Introduction to **Audio Content Analysis**

Module 7.3.3: Fundamental Frequency Detection in Monophonic Signals

alexander lerch

Georgia Tech Center for Music Technology
College of Design

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 $\hat{f}_0$-trackers and describe one approach in detail
  - summarize the principles of frequency-domain $\hat{f}_0$-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 $\hat{f}_0$-trackers and describe one approach in detail
  - summarize the principles of frequency-domain $\hat{f}_0$-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

$$x(t) \approx x(t + \hat{T}_0)$$

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

$\hat{f}_0$: musically, perceptually most "relevant" frequency

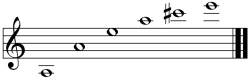
fundamental frequency
introduction

remember

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

$$x(t) \approx x(t + \hat{T}_0)$$
$$x(t) \approx \sum_{k=-\infty}^{\infty} a(k) e^{\mathrm{j}\omega_0 k t}$$

$\hat{f}_0$: musically, perceptually most "relevant" frequency



pitch detection
task definition

- detect the **fundamental frequency** $\hat{f}_0$
- 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

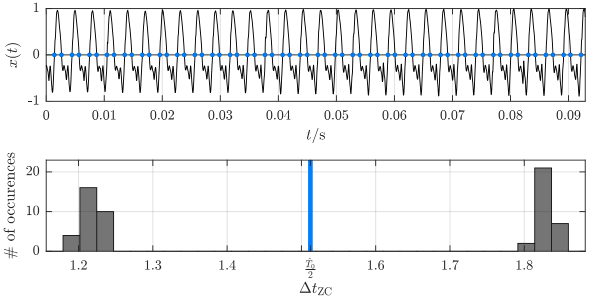
- **ZCR per block (bad)**

$$\hat{T}_0(n) = \frac{2 \cdot \left(i_\mathrm{e}(n) - i_\mathrm{s}(n)\right)}{f_\mathrm{S} \cdot \sum\limits_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \left|\mathrm{sign}\left[x(i)\right] - \mathrm{sign}\left[x(i-1)\right]\right|}$$

- **average period length**

$$\hat{T}_0(n) = \frac{2}{\mathcal{Z}-1} \sum_{j=0}^{\mathcal{Z}-2} \Delta t_\mathrm{ZC}(j).$$

- **variants**:
  - create distance histogram and choose maximum
  - also use distances between local extrema



matlab source: plotF0Zcr.m

monophonic fundamental frequency detection

zero crossing rate

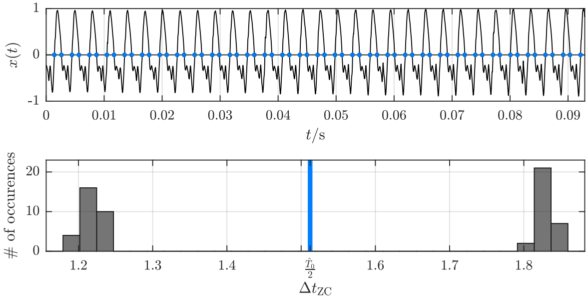
- **ZCR per block (bad)**

$$\hat{T}_0(n) = \frac{2 \cdot (i_e(n) - i_s(n))}{f_S \cdot \sum\limits_{i=i_s(n)}^{i_e(n)} |\mathrm{sign}\,[x(i)] - \mathrm{sign}\,[x(i-1)]|}$$

- **average period length**

$$\hat{T}_0(n) = \frac{2}{\mathcal{Z} - 1} \sum_{j=0}^{\mathcal{Z}-2} \Delta t_{ZC}(j).$$

- **variants**:
  - create distance histogram and choose maximum
  - also use distances between local extrema



matlab source: plotF0Zcr.m

monophonic fundamental frequency detection
zero crossing rate

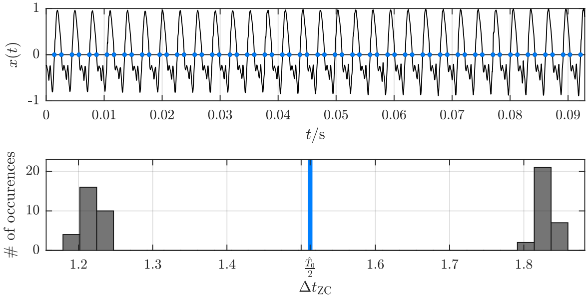
- **ZCR per block (bad)**

$$\hat{T}_0(n) = \frac{2 \cdot \left(i_\mathrm{e}(n) - i_\mathrm{s}(n)\right)}{f_\mathrm{S} \cdot \sum\limits_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \left|\mathrm{sign}\left[x(i)\right] - \mathrm{sign}\left[x(i-1)\right]\right|}$$

- **average period length**

$$\hat{T}_0(n) = \frac{2}{\mathcal{Z}-1}\sum_{j=0}^{\mathcal{Z}-2}\Delta t_\mathrm{ZC}(j).$$

- **variants**:
  - create distance histogram and choose maximum
  - also use distances between local extrema



matlab source: plotF0Zcr.m

monophonic fundamental frequency detection

auto correlation function

- find **lag of ACF maximum**

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

$$\hat{T}_0(n) = \mathrm{argmax}\left(r_{xx}(\eta, n)\right)$$

- **variants**:
  - center clipping

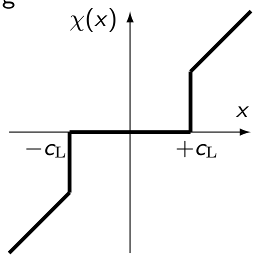
monophonic fundamental frequency detection
auto correlation function

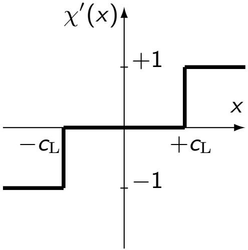
- find **lag of ACF maximum**

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

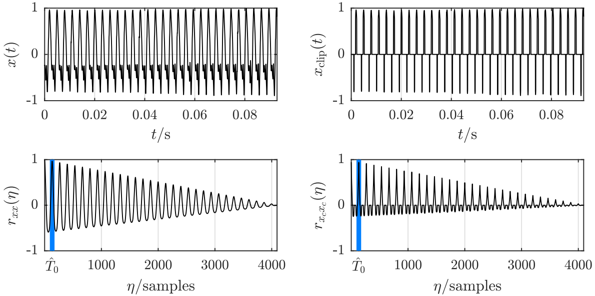
$$\hat{T}_0(n) = \mathrm{argmax}\left(r_{xx}(\eta, n)\right)$$

- **variants**:
  - center clipping





monophonic fundamental frequency detection
auto correlation function



matlab source: plotF0Acf.m

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

- find **lag of AMDF minimum**

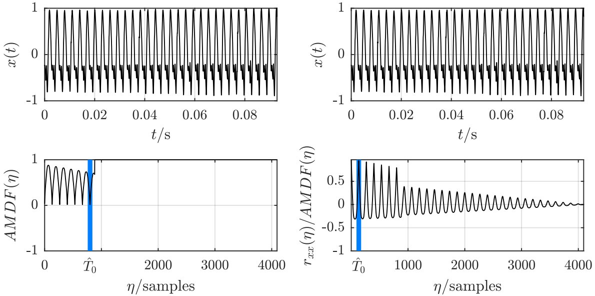
$$\mathrm{AMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)-\eta} |x(i) - x(i+\eta)|$$

- **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



matlab source: plotF0Amdf.m

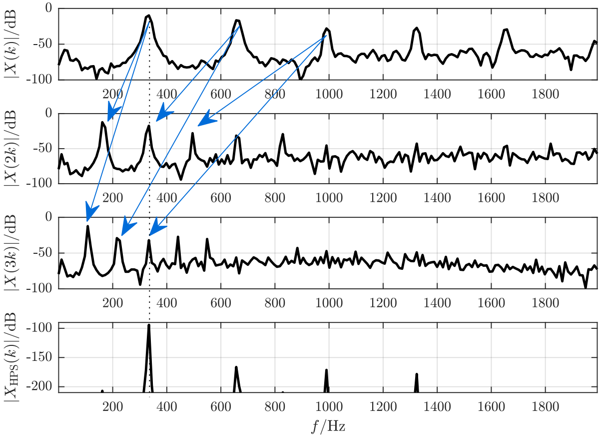
monophonic fundamental frequency detection

harmonic product spectrum 1/2

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

$$\hat{f}_0(n) = \operatorname{argmax}\left(X_{\mathrm{HPS}}(k,n)\right)$$

first published in the 1960s by Noll[1]

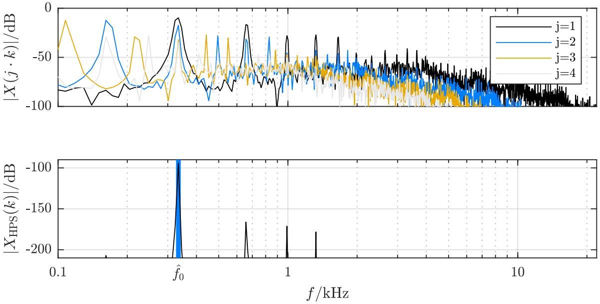


matlab source: plotF0HpsMethod.m

[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

monophonic fundamental frequency detection

harmonic product spectrum 2/2



matlab source: plotF0Hps.m

monophonic fundamental frequency detection

harmonic sum spectrum

- sum instead product sum

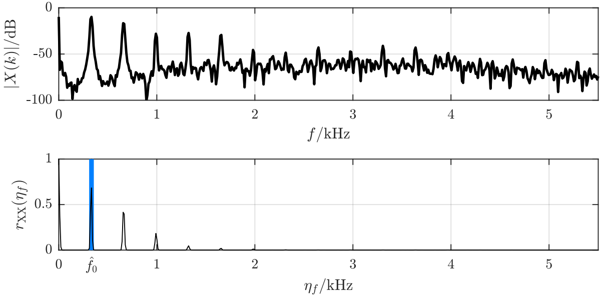
$$X_{\mathrm{HSS}}(k, n) = \sum_{j=1}^{\mathcal{O}} |X(j \cdot k, n)|^2$$

- **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**

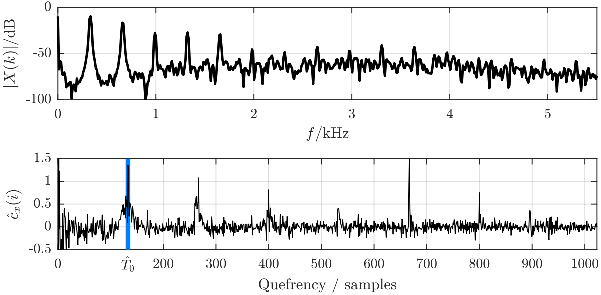


matlab source: plotF0AcfOfFft.m

monophonic fundamental frequency detection

cepstral pitch detection

1. compute cepstrum
2. detect periodicities



matlab source: plotF0Cepstrum.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

monophonic fundamental frequency detection

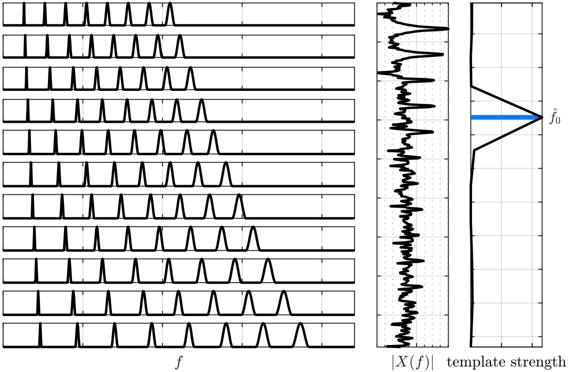
spectral maximum likelihood

- 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

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



matlab source: plotF0Template.m

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

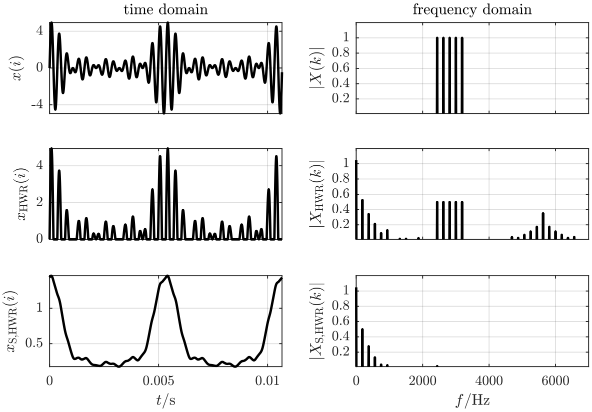
auditory-motivated pitch tracking 1/2

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



matlab source: plotF0Auditory.m

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