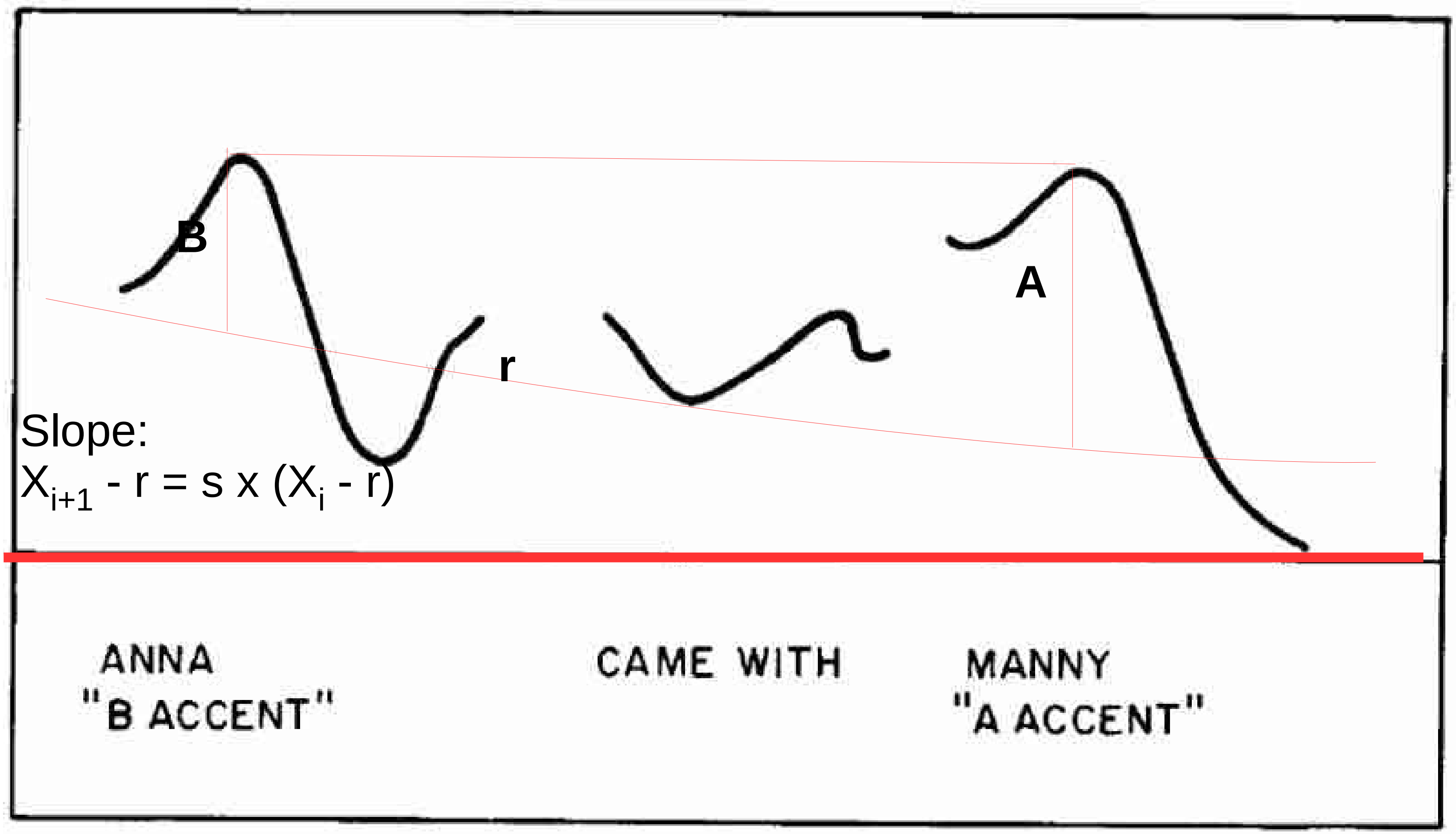
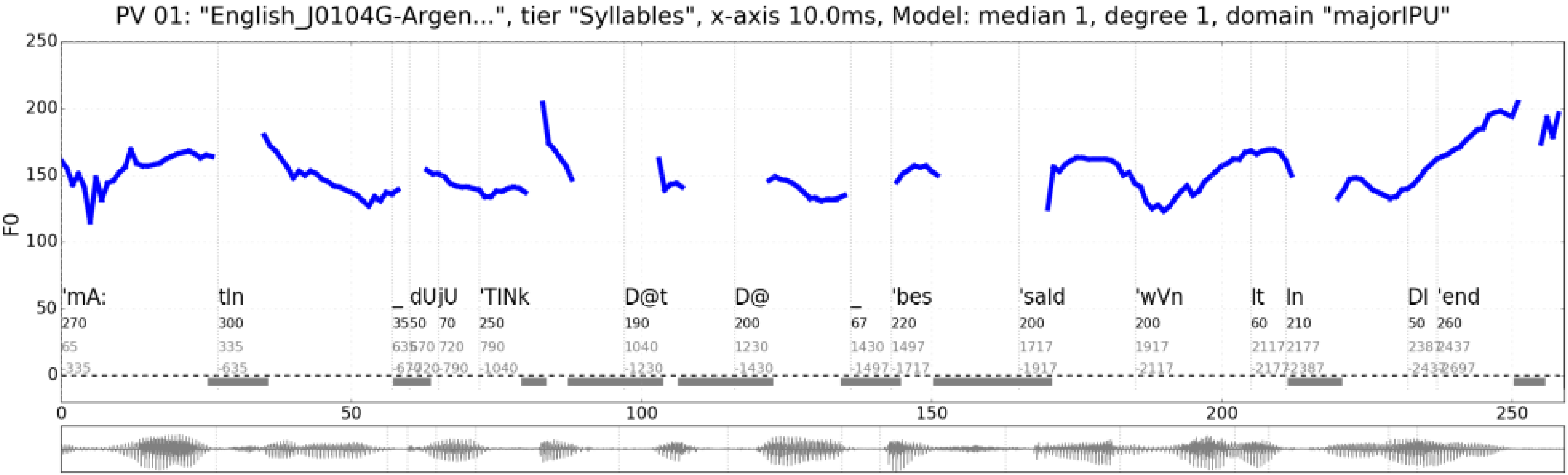


Figure 10
An F0 contour for *Anna came with Manny*, produced as a response to *What about Anna? Who did she come with?*

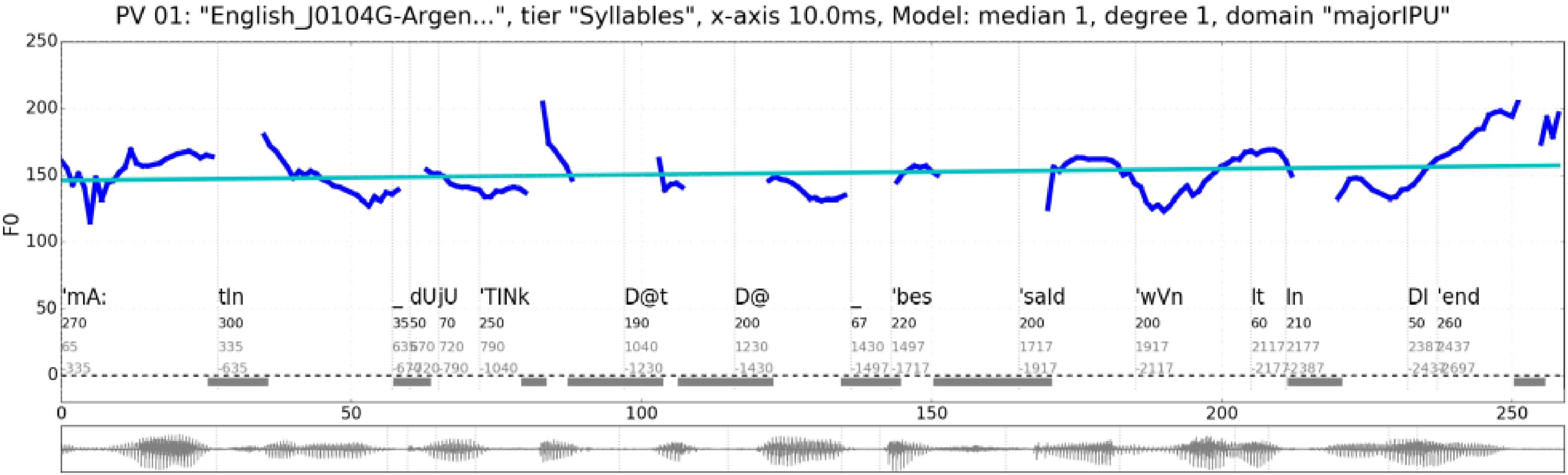
Orientation: The F0 Curve (Pitch Curve) as a Time Function



The F0 curve of an utterance is a model of the melody of this utterance:

- It is a simplified representation of reality
- It is an abstraction, and ignores many aspects of reality
- It is a complex function:
  - Overall shape from beginning to end
  - Pitch accent shape and position
  - Patterns of weak syllables
  - Effects of vowels and consonants (perturbations)

Orientation: The F0 Curve (Pitch Curve) as a Time Function

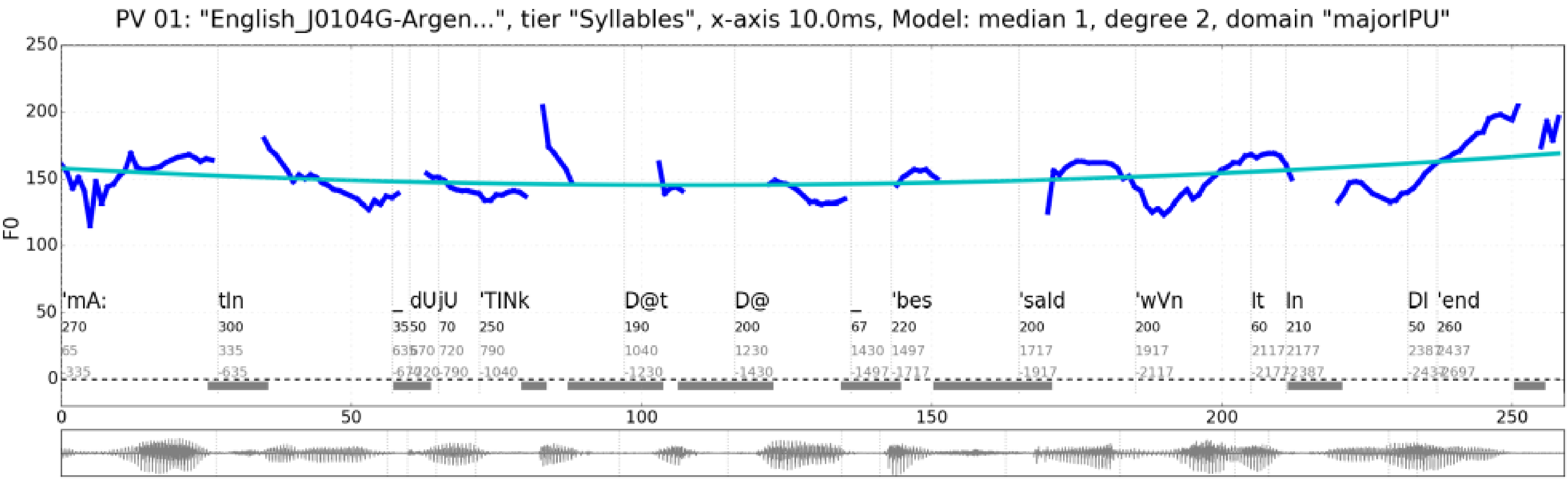


Identification of the main factors with partial models:

- Even more simplifications of reality!
- Many methods:
  - IPO method with local linear models
  - **Huber's method: utterance-size linear models**
  - Hirst's "MoMel" with local quadratic splines
  - Wavelet analysis

Question:
What does this function model? Is it a good model?

Orientation: The F0 Curve (Pitch Curve) as a Time Function

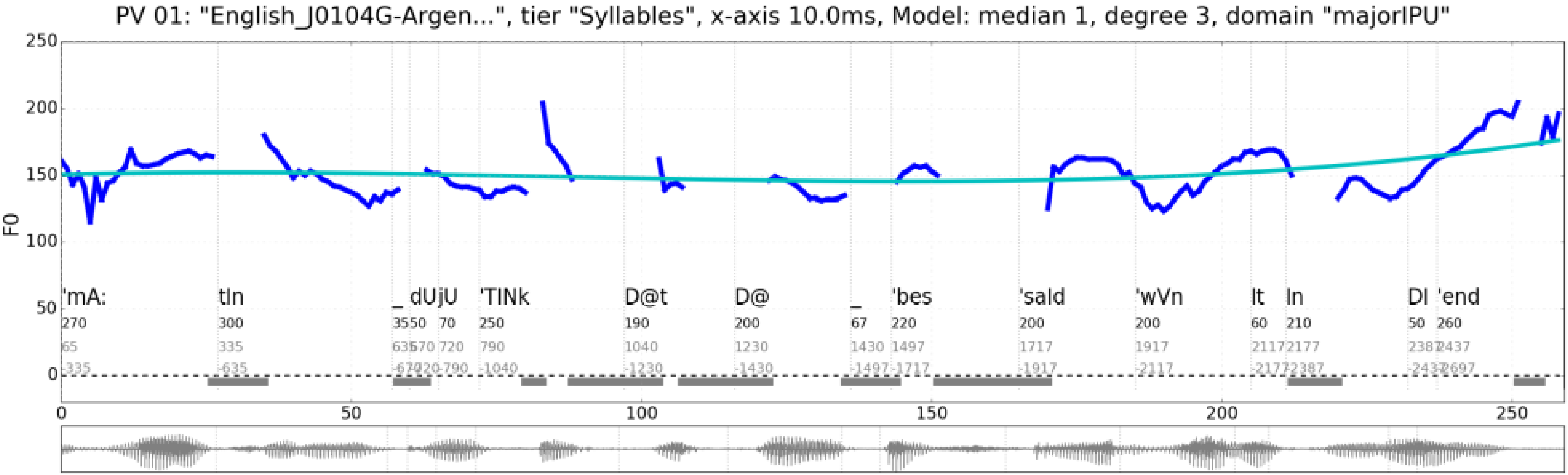


Identification of the main factors with partial models:

- Even more simplifications of reality!
- Many methods:
  - IPO method with local linear models
  - Huber's method with utterance-size linear models
  - Hirst's "MoMel" with local quadratic splines
  - Wavelet analysis

Question:
What does this function model? Is it a good model?

Orientation: The F0 Curve (Pitch Curve) as a Time Function

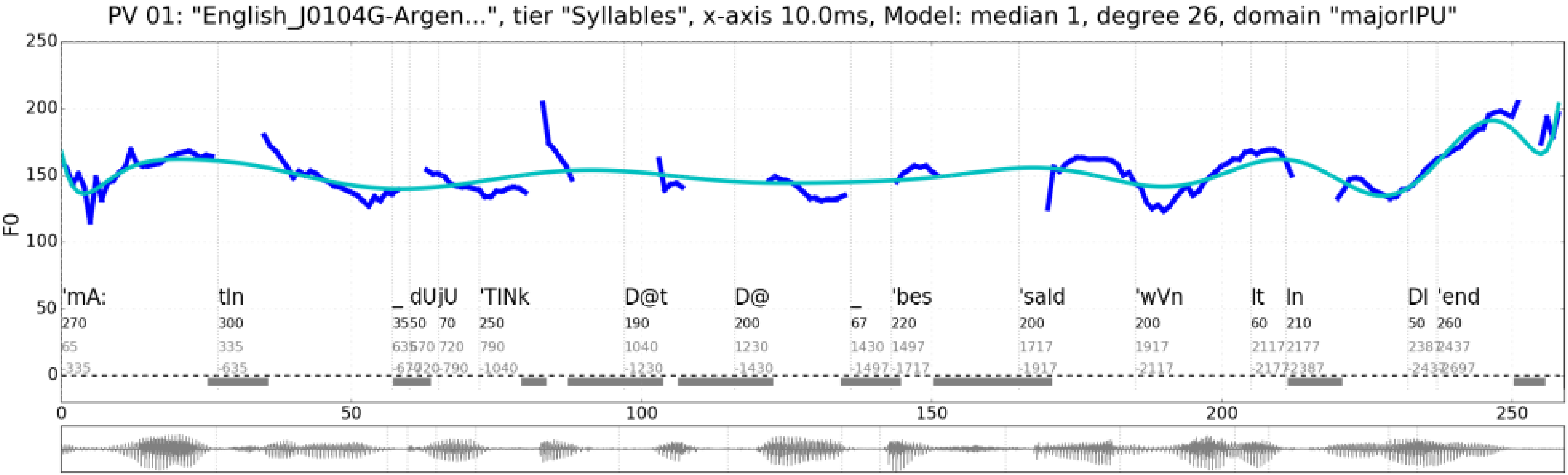


Identification of the main factors with partial models:

- Even more simplifications of reality!
- Many methods:
  - IPO method with local linear models
  - Huber’s method with utterance-size linear models
  - Hirst’s “MoMel” with local quadratic splines
  - Wavelet analysis

Question:
What does this function model? Is it a good model?

Orientation: The F0 Curve (Pitch Curve) as a Time Function

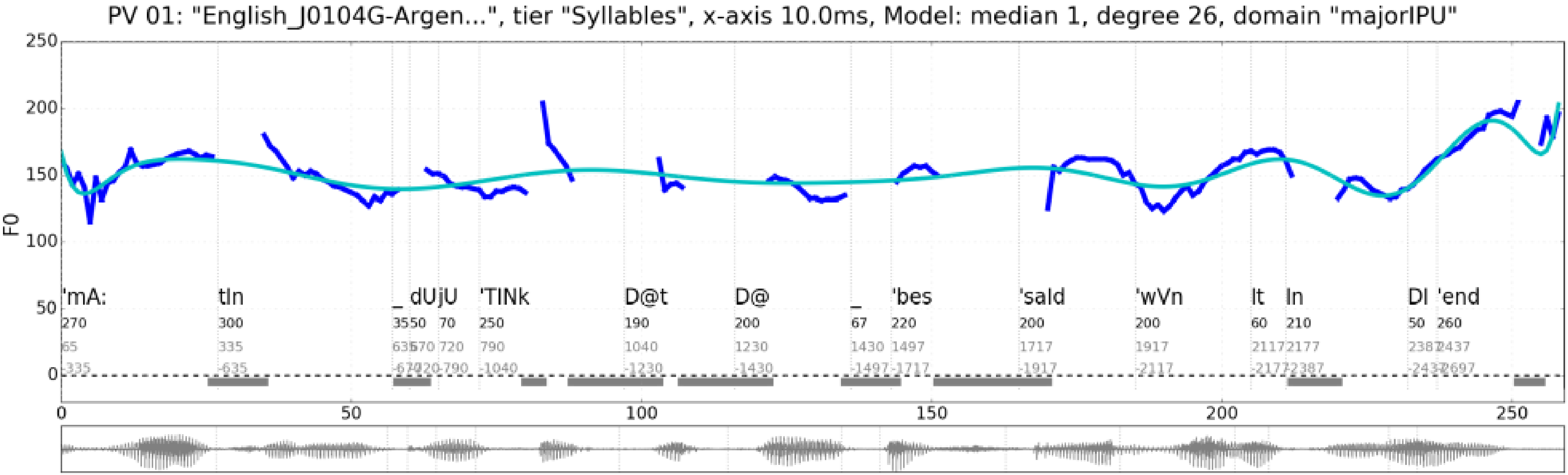


Identification of the main factors with partial models:

- Even more simplifications of reality!
- Many methods:
  - IPO method with local linear models
  - Huber's method with utterance-size linear models
  - Hirst's "MoMel" with local quadratic splines
  - Wavelet analysis

Question:
What does this function model? Is it a good model?

Orientation: The F0 Curve (Pitch Curve) as a Time Function

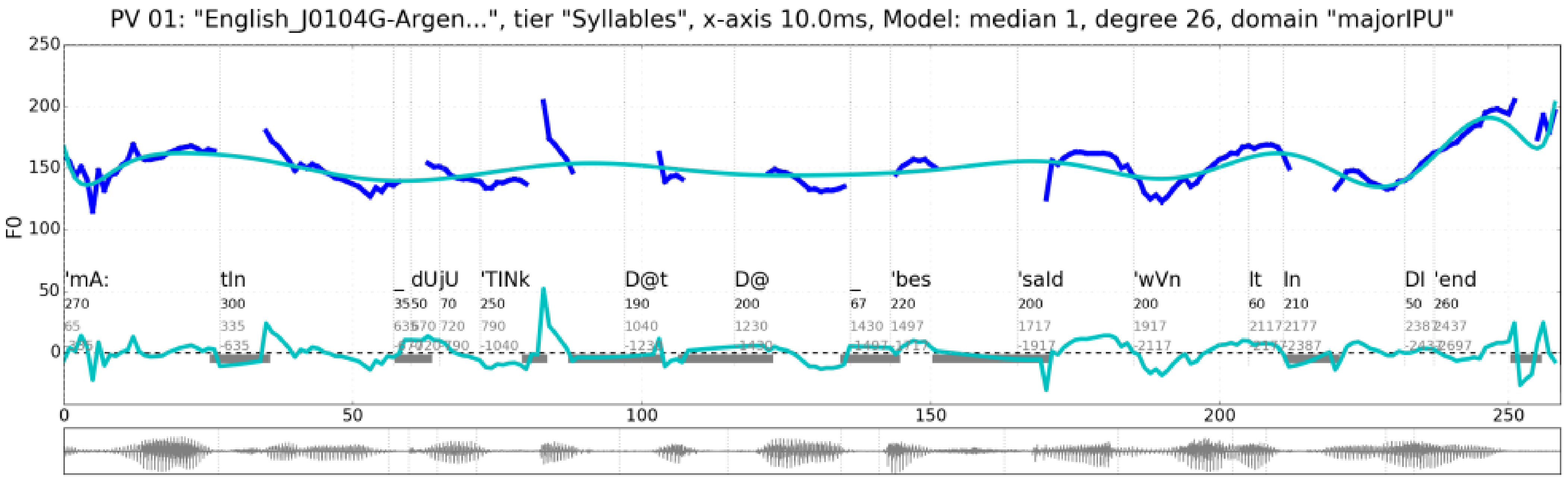


Identification of the main factors with partial models:

- Even more simplifications of reality!
- Many methods:
  - IPO method with local linear models
  - Huber’s method with utterance-size linear models
  - Hirst’s “MoMel” with local quadratic splines
  - Wavelet analysis

Question:
What does this function model? Is it a good model?

Orientation: The F0 Curve (Pitch Curve) as a Time Function



Procedure:

1. Calculate F0

   In these examples: Talkin's RAPT algorithm:
   "A Robust Algorithm for Pitch Tracking"

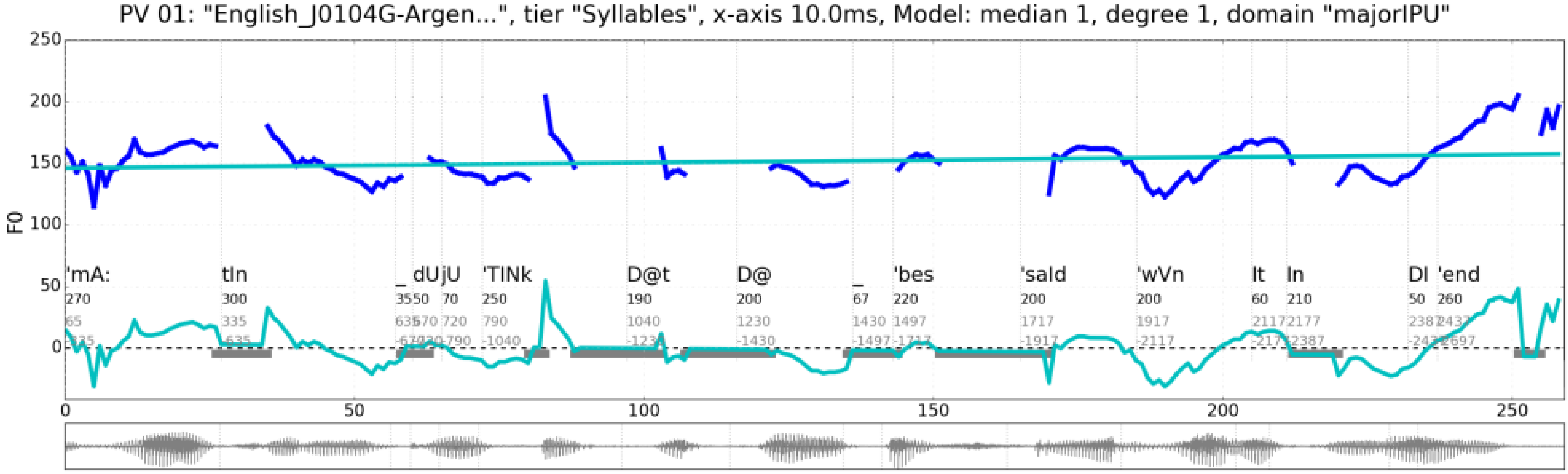
2. Calculate polynomial regression line
3. Calculate residuals:

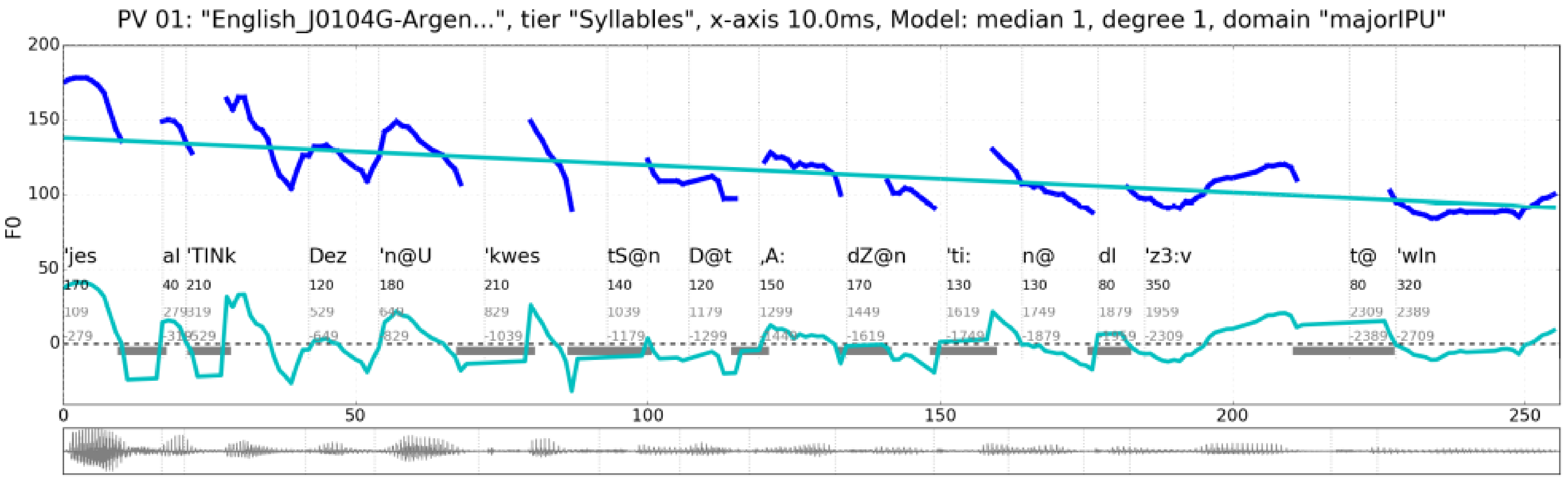
   d – p (subtract polynomial values from data values)

Question:

What do the residuals model? Is this a good model?

Orientation: The F0 Curve in Dialogue

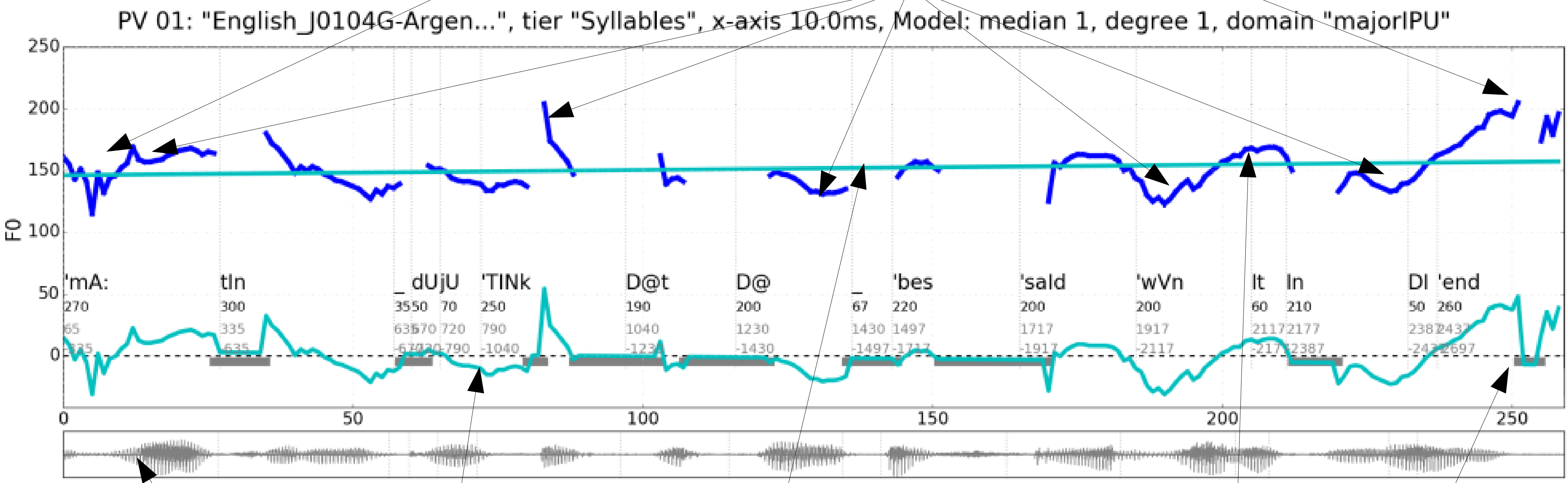




Overview of the Main Components of the F0 Time Function

boundary tone

local pitch accents



waveform amplitude

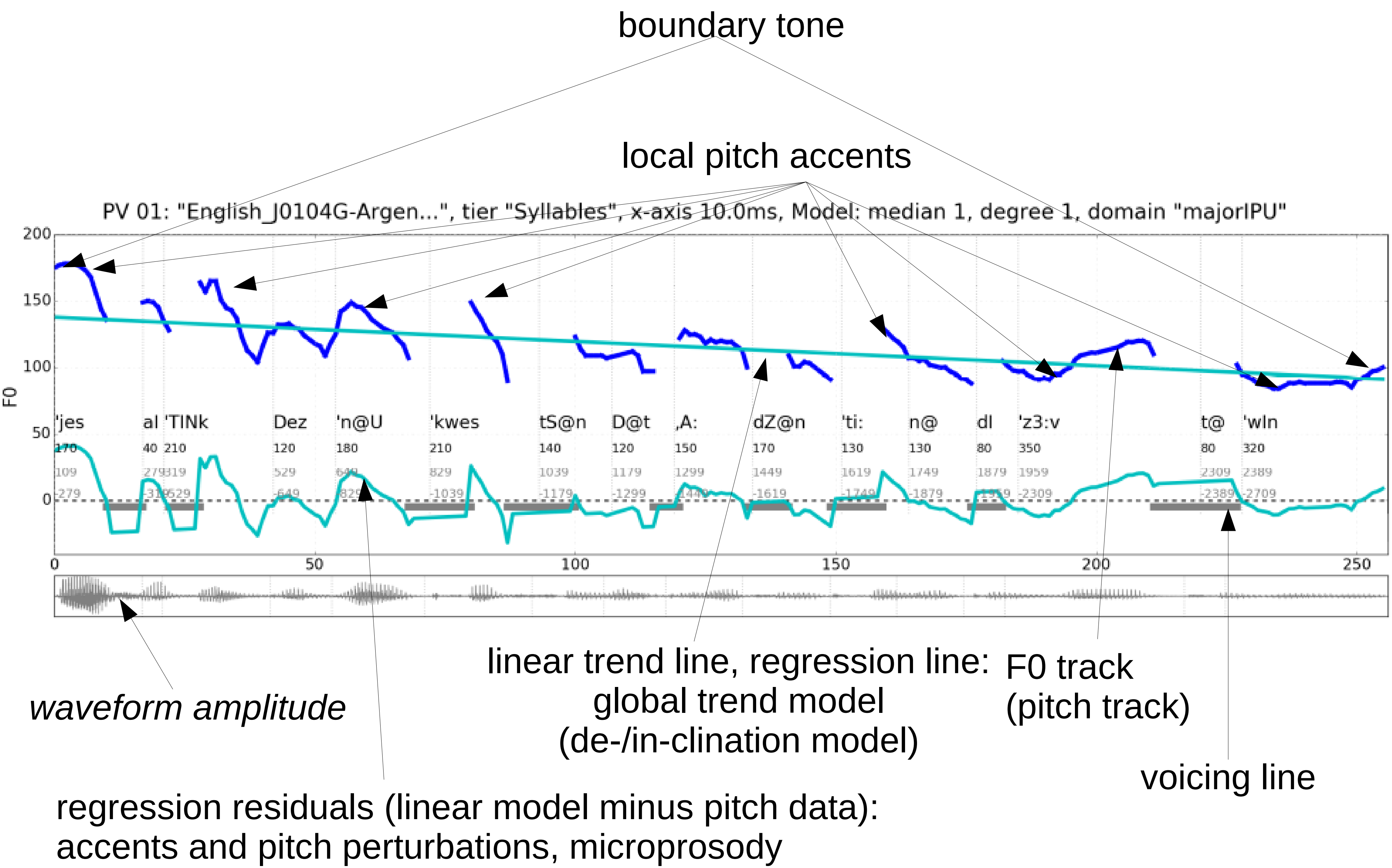
linear trend line, regression line:
global trend model
(de-/in-clination model)

F0 track
(pitch track)

voicing line

regression residuals (linear model minus pitch data):
accents and pitch perturbations, microprosody

Components of the F0 Time Function



Evaluation of stylised contours – 2 methods:

Difference between F0 and stylised contour

Difference between contours in perception test

From:
Demenko Grażyna, Wagner Agnieszka (2006). The Stylization of Intonation Contours. *Proceedings of Speech Prosody 3,* May 2-5, 2006, Dresden, Germany.

Evaluation of stylised contours: Demenko & Wagner

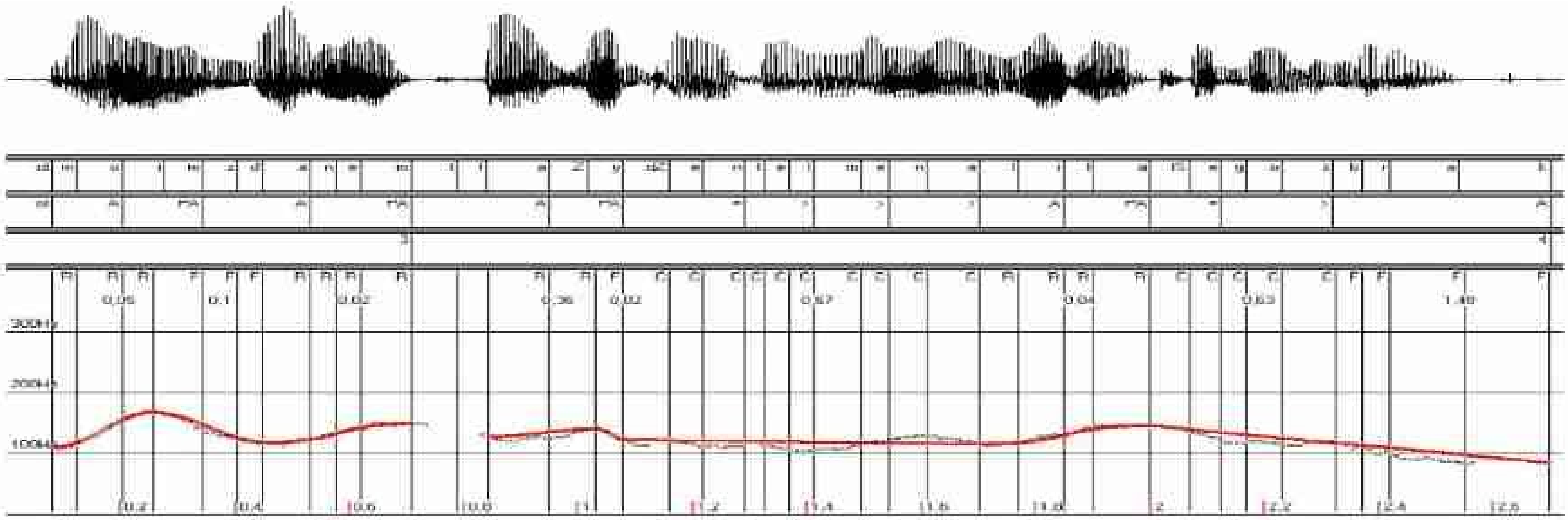


Figure 1: *Sentence:* In my opinion, the face of the lilac gentleman lacks something. *From top to bottom of the picture: waveform, .lab, .syl and .break tiers, and the stylization window. The original F0 contour is marked by dotted black line and the stylized F0 contour in red line.*

D&W 2006 stylisation model (SP3):

$IP \rightarrow IE^{+}$

$IE_i + SL_{i+1} + IE_{i+1}$

IP: Intonation Phrase
IE: Intonation Event
SL: Straight Line

$IE \in \{R, F, C\}$
IE parameters:
- slope
- Fp (F0 at start of event)
- range of F0 change
- shape coefficient of curve:
  - $y = y^{\gamma}$ for $0<x<1$
  - $y = 2-(2-x)y^{\gamma}$ for $1<x<2$

Evaluation of stylised contours: Demenko & Wagner

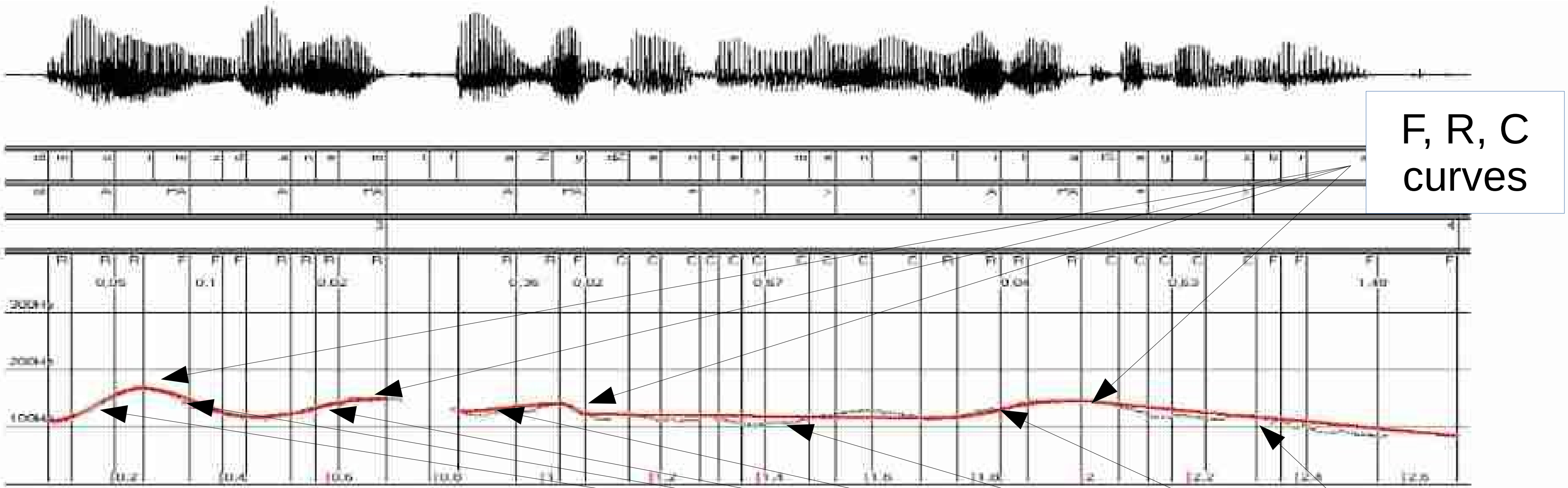


Figure 1: *Sentence:* In my opinion, the face of the lilac gentleman lacks something. *From top to bottom of the picture: waveform, .lab, .syl and .break tiers, and the stylization window. The original F0 contour is marked by dotted black line and the stylized F0 contour in red line.*

D&W 2006 stylisation model (SP3):

$IP \rightarrow IE^{+}$

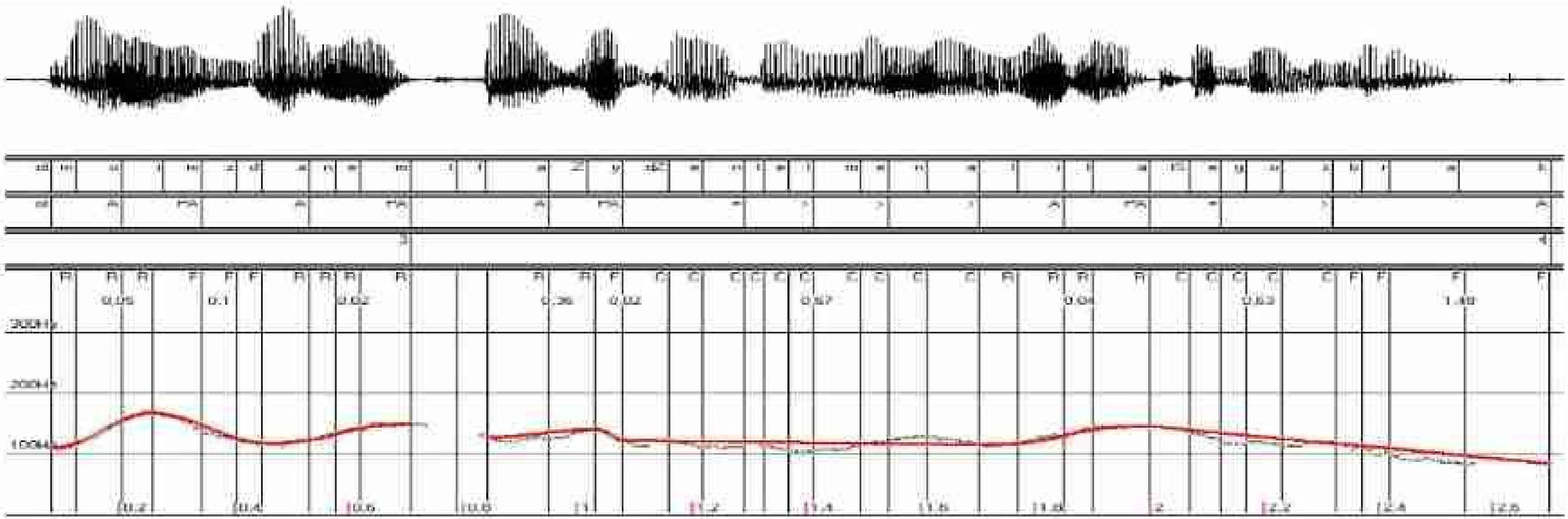
$IE_i + SL_{i+1} + IE_{i+1}$

IP: Intonation Phrase
IE: Intonation Event
SL: Straight Line

$IE \in \{R, F, C\}$
IE parameters:

- slope
- Fp (F0 at start of event)
- range of F0 change
- shape coefficient of curve:
  - $y = y^{\gamma}$ for $0<x<1$
  - $y = 2-(2-x)y^{\gamma}$ for $1<x<2$

Evaluation of stylised contours: Demenko & Wagner



Evaluation 1: goodness of fit

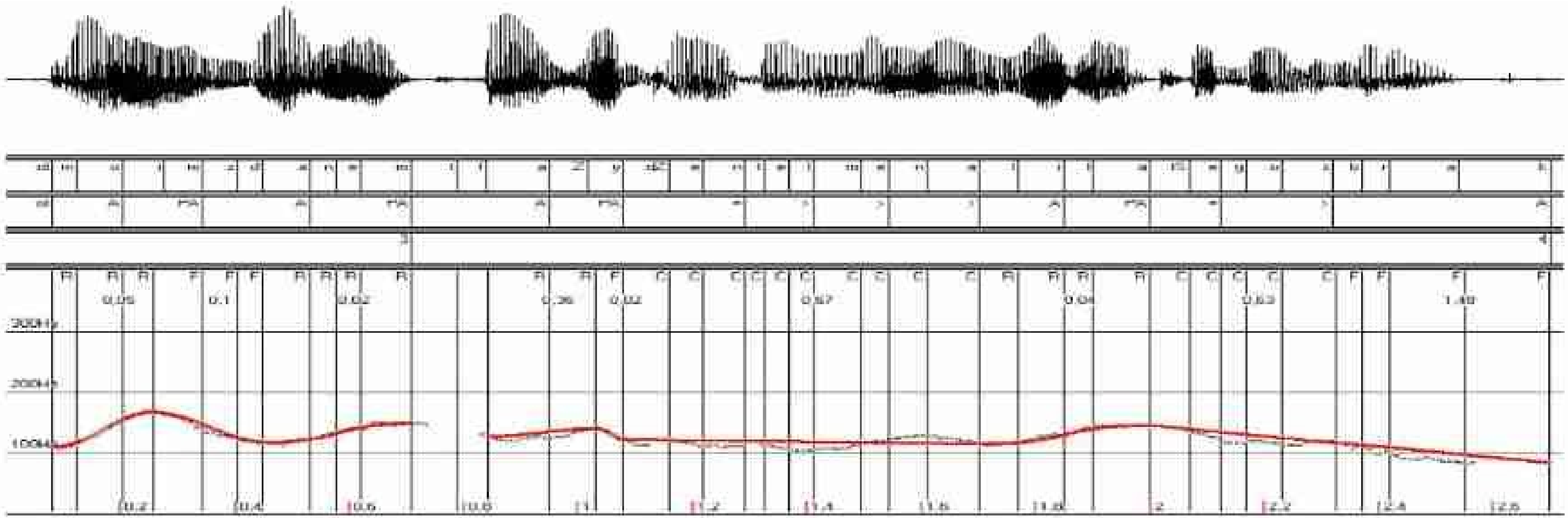
Compare F0 with stylised function with Normalised Mean Square Error:

$$NMSE(t) = \frac{\overline{(F_0(t_i) - Sty(t_i))^2}}{\overline{F_0(t)} \cdot \overline{Sty(t)}}$$

Accented syllable					
	Slope (Hz/s)	Fp (Hz)	Range (Hz)	bend	error
median	58,51	112	14,2	1,51	0,01
min	-357,5	70,1	-96,8	1	0
max	401,7	176,7	93,7	9,549	0,64
Post-accented syllable					
	Slope (Hz/s)	Fp (Hz)	Range (Hz)	bend	error
median	-64,5	129	-13,1	1,05	0,001
min	-529,4	65	-135,6	1	0
max	364,7	208,5	86,6	9,549	0,25

Table1. *The range of variability of parameters describing accented and post-accented syllables.*

Evaluation of stylised contours: Demenko & Wagner



Evaluation 2: perception test

1 (identical: F0 & Sty perceived as same)
2 (a bit different: small differences in pitch height (<10Hz) perceived between F0 & Sty (e.g. pitch too high at stylized phrase end), from microprosody, errors in F0 extraction or phone or syllable segmentation.
3 (very different: F0 & Sty differ significantly – different melody, from unrecognized accents (i.e. syllable accented but not labelled “A”; cf. also #2).
Subjects could listen as often as necessary.

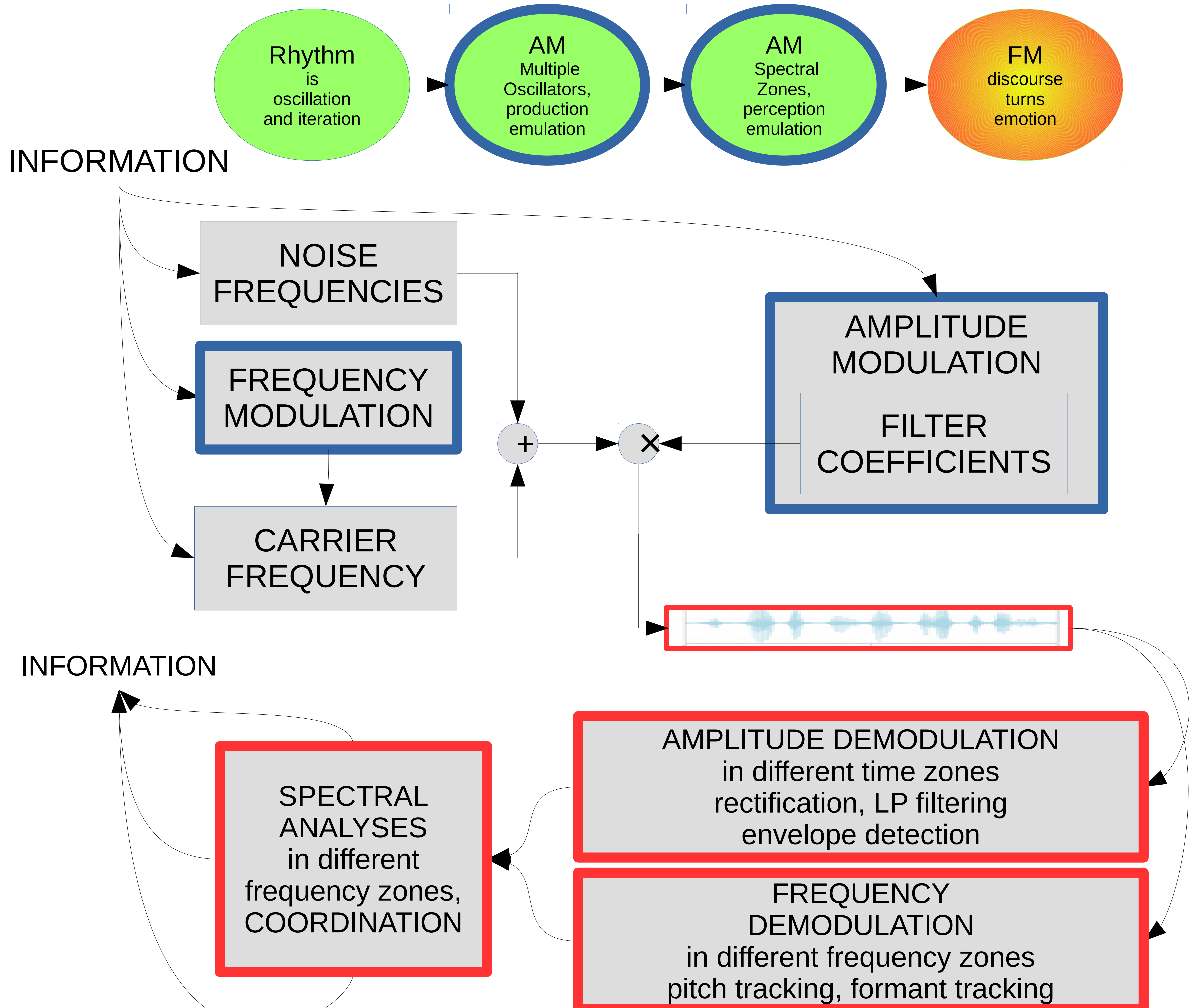
Result:

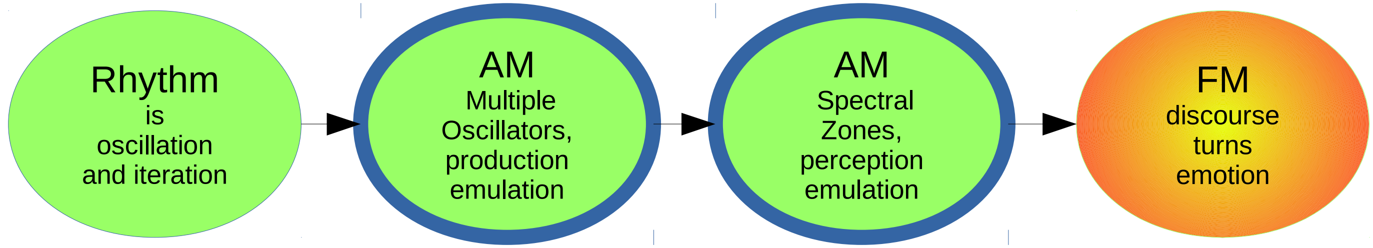
n=400	**Test**
Score 1:	256
Score 2:	68
Score 3:	76

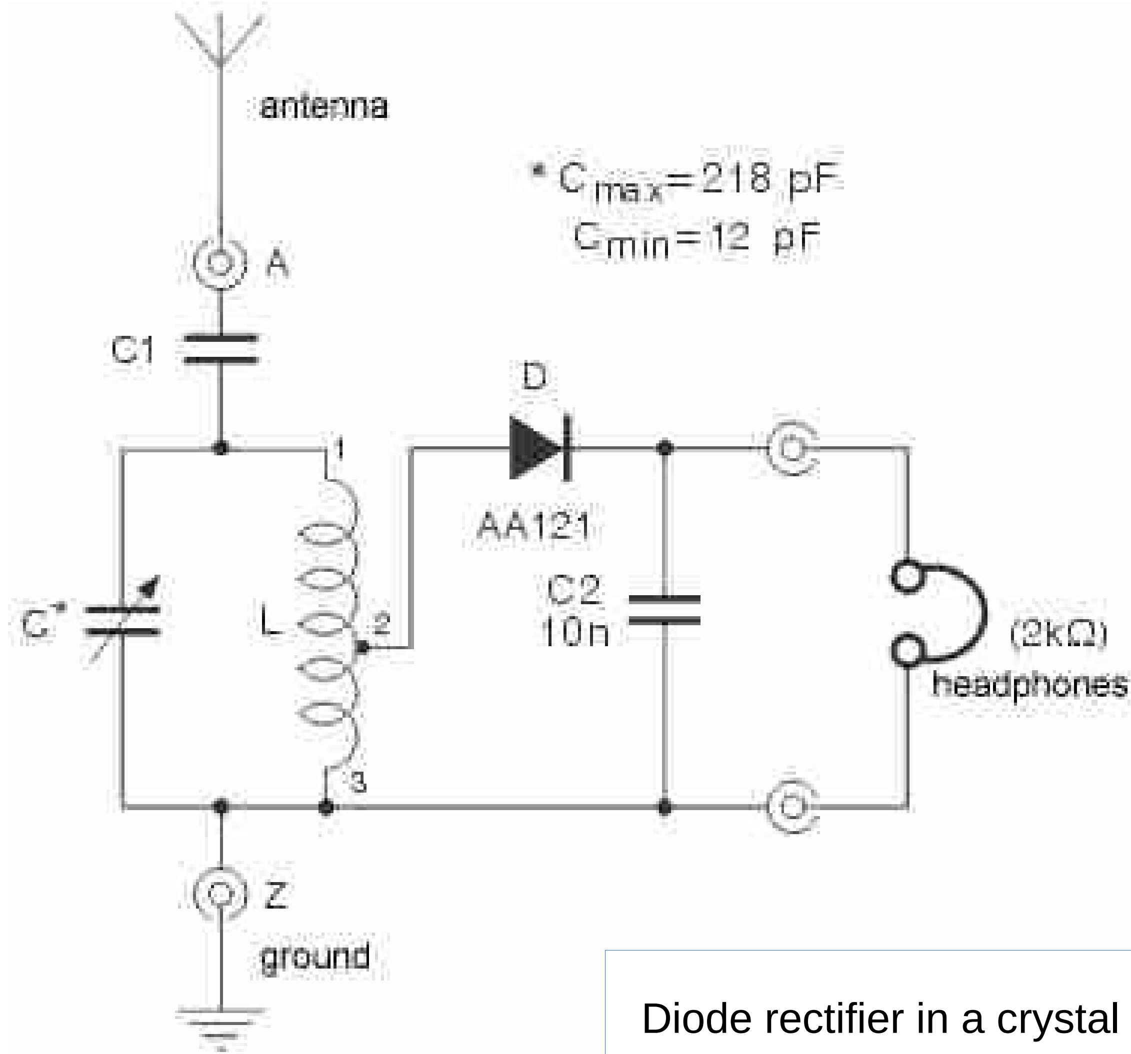
After revision of stylisation criteria, items with score 3 re-tested:
30% still with score 3.

Rhythm again: Phonetic Oscillators as Modulators

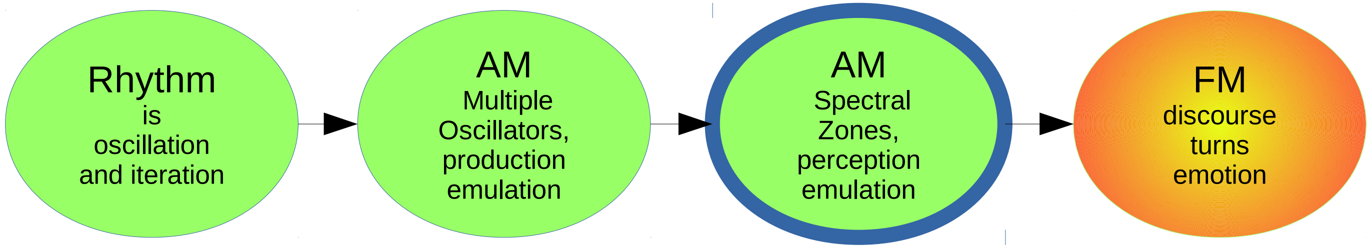
Amplitude Modulation
Frequency Modulation


Rhythm
is
oscillation
and iteration
AM
Multiple
Oscillators,
production
emulation
AM
Spectral
Zones,
perception
emulation
FM
discourse
turns
emotion
INFORMATION
NOISE
FREQUENCIES
FREQUENCY
MODULATION
CARRIER
FREQUENCY
+
×
AMPLITUDE
MODULATION
FILTER
COEFFICIENTS
INFORMATION
SPECTRAL
ANALYSES
in different
frequency zones,
COORDINATION
AMPLITUDE DEMODULATION
in different time zones
rectification, LP filtering
envelope detection
FREQUENCY
DEMODULATION
in different frequency zones
pitch tracking, formant tracking





Diode rectifier in a crystal set



Selected Work on Amplitude Envelope Demodulation Spectra

☛ **[1] Cummins**, Fred, Felix **Gers** and Jürgen **Schmidhuber**. "Language identification from prosody without explicit features." *Proc. Eurospeech*. 1999.

[2] He, Lei and Volker **Dellwo**. "A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform." In: *Proc. Interspeech* 2016, San Francisco, pp. 530-534, 2016.

[3] Hermansky, Hynek. "History of modulation spectrum in ASR." *Proc. ICASSP* 2010.

[4] Leong, Victoria and Usha **Goswami**. "Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech." *PLoS One* 10(12), 2015.

[5] Leong, Victoria, Michael A. **Stone**, Richard E. **Turner**, and Usha **Goswami**. "A role for amplitude modulation phase relationships in speech rhythm perception." *JAcSocAm*, 2014.

[6] Liss, Julie M., Sue **LeGendre**, and Andrew J. **Lotto**. "Discriminating Dysarthria Type From Envelope Modulation Spectra." *Journal of Speech, Language and Hearing Research* 53(5):1246–1255, 2010.

[7] Ludusan, Bogdan Antonio **Origlia**, Francesco **Cutugno**. "On the use of the rhythmogram for automatic syllabic prominence detection." *Proc. Interspeech*, pp. 2413-2416, 2011.

[8] Ojeda, Ariana, Ratree **Wayland**, and Andrew **Lotto**. "Speech rhythm classification using modulation spectra (EMS)." Poster presentation at the 3rd Annual Florida Psycholinguistics Meeting, 21.10.2017, U Florida. 2017.

[9] Tilsen Samuel and Keith **Johnson**. "Low-frequency Fourier analysis of speech rhythm." *Journal of the Acoustical Society of America*. 2008; 124(2):EL34–EL39. [PubMed: 18681499]

[10] Tilsen, Samuel and Amalia **Arvaniti**. "Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages." *The Journal of the Acoustical Society of America* 134, p. 628 .2013.

☛ **[11] Todd**, Neil P. McAngus and Guy J. Brown. "A computational model of prosody perception." *Proc. ICSLP* 94, pp. 127-130, 1994.

[12] Varnet, Léo, Maria Clemencia **Ortiz-Barajas**, Ramón Guevara **Erra**, Judit **Gervain**, and Christian **Lorenzi**. "A cross-linguistic study of speech modulation spectra." *JAcSocAm* 142 (4), 1976–1989, 2017.

Amplitude Modulation and Demodulation

Amplitude Envelope Modulation

↓

Amplitude Envelope Demodulation

absolute value of Hilbert transform

(or rectification & peak-picking / LP filtering)

↓

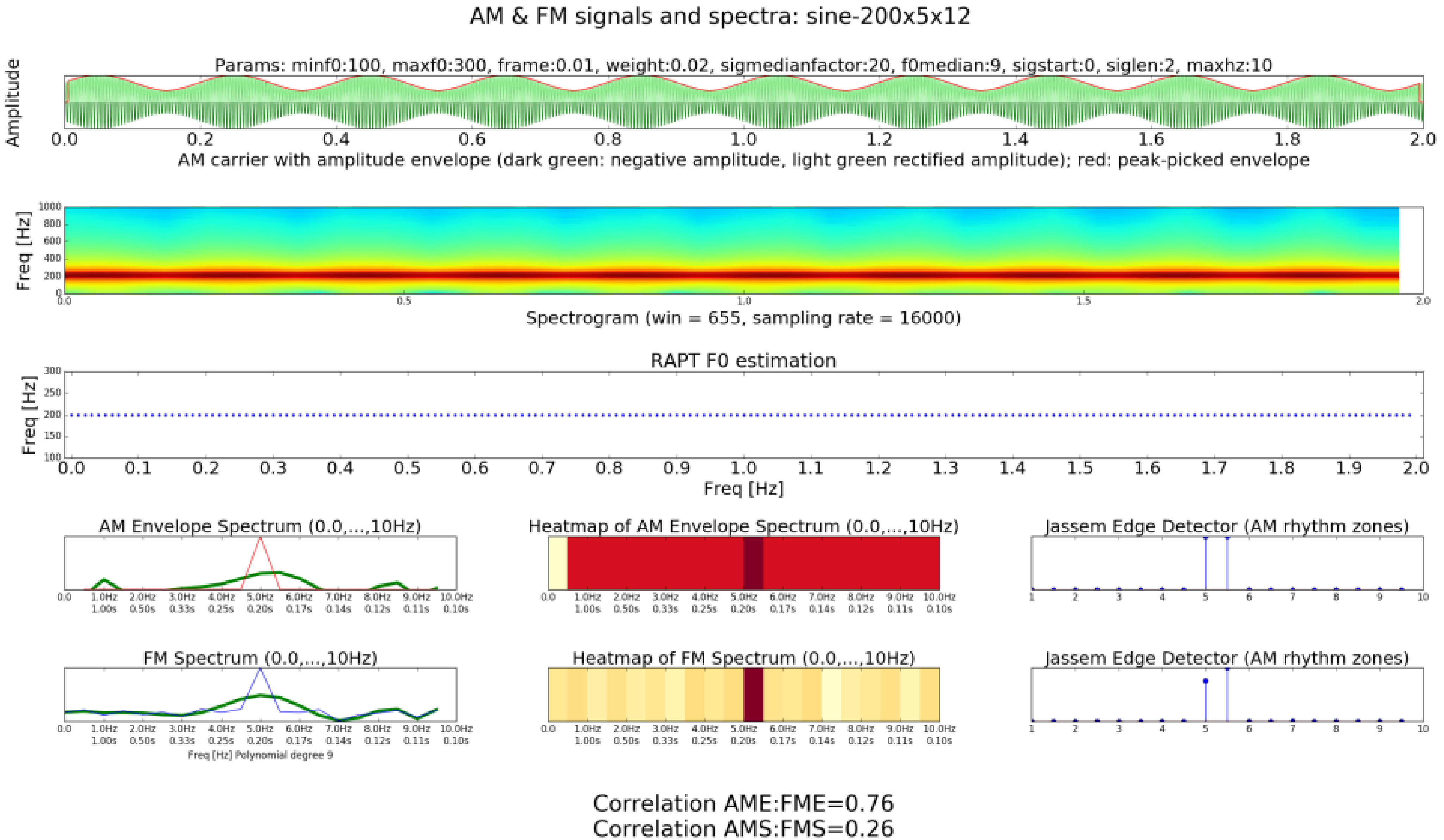
Spectral slice (FFT)

↓

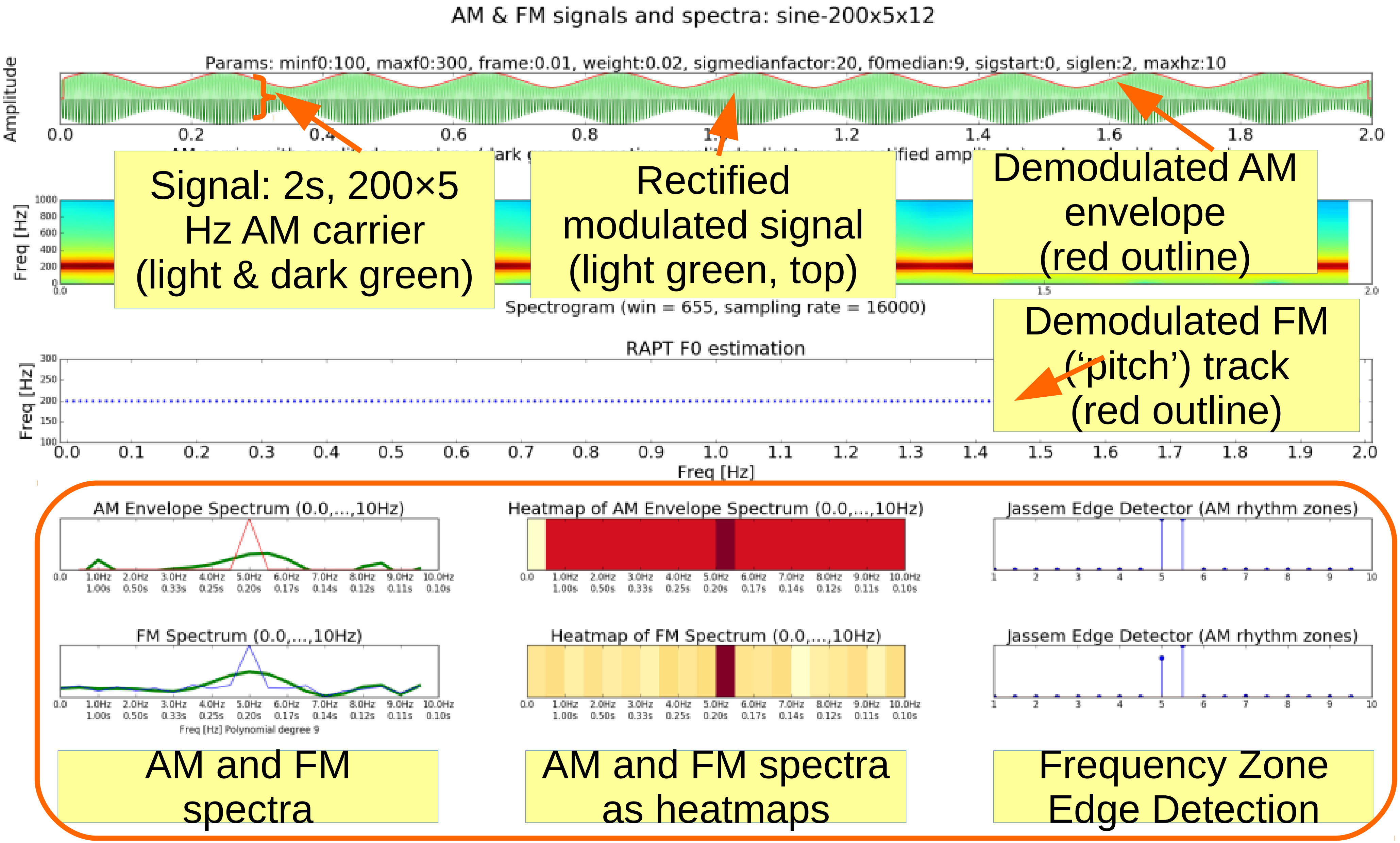
Spectral Zone Edge Detection

Amplitude Envelope Modulation Spectrum (AEMS, AMS, EMS) Frequency Zones

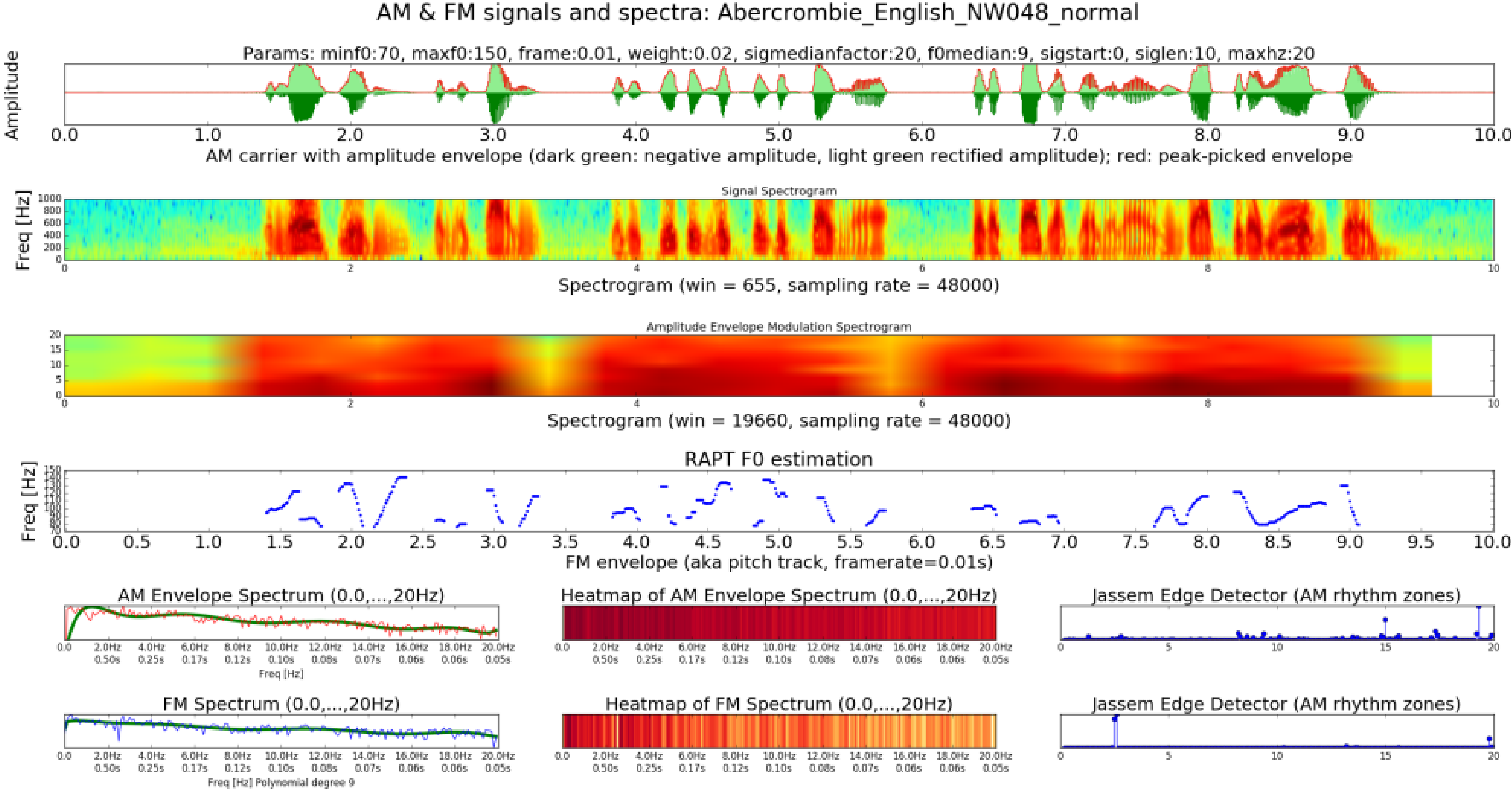
Amplitude Modulation and Demodulation



Amplitude Modulation and Demodulation

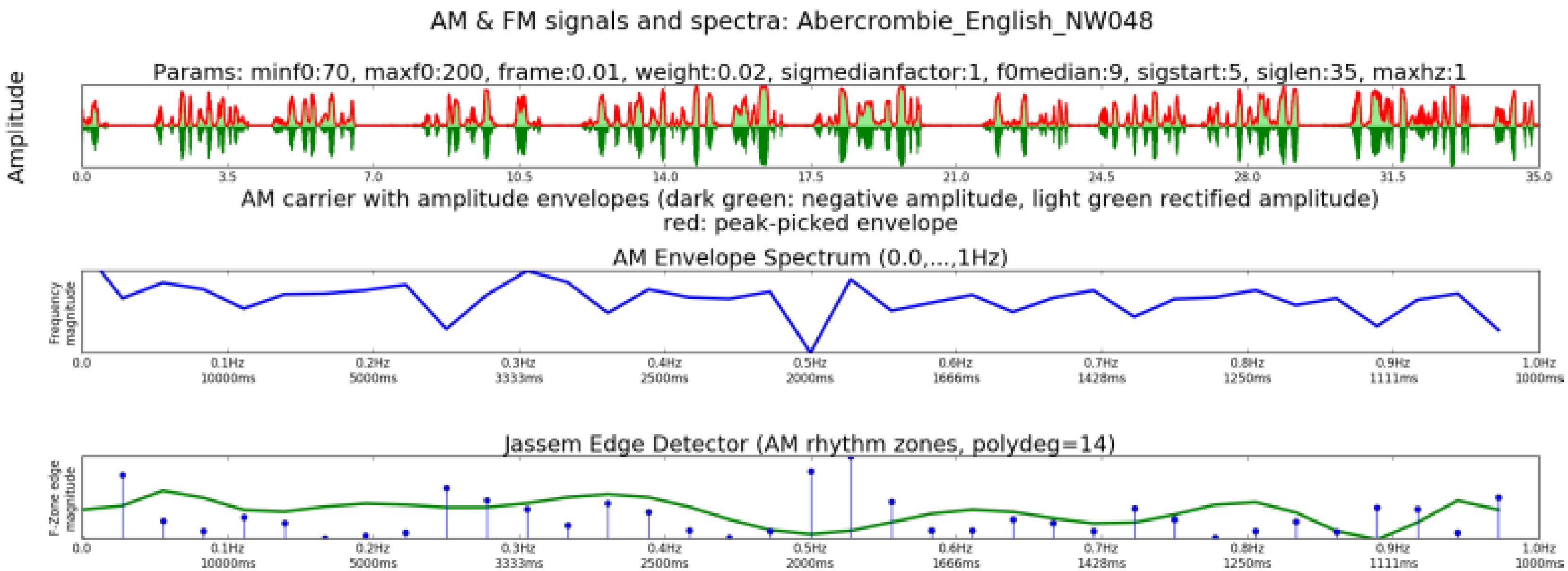


Amplitude Modulation and Demodulation



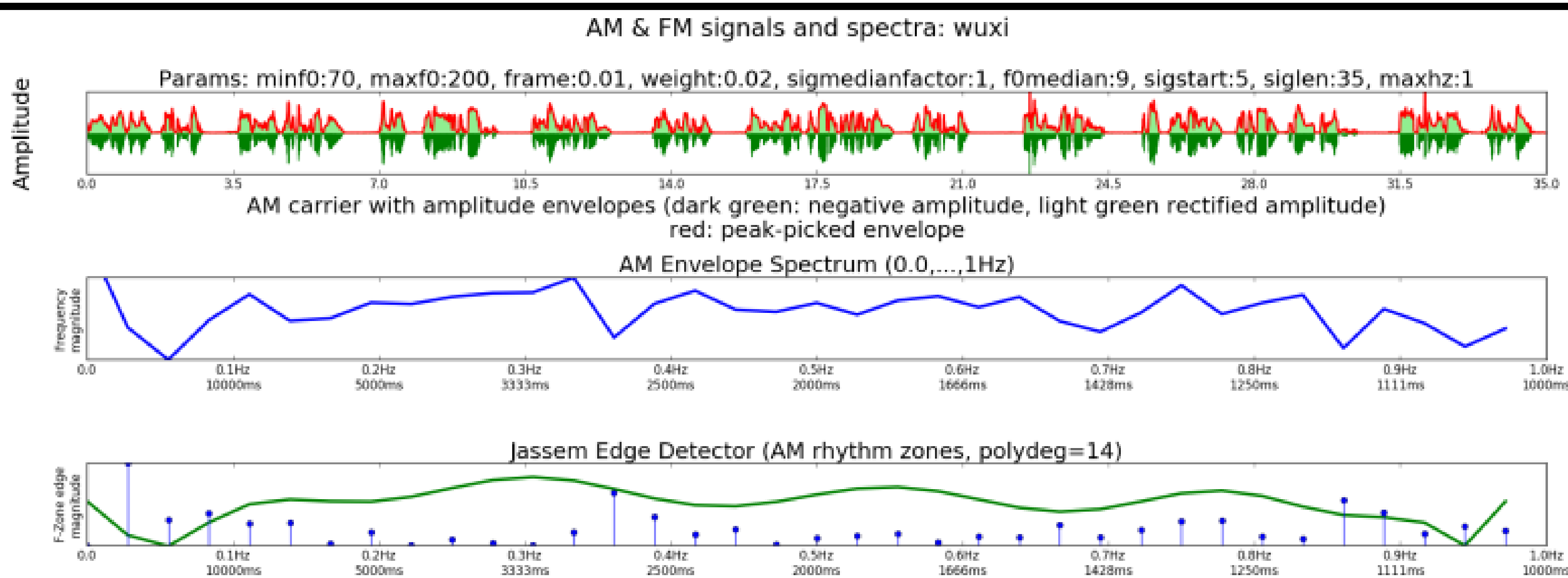
Correlation AME:FME=0.59
Correlation AMS:FMS=0.51

Amplitude Envelope Modulation Spectrum



English (RP)
Edinburgh corpus

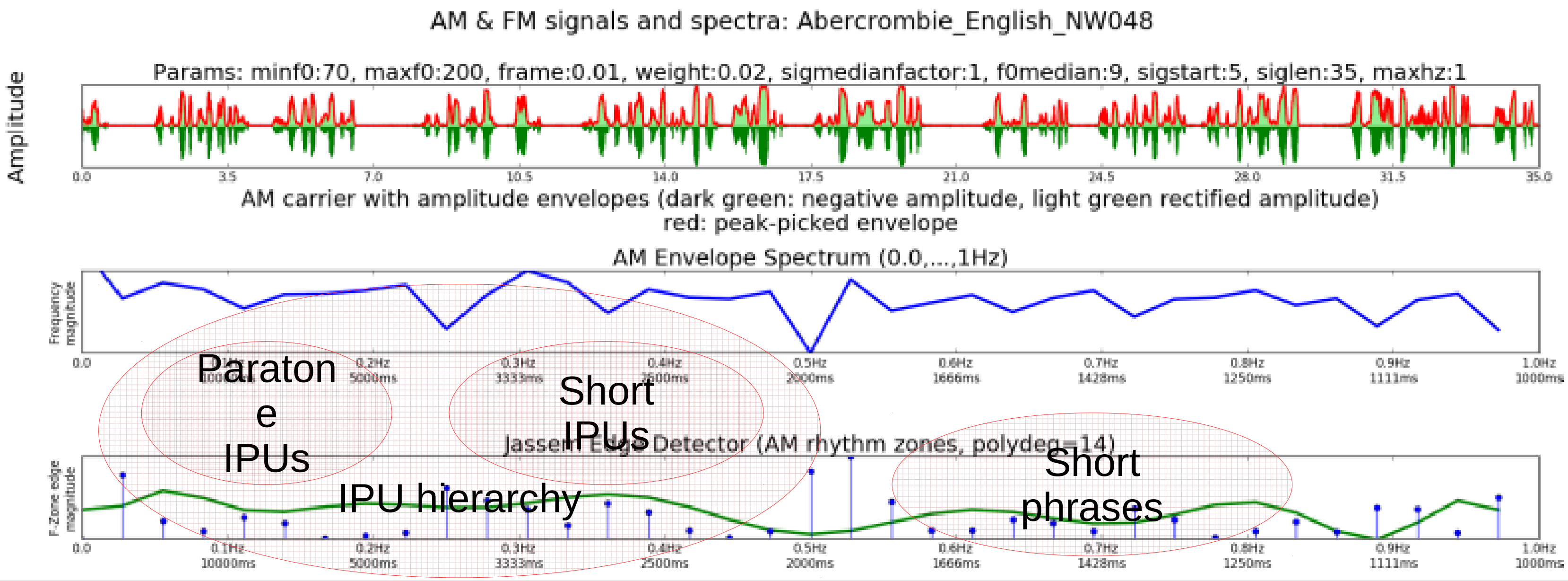
“The North Wind and the Sun”



Beijing Mandarin
Yu corpus

“bei3 feng1 gen1 tai4 yang2”

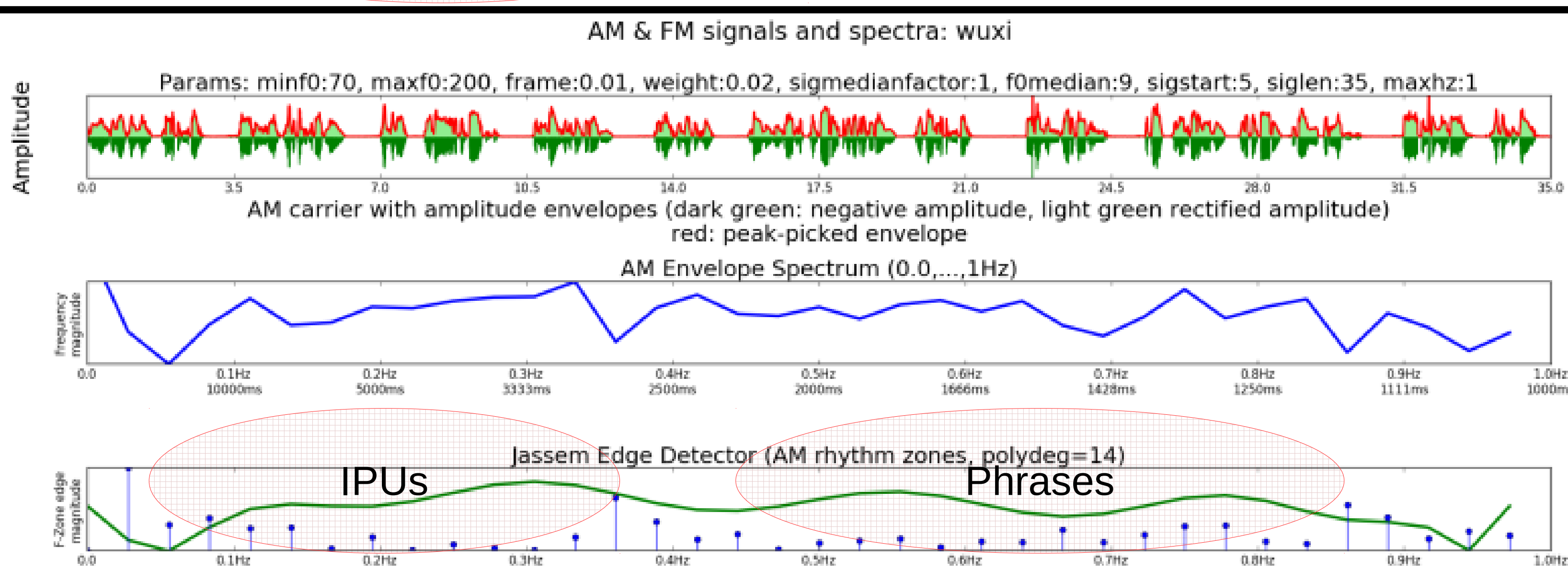
Amplitude Envelope Modulation Spectrum



English (RP)
Edinburgh corpus

"The North Wind and the Sun"

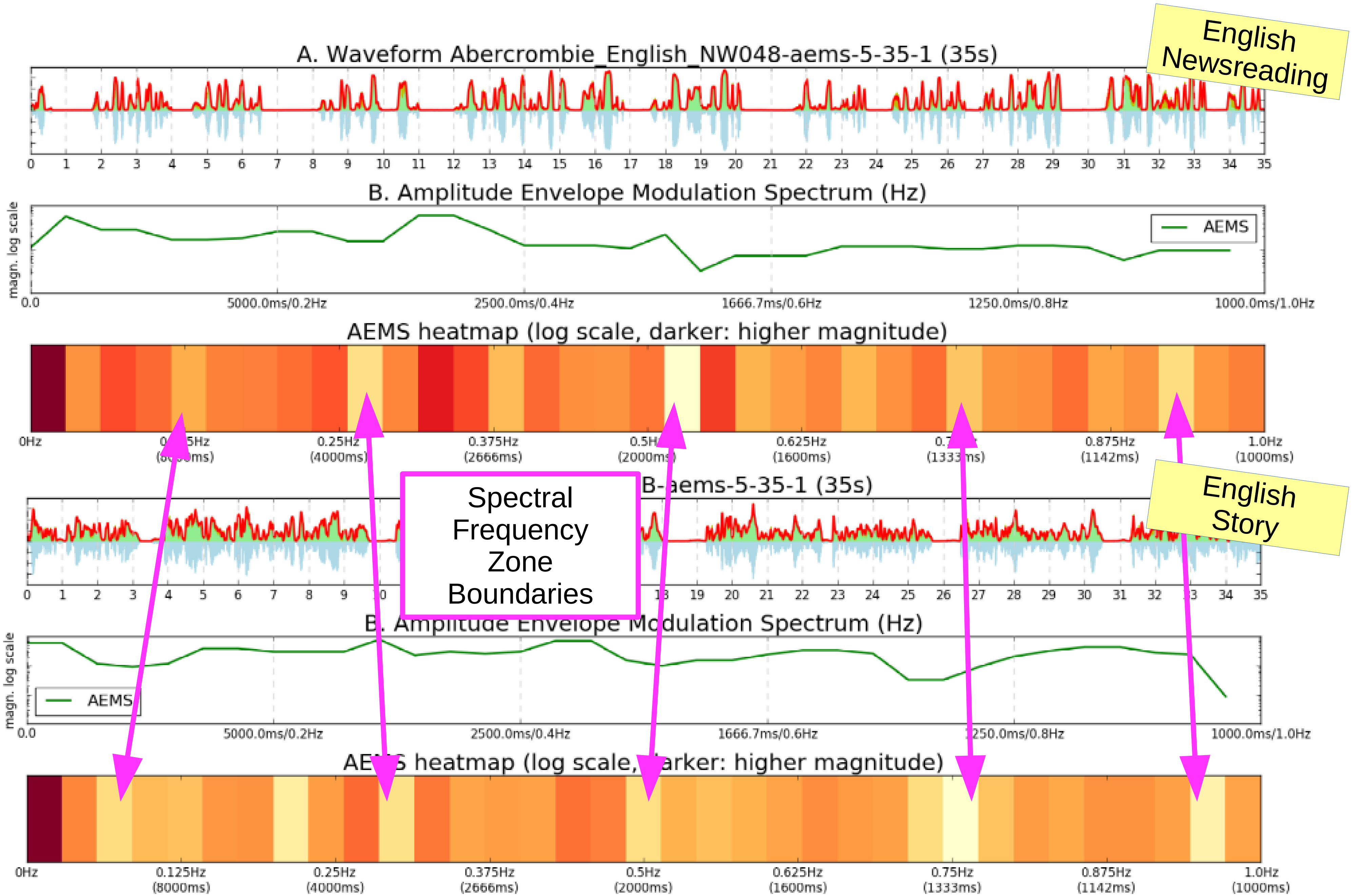
1 Hz



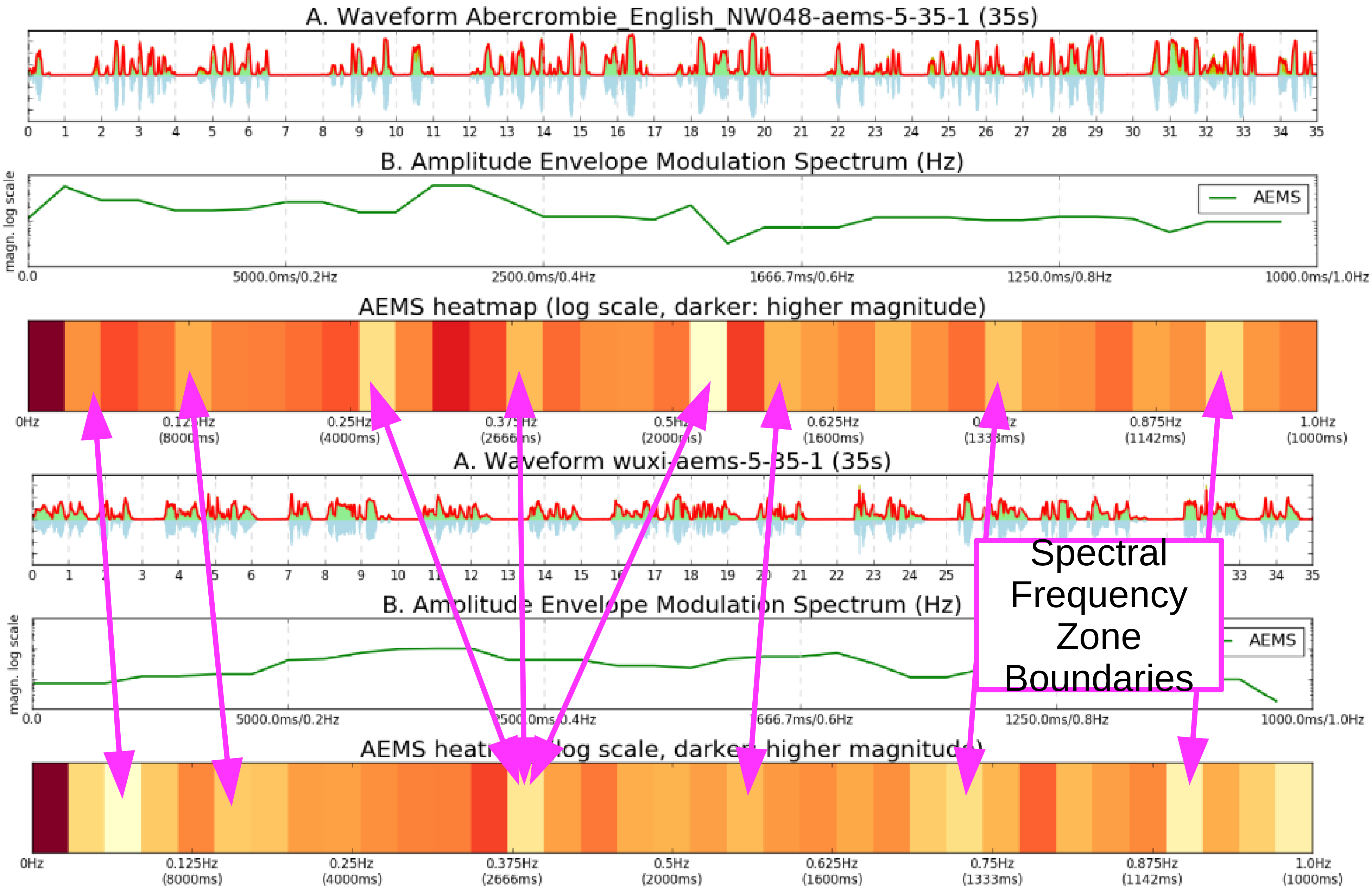
Beijing Mandarin
Yu corpus

"bei3 feng1 gen1 tai4 yang2"

Amplitude Envelope Modulation Spectrum

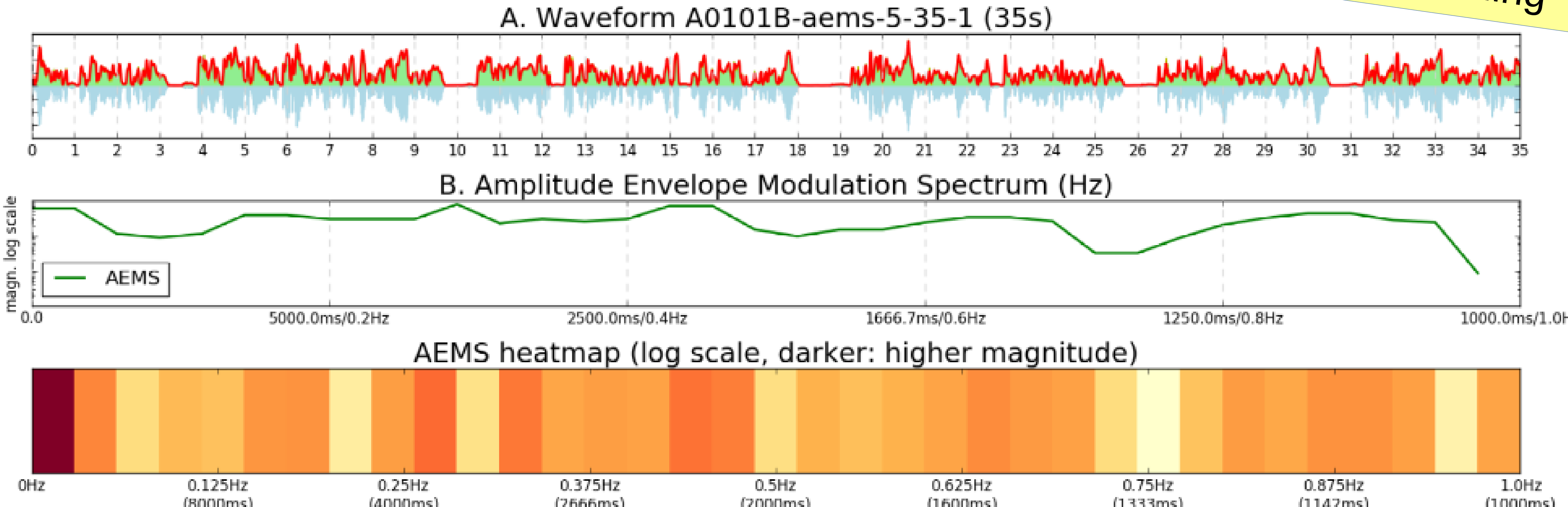


Amplitude Envelope Modulation Spectrum

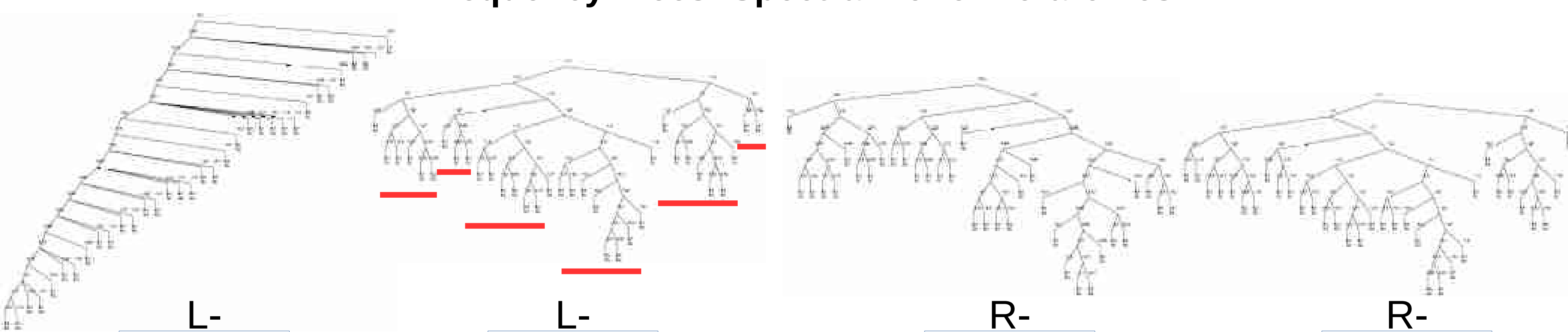


Amplitude Envelope Modulation Spectrum

English Newsreading



Frequency Trees: Spectral Zone Hierarchies



L- strong, >

L- strong, <

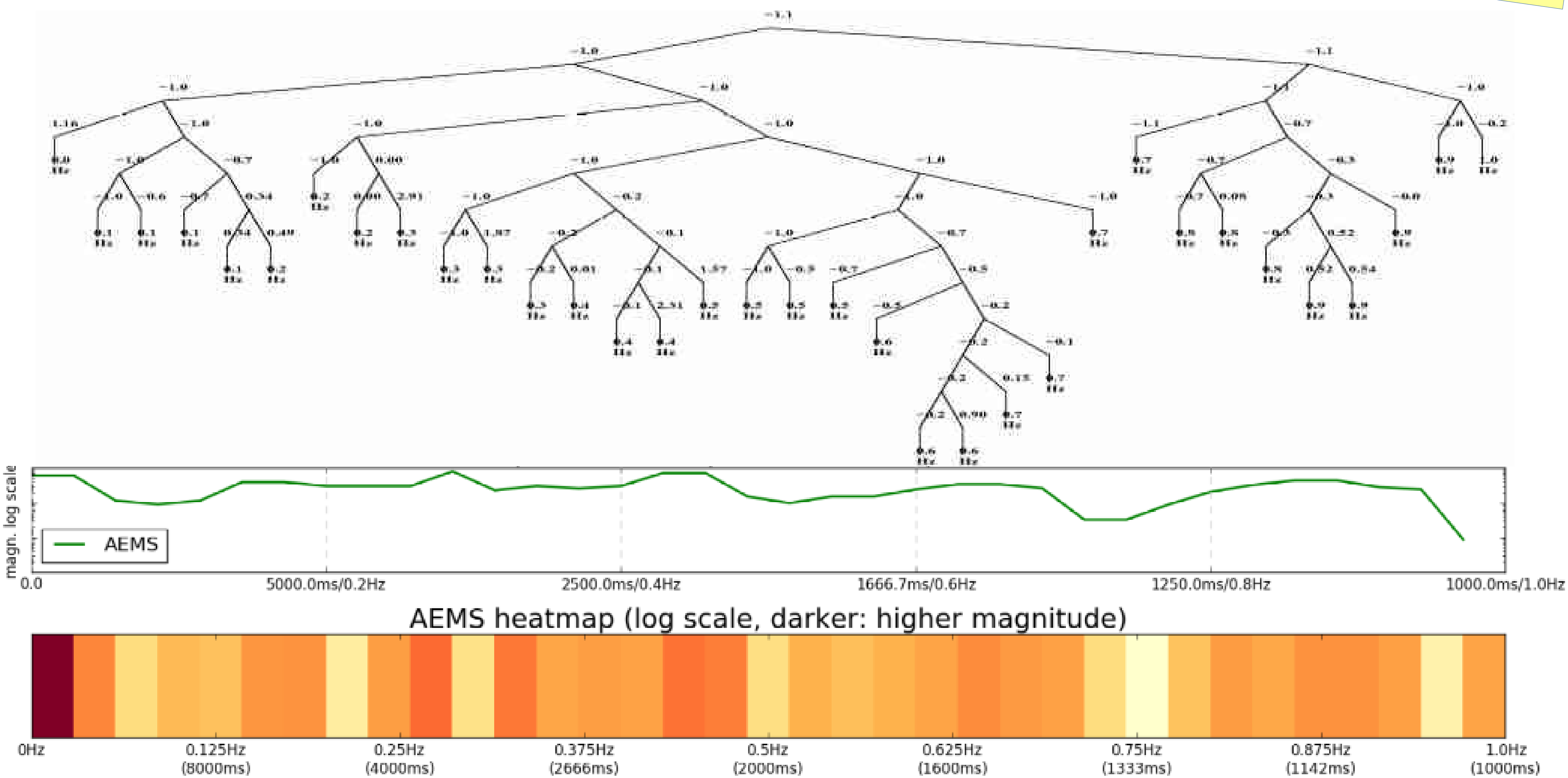
R- strong, >

R- strong, <

Amplitude Envelope Modulation Spectrum

English Newsreading

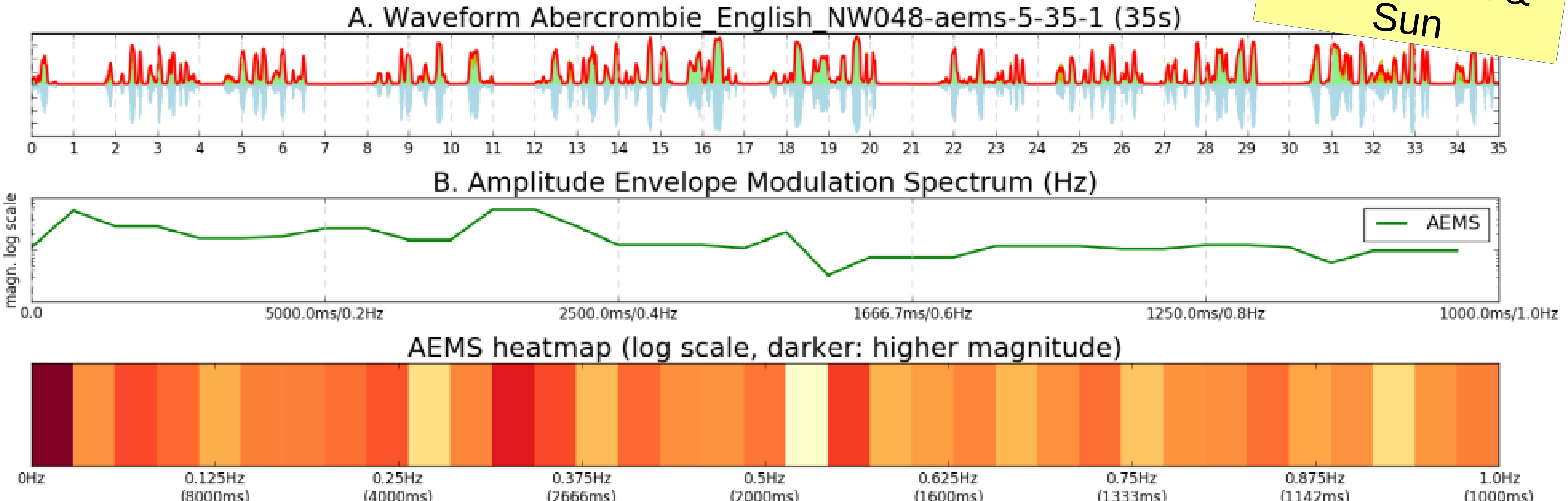
AEMS Frequency Tree



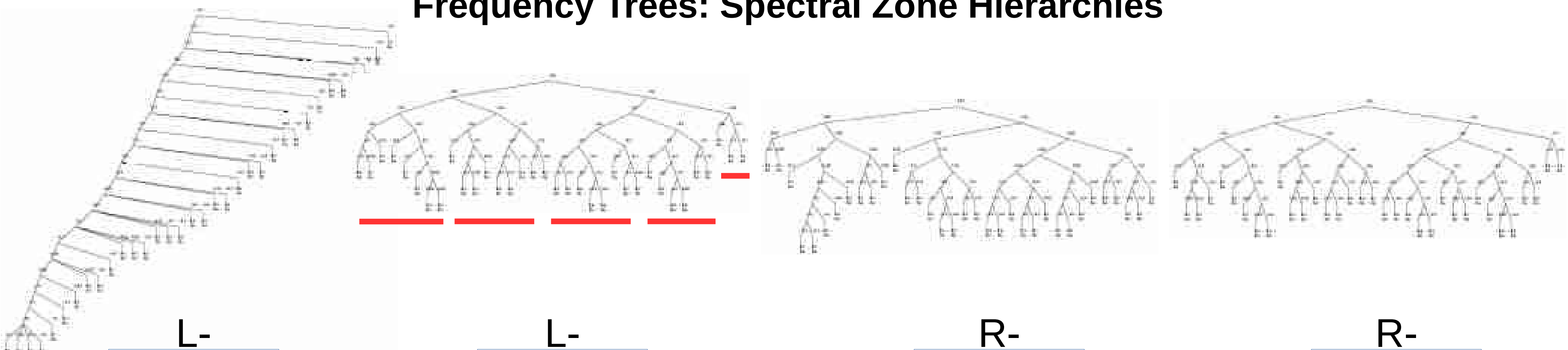
L- strong, <

Amplitude Envelope Modulation Spectrum

English North Wind & Sun



Frequency Trees: Spectral Zone Hierarchies



L-strong, >

L-strong, <

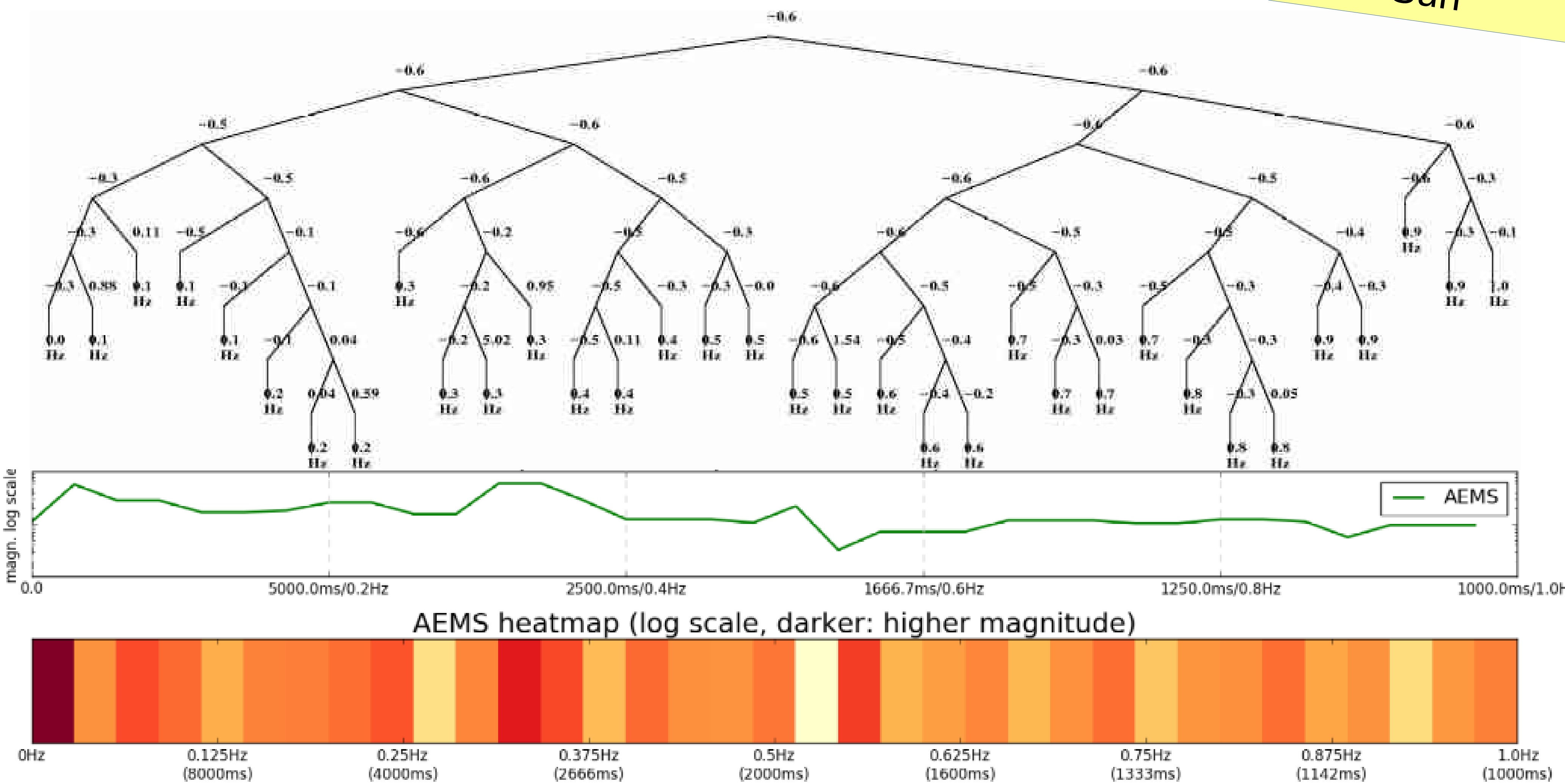
R-strong, >

R-strong, <

Amplitude Envelope Modulation Spectrum

English
North Wind &
Sun

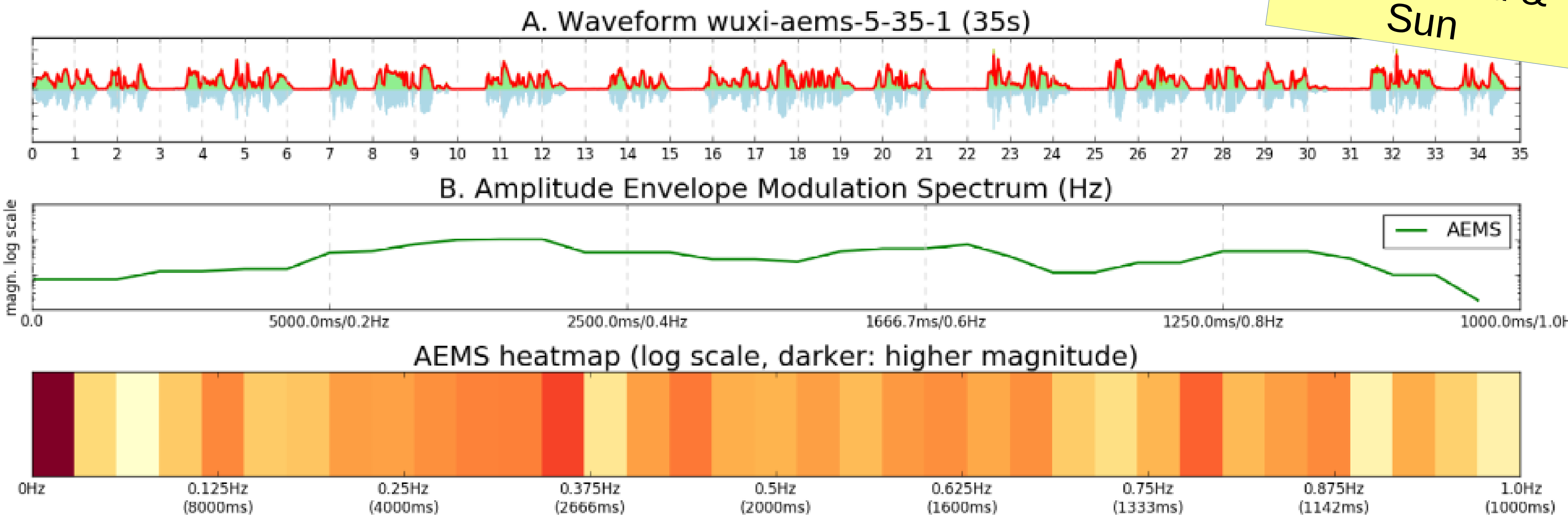
AEMS Frequency Tree



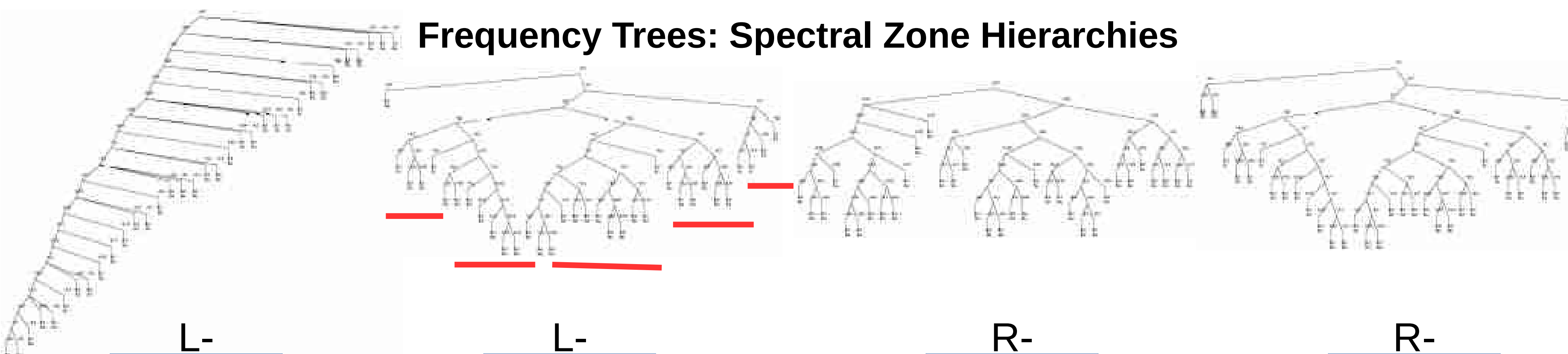
L-
strong,
<

Amplitude Envelope Modulation Spectrum

Mandarin North Wind & Sun



Frequency Trees: Spectral Zone Hierarchies



L-strong, >

L-strong, <

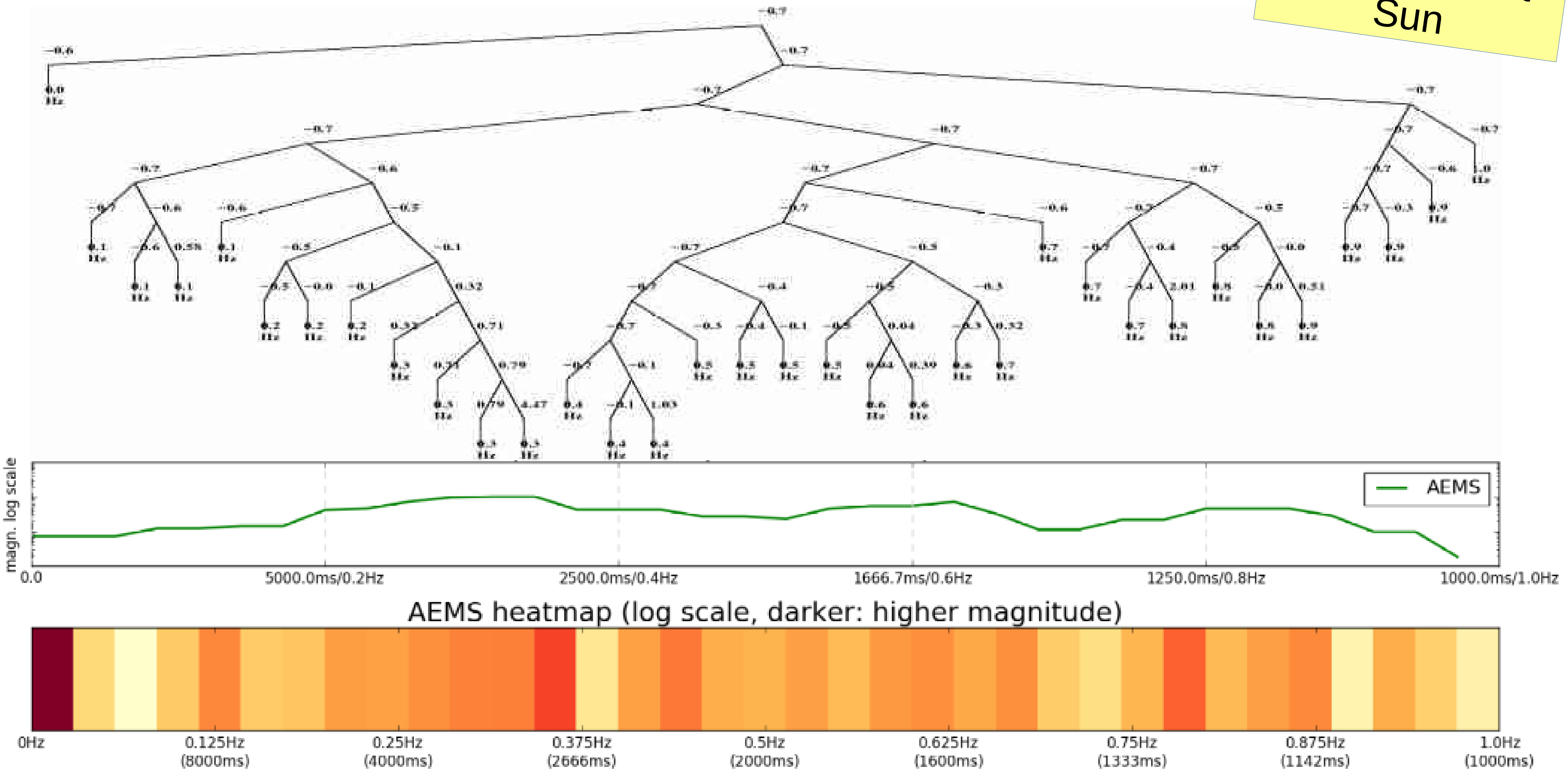
R-strong, >

R-strong, <

Amplitude Envelope Modulation Spectrum

Mandarin North Wind & Sun

AEMS Frequency Tree

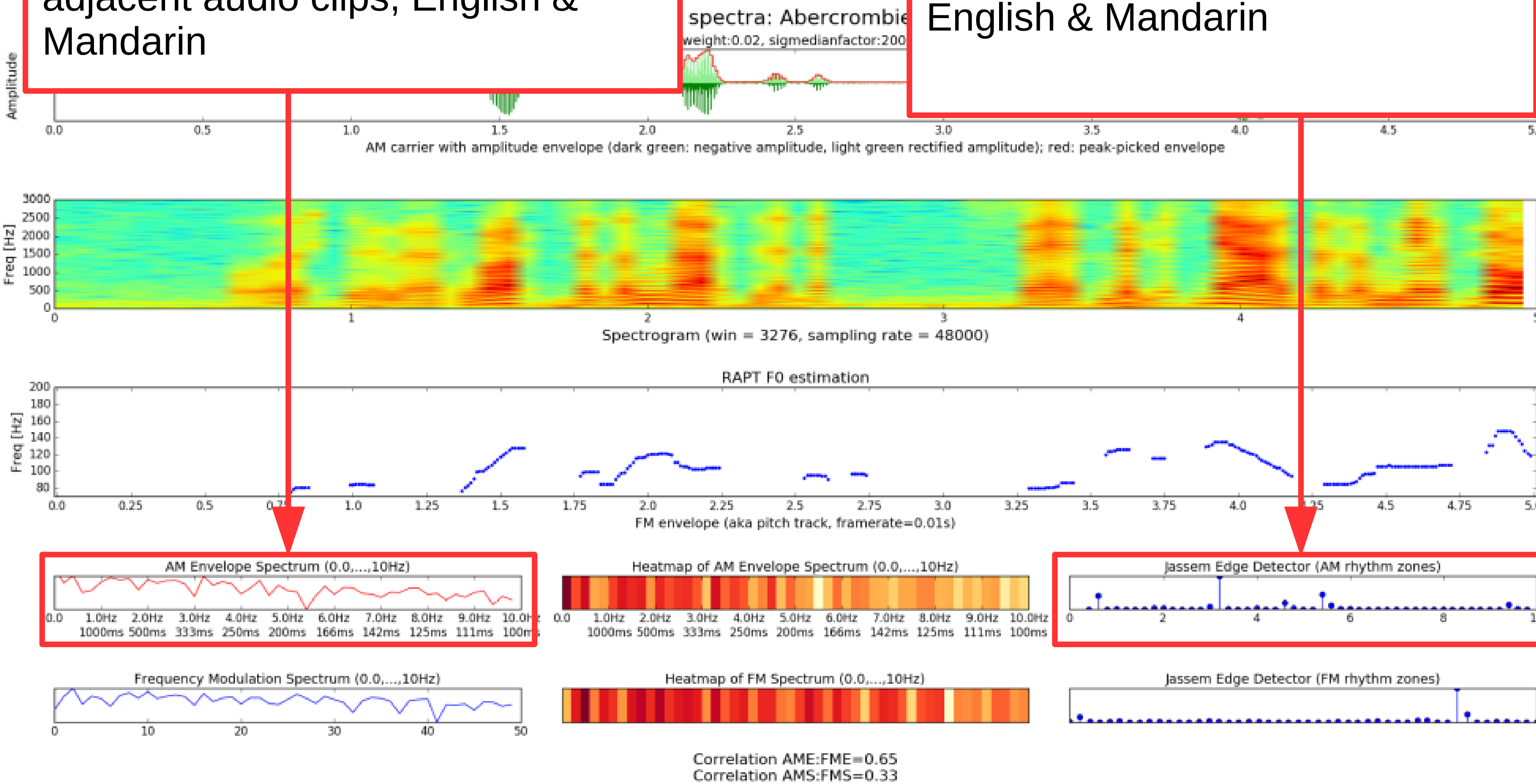


L-strong, <

Amplitude Envelope Modulation Spectrum

Next step:
Distance analysis of AEMS of 5s adjacent audio clips, English & Mandarin

Next but one step:
Conventional analysis of AEMS edges in 5 second audio clips, English & Mandarin



Amplitude Envelope Modulation Spectrum

Data:

“The North Wind and the Sun”

Male, English: 40s

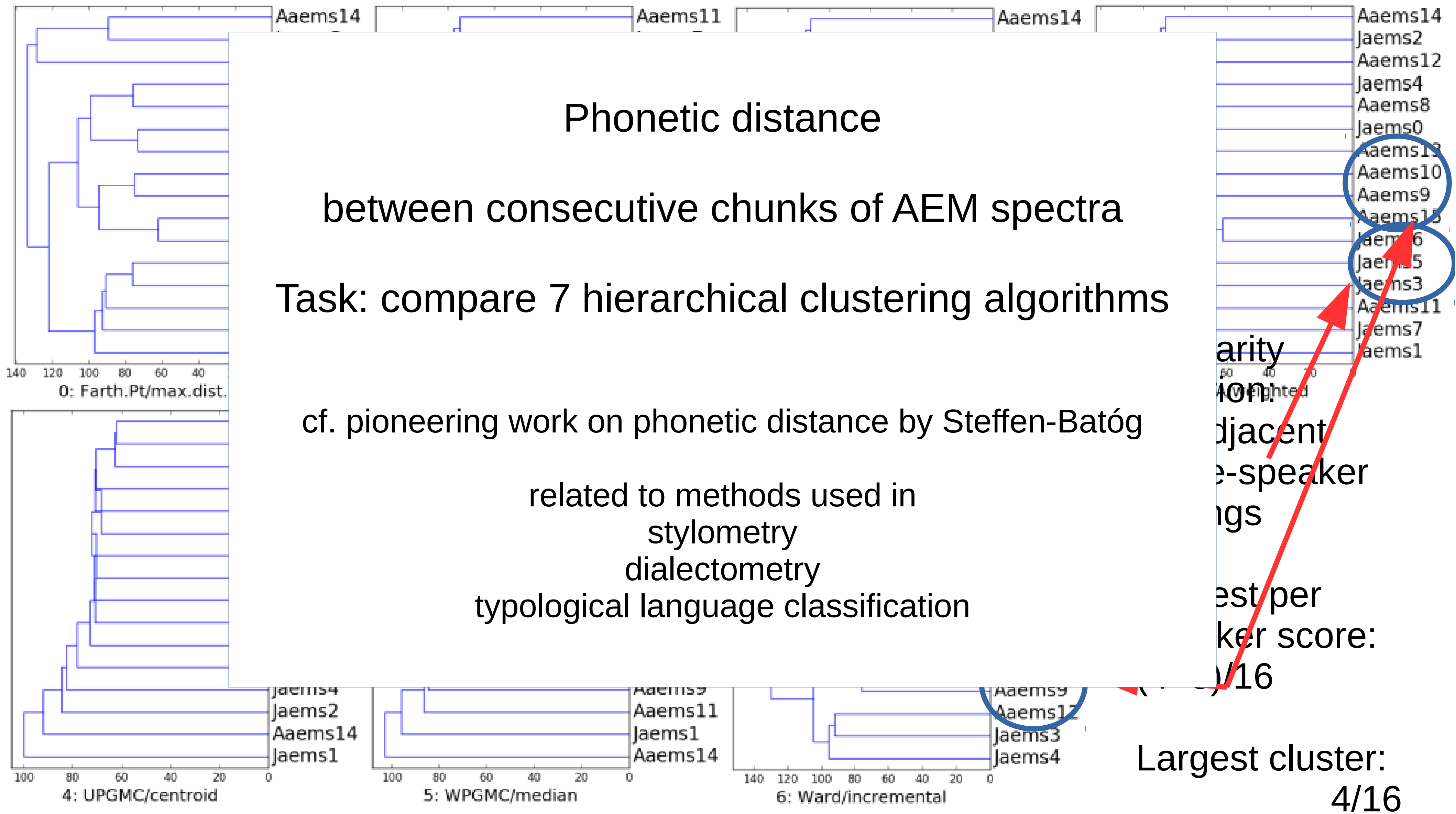
Female, Mandarin: 40s

Method:

Comparison of non-overlapping adjacent 5s audio chunks

- offsets into recording: 0, 5, 10, 15, 20, 25, 30, 35
- AEMS for each chunk
- Inter-speaker comparison (AEMS pointwise means, *r*=0.82)
- Comparison by hierarchical similarity / distance

Amplitude Envelope Modulation Spectrum



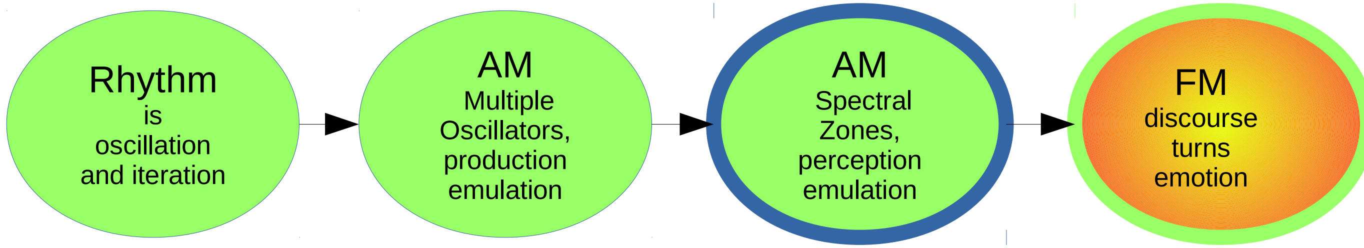
Phonetic distance

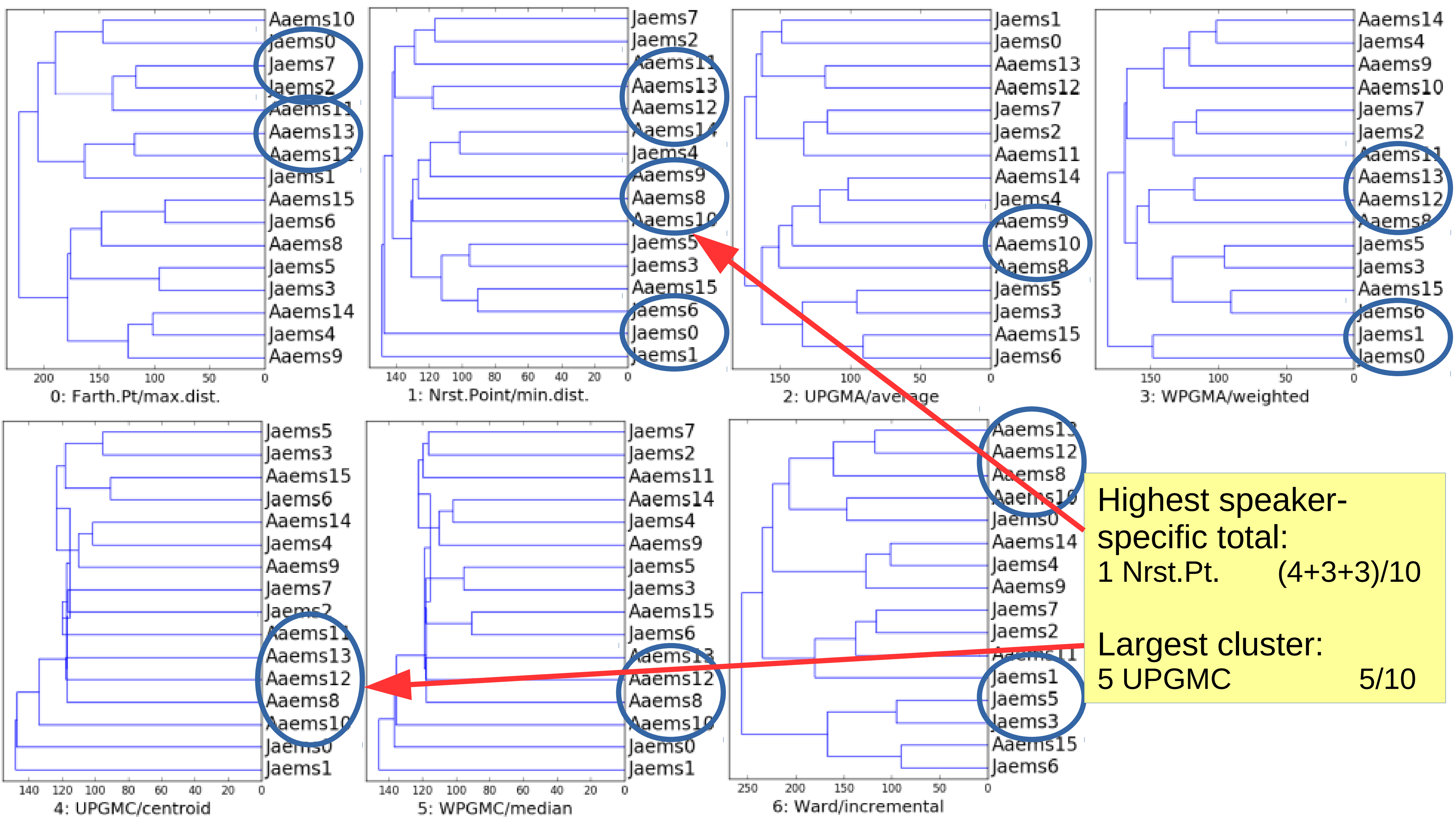
between consecutive chunks of AEM spectra

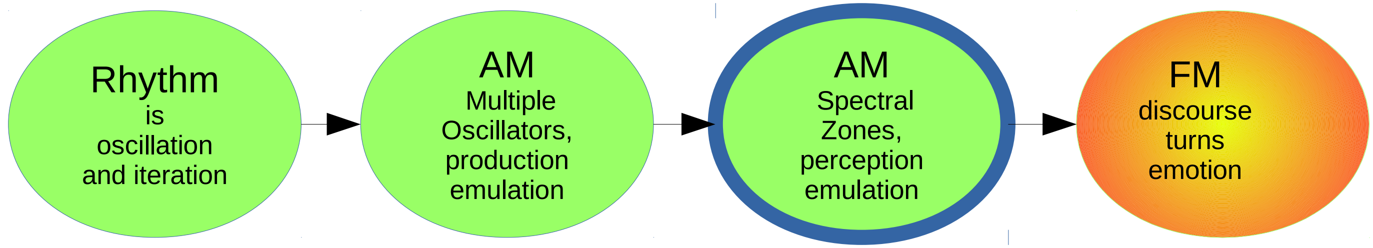
Task: compare 7 hierarchical clustering algorithms

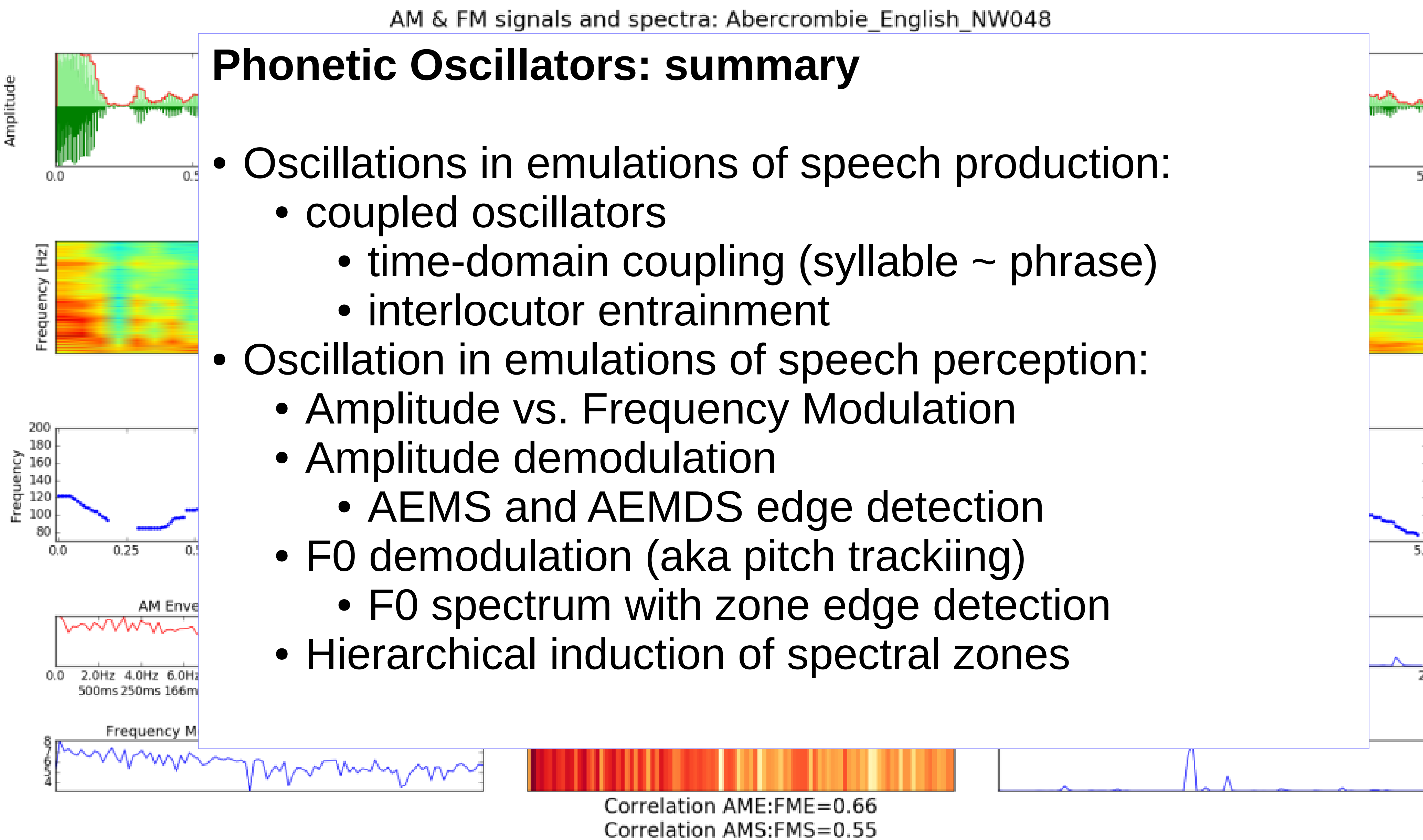
cf. pioneering work on phonetic distance by Steffen-Batóg

related to methods used in
stylometry
dialectometry
typological language classification


Rhythm
is
oscillation
and iteration
AM
Multiple
Oscillators,
production
emulation
AM
Spectral
Zones,
perception
emulation
FM
discourse
turns
emotion

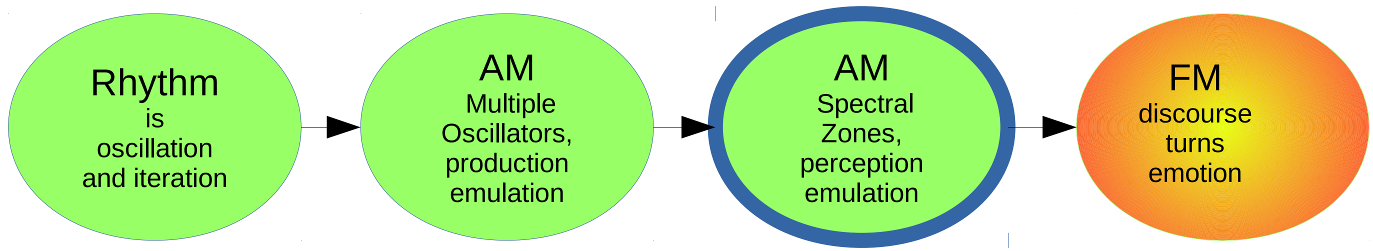

0: Farth.Pt/max.dist.
1: Nrst.Point/min.dist.
2: UPGMA/average
3: WPGMA/weighted
4: UPGMC/centroid
5: WPGMC/median
6: Ward/incremental
Highest speaker-specific total:
1 Nrst.Pt. (4+3+3)/10
Largest cluster:
5 UPGMC 5/10

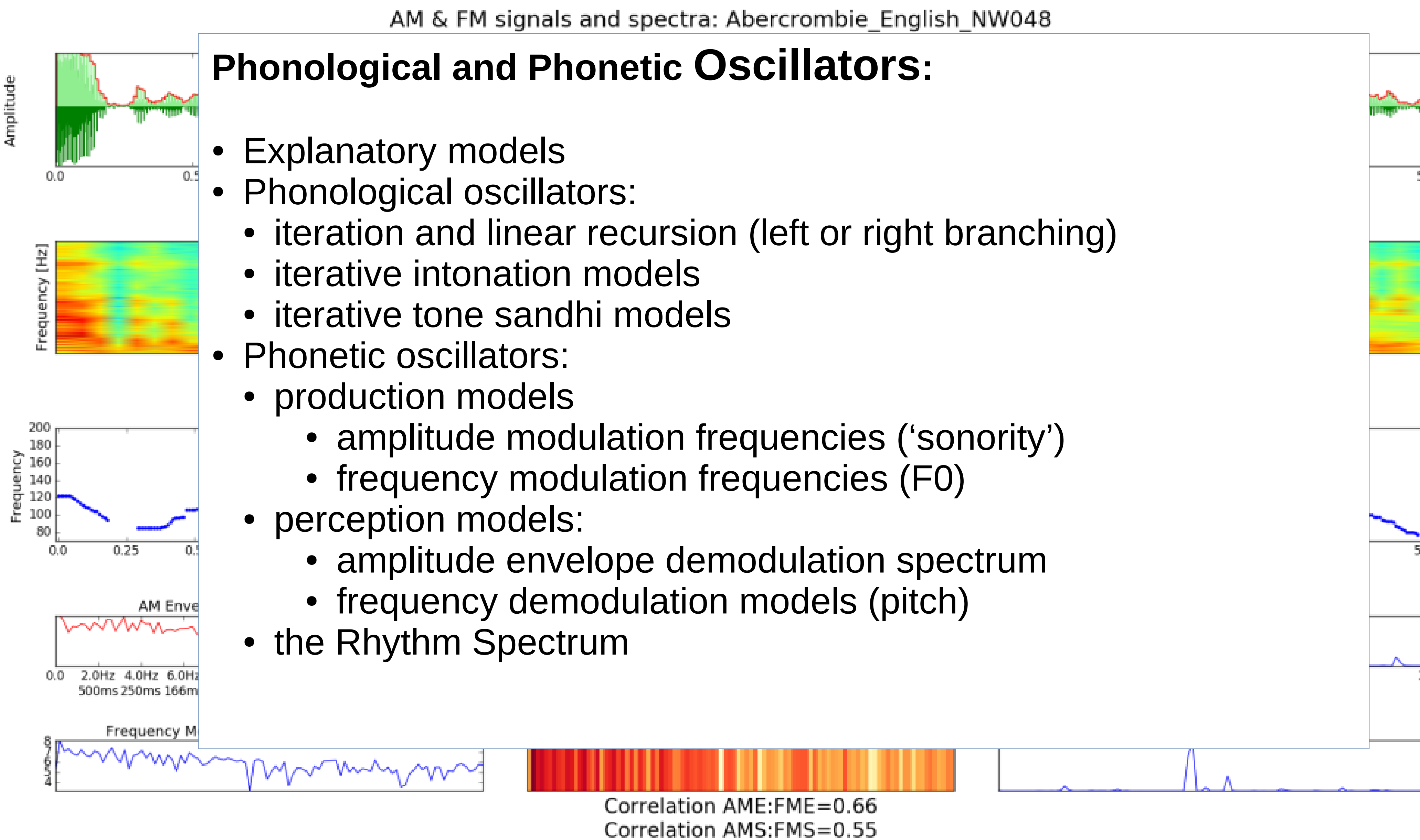




Phonetic Oscillators: summary

- Oscillations in emulations of speech production:
  - coupled oscillators
    - time-domain coupling (syllable ~ phrase)
    - interlocutor entrainment
- Oscillation in emulations of speech perception:
  - Amplitude vs. Frequency Modulation
  - Amplitude demodulation
    - AEMS and AEMDS edge detection
  - F0 demodulation (aka pitch trackiing)
    - F0 spectrum with zone edge detection
  - Hierarchical induction of spectral zones





Phonological and Phonetic Oscillators:

- Explanatory models
- Phonological oscillators:
  - iteration and linear recursion (left or right branching)
  - iterative intonation models
  - iterative tone sandhi models
- Phonetic oscillators:
  - production models
    - amplitude modulation frequencies (‘sonority’)
    - frequency modulation frequencies (F0)
  - perception models:
    - amplitude envelope demodulation spectrum
    - frequency demodulation models (pitch)
  - the Rhythm Spectrum

Summary:

From the Phonology of Prosody
to
the Phonetics of Prosody

Melody: Pitch Patterns

From Time to Frequency:
the Rhythm Spectrum

Conclusion:

… thinking outside the box

Summary:

From the Phonology of Prosody
to
the Phonetics of Prosody

Melody: Pitch Patterns
The Rhythm Spectrum

Conclusion:

… thinking outside the box

Thank you!

谢谢！