Österreichisches Forschungsinstitut für Artificial Intelligence

Intonation Modelling

(Fujisaki and more)

Hannes Pirker
Austrian Research Inst. for Artificial Intelligence (OFAI)

presentation given XI.2004 at the
Technical University Graz, Austria

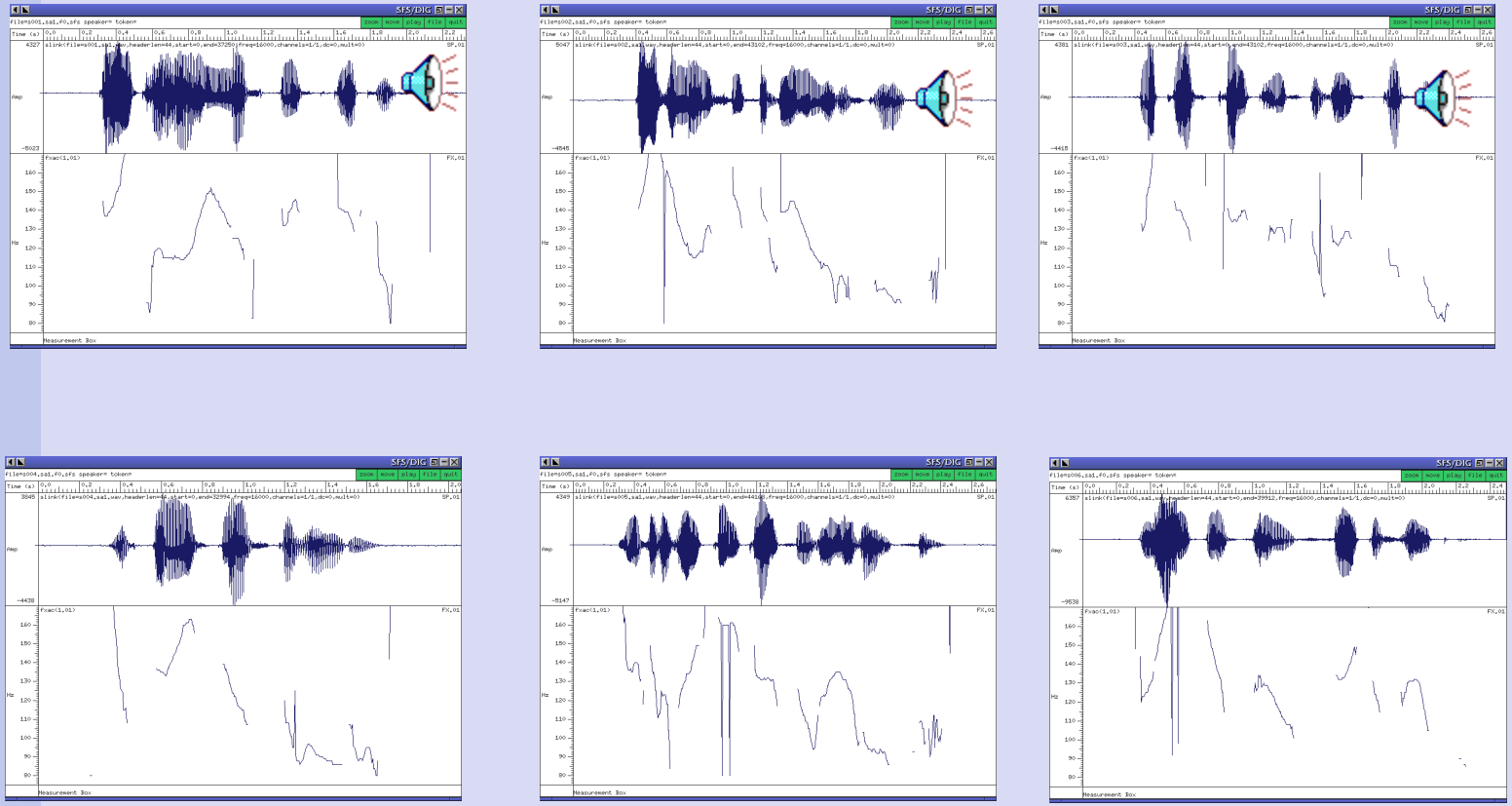
Intonation
Linguistic Functions

- Emphasisis
  - (Word) Stress
  - Accentuation
- Grouping together
  - Phrasing
- Sentence Mode
  - declarative vs. interrogative
  - (continuing vs. terminating)

Intonation
Paralinguistic & Nonlinguistic factors

- Speaking style
  - e.g. spontaneous vs. read
  - fairy tail vs. Newsreader
  - social status
- Emotion
  - e.g. aroused vs. bored
- Individual Factors
  - sex
  - age ...

Intonation/Pitch/F0



Österreichisches Forschnungsinstitut für
Artificial Intelligence

Application of Intonation Models

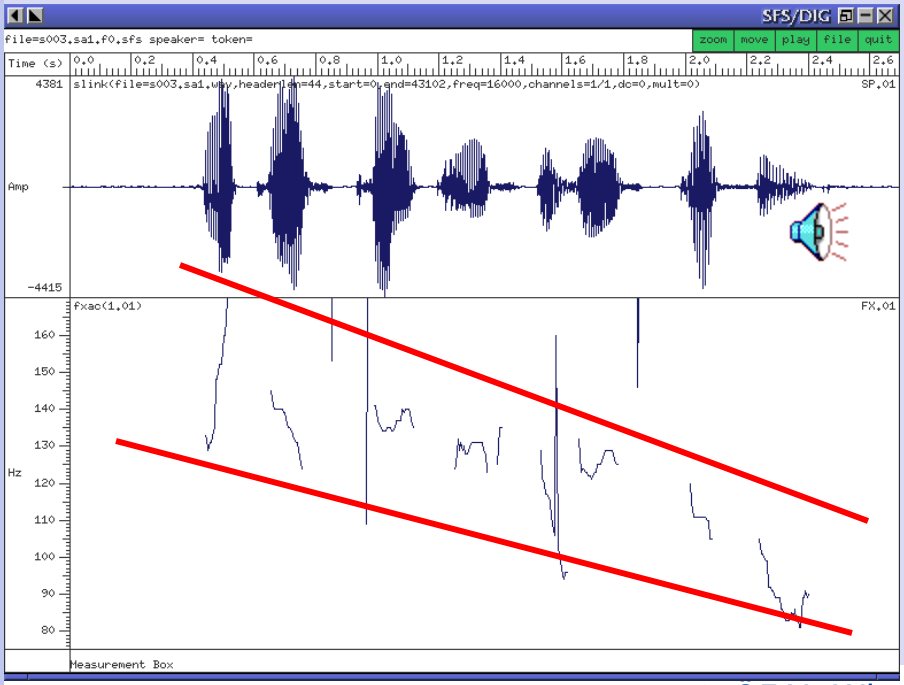
- Speech Synthesis
  - How to map linguistic function to intonation contours?
  - Aim for adequacy and naturalness
- Speech Recognition
  - Spot accents, focus stucture, sentence mode...
  - Analyse paralinguistic factors

Properties of F0-contours

- Microprosodic variaton
  - “dip” in contour at /l/
  - voiced/unvoiced transitions...

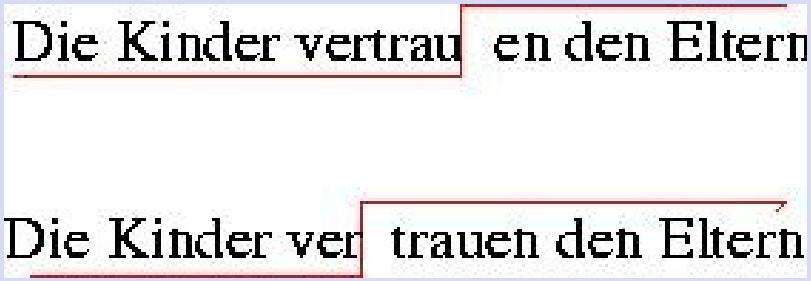
Properties of F0-contours: Declination

- Overall downtrend
  - of base- and topline.
  - reset at major phrase boundaries



Models of Intonation: Isacenko&Schaedlich 1964

- Simple switching of f0
  - between 150 : 178.6 Hz
- High correlation in listener's rating of linguistic function



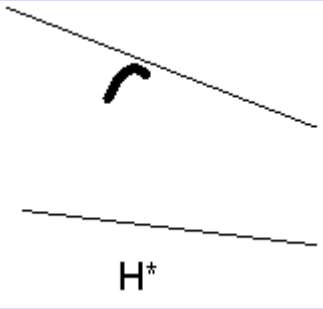
- question
- unfinished

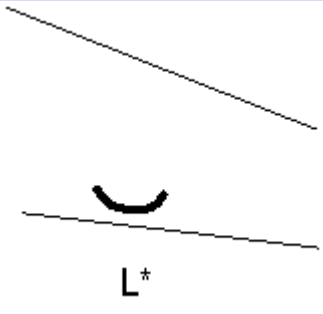
ToBI: Tones and Breaks Indices
Pierrehumbert, Hirschberg, Beckman

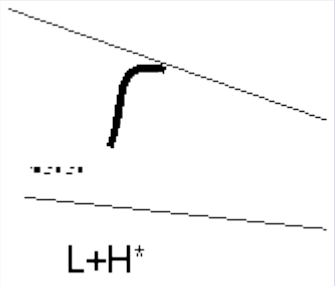
- Intonation described as series of H(igh) and L(ow) target tones
- Accent Tones
  - H*, L*, L+H*, H+L, ...
- Phrasal Tones
  - H, L
- Boundary Tones
  - H%, L%

ToBI: Tones and Breaks Indices
Accent Tones

- * denotes alignment with stressed syllable
- No direct quantitative information
  - e.g. H* can denote be a steep and high hill or a gentle slope



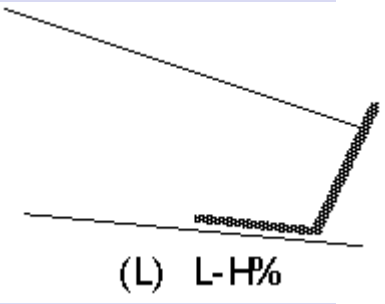


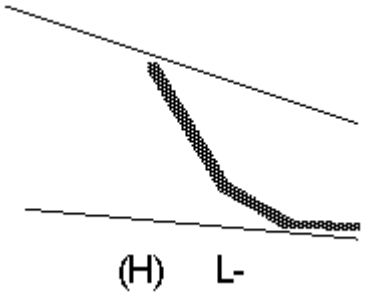


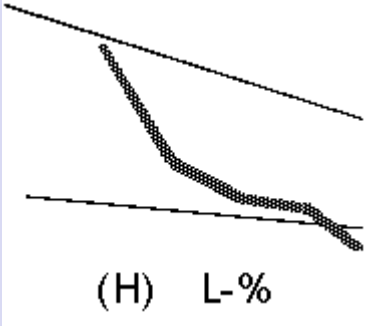
ToBI: Tones and Breaks Indices Boundaries

- Boundary tones H% and L%
- Combined with L- H-

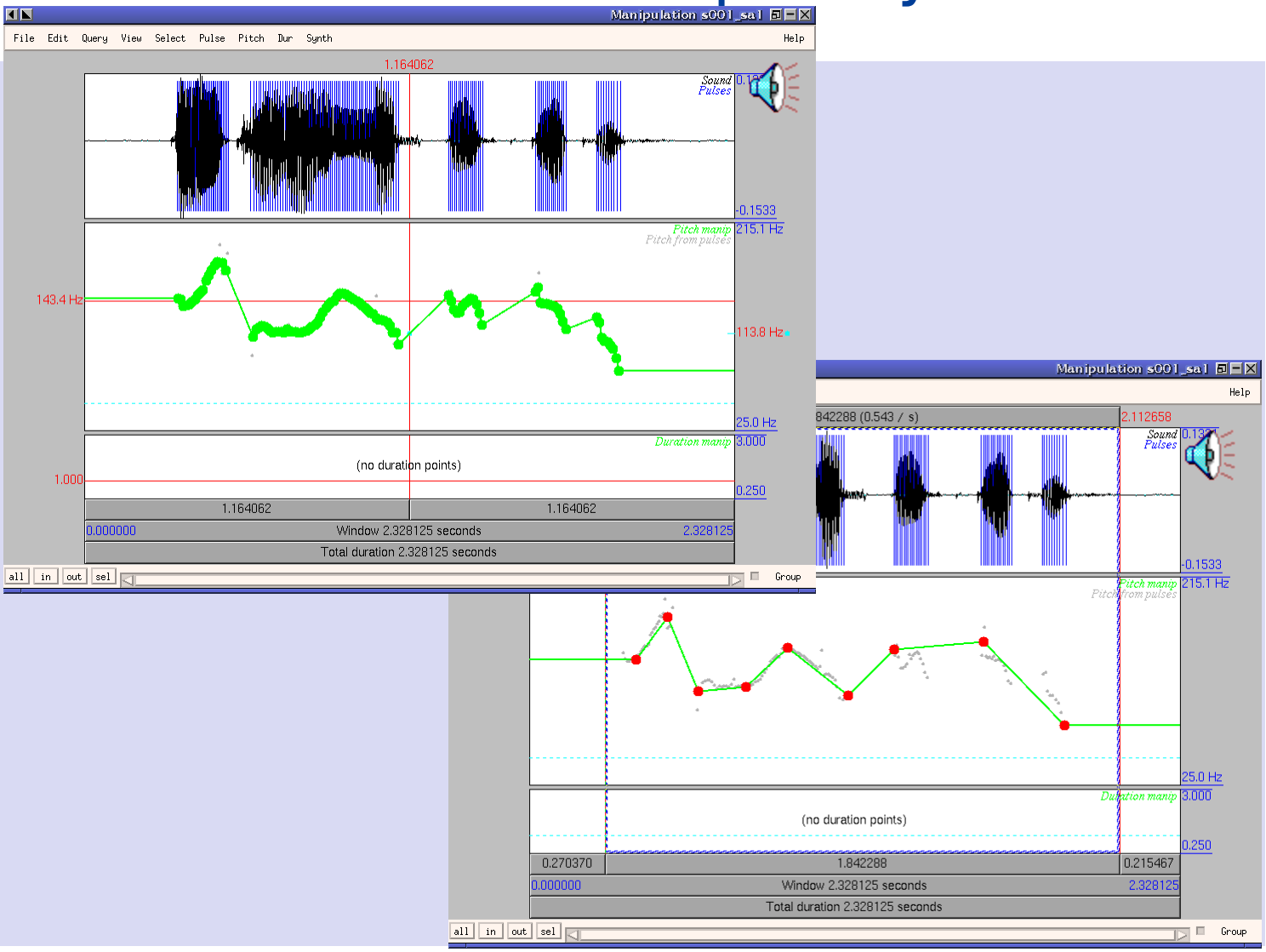
- E.g.
  - L-L% : typical final fall in declarative sentenes
  - H-H%: typical rise in questions





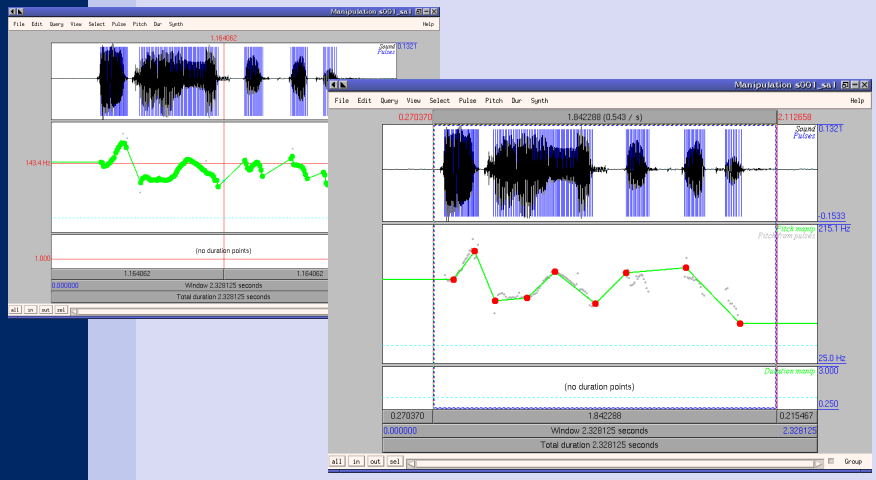


F0 - stylisation: getting rid of microprosdy, flaws, ...



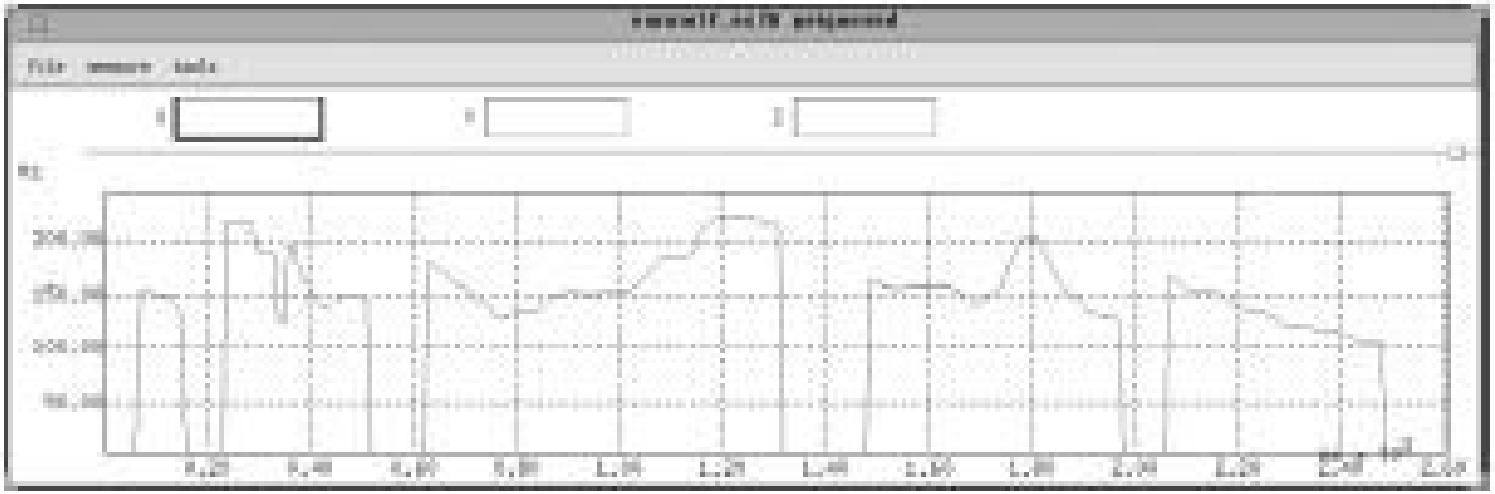
Quantitative Models of Intonation: IPO Model (tHart/Collier)

- 1. stylise to “perceptually identic”



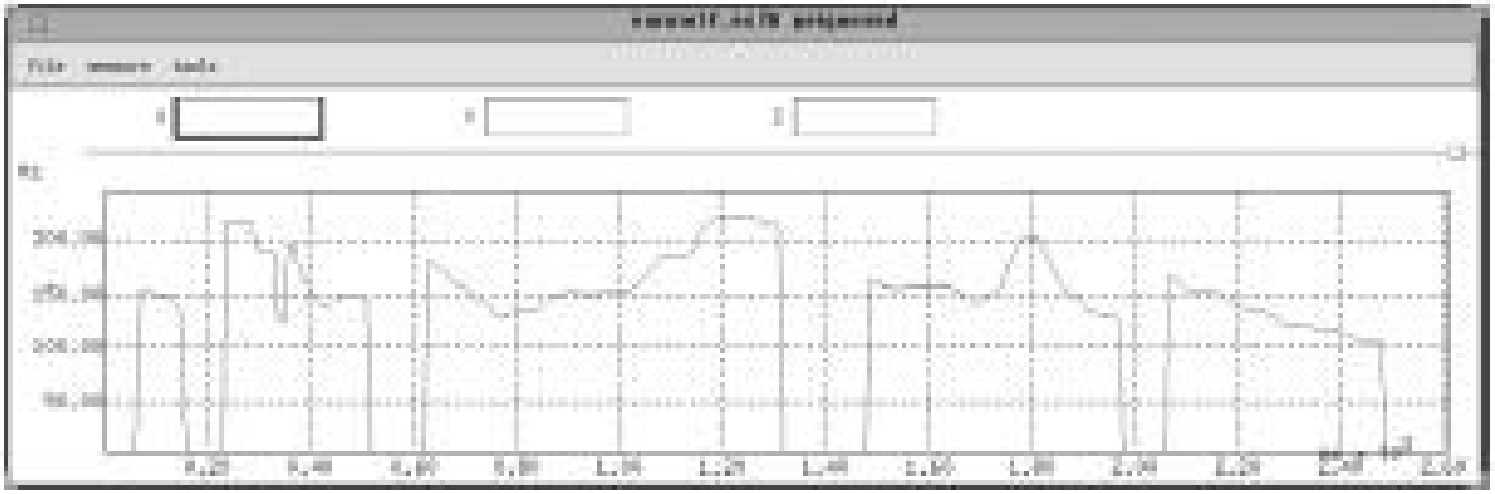
- 2. Functionally identic
- 3. Inventory of 11 accent-lending and phrase-marking movements

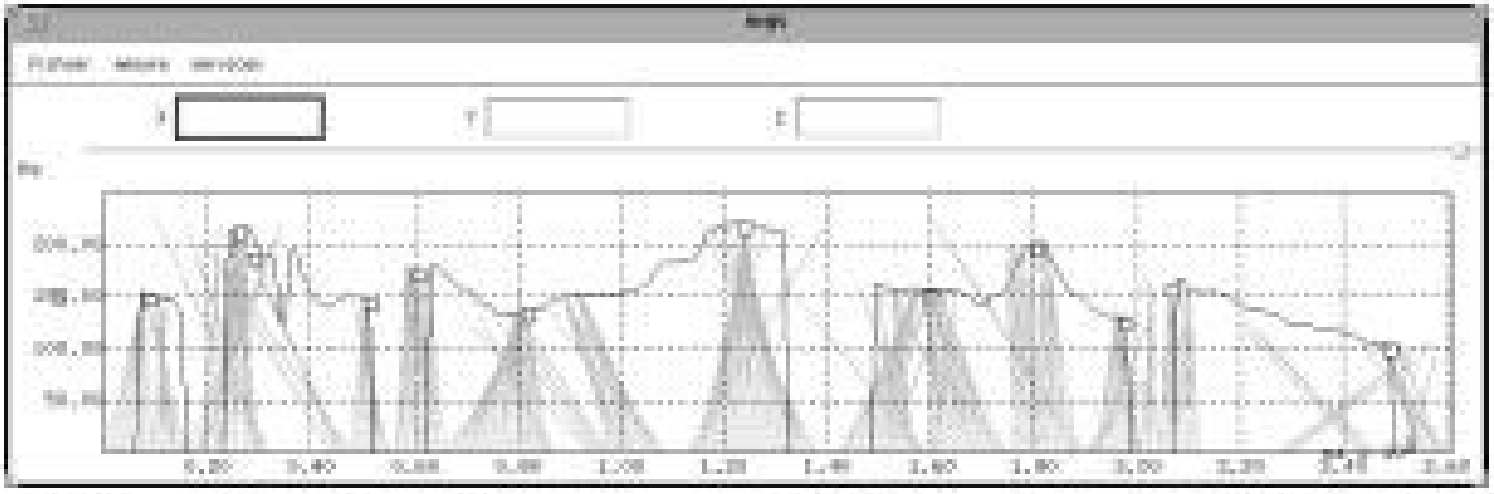
MOMEL
Melodic Modellisation (Hirst 1991)



- Modelling contour via quadratic splines
- Claimed universal (Language independent)

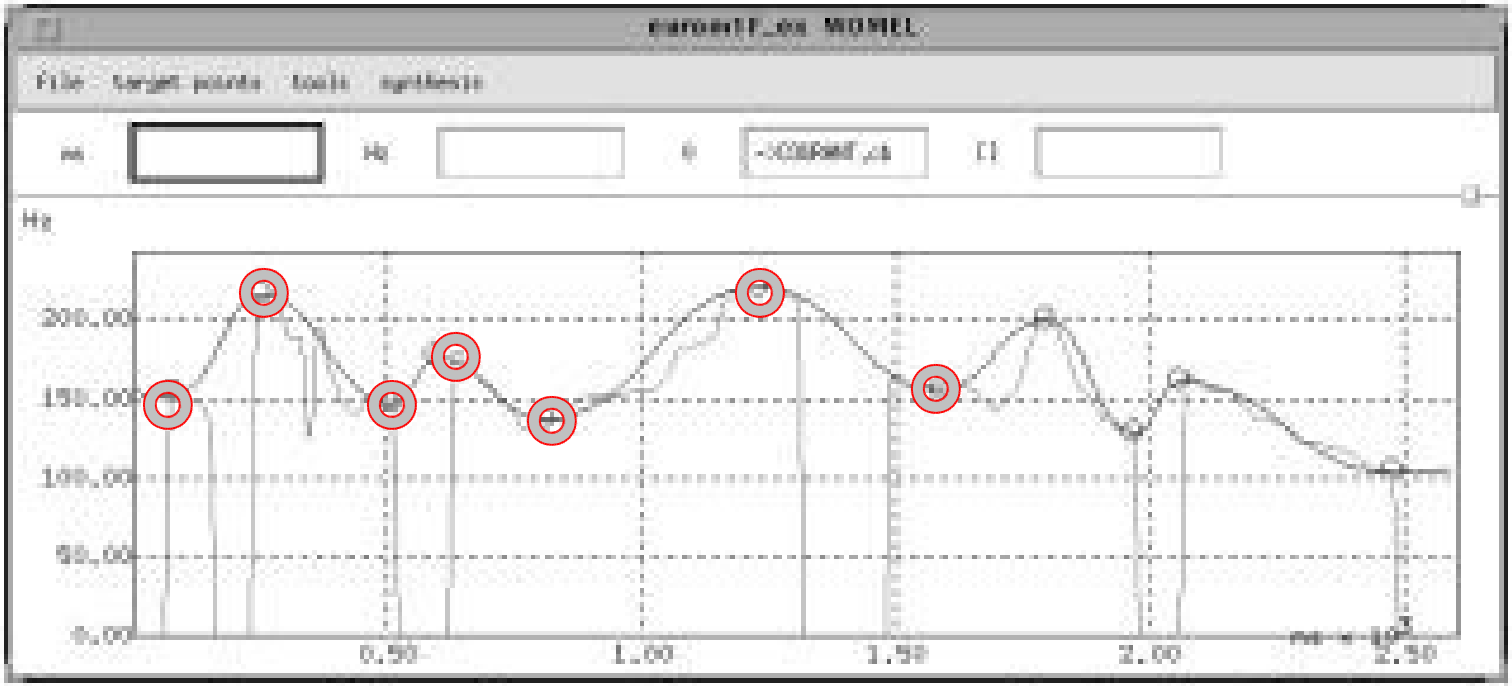
MOMEL
Melodic Modellisation (Hirst 1991)
Quadratic Splines





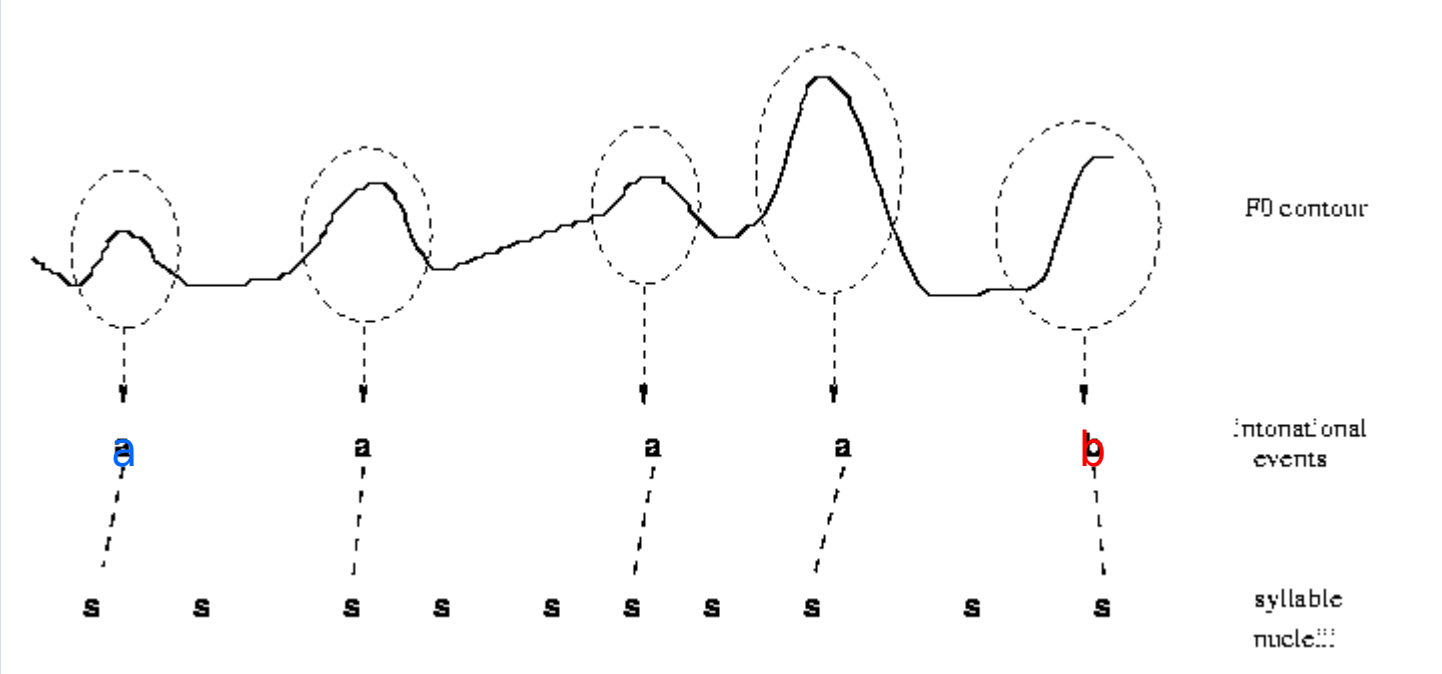
MOMEL
Melodic Modellisation (Hirst 1991)

- Freely available
- Valid smoothing method

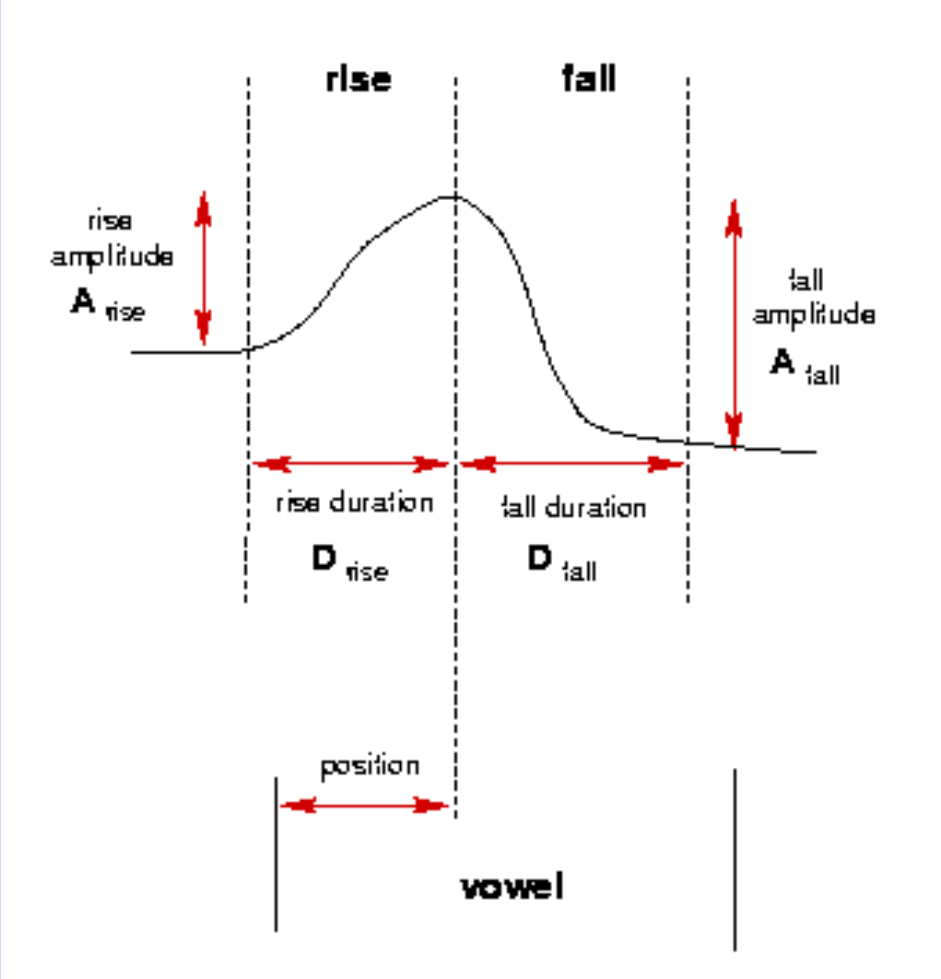


Tilt (Taylor & Black/ EST)

- Intonation contour as a series of (a)ccent and (b)oundary events



Events modelled by Rise-Fall-Coefficients (RFC)



- Amplitude
  - A_{rise}
  - A_{all}
- Duration
  - D_{rise}
  - D_{fall}
- *„Absolute Position“*
  - *Some absolute f0 value (peak, start)*
  - *Some absolut positon in timeline*

Tilt –value:
Ratio between difference and sum

$$tilt_{amp} = \frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|}$$

$$tilt_{dur} = \frac{D_{rise} - D_{fall}}{D_{rise} + D_{fall}}$$

- Tilt values
  - +1 rise component only
  - -1 fall component only
  - 0 rise and fall symetrical

Tilt –model
Reduction of necessay parameters to 3

- Intonation events encoded via:
- Dur_{event} (sum of fall and rise)
- $Amplitude_{event}$ (sum of fall and rise)
- $Tilt_{event}$
- (absolute positioning)

Combined into global Tilt value

- tiltAmp and tiltDur highly correlated
- Combined into:
- tilt = (tiltAmp + tiltDur) / 2

$$tilt = \frac{|A_{rise}| - |A_{fall}|}{2(|A_{rise}| + |A_{fall}|)} + \frac{D_{rise} - D_{fall}}{2(D_{rise} + D_{fall})}$$

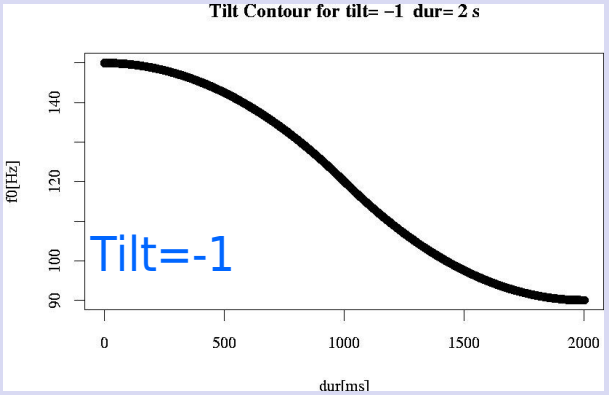
Tilt –value:
Ratio between difference and sum

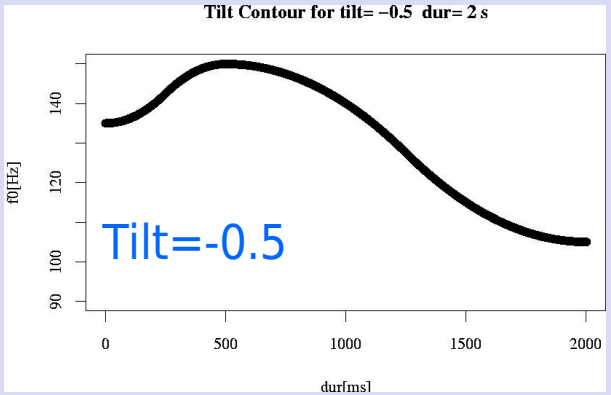
$$tilt_{amp} = \frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|}$$

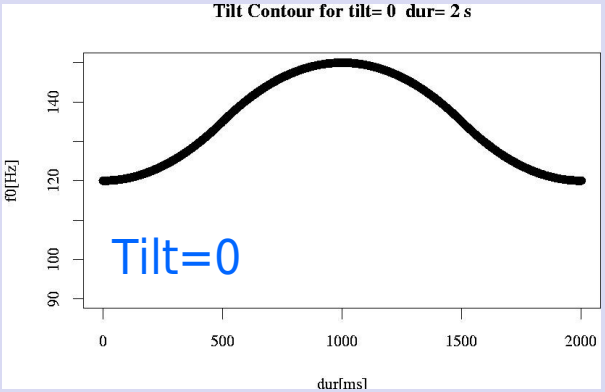
$$tilt_{dur} = \frac{D_{rise} - D_{fall}}{D_{rise} + D_{fall}}$$

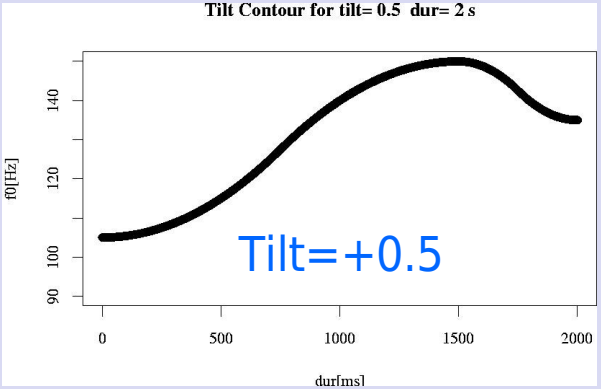
$$tilt = \frac{|A_{rise}| - |A_{fall}|}{2(|A_{rise}| + |A_{fall}|)} + \frac{D_{rise} - D_{fall}}{2(D_{rise} + D_{fall})}$$

Tilt
Dur=2sec, Amp=60, F0peak=150



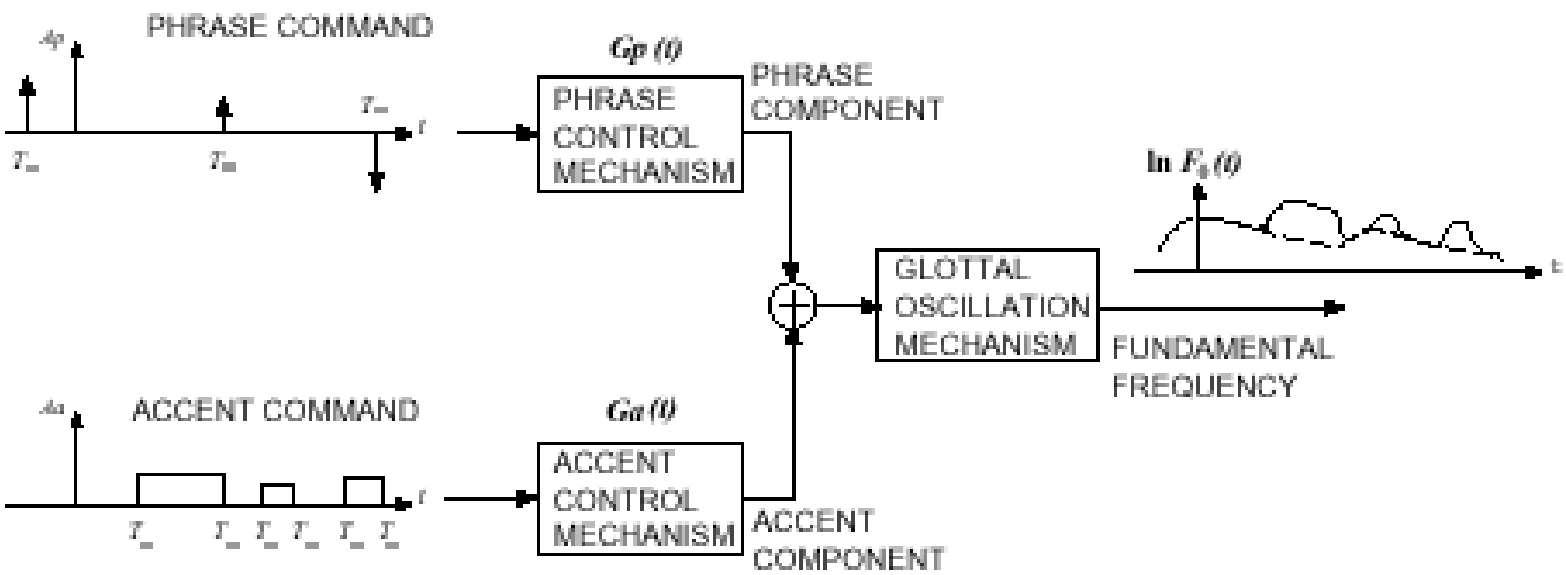






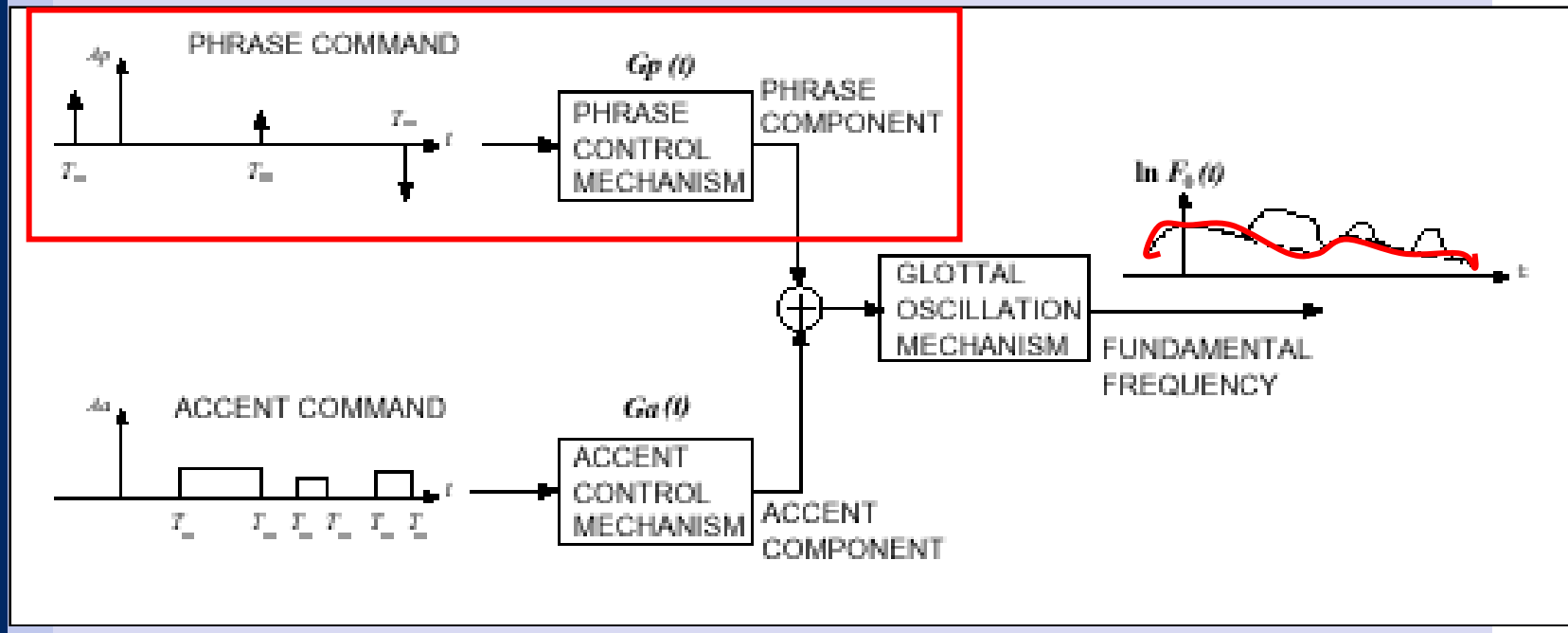
Fujisaki-Model
Hiroya Fujisaki 1984

- Superpositional Model: F0 production modelled by 2 separate components



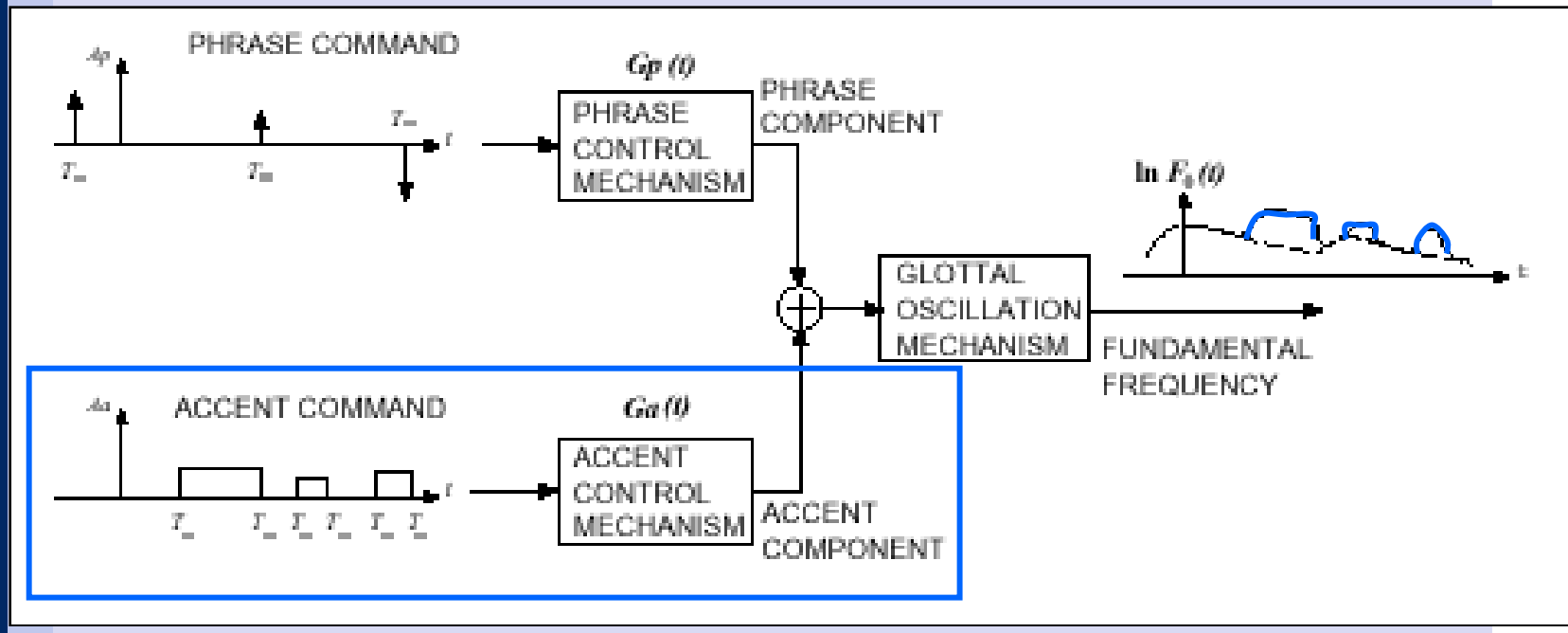
Fujisaki-Model
Hiroya Fujisaki 1984

- PHRASE component driven by:
  - Phrase commands: Tp and Ap



Fujisaki-Model
Hiroya Fujisaki 1984

- ACCENT component driven by:
  - Accent commands: switch on and off at T1, and T2, Aa



Addition in the logarithmic domain

- F0 = Baseline + PhraseComponent + AccentComponent

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj}[G_a(t - T_{1j}) - G_a(t - T_{2j})].$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

$$G_a(t) = \begin{cases} \min\left[1 - (1 + \beta t) \exp(-\beta t), \gamma\right], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

Addition in the logarithmic domain

- F0 = Baseline + PhraseComponent + AccentComponent

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj}[G_a(t - T_{1j}) - G_a(t - T_{2j})].$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

$$G_a(t) = \begin{cases} \min\left[1 - (1 + \beta t) \exp(-\beta t), \gamma\right], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

Addition in the logarithmic domain

- F0 = Baseline + PhraseComponent + AccentComponent

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj}[G_a(t - T_{1j}) - G_a(t - T_{2j})].$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

$$G_a(t) = \begin{cases} \min\left[1 - (1 + \beta t) \exp(-\beta t), \gamma\right], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

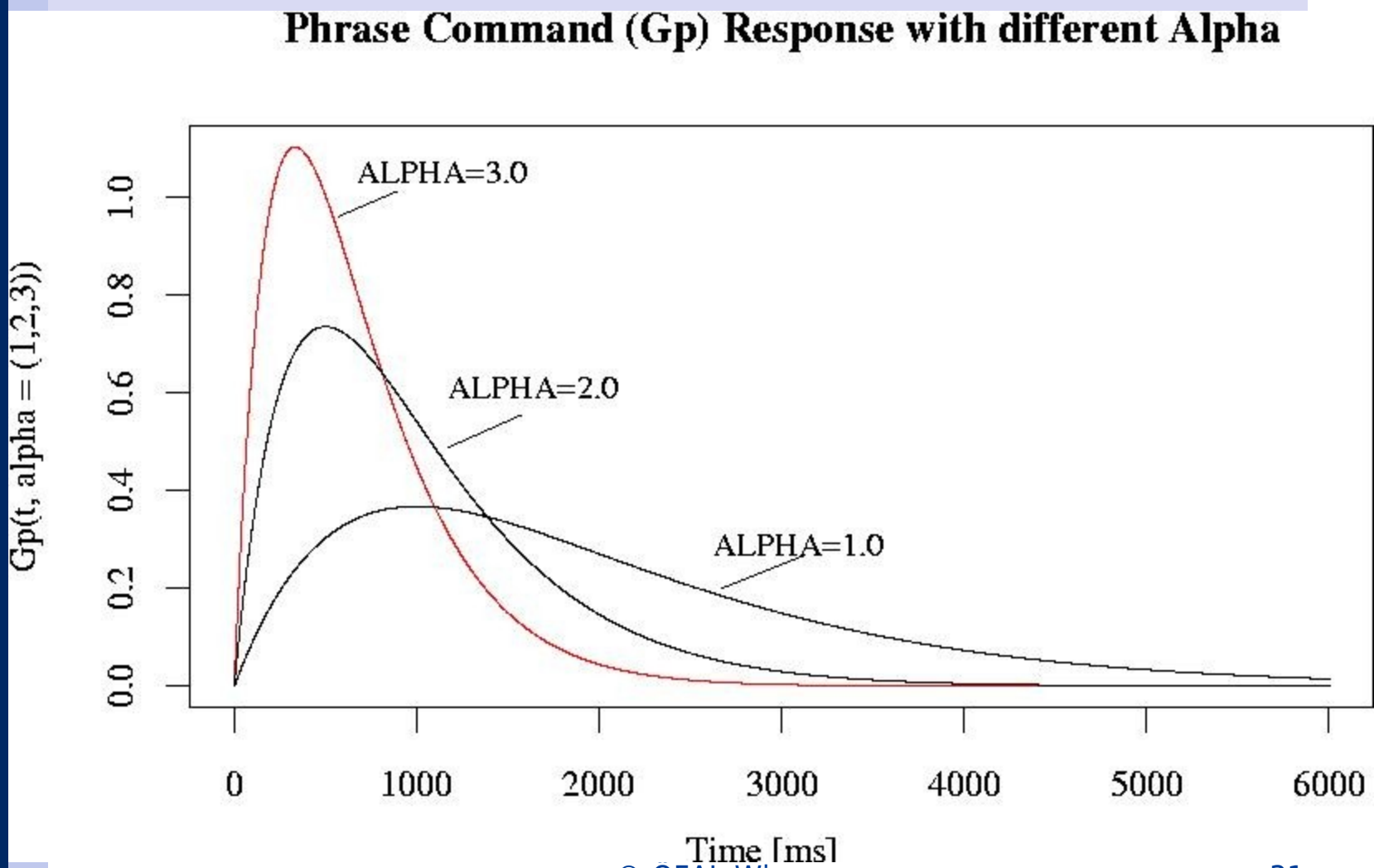
Addition in the logarithmic domain

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{aj} [G_a(t - T_{1j}) - G_a(t - T_{2j})].$$

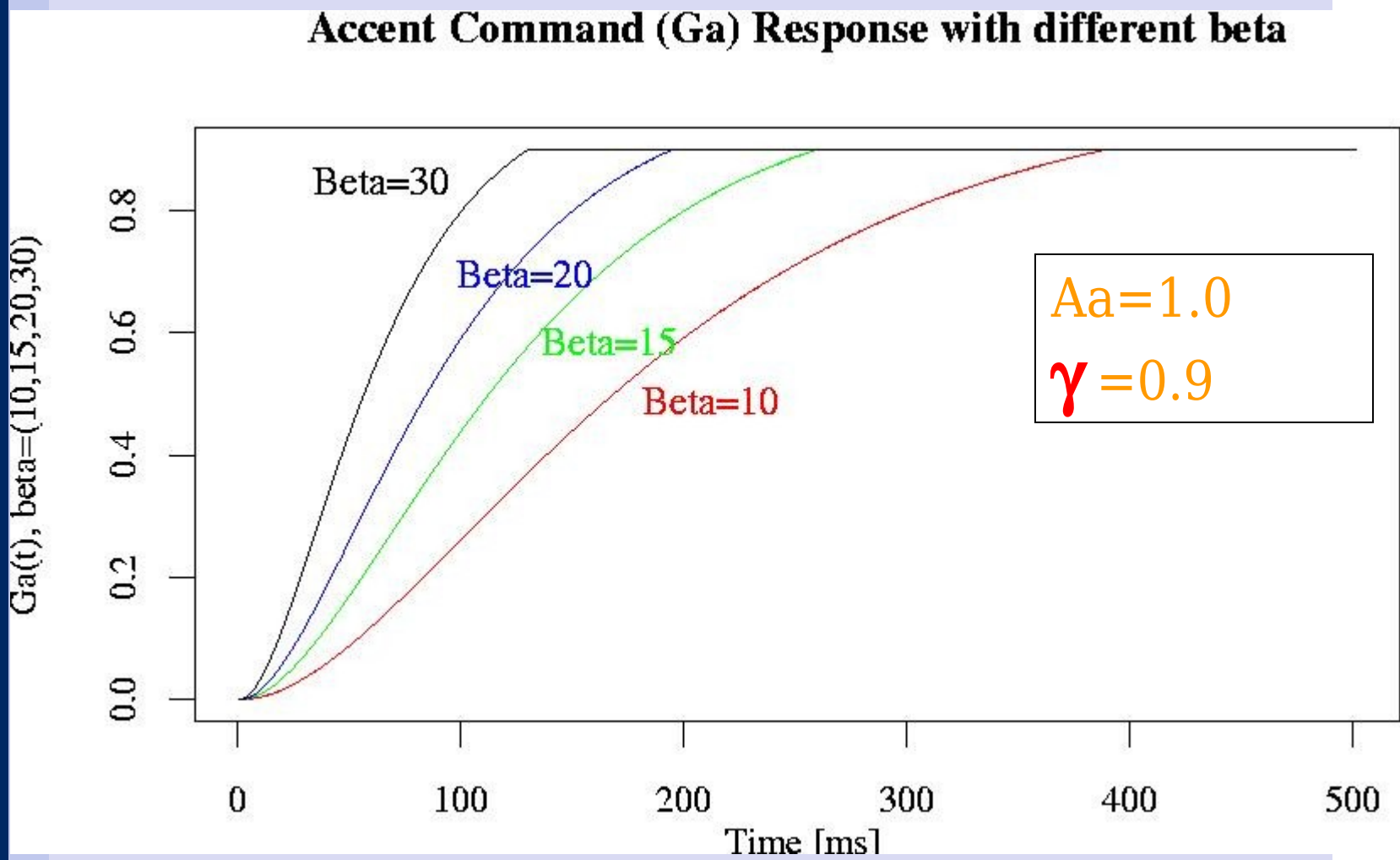
$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

$$G_a(t) = \begin{cases} \min\left[1 - (1 + \beta t) \exp(-\beta t), \gamma\right], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

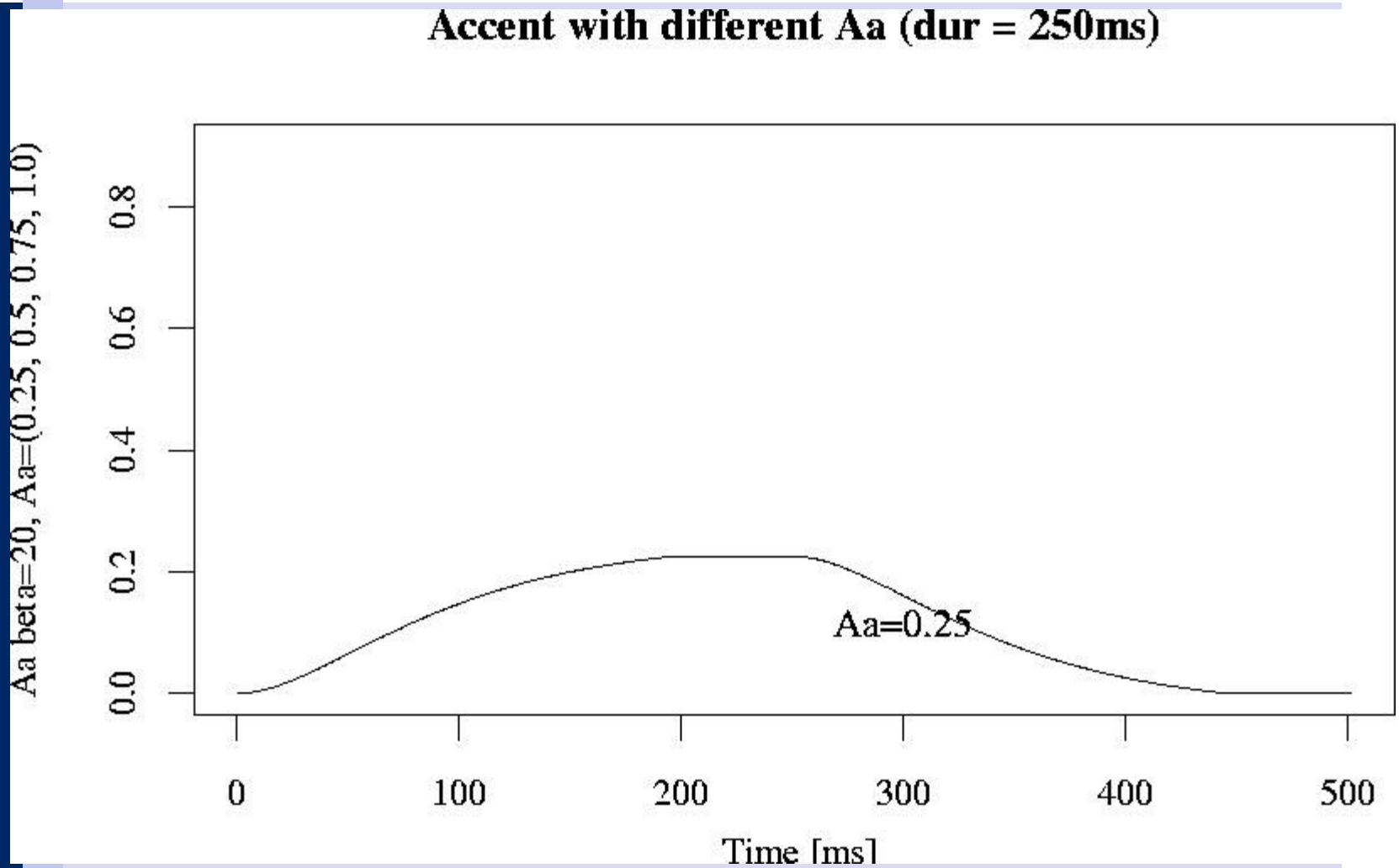
Phrase Component
$Gp(t) = \alpha^2 t \exp(-\alpha t))$



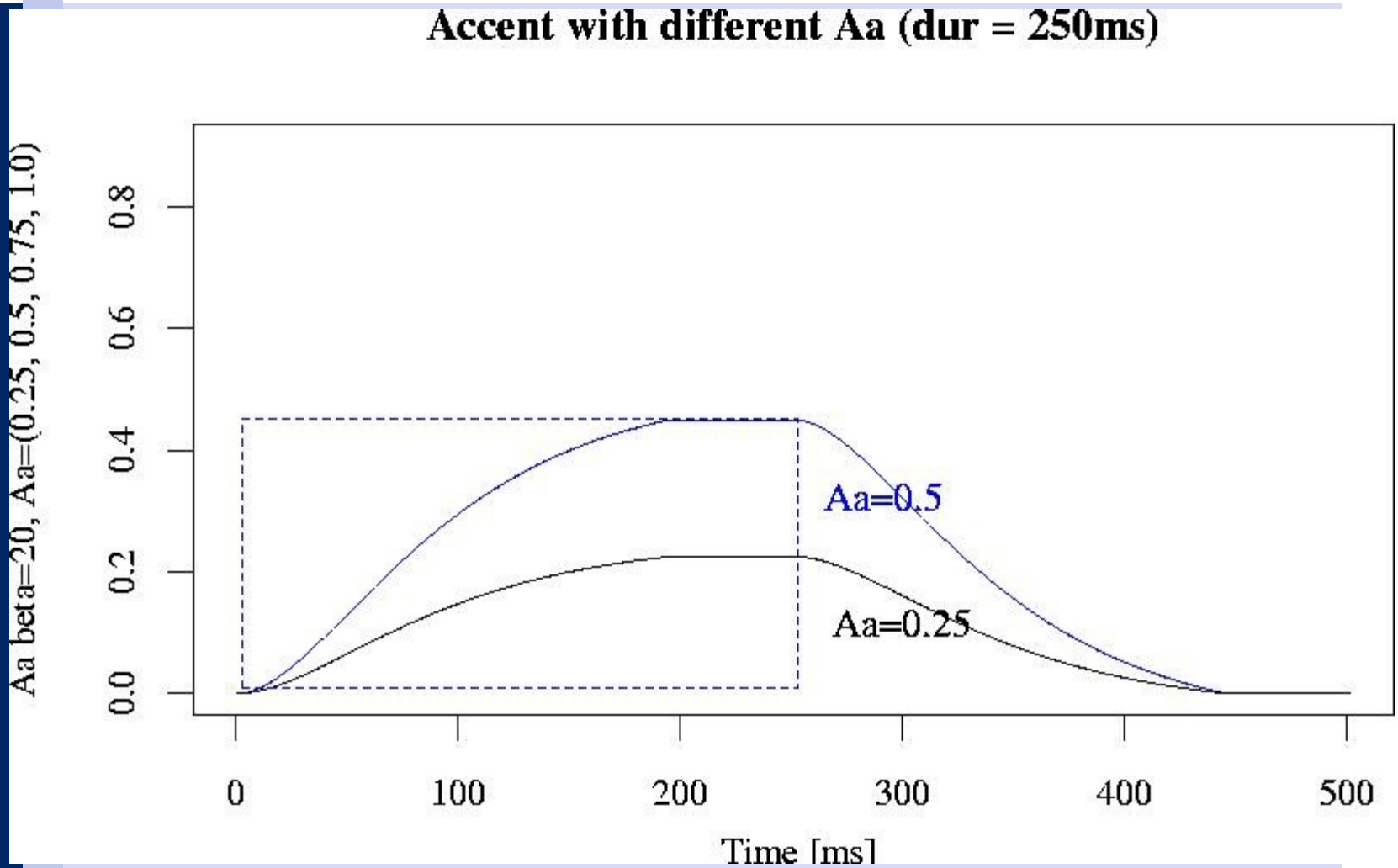
Accent Command:
$Ga(t)=\min[1 - (1+\beta t) * \exp(-\beta t), \gamma]$



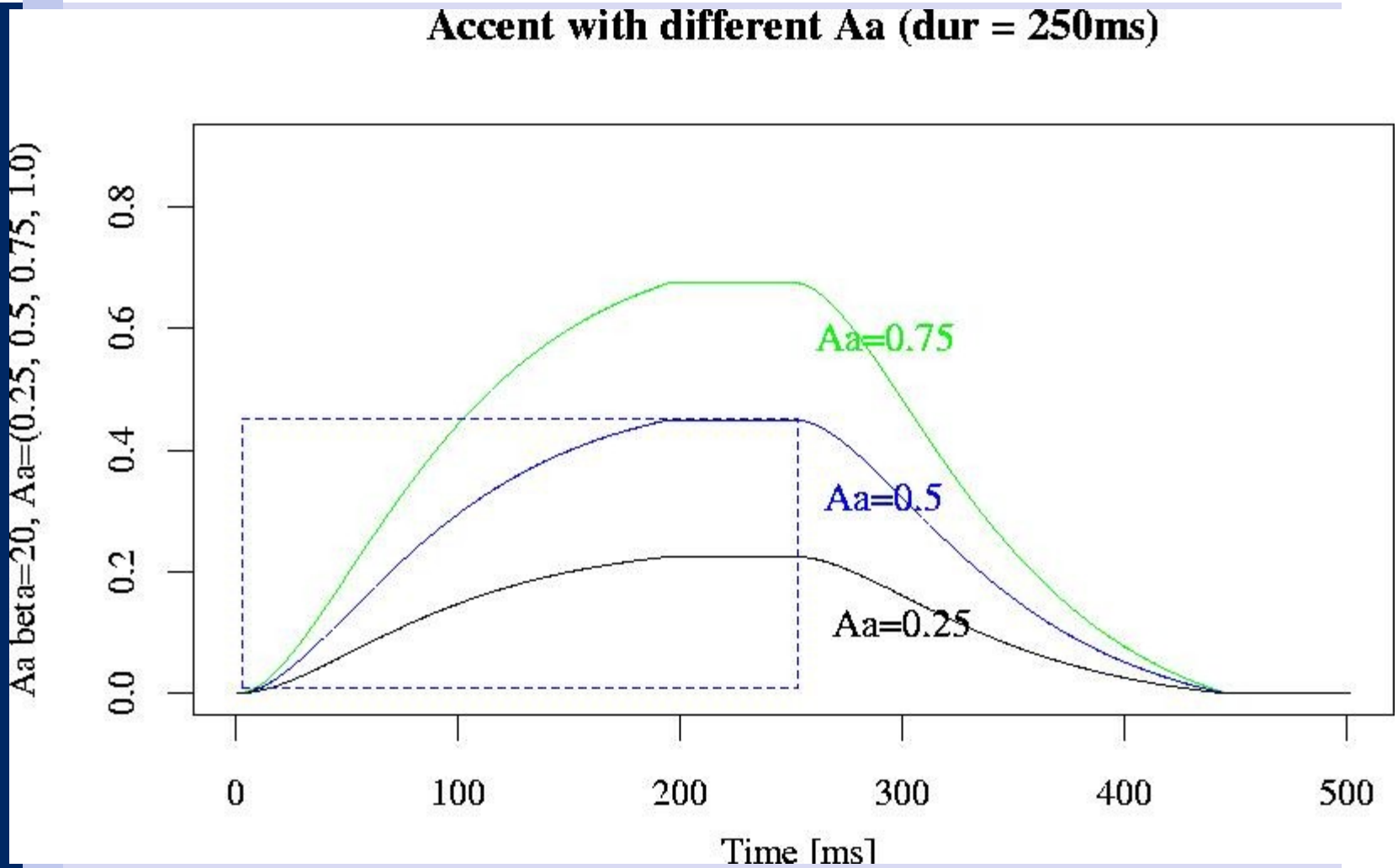
Accent form with different Amplitude Aa



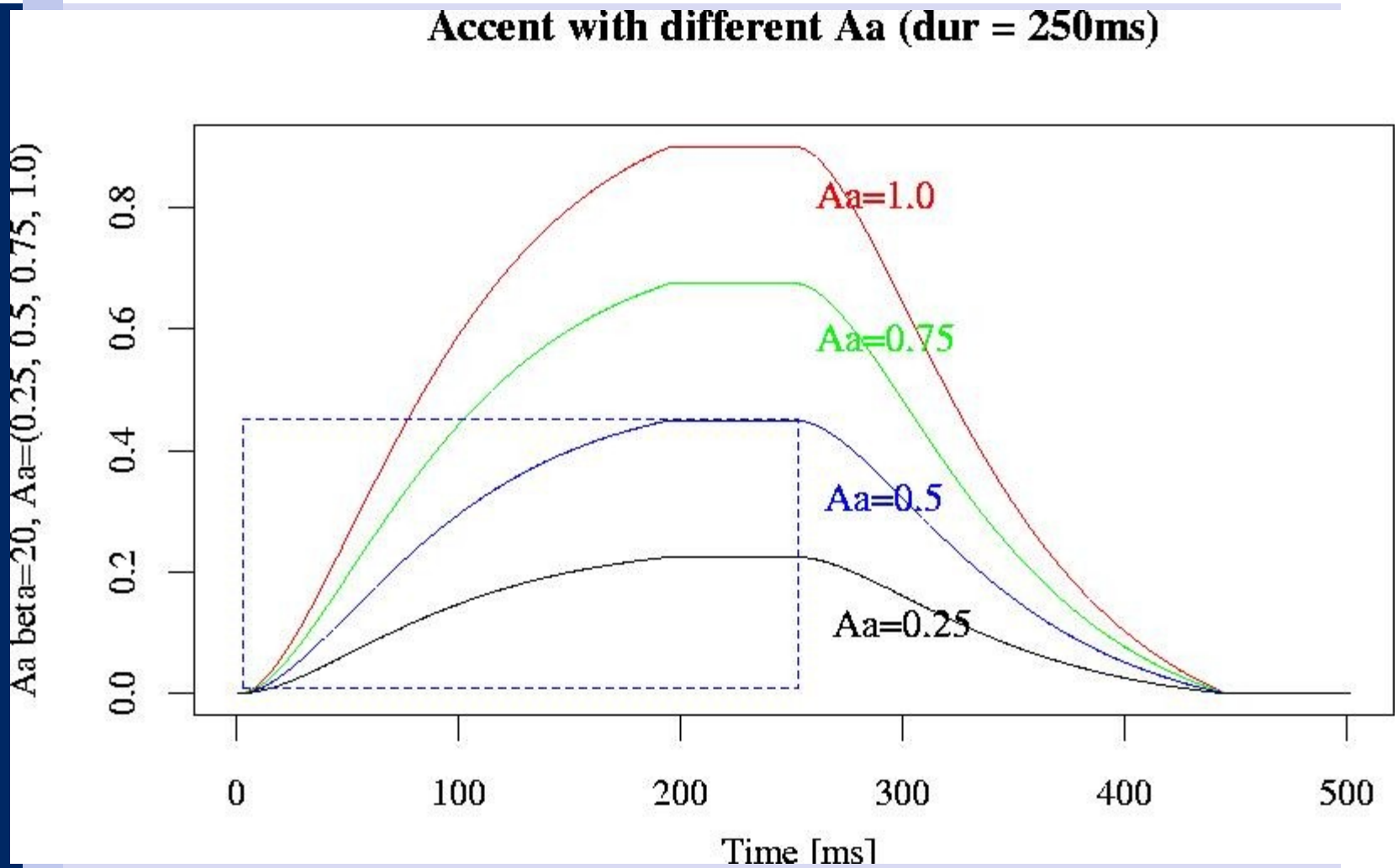
Accent form with different Amplitude Aa



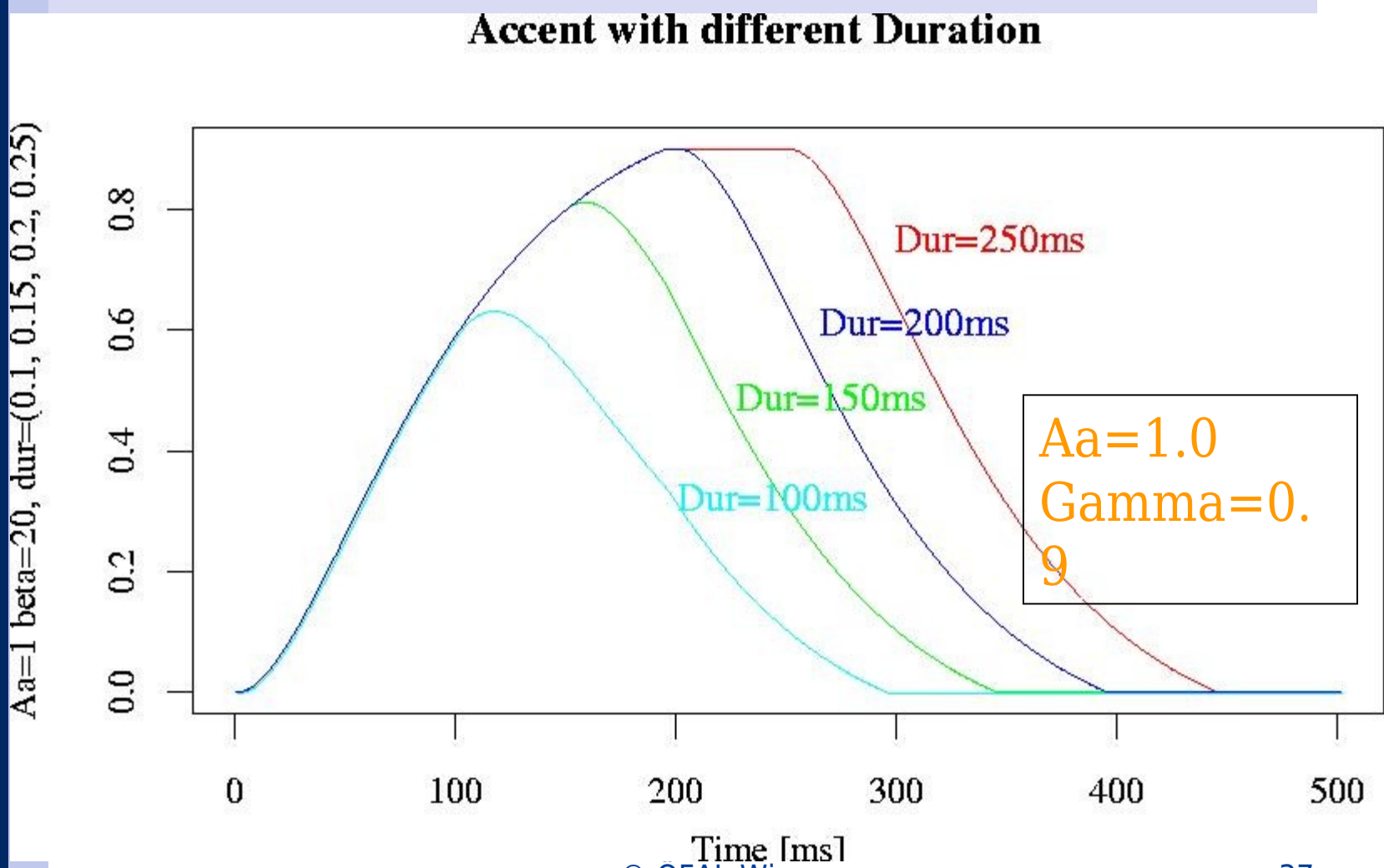
Accent form with different Amplitude Aa



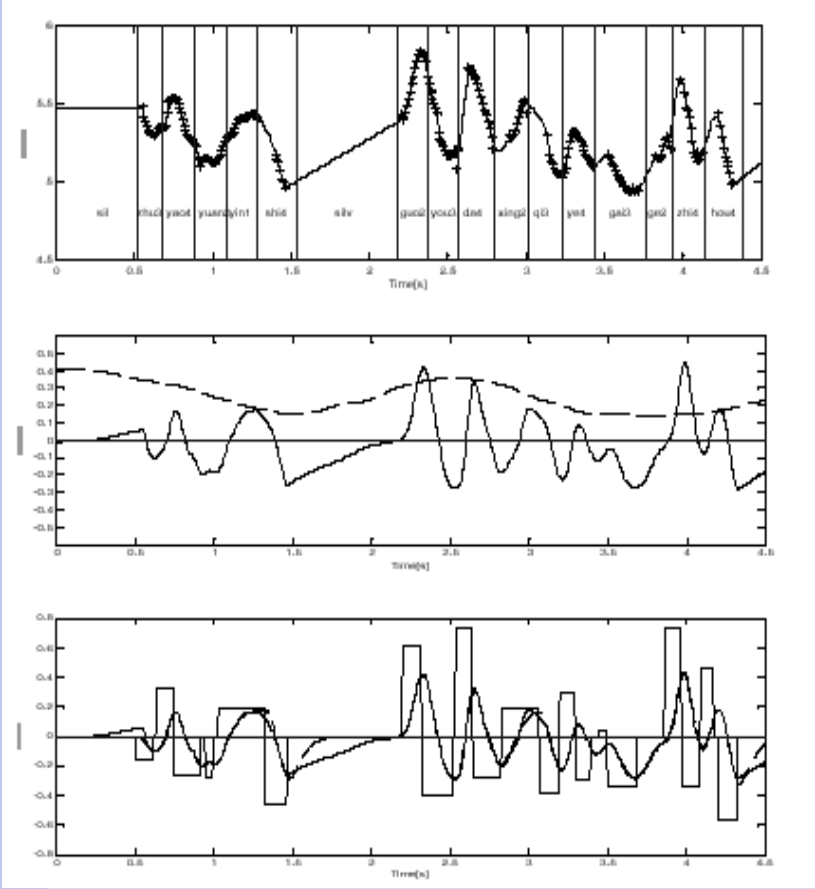
Accent form with different Amplitude Aa



Accent form with different Duration



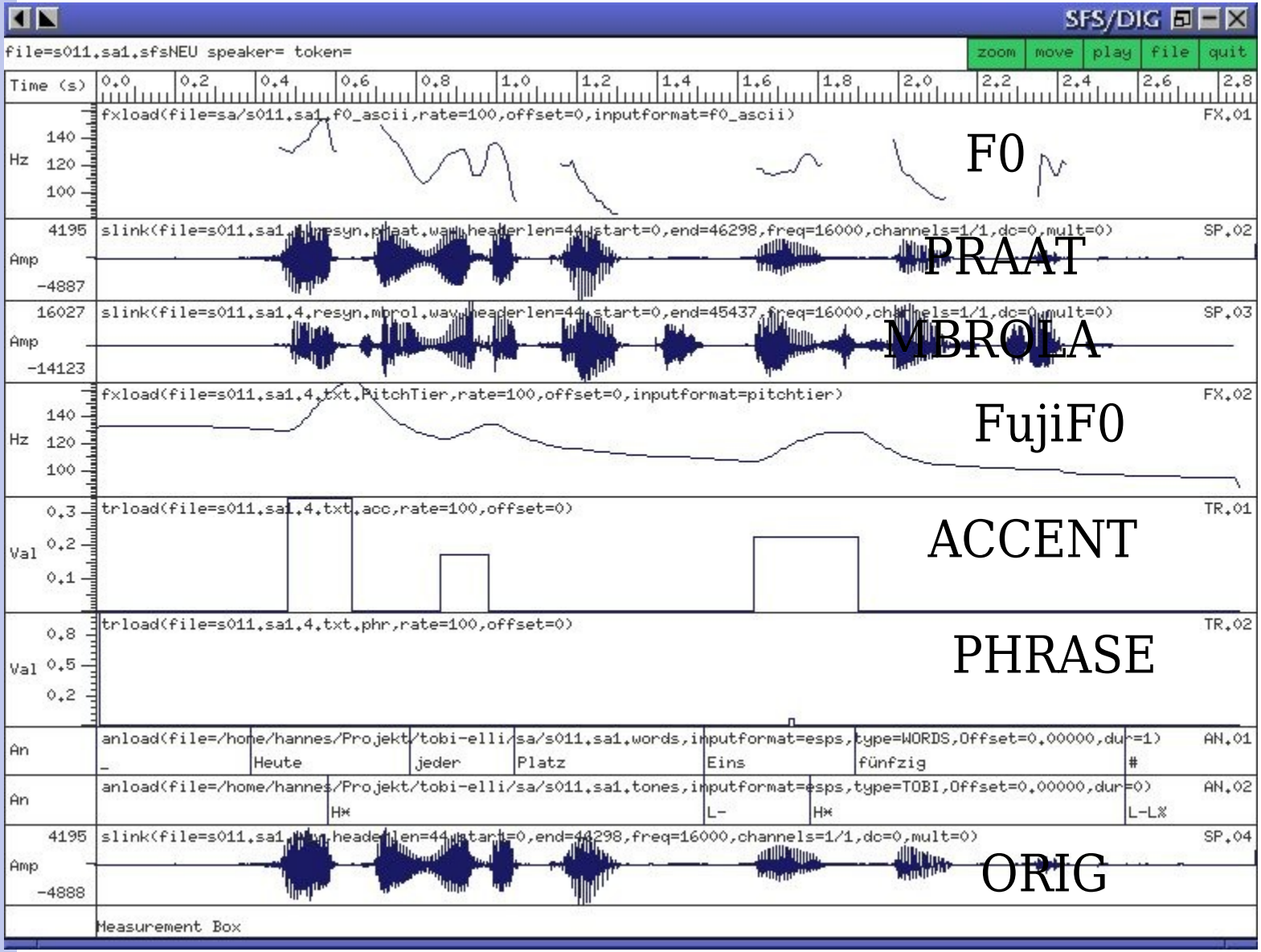
How to extract Phrase- and Accent-Commands



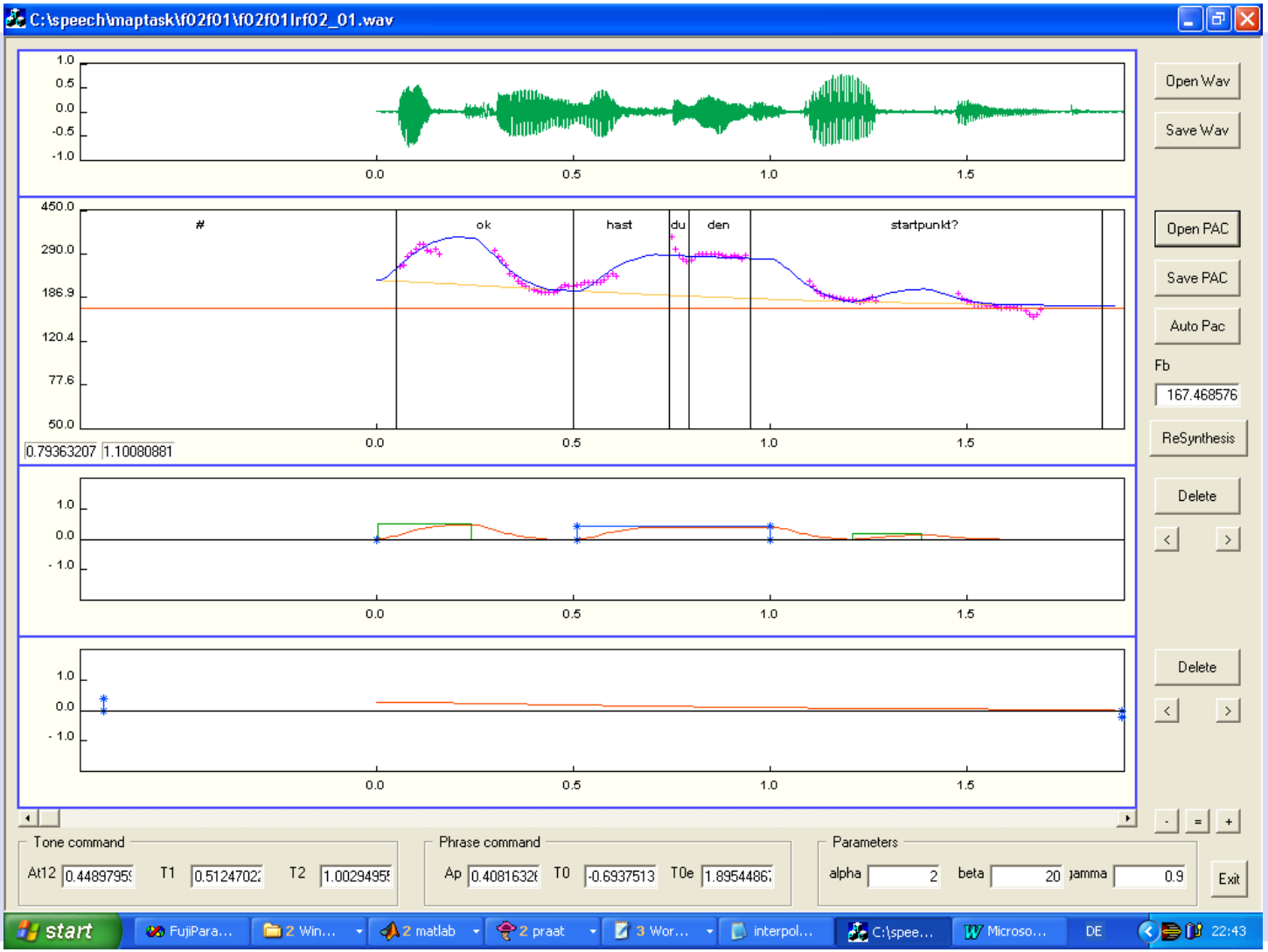
- 1. Smoothing
- 2. Highpassfilt (0.5 Hz): HFC
- 3. Subtract: LFC
  - Minima ->Tp
  - Maxima -> ~ Ap
- 4. HFC
  - Minima -> Ta1
  - Maxima -> Ta2
- 5. Hillclimb search

Österreichisches Forschungsinstitut für Artificial Intelligence

Example



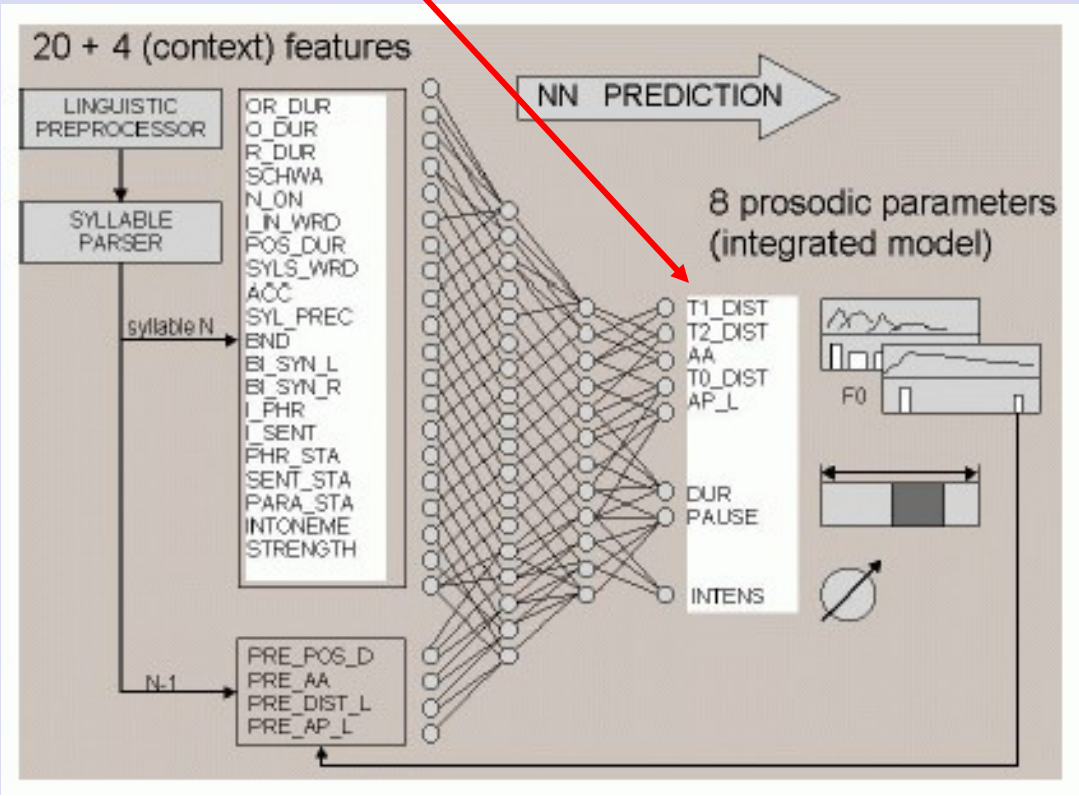
FujiParamEditor



Österreichisches Forschungsinstitut für Artificial Intelligence

Application Example: Using Fujisaki-Model in DRESS

- Fuji params predicted together with Duration and Intensity



Summary

- Overview on some quantitative models of intonation
- IPO
- MOMEL
- Tilt
- Fujisaki

- http://www.ofai.at/~hannes.pirker

Resources, Literature etc.

- Homepage of Hansjoerg Mixdorff where you find lots of references to his work on using Fujisaki's model for German and other languages and can download the FujisakiEditor http://www.tfh-berlin.de/~mixdorff/Research.htm
- Praat: The indispensible tool for speech analysis http://www.fon.hum.uva.nl/praat/
- A praat implementation for MOMEL http://www.icp.inpg.fr/~loeven/Praat/momel_english.html
- The Edinburgh Speech Tools (EST) which contain the Tilt-model. http://festvox.org/docs/speech_tools-1.2.0/