Machine Learning in Signal Processing

Pitch and Intonation

F0 and Intonation

- What is F0
- What it typically looks like
- How to extract it from Speech
- How to model if
- How to model what it means

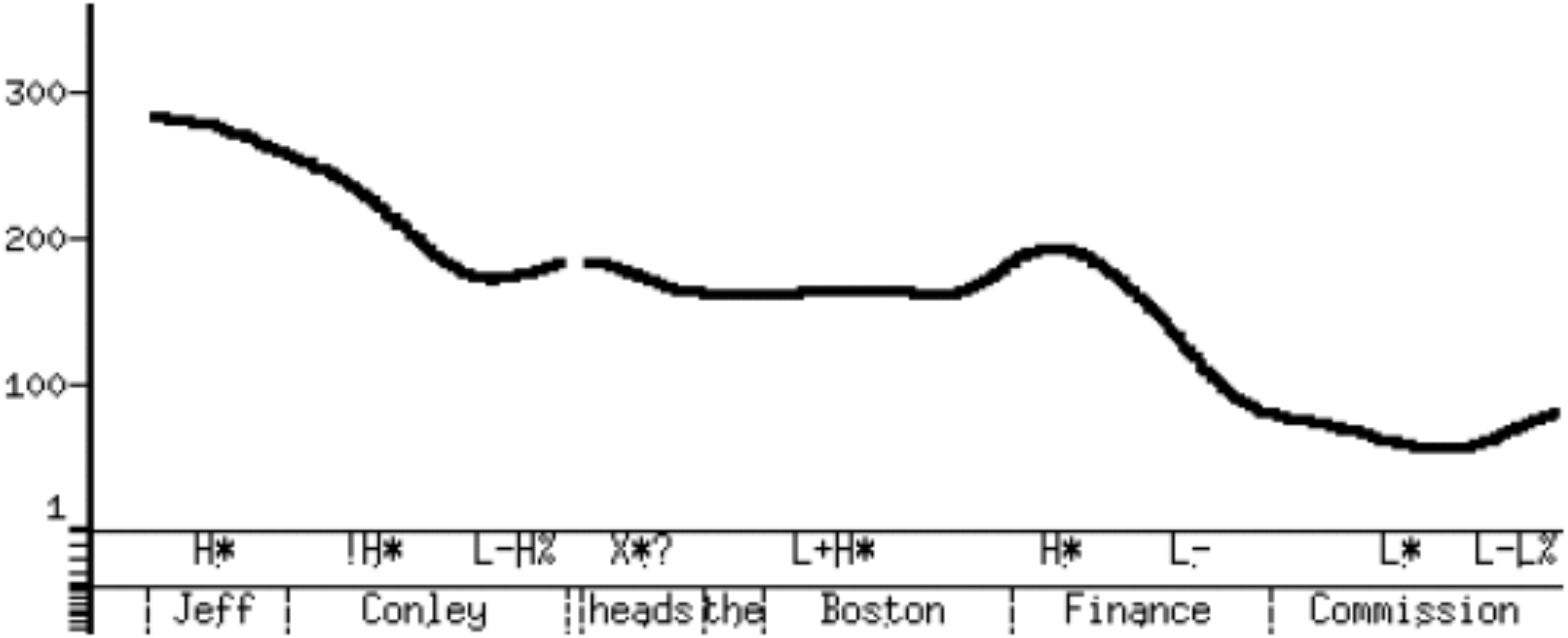
Prosody

- *How the phonemes will be said*
- *Four aspects of prosody*
  - *Phrasing: where the breaks will be*
  - *Intonation: pitch accents and F0 generation*
  - *Duration: how long the phonemes will be*
  - *Power: energy in signal*

Intonation

- *The fundamental tune*
  - *Accents (highlighting important parts)*
  - *F0 generation (the tune itself)*

Intonation Contour



Intonation Information

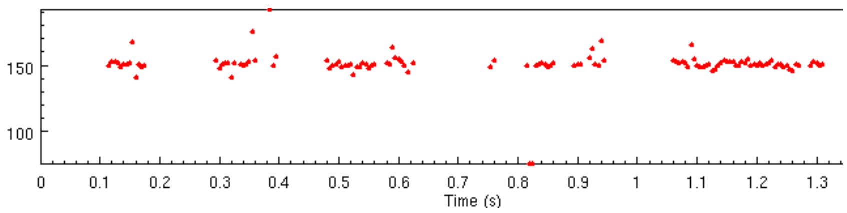
- Large pitch range (female)
- Authoritative since goes down at the end
  - News reader
- Emphasis for Finance H*
- Final has a raise – more information to come

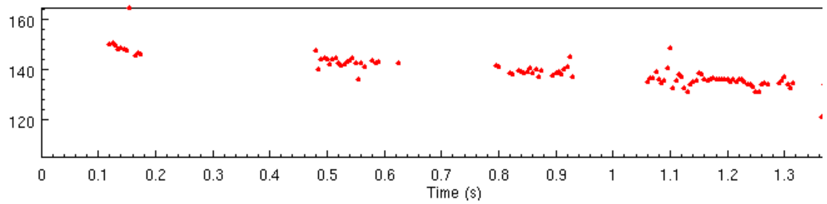
- Female American newsreader from WBUR
- (Boston University Radio)

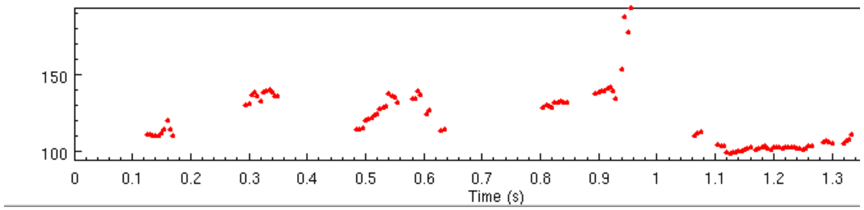
Intonation Examples

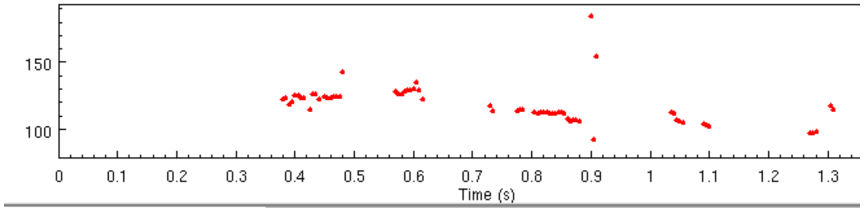
- *Fixed durations, flat F0.*
- *Decline F0*
- *“hat” accents on stressed syllables*
- *accents and end tones*
- *statistically trained*

F0 Examples

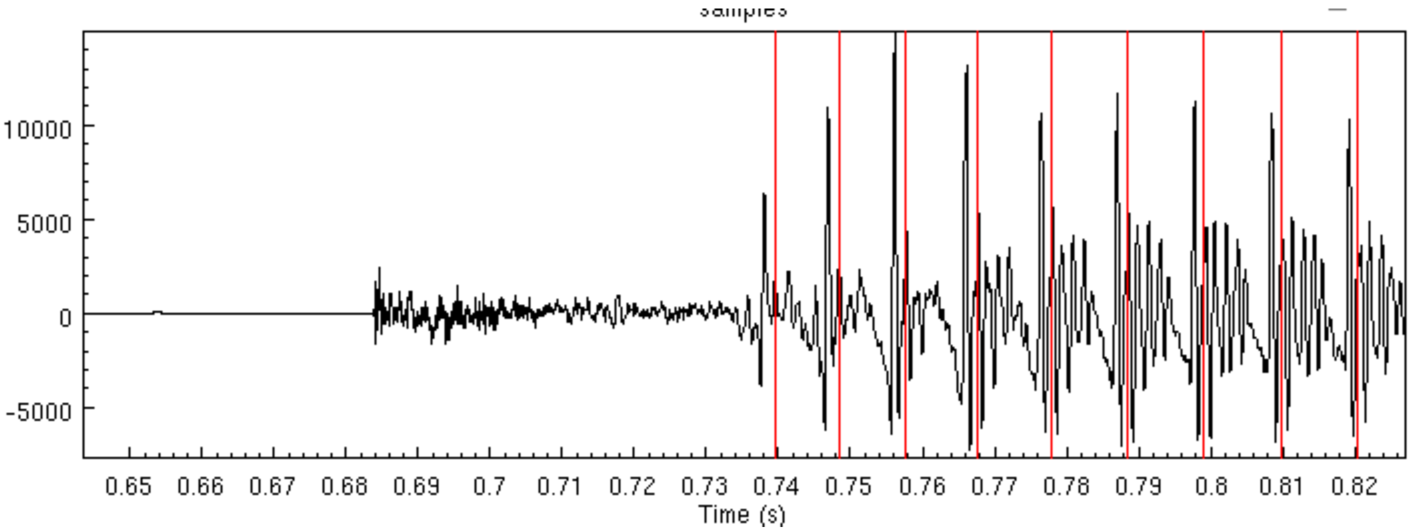

150
100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
Time (s)


160
140
120
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
Time (s)

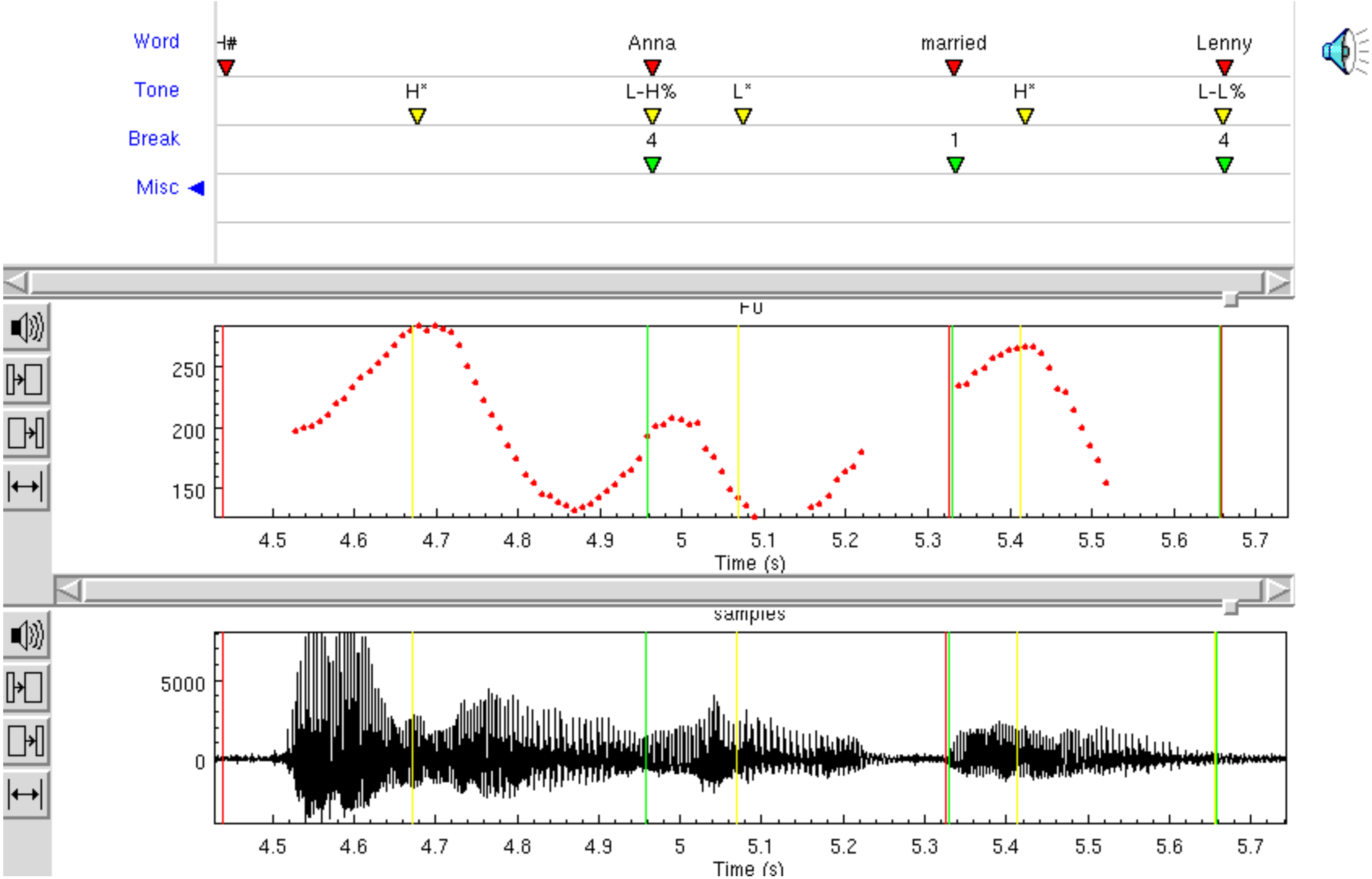

150
100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
Time (s)


150
100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
Time (s)

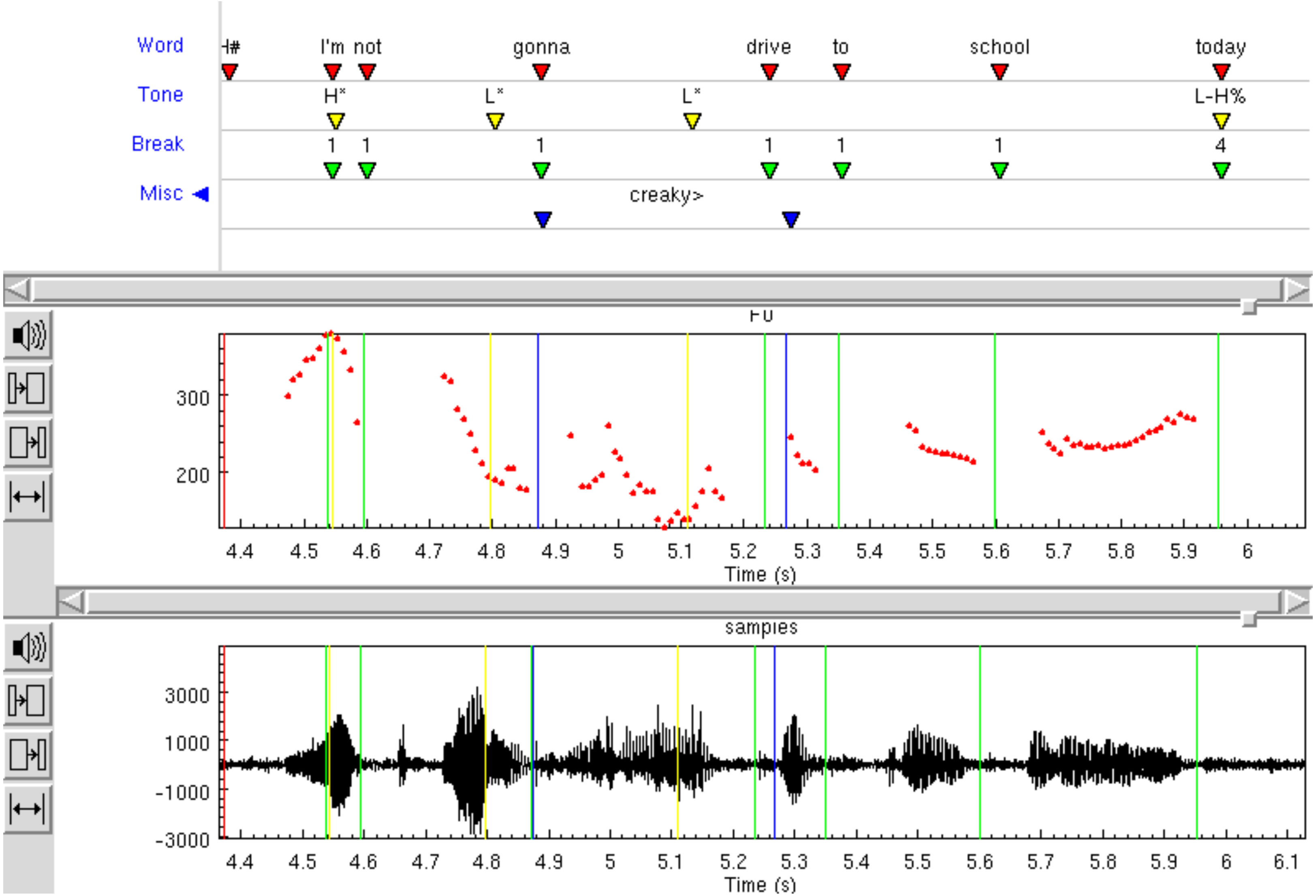
Finding Pitch



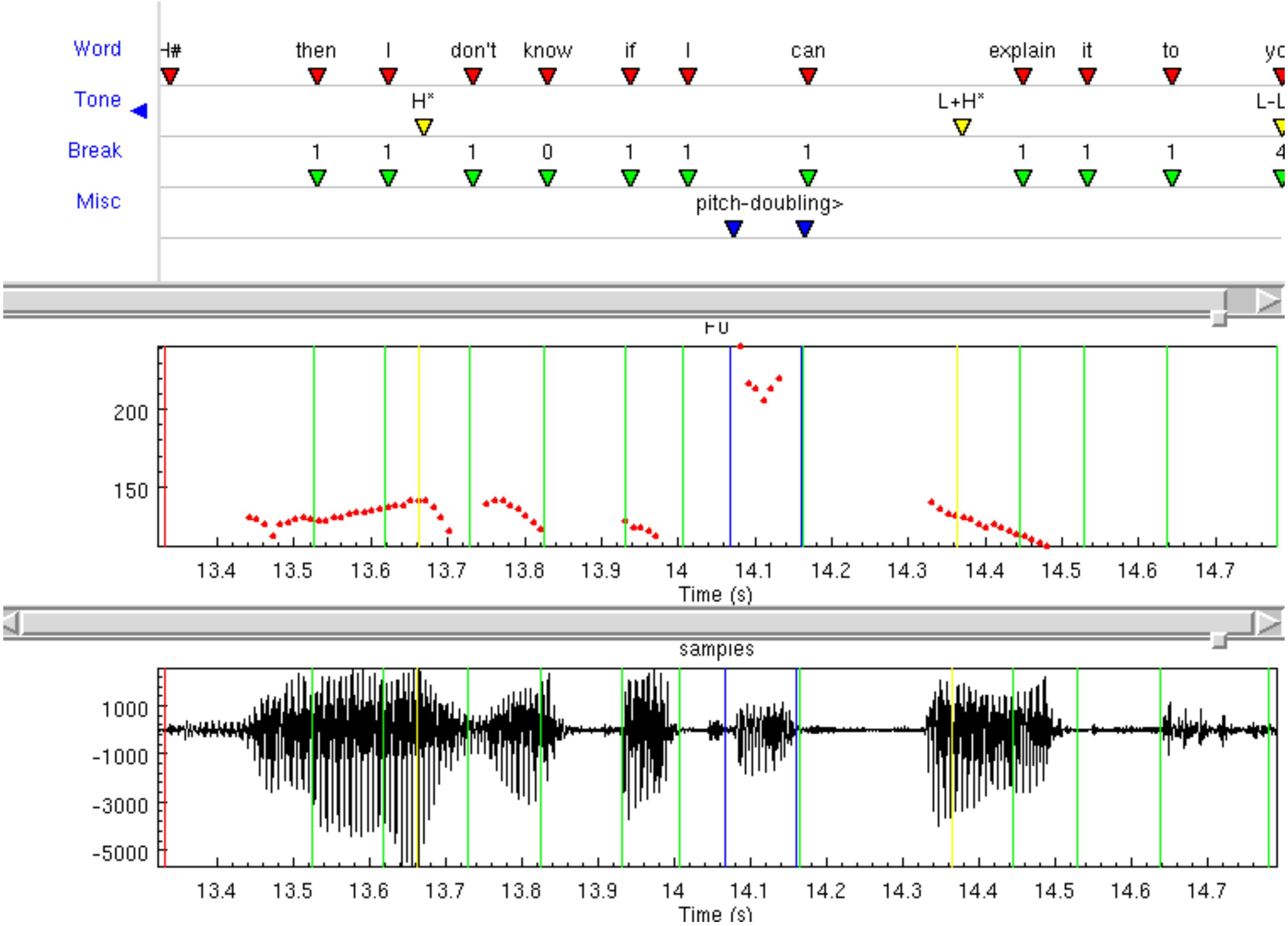
F0 Example


Word
Tone
Break
Misc
H#
Anna
married
Lenny
H*
L-H%
L*
H*
L-L%
4
1
4
F0
250
200
150
4.5
4.6
4.7
4.8
4.9
5
5.1
5.2
5.3
5.4
5.5
5.6
5.7
Time (s)
samples
5000
0
4.5
4.6
4.7
4.8
4.9
5
5.1
5.2
5.3
5.4
5.5
5.6
5.7
Time (s)

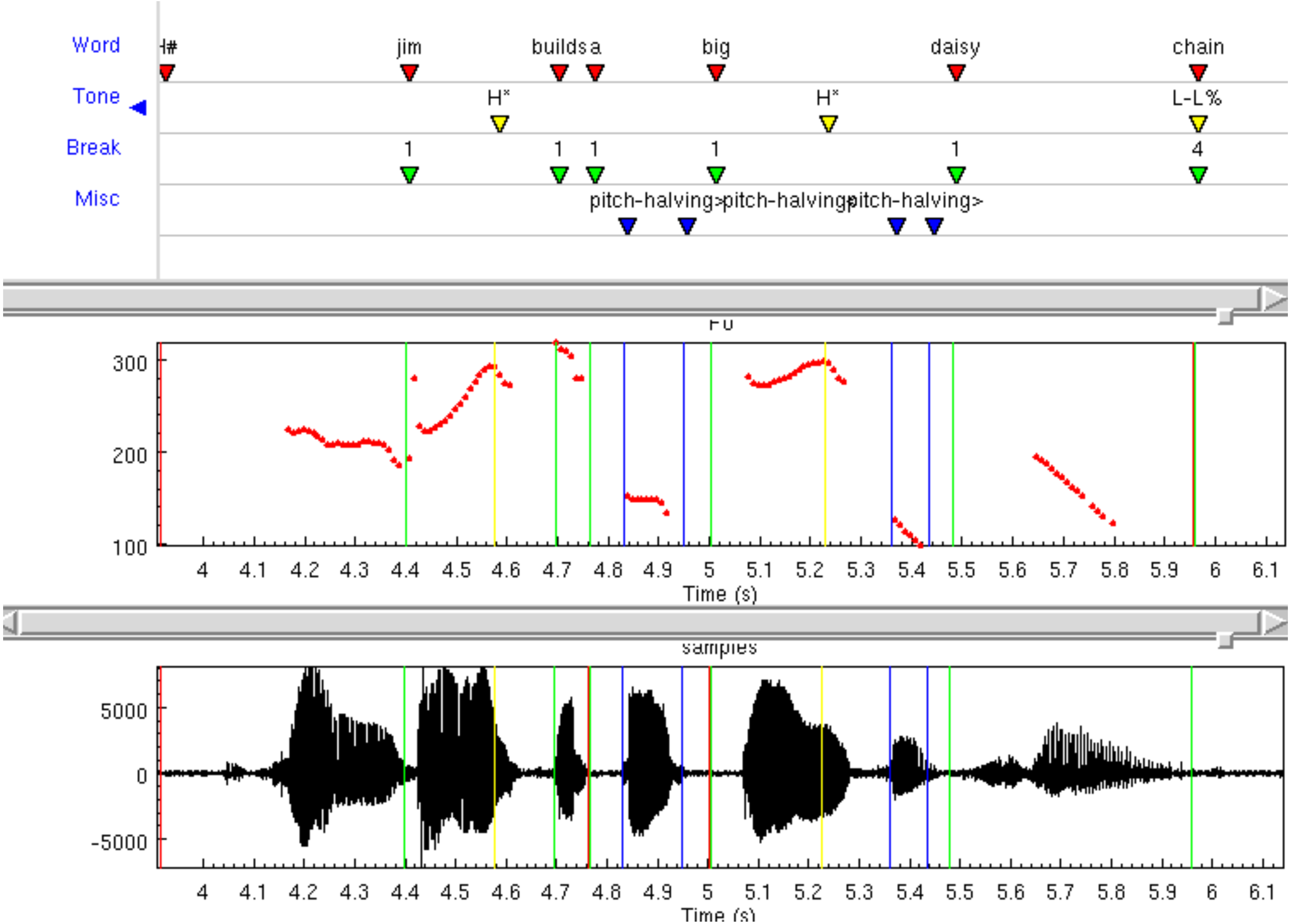
Creaky Voice



Pitch Doubling



Pitch Halving

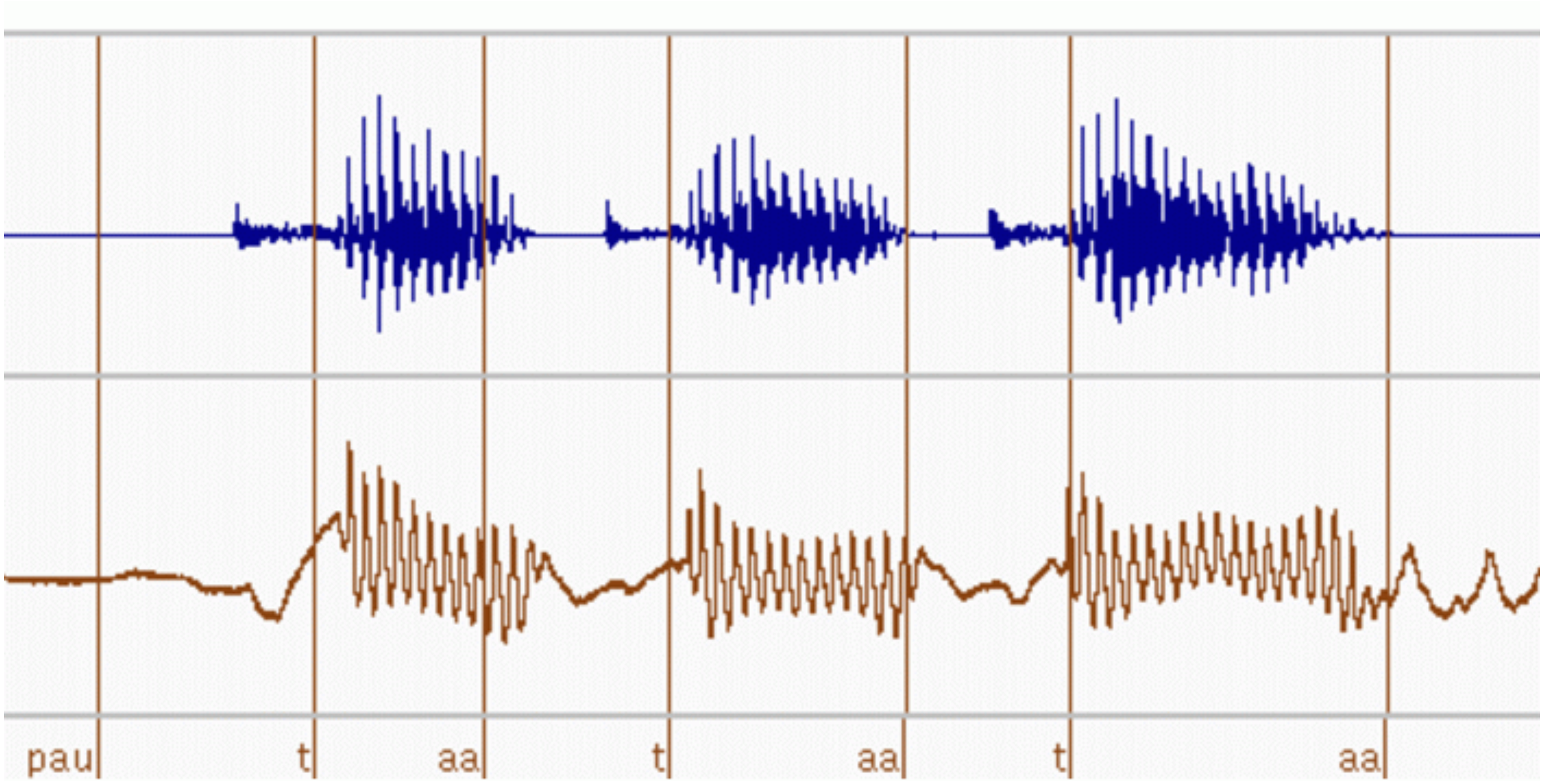


Finding Pitch

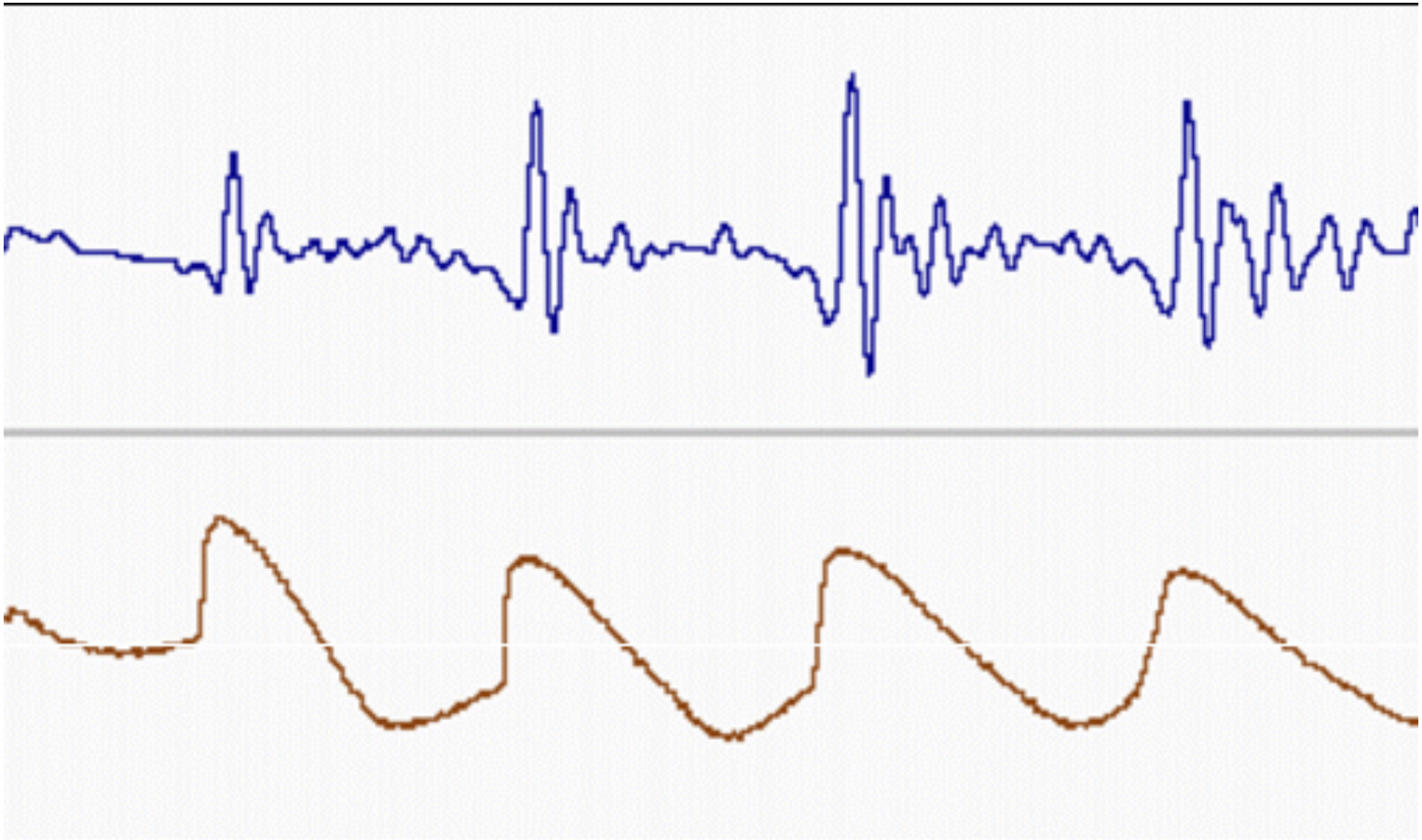
- *Know what you are looking for and look*
- *Low Pass filter*
  - *Pitch will be in range 60-300Hz*
- *LPC and residual*
  - *Peaks will be clearer in residual*
- *Use autocorrelation*
  - *Find common frequency*
  - *Though pitch changes over time*
- *Use *my* method it works best*
  - *ESPS get_f0*
  - *PDA*
  - *TEMPO (YIN)*

Use Electroglottograph

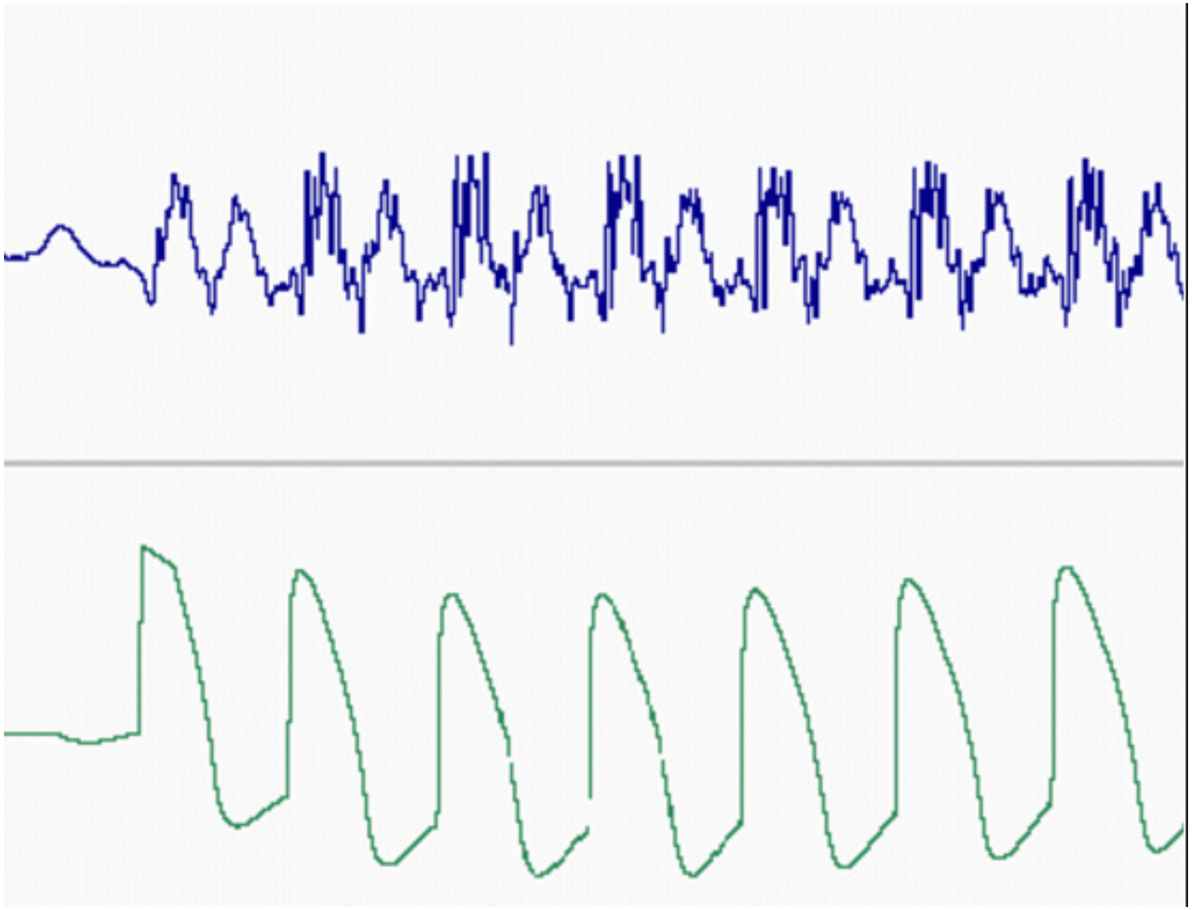
- EGG/Larynograph



EGG



EGG



What do you do with it?

- We’d like to model it
  - Predict it from text
  - Use it to find “focus” in speech
- Normalize it
  - Interpolate through unvoiced regions
  - Smooth it
  - Parameterize it

One strategy

- Find Pitch Periods
  - Low pass filter, use LPC residual
  - Use autocorrelation
  - Prune in expected range
- Interpolate through unvoiced regions
- Convert to F0
  - 1/pitch period
- Smooth
  - Or curve fit

F0 Generation

- Contour from accents (and durations)
- Piece together shapes of different accents
- Generated
  - By rule
  - Trained from data

Three Point Model

- Find F0 at
  - Syllable start
  - Voicing onset
  - Syllable end
- Predict these values with
  - CART/Linear Regression
- Sort of reasonable
  - RMS: 34.8
  - Correlation: 0.62

Find Structures/Shapes in F0

- Tilt Theory of Intonation
  - Describe shapes with 5 parameters
- Moeller Vector Quantized Shapes
  - 8 shapes
- Klabbers et al, Superpostional model
  - Parameters per “foot”

Intonational Phonology

- Accents and Boundaries
  - Where are the important changes in F0
- Accents on syllables
  - Identifies "important" words
    - It will be RAINY today in Boston
    - It will be rainy TODAY in Boston
    - It will BE rainy today IN Boston (strange)

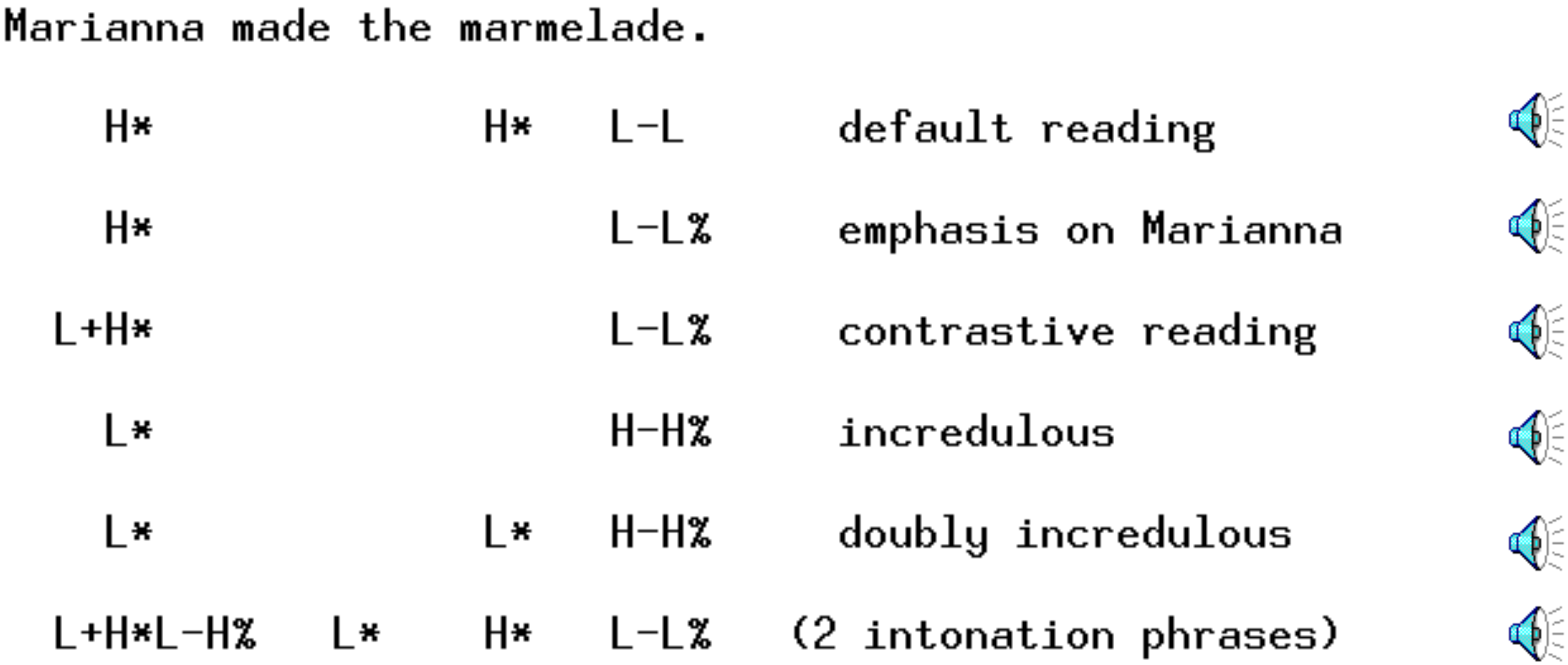
Where do the accents go?

- On important words
- First approximation
  - On stressed syllables in content words
    - It WILL be RAINY TODAY in BOSTON
  - About 80% correct on news reader speech
- CART training on more features
  - Content, proper nouns, POS, position in text
  - (not semantic information)

ToBI

- Tones and Break Indices
  - A labeling for intonation (English)
- Different accent types
  - H*, !H, L*, L+H*
- Different boundary types
  - L+L%, L+H%, H+H%,

ToBI examples



```
Marianna made the marmelade.

    H*             H*   L-L      default reading

    H*                  L-L%     emphasis on Marianna

  L+H*                  L-L%     contrastive reading

    L*                  H-H%     incredulous

    L*             L*   H-H%     doubly incredulous

  L+H*L-H%   L*    H*   L-L%   (2 intonation phrases)
```

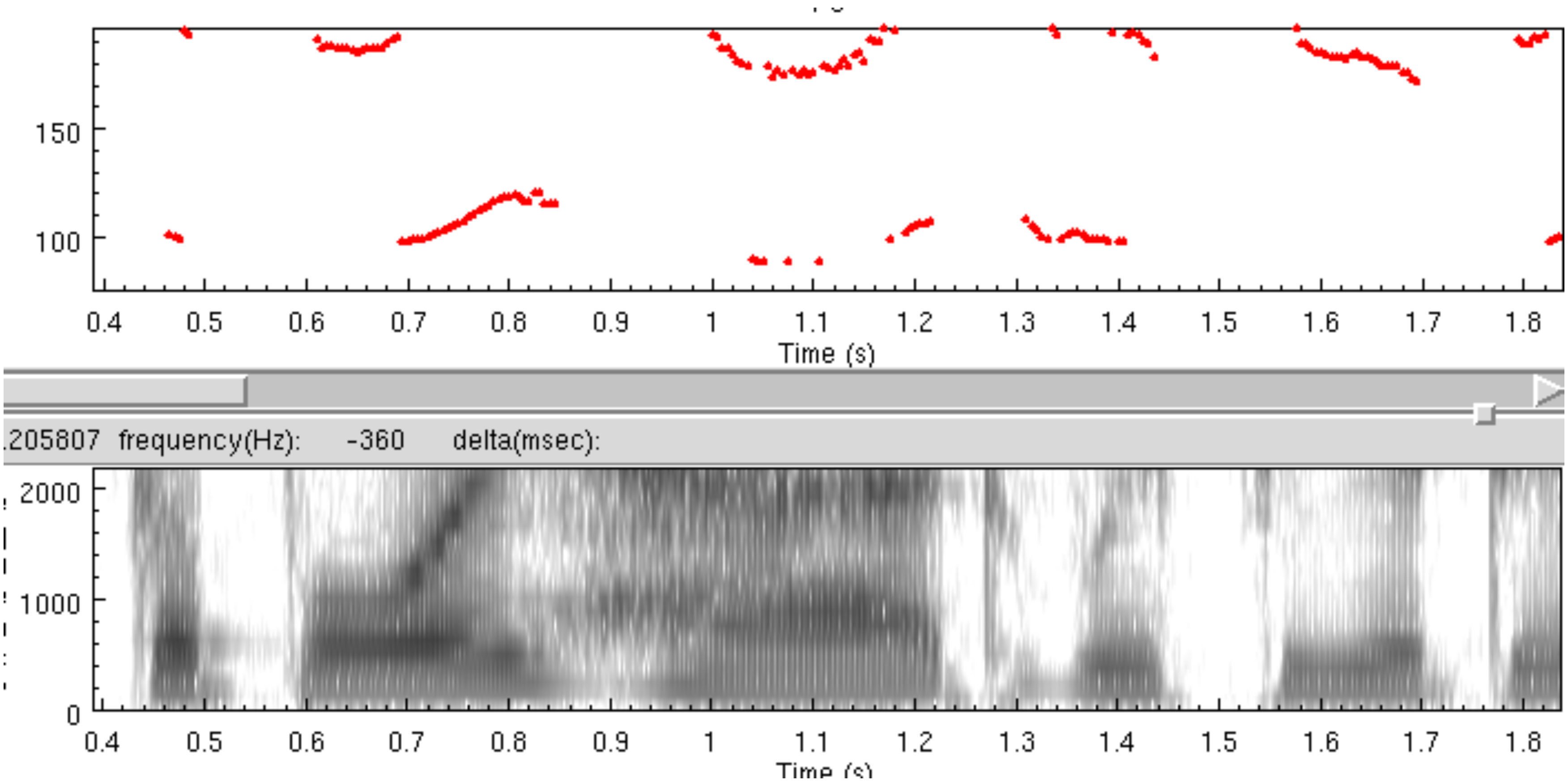
Using real contours

- From a data base of different contours
  - Select most appropriate one
- Record lots of different intonation examples
  - He DID then KNOW what HAD occurred
  - TARZAN and JANE raised THEIR heads
  - …
- Label them and select the contours when you want emphasis

Emphasis Synthesis

- *This is a short example*
- *THIS is a short example*
- *This IS a short example*
- *This is A short example*
- *This is a SHORT example*
- *This is a short EXAMPLE*

Extracting F0 from "real" speech



.205807 frequency(Hz): -360 delta(msec):

Summary

- *Extracting F0 from speech*
- *Modeling F0*
  - *Low level to high level*
- *Intonational accents*
  - *How to predict where the go*
- *Problems in moving from lab to real speech*