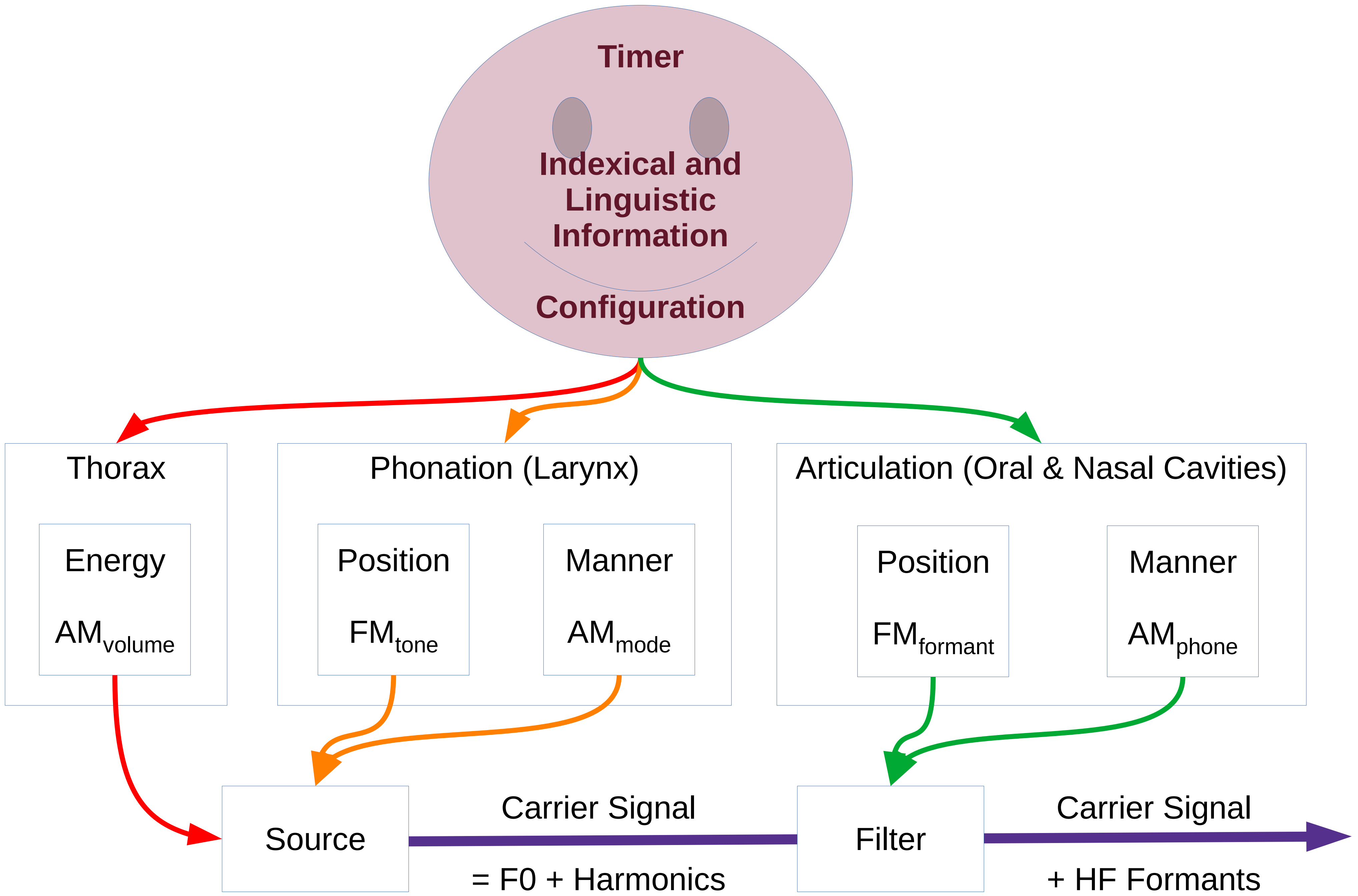
Lecture 3: Rhythm
3A: The F0 component

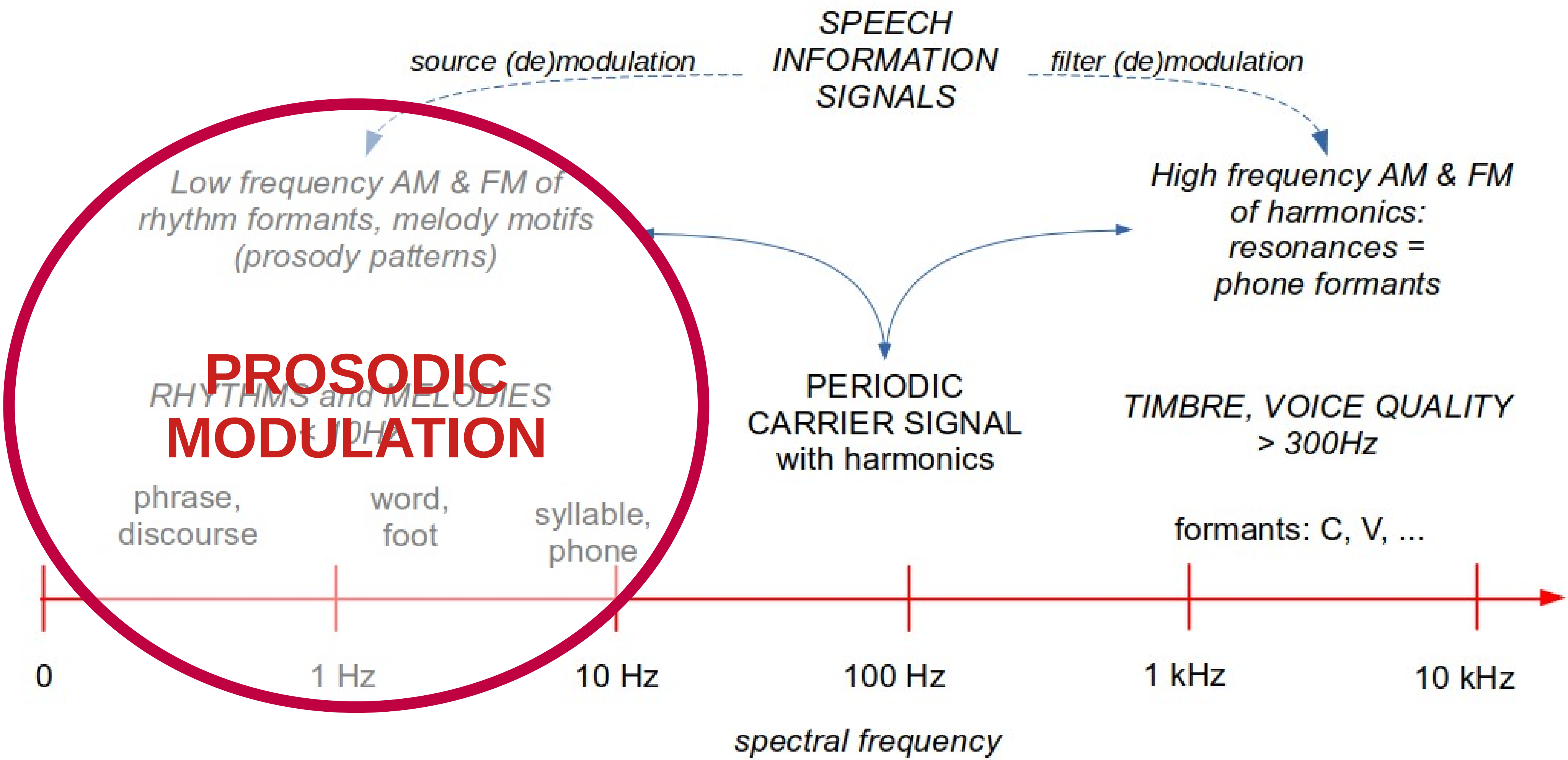
Dafydd Gibbon
Bielefeld University, Germany
2022-04-29

II Brazilian Congress of Prosody
Minicourse 9: 25, 27, 29 April 2022
(09:00-11:30 Brazilian Standard Time)

Encoding Information by Modulating a Carrier Signal



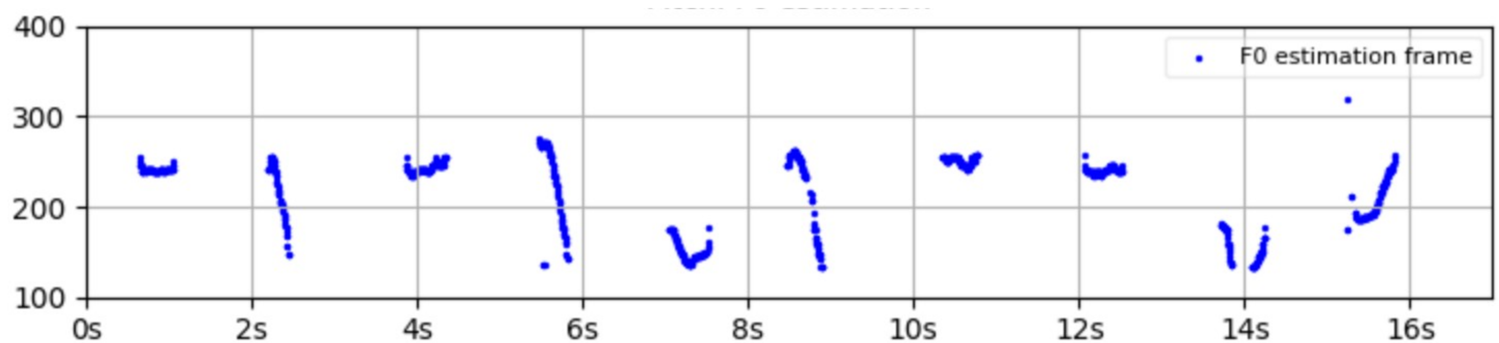
Modulation Codes: AM and FM



FM Code: Tone, Pitch Accent, Intonation

Sino-Tibetan
Pǔtōnghuà
ISO-693-3 *cmn*

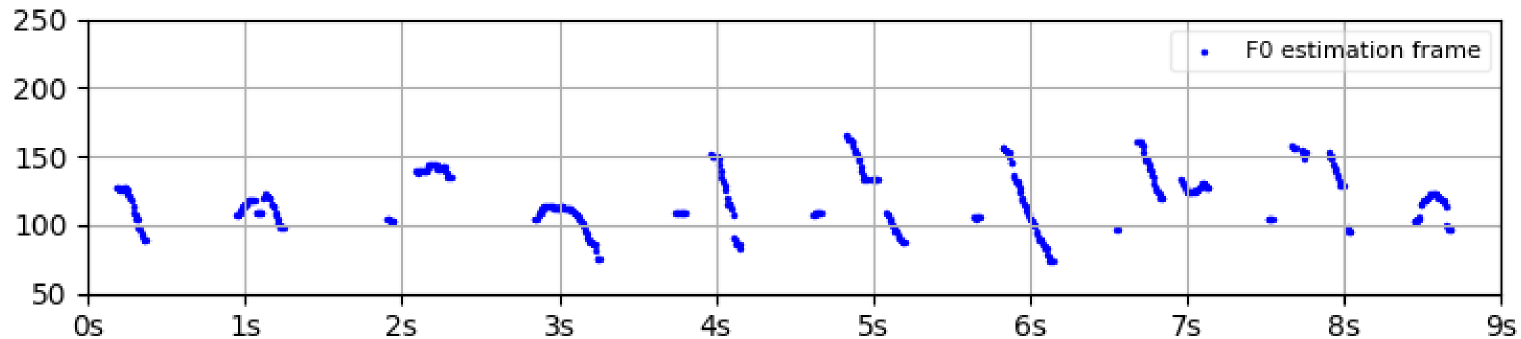
lexical tone



yī èr sān sì wǔ liù qī bā jiǔ shí

Niger-Congo
Ibibio
ISO-693-3 *ibb*

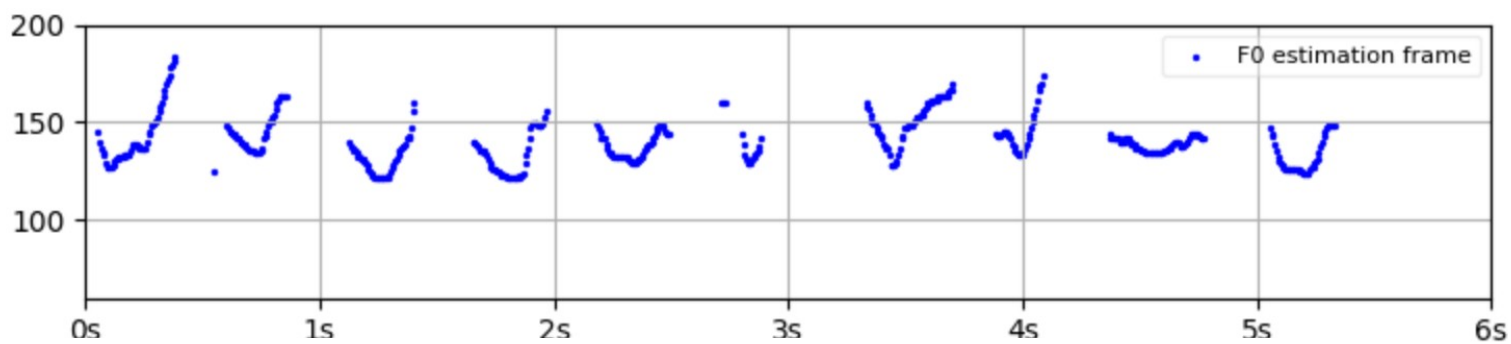
lexical and morphological tone



kèèd ìbà ìtá ìnààñ ìtíòn ìtíòkèèd ìtíàbà ìtiáìtá ùsúkkéèd dùòp

Indo-Germanic
English
ISO 693-3 *eng*

stress-pitch accent & intonation

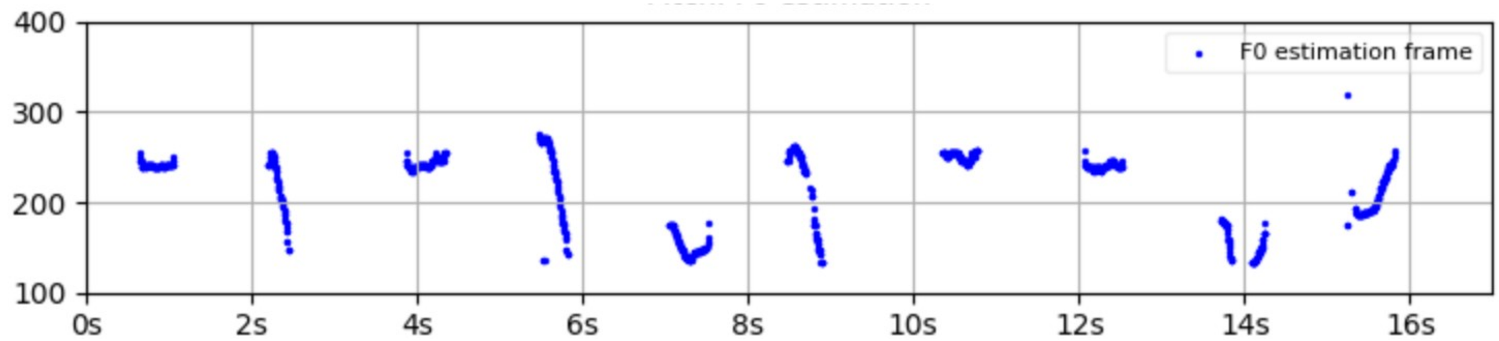


one two three four five six seven eight nine ten

FM Code: Tone, Pitch Accent, Intonation

Sino-Tibetan
Pŭtōnghuà
ISO-693-3 *cmn*

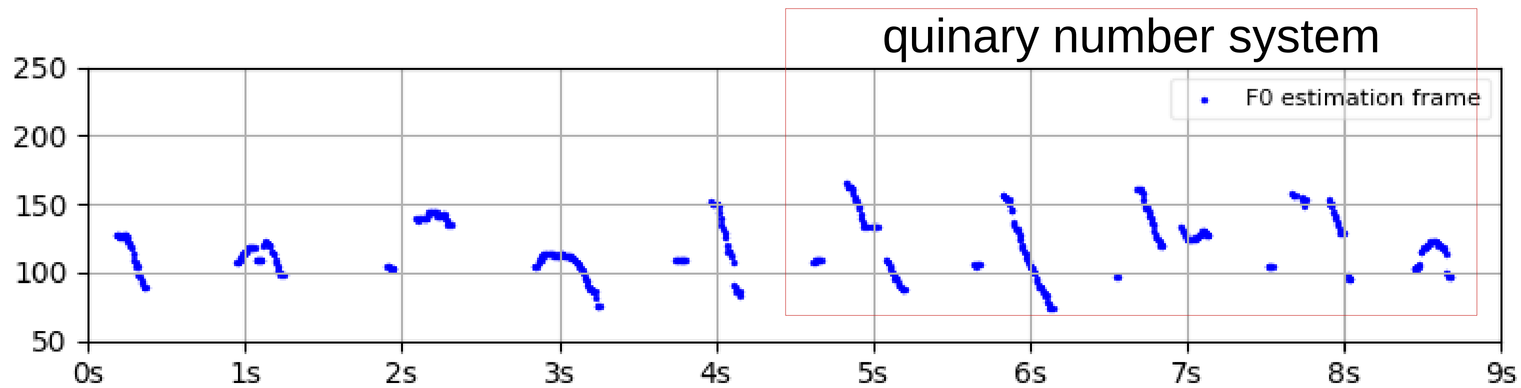
lexical tone



yī èr sān sì wŭ liù qī bā jiŭ shí

Niger-Congo
Ibibio
ISO-693-3 *ibb*

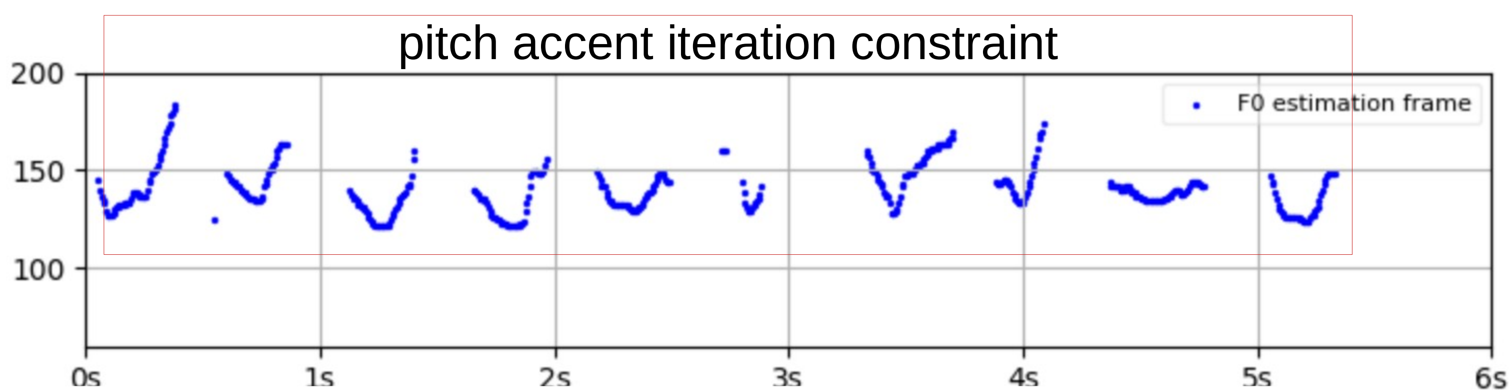
lexical and morphological tone



kèèd ìbà ìtá ìnààñ ìtíòn ìtíòkèèd ìtíàbà ìtiáìtá ùsúkkéèd dùòp

Indo-Germanic
English
ISO 693-3 *eng*

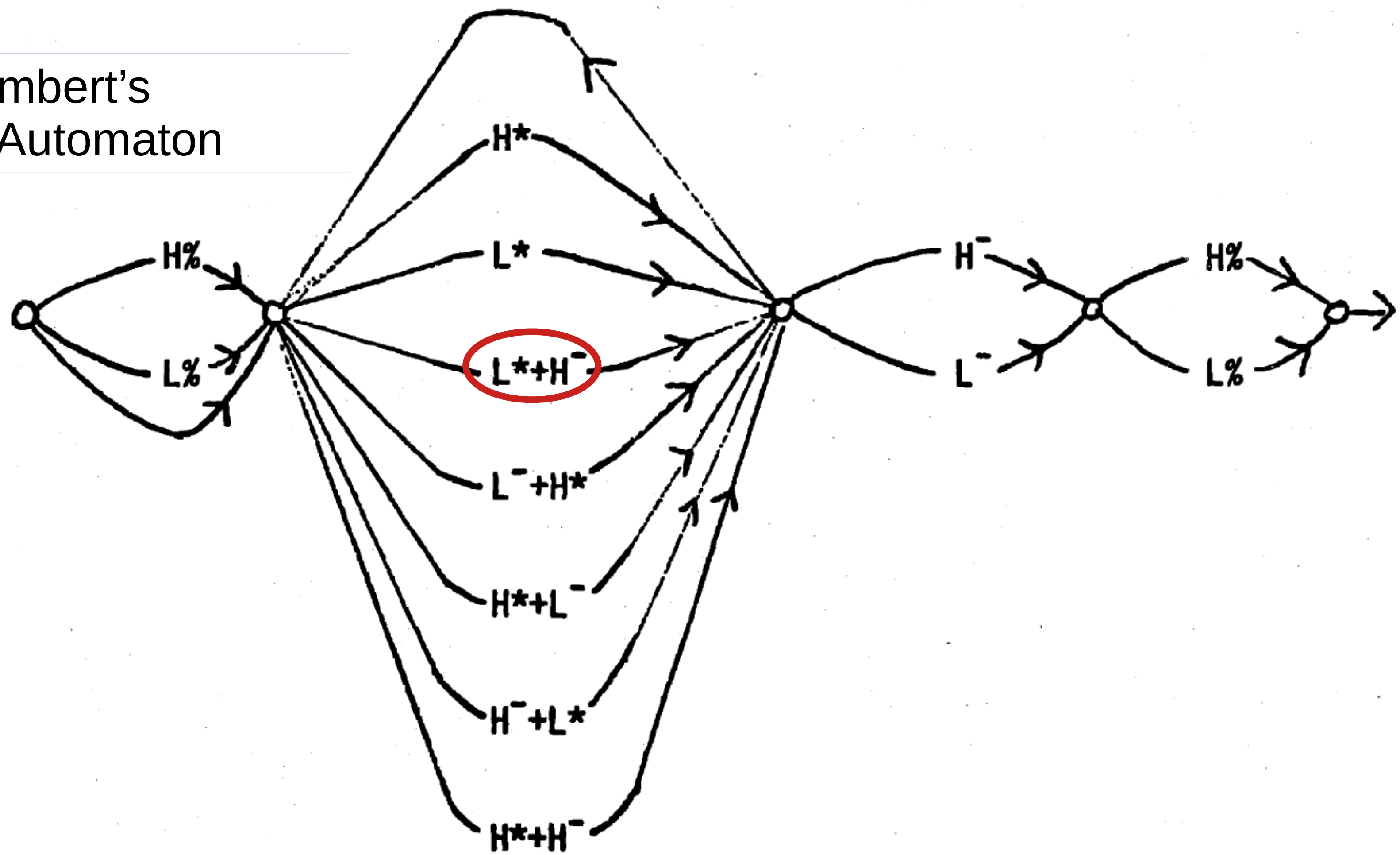
stress-pitch accent & intonation



one two three four five six seven eight nine ten

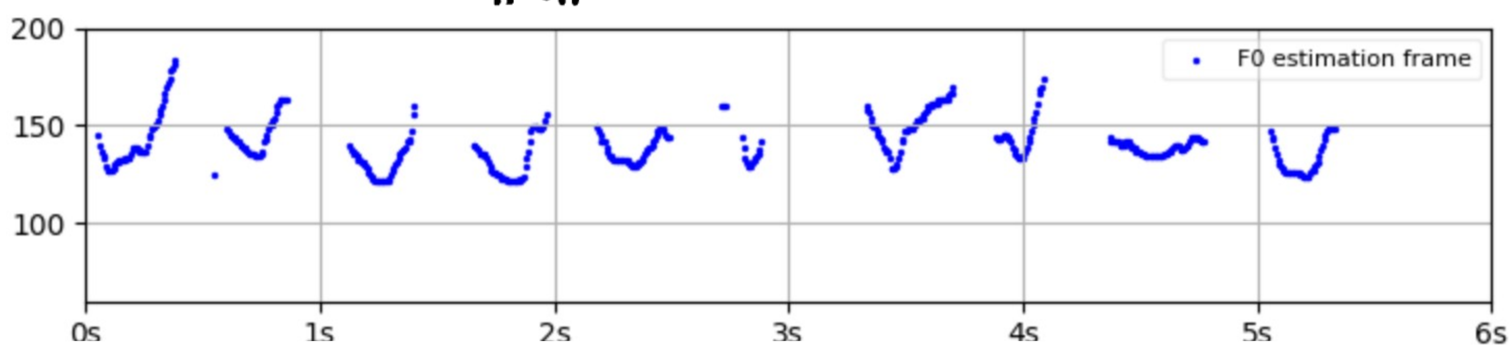
FM Code: Tone, Pitch Accent, Intonation

Pierrehumbert's
Finite State Automaton



Indo-Germanic
English
ISO 693-3 *eng*

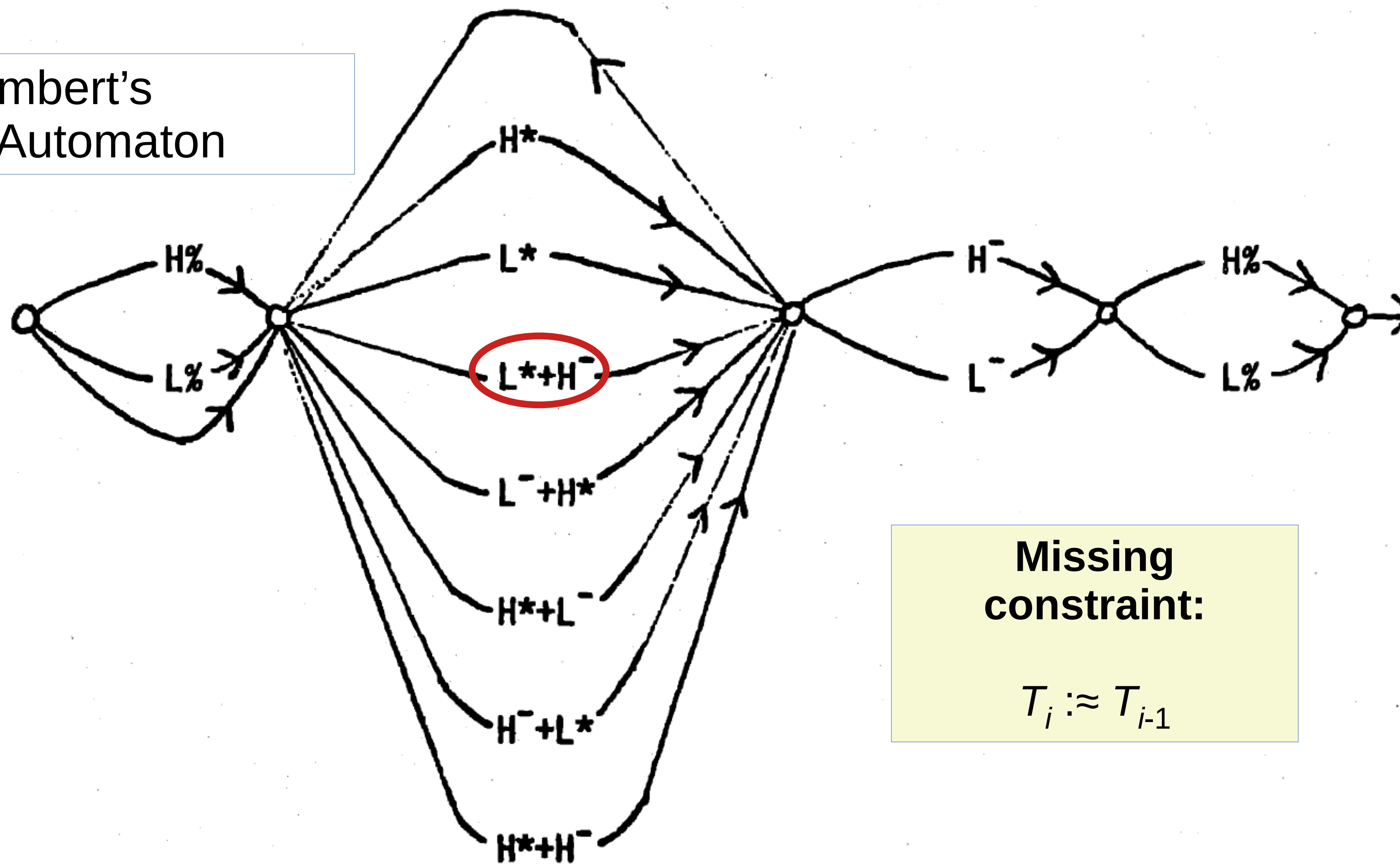
stress-pitch accent & intonation



one two three four five six seven eight nine ten

FM Code: Tone, Pitch Accent, Intonation

Pierrehumbert's
Finite State Automaton

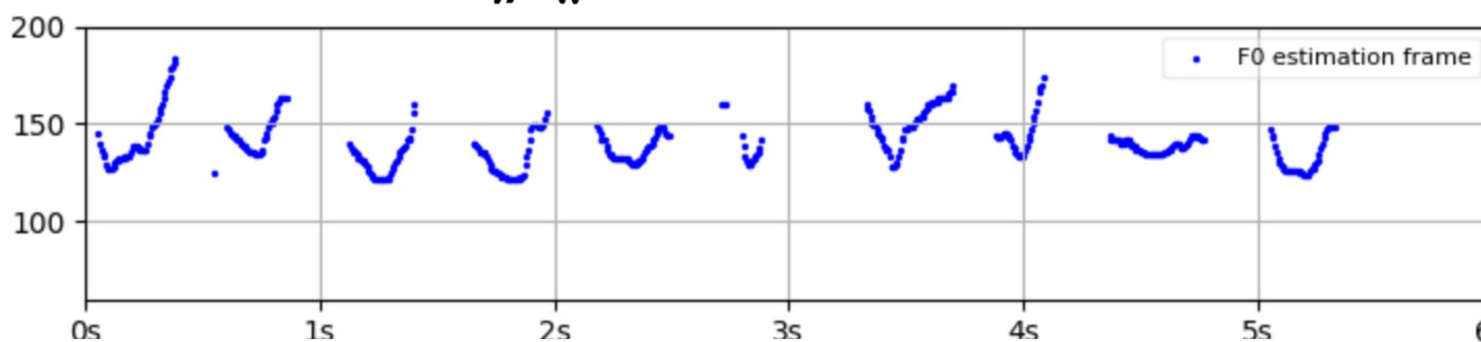


Missing constraint:

$T_i :\approx T_{i\text{-}1}$

Indo-Germanic
English
ISO 693-3 *eng*

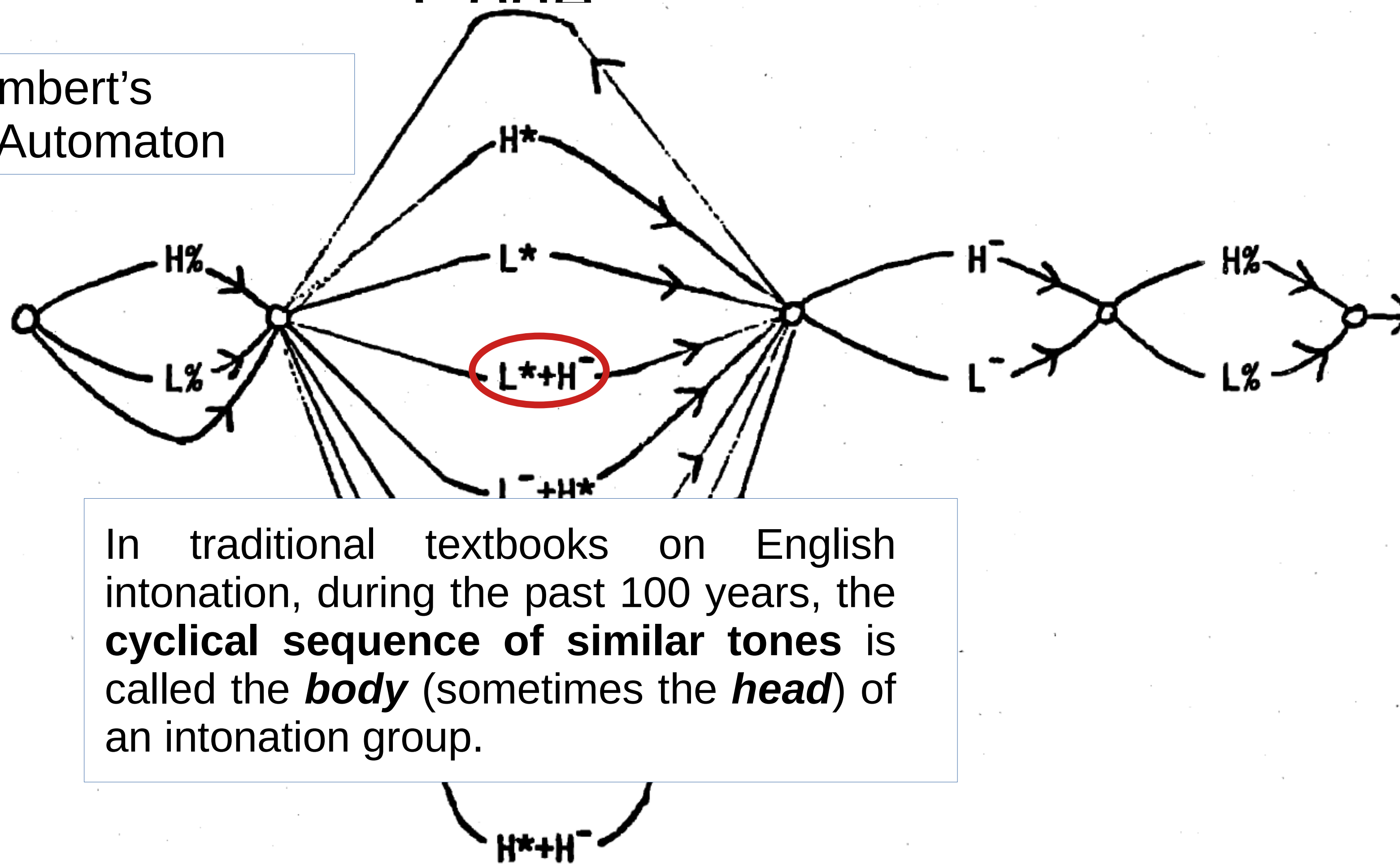
stress-pitch accent & intonation



one two three four five six seven eight nine ten

Tones, Pitch Accents and Intonation: the 'Modulation Code'

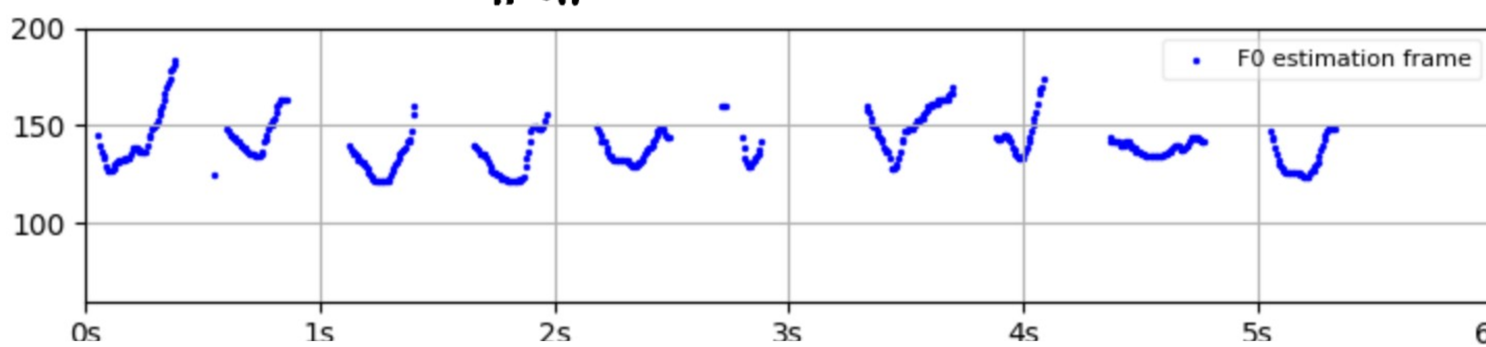
Pierrehumbert's
Finite State Automaton



In traditional textbooks on English intonation, during the past 100 years, the **cyclical sequence of similar tones** is called the ***body*** (sometimes the ***head***) of an intonation group.

Indo-Germanic
English
ISO 693-3 *eng*

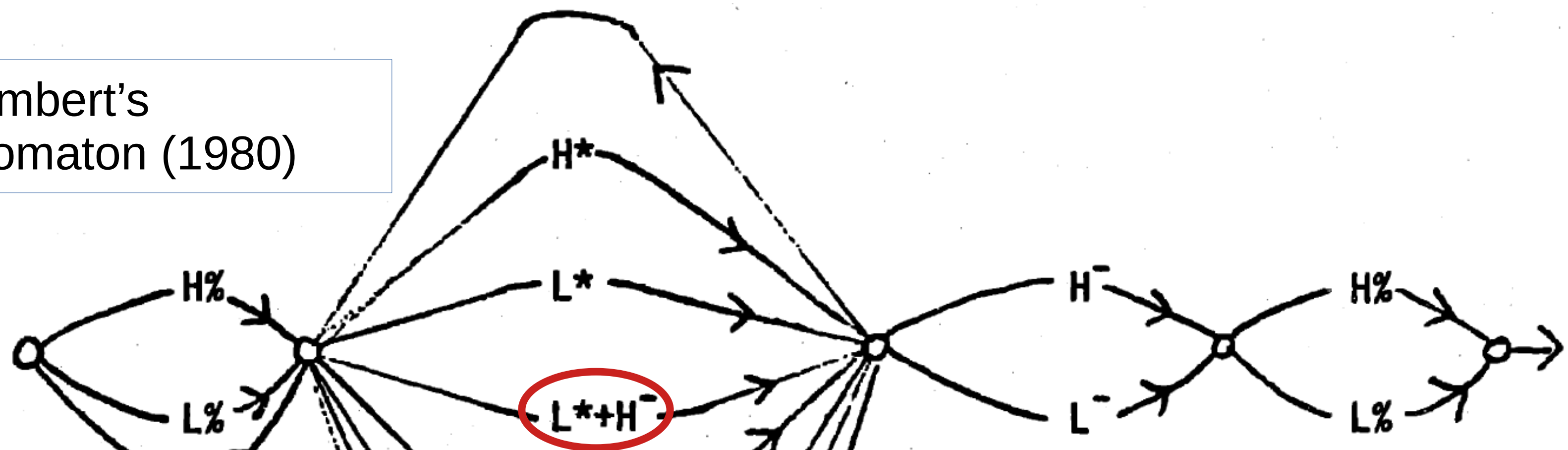
stress-pitch accent & intonation



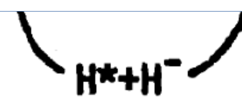
one two three four five six seven eight nine ten

FM Code: Tone, Pitch Accent, Intonation

Pierrehumbert's
Finite State Automaton (1980)

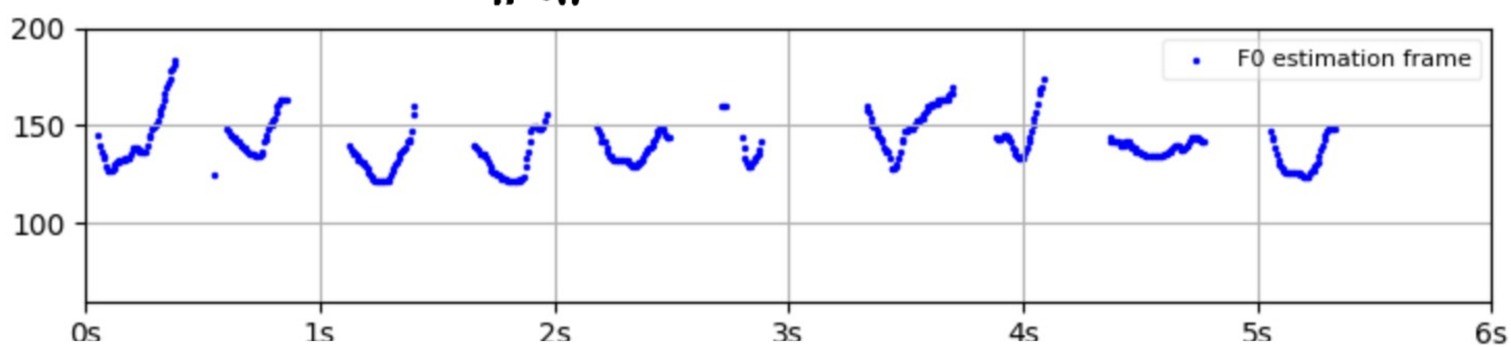


Dilley (1997: 87ff.)
- proposed an **accent sequence similarity constraint** for the head pattern,
- in order to explain such sequential pitch accent patterns as **correlate of coherent grammatical patterns** and
- as a means of **entraining the attention of listeners** to expect pattern changes such as nuclear tones.



Indo-Germanic
English
ISO 693-3 *eng*

stress-pitch accent & intonation



one two three four five six seven eight nine ten

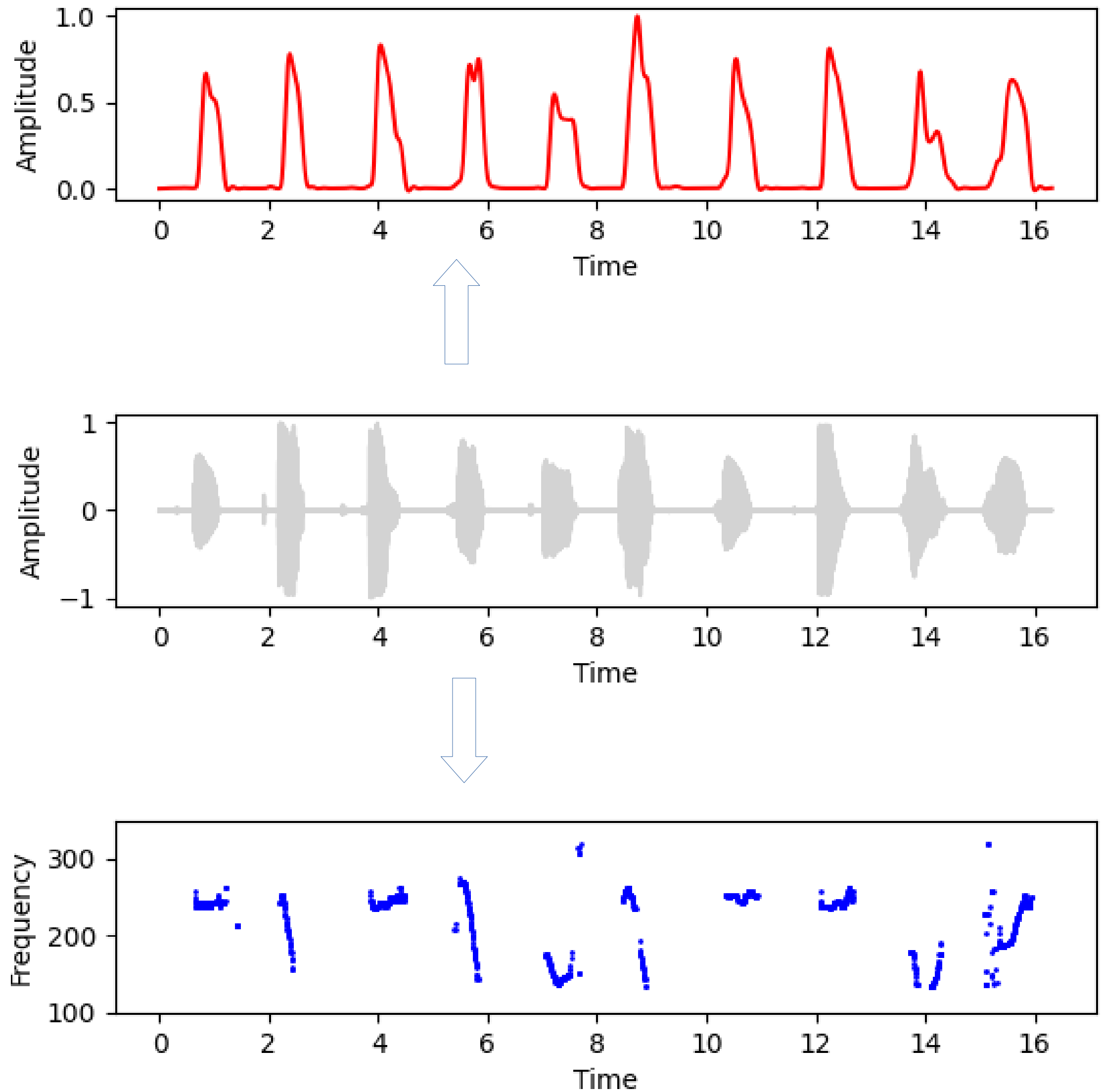
F0 Modulation (FM) and F0 Demodulation

Low Frequency AM and FM Demodulation

AM envelope demodulation:
- phonetics:
  amplitude curve, syllable, stress-accent
- phonology:
  sonority curve, syllables, stress

Modulated carrier signal

FM envelope demodulation:
- phonetics:
  F0, pitch track
- phonology:
  tones, pitch accents, intonation



F0 Modulation (FM) and F0 Demodulation

The FM and AM modes have their general signal processing meanings:

Basic sinusoid carrier signal: $S_{car} = A_{car}\cos(2\pi f_{car}t+\varphi_{car})$

Basic sinusoid modulation signal: $S_{mod} = A_{mod}\cos(2\pi f_{mod}t+\varphi_{mod})$

But while an unmodulated radio carrier signal is close to a pure sine wave, the unmodulated speech signal is close to a <u>triangular 'sawtooth' wave, with a series of harmonics.</u>

In FM, the values of f_{car}, the frequency component, are modified
(mainly for the functions of lexical tone, pitch accent and in intonation)
by addition with f_{mod} (in this case: f_{modfm}) values at source.

In AM the component A_{car} (amplitude) of S_{car} is modified
(mainly for vowel and consonant sequences)
by multiplication with A_{mod} (in this case: A_{modam})

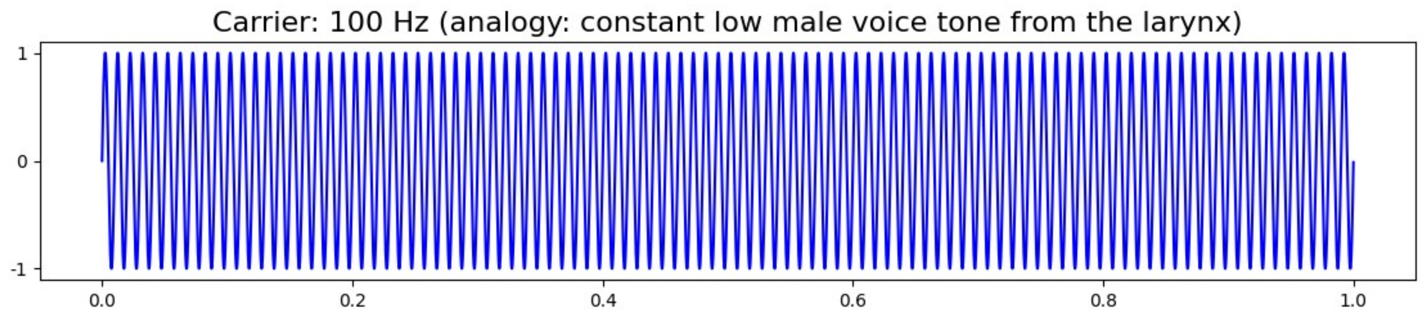
The interruption of the signal by stop/plosive consonants is a special case of AM.

Phase modulation:

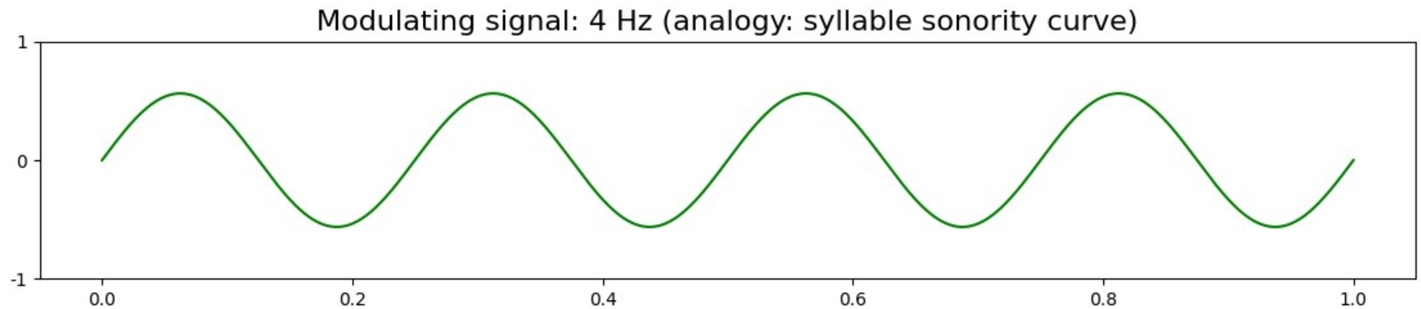
For present purposes, the phases φ of carrier and modulation signals are not considered, though changes in tempo imply phase modulation. However, PM and FM are closely related.

F0 Modulation (FM) and F0 Demodulation

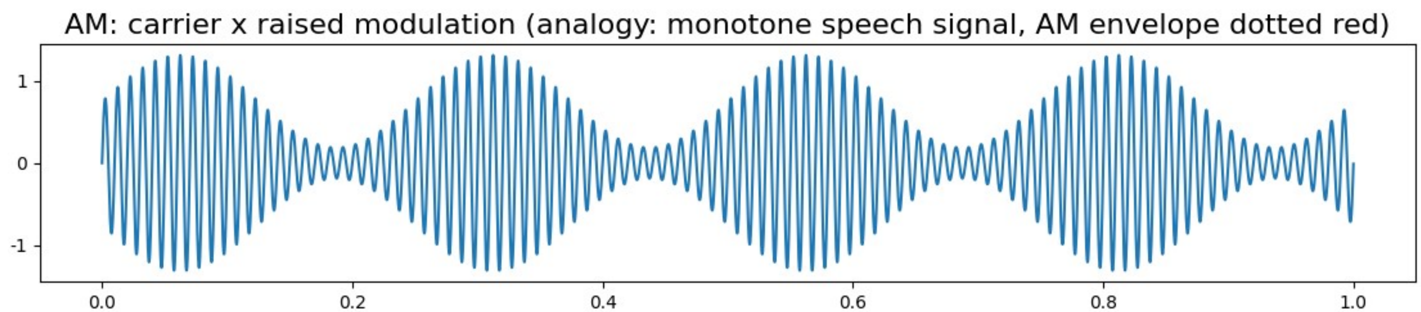
Carrier signal:



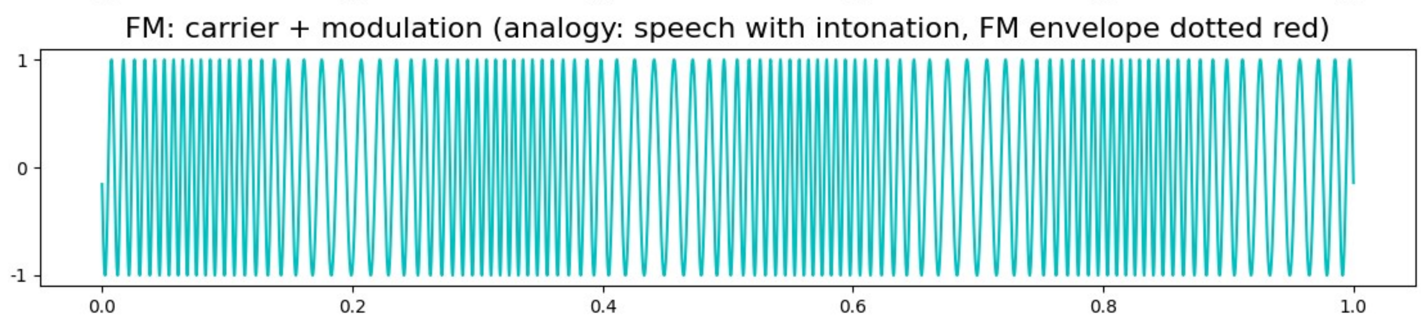
Modulation signal:



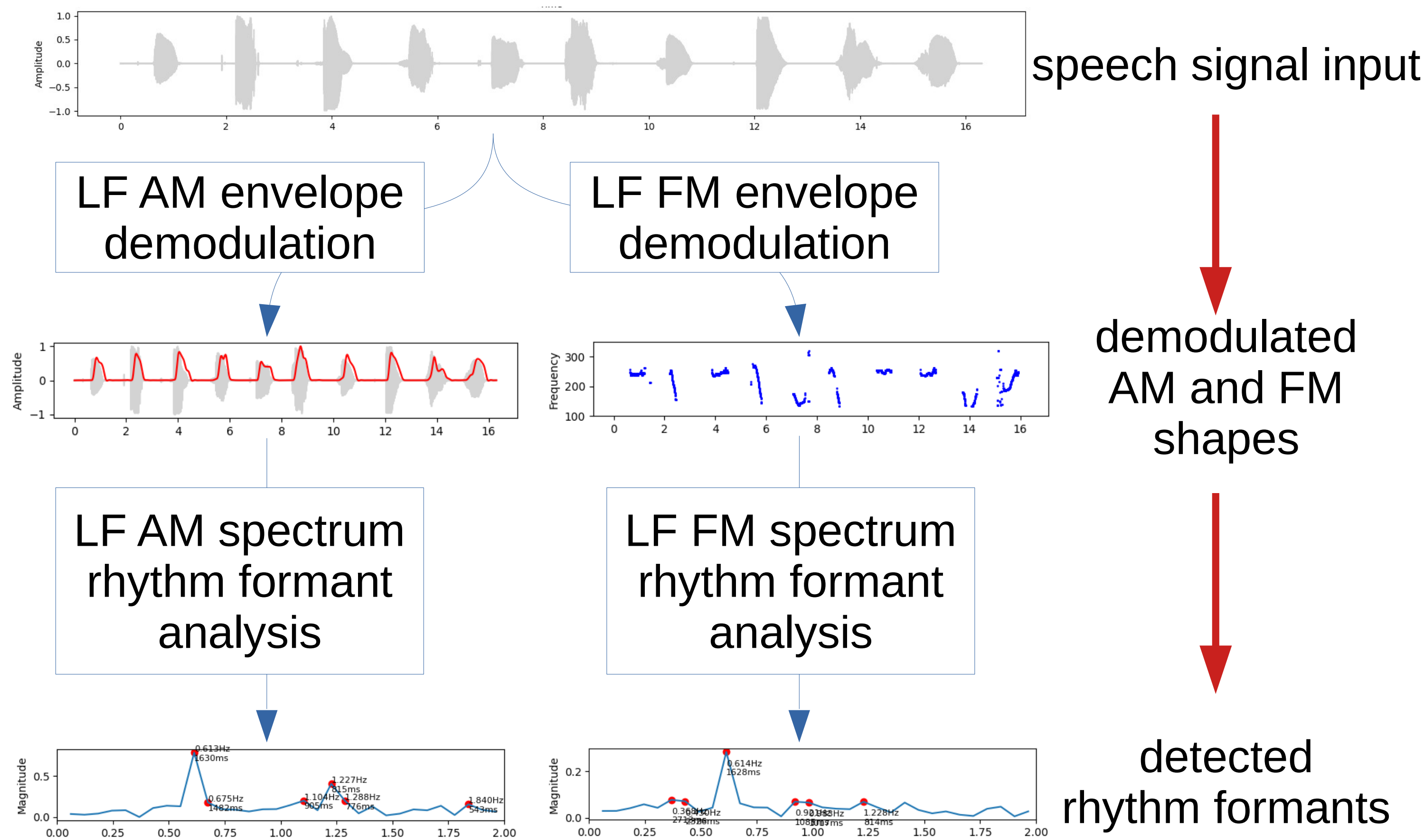
Amplitude modulated signal:



Frequency modulated signal:

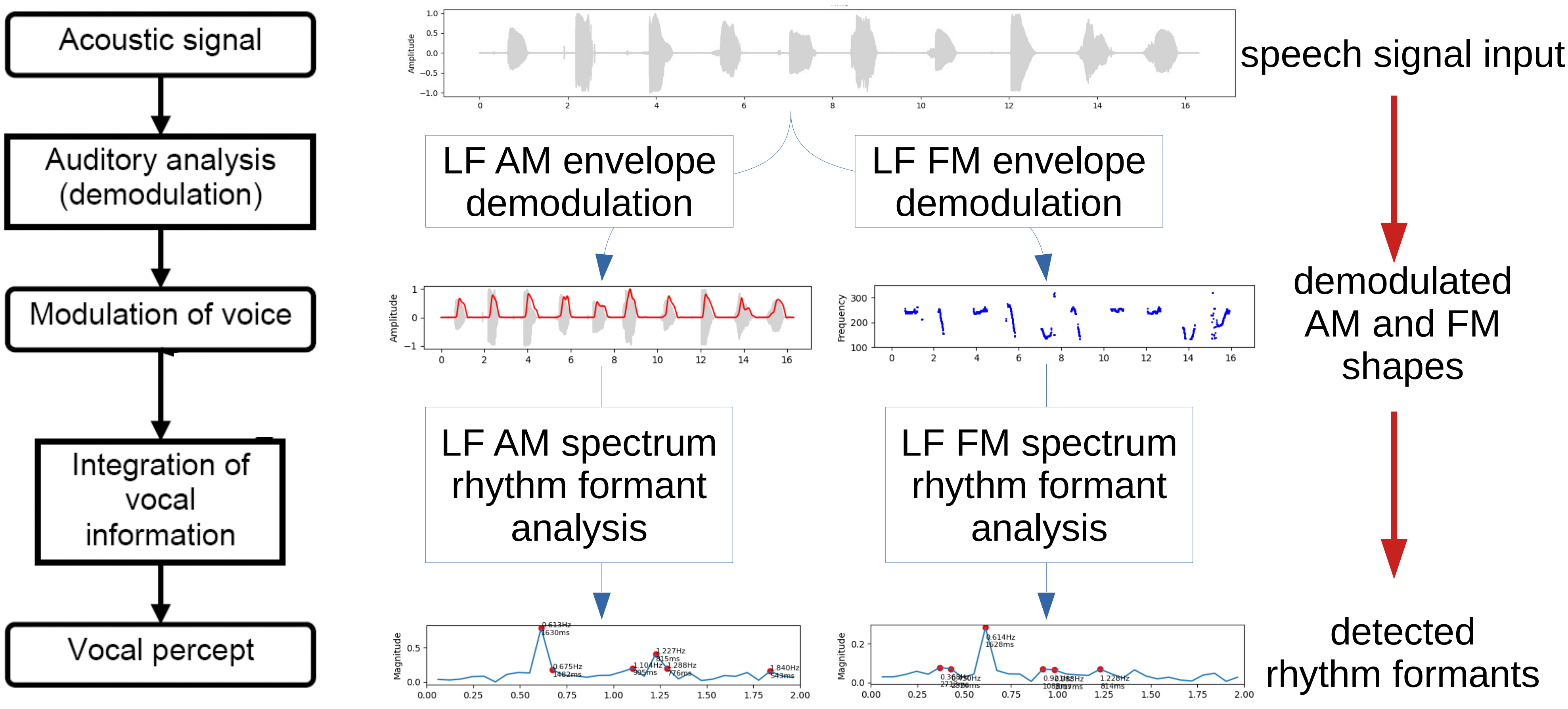


AM and FM demodulation and detection of rhythm



Hartmut Traunmüller (2007) "Demodulation, mirror neurons and audiovisual perception nullify the motor theory" Contr. to Fonetik 2007, TMH-QPSR 50: 17-20. Detpt. of Speech, Music and Hearing, Royal Inst. of Technology, Stockholm.

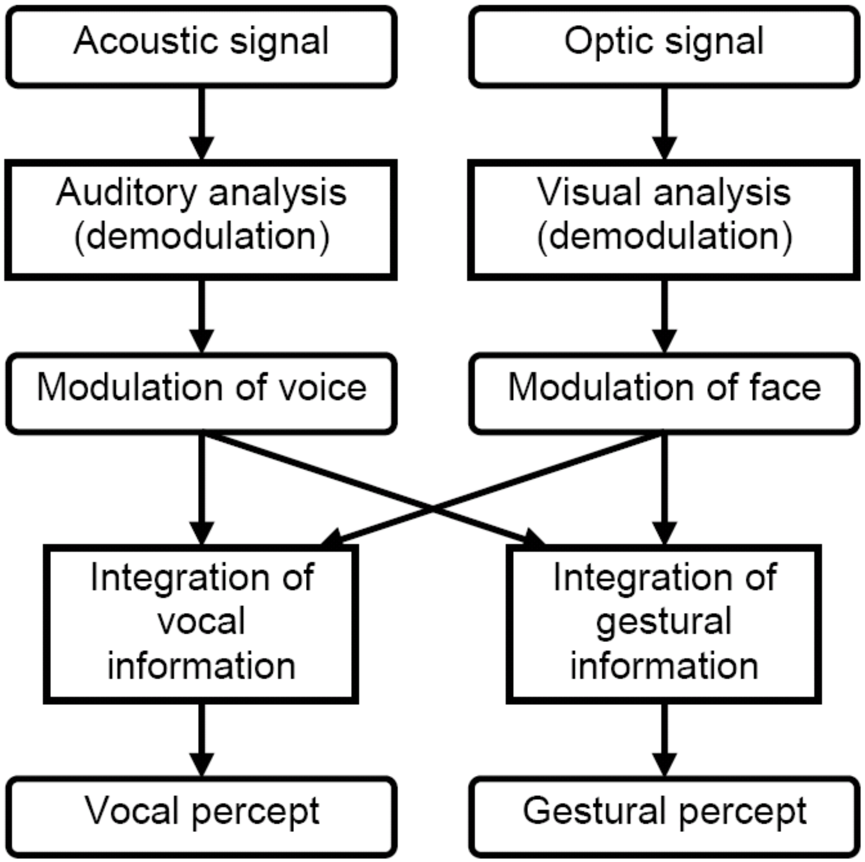
Comparison with Traunmüller's demodulation model



Hartmut Traunmüller (1994) "Conventional, biological, and environmental factors in speech communication: A modulation theory" Phonetica 51: 170-183. doi (Also in PERILUS XVIII: 92-102.)

Hartmut Traunmüller (2007) "Demodulation, mirror neurons and audiovisual perception nullify the motor theory" Contr. to Fonetik 2007, TMH-QPSR 50: 17-20. Detpt. of Speech, Music and Hearing, Royal Inst. of Technology, Stockholm.

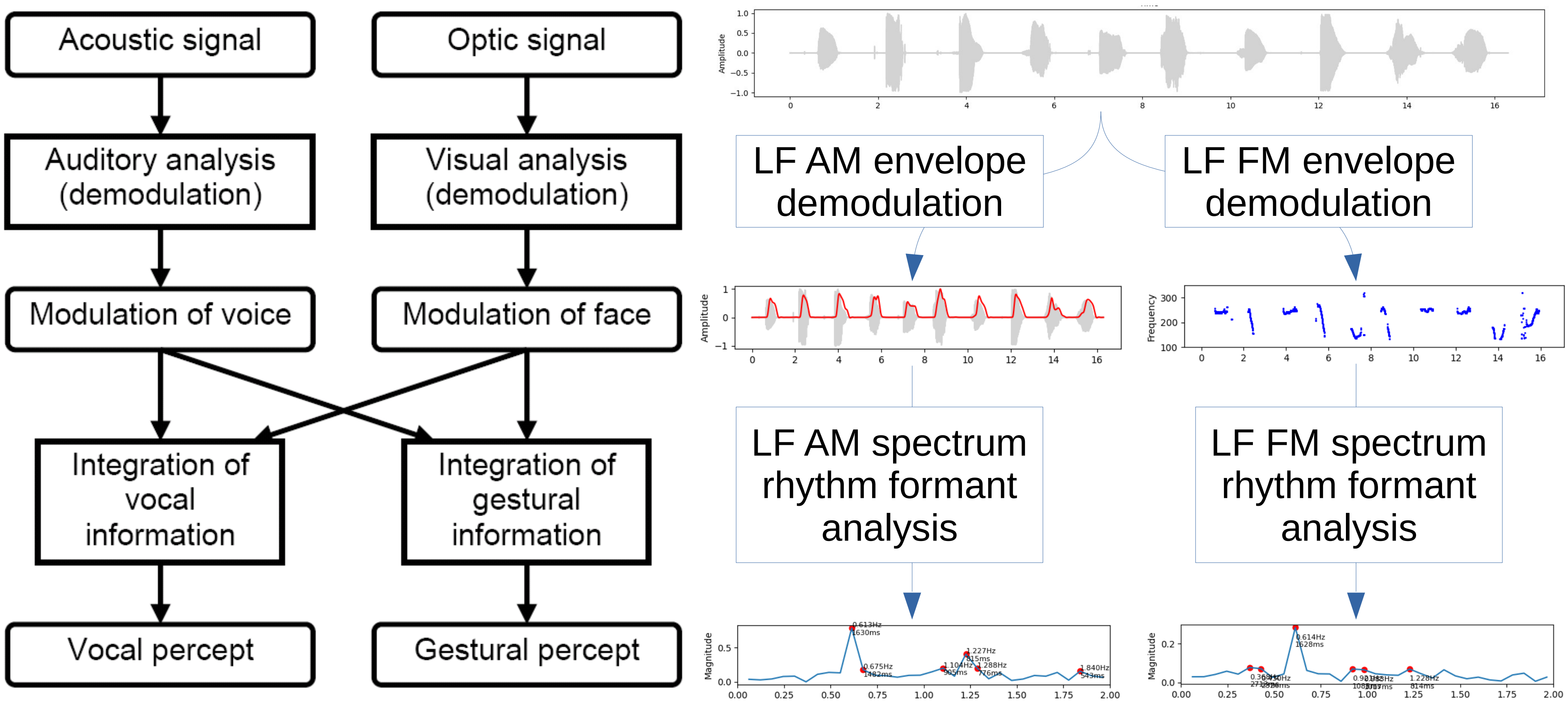
Traunmüller: audiovisual perception (2007)



Hartmut Traunmüller (1994) "Conventional, biological, and environmental factors in speech communication: A modulation theory" Phonetica 51: 170-183. doi (Also in PERILUS XVIII: 92-102.)

Hartmut Traunmüller (2007) "Demodulation, mirror neurons and audiovisual perception nullify the motor theory" Contr. to Fonetik 2007, TMH-QPSR 50: 17-20. Detpt. of Speech, Music and Hearing, Royal Inst. of Technology, Stockholm.

Traunmüller: audiovisual perception (2007)

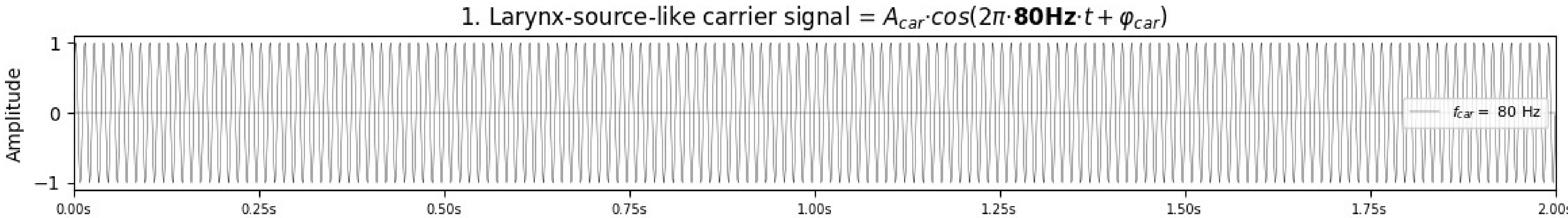


Hartmut Traunmüller (1994) "Conventional, biological, and environmental factors in speech communication: A modulation theory" Phonetica 51: 170-183. doi (Also in PERILUS XVIII: 92-102.)

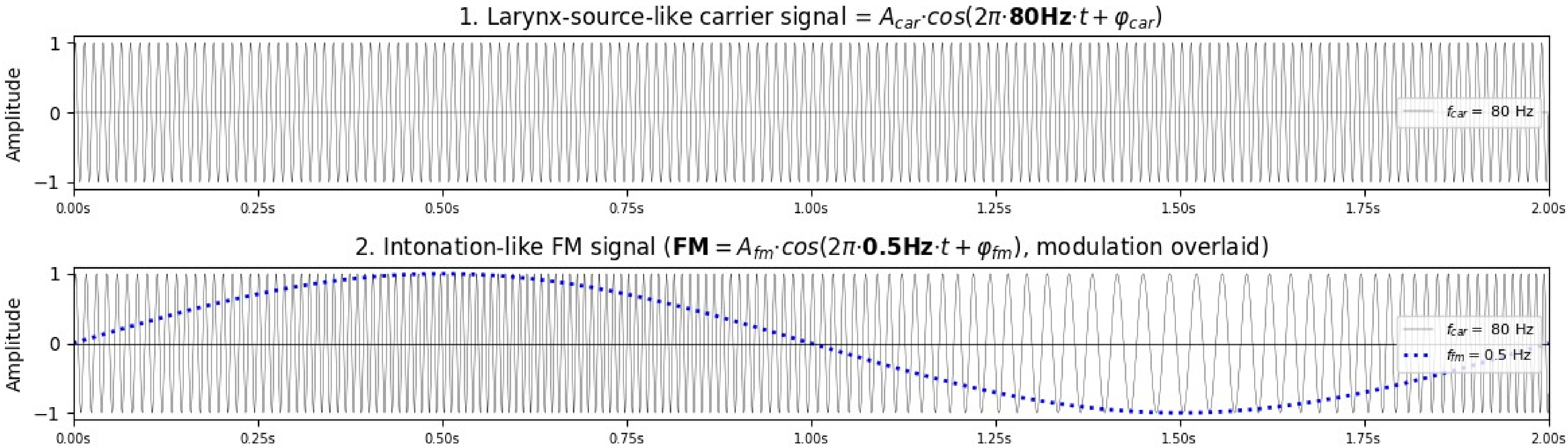
Hartmut Traunmüller (2007) "Demodulation, mirror neurons and audiovisual perception nullify the motor theory" Contr. to Fonetik 2007, TMH-QPSR 50: 17-20. Detpt. of Speech, Music and Hearing, Royal Inst. of Technology, Stockholm.

AM and FM modulation step by step

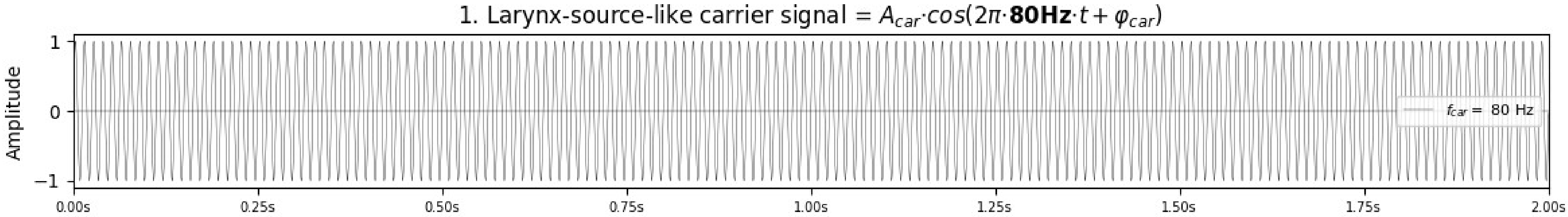
Modulation: carrier signal

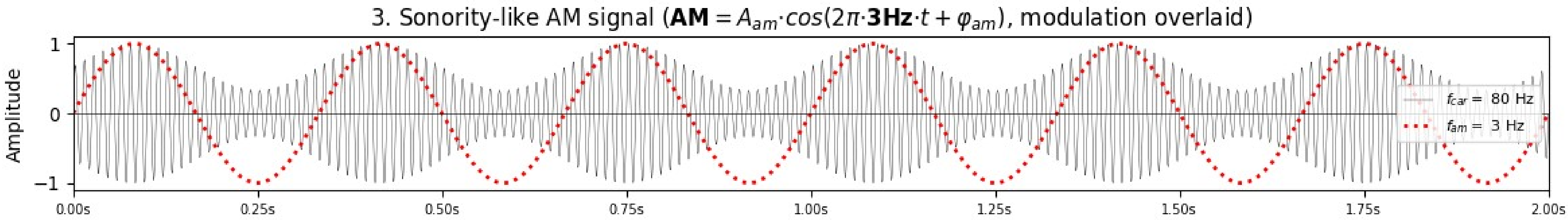


Modulation: FM signal with low frequency information



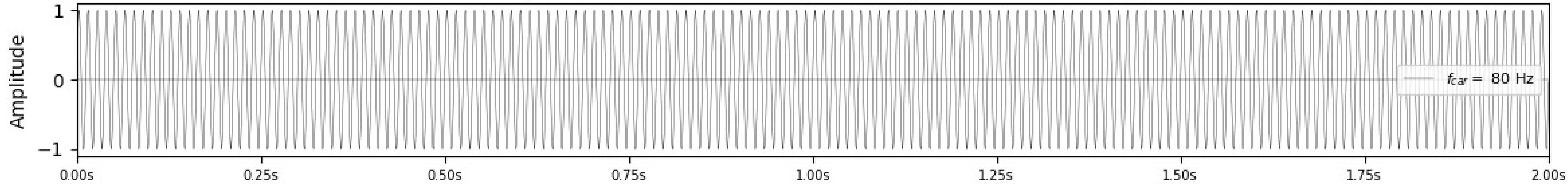
Modulation: AM signal with low frequency information



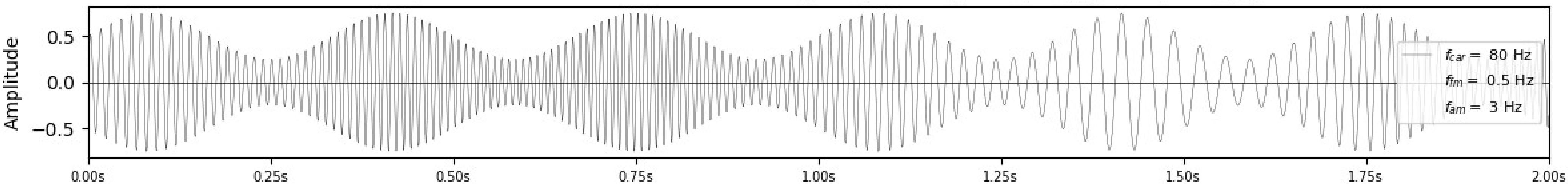


Modulation Theory

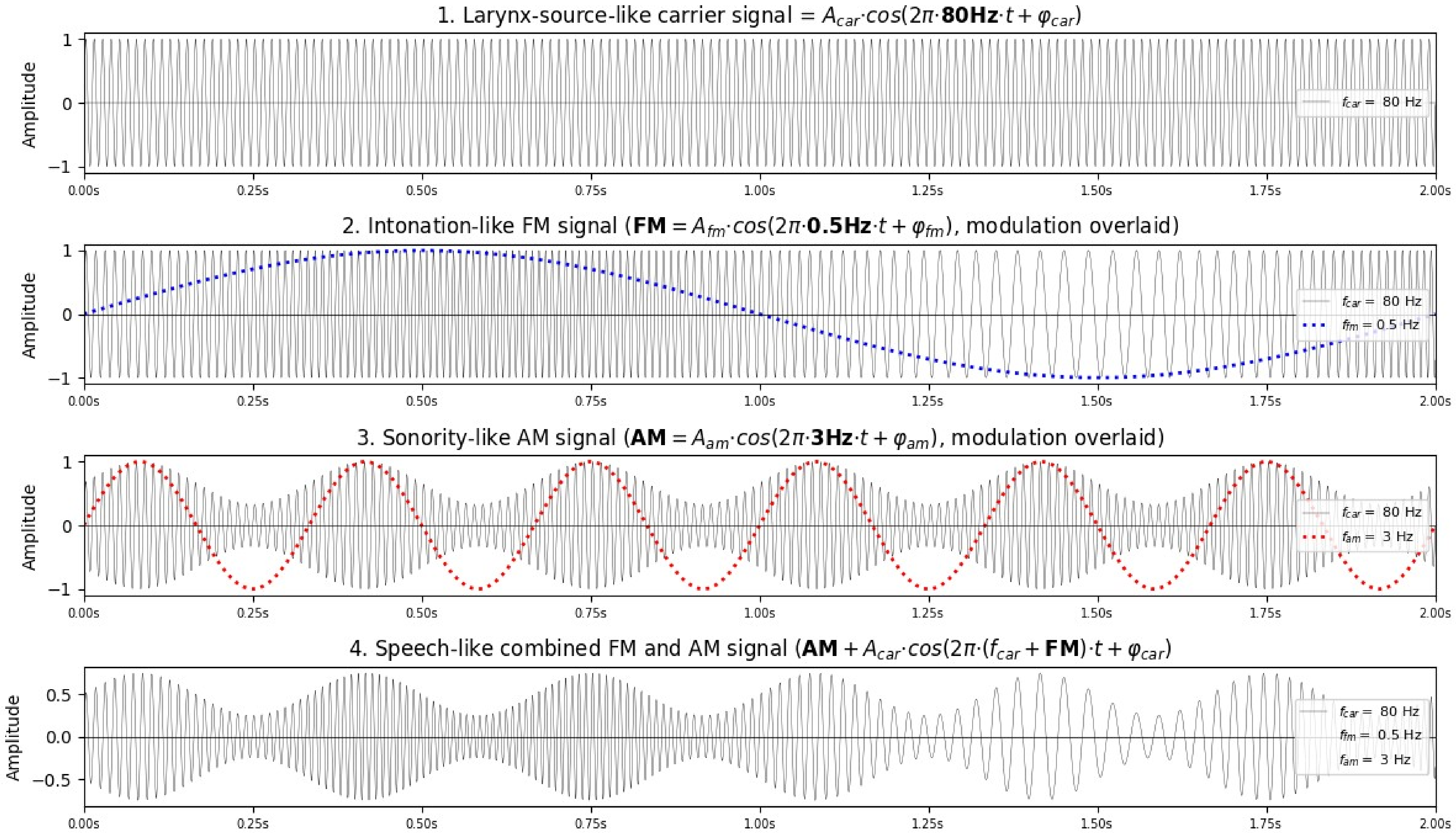
1. Larynx-source-like carrier signal = $A_{car} \cdot cos(2\pi \cdot \mathbf{80Hz} \cdot t + \varphi_{car})$



4. Speech-like combined FM and AM signal $(\mathbf{AM} + A_{car} \cdot cos(2\pi \cdot (f_{car} + \mathbf{FM}) \cdot t + \varphi_{car})$



Modulation Theory



Demodulation and analysis procedures

- Time domain procedures:
  - Envelope extraction
  - Fundamental frequency estimation (‘pitch’ extraction)
- Frequency domain procedures:
  - Spectral analysis
  - Spectrogram analysis
  - there are also frequency domain procedures for F0 estimation
- Comparison using distance metrics
  - distance calculation with different distance metrics
  - hierarchical clustering with distance and different clustering criteria
- Output:
  - Graphical display
  - Numerical files and figure files

FM Demodulation

FM Demodulation – F0 estimation (‘pitch’ extraction)

There are many algorithms for F0 estimation, for example:

Time domain algorithms:

autocorrelation (AC), average magnitude difference function (AMDF), average squared difference function (ASDF) ...

Frequency domain algorithms:

harmonic peak detection, spectral comb, ...

The AMDF algorithm:

1. Divide the speech signal into equal time frames.
2. Make a copy of the first frame, noting the start position.
3. Move the copy through the first frame:
   - compare with the signal at each point
   - save the differences in a list
4. Find the first smallest difference in the list:
   - find its position in the signal
   - find the fundamental period (P0) by subtracting the start position from this position and divide by the sampling frequency.
   - then the fundamental frequency in this frame is: F0 = 1/P0
5. Move to the next frame and repeat until the last frame.

FM Demodulation with AMDF

Average Magnitude Difference Function

Move a window (frame) through the signal, at each step:

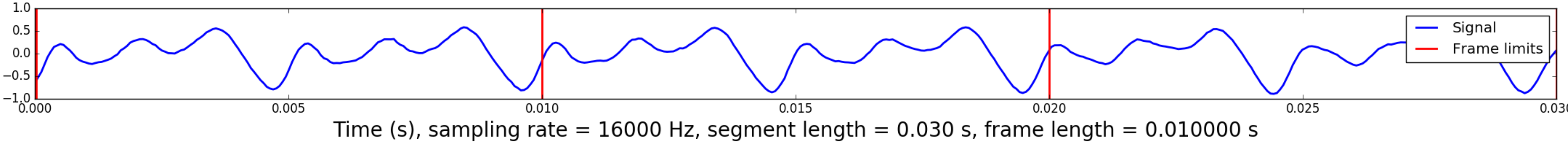
Copy the frame.

Move the copy through the frame, at each step:
Subtract the copy from the frame position by position.
Average (or sum) the absolute (or squared) differences.

Find the position of the first smallest difference for the frame.
Calculate period from start of frame to this position.

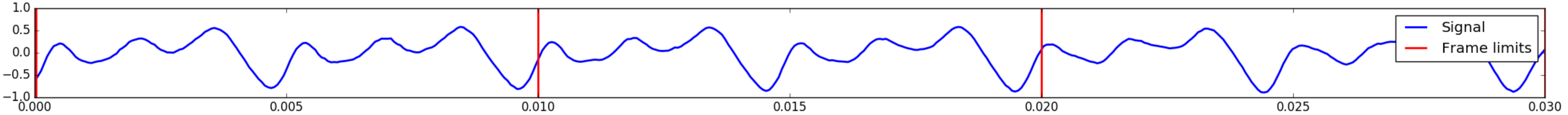
Convert period to frequency (estimation of F0).
Collect the frequencies (estimation of F0 track).

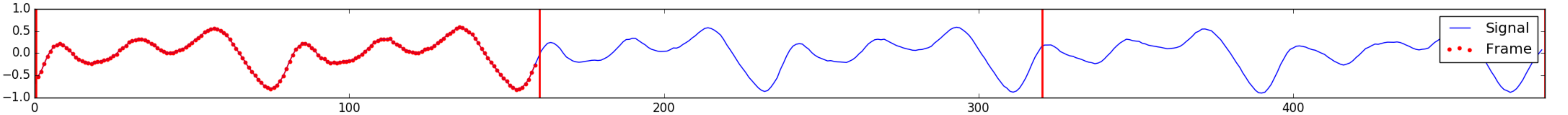
For all algorithms: divide the signal into equal time frames



The duration of the time frame depends on the lowest frequency to be measured.

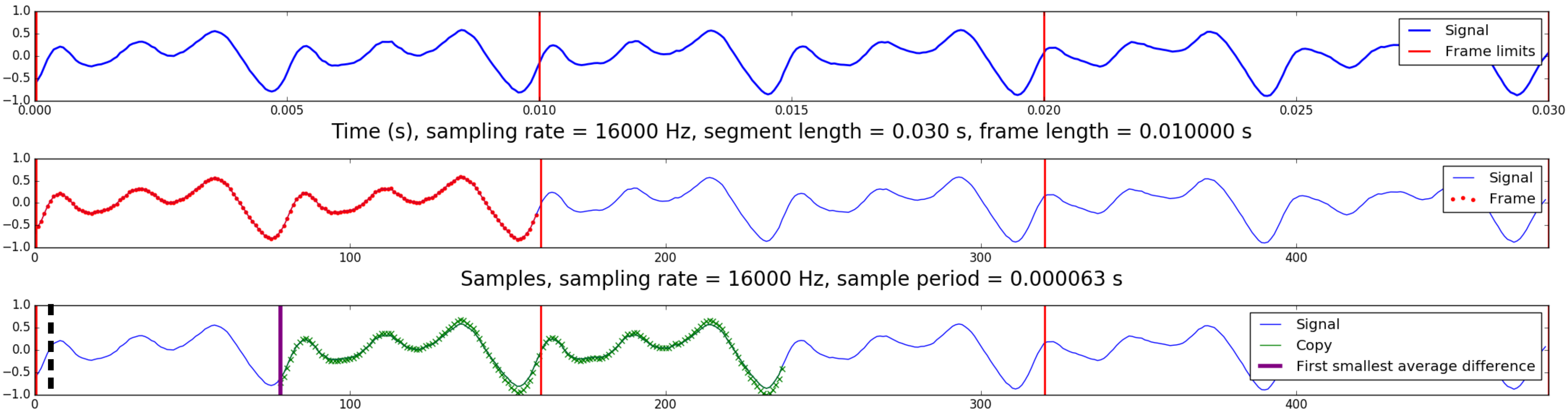
AMDF: make a copy of the first time frame





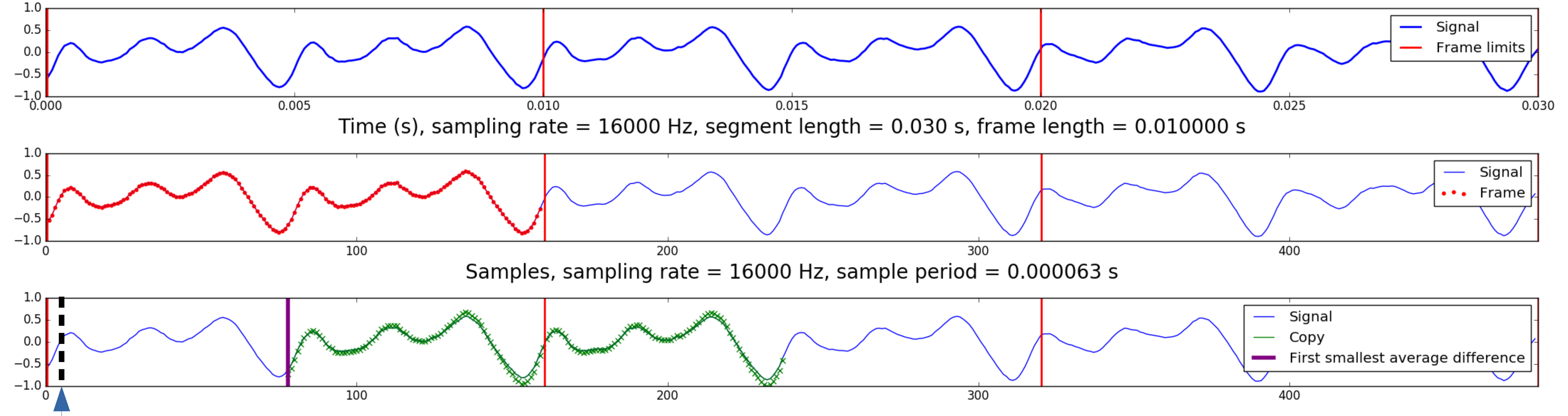
Note the start position of the time frame in the signal.

AMDF: move copy through first time frame



1. Compare the copy with the signal point by point at each position in the frame
2. Save each difference in a list, together with its current position in the frame
3. When finished with comparisons at all positions in the frame:
   search the list for the smallest difference with the copy and its position.

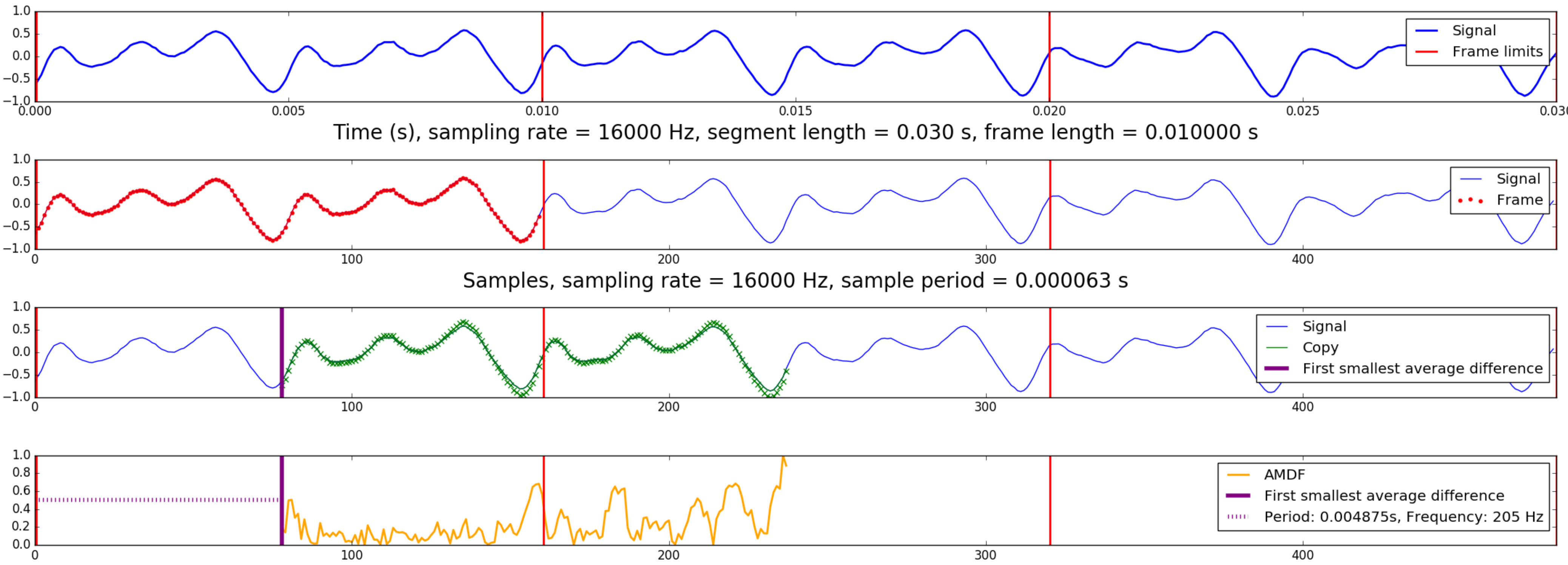
AMDF: move the copy through the first frame to the end



1. Compare the copy with the signal point by point at each position in the frame
2. Save each difference in a list, together with its current position in the frame
3. When finished with comparisons at all positions in the frame:
   search the list for the smallest difference with the copy and its position.

In practice, comparison of the copy with the signal starts with an offset slightly after the first position in the frame otherwise the smallest difference would always be zero! The position of the offset depends on the highest frequency to be measured.

AMDF: calculate differences, minimal difference, T, F0



1. Note the position of the minimal difference between copy and signal
2. Calculate time period *T* of the frame as the difference between
   - the beginning of the frame and
   - the position of the minimal difference

     (in this case: 0.004875 s, i.e. 4.875 ms) divided by the sampling frequency *fs*
3. Calculate the frequency from the period: *F0* = 1 / *T*

   (in this case: 1 / 0.04875 = 205 Hz)

Move to the next frame and repeat the procedure for the remaining frames

Definition of AMDF

τ is the lag, a period for which the average differences are calculated.

$$D(\tau) = \frac{1}{N-1-\tau} \sum_{n=0}^{N-1-\tau} \left| x(n) - x(n+\tau) \right|, 0 \leq \tau \leq N-1$$

Definition of AMDF

τ is the lag, a period for which the average differences are calculated.

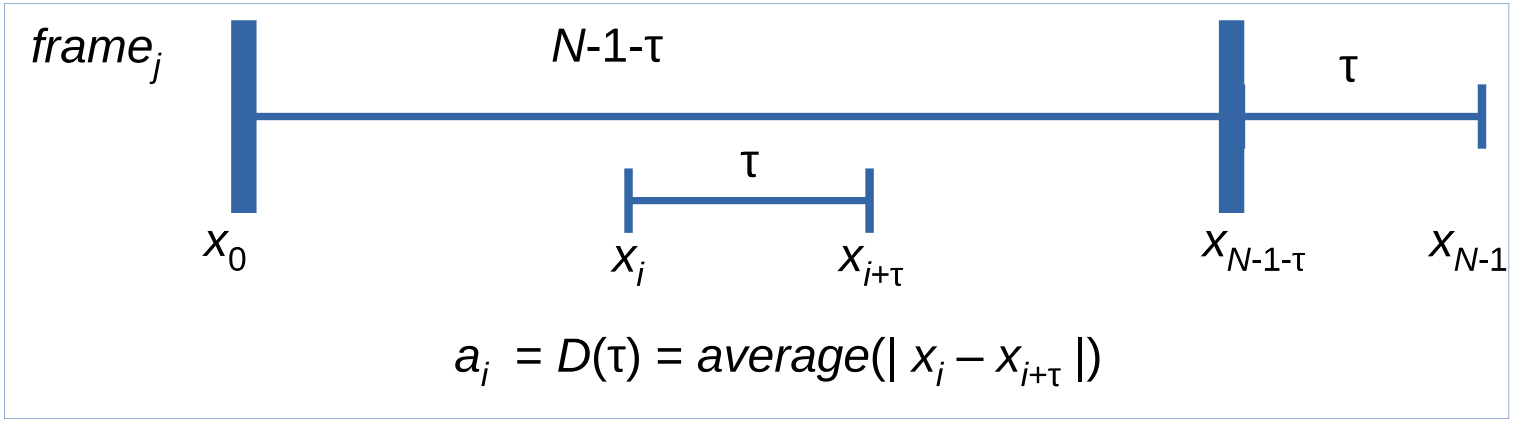
$$D(\tau)=\frac{1}{N-1-\tau}\sum_{n=0}^{N-1-\tau}\left|x(n)-x(n+\tau)\right| \quad 0\leq\tau\leq N-1$$

Does this remind you of the *nPVI* and the Manhattan Distance metric?

Definition of AMDF

τ is the lag, a period for which the average differences are calculated.

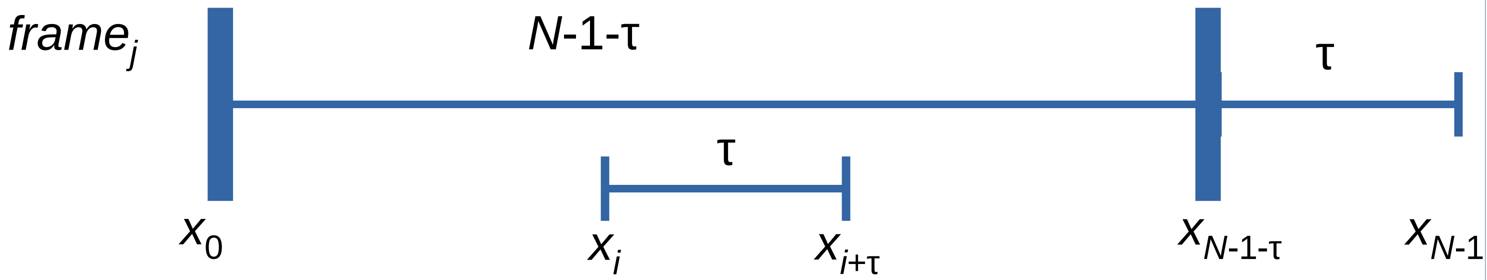
$$D(\tau)=\frac{1}{N-1-\tau}\sum_{n=0}^{N-1-\tau}\left|x(n)-x(n+\tau)\right|, 0\leq\tau\leq N-1$$



Frame by frame F0 estimation with AMDF

τ is the lag, a period for which the average differences are calculated.

$$D(\tau)=\frac{1}{N-1-\tau}\sum_{n=0}^{N-1-\tau}\left|x(n)-x(n+\tau)\right|, 0\leq\tau\leq N-1$$

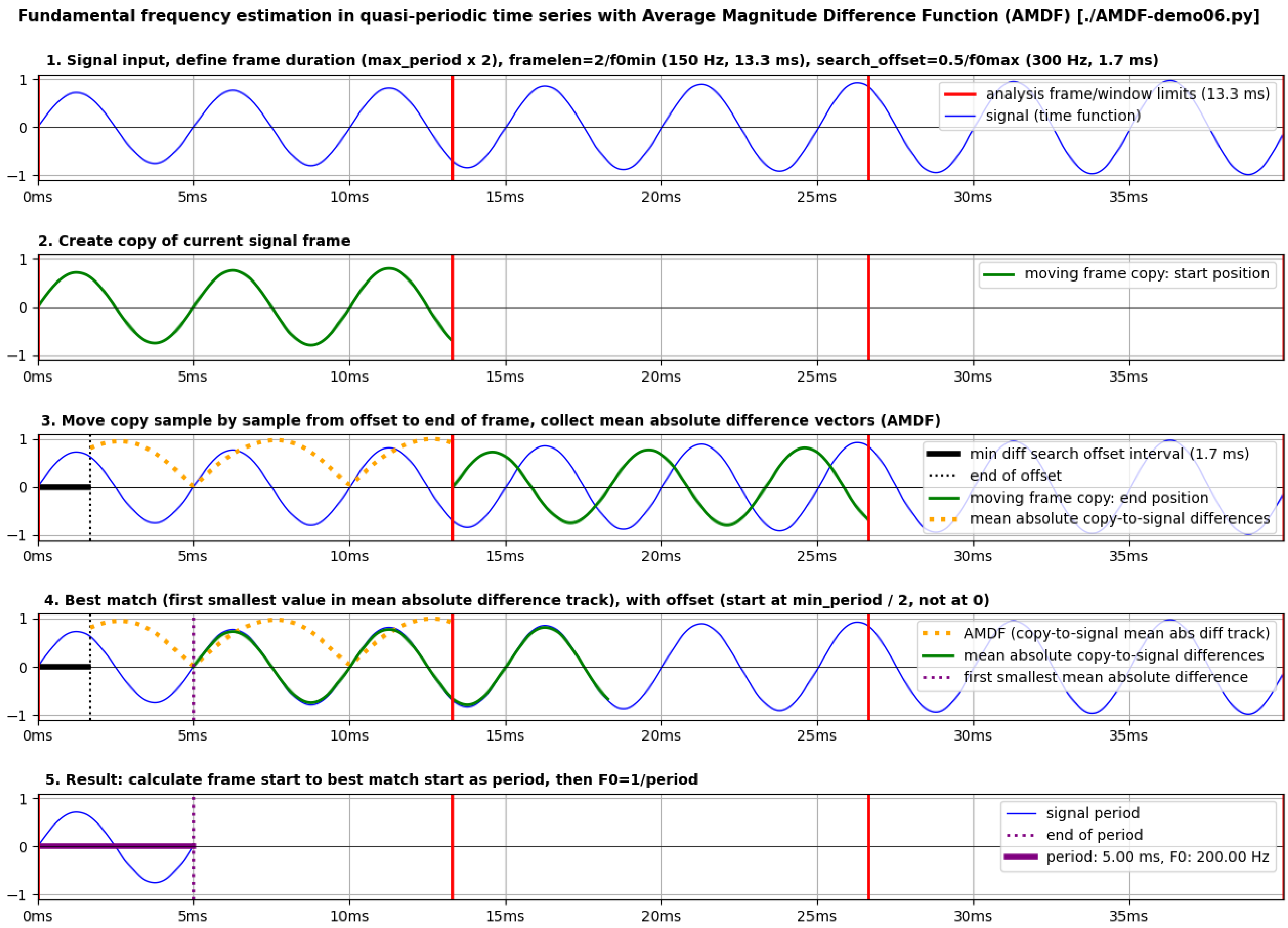


$a_i = D(\tau) = average(|\ x_i - x_{i+\tau}\ |)$

1. Collect the *average* of all subtractions of signal magnitudes in each period $x_i - x_{i+\tau}$
2. Find the *position* of the *first smallest* average distance for all steps in $frame_j$:

$$pos_j = position(first(smallest(a_{0+offset}, \ldots, a_{N-1-\tau})))$$

3. Calculate the difference (number of steps) p_j between pos_0, the start of $frame_j$ and pos_j.
4. Divide the difference by the sampling frequency to get the period in seconds.
5. Calculate the F0 estimation for $frame_j$ as the inverse of the period: $1/p_j$.
6. Do the same for all frames through the signal.


Fundamental frequency estimation in quasi-periodic time series with Average Magnitude Difference Function (AMDF) [./AMDF-demo06.py]
1. Signal input, define frame duration (max_period x 2), framelen=2/f0min (150 Hz, 13.3 ms), search_offset=0.5/f0max (300 Hz, 1.7 ms)
analysis frame/window limits (13.3 ms)
signal (time function)
2. Create copy of current signal frame
moving frame copy: start position
3. Move copy sample by sample from offset to end of frame, collect mean absolute difference vectors (AMDF)
min diff search offset interval (1.7 ms)
end of offset
moving frame copy: end position
mean absolute copy-to-signal differences
4. Best match (first smallest value in mean absolute difference track), with offset (start at min_period / 2, not at 0)
AMDF (copy-to-signal mean abs diff track)
mean absolute copy-to-signal differences
first smallest mean absolute difference
5. Result: calculate frame start to best match start as period, then F0=1/period
signal period
end of period
period: 5.00 ms, F0: 200.00 Hz
1
0
−1
0ms
5ms
10ms
15ms
20ms
25ms
30ms
35ms

Now check the code!