Periodicity detection

Juan Pablo Bello
EL9173 Selected Topics in Signal Processing: Audio Content Analysis
NYU Poly
Periodicity detection

- Formally, a periodic signal is defined as:

  \[ x(t) = x(t + T_0), \forall t \]

- Detect the fundamental period/frequency (and phase)
Applications

• At short (pitch) and long (rhythm) time scales:
  • pitch-synchronous analysis
  • voice/sound identification
  • prosodic analysis
  • bioacoustics
  • melodic analysis
  • note transcription
  • beat tracking, segmentation
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with $f_0$
- Transients and noise
- Polyphonies: information overlap
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- Transients and noise
- Polyphonies: information overlap
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- Transients and noise
- Polyphonies: information overlap
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- Transients and noise
- Polyphonies: information overlap
Difficulties

- Quasi-periodicities, temporal variations
- Multiple periodicities associated with f0
- Transients and noise
- Polyphonies: information overlap
Overlap

(a) C4 and E4 - a third

(b) C4 and G4 - a fifth

(c) C4 and C5 - an octave
Architecture

Audio → Break into blocks → Detection function + peak-picking → $f_0$ → Integration over time

Time smoothing
Overview of Methods

• DFT

• Autocorrelation

• Spectral Pattern Matching

• Cepstrum

• Spectral Autocorrelation

• YIN

• Auditory model
DFT

Waveform

Time (secs)

Magnitude (dB)

Frequency (Hz)
DFT
DFT
DFT

The top graph represents the waveform over time, with a highlighted segment marked as $T_0$. The bottom graph shows the magnitude of the signal in dB over frequency, with several highlighted peaks marked as $f_0$. The x-axis represents time in seconds, ranging from 0.25 to 0.33, and the y-axis represents the waveform magnitude. The x-axis of the bottom graph represents frequency in Hz, ranging from 0 to 3500.
Autocorrelation

• Cross-product measures similarity across time

• Cross-correlation of two real-valued signals $x$ and $y$:

$$r_{xy}(l) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)y(n + l) \mod N$$

$$l = 0, 1, 2, \ldots, N - 1$$

• Unbiased (short-term) Autocorrelation Function (ACF):

$$r_x(l) = \frac{1}{N - l} \sum_{n=0}^{N-1-l} x(n)x(n + l)$$

$$l = 0, 1, 2, \ldots, L - 1$$
Autocorrelation
Autocorrelation

• The short-term ACF can also be computed as:

\[
rx(l) = \left(\frac{1}{N-l}\right) \text{real}(IFTT(|X|^2))
\]

\[X \rightarrow FFT(x)\]

\[x \text{ zero-padded to next power of 2 after } (N + L) - 1\]
Autocorrelation

- This is equivalent to the following correlation:

\[ r_x(l) = \frac{1}{K - l} \sum_{k=0}^{K-1} \cos(2\pi lk/K)|X(k)|^2 \]
Pattern Matching

• Comb filtering is a common strategy

• Any other template that realistically fits the magnitude spectrum

• Templates can be specific to sound sources

• Matching strategies vary: correlation, likelihood, distance, etc.
Pattern Matching
Cepstrum

- Treat the log magnitude spectrum as if it were a signal - take its (I)DFT
- Measures rate of change across frequency bands (Bogert et al., 1963)
- Cepstrum -> Anagram of Spectrum (same for quefrency, liftering, etc)
- For a real-valued signal is defined as:

$$c_x(l) = \text{real}(\text{IFFT}(\log(|FFT(x)|)))$$
Cepstrum
Spectral ACF

• Spectral location -> sensitive to quasi-periodicities

• (Quasi-)Periodic Spectrum, Spectral ACF.

\[
rx(l_f) = \frac{1}{N - l_f} \sum_{k=0}^{N-1-l_f} |X(k)||X(k + l_f)|
\]

\[
l_f = 0, 1, 2, \cdots, L - 1
\]

• Exploits intervalic information (more stable than locations of partials), while adding shift-invariance.
Spectral ACF
• Alternative to the ACF that uses the squared difference function (deCheveigne, 02):

\[
d(l) = \sum_{n=0}^{N-1-l} (x(n) - x(n + l))^2
\]

• For (quasi-)periodic signals, this functions cancel itself at \(l = 0, l_0\) and its multiples. Zero-lag bias is avoided by normalizing as:

\[
\hat{d}(l) = \begin{cases} 
1 & l = 0 \\
\frac{d(l)}{[(1/l) \sum_{u=1}^l d(u)]} & \text{otherwise}
\end{cases}
\]
Auditory model

$\begin{align*}
\mathbf{x}(n) & \xrightarrow{\text{Auditory filterbank}} \\
\text{Inner hair-cell model} & \quad \cdots \quad \text{Inner hair-cell model} \\
\text{Periodicity} & \quad \cdots \quad \text{Periodicity} \\
\text{Summarization} & \xrightarrow{\text{Summary periodicity function}} \\
\mathbf{x}_c(n) & \quad \cdots \\n\mathbf{z}_c(n) & \\n\mathbf{r}_c(n) & 
\end{align*}$
Auditory model

- Auditory filterbank: gammatone filters (Slaney, 93; Klapuri, 06):

![Diagram of filter responses](image-url)
Auditory model

The Equivalent Rectangular Bandwidths (ERB) of the filters:

\[ b_c = 0.108 f_c + 24.7 \]

\[ f_c = 229 \times (10^{\psi/21.4} - 1) \]

\[ \psi = \psi_{\text{min}} : (\psi_{\text{max}} - \psi_{\text{min}})/F : \psi_{\text{max}} \]

\[ \psi_{\text{min}/\text{max}} = 21.4 \times \log_{10}(0.00437 f_{\text{min}/\text{max}} + 1) \]

\[ F = \text{number of filters}. \]
Auditory model

• Beating: interference between sounds of frequencies $f_1$ and $f_2$

• Fluctuation of amplitude envelope of frequency $|f_2 - f_1|$

• The magnitude of the beating is determined by the smaller of the two amplitudes
Auditory model

• Inner hair-cell (IHC) model:
Auditory model

- Sub-band periodicity analysis using ACF

- Summing across channels (Summary ACF)

- Weighting of the channels changes the topology of the SACF
Auditory model
Comparing detection functions
Comparing detection functions
Comparing detection functions
Comparing detection functions
Tempo

• Tempo refers to the pace of a piece of music and is usually given in beats per minutes (BPM).

• Global quality vs time-varying local characteristic.

• Thus, in computational terms we differentiate between tempo estimation and tempo (beat) tracking.

• In tracking, beats are described by both their rate and phase.

• Vast literature: see, e.g. Hainsworth, 06; or Goto, 06 for reviews.
**Tempo estimation and tracking (Davies, 05)**

- Novelty function (NF): remove local mean + half-wave rectify

- Periodicity: dot multiply ACF of NF with a weighted comb filterbank

\[ R_w(l) = \left( \frac{l}{b^2} \right) e^{-\frac{l^2}{2b^2}} \]

*From Davies and Plumbley, ICASSP 2005*
Tempo estimation and tracking (Davies, 05)

- Choose lag that maximizes the ACF
Tempo estimation and tracking (Davies, 05)

- Choose filter that maximizes the dot product
Tempo estimation and tracking (Davies, 05)

- Phase: dot multiply DF with shifted versions of selected comb filter
Tempo estimation and tracking (Grosche, 09)

- DFT of novelty function $\gamma(n)$ for frequencies: $\omega \in [30 : 600]/(60 \times f_{s_{\gamma}})$

- Choose frequency that maximizes the magnitude spectrum at each frame

- Construct a sinusoidal kernel: $\kappa(m) = w(m - n)\cos(2\pi(\hat{\omega}m - \hat{\phi}))$

- In Grosche, 09 phase is computed as:

  $$\hat{\phi} = \frac{1}{2\pi} \arccos\left(\frac{\text{Re}(F(\hat{\omega}, n))}{|F(\hat{\omega}, n)|}\right)$$

- Alternatively, we can find the phase that maximizes the dot product of $\gamma(n)$ with shifted versions of the kernel, as before.
Tempo estimation and tracking (Grosche, 09)

- tracking function: Overlap-add of optimal local kernels + half-wave rectify

*From Grosche and Mueller, WASPAA 2009*
Tempo estimation and tracking (Davies, 05)
Tempo estimation and tracking (Grosche, 09)

The Grid - Swamp Thing

Detection Function

Post-processed DF

Tempogram
Tempo estimation and tracking (Davies, 05)
Tempo estimation and tracking (Grosche, 09)

Groove Armada - Whatever, Whenever

Detection Function

Post-processed DF

Tempogram

PLP
Tempo estimation and tracking (Davies, 05)
Tempo estimation and tracking (Grosche, 09)

Herbie Hancock - Cantaloup Island

Detection Function

Post-processed DF

Tempogram

PLP
Tempo estimation and tracking (Davies, 05)
Tempo estimation and tracking (Grosche, 09)
References


• This lecture borrows heavily from: Emmanuel Vincent’s lecture notes on pitch estimation (QMUL - Music Analysis and Synthesis); and from Anssi Klapuri’s lecture notes on F0 estimation and automatic music transcription (ISMIR 2004 Graduate School: http://ismir2004.ismir.net/graduate.html)