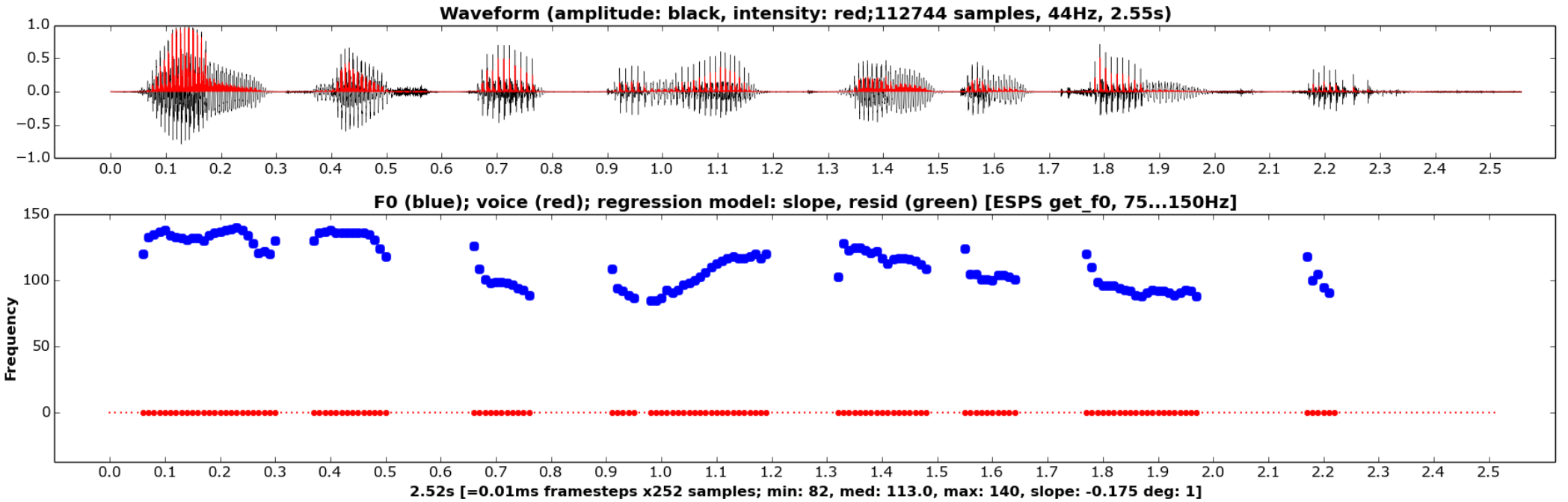
Prosody: Speech Rhythms and Melodies

6. Generalising Pitch: Stylisation

Dafydd Gibbon

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

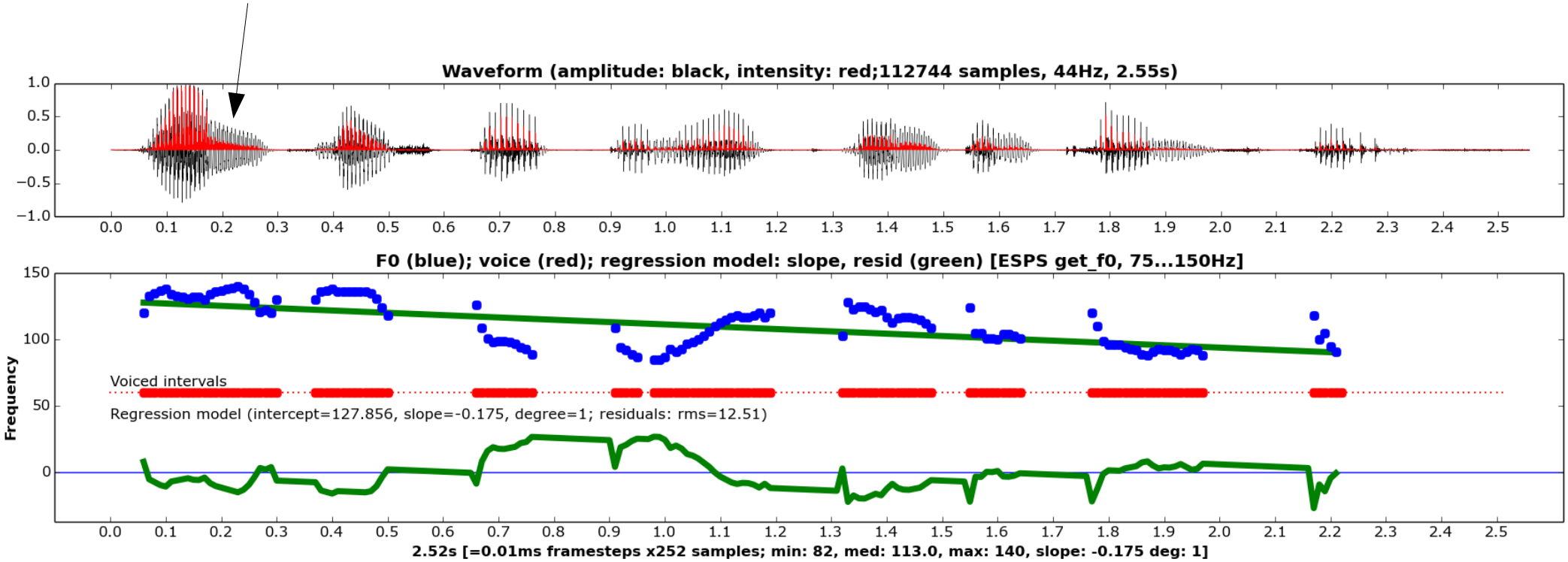
Revision: What exactly does this graph show?



Endlich gab der Nordwind den Kampf auf.

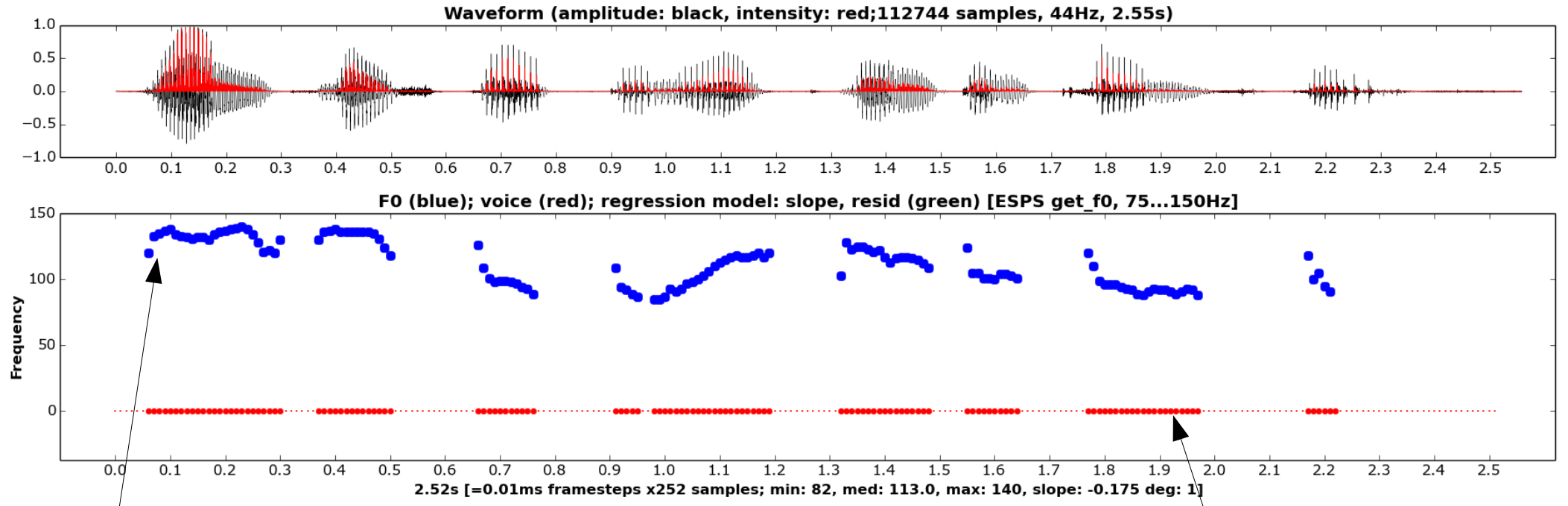
Phonetic correlates of phonological categories

amplitude; *intensity* = *f*(*amplitude*^{2}) → factor in stress, focus, contrast, emphasis



Endlich gab der Nordwind den Kampf auf.

Representations of phonetic time functions

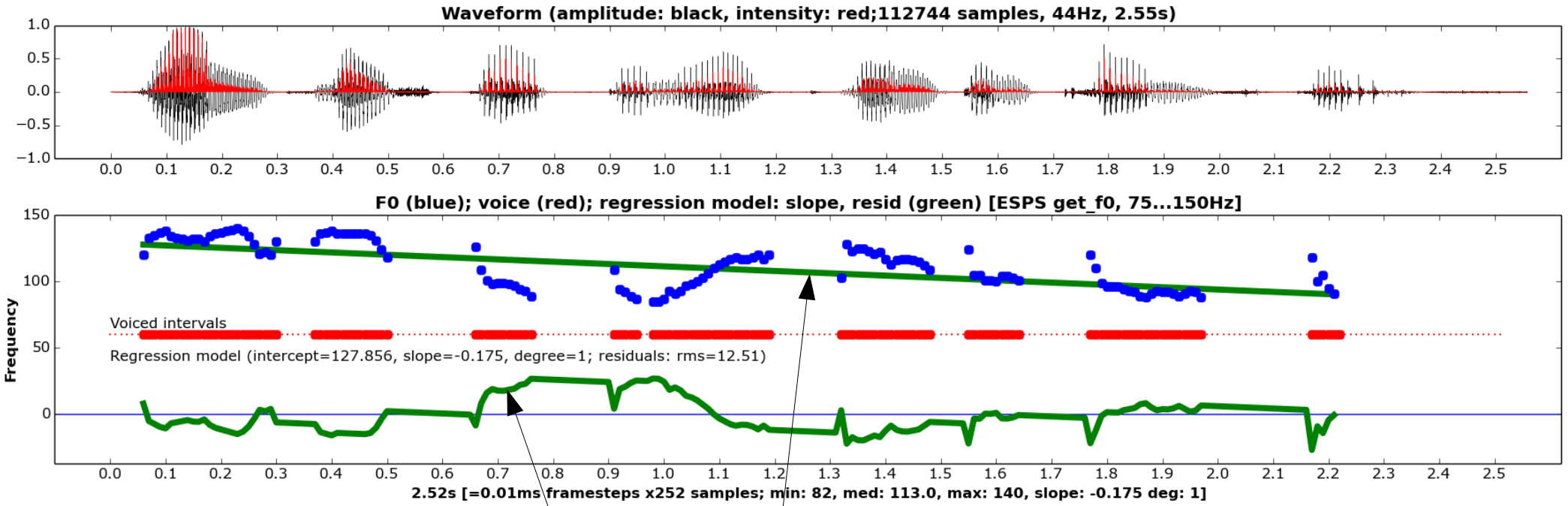


pitch track: fundamental frequency, f0

voicing

Endlich gab der Nordwind den Kampf auf.

Phonetic correlates of phonological categories

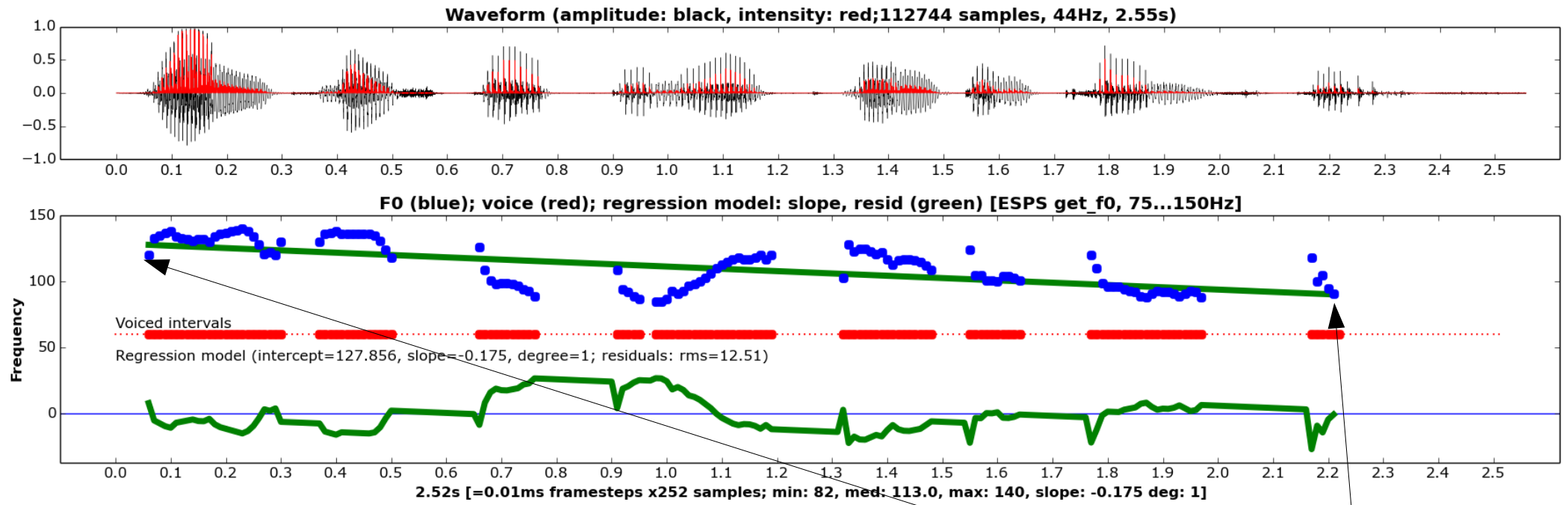


global downtrend model
→ declination model

pitch perturbation model → accent/tone and microprosody model

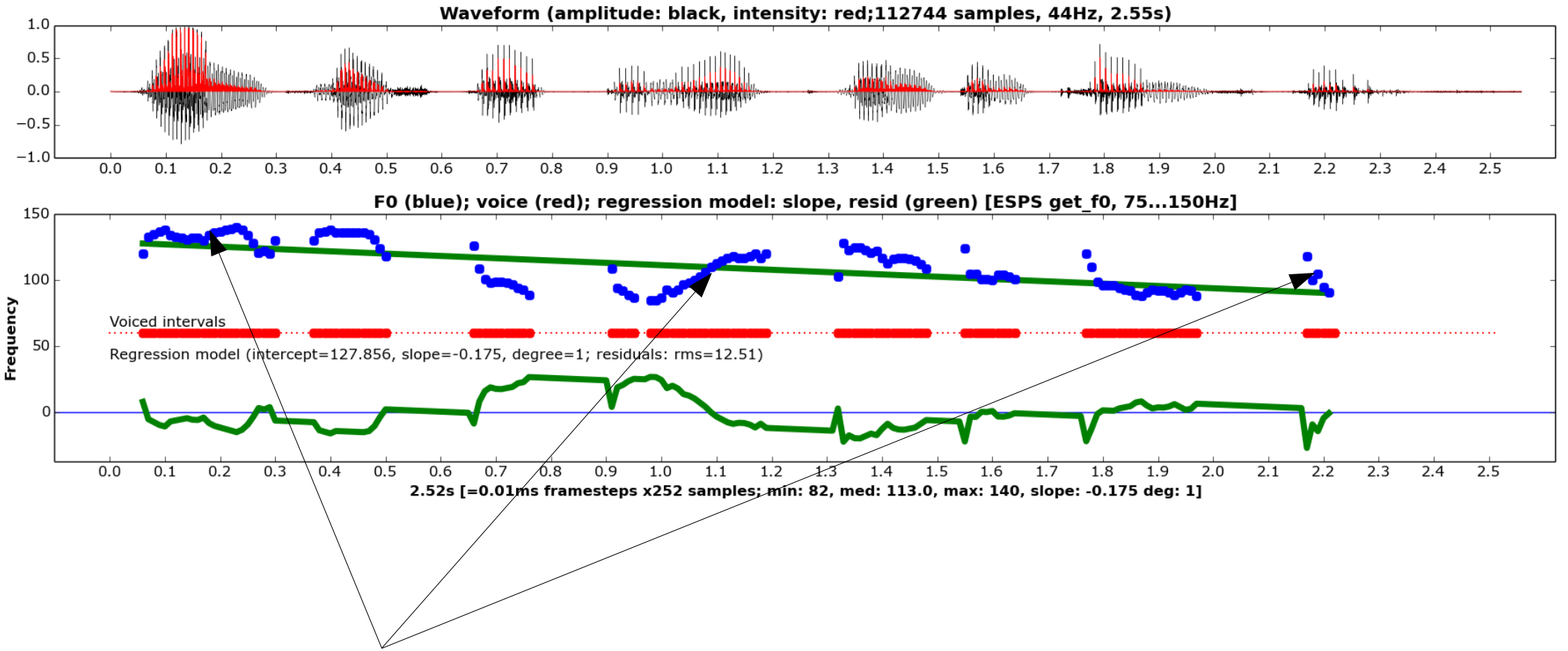
Endlich gab der Nordwind den Kampf auf.

Phonetic correlates of phonological categories



initial & final pitch
→ boundary tone

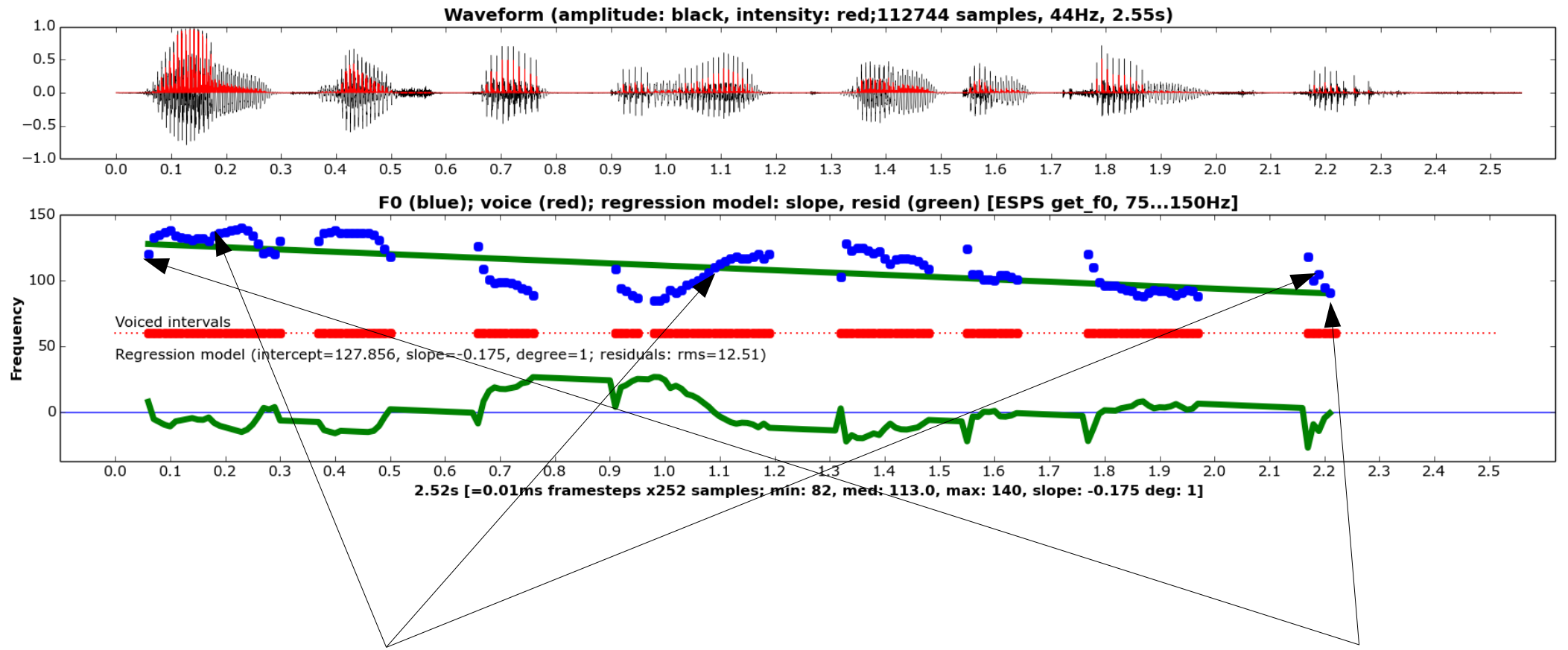
Phonetic correlates of phonological categories



pitch deviations → local pitch accents (or tones)

Endlich gab der Nordwind den Kampf auf.

Phonetic correlates of phonological categories



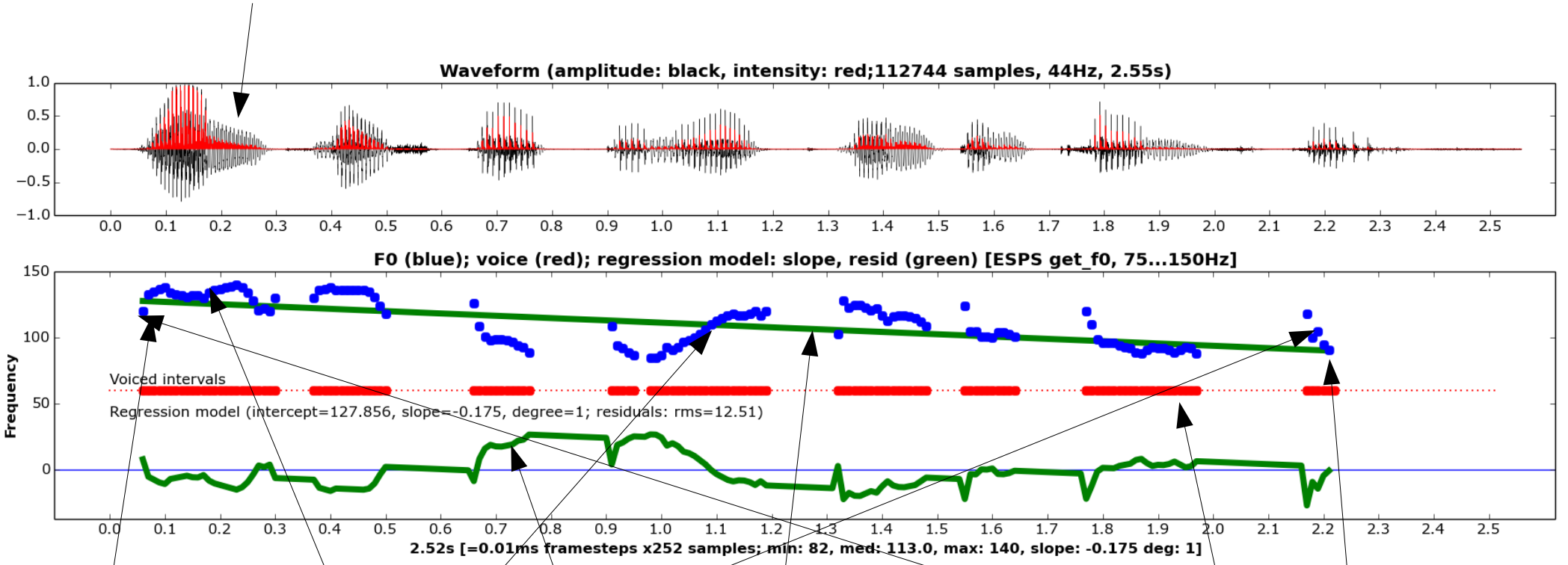
pitch deviations → local pitch accents (or tones)

initial & final pitch
→ boundary tone

Endlich gab der Nordwind den Kampf auf.

Phonetic correlates of phonological categories

amplitude; *intensity* = *f*(*amplitude*^{2}) → factor in stress, focus, contrast, emphasis



pitch track
f0

global downtrend model
→ declination model

voicing

pitch deviations → local pitch accents (or tones)

boundary tone

pitch perturbation model → accent/tone and microprosody model

Endlich gab der Nordwind den Kampf auf.

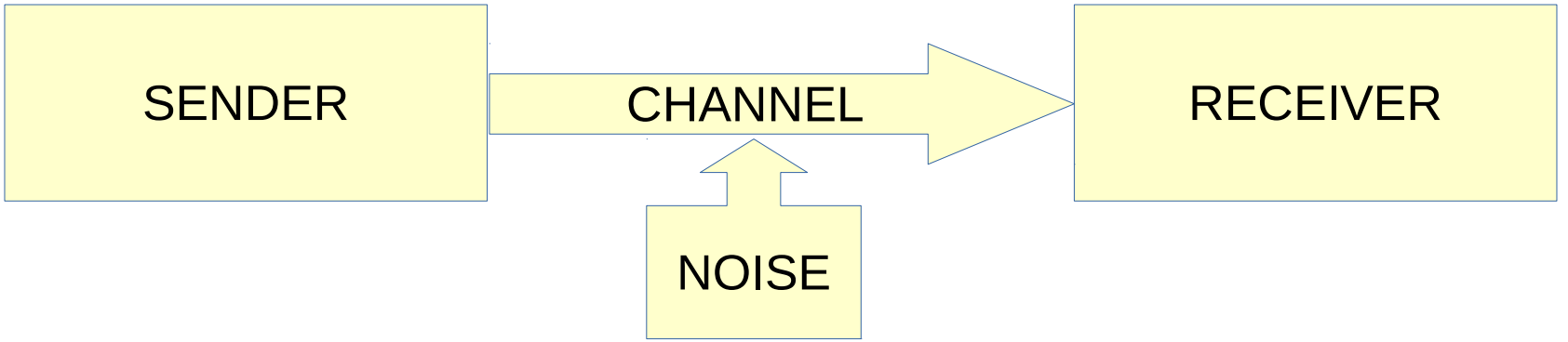
Phonetic interpretation: parameters and trajectories

- Phonetic domain parameters and trajectories
  - *melody*:
    - variation of fundamental frequency properties in time
  - *volume*:
    - variation of intensity properties in time
  - *duration*:
    - variation of unit duration properties in time
      - Note that *duration* has two temporal dimensions

- Phonological domains
  - structural and functional units and patterns

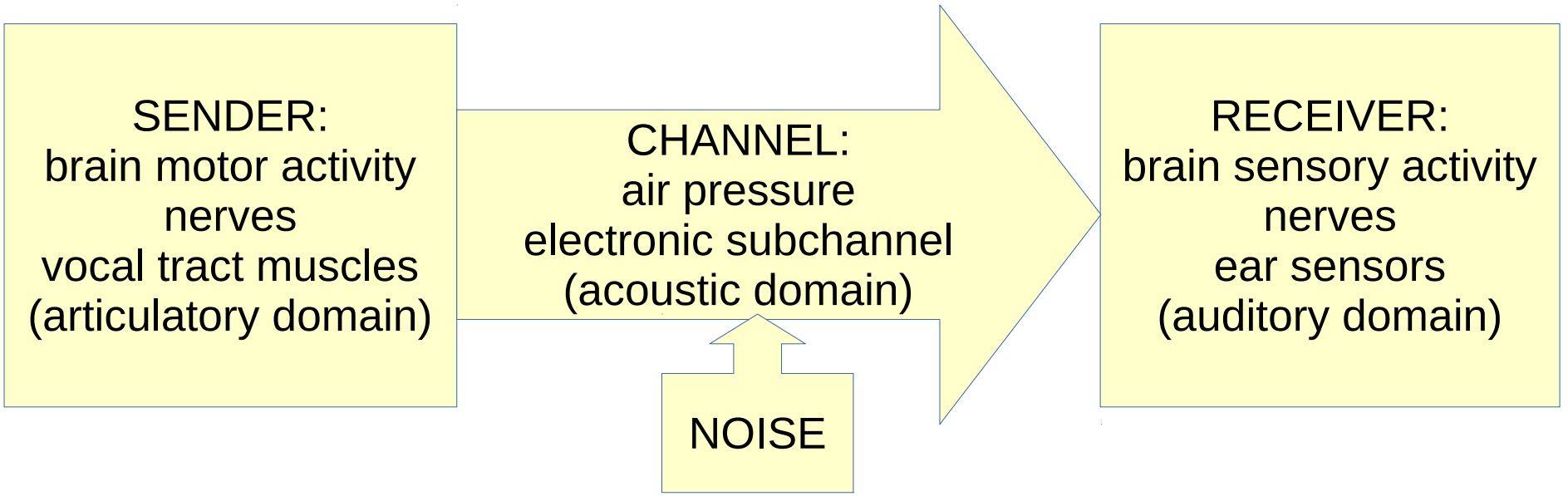
Phonetic interpretation: phases

- speaker, production, articulatory phonetics:
  - articulation rate - effort
- channel, acoustic phonetics:
  - fundamental frequency - intensity
- hearer, reception, auditory phonetics:
  - pitch - loudness



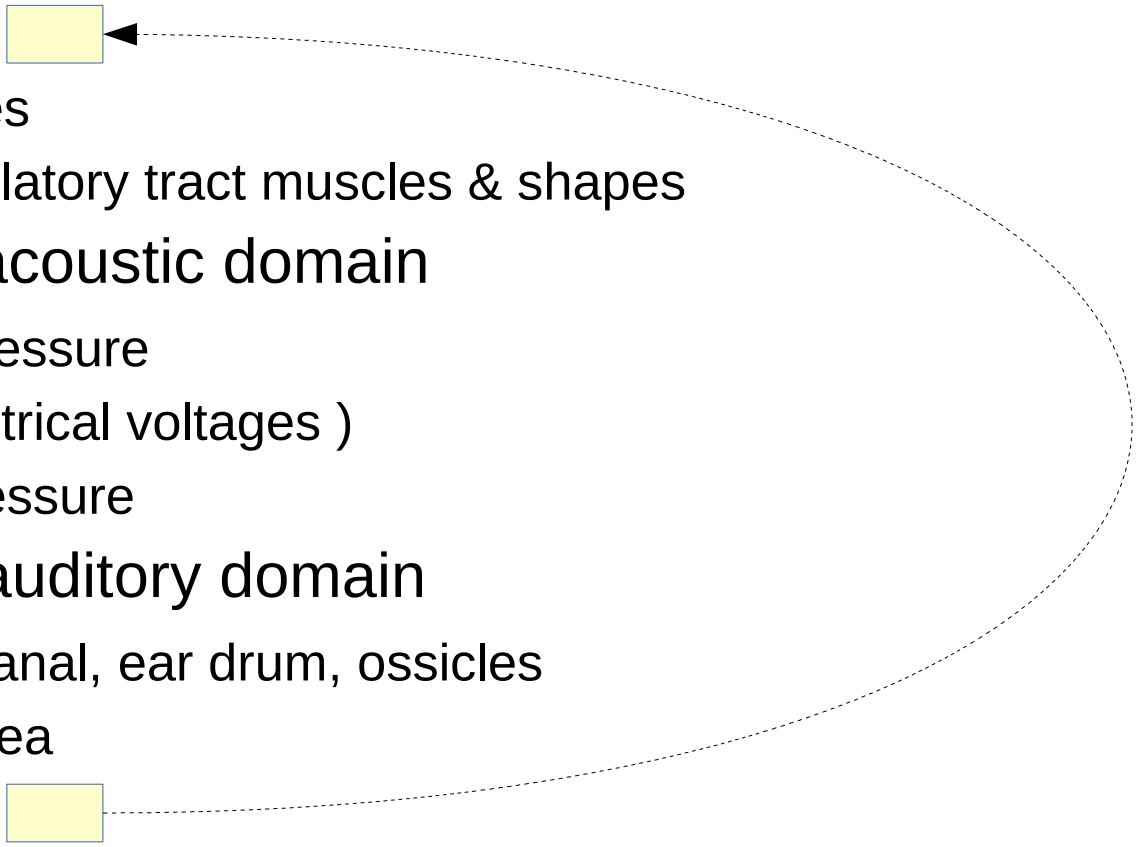
Forms of prosody: phases and subphases

- 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



Phonetic interpretation: phases

- phonetic domains: information theoretic phases:
  - sender → channel → receiver
    - sender: articulatory domain
      - → brain
      - → nerves
      - → articulatory tract muscles & shapes
    - channel: acoustic domain
      - → air pressure
      - (→ electrical voltages)
      - → air pressure
    - receiver: auditory domain
      - → ear canal, ear drum, ossicles
      - → cochlea
      - → brain



Phonetic interpretation: methods

observations – measurements – models

Observations and measurements

- analysis methods:
  - observational methods
    - perceptual
      - quantitative measurements
      - interpretative judgments
    - instrumental
      - quantitative measurements + interpretative judgments
  - experimental methods
    - production
      - quantitative analysis of elicited corpus
      - quantitative analysis of authentic corpus
    - perception
      - quantitative analysis of same-different judgments, reaction times, ...

Observations and measurements

- analysis methods:
  - observational
    - perceptual
      - quantitative measurements
      - interpretative judgments
    - instrumental
      - quantitative measurements + interpretative judgments
  - experimental
    - production
      - quantitative analysis of elicited corpus
      - quantitative analysis of authentic corpus
    - perception
      - quantitative analysis of same-different judgments, reaction times, ...

From phonetic measurements to phonetic models

- First steps: collect data, extract F0
- Induce a prosodic model
- Evaluate prosodic model:
  - Method 1, machine learning:
    - use new data, predict goodness of fit of new data
  - Method 2, perception:
    - re-synthesise prosodic model
    - test results with perception experiments
      - same-different comparison
      - naturalness judgments
      - comprehensibility judgments

First steps: from waveform to F0

From waveform to F0

- Time domain methods
  - frequency = 1 / period
  - peak picking: intervals between peaks
  - intervals between zero crossing measurement
  - autocorrelation
- Frequency domain methods
  - overtone differences
  - spectral comb
  - cepstrum

Check Moebius slides

From phonetic measurements to phonetic models

From phonetic measurements to phonetic models

- Phonetic models:
  - smoothing models:
    - median smoothing
    - global regression
    - local regression (IPO)
  - segment models
    - voiced signal segments
    - quadratic interpolation between reference points
  - structured models
    - Fujisaki model
    - Liberman and Pierrehumbert model
    - Hirst model

From phonetic measurements to phonetic models

- Phonetic models:
  - smoothing models:
    - median smoothing
    - global regression
    - local regression (IPO)
  - segment models
    - voiced signal segments
    - quadratic interpolation between reference points
  - structured models
    - Fujisaki model
    - Liberman and Pierrehumbert model
    - Hirst model

From phonetic measurements to phonetic models

- Phonetic stylisation models:
  - smoothing models:
    - median; global (Huber) and local (IPO) regression
  - segment models
    - voiced segment smoothing
    - quadratic spline segment interpolation (Hirst)
- F0 stylisation is the simplification of the F0 trajectory to remove
  - irrelevant properties
  - noise

First steps to stylisation: smoothing filters

Regression smoothing

F0 smoothing: general procedures

1. Identify voiced and unvoiced intervals, extract F0
2. Smoothing and 'stylisation' modelling procedures

Local sequencing procedures:

- level F0 sequences, e.g. based on median of a sequence: robotic!
- median smoothing:

  for each F0 (t_i) measurement :

  $F0_{smooth}$ (t_i) mean <F0 (t_i), …, F0 (t_n)>
- quadratic spline sequences
  - (Hirst)

Global reference plus accent/tone excursion procedures

- Regression: log, linear, quadratic, … ,
  - (Fujisaki, Pierrehumbert & Lieberman, Tilt model)

F0 smoothing: different approaches

- Smoothing by median filter:
  - the median of sequences of 3 measurements
- Smoothing by linear regression

$y = a_0 + a_1 x + \varepsilon$

- Smoothing by polynomial regression:

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

- Smoothing by asymptotic descent, effectively log(x):

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

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

Smoothing: different approaches

- Smoothing by median filter:
  - the median of sequences of 3 measurements
- Smoothing by linear regression

  $y = a_0 + a_1 x + \varepsilon$
- Smoothing by polynomial regression:

  $y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 x^3 + \ldots + a_0 x^n + \varepsilon$
- Smoothing by asymptotic descent:

  $y \in \langle x_1, \ldots, x_1 \rangle : x_i = m \cdot x_{i-1} + \varepsilon$

  $y \in \langle x_1, \ldots, x_1 \rangle : a + x_i = a + m \cdot x_{i-1} + \varepsilon$

Regression smoothing examples

1. Identify voiced intervals
2. Extract F0
3. Interpolate silent intervals
   - Simplified in the following examples:
     - 3rd quartile (75th percentile)
4. Calculate smoothing (declination / accent model)
   - Linear, quadratic etc. (polynomial) regression over interpolated F0 sequence
5. Calculate residuals (microprosody model):
   - Subtract regression values from F0 values

F0 smoothing: general procedures

1. Identify voiced and unvoiced intervals, extract F0
2. Smoothing and 'stylisation' modelling procedures

Local sequencing procedures:

- level F0 sequences, e.g. based on median of a sequence: robotic!
- median smoothing:

  for each F0 (t_i) measurement :

  $F0_{smooth}$ (t_i) mean <F0 (t_i), …, F0 (t_n)>
- quadratic spline sequences
  - (Hirst)

Global reference plus accent/tone excursion procedures

- Regression: log, linear, quadratic, … ,
  - (Fujisaki, Pierrehumbert & Lieberman, Tilt model)

F0 smoothing: global procedures

F0 smoothing: procedure

1. Identify voiced and unvoiced intervals, extract F0
2. Smoothing and ‘stylisation’ modelling:
   - Local smoothing:
     - linear, median; quadratic spline (Hirst)
   - Global reference plus discrete deviant values for different accents or tones
     - Fujisaki model, Liberman & Pierrehumbert’s invariance model, Taylor’s ‘Tilt’ model
   - Global smoothing with regression:
     - log, linear, quadratic, polynomial of degree *n*

       (used for illustration in the following examples)

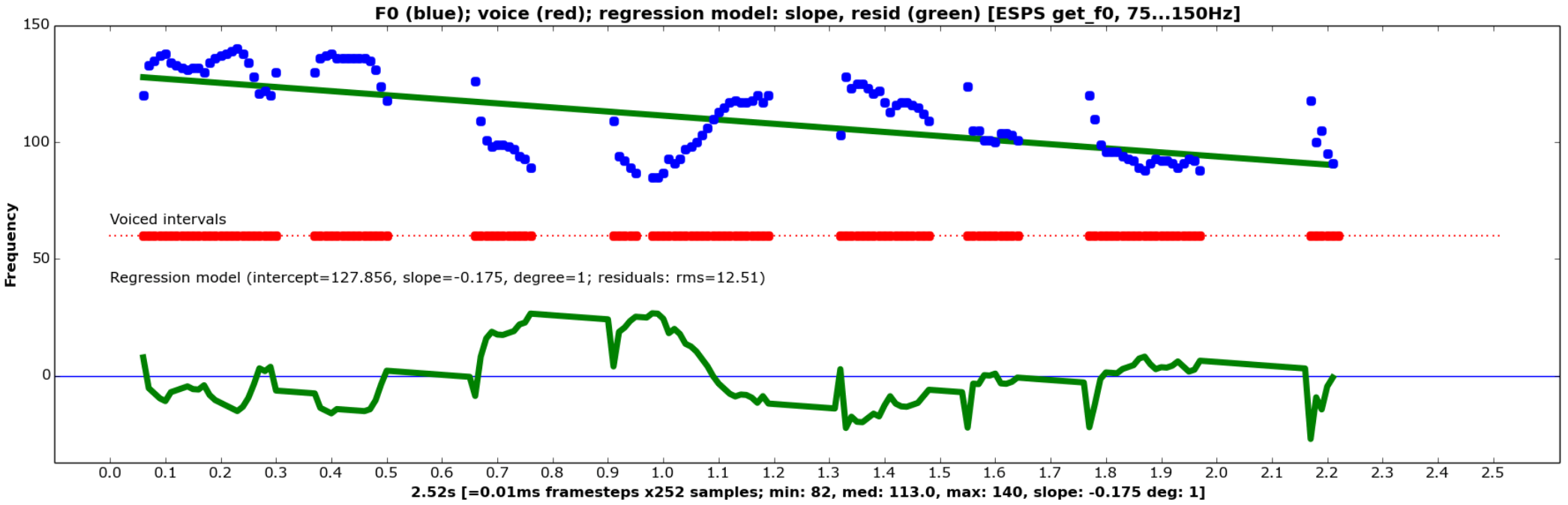
Smoothing: different approaches, different goals

- Smoothing by polynomial regression (degree *n)*:

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

- Smoothing by linear regression (degree 1)

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



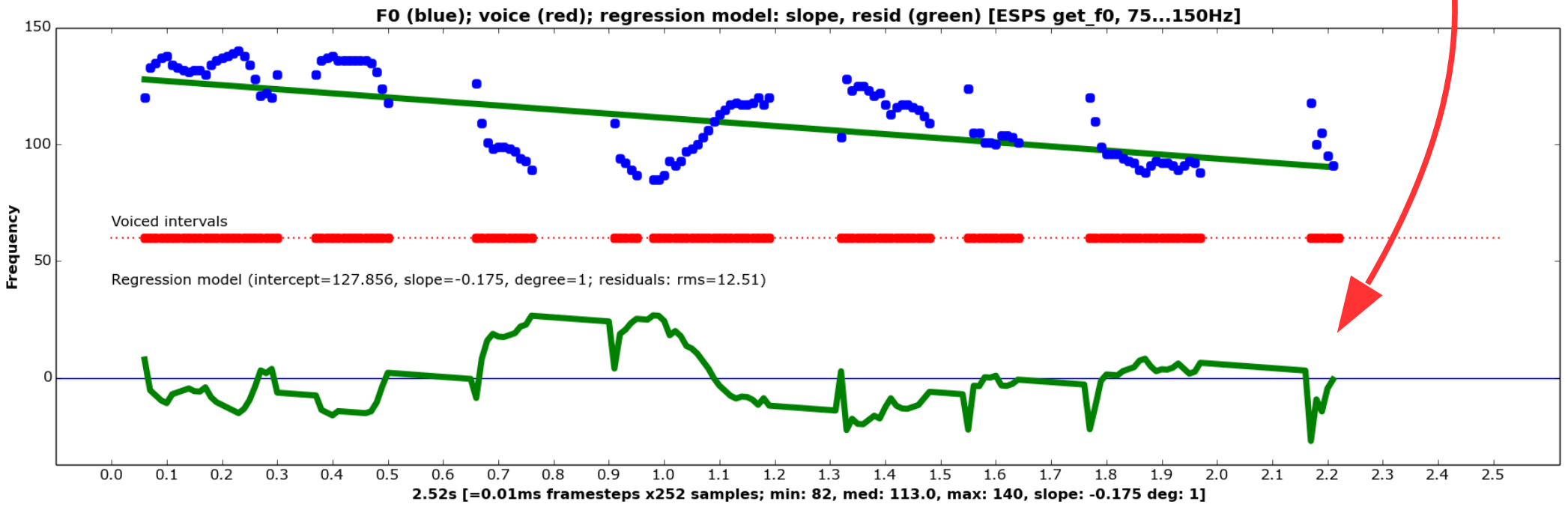
Smoothing: different approaches, different goals

- Smoothing by polynomial regression (degree *n)*:

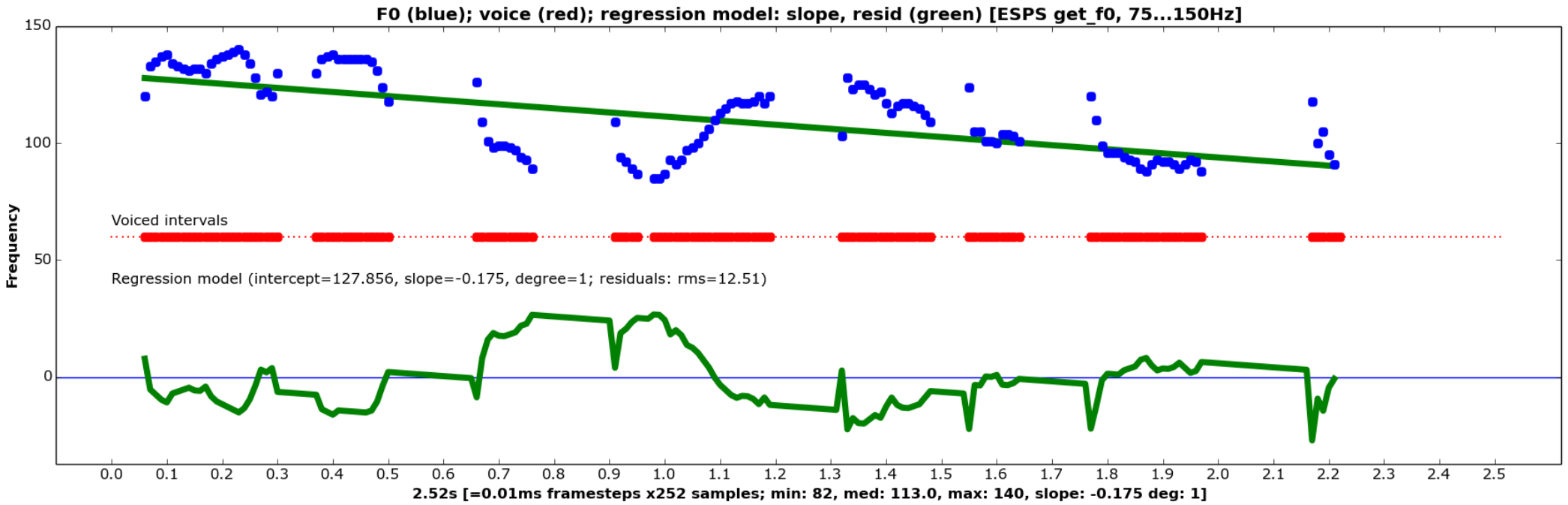
$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 x^3 + \ldots + a_0 x^n + \varepsilon$$

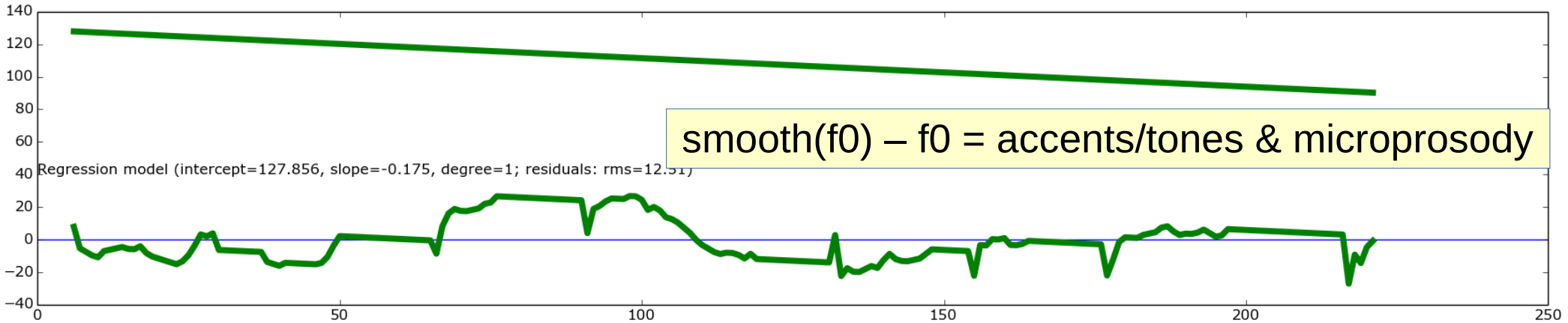
- Smoothing by linear regression (degree 1)

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



Global linear regression contour

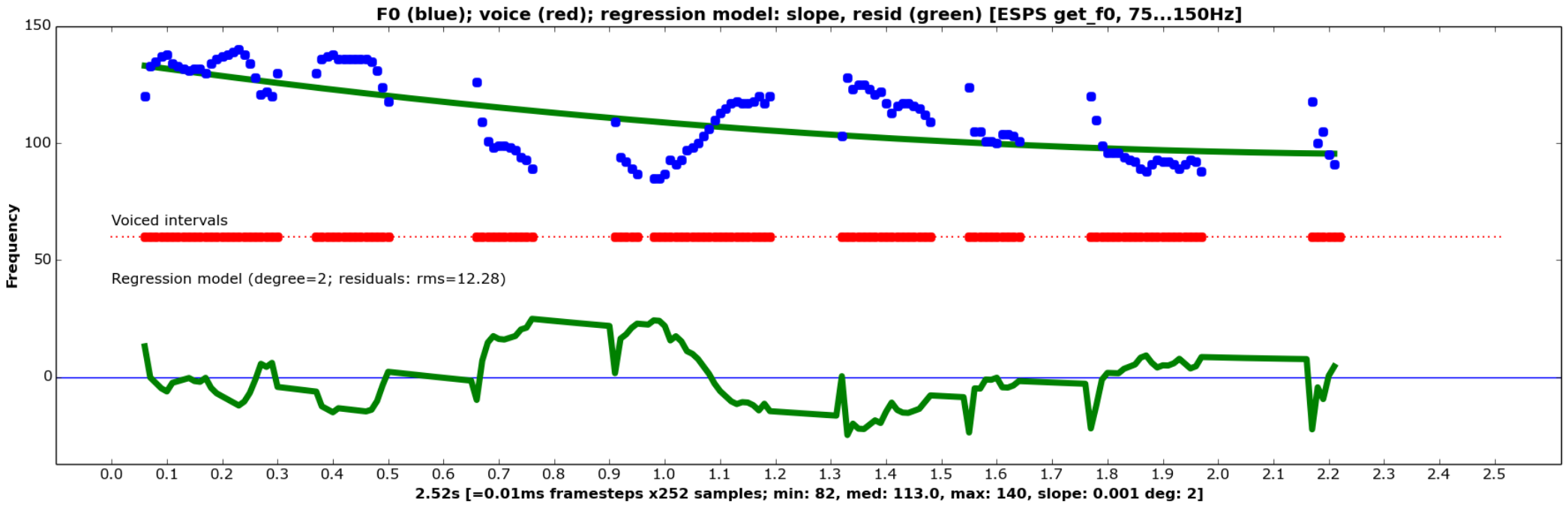


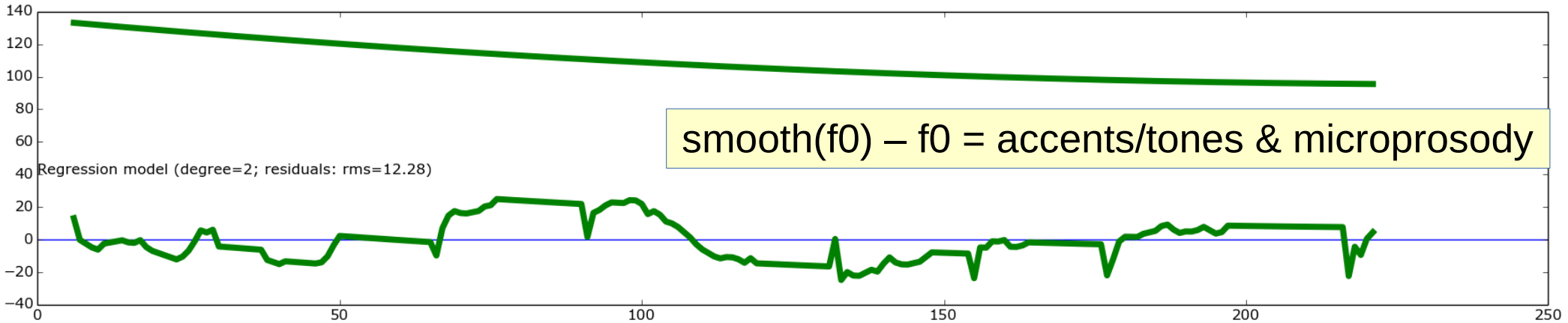


smooth(f0) – f0 = accents/tones & microprosody

Endlich gab der Nordwind den Kampf auf.

Global quadratic regression contour

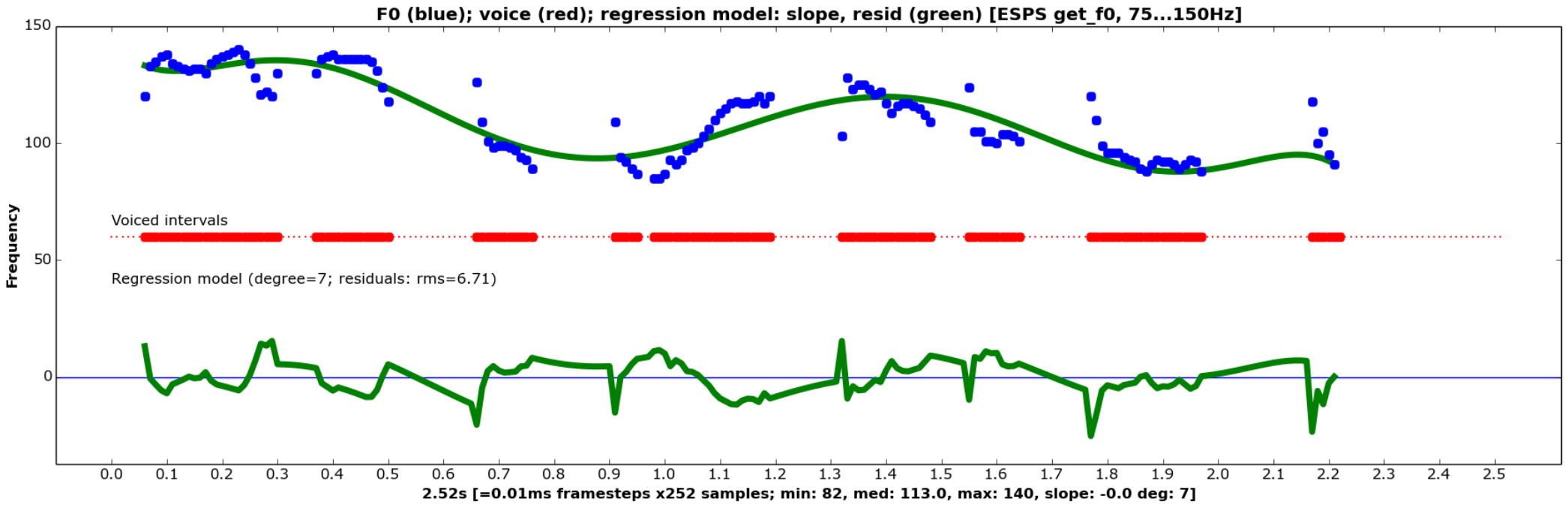


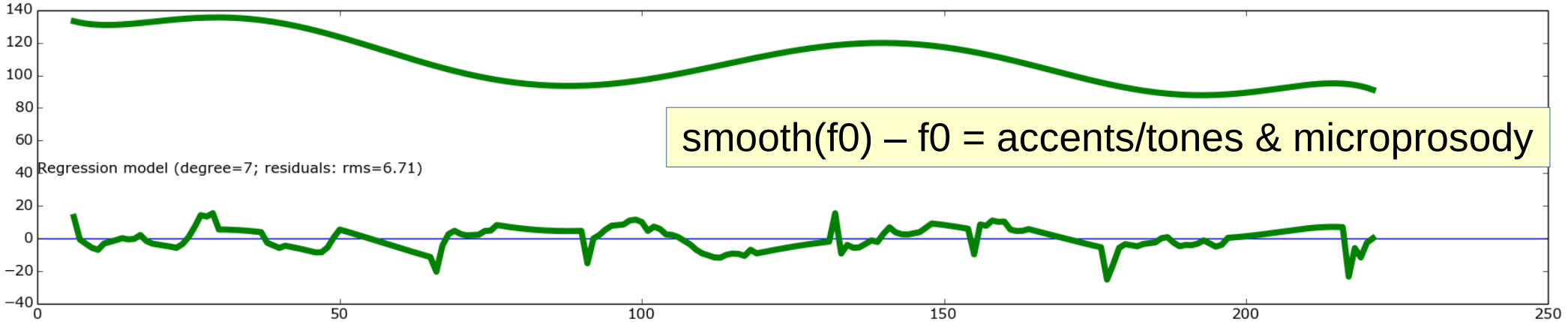


smooth(f0) – f0 = accents/tones & microprosody

Endlich gab der Nordwind den Kampf auf.

Global regression contour, degree 7

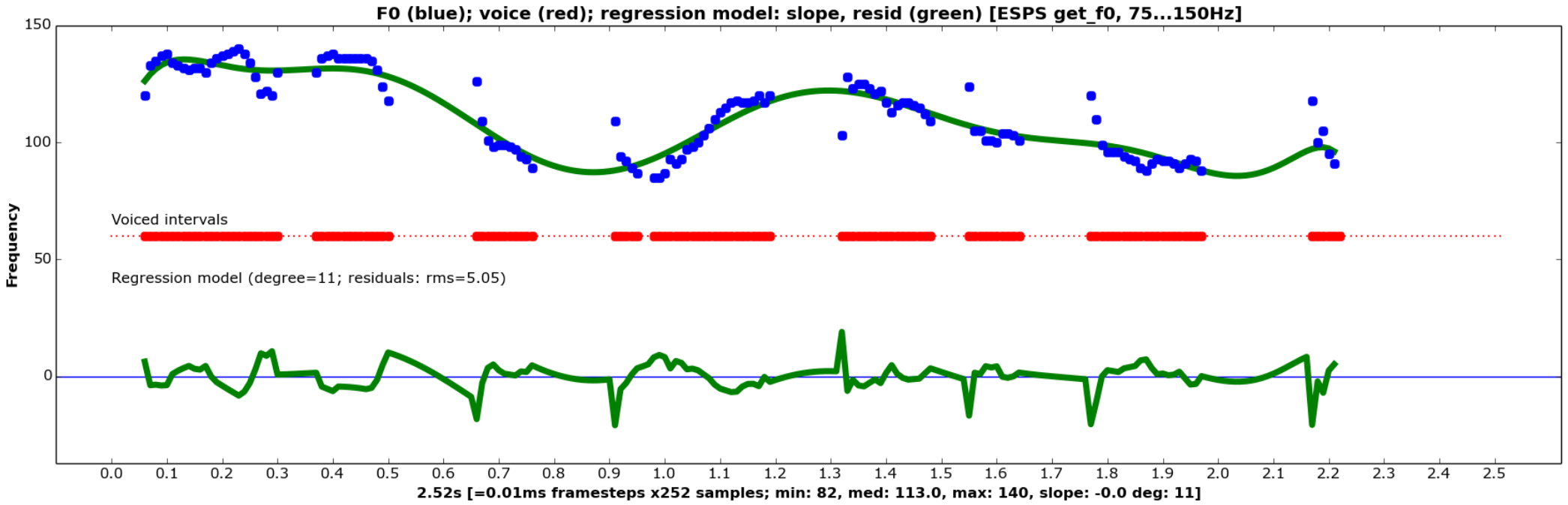


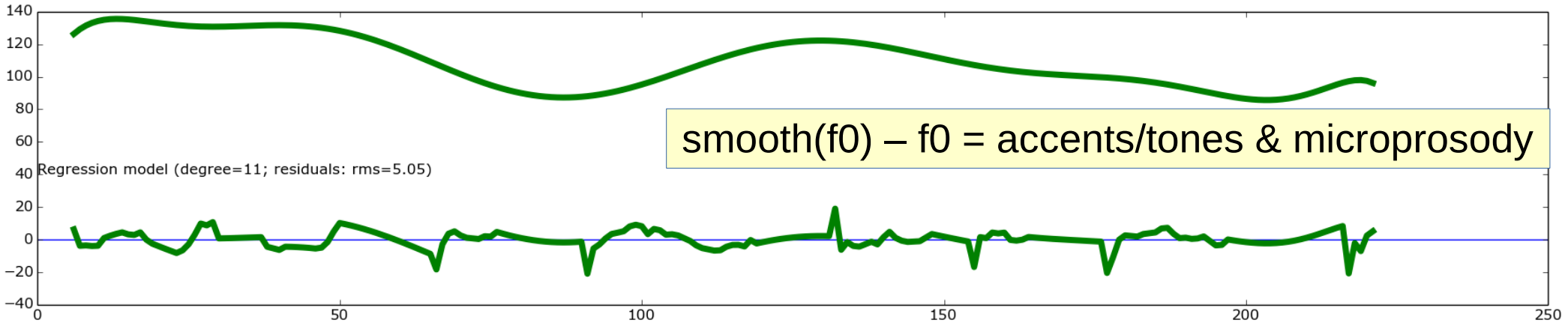


smooth(f0) – f0 = accents/tones & microprosody

Endlich gab der Nordwind den Kampf auf.

Global regression contour, degree 11

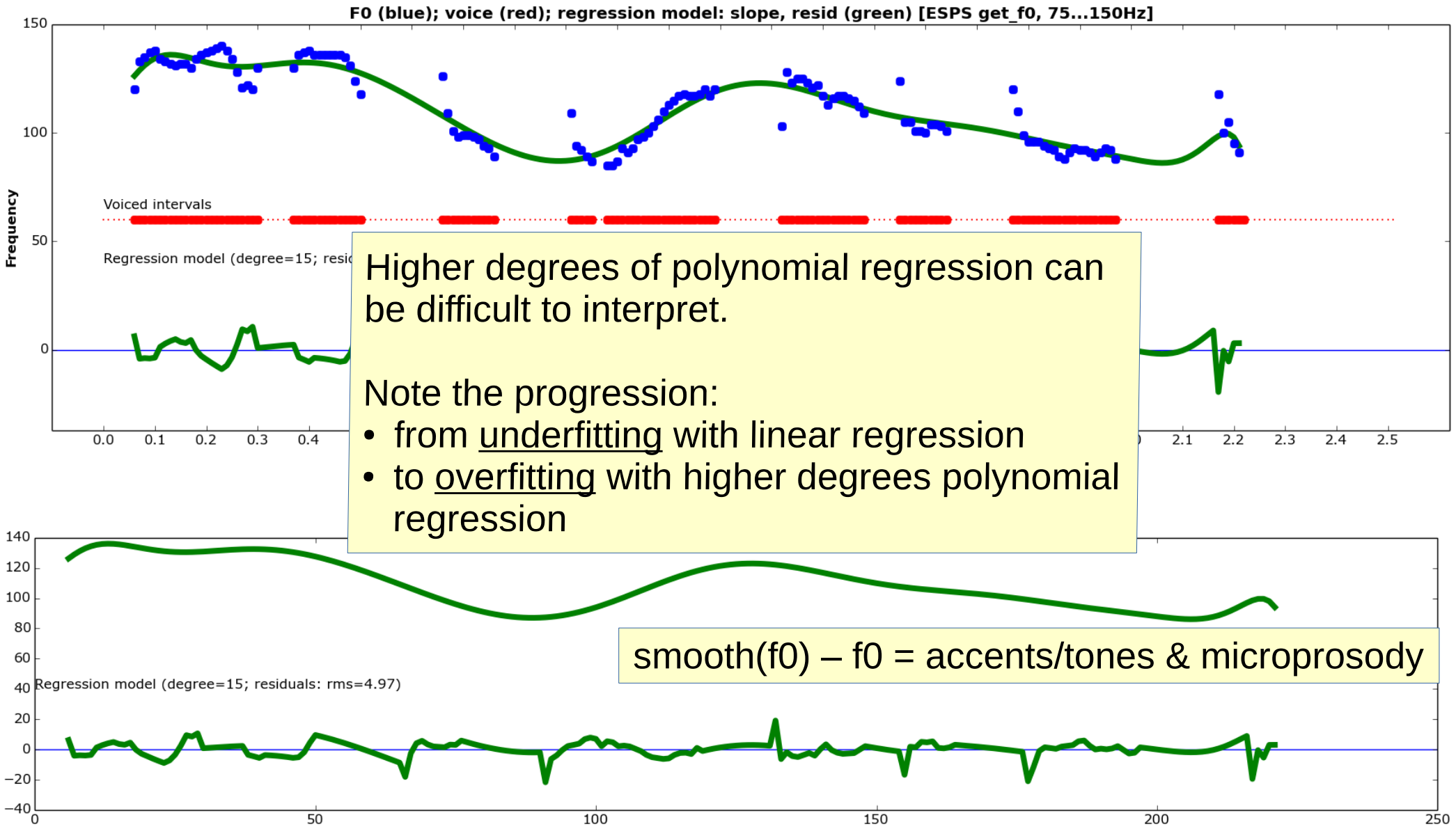




smooth(f0) – f0 = accents/tones & microprosody

Endlich gab der Nordwind den Kampf auf.

Global regression contour, degree 15



Higher degrees of polynomial regression can be difficult to interpret.

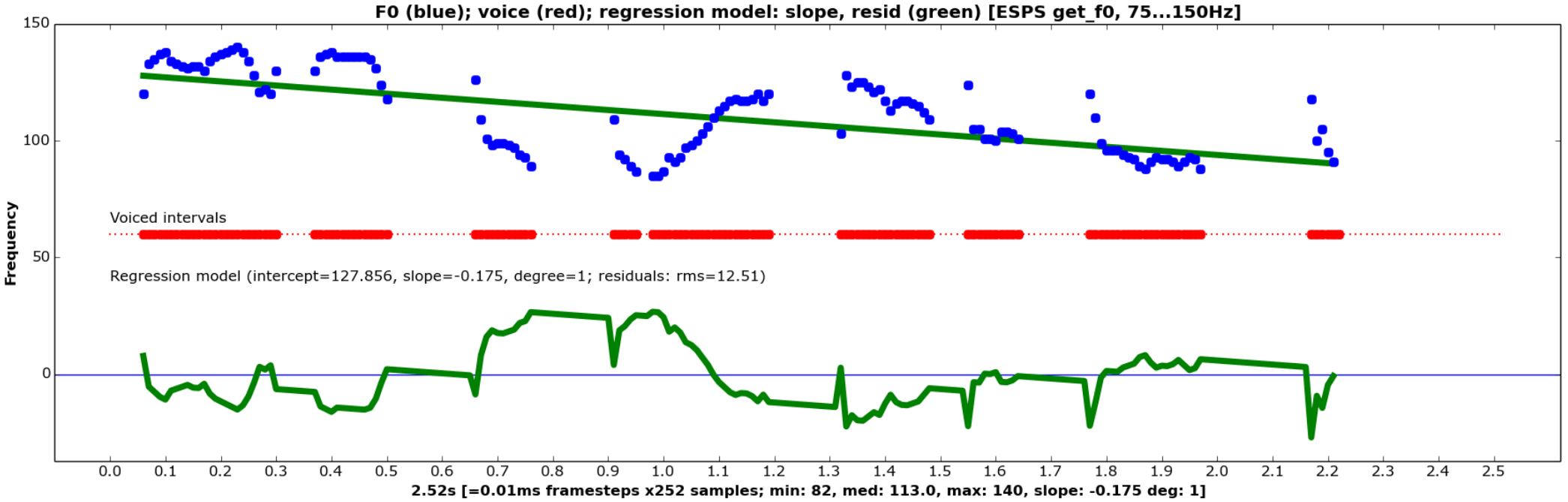
Note the progression:

- from underfitting with linear regression
- to overfitting with higher degrees polynomial regression

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

Endlich gab der Nordwind den Kampf auf.

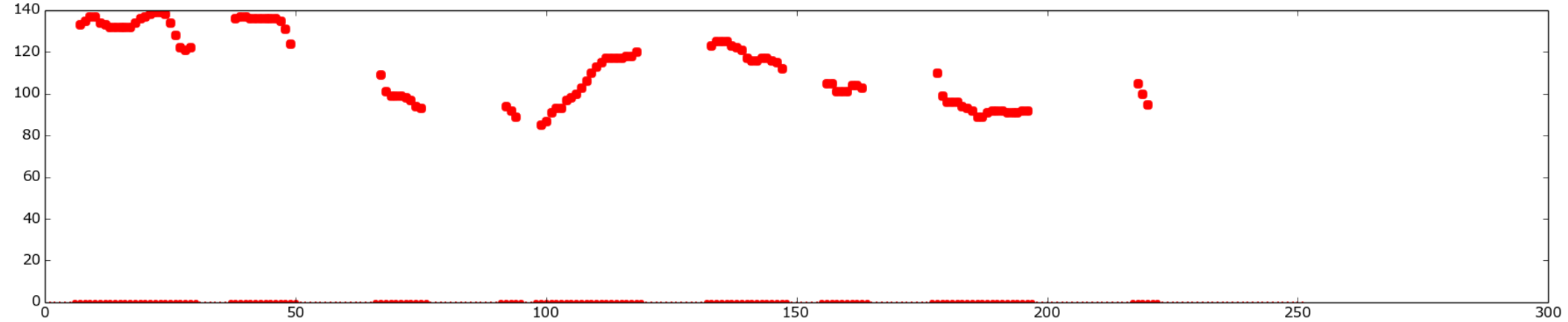
Global regression contours, up to degree 20



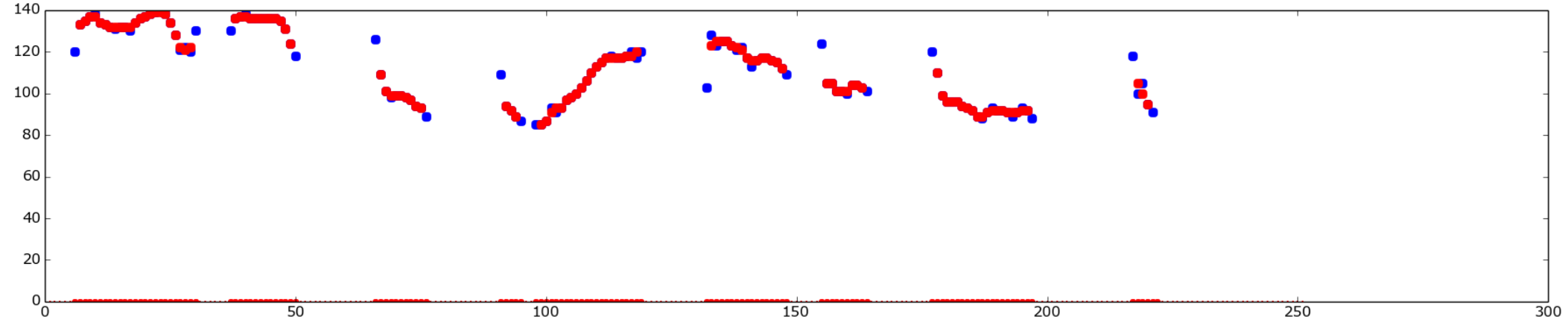
Endlich gab der Nordwind den Kampf auf.

F0 smoothing: local procedures

Simple median filter (scope: 3), often used

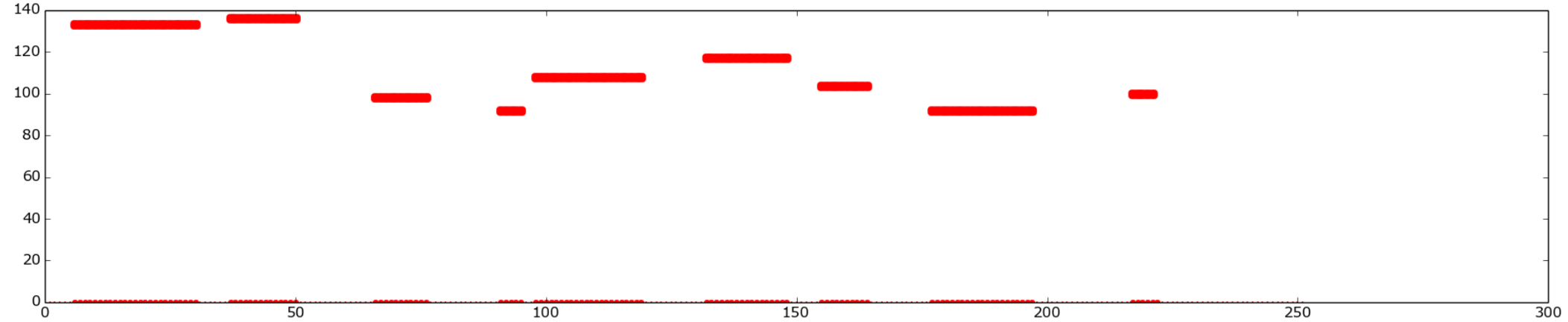


Endlich gab der Nordwind den Kampf auf.

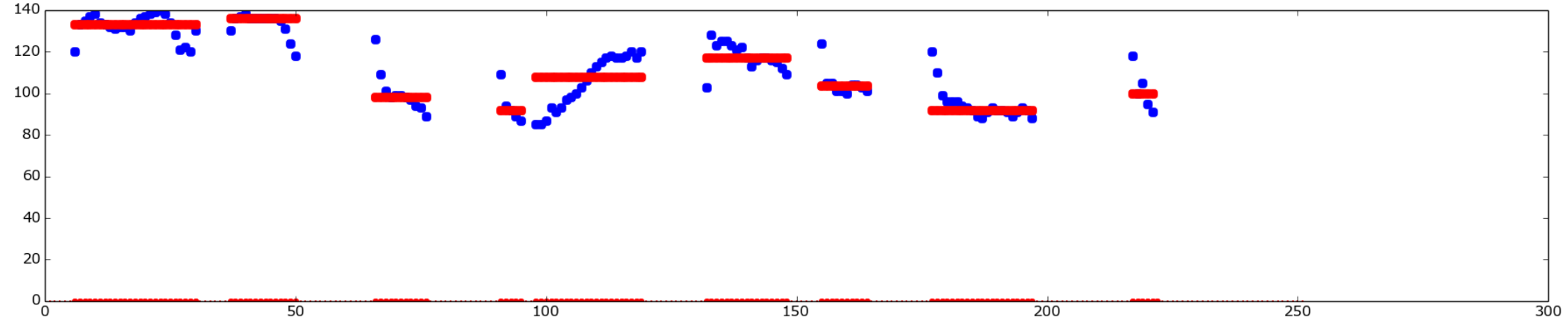


Each F0 value is normalised to the median F0 value of its immediate neighbours

Simple local median levelling filter – robotic!

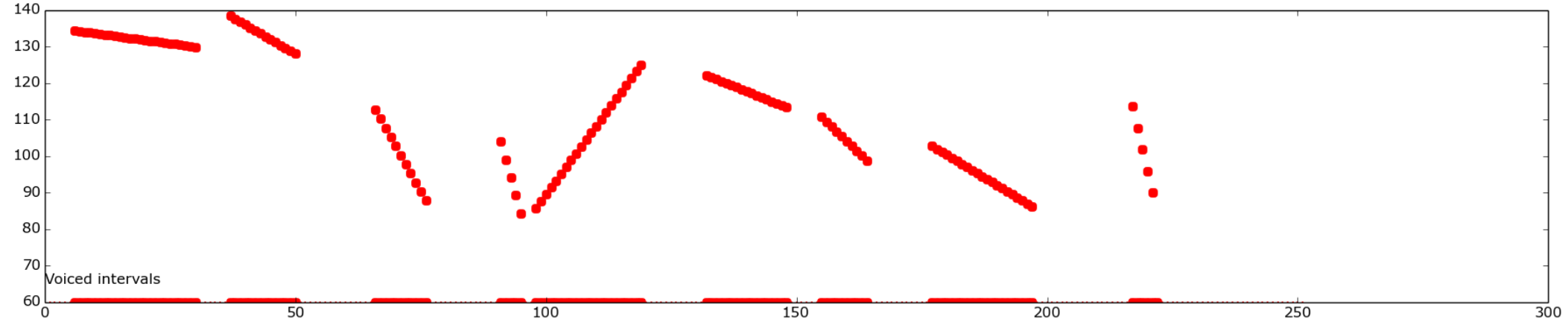


Endlich gab der Nordwind den Kampf auf.

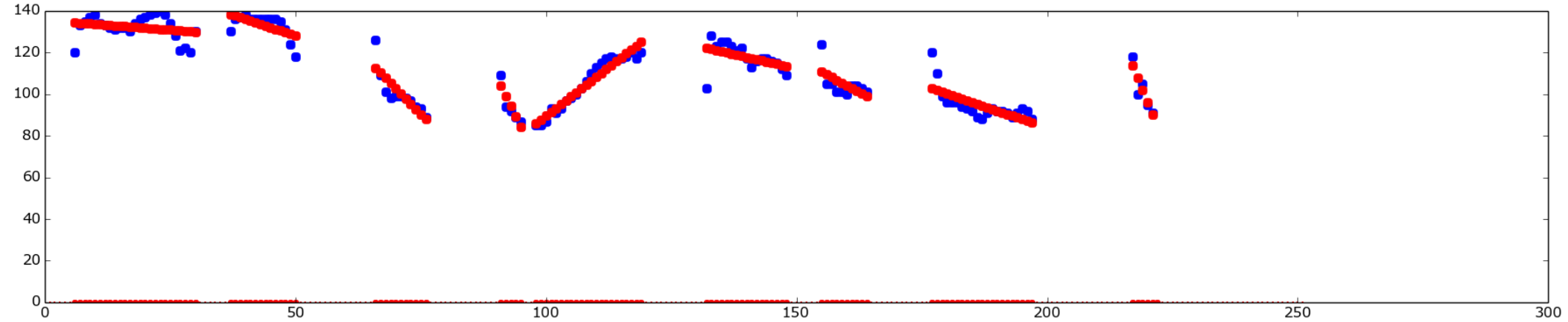


Each F0 value in a sequence is normalised to the median F0 value for the sequence

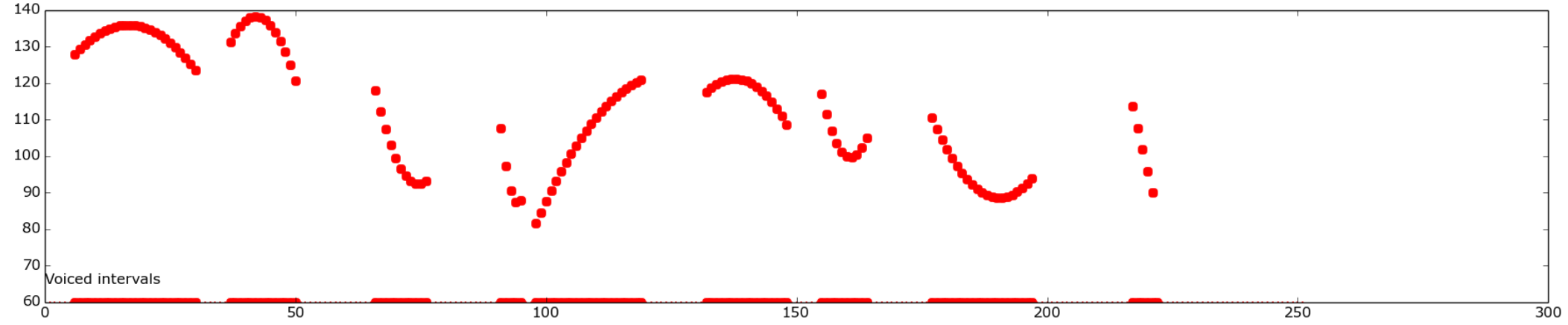
Local voicing regression contours, degree 1



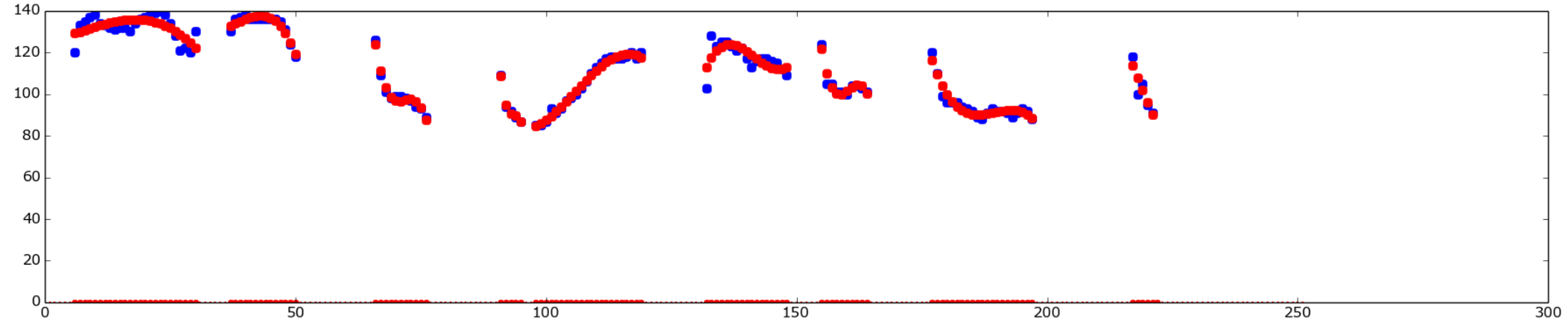
Endlich gab der Nordwind den Kampf auf.



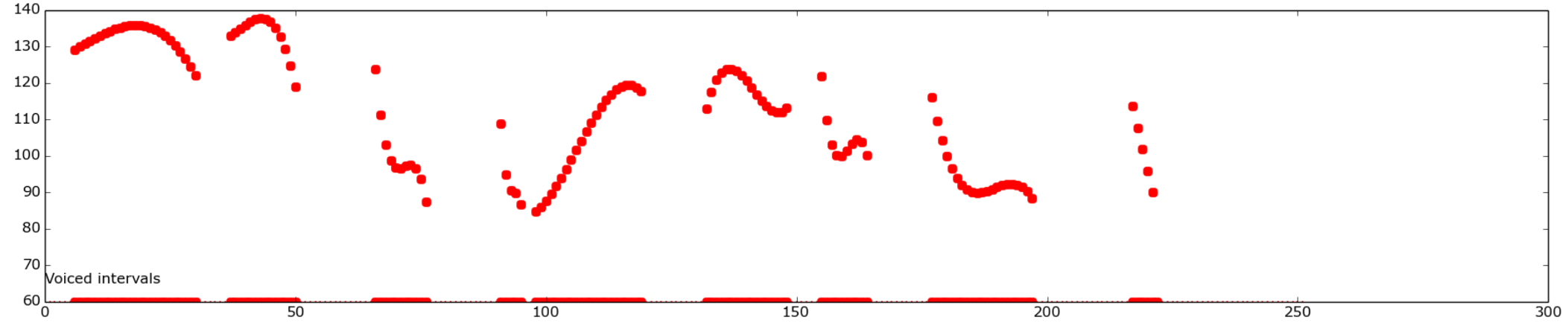
Local voicing regression contours, degree 2



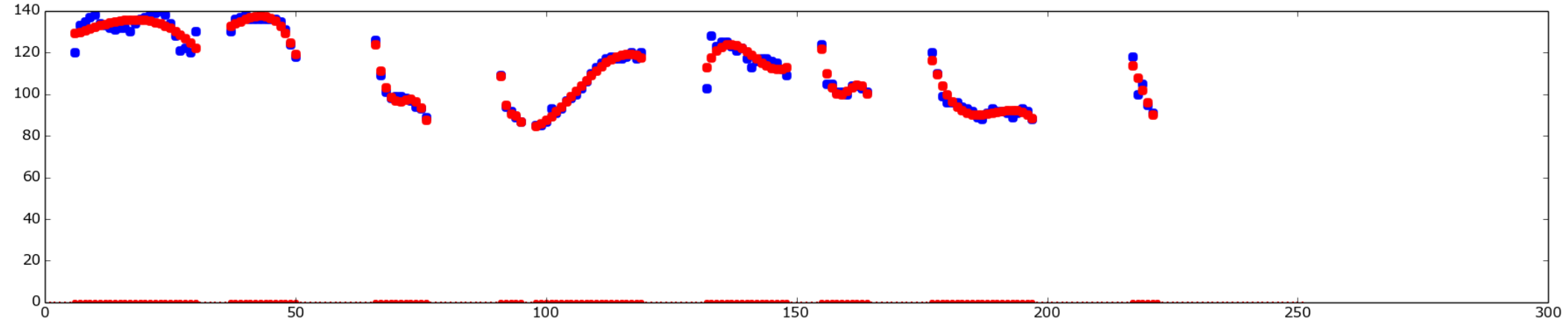
Endlich gab der Nordwind den Kampf auf.



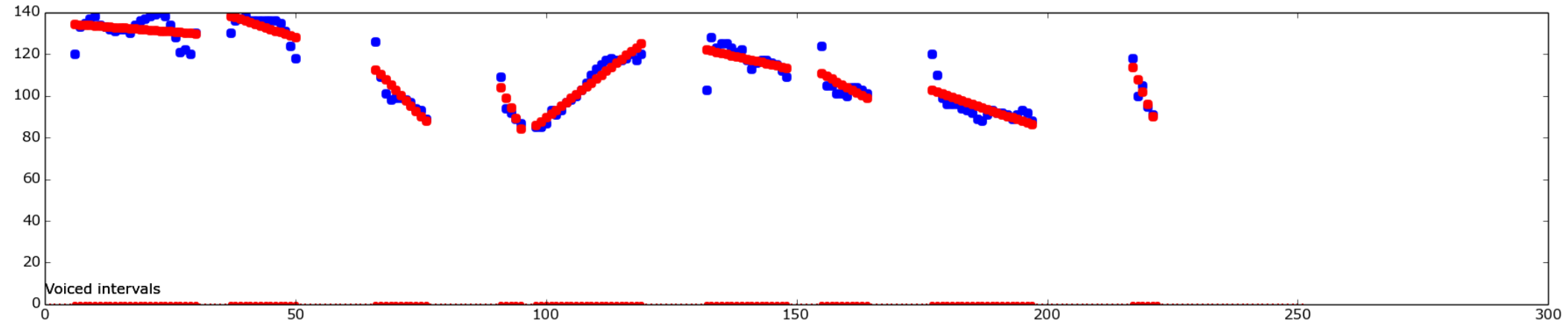
Local voicing regression contours, degree 3



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 underfitting with linear regression
- to overfitting with higher degrees polynomial regression

Models of f0 patterning: Hirst

- Intsint
- Momel
- ProZed

Hirst: quadratic spline - 'piecewise quadratic function'

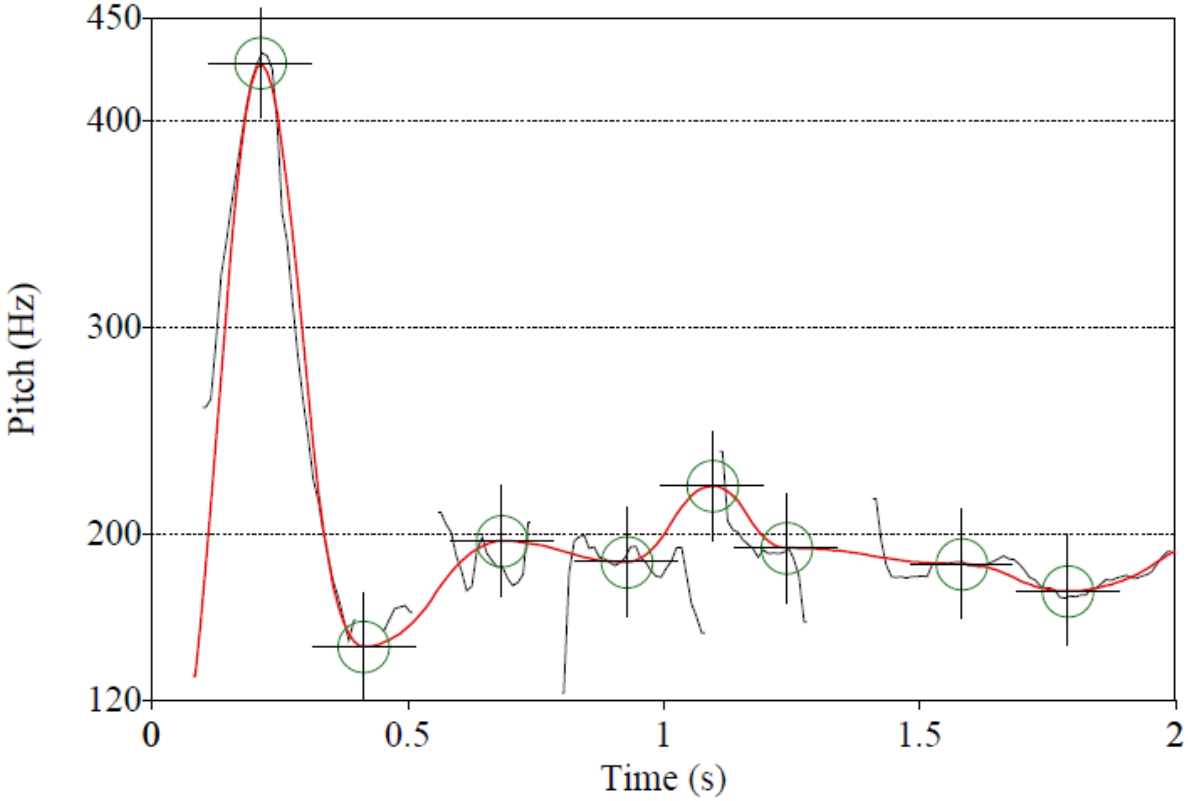


Fig. 6.7 Macromelodic profile (red) for a two-second extract from recording A01, defined as quadratic transitions between anchor points (green).

Hirst: quadratic spline - 'piecewise quadratic function'

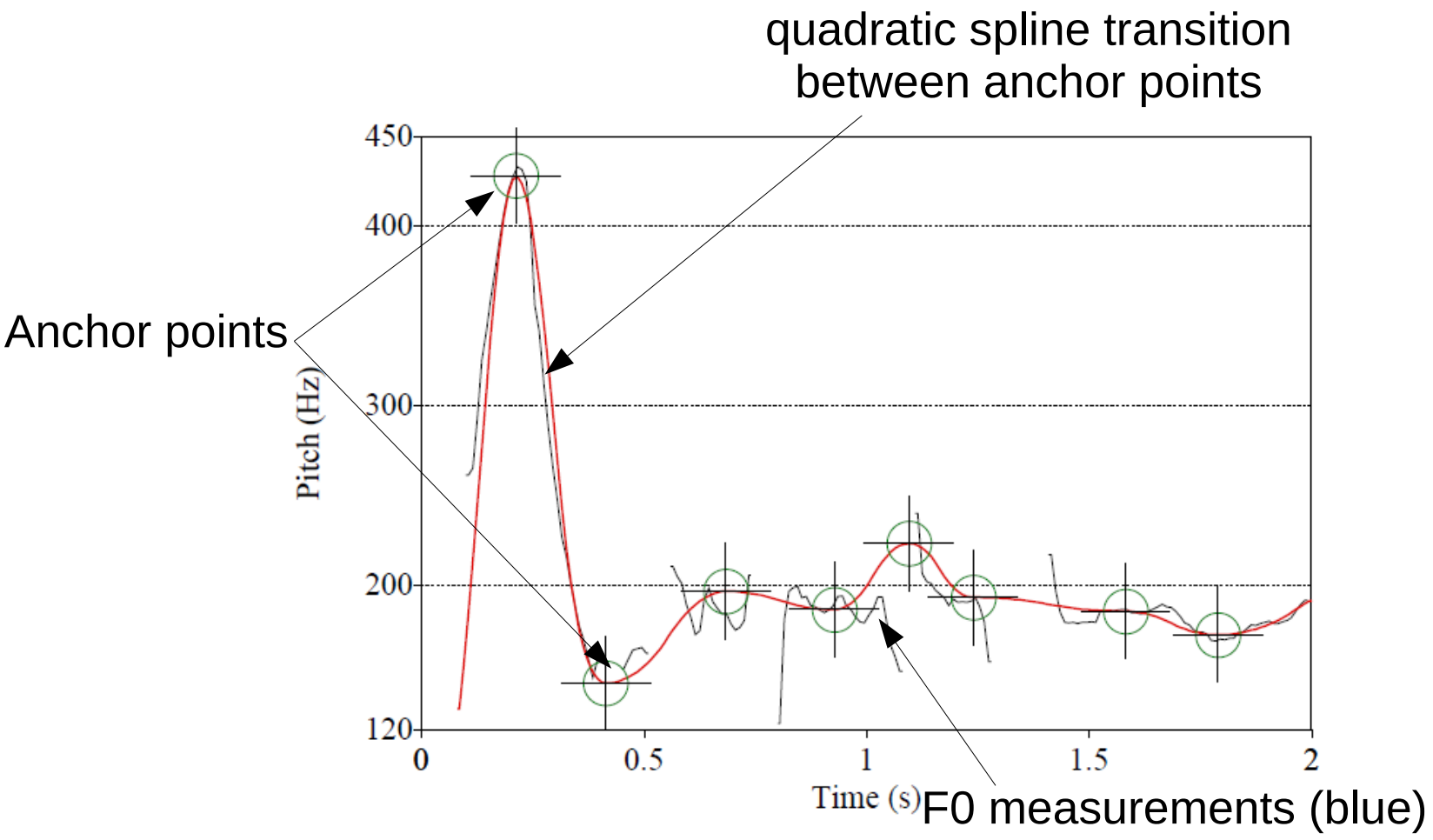
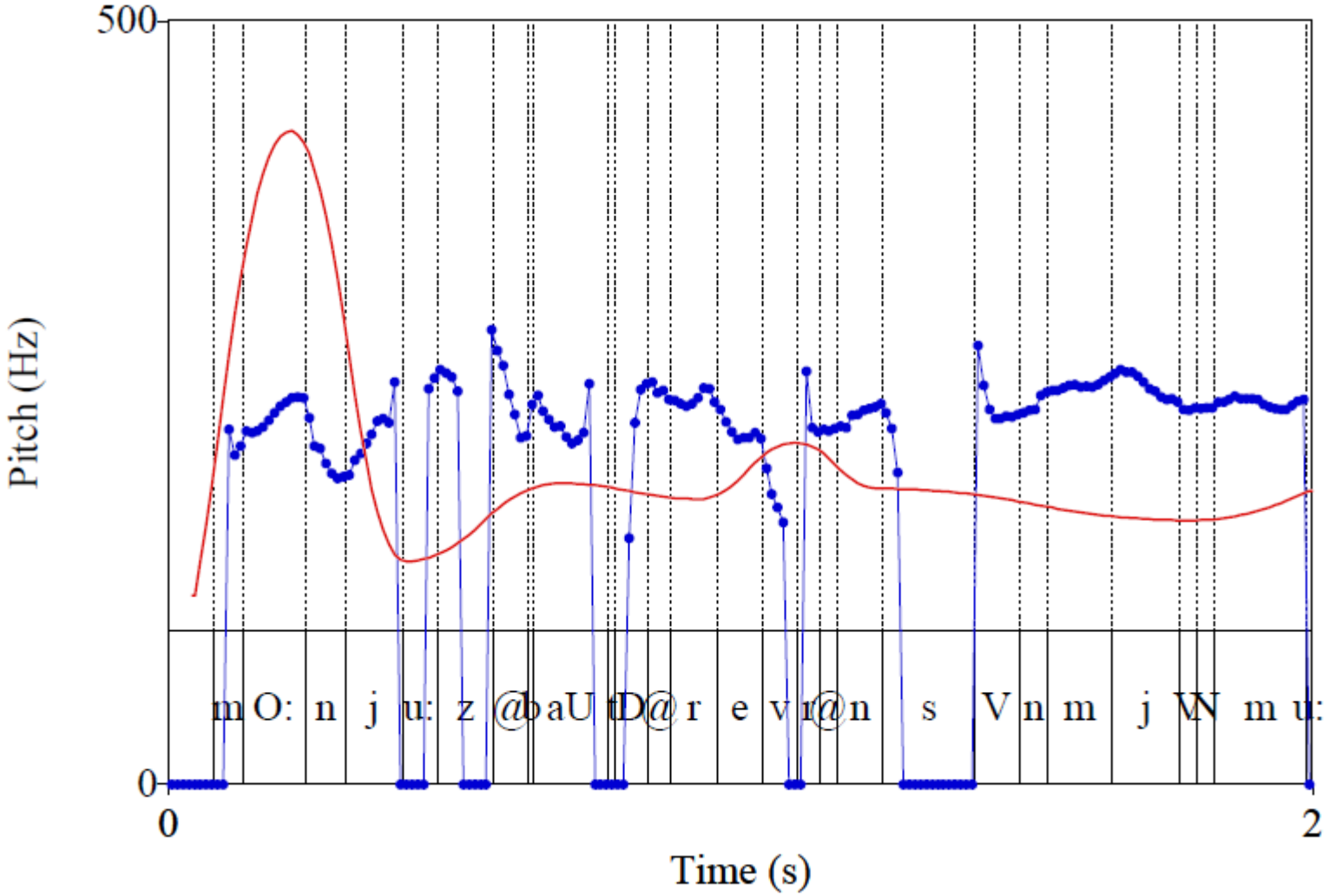


Fig. 6.7 Macromelodic profile (red) for a two-second extract from recording A01, defined as quadratic transitions between anchor points (green).

Hirst: micromelody = F0 / quadratic spline function

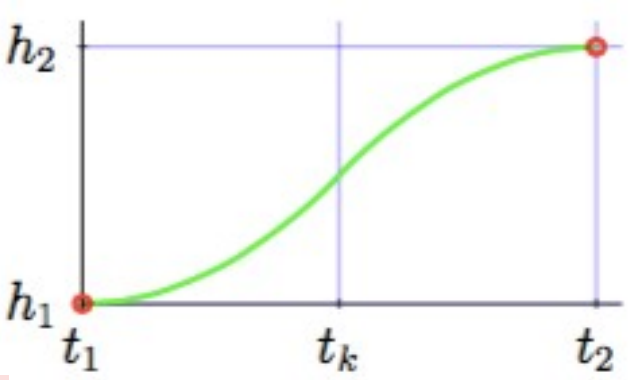


Macromelody (red), micromelody (blue): micromelody = F0 / spline model

Smoothing by local spline interpolation (Hirst)

Momel:
Quadratic splines:

- changing an anchor point only affects neighbouring transitions
- anchor points correspond to zeros on the first derivative of the spline
- the transition between two anchor points:
  - symmetrical
  - maximum slope at the spline "knot" - half way between two anchor points.



Hirst's f0 formulas:

$$t_i \in [t_1 \ldots t_k] : h_i = h_1 + \frac{(h_2 - h_1) \cdot (t_i - t_1)^2}{(t_k - t_1)(t_2 - t_1)}$$

$$t_i \in [t_k \ldots t_2] : h_i = h_2 + \frac{(h_1 - h_2) \cdot (t_i - t_2)^2}{(t_k - t_2)(t_1 - t_2)}$$

Cubic spline problem, so not used in Momel:

Changing one anchor point can affect the whole curve.

Many other methods ...

- Straight lines (IPO)
- Baseline + pulse modulation
- Gårding
- Grønnum (Thorsen)
- Asymptotic descent (Liberman & Pierrehumbert)
  - Tilt
- Spline sequence interpolation (Hirst)

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

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 $l < 1$

Modified transform for model 1

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

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

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

Model 1

a. General F0 transform

$T(P) = P - r$

P and r in Hz

Subtract the reference line from the F0 trajectory

b. Downstep

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

where P_i is the F0 targ… tion i, down-stepped with respect to the previous accent target T_{i-1}

Define the asymptotic declination line

c. Answer-background relation

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

where P_A is the F0 targ… the B accent

Define the relation between focus and non-focus accent types

d. Relation of r to initial accent target

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

where P_0 is the target i… are constants

Define the relation between first pitch accent and reference line

e. Final Lowering

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

where $l < 1$

for rule (5e) in model 1.

Define final lowering

Modified transform for model 1

…)

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

Model 1A

Substitute

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

for equation (5d) in model 1.

d, e, f, and b

Model 1B

Substitute

$r = f \cdot P_0 + d$

for equation (5d) in model 1.

Models of f0 patterning: Liberman & Pierrehumbert

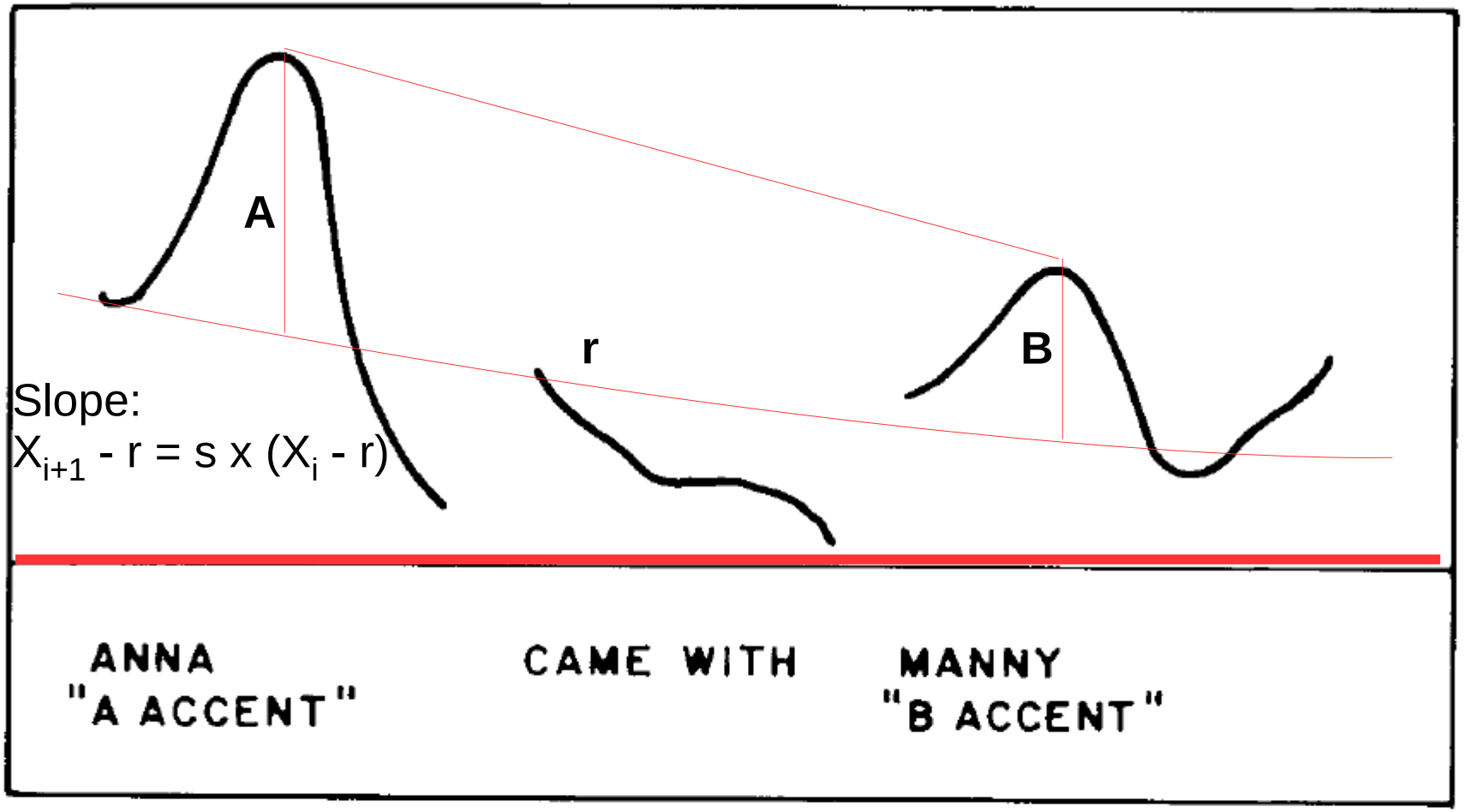


Figure 9
An F0 contour for *Anna came with Manny*, produced as a response to *What about Manny? Who came with him?*

Models of f0 patterning: Liberman & Pierrehumbert

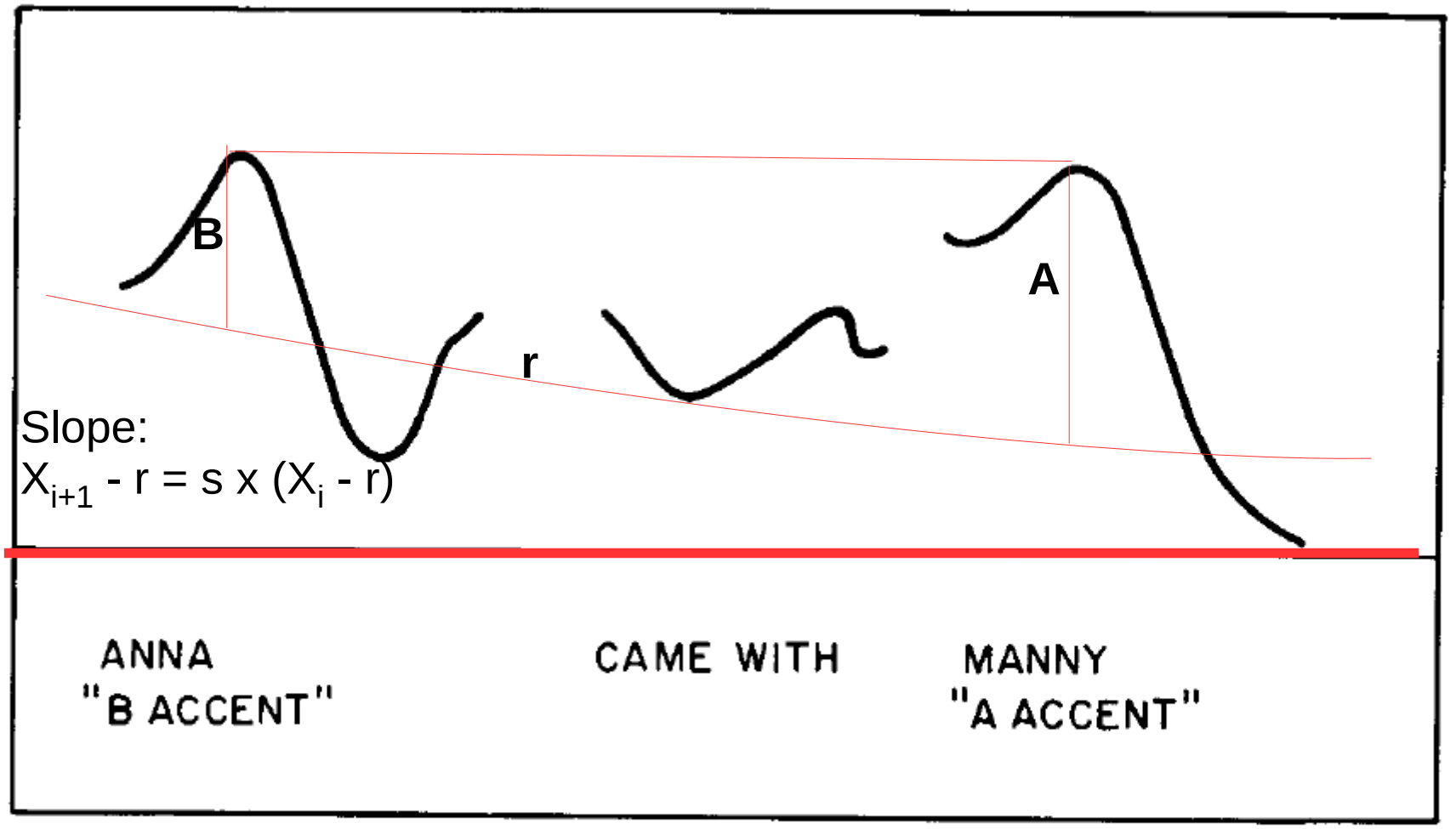


Figure 10
An F0 contour for *Anna came with Manny*, produced as a response to *What about Anna? Who did she come with?*

Models of f0 patterning: Liberman & Pierrehumbert

Model 1

a. General F0 transform

$T(P) = P - r$

P and r in Hz

Subtract the reference line

odified transform for model 1

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

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

b. Downstep

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

Define the asymptotic declination line

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 the B accent

Define the relation between focus and non-focus accent types

lodel 1A

Substitute

d. Relation of r to initial accent

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

where P_0 is the target in Hz

are constants

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

r equation (5d) in model 1.

l, e, f, and *b*

Define the relation between first pitch accent and reference line

Model 1B

Model 1C

e. Final Lowering

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

where $l < 1$

Su

P

Define final lowering

bstitute

$= f \cdot P_0 + d$

for rule (5e) in model 1.

for equation (5d) in model 1.

Models of f0 patterning: Liberman & Pierrehumbert

Model 1A

Substitute

$r = f \cdot (\mathrm{P}_0)^e + d$

for equation (5d) in model 1.

Model 1B

Substitute

$r = f \cdot \mathrm{P}_0 + d$

for equation (5d) in model 1.

Model 1C

Substitute

$\mathrm{P} \rightarrow l \cdot \mathrm{P} / \_\_\_\_\$$

for rule (5e) in model 1.

Modified transform for model 1

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

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

Model 2

Substitute

$\mathrm{T(P)} = \log((\mathrm{P} - b) / (r - b))$

for equation (5a) in model 1.

Models of f0 patterning: Liberman & Pierrehumbert

- Zero asymptote:
- X_i+1 = s x X_i
-
- X_i+1 - r = s x (X_i -r)
-
- F0 transform: converts measured F0 values into a new set of values that are assumed to behave in a simpler way - coser to underlying phonetic control parameters for intonation.
-
- Answer-background relation: taken to be constant ratio in transformed F0 values: k
-
- Downstep relation: taken to be constant ratio in transformed F0 values: s
-
- Lowering of F0 targets in utterance-final position; final lowering constant: l, utterance-final is bottom of entire system: b
-
- Transformed value of F0 target P depends on pitch range; reference level for each phrase: r
-
- Transformed value of P is its distance above r
-
- r constrained to remain is above final F0 value: b + d

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

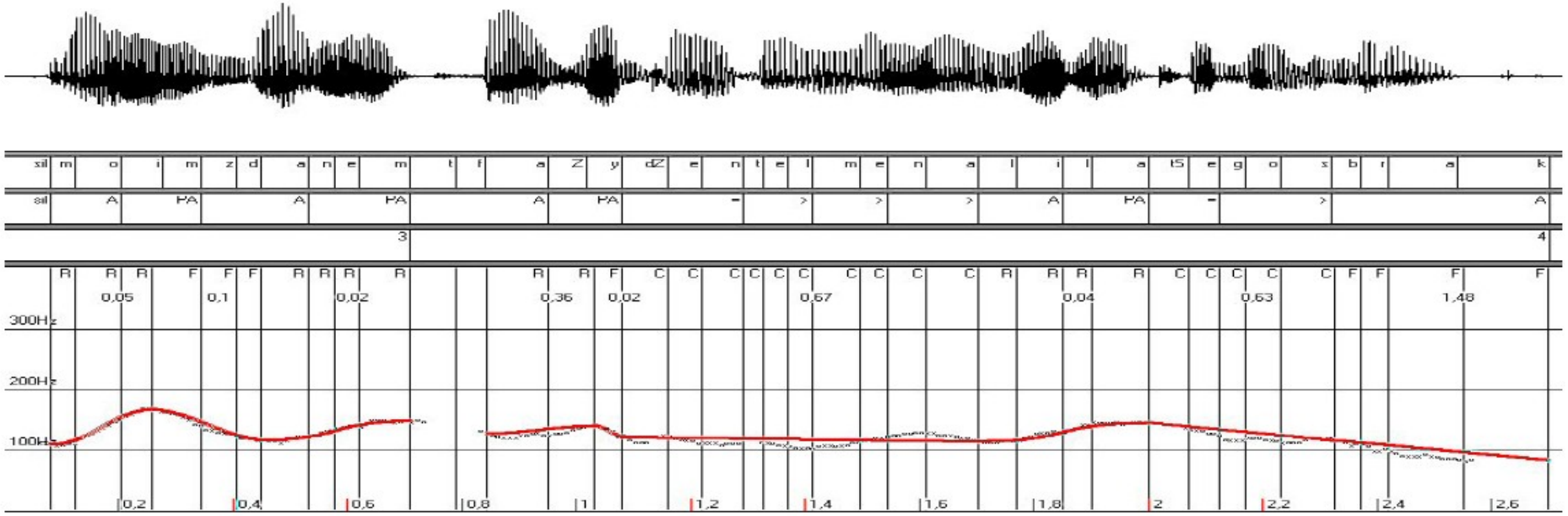


Figure 1: *Sentence:* In my opinion, the face of the lilac gentleman lacks something. *From top to bottom of the picture: waveform, .lab, .syl and .break tiers, and the stylization window. The original F0 contour is marked by dotted black line and the stylized F0 contour in red line.*

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

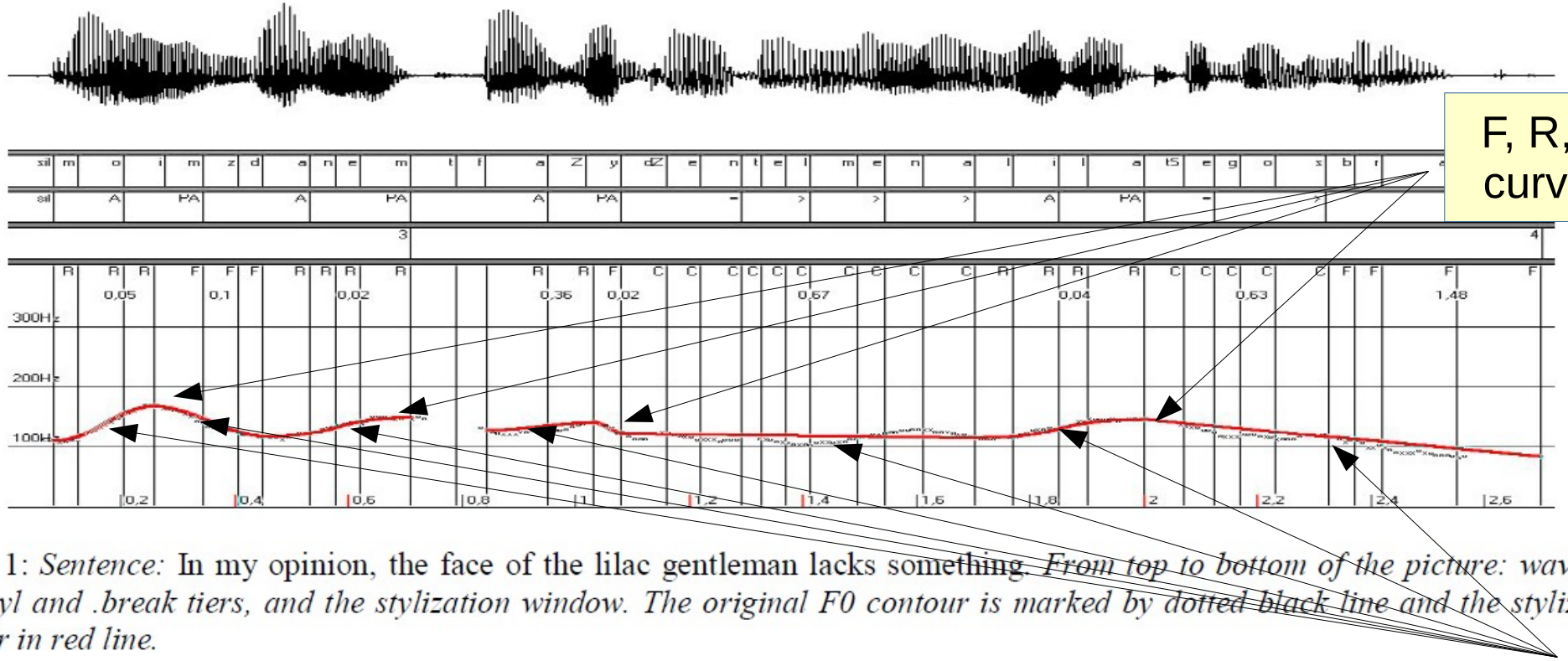


Figure 1: *Sentence:* In my opinion, the face of the lilac gentleman lacks something. *From top to bottom of the picture: waveform, .lab, .syl and .break tiers, and the stylization window. The original F0 contour is marked by dotted black line and the stylized F0 contour in red line.*

D&W 2006 stylisation model (SP3):

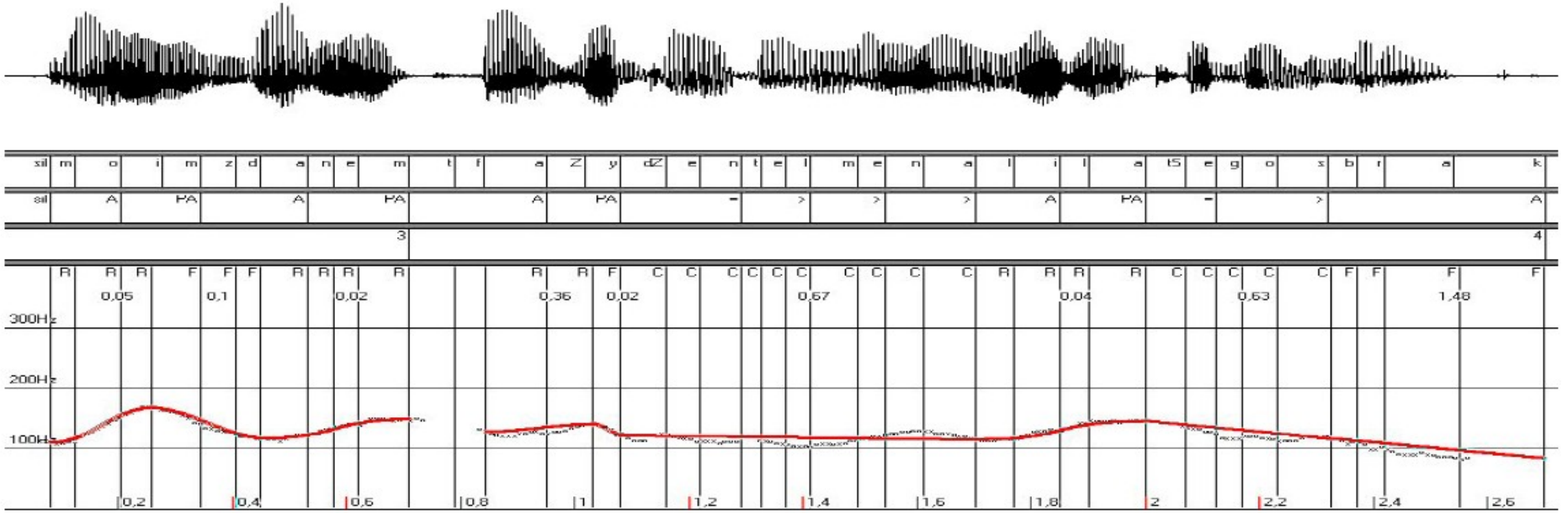
$IP \rightarrow IE^{+}$

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

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

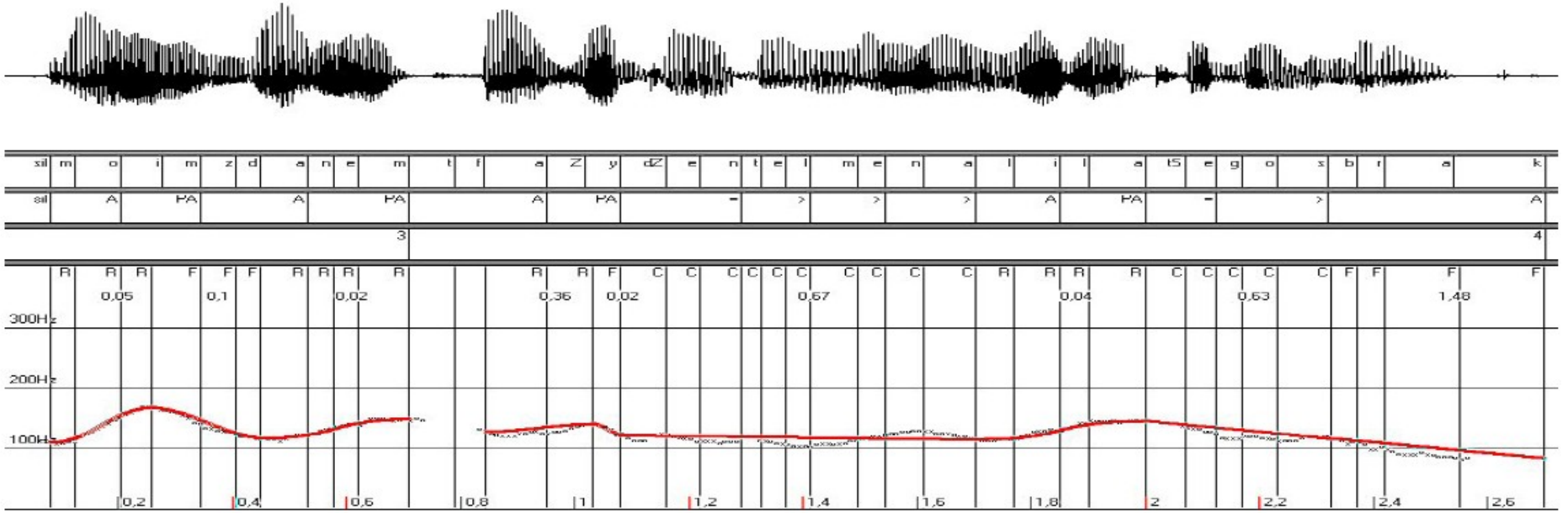
Compare F0 with stylised function with Normalised Mean Square Error:

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

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.

From phonetic models to phonological models

Phonology: representation systems

- Reminder:
  - Tonetic
  - Conversation analysis
  - Levels
    - 4 levels + junctures: Pike
    - 2 levels + break indices: ToBI (Pierrehumbert)
  - Relations

http://mi.eng.cam.ac.uk/~pat40/examples.html

Phonology: representation systems

- Reminder:
  - Tonetic
  - Conversation analysis
  - Levels
    - 4 levels + junctures: Pike
    - 2 levels + break indices: ToBI (Pierrehumbert)
  - Relations

I will leave the details of relating phonetics and phonology to your own research!

http://mi.eng.cam.ac.uk/~pat40/examples.html

Rank-Interpretation Architecture of Language