Phrasing

- ☐ Chunking utterances into breath sized pieces
- ☐ First approximation: punctuation
  – too little
- ☐ Second at content/function words
  – too much

Next week, some inmates released early from the Hampton County jail in Springfield, will be wearing a wristband that hooks up with a special jack on their home phones.

Next week | some inmates released early | from the Hampton County jail | in Springfield | will be wearing | a wristband | that hooks | up with a special jack | on their home phones.

Phrasing

- Banchenko and Fitzpatrick 90:
  - rule driven with punc, POS and syntax
  - balanced phrasing
  - (the boy saw) (the girl in the park)
  - (the boy in the park) (saw the girl)
- Hirschberg and Prieto 94:
  - CART trees
  - 95% for Spanish
- Ostendorf and Veilleux 94:
  - hierarchical statistical model
  - Multilevel breaks.

Taylor and Black 97

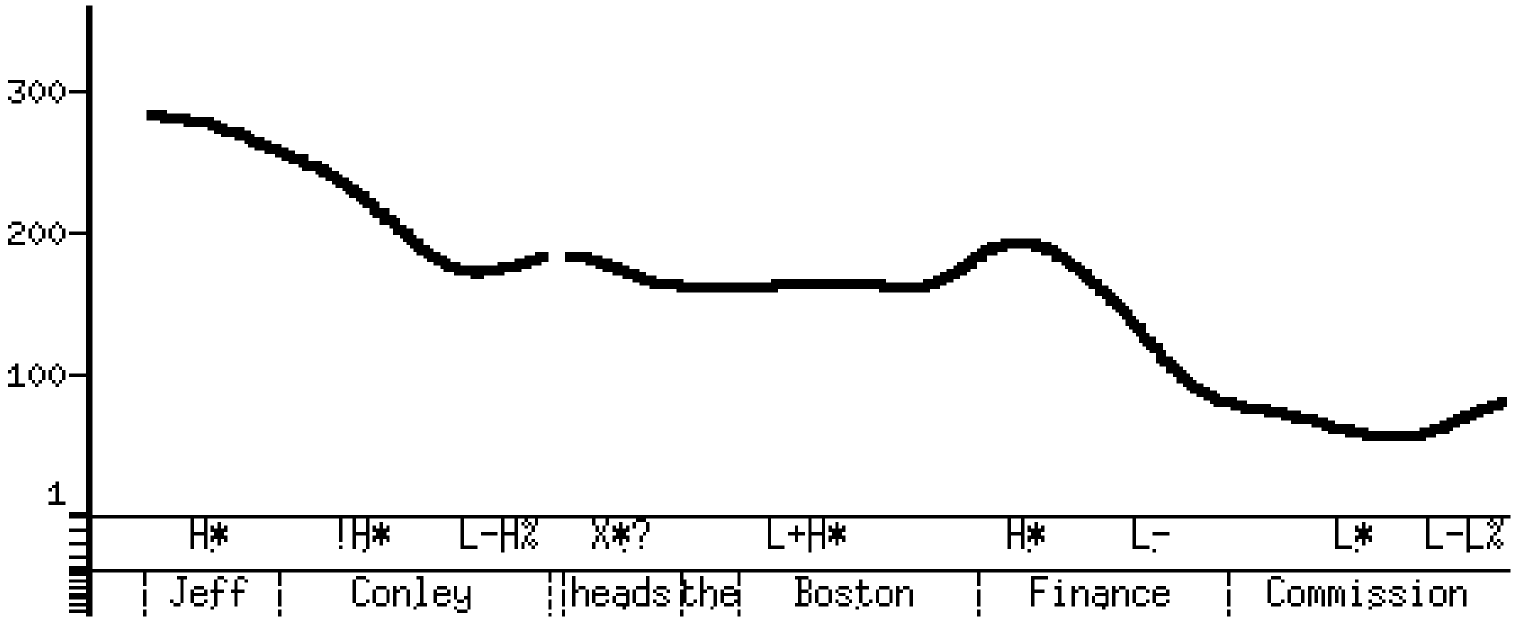
- Keeping balanced phrases
- two part:
  - Predict prob of break at point by CART
  - Base choice on previous break/non-break selections
- $$\prod_{k=1}^{n} \frac{P(B_k \mid B_{k-1}, \ldots, B_{k-N+1}) P(T_{k-N,\ldots,k+1} \mid B_k)}{P(T_{k-N,\ldots,k+1})}$$
- Trained on BBC Radio 4 (like NPR)
  - 31,707 words, 6,346 breaks
  - 91% correct with 6-gram

Correctness

When is phrasing correct/wrong?

- □ Multiple *acceptable* phrasing exists
- □ Not all possible phrasing is acceptable:
  – but possible in some context
- □ Ostendorf and Veilleux 94:
  – same utterance by multiple speakers
  – if predicted utt matches any speaker its correct
- □ Some choices are arbitrary, some not

Intonation


300
200
100
1
H*
!H*
L-H%
X*?
L+H*
H*
L-
L*
L-L%
Jeff
Conley
heads
the
Boston
Finance
Commission

Intonation

- Predict:
  - accents, boundary tones
  - F0 contour
- More theories than researchers

Intonation Examples

- ☐ Fixed durations, flat F0.
- ☐ Decline F0
- ☐ "hat" accents on stressed syllables
- ☐ accents and end tones
- ☐ statistically trained

Theory neutral model

(not really neutral)

- ☐ Where do accents go?
- ☐ Where do boundaries go?
- ☐ What shape are they?
- ☐ What size, length, position are they

Where do accents go?

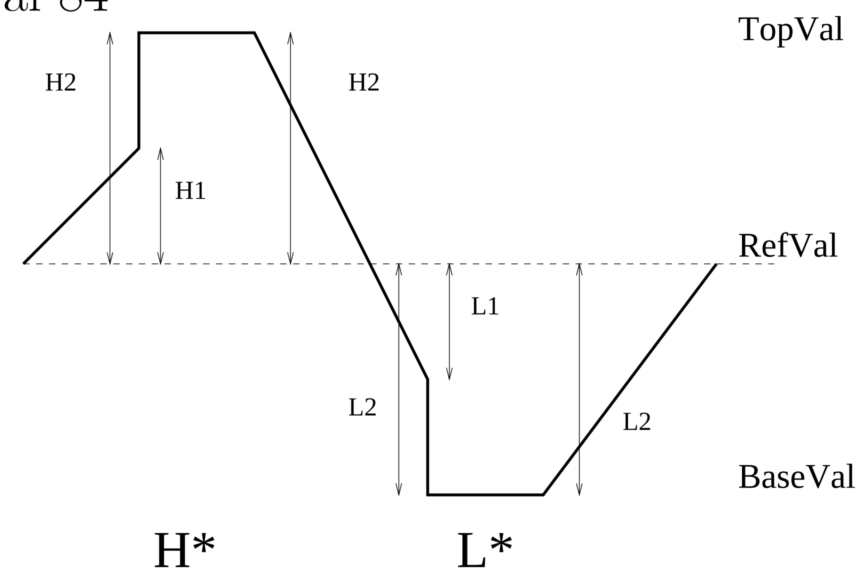
- On the important words.
- First approximation:
  - on stressed syllables in content words
  - 80% correct
- Hirschberg 92
  - hand written rules
  - compound/proper noun
  - phrase position etc
- Festival
  - uses CART on "Hirschberg" features

What shape are they?

- ToBI (Silverman et al 92):
  - Tones and Break Indices
  - Labelling standard *not* computational model
- 6 basic accent types:
  - H*, !H, L+H*, L*, L*+H
- 4 basic end tone types:
  - L-L%, L-H%, H-H%, H-L%
- Break level
  - 1, 2, 3, 4 (larger is bigger break)
- Disadvantages
  - no autolabeller
  - no F0 generator (but ...)
  - Why 6 ?

ToBI F0 generation

- Anderson et al 84



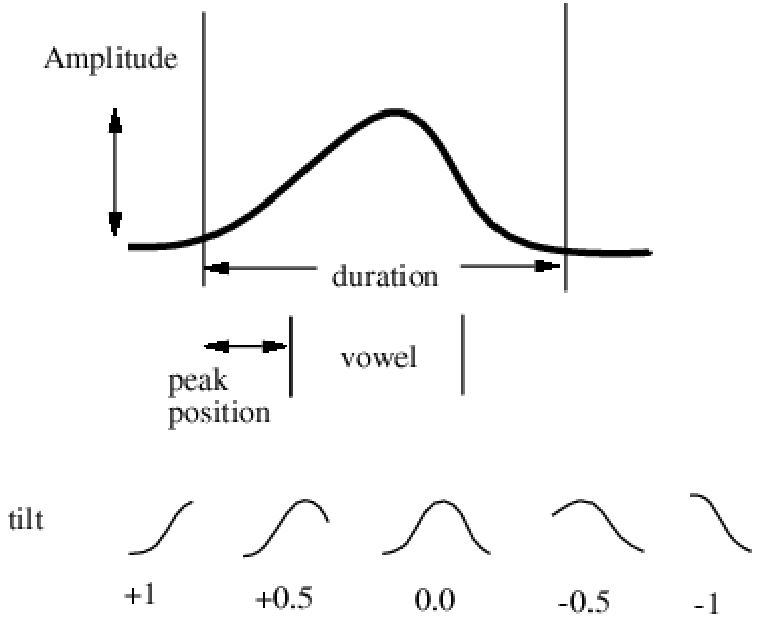
- Three point model (Black and Hunt 96)
  - Linear regression
  - predict start, mid vowel, end on syls
  - smooth result

From the other end ...

- Data driven approach
  - Build models from F0 contours
  - Extract F0
  - Smooth F0
  - Parameterize F0
- Models are good represenation of F0
  - small RMSE error (a few Hz)
- But can parameters be predicted?

Tilt Model: Taylor 97

- F0 derived:
  - accents+boundaries
  - 5 params per accent



Predicting Tilt Params

Dusterhoff (PhD 2000)

- □ CART for each Tilt param.
- □ Targeted F0 comparison.
- □ Comparison with ToBI (LR)
  Tilt: RMSE 32.5Hz and correlation of 0.60
  ToBI: RMSE 34.5Hz and correlation of 0.62
  Dynamical System (Ross): RMSE 33Hz

All on BU FM Radio data f2b

Other Intonation systems

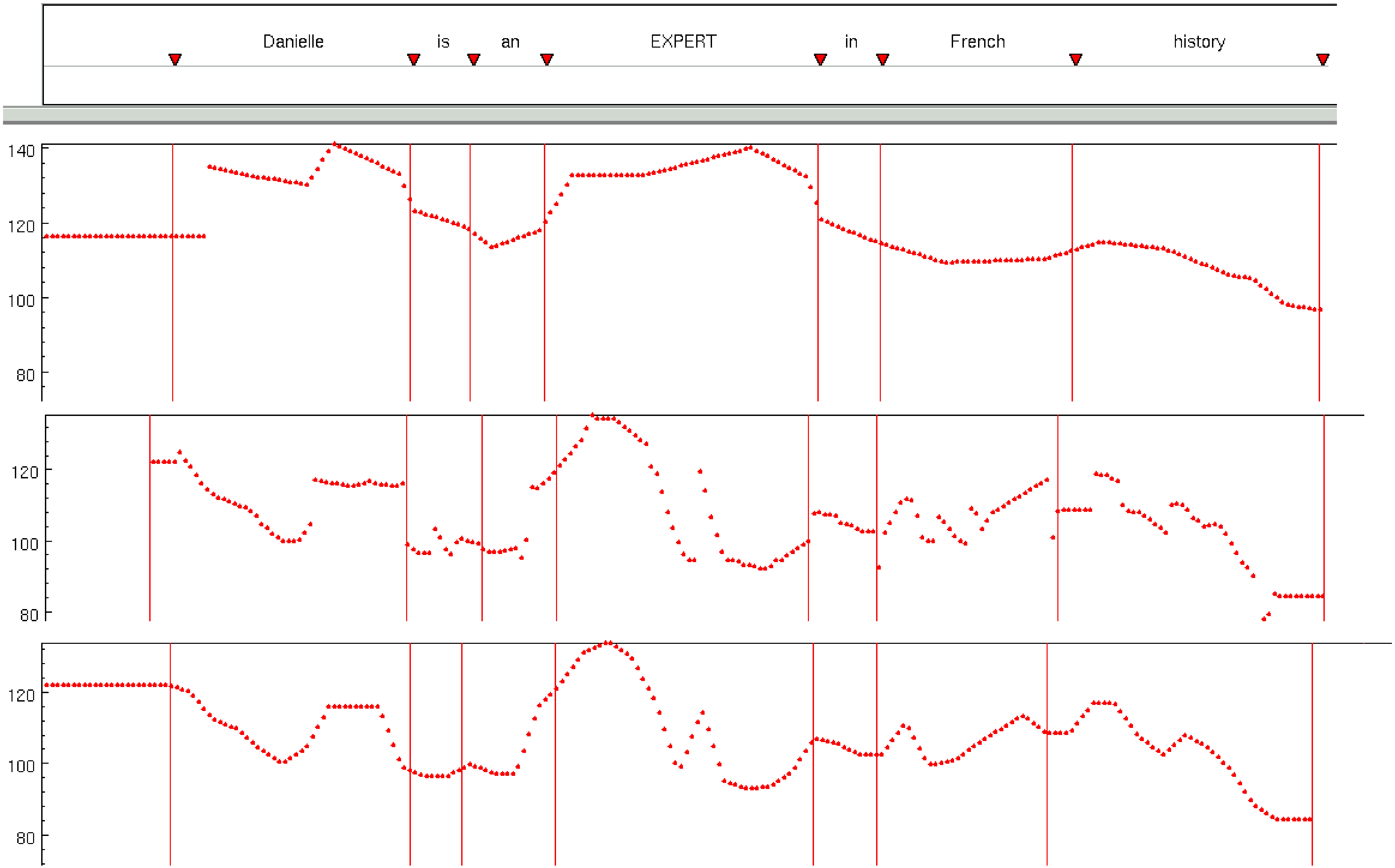
- Fujisaki:
  - physiologically based F0 generation
  - Japanese and German
  - hard to predict
- van Santen
  - six point model
  - over intonation phrases
- Möhler
  - Vector Quantization
  - accent types auto built from data
  - from (sort of) tilt-like parameters
- Malfrere and Dutoit
  - Select natural contours from database

Unit Selection of F0 contours

Raux and Black ASRU2003

- Database of natrual speech:
  - Extract F0 contours
  - Label each phone-sized segment with features
  - Phone/syllable/phrasal context
  - Select string of segments that match target

Rule vs Data F0 Generation


Danielle
is
an
EXPERT
in
French
history
140
120
100
80
120
100
80
120
100
80

Measuring Intonation

- ☐ Multiple acceptable ways
- ☐ RMSE and correlation
  - – insensitive to small errors
  - – can be swamped by uninteresting parts
  - – what about "microprosody"
  - – Absolute, log, zscores
- ☐ Human perception tests
  - – expensive to run
  - – not very exact

Intonation Theory Wishlist

- □ From "accents to F0"
- □ From "F0 to accents"
- □ Easily trainable to new styles

But ...

- □ Can't really be done in isolation from
  - phrasing, duration and power

Intonation Summary

- ☐ Position of accents on syllables
- ☐ Type of accents/boundaries
- ☐ F0 contour generation