

Introduction to Audio Content Analysis

Module 7.3.3: Fundamental Frequency Detection in Monophonic Signals

alexander lerch

introduction

overview

corresponding textbook section

Section 7.3.3

■ 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



corresponding textbook section

Section 7.3.3

■ 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

remember

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

$$x(t) \approx x(t + T_0)$$

$$x(t) \approx \sum_{k=-\infty}^{\infty} a(k) e^{j\omega_0 k t}$$

f_0 : musically, perceptually most “relevant” frequency

fundamental frequency

introduction

remember

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

$$x(t) \approx x(t + T_0)$$

$$x(t) \approx \sum_{k=-\infty}^{\infty} a(k) e^{j\omega_0 k t}$$

f_0 : musically, perceptually most “relevant” frequency



pitch detection

task definition

- detect the **fundamental frequency** f_0
- monophonic: only **one** fundamental frequency at a time
- related **subtasks**:
 - *voice activity*: detect when there is no voice/no fundamental frequency
 - *note segmentation*
 - ▶ note start time and stop time
 - ▶ average note frequency
 - ▶ average note velocity
 - ▶ vibrato detection
 - frequency to *pitch mapping*

monophonic fundamental frequency detection

zero crossing rate

■ number of zero crossings per block (inaccurate)

$$T_0(n) = \frac{2 \cdot (i_e(n) - i_s(n))}{f_S \cdot \sum_{i=i_s(n)}^{i_e(n)} |\text{sign}[x(i)] - \text{sign}[x(i-1)]|}$$

■ average period length

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

■ variants:

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

monophonic fundamental frequency detection

zero crossing rate

■ number of zero crossings per block (inaccurate)

$$T_0(n) = \frac{2 \cdot (i_e(n) - i_s(n))}{f_S \cdot \sum_{i=i_s(n)}^{i_e(n)} |\text{sign}[x(i)] - \text{sign}[x(i-1)]|}$$

■ average period length

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

■ variants:

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

monophonic fundamental frequency detection

zero crossing rate

■ number of zero crossings per block (inaccurate)

$$T_0(n) = \frac{2 \cdot (i_e(n) - i_s(n))}{f_S \cdot \sum_{i=i_s(n)}^{i_e(n)} |\text{sign}[x(i)] - \text{sign}[x(i-1)]|}$$

■ average period length

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

■ variants:

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

monophonic fundamental frequency detection

auto correlation function

■ 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

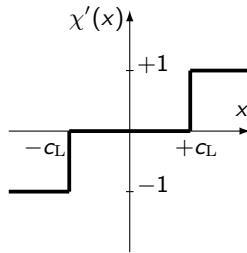
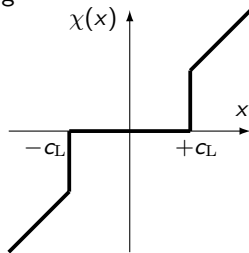
auto correlation function

■ find **lag** of **ACF** maximum

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

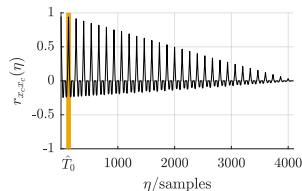
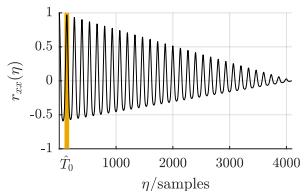
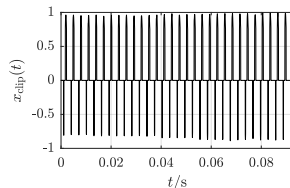
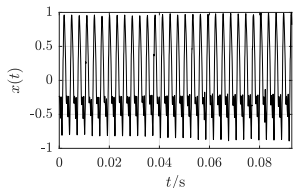
■ **variants:**

- center clipping



monophonic fundamental frequency detection

auto correlation function



monophonic fundamental frequency detection

average magnitude difference function

■ find **lag of AMDF minimum**

$$\text{AMDF}_{xx}(\eta, n) = \frac{1}{i_e(n) - i_s(n) + 1} \sum_{i=i_s(n)}^{i_e(n)-\eta} |x(i) - x(i + \eta)|$$

■ **variants:**

- AMDF-weighted ACF

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

monophonic fundamental frequency detection

average magnitude difference function

■ find **lag of AMDF minimum**

$$\text{AMDF}_{xx}(\eta, n) = \frac{1}{i_e(n) - i_s(n) + 1} \sum_{i=i_s(n)}^{i_e(n)-\eta} |x(i) - x(i + \eta)|$$

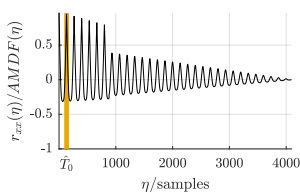
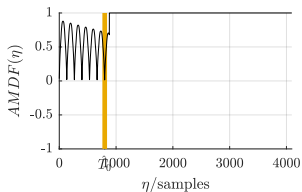
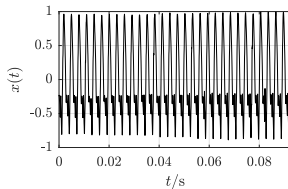
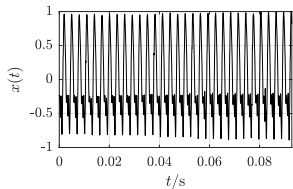
■ **variants:**

- AMDF-weighted ACF

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

monophonic fundamental frequency detection

average magnitude difference function

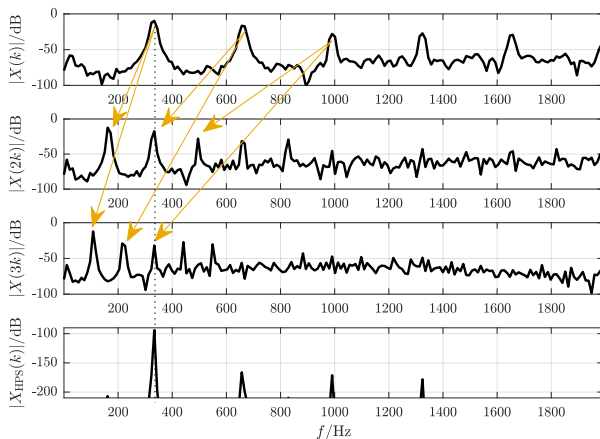


monophonic fundamental frequency detection

harmonic product spectrum 1/2

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

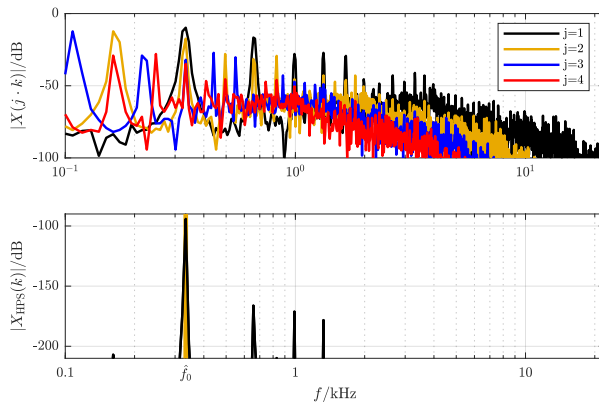
first published in the 1960s by Noll



¹A. 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 City University of New York, 1968, pp. 177-183.

monophonic fundamental frequency detection

harmonic product spectrum 2/2



monophonic fundamental frequency detection

harmonic sum spectrum

- sum instead product sum

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

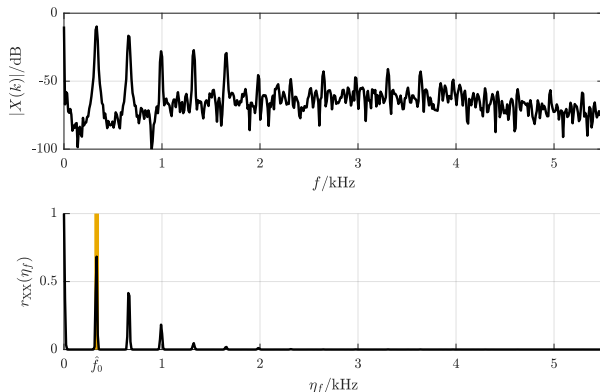
- **advantage**
 - ▶ robust against missing harmonics
- **disadvantage**
 - ▶ less pronounced peak

monophonic fundamental frequency detection

ACF of magnitude spectrum

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

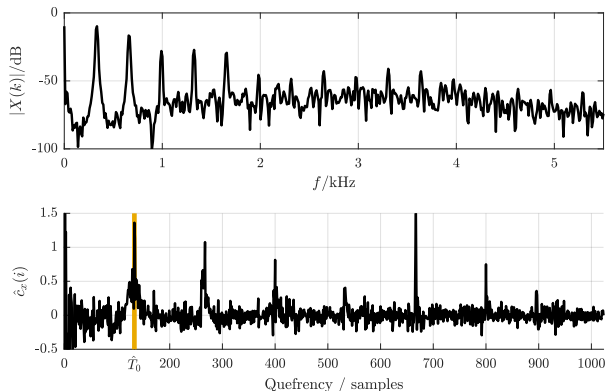
⇒ **detect maximum location**



monophonic fundamental frequency detection

cepstral pitch detection

- 1 compute cepstrum
- 2 detect periodicities



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 see²)

²P. 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

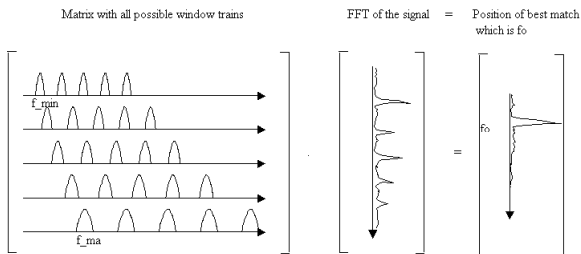
- create **template matrix** with (smoothed) delta pulses for all possible frequencies
- compute the **cross correlation** ($lag = 0$) between spectrum and all templates
- pick the result with the **highest correlation** \Rightarrow frequency estimate (graph see²)

²P. 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

- 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 see²)



²P. 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 pass 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 pass 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 pass 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 pass 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

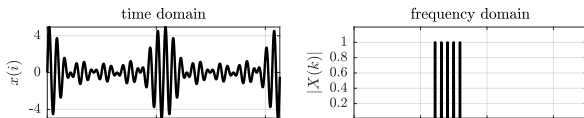
auditory-motivated pitch tracking 1/2

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

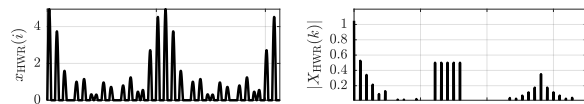
monophonic fundamental frequency detection

auditory-motivated pitch tracking 2/2

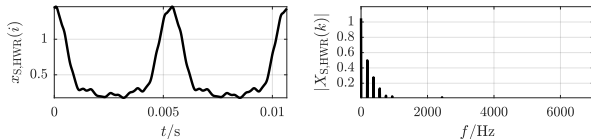
1 filterbank output



2 half wave rectification



3 smoothed output



summary

lecture content

■ basic commonality

- 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
- combinations of different approaches can be robust

