ÖFAI

Österreichisches Forschnungsinstitut für
Artificial Intelligence

Intonation Modelling
(Fujisaki and more)

Hannes Pirker
Austrian Research Inst. for Artificial Intelligence (ÖFAI)
hannes@oefai.at

presentation given XI.2004 - Graz

Intonation
Linguistic Functions

Österreichisches Forschungsinstitut für
Artificial Intelligence

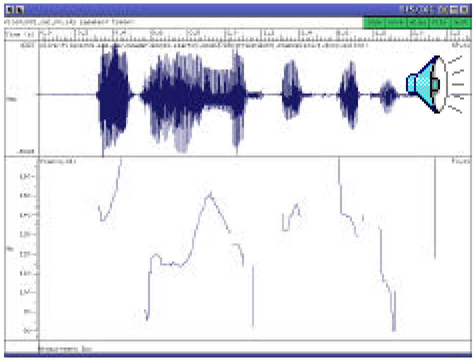
- Emphasisis
  - (Word) Stress
  - Accentuation
- Grouping together
  - Phrasing
- Sentence Mode
  - declarative vs. interrogative
  - (continuing vs. terminating)

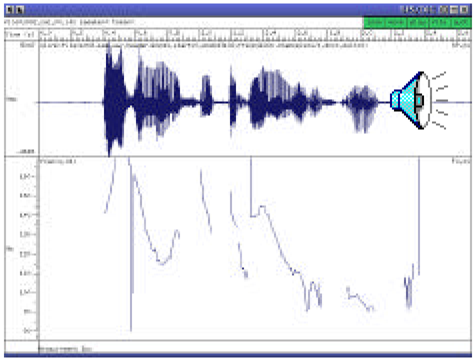
Intonation
Paralinguistic & Nonlinguistic factors

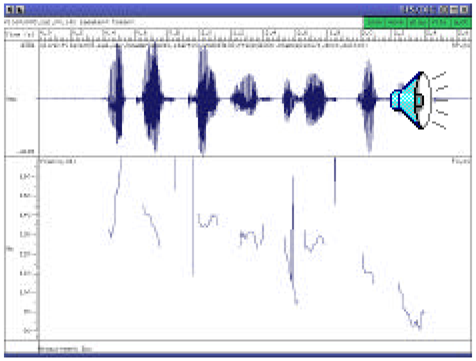
- Speaking style
  - e.g. spontaneous vs. read
  - fairy tail vs. Newsreader
  - social status
- Emotion
  - e.g. aroused vs. bored
- Individual Factors
  - sex
  - age ...

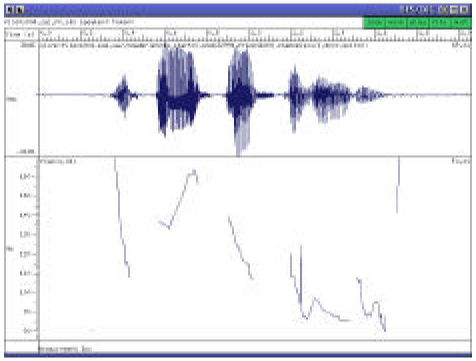
Intonation/Pitch/F0

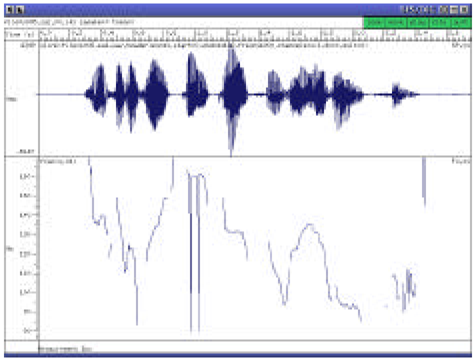
Österreichisches Forschungsinstitut für Artificial Intelligence

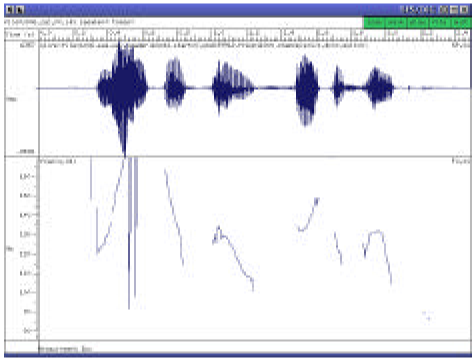










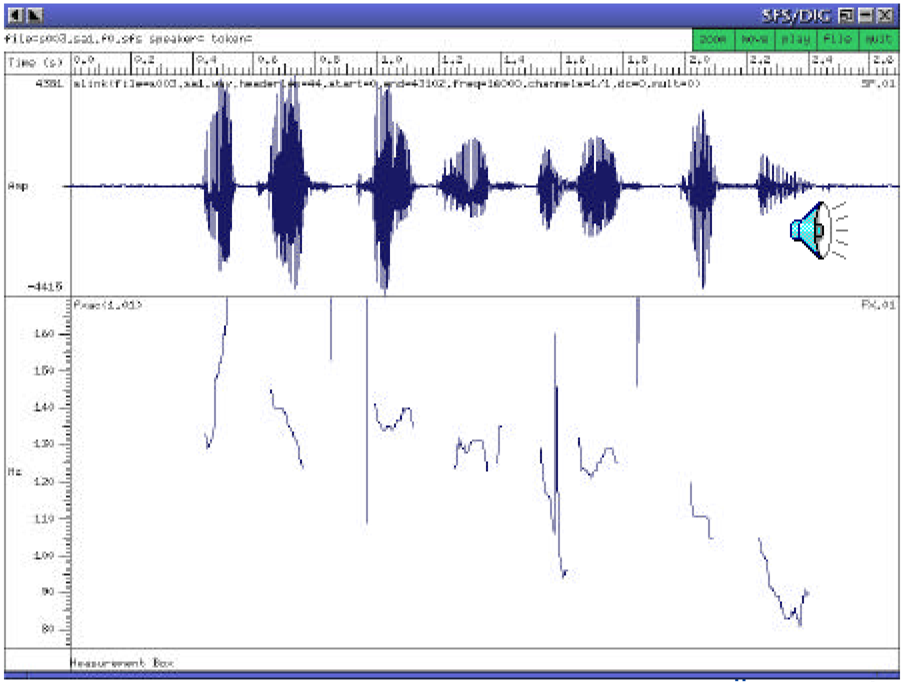


Application of Intonation Models

- 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

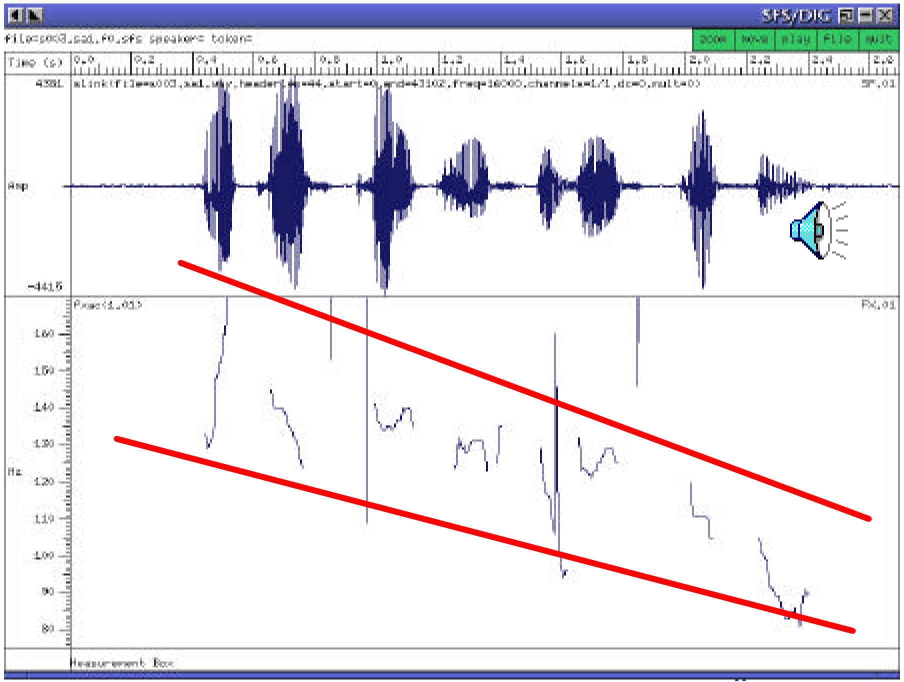
- Microprosodic variaton
  - “dip” in contour at /l/
  - voiced/unvoiced transitions...



Österreichisches Forschungsinstitut für Artificial Intelligence

Properties of F0-contours: Declination

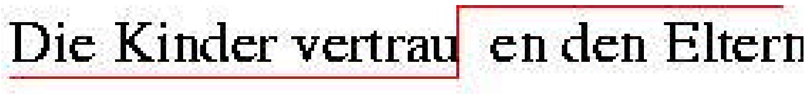
- Overall downtrend
  - of base- and topline.
  - reset at major phrase boundaries



Models of Intonation: Isacenko&Schaedlich 1964

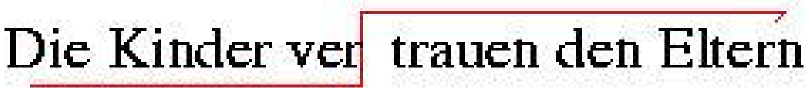
Österreichisches Forschungsinstitut für Artificial Intelligence

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



Die Kinder vertrau en den Eltern

- question



Die Kinder ver trauen den Eltern

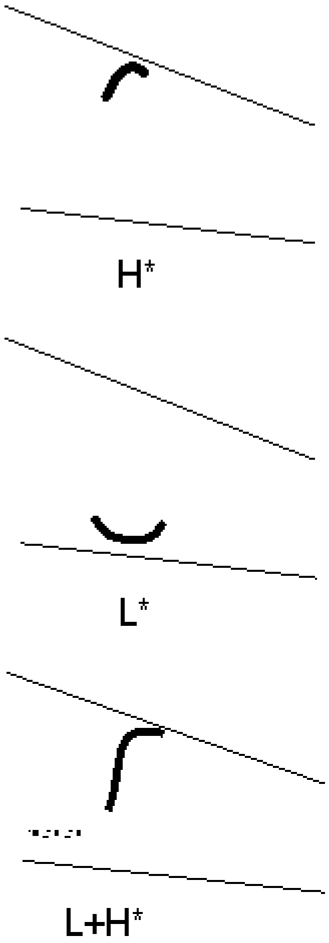
- unfinished

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%

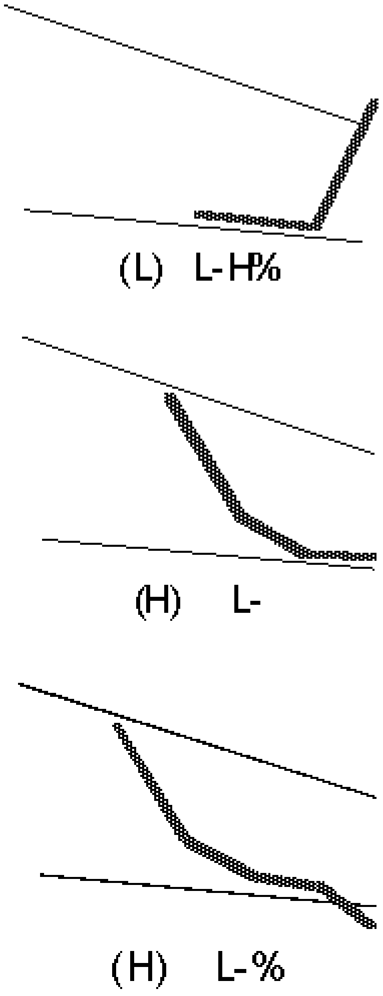
ToBI: Tones and Breaks Indices
Accent Tones

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



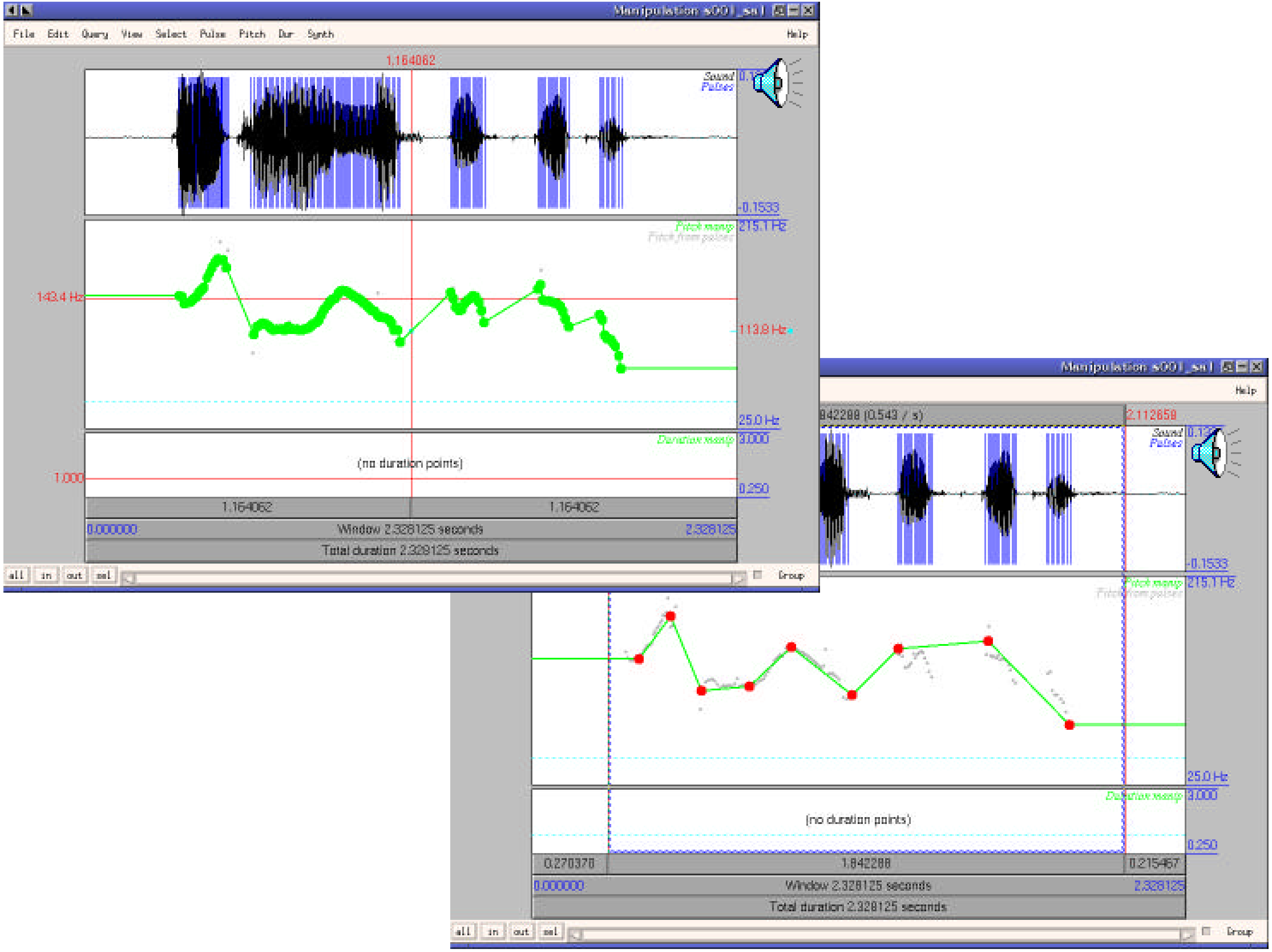
ToBI: Tones and Breaks Indices Boundaries

- 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



Österreichisches Forschungsinstitut für Artificial Intelligence

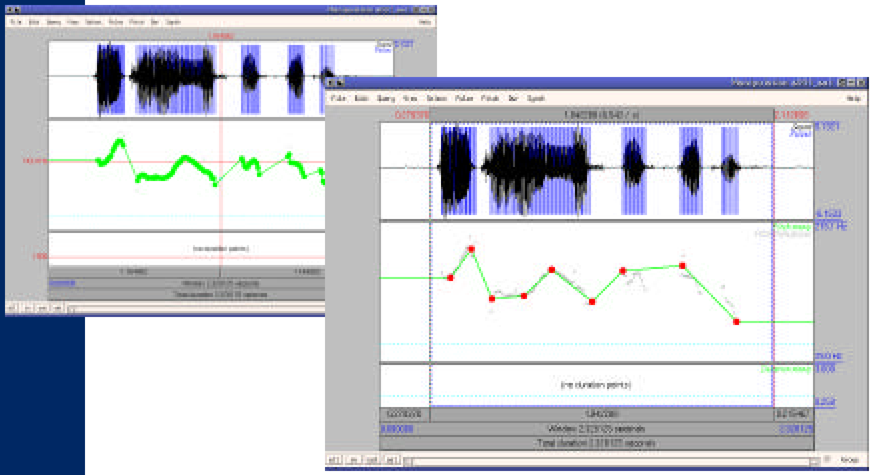
F0 - stylisation: getting rid of microprosdy, flaws, ...



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

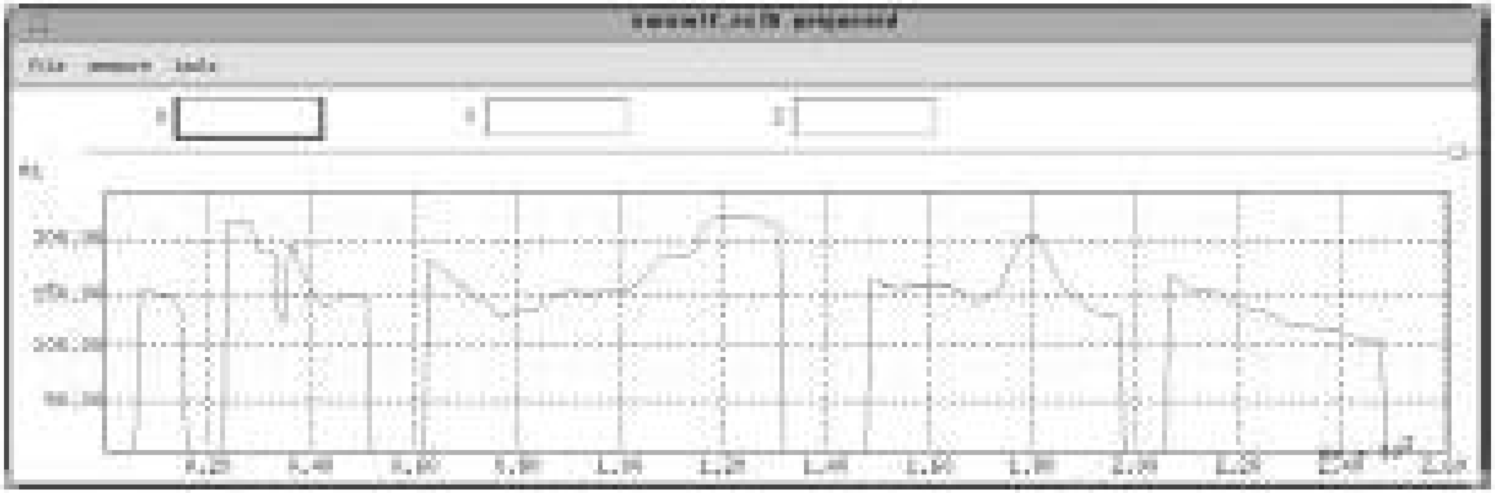
Österreichisches Forschungsinstitut für Artificial Intelligence

- 1. stylise to “perceptually identic”



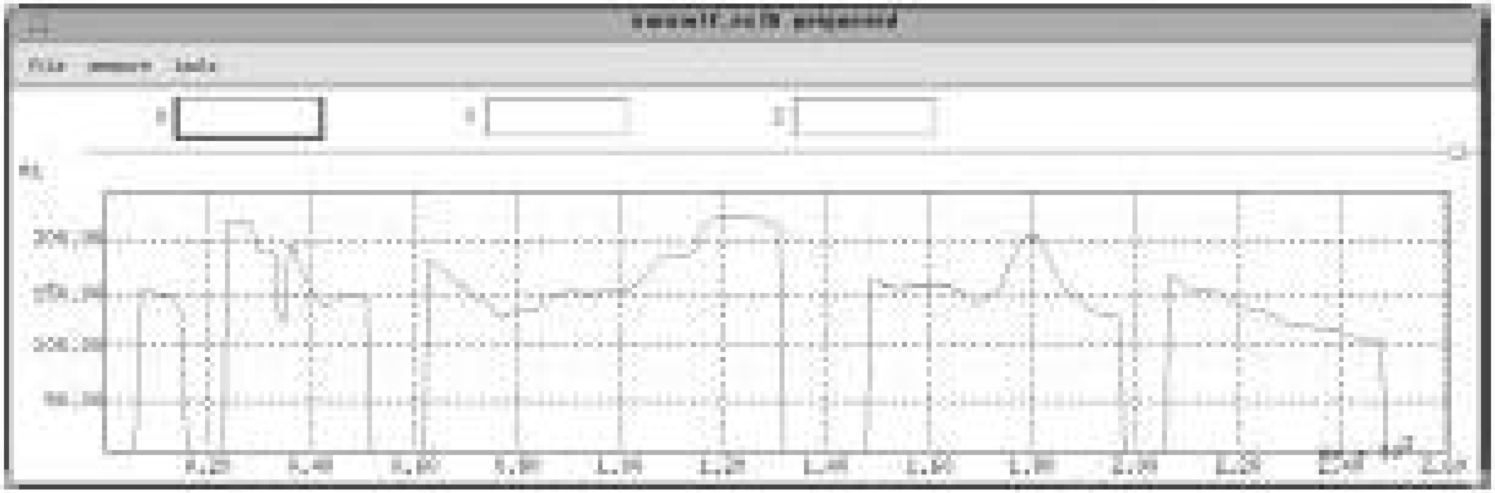
- 2. Functionally identic
- 3. Inventory of 11 accent-lending and phrase-marking movements

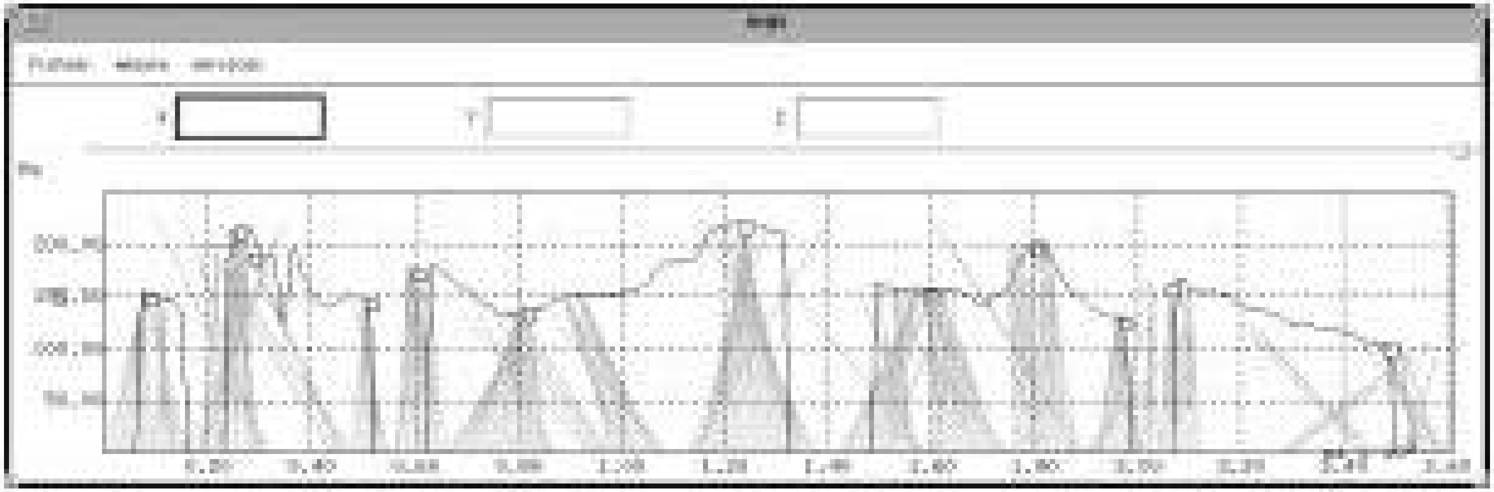
MOMEL
Melodic Modellisation (Hirst 1991)



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

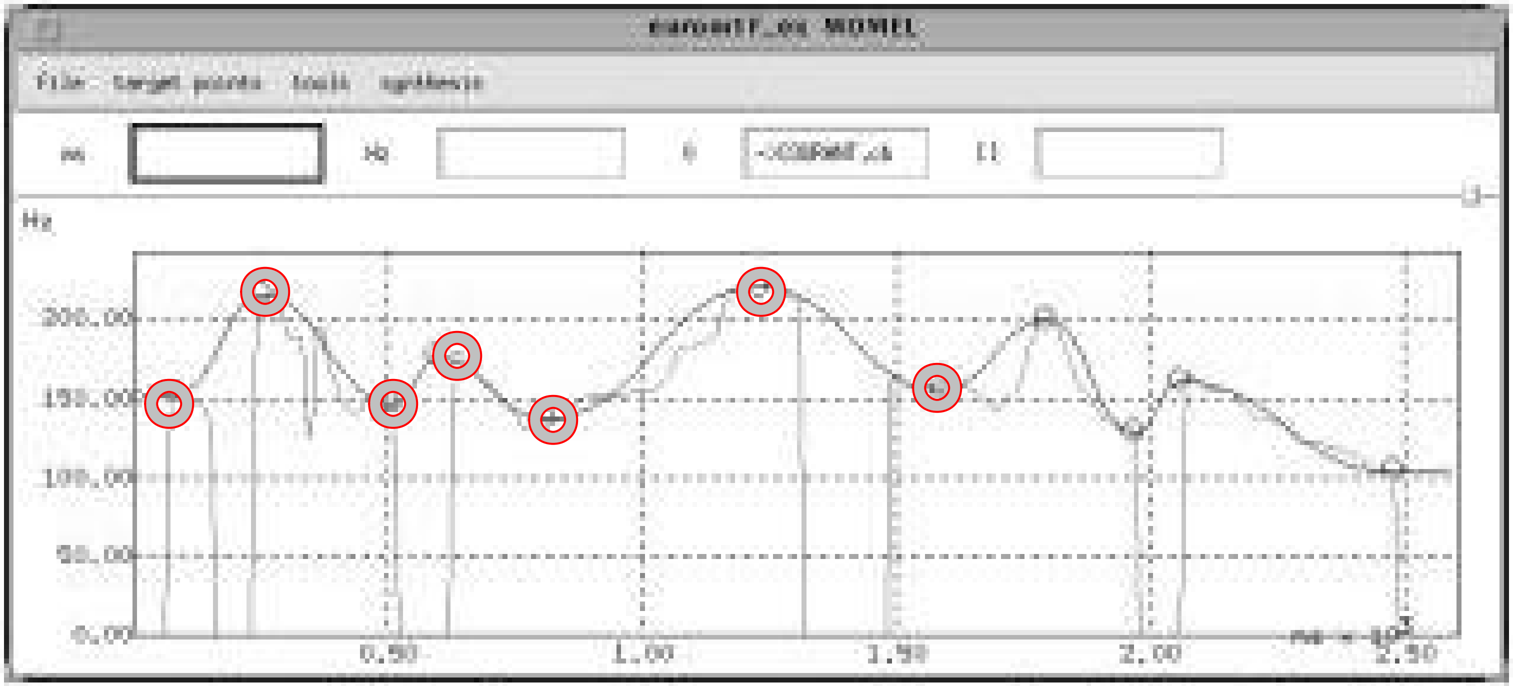
MOMEL
Melodic Modellisation (Hirst 1991)
Quadratic Splines





MOMEL
Melodic Modellisation (Hirst 1991)

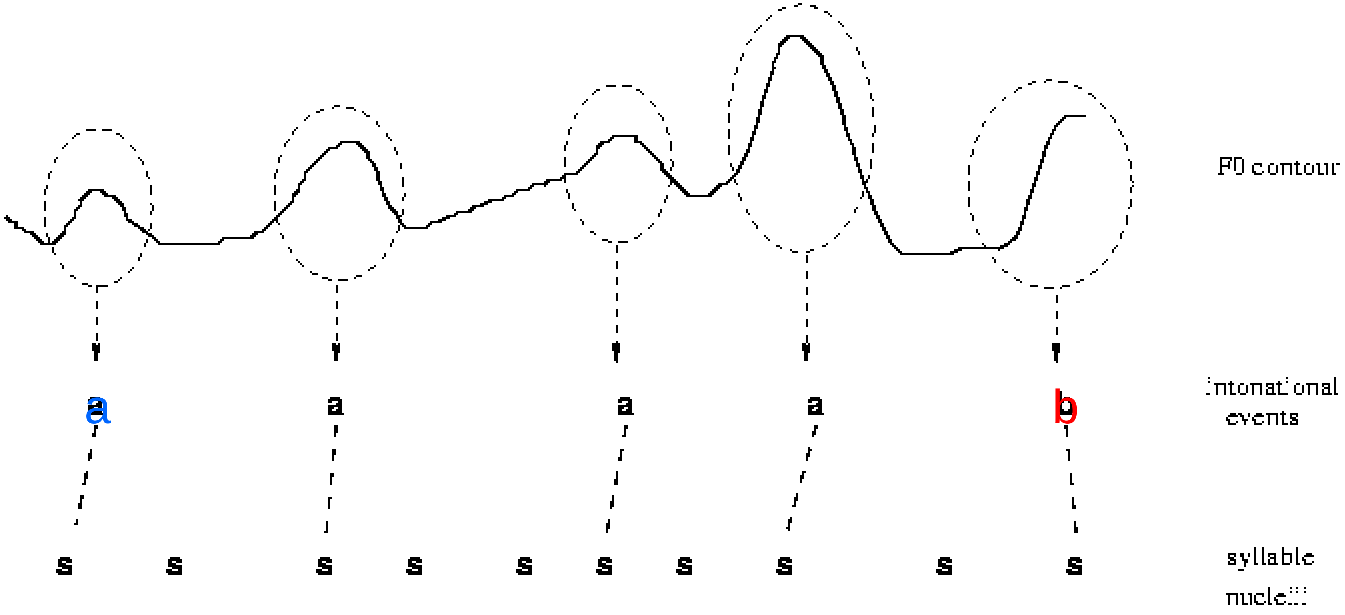
- Freely available
- Valid smoothing method



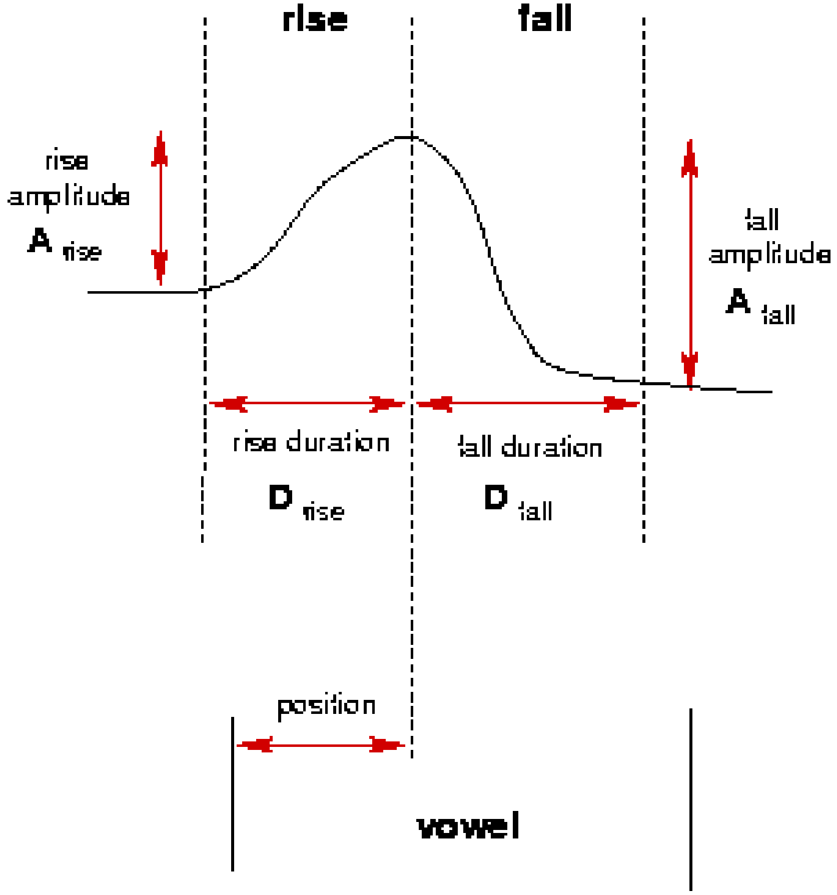
Österreichisches Forschungsinstitut für Artificial Intelligence

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*

Österreichisches Forschungsinstitut für Artificial Intelligence

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:
- $\text{Dur}_{\text{event}}$ (sum of fall and rise)
- $\text{Amplitude}_{\text{event}}$ (sum of fall and rise)
- $\text{Tilt}_{\text{event}}$
- (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

Österreichisches Forschungsinstitut für Artificial Intelligence

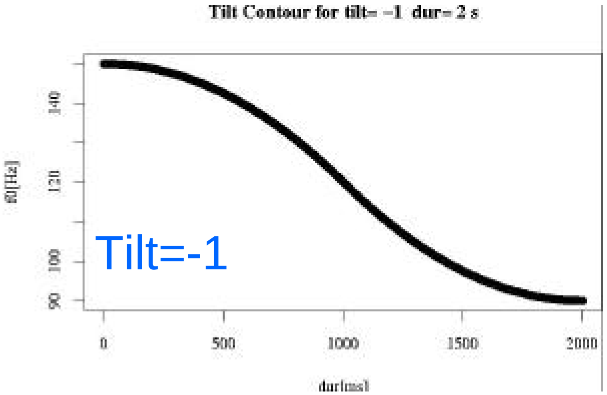
$$tilt_{amp} = \frac{|A_{rise}| - |A_{fall}|}{|A_{rise}| + |A_{fall}|}$$

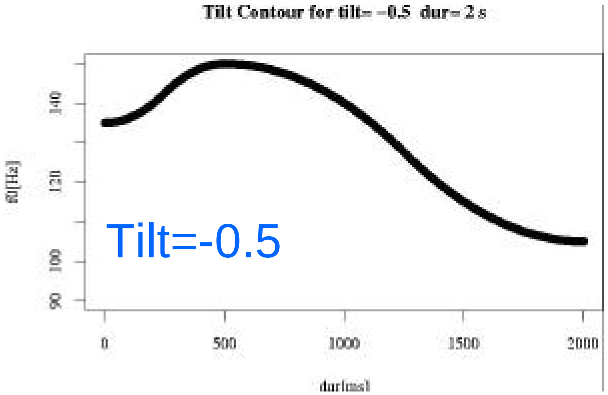
$$tilt_{dur} = \frac{D_{rise} - D_{fall}}{D_{rise} + D_{fall}}$$

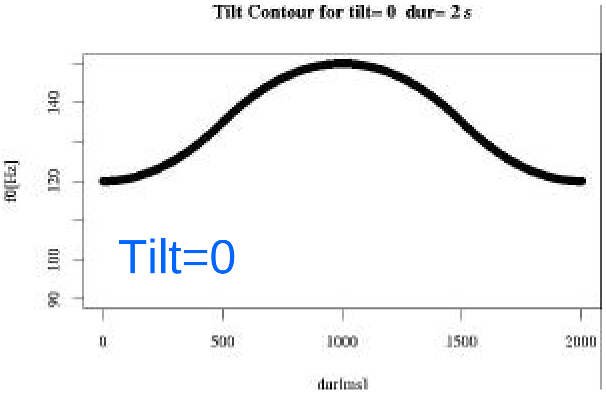
$$\boxed{tilt = \frac{|A_{rise}| - |A_{fall}|}{2(|A_{rise}| + |A_{fall}|)} + \frac{D_{rise} - D_{fall}}{2(D_{rise} + D_{fall})}}$$

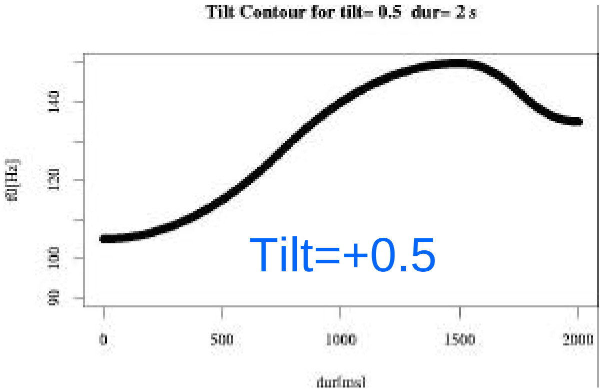
Tilt
Dur=2sec, Amp=60, F0peak=150

Österreichisches Forschungsinstitut für Artificial Intelligence





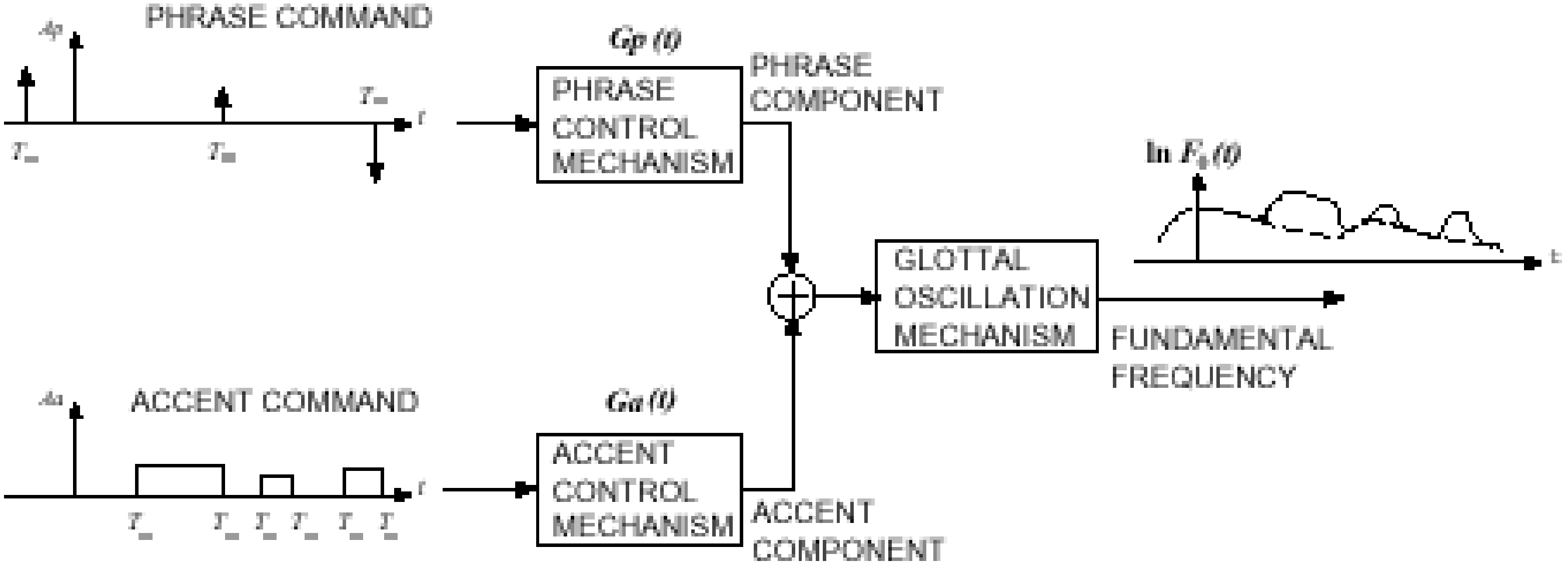




Fujisaki-Model
Hiroya Fujisaki 1984

Österreichisches Forschungsinstitut für Artificial Intelligence

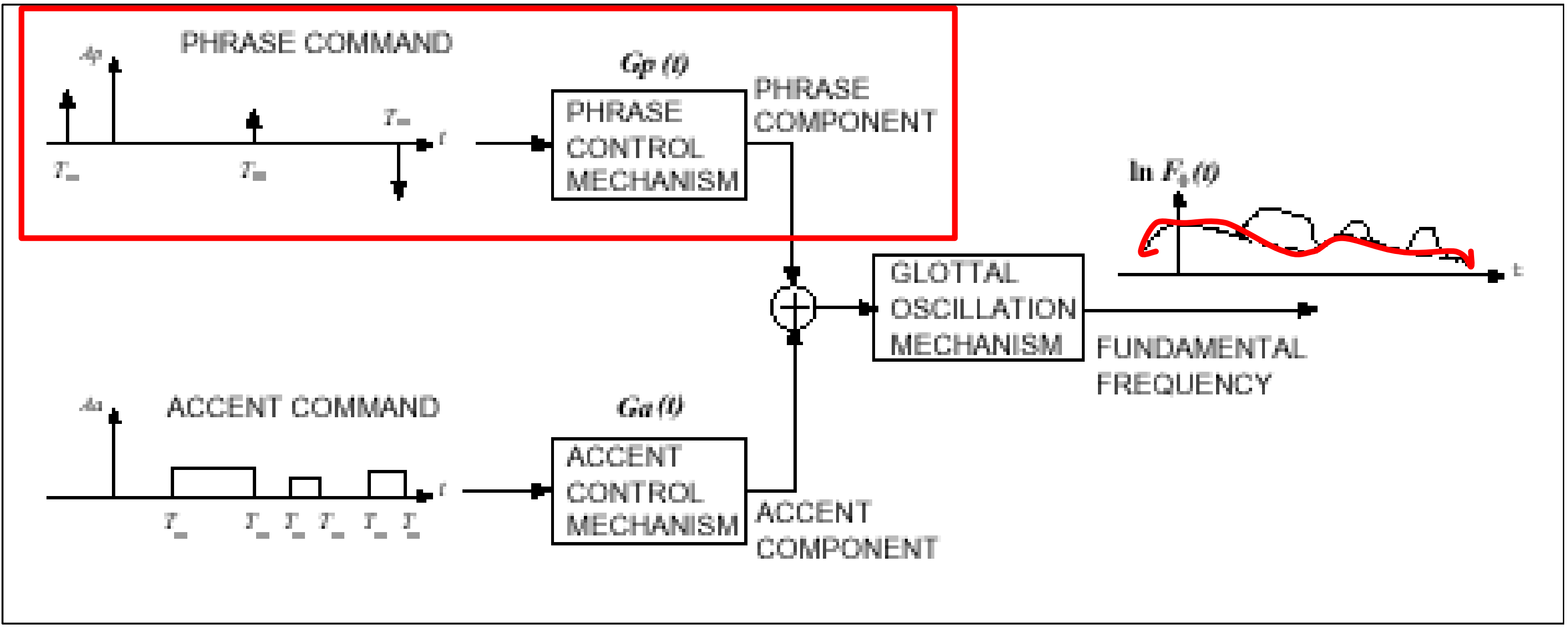
- Superpositional Model: F0 production modelled by 2 separate components



Fujisaki-Model
Hiroya Fujisaki 1984

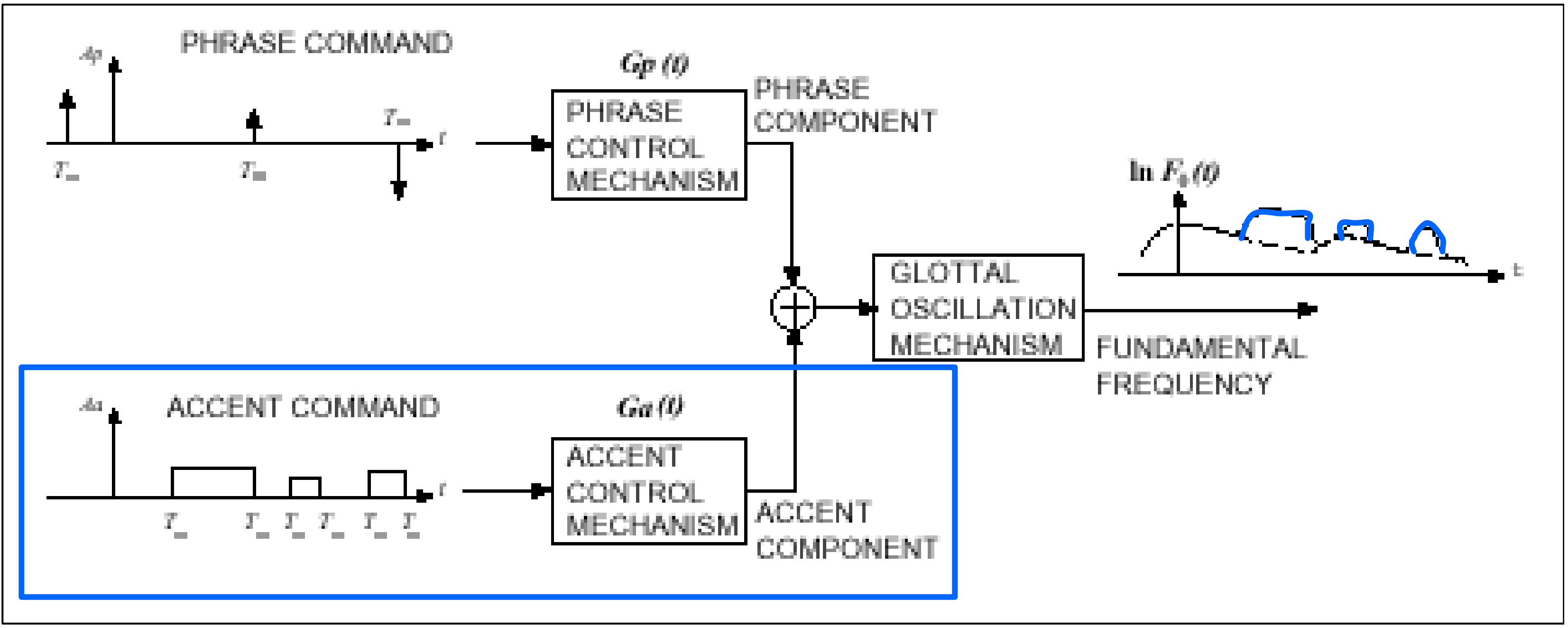
Österreichisches Forschungsinstitut für Artificial Intelligence

- 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



Österreichisches Forschungsinstitut für Artificial Intelligence

Addition in the logarithmic domain

- F0 = Baseline + PhraseComponent + AccentComponent

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

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

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

Österreichisches Forschungsinstitut für Artificial Intelligence

Addition in the logarithmic domain

- F0 = Baseline + PhraseComponent + AccentComponent

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j [Ga(t - T_{1j}) - Ga(t - T_{2j})].$$

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

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

Addition in the logarithmic domain

Österreichisches Forschungsinstitut für Artificial Intelligence

- F0 = Baseline + PhraseComponent + AccentComponent

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j [Ga(t - T_{1j}) - Ga(t - T_{2j})].$$

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

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

Addition in the logarithmic domain

Österreichisches Forschungsinstitut für Artificial Intelligence

$$\ln F_o(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Cp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j [Ca(t - T_{1j}) - Ca(t - T_{2j})].$$

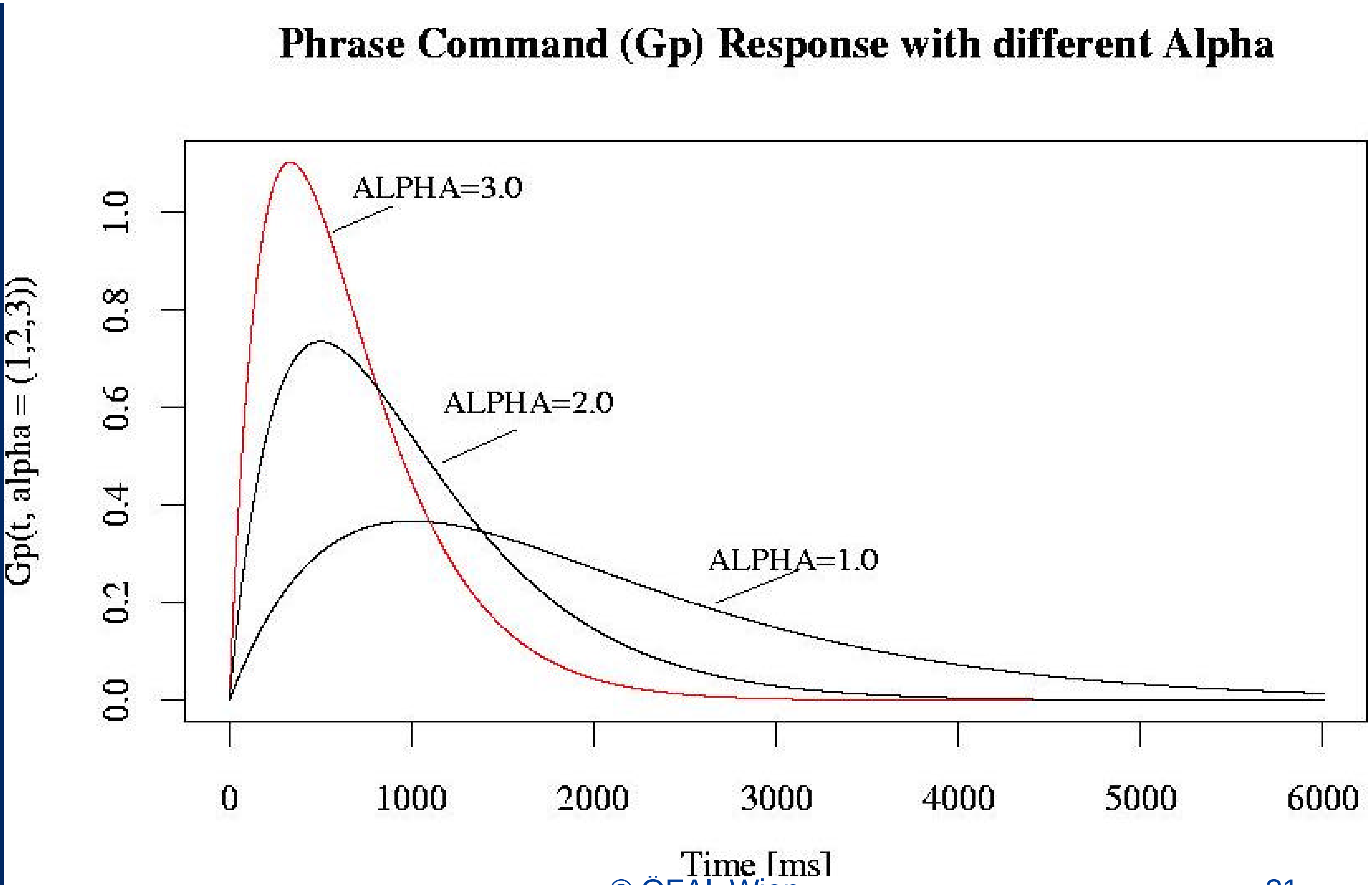
$$Cp(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

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

Phrase Component

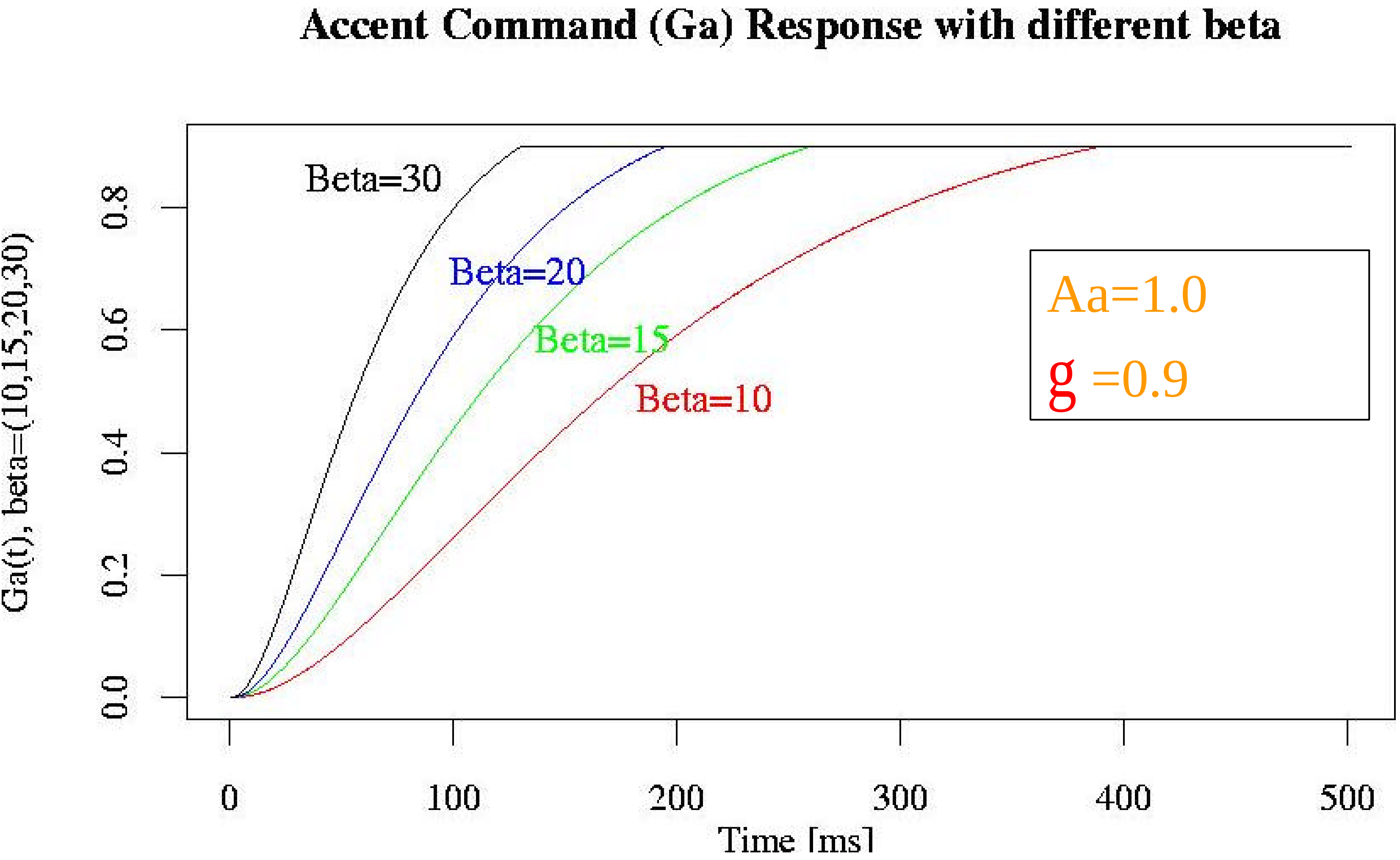
Gp(t) = $a^{2} t$ exp(- $a t$))

Österreichisches Forschungsinstitut für Artificial Intelligence



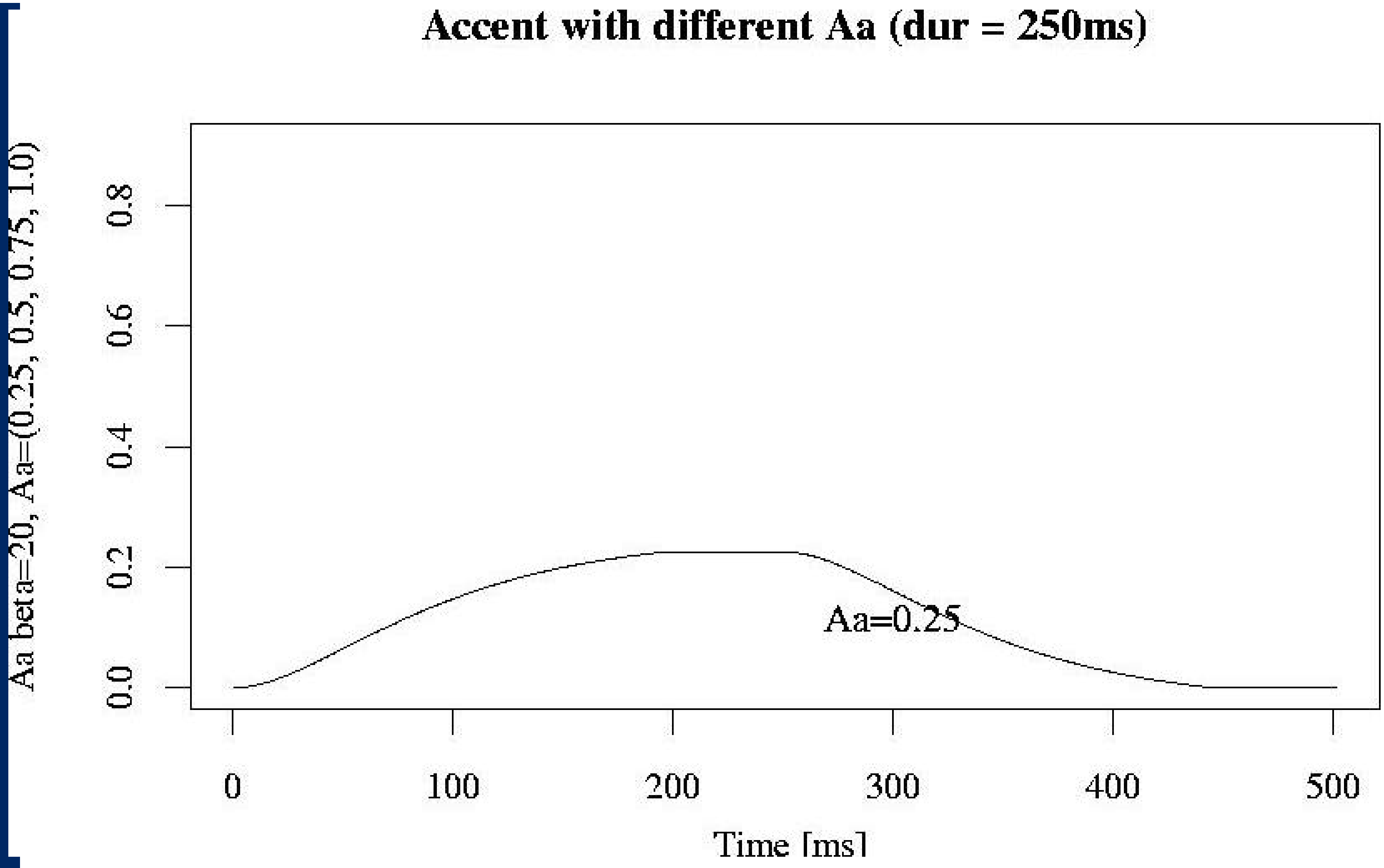
Accent Command: Ga(t)=min[1 - (1+bt) * exp(-bt), g]

Österreichisches Forschungsinstitut für Artificial Intelligence

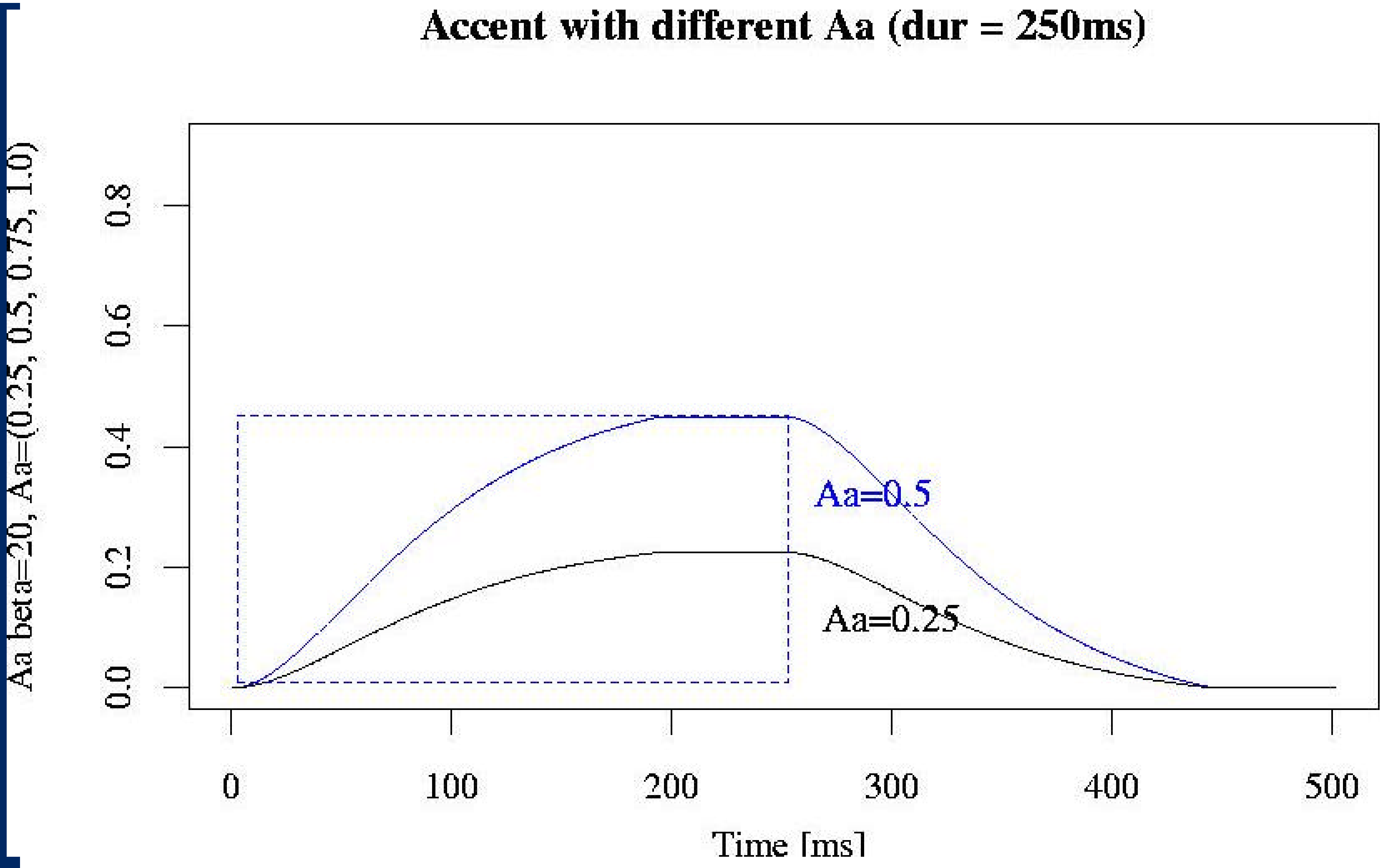


Accent form with different Amplitude Aa

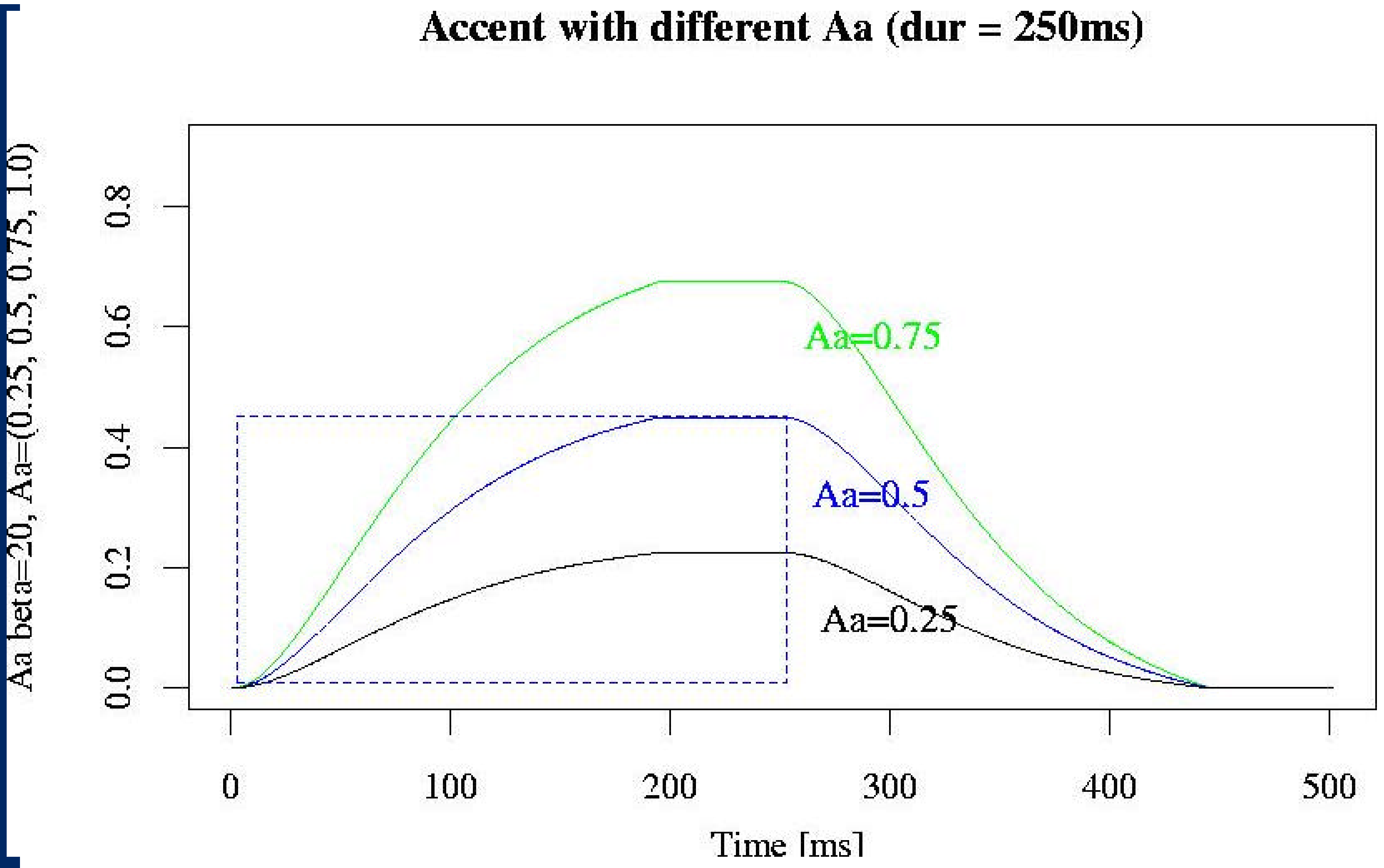
Österreichisches Forschnungsinstitut für Artificial Intelligence



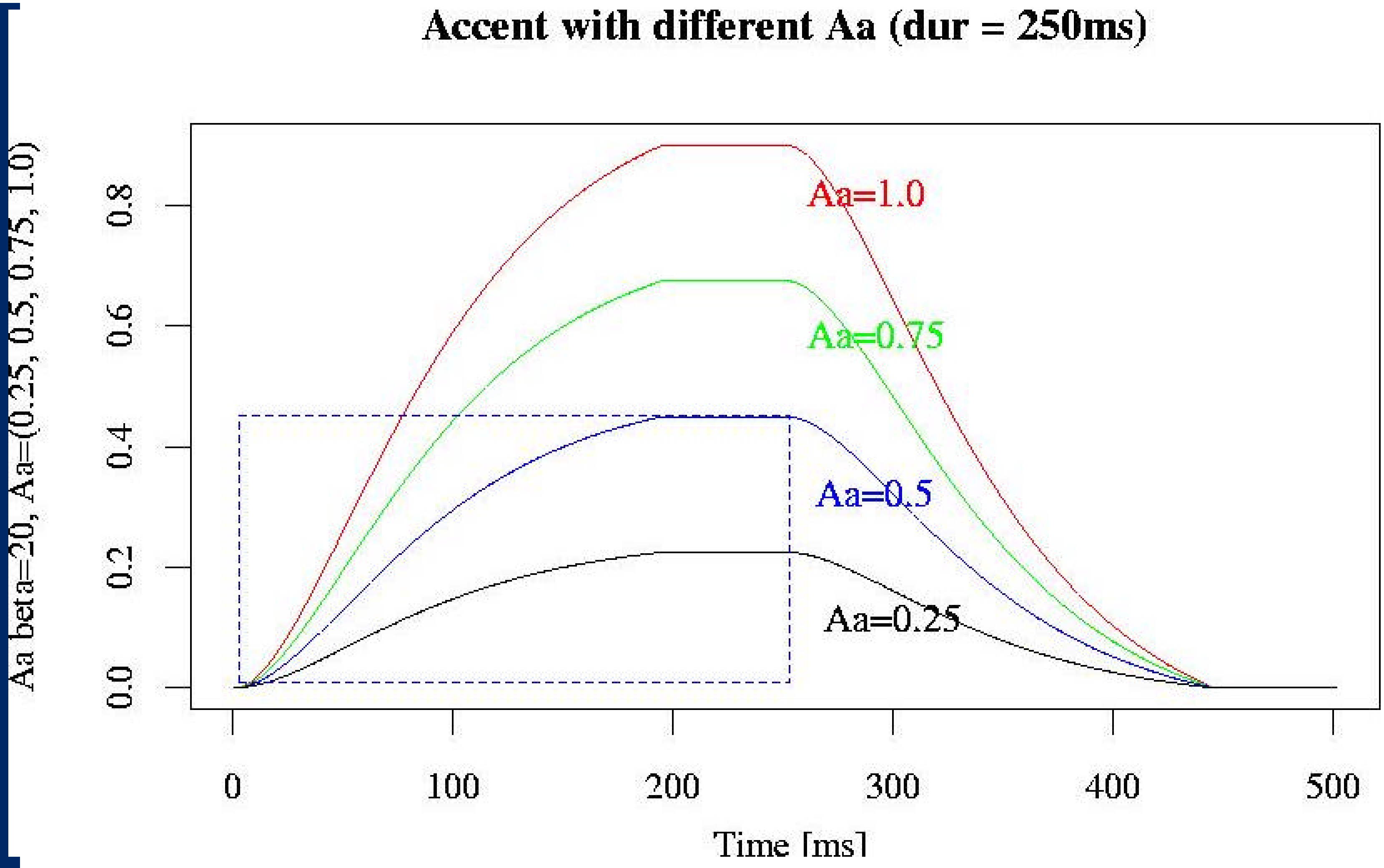
Accent form with different Amplitude Aa



Accent form with different Amplitude Aa



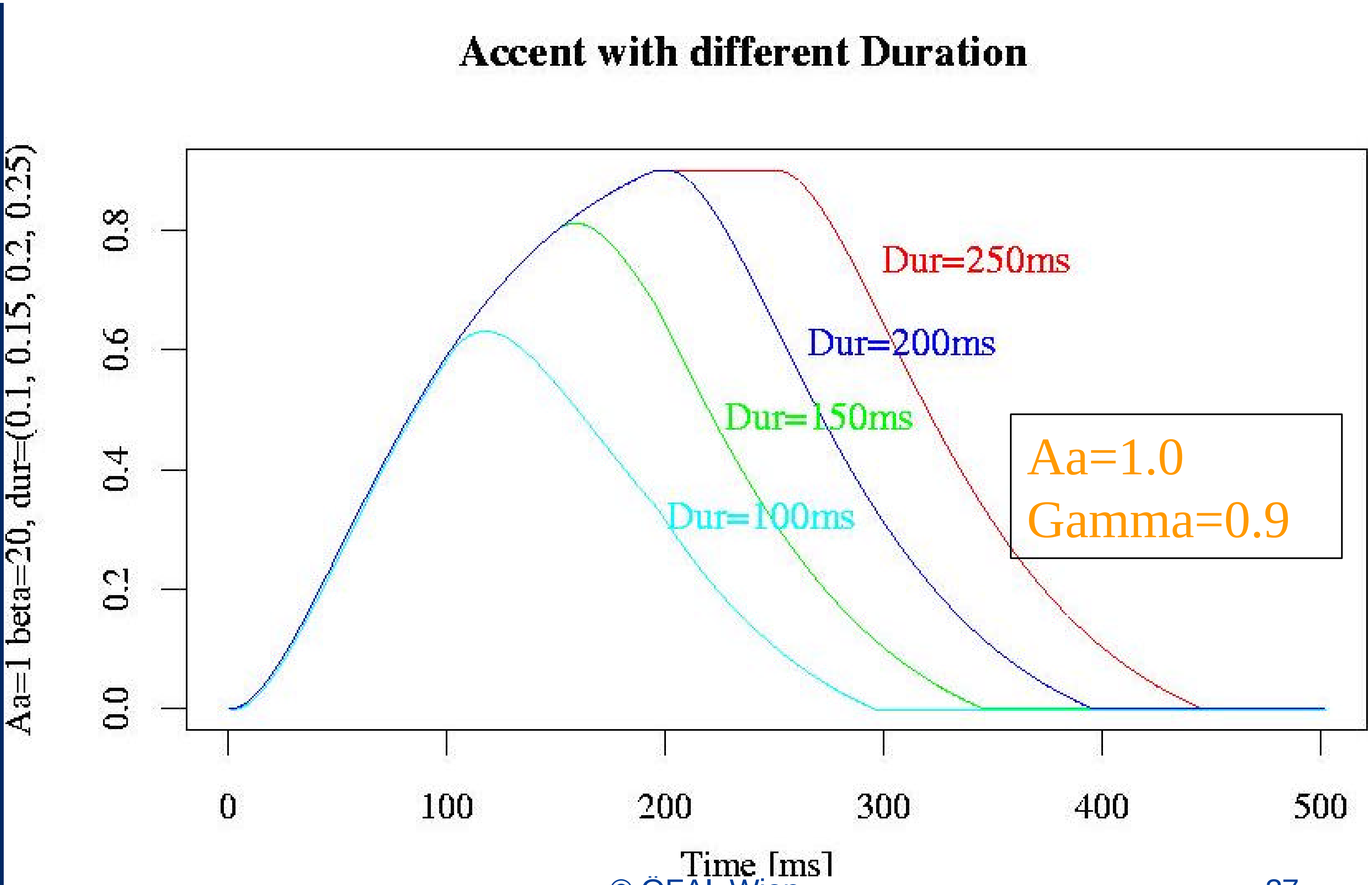
Accent form with different Amplitude Aa



Österreichisches Forschungsinstitut für Artificial Intelligence

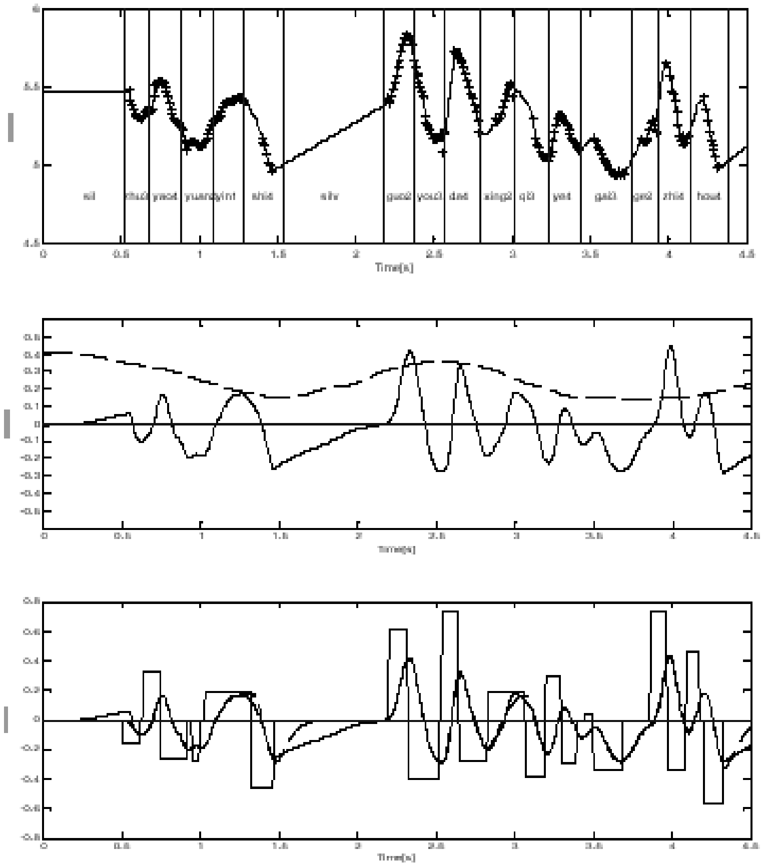
Accent form with different Duration

Österreichisches Forschungsinstitut für Artificial Intelligence



How to extract Phrase- and Accent-Commands

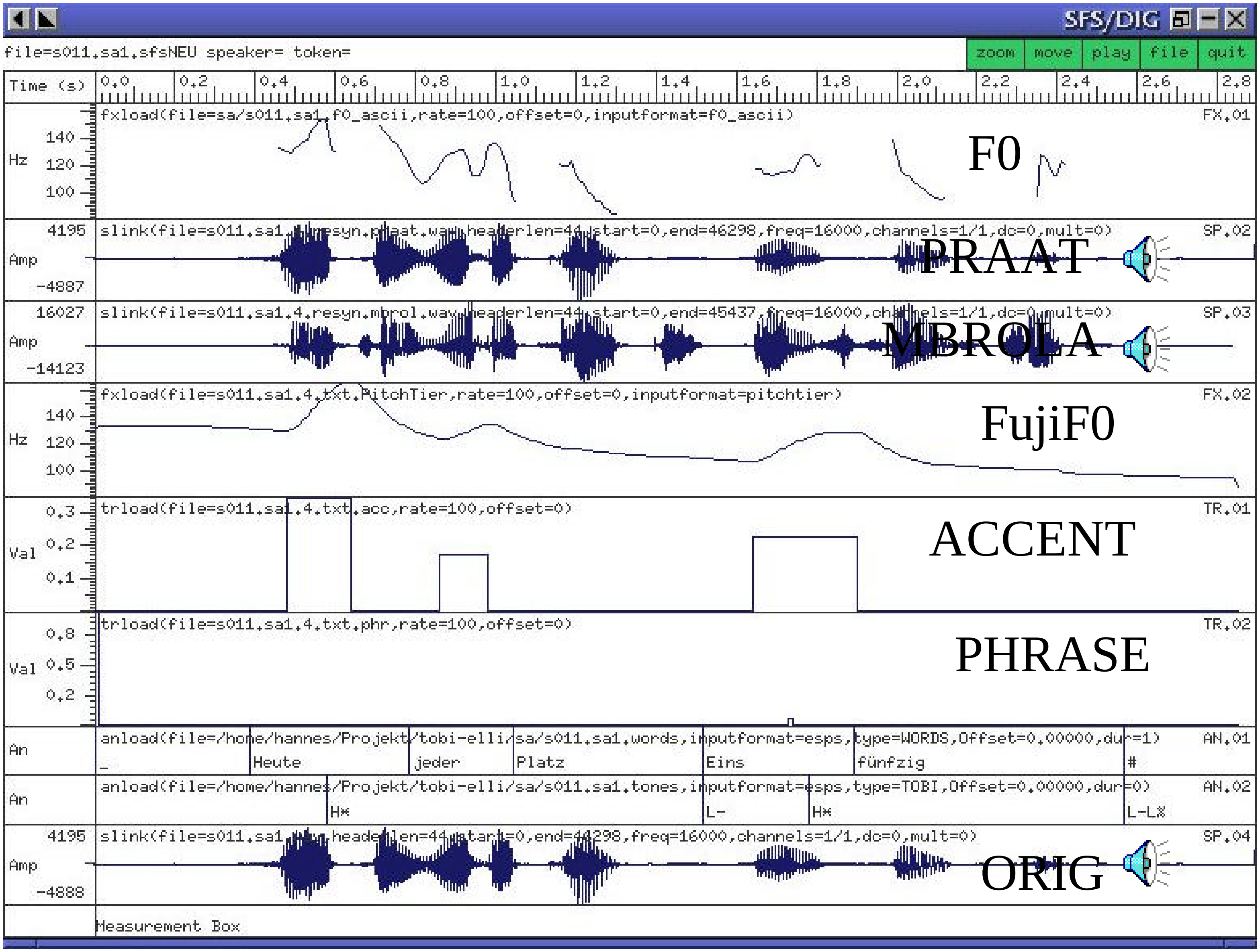
Österreichisches Forschungsinstitut für Artificial Intelligence



- 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

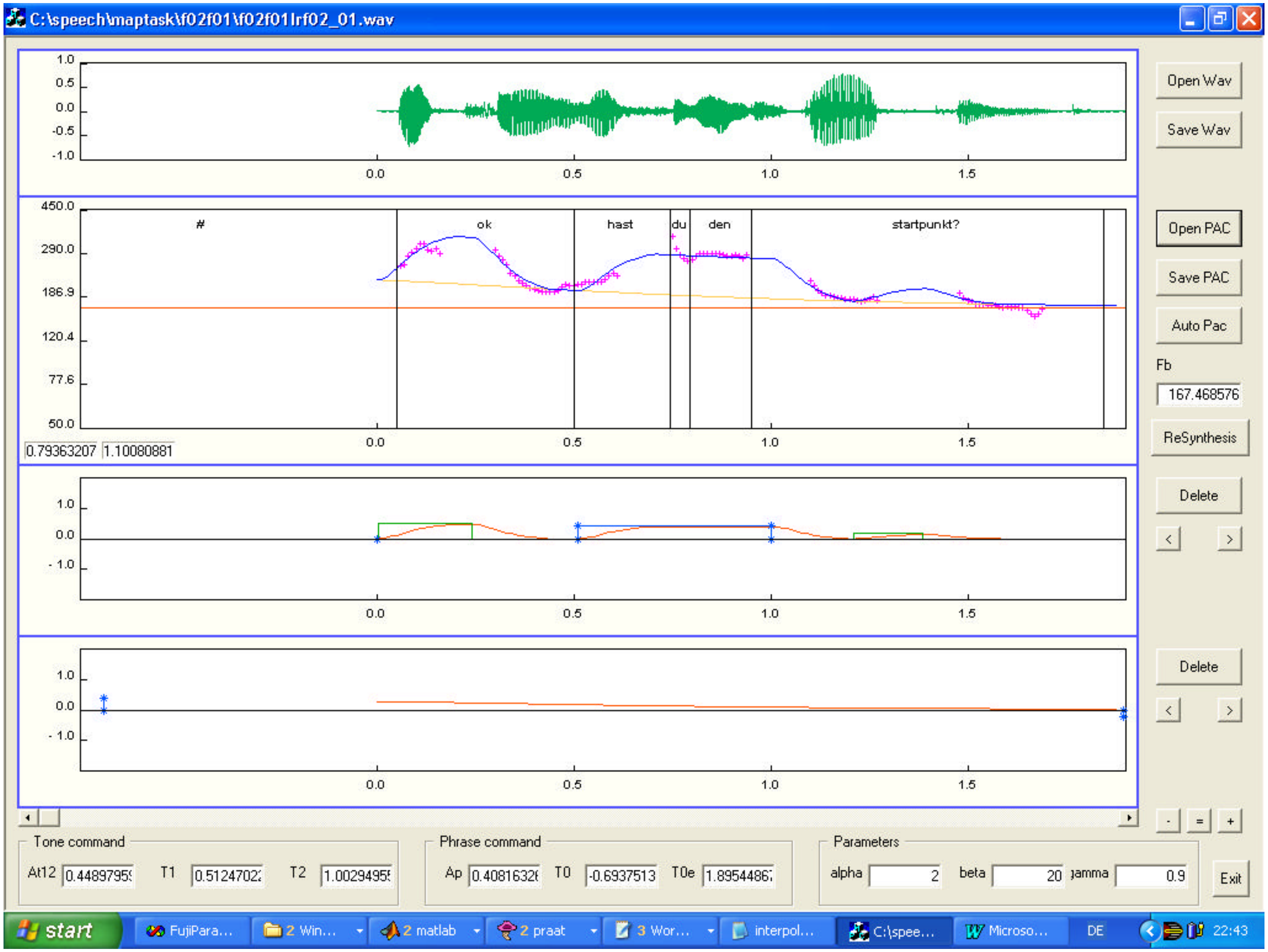
Example



```
file=s011.sa1.sfsNEU speaker= token=
fxload(file=sa/s011.sa1.f0_ascii,rate=100,offset=0,inputformat=f0_ascii)                FX.01
slink(file=s011.sa1.4.resyn.praat.wav,headerlen=44,start=0,end=46298,freq=16000,channels=1/1,dc=0,mult=0)    SP.02
slink(file=s011.sa1.4.resyn.mbrol.wav,headerlen=44,start=0,end=45437,freq=16000,channels=1/1,dc=0,mult=0)    SP.03
fxload(file=s011.sa1.4.txt.PitchTier,rate=100,offset=0,inputformat=pitchtier)          FX.02
trload(file=s011.sa1.4.txt.acc,rate=100,offset=0)                                       TR.01
trload(file=s011.sa1.4.txt.phr,rate=100,offset=0)                                       TR.02
anload(file=/home/hannes/Projekt/tobi-elli/sa/s011.sa1.words,inputformat=esps,type=WORDS,Offset=0.00000,dur=1)    AN.01
_   Heute   jeder   Platz   Eins   fünfzig   #
anload(file=/home/hannes/Projekt/tobi-elli/sa/s011.sa1.tones,inputformat=esps,type=TOBI,Offset=0.00000,dur=0)     AN.02
H*   L-   H*   L-L%
slink(file=s011.sa1.wav,headerlen=44,start=0,end=44298,freq=16000,channels=1/1,dc=0,mult=0)    SP.04
Measurement Box
```

Österreichisches Forschungsinstitut für Artificial Intelligence

FujiParamEditor



Fb 167.468576

Tone command

At12 0.44897959 T1 0.51247022 T2 1.00294955

Phrase command

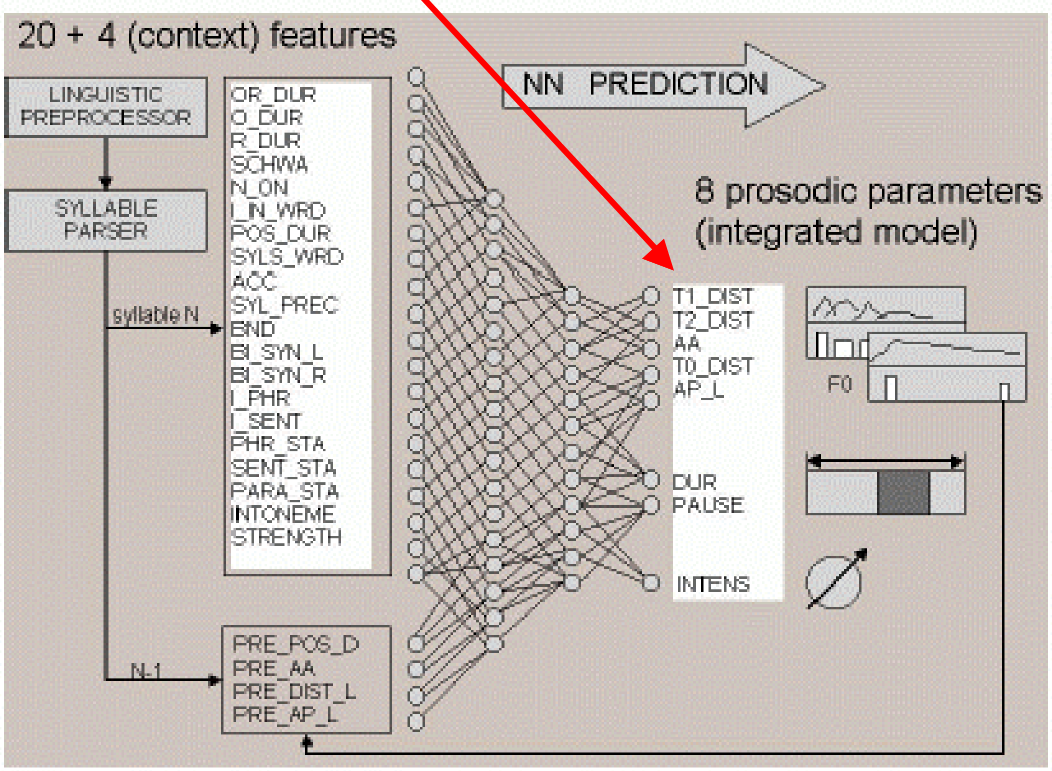
Ap 0.40816326 T0 -0.6937513 T0e 1.89544867

Parameters

alpha 2 beta 20 gamma 0.9

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.oefai.at/~hannes

Resources, Literature etc.

Österreichisches Forschungsinstitut für Artificial Intelligence

- 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/