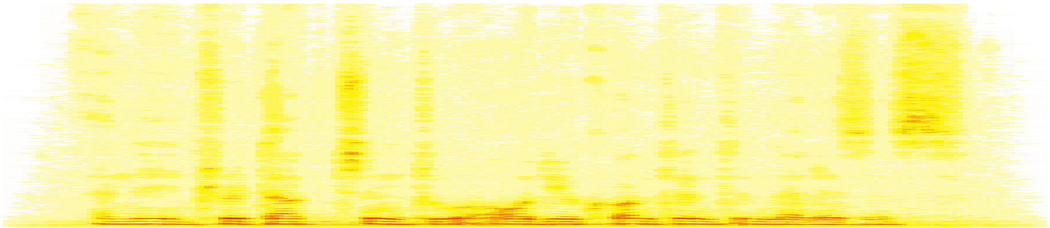
Introduction to Audio Content Analysis

Module 5.4: Fundamental Frequency Detection in Polyphonic Signals

alexander lerch



Georgia Tech Center for Music Technology
College of Design

introduction
overview

corresponding textbook section

Chapter 5 — Tonal Analysis: pp. 103–106

- **lecture content**
  - overview of "historic" methods for polyphonic pitch detection

- **learning objectives**
  - describe the task and challenges of polyphonic pitch detection
  - list the processing steps of iterative subtraction and relate them to the introduced approaches
  - discuss other approaches to polyphonic pitch tracking

introduction
overview

Georgia Tech Center for Music Technology College of Design

corresponding textbook section

Chapter 5 — Tonal Analysis: pp. 103–106

- **lecture content**
  - overview of "historic" methods for polyphonic pitch detection
- **learning objectives**
  - describe the task and challenges of polyphonic pitch detection
  - list the processing steps of iterative subtraction and relate them to the introduced approaches
  - discuss other approaches to polyphonic pitch tracking

polyphonic pitch tracking

problem statement

- **monophonic** fundamental frequency detection:
  - exactly one fundamental frequency with sinusoidals at multiples of f_0 (harmonics)
- **polyphonic** fundamental frequency detection:
  - multiple/unknown number of fundamental frequencies with harmonics
  - number of voices might change over time
  - complex mixture with overlapping frequency content

polyphonic pitch tracking
iterative subtraction: introduction

Georgia Tech | Center for Music Technology
College of Design

- **principle**
  1. find most salient fundamental frequency
     - e.g., with monophonic pitch tracking
  2. remove this frequency and related frequency components
     - e.g., mask or subtraction
  3. repeat until termination criterion
     - e.g., number of voices

- **challenges**
  - reliably *identify fundamental frequency* in a mixture
  - *identify/group components* and amount to subtract
    - overlapping components
    - spectral leakage
  - define *termination criterion*
    - e.g., unknown number of voices or overall energy

polyphonic pitch tracking

iterative subtraction: introduction

Georgia Tech Center for Music Technology College of Design

- **principle**
  1. find most salient fundamental frequency
     - e.g., with monophonic pitch tracking
  2. remove this frequency and related frequency components
     - e.g., mask or subtraction
  3. repeat until termination criterion
     - e.g., number of voices

- **challenges**
  - reliably *identify fundamental frequency* in a mixture
  - *identify/group components* and amount to subtract
    - overlapping components
    - spectral leakage
  - define *termination criterion*
    - e.g., unknown number of voices or overall energy

polyphonic pitch tracking

iterative subtraction: Cheveigné

Georgia Tech Center for Music Technology College of Design

1. compute squared AMDF

$$\mathrm{ASMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \left(x(i) - x(i+\eta)\right)^2$$

2. find fundamental frequency

$$\eta_\mathrm{min} = \mathrm{argmin}\left(\mathrm{ASMDF}_{xx}(\eta, n)\right)$$

3. apply comb cancellation filter, IR:

$$h(i) = \delta(i) - \delta(i - \eta_\mathrm{min})$$

4. repeat process

polyphonic pitch tracking
iterative subtraction: Cheveigné

Georgia Tech Center for Music Technology College of Design

1. compute squared AMDF

$$\mathrm{ASMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \left(x(i) - x(i+\eta)\right)^2$$

2. find fundamental frequency

$$\eta_\mathrm{min} = \mathrm{argmin}\left(\mathrm{ASMDF}_{xx}(\eta, n)\right)$$

3. apply comb cancellation filter, IR:

$$h(i) = \delta(i) - \delta(i - \eta_\mathrm{min})$$

4. repeat process

polyphonic pitch tracking

iterative subtraction: Cheveigné

1. compute squared AMDF

$$\mathrm{ASMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \big(x(i) - x(i+\eta)\big)^2$$

2. find fundamental frequency

$$\eta_\mathrm{min} = \mathrm{argmin}\left(\mathrm{ASMDF}_{xx}(\eta, n)\right)$$

3. apply comb cancellation filter, IR:

$$h(i) = \delta(i) - \delta(i - \eta_\mathrm{min})$$

4. repeat process

polyphonic pitch tracking

iterative subtraction: Cheveigné

Georgia Tech Center for Music Technology College of Design

1. compute squared AMDF

$$\mathrm{ASMDF}_{xx}(\eta, n) = \frac{1}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} \big(x(i) - x(i+\eta)\big)^2$$

2. find fundamental frequency

$$\eta_\mathrm{min} = \mathrm{argmin}\left(\mathrm{ASMDF}_{xx}(\eta, n)\right)$$

3. apply comb cancellation filter, IR:

$$h(i) = \delta(i) - \delta(i - \eta_\mathrm{min})$$

4. repeat process

polyphonic pitch tracking
iterative subtraction: Meddis

1. auditory pitch tracking:

$$r_{zz}(c,n,\eta) = \sum_{i=0}^{\mathcal{K}-1} z_c(i) \cdot z_c(i+\eta)$$

2. detect most likely frequency for all bands
3. remove all bands with a max at detected frequency
4. reiterate until most bands have eliminated

polyphonic pitch tracking

iterative subtraction: Meddis

1. auditory pitch tracking:

$$r_{zz}(c, n, \eta) = \sum_{i=0}^{\mathcal{K}-1} z_c(i) \cdot z_c(i + \eta)$$

2. detect most likely frequency for all bands
3. remove all bands with a max at detected frequency
4. reiterate until most bands have eliminated

polyphonic pitch tracking

iterative subtraction: Meddis

Georgia Tech | Center for Music Technology | College of Design

1. auditory pitch tracking:

$$r_{zz}(c, n, \eta) = \sum_{i=0}^{\mathcal{K}-1} z_c(i) \cdot z_c(i + \eta)$$

2. detect most likely frequency for all bands
3. remove all bands with a max at detected frequency
4. reiterate until most bands have eliminated

polyphonic pitch tracking

iterative subtraction: Meddis

1. auditory pitch tracking:

$$r_{zz}(c, n, \eta) = \sum_{i=0}^{\mathcal{K}-1} z_c(i) \cdot z_c(i + \eta)$$

2. detect most likely frequency for all bands
3. remove all bands with a max at detected frequency
4. reiterate until most bands have eliminated

polyphonic pitch tracking
iterative subtraction: spectral

Georgia Tech Center for Music Technology College of Design

1. find salient fundamental frequency (e.g. auditory approach, HPS)
2. estimate current model for harmonic magnitudes
3. subtract the model spectrum
4. repeat process

polyphonic pitch tracking
iterative subtraction: spectral

Georgia Tech Center for Music Technology College of Design

1. find salient fundamental frequency (e.g. auditory approach, HPS)
2. estimate current model for harmonic magnitudes
3. subtract the model spectrum
4. repeat process

polyphonic pitch tracking
iterative subtraction: spectral

Georgia Tech Center for Music Technology College of Design

1. find salient fundamental frequency (e.g. auditory approach, HPS)
2. estimate current model for harmonic magnitudes
3. subtract the model spectrum
4. repeat process

polyphonic pitch tracking
iterative subtraction: spectral

1. find salient fundamental frequency (e.g. auditory approach, HPS)
2. estimate current model for harmonic magnitudes
3. subtract the model spectrum
4. repeat process

polyphonic pitch tracking

exhaustive search

Georgia Tech Center for Music Technology College of Design

1. define set of all possible fundamental frequencies
2. compute all possible pairs of fundamental frequency
3. repeatedly filter the signal with two comb cancellation filters (all combinations)
4. find combination with minimal output energy

polyphonic pitch tracking

exhaustive search

Georgia Tech Center for Music Technology College of Design

1. define set of all possible fundamental frequencies
2. compute all possible pairs of fundamental frequency
3. repeatedly filter the signal with two comb cancellation filters (all combinations)
4. find combination with minimal output energy

polyphonic pitch tracking

exhaustive search

1. define set of all possible fundamental frequencies
2. compute all possible pairs of fundamental frequency
3. repeatedly filter the signal with two comb cancellation filters (all combinations)
4. find combination with minimal output energy

polyphonic pitch tracking

exhaustive search

1. define set of all possible fundamental frequencies
2. compute all possible pairs of fundamental frequency
3. repeatedly filter the signal with two comb cancellation filters (all combinations)
4. find combination with minimal output energy

polyphonic pitch tracking
Karjalainen and Tolonen 1/3

1. pre-whitening by frequency warped linear prediction
2. filter bank: low-pass and high-pass band (cut-off: 1 kHz)
3. HWR and smoothing
4. generalized ACF ($\beta = 2/3$):

$$r_{xx}^{\beta}(\eta, n) = \mathfrak{F}^{-1}\left\{|X(\mathrm{j}\omega)|^{\beta}\right\}$$

polyphonic pitch tracking

Karjalainen and Tolonen 1/3

1. pre-whitening by frequency warped linear prediction
2. filter bank: low-pass and high-pass band (cut-off: 1 kHz)
3. HWR and smoothing
4. generalized ACF ($\beta = 2/3$):

$$r_{xx}^{\beta}(\eta, n) = \mathfrak{F}^{-1}\left\{|X(\mathrm{j}\omega)|^{\beta}\right\}$$

polyphonic pitch tracking

Karjalainen and Tolonen 1/3

1. pre-whitening by frequency warped linear prediction
2. filter bank: low-pass and high-pass band (cut-off: 1 kHz)
3. HWR and smoothing
4. generalized ACF ($\beta = 2/3$):

$$r_{xx}^{\beta}(\eta, n) = \mathfrak{F}^{-1}\left\{|X(\mathrm{j}\omega)|^{\beta}\right\}$$

polyphonic pitch tracking

Karjalainen and Tolonen 1/3

Georgia Tech Center for Music Technology College of Design

1. pre-whitening by frequency warped linear prediction
2. filter bank: low-pass and high-pass band (cut-off: 1 kHz)
3. HWR and smoothing
4. generalized ACF ($\beta = 2/3$):

$$r_{xx}^{\beta}(\eta, n) = \mathfrak{F}^{-1}\left\{|X(\mathrm{j}\omega)|^{\beta}\right\}$$

polyphonic pitch tracking

Karjalainen and Tolonen 2/3

1. summary ACF
2. harmonic ACF processing:
   1. define temporary function:

      $$r'(\eta) = HWR(r^{\beta}_{xx}(\eta, n))$$

   2. resample (e.g. linear interpolation):

      $$\eta' = \frac{\eta}{m}$$

   3. update $r(\eta)$

      $$r'(\eta) = HWR\big(r'(\eta) - HWR(r'_m(\eta))\big)$$

polyphonic pitch tracking

Karjalainen and Tolonen 2/3

1. summary ACF
2. harmonic ACF processing:
   1. define temporary function:

$$r'(\eta) = HWR(r_{xx}^{\beta}(\eta, n))$$

   2. resample (e.g. linear interpolation):

$$\eta' = \frac{\eta}{m}$$

   3. update $r(\eta)$

$$r'(\eta) = HWR(r'(\eta) - HWR(r'_m(\eta)))$$

polyphonic pitch tracking

Karjalainen and Tolonen 2/3

Georgia Tech Center for Music Technology College of Design

1. summary ACF
2. harmonic ACF processing:
   1. define temporary function:

$$r'(\eta) = HWR(r_{xx}^{\beta}(\eta, n))$$

   2. resample (e.g. linear interpolation):

$$\eta' = \frac{\eta}{m}$$

   3. update $r(\eta)$

$$r'(\eta) = HWR\left(r'(\eta) - HWR(r'_m(\eta))\right)$$

polyphonic pitch tracking

Karjalainen and Tolonen 2/3

Georgia Tech Center for Music Technology College of Design

1. summary ACF
2. harmonic ACF processing:
   1. define temporary function:

      $$r'(\eta) = HWR(r_{xx}^{\beta}(\eta, n))$$

   2. resample (e.g. linear interpolation):

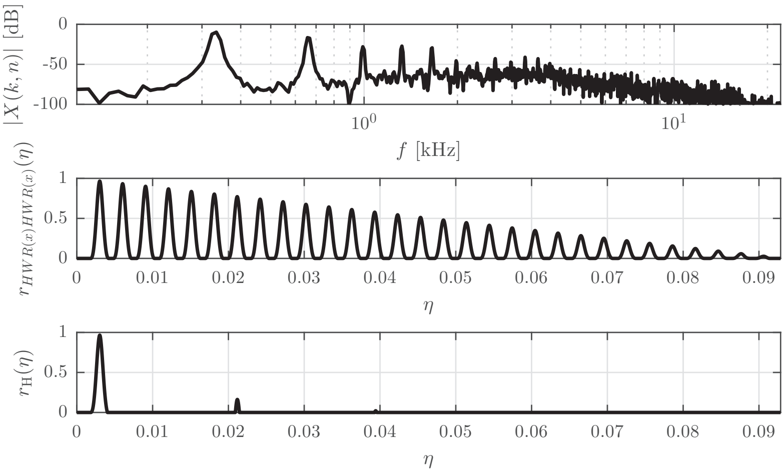
      $$\eta' = \frac{\eta}{m}$$

   3. update $r(\eta)$

      $$r'(\eta) = HWR\big(r'(\eta) - HWR(r'_m(\eta))\big)$$

polyphonic pitch tracking
Karjalainen and Tolonen 3/3

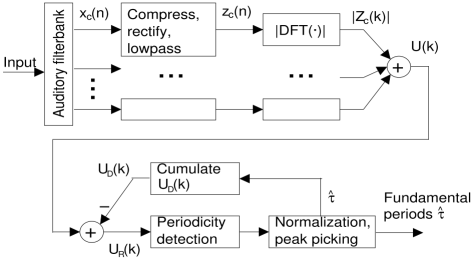
Georgia Tech Center for Music Technology College of Design



matlab source: matlab/displayKarjalainen.m

polyphonic pitch tracking
klapuri

1. gammatone **filterbank** (100 bands)
2. **normalization**, HWR, smoothing, ...
3. **STFT** per filter channel (magnitude)
4. use **delta pulse templates** to detect frequency patterns
5. **pick most salient frequencies**, remove them

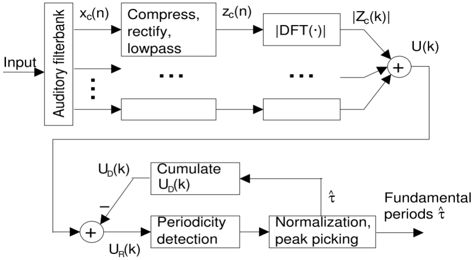


graph from[1]

[1] A. P. Klapuri, "A Perceptually Motivated Multiple-F0 Estimation Method," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2005.

polyphonic pitch tracking
klapuri

1. gammatone **filterbank** (100 bands)
2. **normalization**, HWR, smoothing, ...
3. **STFT** per filter channel (magnitude)
4. use **delta pulse templates** to detect frequency patterns
5. **pick most salient frequencies**, remove them

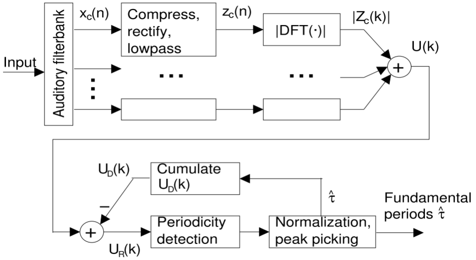


graph from[1]

[1] A. P. Klapuri, "A Perceptually Motivated Multiple-F0 Estimation Method," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2005.

polyphonic pitch tracking
klapuri

1. gammatone **filterbank** (100 bands)
2. **normalization**, HWR, smoothing, ...
3. **STFT** per filter channel (magnitude)
4. use **delta pulse templates** to detect frequency patterns
5. **pick most salient frequencies**, remove them

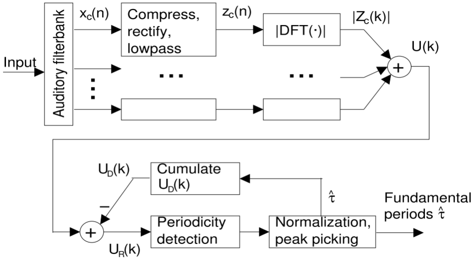


graph from[1]

[1] A. P. Klapuri, "A Perceptually Motivated Multiple-F0 Estimation Method," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2005.

polyphonic pitch tracking
klapuri

1. gammatone **filterbank** (100 bands)
2. **normalization**, HWR, smoothing, ...
3. **STFT** per filter channel (magnitude)
4. use **delta pulse templates** to detect frequency patterns
5. pick most salient frequencies, remove them



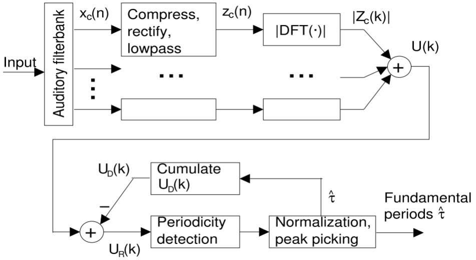
graph from[1]

[1] A. P. Klapuri, "A Perceptually Motivated Multiple-F0 Estimation Method," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2005.

polyphonic pitch tracking

klapuri

1. gammatone **filterbank** (100 bands)
2. **normalization**, HWR, smoothing, ...
3. **STFT** per filter channel (magnitude)
4. use **delta pulse templates** to detect frequency patterns
5. **pick most salient frequencies**, remove them



graph from[1]

[1] A. P. Klapuri, "A Perceptually Motivated Multiple-F0 Estimation Method," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2005.

summary
lecture content

- **polyphonic pitch detection**
  - highly challenging task with
    - unknown number of sources
    - unknown harmonic structure
    - spectral overlap of sources
    - time-varying mixture

- **typical approaches**
  - iterative subtraction (detect one pitch, remove it, repeat analysis)
  - multi-band processing