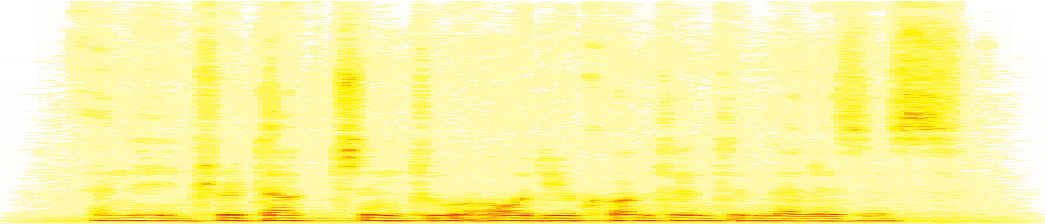
Introduction to Audio Content Analysis

Module 5.2: Fundamental Frequency Detection in Monophonic Signals

alexander lerch



Georgia Tech Center for Music Technology
College of Design

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 5 — Tonal Analysis: pp. 91–103

- **lecture content**
  - established approaches to monophonic pitch tracking in
    - time domain
    - frequency domain
- **learning objectives**
  - define the task of monophonic pitch tracking
  - summarize the principles of time-domain f_0-trackers and describe one approach in detail
  - summarize the principles of frequency-domain f_0-trackers and describe one approach in detail

Georgia Tech Center for Music Technology College of Design

corresponding textbook section

Chapter 5 — Tonal Analysis: pp. 91–103

- **lecture content**
  - established approaches to monophonic pitch tracking in
    - time domain
    - frequency domain
- **learning objectives**
  - define the task of monophonic pitch tracking
  - summarize the principles of time-domain f_0-trackers and describe one approach in detail
  - summarize the principles of frequency-domain f_0-trackers and describe one approach in detail

fundamental frequency

introduction

Georgia Tech Center for Music Technology
College of Design

remember

Fourier series: every (quasi-)periodic sound is a combination of sinusoidals with integer frequency ratios

$$
\begin{aligned}
x(t) &\approx x(t+T_0) \\
x(t) &\approx \sum_{k=-\infty}^{\infty} a(k) e^{\mathrm{j}\omega_0 kt}
\end{aligned}
$$

f_0: musically, perceptually most "relevant" frequency

fundamental frequency

introduction

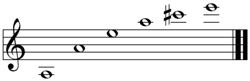
Georgia Tech Center for Music Technology
College of Design

remember

Fourier series: every (quasi-)periodic sound is a combination of sinusoidals with integer frequency ratios

$$
\begin{aligned}
x(t) &\approx x(t+T_0) \\
x(t) &\approx \sum_{k=-\infty}^{\infty} a(k) e^{\mathrm{j}\omega_0 k t}
\end{aligned}
$$

f_0: musically, perceptually most "relevant" frequency



pitch detection
task definition

Georgia Tech Center for Music Technology College of Design

- detect the fundamental frequency f_0
- there is only one fundamental frequency at a time

- related subtasks:
  - detect when there is no fundamental frequency
  - segment into notes
    - start time and stop time
    - average note frequency
    - vibrato detection
  - map to pitch scale

monophonic fundamental frequency detection

zero crossing rate

Georgia Tech Center for Music Technology College of Design

- **number of zero crossings per block** (inaccurate)

$$T_0(n) = \frac{2 \cdot \left(i_e(n) - i_s(n)\right)}{f_S \cdot \sum\limits_{i=i_s(n)}^{i_e(n)} \left|\mathrm{sign}\left[x(i)\right] - \mathrm{sign}\left[x(i-1)\right]\right|}$$

- **average period length**

$$T_0(n) = \frac{2}{\mathcal{Z}-1} \sum_{j=0}^{\mathcal{Z}-2} \Delta t_{ZC}(j).$$

- **variants**:
  - create histogram with distances and choose maximum
  - use not (only) ZC but distance between local extrema

monophonic fundamental frequency detection

zero crossing rate

Georgia Tech | Center for Music Technology, College of Design

- **number of zero crossings per block** (inaccurate)

$$T_0(n) = \frac{2 \cdot \left(i_\mathrm{e}(n) - i_\mathrm{s}(n)\right)}{f_\mathrm{S} \cdot \sum\limits_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \left|\mathrm{sign}\left[x(i)\right] - \mathrm{sign}\left[x(i-1)\right]\right|}$$

- **average period length**

$$T_0(n) = \frac{2}{\mathcal{Z} - 1} \sum_{j=0}^{\mathcal{Z}-2} \Delta t_\mathrm{ZC}(j).$$

- **variants**:
  - create histogram with distances and choose maximum
  - use not (only) ZC but distance between local extrema

monophonic fundamental frequency detection

zero crossing rate

- **number of zero crossings per block** (inaccurate)

$$T_0(n) = \frac{2 \cdot \left(i_\mathrm{e}(n) - i_\mathrm{s}(n)\right)}{f_\mathrm{S} \cdot \sum\limits_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \left|\mathrm{sign}\left[x(i)\right] - \mathrm{sign}\left[x(i-1)\right]\right|}$$

- **average period length**

$$T_0(n) = \frac{2}{\mathcal{Z}-1} \sum_{j=0}^{\mathcal{Z}-2} \Delta t_\mathrm{ZC}(j).$$

- **variants**:
  - create histogram with distances and choose maximum
  - use not (only) ZC but distance between local extrema

monophonic fundamental frequency detection

auto correlation function

- find **lag of ACF maximum**

$$r_{xx}(\eta, n) = \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)-\eta} x(i) \cdot x(i+\eta)$$

- **variants**:
    - center clipping

monophonic fundamental frequency detection
auto correlation function

Georgia Tech Center for Music Technology College of Design

- find **lag of ACF maximum**

$$r_{xx}(\eta, n) = \sum_{i=i_s(n)}^{i_e(n)-\eta} x(i)\cdot x(i+\eta)$$

- **variants**:
  - center clipping

monophonic fundamental frequency detection

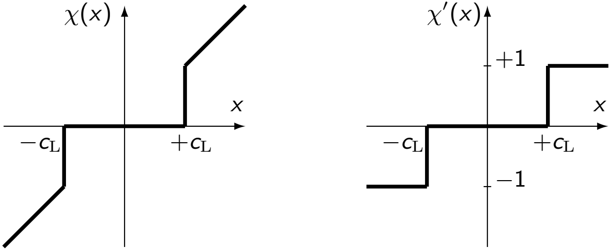
auto correlation function

Georgia Tech Center for Music Technology College of Design

- find **lag of ACF maximum**

$$r_{xx}(\eta, n) = \sum_{i=i_s(n)}^{i_e(n)-\eta} x(i)\cdot x(i+\eta)$$

- **variants**:
    - center clipping



monophonic fundamental frequency detection
average magnitude difference function

- find **lag of AMDF minimum**

$$\mathrm{AMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)-\eta} |x(i) - x(i+\eta)|$$

- **variants**:
  - AMDF-weighted ACF

$$r'_{xx}(\eta, n) = \frac{r_{xx}(\eta, n)}{\mathrm{AMDF}_{xx}(\eta, n) + 1}$$

monophonic fundamental frequency detection
average magnitude difference function

Georgia Tech Center for Music Technology College of Design

- find **lag of AMDF minimum**

$$\mathrm{AMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)-\eta} |x(i) - x(i+\eta)|$$

- **variants**:
  - AMDF-weighted ACF

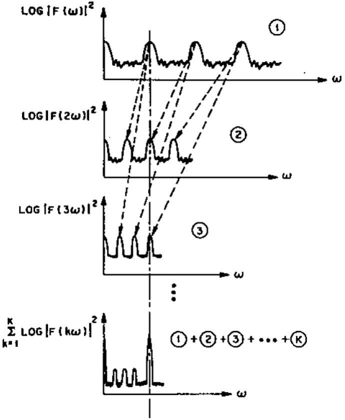
$$r'_{xx}(\eta, n) = \frac{r_{xx}(\eta, n)}{\mathrm{AMDF}_{xx}(\eta, n) + 1}$$

monophonic fundamental frequency detection

harmonic product spectrum 1/2

$$X_{\mathrm{HPS}}(k,n) = \prod_{j=1}^{\mathcal{O}} |X(j\cdot k, n)|^2$$

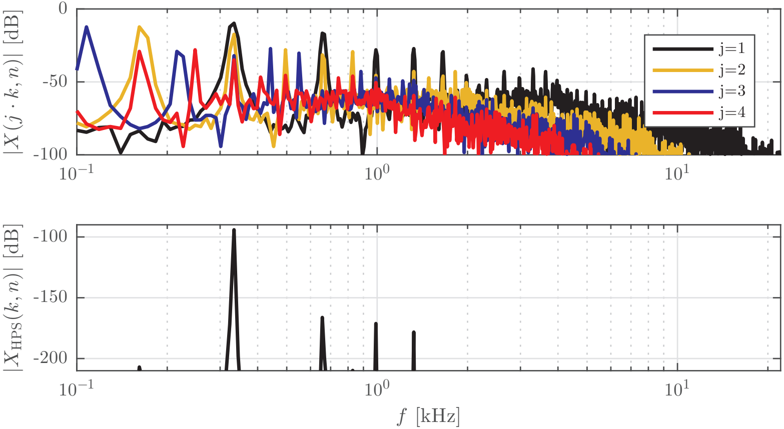
first published in the 1960s by Noll (graph from that paper)[1]



[1] A. M. Noll, "Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum, and a Maximum Likelihood Estimate," in *Proceedings of the Symposium on Computer Processing in Communications*, vol. 19, Brooklyn: Polytechnic Press of the University of Brooklyn, 1969, pp. 779–797.

monophonic fundamental frequency detection

harmonic product spectrum 2/2



matlab source: matlab/displayHarmonicProductSpectrum.m

monophonic fundamental frequency detection

harmonic sum spectrum

Georgia Tech Center for Music Technology College of Design

- sum instead product sum

$$X_{\mathrm{HSS}}(k, n) = \sum_{j=1}^{\mathcal{O}} |X(j \cdot k, n)|^2$$

- **advantage**
  - robust against missing harmonics
- **disadvantage**
  - less pronounced peak

monophonic fundamental frequency detection
ACF of magnitude spectrum

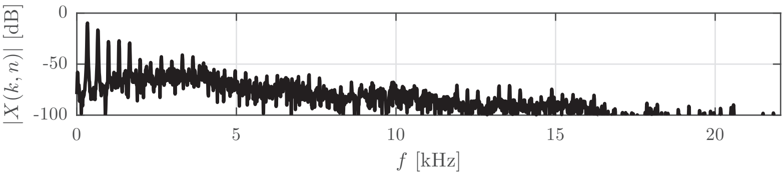
$$r_{XX}(\eta, n) = \sum_{k=-\mathcal{K}/2}^{\mathcal{K}/2-1} |X(k,n)| \cdot |X(k+\eta, n)|$$

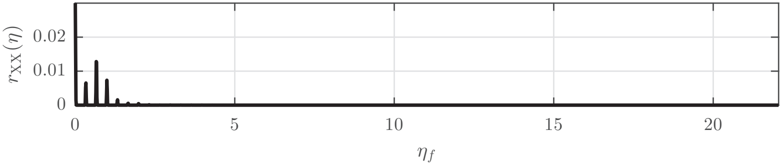
$\Rightarrow$ **detect maximum location**

monophonic fundamental frequency detection
ACF of magnitude spectrum

$$r_{XX}(\eta, n) = \sum_{k=-\mathcal{K}/2}^{\mathcal{K}/2-1} |X(k,n)| \cdot |X(k+\eta, n)|$$

$\Rightarrow$ **detect maximum location**





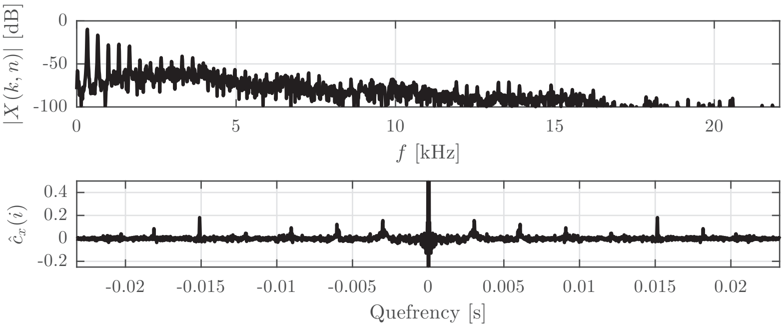
matlab source: matlab/displayAcfOfFft.m

monophonic fundamental frequency detection

cepstral pitch detection

Georgia Tech Center for Music Technology College of Design

1. compute cepstrum
2. detect periodicities



matlab source: matlab/displayCepstrum.m

monophonic fundamental frequency detection

spectral maximum likelihood

- create **template matrix** with (smoothed) delta pulses for all possible frequencies
- compute the **cross correlation** ($lag = 0$) between spectrum and all templates
- pick the result with the **highest correlation** $\Rightarrow$ frequency estimate (graph see[2])

[2] P. de la Cuadra, *Pitch detection methods review*, [Online]. Available: https://ccrma.stanford.edu/~pdelac/154/m154paper.htm (visited on 08/04/2015).

monophonic fundamental frequency detection

spectral maximum likelihood

Georgia Tech Center for Music Technology College of Design

- create **template matrix** with (smoothed) delta pulses for all possible frequencies
- compute the **cross correlation** ($lag = 0$) between spectrum and all templates
- pick the result with the **highest correlation** $\Rightarrow$ frequency estimate (graph see[2])

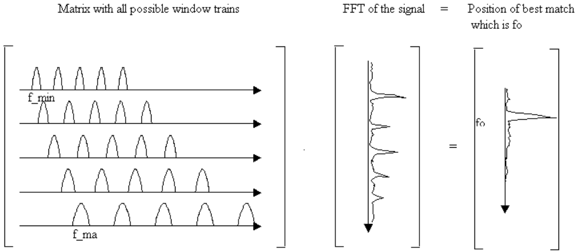
[2] P. de la Cuadra, *Pitch detection methods review*, [Online]. Available: https://ccrma.stanford.edu/~pdelac/154/m154paper.htm (visited on 08/04/2015).

monophonic fundamental frequency detection

spectral maximum likelihood

Georgia Tech Center for Music Technology College of Design

- create **template matrix** with (smoothed) delta pulses for all possible frequencies
- compute the **cross correlation** ($lag = 0$) between spectrum and all templates
- pick the result with the **highest correlation** $\Rightarrow$ frequency estimate (graph see[2])



[2] P. de la Cuadra, *Pitch detection methods review*, [Online]. Available: https://ccrma.stanford.edu/~pdelac/154/m154paper.htm (visited on 08/04/2015).

monophonic fundamental frequency detection

auditory-motivated pitch tracking 1/2

1. **filterbank** of band bass filters (e.g., mel scale)
2. **HWR**
3. **smoothing**
4. within-band periodicity estimate (e.g. **ACF**)
5. **combination** of bands

monophonic fundamental frequency detection

auditory-motivated pitch tracking 1/2

Georgia Tech Center for Music Technology College of Design

1. **filterbank** of band bass filters (e.g., mel scale)
2. **HWR**
3. **smoothing**
4. within-band periodicity estimate (e.g. **ACF**)
5. **combination** of bands

monophonic fundamental frequency detection

auditory-motivated pitch tracking 1/2

1. **filterbank** of band bass filters (e.g., mel scale)
2. **HWR**
3. **smoothing**
4. within-band periodicity estimate (e.g. **ACF**)
5. **combination** of bands

monophonic fundamental frequency detection

auditory-motivated pitch tracking 1/2

Georgia Tech Center for Music Technology College of Design

1. **filterbank** of band bass filters (e.g., mel scale)
2. **HWR**
3. **smoothing**
4. within-band periodicity estimate (e.g. **ACF**)
5. **combination** of bands

monophonic fundamental frequency detection

auditory-motivated pitch tracking 1/2

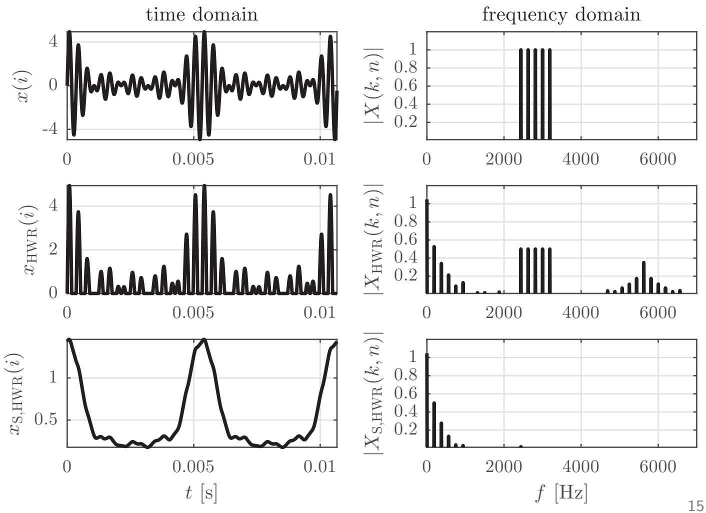
1. **filterbank** of band bass filters (e.g., mel scale)
2. **HWR**
3. **smoothing**
4. within-band periodicity estimate (e.g. **ACF**)
5. **combination** of bands

monophonic fundamental frequency detection

auditory-motivated pitch tracking 2/2

Georgia Tech | Center for Music Technology, College of Design

1. filterbank output
2. half wave rectification
3. smoothed output



matlab source: matlab/displayAuditoryPitchTracking.m

summary
lecture content

Georgia Tech Center for Music Technology College of Design

- **basic approach**
  - all approaches look for **periodicity**
    - waveform similarity in time domain
    - equidistant harmonics/peaks in freq domain
- **state-of-the-art**
  - despite the age of the presented methods, tweaked versions of the presented approaches are still often considered state-of-the-art
  - especially combinations of different approaches can be very successful