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

• Emphasis
  • (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
Application of Intonation Models

• Speech Synthesis
  • How to map linguistic function to intonation contours?
  • Aim for adequacy and naturalness

• Speech Recognition
  • Spot accents, focus structure, sentence mode...
  • Analyse paralinguistic factors
Properties of F0-contours

- Microprosodic variation
- "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

\[ \text{H}^* \]
\[ \text{L}^* \]
\[ \text{L+H}^* \]
ToBI: Tones and Breaks Indices

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

E.g.
- L-L% : typical final fall in declarative sentences
- H-H%: typical rise in questions
F0 - stylisation: getting rid of microprosody, 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
• 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
  - $A_{\text{rise}}$
  - $A_{\text{fall}}$

- Duration
  - $D_{\text{rise}}$
  - $D_{\text{fall}}$

- "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 symmetrical
Tilt – model

Reduction of necessary 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:
- \( \text{tilt} = \frac{\text{tiltAmp} + \text{tiltDur}}{2} \)

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

- Tilt=-1
- Tilt=-0.5
- Tilt=0
- Tilt=+0.5
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

- $F_0 = \text{Baseline} + \text{PhraseComponent} + \text{AccentComponent}$

\[
\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_p_i C_p(t-T_{bi}) + \sum_{j=1}^{J} A_a_j [C_a(t-T_{1j}) - C_a(t-T_{2j})].
\]

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

\[
C_a(t) = \begin{cases} 
\min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \text{for } t \geq 0, \\
0, & \text{for } t < 0. 
\end{cases}
\]
Addition in the logarithmic domain

- \( F_0 = \text{Baseline} + \text{PhraseComponent} + \text{AccentComponent} \)

\[
\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_{pi} C_p(t-T_{di}) + \sum_{j=1}^{J} A_{aj} [C_a(t-T_{1j}) - C_a(t-T_{2j})].
\]

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

\[
C_a(t) = \begin{cases} 
\min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \text{for } t \geq 0, \\
0, & \text{for } t < 0.
\end{cases}
\]
Addition in the logarithmic domain

- \( F_0 = \text{Baseline} + \text{PhraseComponent} + \text{AccentComponent} \)

\[
\ln F_0(t) = \ln F_b + \sum_{i=1}^I A_p_i C_p(t-T_{b,i}) + \sum_{j=1}^J A_{a,j}[C_a(t-T_{1,j}) - C_a(t-T_{2,j})].
\]

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

\[
C_a(t) = \begin{cases} 
\min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \text{for } t \geq 0, \\
0, & \text{for } t < 0.
\end{cases}
\]
Addition in the logarithmic domain

\[
\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_p C_p(t-T_{b_i}) + \sum_{j=1}^{J} A a_j [C a(t-T_{1,j}) - C a(t-T_{2,j})].
\]

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

\[
C a(t) = \begin{cases} 
\min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \text{for } t \geq 0, \\
0, & \text{for } t < 0.
\end{cases}
\]
Phrase Component
\[ G_p(t) = \alpha^2 t \exp(- \alpha t) \]
Accent Command:
\[ Ga(t) = \min[1 - (1 + \beta t) \cdot \exp(-\beta t), \gamma] \]

Accent Command (Ga) Response with different beta

\(Aa=1.0\)
\(\gamma=0.9\)
Accent form with different Amplitude Aa

Accent with different Aa (dur = 250ms)

Aa = 0.25
Accent form with different Amplitude $A_a$

Accent with different $A_a$ ($\text{dur} = 250\text{ms}$)

\[ A_a = 0.25, 0.5, 0.75, 1.0 \]

\[ A_a = 0.5 \]

\[ A_a = 0.25 \]

Time [ms]
Accent form with different Amplitude $A_a$

Accent with different $A_a$ (dur = 250ms)
Accent form with different Amplitude Aa

Accent with different Aa (dur = 250ms)
Accent form with different Duration

Accent with different Duration

Aa=1.0
Gamma=0.9

Aa=1.0, beta=20, dur=(0.1, 0.15, 0.2, 0.25)

Dur=250ms
Dur=200ms
Dur=150ms
Dur=100ms
How to extract Phrase- and Accent-Commands

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

F0

PRAAT

MBROLA

FujiF0

ACCENT

PHRASE

ORIG
FujiParamEditor
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](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](http://www.tfh-berlin.de/~mixdorff/Research.htm)
- A praat implementation for MOMEL [http://www.icp.inpg.fr/~loeven/Praat/momel_english.html](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/](http://festvox.org/docs/speech_tools-1.2.0/)