Computational Methods in Phonetics

Visualisation of Waveforms and Frequencies
2019-07-17, 14:30-16:30

Dafydd Gibbon
Bielefeld University
Jinan University, Guangzhou
Types of Computing in Phonetics

Consumer computation
- Praat, Audacity
- Calc, Excel, SPSS

Scripting and Programming
- Praat scripting
- R, Stata, Matlab
- Python

Software development
- C
- Java
- Python
Domain of investigation: Prosody

- The **rhythms** and **melodies** of spoken language
- The focus here is on rhythm more than on melody

Scientific results depend on scientific methods:

- Cognitive or hermeneutic methods:
  - direct observation and analysis of speech based on understanding of speech
  - Classic methods in linguistics
- Physical methods:
  - Physical methods for analysis of speech production, transmission and perception
  - Visualisation
Theses: YARD (Yet Another Discussion of Rhythm)

There are many rhythms in speech (and in music):
- Rhythms are oscillations, not sequences of isochronous states
- Rhythmic oscillations are essentially linear (‘finite state’)  
- Rhythms occur in time domains of different sizes, each of them linear
- The time domains are associated with ranks of units of speech/language from discourse to phoneme

Rhythms appear to be hierarchical, but the hierarchy
- has limited depth
- has linear layers with iterative cycles
- is not a general recursive hierarchy
Linear Layered Discourse Rhythms

57 seconds of a BBC news broadcast, from Aix-MARSEC corpus (A0101B)

Two Frequency zones:
1) Approximately 120 – 220 Hz
2) Approximately 300 – 400 Hz (on ‘paratone’ onsets)

Finite depth linear 2-cycle iterative layered structure
(FS, cf. Pierrhumbert, not a recursive hierarchy)
• Two independently motivated but structurally dependent linear layers
  (like hours & minutes on a clock)
• Each gradually declining and periodically resetting
Rhythm is oscillation

Rhythm is a marker of syntagmatic relations

- Rhythms mark syntagmatic relations in sequences of units which belong together

Levels of abstraction:

1) Abstract oscillation: iterations, modelled by ‘loops’ in Finite State Machines
   - Intonation
     - Pierrehumbert’s model, with added loops
   - Tone
     - Gibbon’s model of Niger-Congo 2-level terraced tone
     - Jansche’s model of Tianjin tone

2) Physical oscillation of the amplitude of speech
   - Syllables
   - Phrases
   - ...
Abstract oscillation
**Phonological iteration as abstract oscillation**

Pierrehumbert’s regular grammar / finite state transition network

- **Empirical overgeneration**
  1. Accents in a sequence tend to be all H* or all L*
  2. Global contours tend to be rising with L* accents, falling with H* accents
  3. Global contours may span more than 1 turn

- **Empirical undergeneration**
  1. Paratone hierarchy not included
  2. No time constraints

Not the first (cf. Reich, ’t Hart et al., Fujisaki, …)
But linguistically the most interesting.
**Phonological iteration as abstract oscillation**

Niger-Congo Iterative Tonal Sandhi  
(the most general case)

At the most abstract level, just one node with H and L cycling around it.

From an allotonic point of view:

- 3 cycles
- 1-tape (1-level) transition network
Phonological iteration as abstract oscillation

Niger-Congo Iterative Tonal Sandhi
(the most general case)

From an allotonic point of view:

- 3 cycles
- 2-tape (= 2-level) transition network
Phonological iteration as abstract oscillation

Niger-Congo Iterative Tonal Sandhi
(the most general case)

From phonetic signal processing point of view:

- 3 cycles
- 3-tape (= 3-level) transition network
Phonological iteration as abstract oscillation

Tianjin Dialect Iterative Tonal Sandhi

Martin Jansche 1998
Tianjin Mandarin tone sandhi
Physical oscillation
**Physical oscillation**

Oscillation means that intervals – cycles – tend to have the same length (duration).

But is it enough to focus only on duration of vowels, syllables, etc., using annotation mining?

What about the alternation feature of rhythm?

- 1-dimensional annotation mining of time-stamp durations
- 2-dimensional annotation mining of time-stamp durations
- 3-dimensional annotation mining of time-stamp durations
One-dimensional because the result of the analysis is a single scale. The results are all comparable to variance or standard deviation, but differ in detail.

For example, with the \( nPVI \), subtraction is between neighbouring data in a moving window (so a kind of AMDF, Average Magnitude Difference Function), not between mean and data, thus factoring out tempo variations to some extent.

\[
\text{Variance} \left( x_{1\ldots n} \right) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

\[
PIM \left( x_{1\ldots n} \right) = \sum_{i \neq j} \left| \log \frac{I_i}{I_j} \right|
\]

\[
PFD \left( d_{1\ldots n} \right) = \frac{\sum_{i=1}^{n} \left| d - d_i \right|}{\sum_{j=1}^{n} d_j} \times 100
\]

\[
nPVI \left( d_{1\ldots n} \right) = \frac{\sum_{k=1}^{n} \left| d_k - d_{k+1} \right|}{\left( \frac{d_k + d_{k+1}}{2} \right)} \times 100
\]

(or Standard Deviation)

where \( I_{i,j} \) are intervals in a given sequence

where \( d \) is typically the duration of a foot

\( d \) refers to duration of vocalic segment, syllable or foot, typically
### 1-dimensional annotation mining of time-stamp durations

One-dimensional because the result of the analysis is a single scale. The results are all comparable to variance or standard deviation, but differ in detail.

For example, with the \( nPVI \), subtraction is between neighbouring data in a moving window (so a kind of AMDF, Average Magnitude Difference Function), not between mean and data, thus factoring out tempo variations to some extent.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \text{Variance}(x_1\ldots n) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} ]</td>
<td>(or Standard Deviation)</td>
</tr>
<tr>
<td>[ \text{PIM}(x_1\ldots n) = \sum_{i \neq j} \left</td>
<td>\log \frac{I_i}{I_j} \right</td>
</tr>
<tr>
<td>[ \text{PFD}(d_1\ldots n) = \frac{\sum_{i=1}^{n}</td>
<td>\bar{d} - d_i</td>
</tr>
<tr>
<td>[ \text{nPVI}(d_1\ldots n) = \frac{\sum_{k=1}^{k-1}</td>
<td>d_k - d_{k+1}</td>
</tr>
</tbody>
</table>
Isochrony as variance: Roach

Textual description hard to figure out, but maybe ...

\[
\text{Mean Foot Length (MFL)} = \frac{\sum_{i=1}^{n} |\text{foot}_i|}{n}
\]

\[
\text{Percentage Foot Deviation (PFD)} = 100 \times \frac{\sum |\text{MFL} - \text{len(foot}_i)|}{n \times \text{MFL}}
\]

Ignore syllables before initial and after final stresses

Calculate:

- \textit{average length} of interstress interval / foot (MFL)
- \textit{percentage deviation} of each interval from MFL, maybe ...
  \[100 \times \frac{\text{mean-interval}_i}{\text{mean}}\]
- \textit{variance} of percentage deviations (?)

This is a global measure:

ignores \underline{alternation} and \underline{iteration} criteria
One-dimensional because the result of the analysis is a single scale. The results are all comparable to variance or standard deviation, but differ in detail.

For example, with the \( nPVI \), subtraction is between neighbouring data in a moving window (so a kind of AMDF, Average Magnitude Difference Function), not between mean and data, thus factoring out tempo variations to some extent.

\[
\text{Variance}(x_{1\ldots n}) = \frac{\sum_{1}^{n} (x_i - \bar{x})^2}{n-1}
\]

\[
PIM(x_{1\ldots n}) = \sum_{i \neq j} \left| \log \frac{I_i}{I_j} \right|
\]

\[
PFD(d_{1\ldots n}) = \frac{\sum_{i=1}^{n} |d - d_i|}{\sum_{j=1}^{n} d_j} \times 100
\]

\[
nPVI(d_{1\ldots n}) = \frac{\sum_{k=1}^{k-1} \left| d_k - d_{k+1} \right|}{(d_k + d_{k+1})/2} \times 100
\]

(or Standard Deviation)

where \( I_{i,j} \) are intervals in a given sequence

where \( d \) is typically the duration of a foot

\( d \) refers to duration of vocalic segment, syllable or foot, typically
Isochrony as ratio: Scott et al.

The Rhythmic Irregularity Measure (RIM) for individual utterances calculates the sum of the ratios of each interval to each other interval.

Perfect isochrony: RIM = 0; non-isochrony is an open-ended log function.

RIM applies to utterances of the same length:

Scott & al. suggest generalising the RIM by dividing by $n$ for interval sequences of length $n$.

This is incorrect: the RIM calculates a (triangular) matrix so a generalised RIM must be divided by $n^2$.

RIM is designed to be “symmetric”:

RIM therefore just measures isochrony, not rhythm, as it ignores rhythm alternation and iteration.
1-dimensional annotation mining of time-stamp durations

One-dimensional because the result of the analysis is a single scale. The results are all comparable to variance or standard deviation, but differ in detail.

For example, with the $nPVI$, subtraction is between neighbouring data in a moving window (so a kind of AMDF, Average Magnitude Difference Function), not between mean and data, thus factoring out tempo variations to some extent.

<table>
<thead>
<tr>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Variance}(x_1...n) = \frac{\sum^n_{i=1} (x_i - \bar{x})^2}{n-1}$</td>
<td>(or Standard Deviation)</td>
</tr>
<tr>
<td>$PIM(x_1...n) = \sum_{i \neq j} \left</td>
<td>\log \frac{I_i}{I_j} \right</td>
</tr>
<tr>
<td>$PFD(d_1...n) = \frac{\sum^n_{i=1}</td>
<td>d_i - d</td>
</tr>
<tr>
<td>$nPVI(d_1...n) = \frac{\sum_{k=1}^{k-1}</td>
<td>d_k - d_{k+1}</td>
</tr>
</tbody>
</table>
Isochrony as local distance: Grabe & al.

\[ nPVI = 100 \times \sum_{k=1}^{m-1} \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} / (m - 1) \]

Normalises locally between neighbouring intervals for speech rate, using a distance measure:
\[ \text{DISTANCE}_i = | \text{INT}_i - \text{INT}_{i+1} | / \text{AVG} (\text{INT}_i, \text{INT}_i) \]
\[ \text{PVI} = 100 \times \text{AVG} (\text{DISTANCE}) \] (range 0...200, asymptote)

Problems:

Magnitude operation:
If PVI = 0, then isochrony holds – this is ok.
But if PVI ≠ 0, then intervals are somehow irregular, use of the absolute value means many sequences (increasing, decreasing, mixed, non-binary, ...) may have the same PVI

Binary comparison (supposes iambs/trochees?), but
Spondaic: *That big black bear swam fast past Jane’s boat.*
Dactylic: *Jonathan Appleby trundled along with a tune on his lips.*
The models should at least correlate...
... but they don’t correlate too well
Ramus:

Figure 3. The measure $\%V$ is plotted on the y-axis, in reverse order. The standard deviation of intervocalic intervals $\Delta C$ is given on the x-axis.
Interval duration approaches and typology

Grabe & al.:
Daniel Hirst pointed out that these relations can be obtained phonologically, by comparing the phonotactics of the languages, rather than phonetically, with measurements.
Summary: 1-dimensional interval duration approaches

There are many other interval duration measures perhaps most prominently in the past 5 years the non-isochronous Ramus model: $\Delta C \times \%V$

Isochrony/irregularity is not a sufficient condition:

cf. Cummins (2002) on Ramus:

\begin{quote}
Where is the bom-di-bom-bom in $\%V$?
\end{quote}

Interval duration isochrony approaches ignore the ordering and directionality, of rhythm, alternation within Rhythm Units and iteration of Rhythm Units.

And

The interval duration approaches assume the relevant event is duration of segmental constructs which it may or may not be
A 2-dimensional interval duration approach
2-dimensional annotation mining of time-stamp durations

Wagner:

- addresses the absolute magnitude problem

- plots $duration_i \times duration_{i+1}$

- creates typologically interpretable clusters

---

**English**

**French**

**Polish**

<table>
<thead>
<tr>
<th>color</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>green</td>
<td>stressed x unstressed</td>
</tr>
<tr>
<td>blue</td>
<td>unstressed x stressed</td>
</tr>
<tr>
<td>red</td>
<td>phrase-final</td>
</tr>
</tbody>
</table>
Following Wagner, two-dimensional duration relations are represented in a z-scored scatter plot, not as a single scale.
Result, visualising the scale in two dimensions:

- Mandarin: means are scattered relatively evenly around the centre
- English: e.g. count(short-short) > count(long-long), not binary!

A 3-dimensional interval duration approach
3-dimensional annotation mining of time-stamp durations

General strategy:
  • take the local difference/distance measure from the PVI
  • do not throw directionality away by taking absolute values of differences
  • but use directionality (polarity) to determine grouping

Specific procedure:
  • using annotation time-stamps, recursively build tree structures (Time Trees):
    • iambic parametrisation:
      if right neighbour is stronger,
        then group
      else stack and wait for a stronger right neighbour
    • trochaic parametrisation:
      if right neighbour is stronger,
        then group
      else stack and wait for a weaker right neighbour
A tiger and a mouse were walking in a field when they saw a big lump of cheese lying on the ground. The mouse said: "Please, tiger, let me have it. You don't even like cheese. Be kind and find something else to eat." But the tiger put his paw on the cheese and said: "It's mine! And if you don't go I'll eat you too." The mouse was very sad and went away.

The tiger tried to swallow all of the cheese at once but it got stuck in his throat and whatever he tried to do he could not move it. After a while, a dog came along and the tiger asked it for help. "There is nothing I can do." said the dog and continued on his way. Then, a frog hopped along and the tiger asked it for help. "There is nothing I can do." said the frog and hopped away.

Finally, the tiger went to where the mouse lived. She lay in her bed in a hole which she had dug in the ground. "Please help me," said the tiger. "The cheese is stuck in my throat and I cannot remove it." "You are a very bad animal," said the mouse. "You wouldn't let me have the cheese, but I'll help you nonetheless. Open your mouth and let me jump in. I'll nibble at the cheese until it is small enough to fall down your throat." The tiger opened his mouth, the mouse jumped in and began nibbling at the cheese. The tiger thought: "I really am very hungry."
Interpreting Time Trees
Two-dimensional annotation mining of time-stamp durations

Three-dimensional because alternative trees are possible, depending on the algorithm settings:
- binary/nonbinary, lower/higher percolated
- related to phrasal and discourse patterns

Induction of compositional tree structures

But the annotation mining approaches ...

- are based on cognitive filtering by linguists (for example with annotated time-stamps, Praat etc.)
- model static durations
- use pairwise relations
- focus on *isochrony*, equal timing of durations

- ignore the essential property of *rhythm*, which is *alternation of units with approximately equal duration*
- in other words: *oscillation*

There is an alternative: speech signal analysis
Rhythms in Amplitude AND Frequency Modulation
Rhythm as Physical Oscillation

Amplitude Modulation and Frequency Modulation
Which information? Semantics
Which structure? Grammar
Which code? Modulations
INFORMATION

SPECTRAL ANALYSES in different frequency zones, COORDINATION

AMPLITUDE DEMODULATION in different time zones
rectification, LP filtering
envelope detection

FREQUENCY DEMODULATION in different frequency zones
pitch tracking, formant tracking
INFORMATION

- NOISE FREQUENCIES
- FREQUENCY MODULATION
- CARRIER FREQUENCY

AMPLITUDE MODULATION
- FILTER COEFFICIENTS

INFORMATION

- SPECTRAL ANALYSES in different frequency zones, COORDINATION

AMPLITUDE DEMODULATION
- in different time zones
- rectification, LP filtering
- envelope detection

FREQUENCY DEMODULATION
- in different frequency zones
- pitch tracking, formant tracking
1. Waveform

2. Amplitude Envelope (Emulation of Absolute Hilbert Transform)

3. Fundamental Frequency (F0, 'pitch') estimate, AMDF algorithm

1. Waveform

2. Amplitude Envelope (Emulation of Absolute Hilbert Transform)

3. Fundamental Frequency (F0, 'pitch') estimate, AMDF algorithm

4. Spectrogram (sequence of short-term Spectra (FFTs))
The Reality of Rhythm!
How to measure physical rhythm

Amplitude Modulation Spectrum with Rhythm Zones
The Role of F0

If a spectrum can be derived from the **AM envelope**, why not derive a spectrum from the **FM track** and see whether they correlate?

**Preliminary answer:**
Yes, they do correlate to some extent, but not overwhelmingly strongly!

This is not very surprising, of course, since they are partly co-extensive locally and globally, though locally not too similar.

I will look at both AM and FM spectra.
AM and FM Demodulation

Mandarin, female
30 sec, < 1Hz

Correlation AME:FME=0.74
Correlation AMS:FMS=0.29
Envelope Demodulation: Extending to Discourse Spectra

English (RP) Edinburgh corpus “The North Wind and the Sun”

AM & FM signals and spectra: Abercrombie_English_NW048

- Params: minfo:70, maxfo:200, frame:0.01, weight:0.02, sigmedianfactor:1, f0median:9, sigstart:5, siglen:35, maxhz:1

AM carrier with amplitude envelopes (dark green: negative amplitude, light green rectified amplitude)
red: peak-picked envelope

AM Envelope Spectrum (0.0,...,1Hz)

Jassem Edge Detector (AM rhythm zones, polydeg=14)

1 Hz
Envelope Demodulation: Extending to Discourse Spectra

English (RP) Edinburgh corpus “The North Wind and the Sun”
Extending to Discourse Spectra: English Genres

English Newsreading

A. Waveform A0101B-aems-5-35-1 (35s)

B. Amplitude Envelope Modulation Spectrum (Hz)

AEMS heatmap (log scale, darker: higher magnitude)

Spectral Rhythm Zone Boundaries (lighter colours)
AM and FM Demodulation and Spectral Tree Induction

English Newsreading

A. Waveform A0101B-aems-5-35-1 (35s)

B. Amplitude Envelope Modulation Spectrum (Hz)

AEMS heatmap (log scale, darker: higher magnitude)

L-strong, >
L-strong, <
R-strong, >
R-strong, <

IKIU Prosody Workshop, April 2019  D. Gibbon, Rhythms and Melodies of Speech
AM and FM Demodulation and Spectral Tree Induction

English Newsreading

Spectral Tree – Rhythm Tree

L-strong, <
AM and FM Demodulation and Spectral Tree Induction

English (RP) Edinburgh corpus “The North Wind and the Sun”
Rhythm Zones

Speech Modulations and Models V04 2019-04-18 DG [file: one-to-thirty-11s]

1. Waveform

2. Amplitude Modulation Envelope

3. Fundamental Frequency (F0) Estimation

4. Spectrogram (sequence of short-term Spectra (FFTs))

5. Rhythms and Rhythm Zones: Long Term Spectrum (FFT)

6. Rhythms and Rhythm Zones: AEMS with rhythm bars

7. Rhythms and Rhythm Zones: FEMS with main rhythm bar

8. Rhythm Zones (LTS, weighted)
Rhythm Zones

Speech Modulations and Models V04 2019-04-18 DG [file: 02-i-will-be-the-greatest-jobs]

1. Waveform

2. Amplitude Modulation Envelope

3. Fundamental Frequency (F0) Estimation

4. Spectrogram (sequence of short-term Spectra (FFTs))

5. Rhythms and Rhythm Zones: Long Term Spectrum (FFT)

6. Rhythms and Rhythm Zones: AEMS with rhythm bars

7. Rhythms and Rhythm Zones: FEMS with main rhythm bar

8. Rhythm Zones (lts, weighted)
Rhythm Zones


1. Waveform

2. Amplitude Modulation Envelope

3. Fundamental Frequency (F0) Estimation

4. Spectrogram (sequence of short-term Spectra (FFTs))

5. Rhythms and Rhythm Zones: Long Term Spectrum (FFT)

6. Rhythms and Rhythm Zones: AEMS with rhythm bars

7. Rhythms and Rhythm Zones: FEMS with main rhythm bar

8. Rhythm Zones (Its, weighted)
Classification of Utterances by Rhythm Zone Similarity

- Data: rhythm zone vectors
- Calculate Manhattan Distance (differences)
  \[ d_i(p, q) = \sum_{i=1}^{n} |p_i - q_i| \]
- Group distances by Average Pair Group Method with Arithmetic Mean (UPGMA)
  \[ d_i(p, q) = \sum_{i,j} \frac{\text{dist}(p_i, q_j)}{|p| \times |q|} \]
Discourse Rhythms: Long FM contours

Thesis: in evolution,
  - frequency modulation and rhythm came first
    • emotional cries
    • turn-taking came before grammar,
      Levinson, “Turn-taking in Human Communication – Origins and Implications for Language Processing”, 2015

Note: in infant speech,
  - frequency modulation and rhythm also come first
    • emotional cries
      Wermke, Sebastian-Galles
    • turn-taking
      cf. the ‘bootstrapping’ literature
      the infant ‘twin-talk’ videos on YouTube 😊
Discourse Rhythms: Long FM contours

Question: rising utterance contour
Answer: falling utterance contour

Question+Answer: rising-falling adjacency pair contour

syntagmatic entrainment
Discourse Rhythms: Long FM contours

H* sequence, global fall

L* sequence, global rise, final rise

Interview start

Question

Response

Continuation

H* sequence, global fall

L* sequence, global fall, final rise

D. Gibbon, Rhythms and Melodies of Speech
But there are Methodological Problems in F0 / Pitch estimation

1. Terminology:
   - *articulation rate* *(production)*
   - *F0* *(acoustic transmission)*
   - *pitch* *(perception)*

2. Measurement:
   - F0 estimation implementations yield slightly different results
     - Autocorrelation
     - Normalised Cross-correlation
     - Average Magnitude Difference Function (AMDF)
     - FFT peak detection
     - Cepstrum
   - Environment differences
     - Preprocessing: low-pass filter; centre-clipping
     - Postprocessing: moving median
RAPT (Robust Algorithm for Pitch Tracking)

David Talkin
RAPT (Python emulation)

Daniel Gaspari
Reaper (Robust Epoch And Pitch Estimator)
David Talkin
Praat
Paul Boersma
YIN (as opposed to YANG, Python emulation)

Patrice Guyot
YAAPT (Yet Another Algorithm for Pitch Tracking, Python emulation)

Bernardo J. B. Schmitt
SWIPE (Square Wave Inspired Pitch Estimator, Python emulation)

Disha Garg
F0 – Pitch: A Constructive Do-It-Yourself Strategy

Time domain:
AMDF

A kind of auto-correlation, but with subtraction minima not correlation maxima

Preprocessing:
• centre-clipper
• low-pass filter
Postprocessing:
• moving median

Simple: no
• voice detection
• candidate weighting

(code on GitHub)

Frequency domain:
FFT+spectral peak-picking

Finding the lowest frequency spectral peak in the Fourier transform

Preprocessing:
• centre-clipper
• low-pass filter
Postprocessing:
• moving median

Simple - no
• voice detection
• candidate weighting

(code on GitHub)
The Average Magnitude Difference Function works by subtracting segments of the signal from each other and picking the smallest difference, giving the fundamental period.

Then: fundamental frequency = 1 / fundamental period
Frequency Demodulation: the AMDF method

Divide the signal into frames into frames, and for each frame (current frame in red):

Define a frame-sized copy of the frame, and move the copy along the signal for the length of the frame (moving window). For each sample in the frame:

At each step collect average differences between frame and moving window

Find the smallest average difference for all average differences in the frame

Frame start subtracted from time of difference = duration of fundamental period

fundamental frequency = 1 / fundamental period
Autocorrelation functions on the same principle, except that copies are multiplied, not subtracted, and the largest product is taken. This is how Praat works:
How about AMDF (Average Magnitude Difference Function, Python)

Dafydd Gibbon
FFTpeak (Simple F0 Tracker, Python)

Dafydd Gibbon*

* inspired by a snippet from ‘Jonathan’, gist.github.com/endolith/255291

IKIU Prosody Workshop, April 2019   D. Gibbon, Rhythms and Melodies of Speech
Comparing F0 estimators with RAPT as ‘gold standard’

<table>
<thead>
<tr>
<th>F0 estimator pair</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAPT:PyRAPT</td>
<td>0.8902</td>
</tr>
<tr>
<td>RAPT:Praat</td>
<td>0.8657</td>
</tr>
<tr>
<td>RAPT:FFTpeak</td>
<td>0.9096</td>
</tr>
<tr>
<td>RAPT:f0AMDF</td>
<td>0.8605</td>
</tr>
<tr>
<td>FFTpeak:RAPT</td>
<td>0.9096</td>
</tr>
<tr>
<td>FFTpeak:Praat</td>
<td>0.8409</td>
</tr>
<tr>
<td>FFTpeak:PyRAPT</td>
<td>0.8352</td>
</tr>
<tr>
<td>FFTpeak:f0AMDF</td>
<td>0.8016</td>
</tr>
</tbody>
</table>

Benchmark against standard F0 estimators

Encouraging for FFTpeak (problems remain, of course):
- correlation ignores some relevant properties such as overall difference in pitch height
- (slightly positively biased) idea of the relationship
- not enough test data
- not as robust as RAPT
- but suggests that RAPT is fit for purpose
Frequency Modulations – Emotive Rhythms

Thesis 1:
In the evolutionary time domain: emotive ‘animal’ modulations came before structural modulations

Thesis 2:
In the beginning was “Wow!” (Or “Aaah!”)

Thesis 3:
Or the wolf whistle (it’s not simply ‘cat-calling’)

Thesis 4:
Other primates wowed, aahed and whistled first. Humans continued the custom.

... I recommend these topics for future M.A. theses!
Selected Work on Amplitude Envelope Demodulation Spectra


Thank you