Pitch Analysis for Active Music Discovery

Justin Salamon\textsuperscript{1,2}

\texttt{@justin_salamon} \texttt{justin.salamon@nyu.edu} \texttt{www.justinsalamon.com}

\textsuperscript{1}Music and Audio Research Laboratory
\textsuperscript{2}Center for Urban Science and Progress
New York University, USA

Collaborators: Juan Pablo Bello, Rachel Bittner, Jordi Bonada, Juan J. Bosch, Jose Miguel Diaz-Bañez, Chris Cannam, Francisco Escobar, Slim Essid, Emilia Gómez, Paco Gómez, Sankalp Gulati, Helga Jiang, Edith Law, Matthias Mauch, Joaquin Mora, Sergio Oramas, Geoffroy Peeters, Aggelos Pikrakis, Axel Röbel, Bruno Rocha, Joan Serrà, Xavier Serra, Tim Tse, Alex Williams

Thursday June 23\textsuperscript{rd} 2016
In A Nutshell
In A Nutshell

- GOAL: active music discovery
In A Nutshell

- GOAL: active music discovery
- MEANS: pitch content analysis
In A Nutshell

- GOAL: active music discovery
- MEANS: pitch content analysis
- CHALLENGE: data scarcity (and solution!)
Active Music Discovery
Music Recommendation
Music Recommendation
Song A → Versions? → Yes! Similarity = 0.92

Song B →

(Salamon, Serrà & Gómez, 2013)
**COVER SONG ID**

Song A  →  Versions?

Song B  →  Yes!

Similarity = 0.92

**(Salamon, Serrà & Gómez, 2013)**

**QUERY-BY-HUMMING**

**(Salamon, Serrà & Gómez, 2013)**
Active Search & Discovery

Cover Song ID

Song A → Versions?

Song B → Yes! Similarity = 0.92

(Salamon, Serrà & Gómez, 2013)

Query-by-Humming

(Salamon, Serrà & Gómez, 2013)

Melodic Pattern Discovery

(Pikrakis et al., 2012)
Active Search & Discovery

COVER SONG ID

Song A  Versions?
Song B  Yes!
Similarity = 0.92

(Salamon, Serrà & Gómez, 2013)

QUERY-BY-HUMMING

MELODIC PATTERN DISCOVERY

(Pikrakis et al., 2012)

TONIC ID

(Salamon, Gulati & Serra, 2012)
Active Search & Discovery

**Cover Song ID**

Song A 
Versions? 
Yes!
Similarity = 0.92

(Salamon, Serrà & Gómez, 2013)

**Query-by-Humming**

(Salamon, Serrà & Gómez, 2013)

**Melodic Pattern Discovery**

(Pikrakis et al., 2012)

**Tonic ID**

(Pikrakis et al., 2012)

**Singing Style Classification**

(Salamon, Rocha & Gómez, 2012)
Machine Learning for Pitch Analysis
Melody Extraction
Melody Extraction
Melody Extraction

![Image of a song cover with frequency and time graph]

- Frequency (Hz) on the y-axis
- Time (s) on the x-axis
- Graph showing frequency changes over time
Melodia (Salamon & Gómez, 2012)
Melodia (Salamon & Gómez, 2012)
Melodia (Salamon & Gómez, 2012)

Contour Extraction → Contour Selection → Melody $f_0$
Melodia (Salamon & Gómez, 2012)

Contour Extraction → Contour Selection → Melody $f_0$
Melodia (Salamon & Gómez, 2012)

Contour Extraction → Contour Selection → Melody $f_0$
Melodia (Salamon & Gómez, 2012)

Contour Extraction → Contour Selection → Melody $f_0$
Melodia (Salamon & Gómez, 2012)

**Contour Extraction**
- Audio signal
  - Equal loudness filter
  - Spectral transform
  - Frequency/amplitude correction
- Sinusoid extraction
  - Spectral peaks
- Salience function
  - Bin salience mapping with harmonic weighting
  - Time-frequency salience
- Pitch contour creation
  - Peak filtering
  - Peak streaming
  - Contour characterisation

**Contour Selection**
- Pitch contours
- Voicing detection
  - Melody pitch mean
  - Octave error removal
  - Melody pitch mean
  - Pitch outlier removal
- Melody peak selection
- Melody $f_0$ sequence
- Melody $f_0$ sequence

**Melody $f_0$**
Melodia (Salomon & Gómez, 2012)

Contour Extraction

Contour Selection

Melody $f_0$

Audio signal

Sinusoid extraction

Salience function

Pitch contour creation

DSP

Melody selection

Iterative

Melody peak selection

Voice detection

Pitch outlier removal

Melody pitch mean

Octave error removal

Melody pitch mean

Pitch mean

Melody $f_0$ sequence
Melodia (Salamon & Gómez, 2012)
Melodia (Salamon & Gómez, 2012)
Melodia (Salamon & Gómez, 2012)
Melodia (Salamon & Gómez, 2012)
Melodia (Salamon & Gómez, 2012)

Contour Extraction

Audio signal

- Equal loudness filter
- Spectral transform
- Frequency/amplitude correction

Sinusoid extraction

- Equal loudness filter
- Spectral transform
- Frequency/amplitude correction

Salience function

- Spectral transformation
- Time-frequency salience

Pitch contour creation

- Peak filtering
- Peak streaming
- Contour characterisation

Contour Selection

- Pitch contours
- Melody pitch mean
- Octave error removal
- Melody pitch mean
- Pitch outlier removal
- Melody peak selection

Melody selection

- Iterative
- Melody $f_0$ sequence

DSP
Melodia (Salamon & Gómez, 2012)

Contour Extraction → Contour Selection → Melody $f_0$

Audio signal

- Equal loudness filter
- Spectral transform
- Frequency/amplitude correction

Sinusoid extraction

- Spectral salience mapping
- Time-frequency salience

Salience function

- Bin salience mapping with harmonic weighting

Peak filtering → Peak streaming → Contour characterisation

Pitch contour creation

- Melody peak selection
- Pitch outlier removal
- Melody pitch mean
- Octave error removal
- Voicing detection
Melodia (Salamon & Gómez, 2012)

Contour Extraction → Contour Selection → Melody $f_0$

Audio signal

- Equal loudness filter
- Spectral transform
- Frequency/amplitude correction

Sinusoid extraction

- Spectral peaks
- Time-frequency salience
  - Bin salience mapping with harmonic weighting

Salience function

- Peak filtering
- Peak streaming
- Contour characterisation

Pitch contour creation

MIREX: 75%
MedleyDB: 57%
Contour Classification (Salamon, Peeters & Röbel, 2012)

Audio signal → Contour Extraction → Contour Selection → Melody $f_0$

### Contour Extraction

- **Sinusoid extraction**
  - Equal loudness filter
  - Spectral transform
  - Frequency/amplitude correction

### Contour Selection

- **Salience function**
  - Spectral peaks
  - Time-frequency salience
  - Bin salience mapping with harmonic weighting

- **Pitch contour creation**
  - Peak filtering
  - Peak streaming
  - Contour characterisation

- **Pitch contours**
- **Melody contour features**
- **Accompaniment contour features**
Contour Classification (Bittner et al., 2015)

Contour Extraction

Audio signal
- Equal loudness filter
- Spectral transform
- Frequency/amplitude correction

Sinusoid extraction
- Spectral bin salience mapping with harmonic weighting
- Time-frequency salience

Salience function
- Peak filtering
- Peak streaming
- Contour characterisation

Pitch contour creation

Contour Selection

Melody $f_0$

Melody peak selection
- Pitch outlier removal
- Melody pitch mean
- Octave error removal

Voicing detection

Contour Extraction 

Contour Selection
Contour Classification (Bittner et al., 2015)

Contour Extraction → Contour Selection → Melody $f_0$

- Melody
- Audio signal
- Sinusoid extraction
- Salience function
- Pitch contour creation
- Melody selection
- Iterative
- Melody $f_0$ sequence
- Spectral peaks
- Time-frequency salience
- Bin salience mapping with harmonic weighting
- Contour characterisation
- Peak streaming
- Peak filtering
- Frequency/amplitude correction
- Spectral transform
- Equal loudness filter
- Melody peak selection
- Voicing detection
- Melody $f_0$ pitch mean
- Octave error removal
- Pitch outlier removal
- Melody $f_0$ pitch mean

Graphical representation:
- Time (s) vs. Frequency (Hz) chart
- Tree-like structure representing contour selection

Legend:
- MARL
- CUSP
Contour Classification (Bosch et al., 2016)

Contour Extraction $\rightarrow$ Contour Selection $\rightarrow$ Melody $f_0$
Contour Classification (??? , 2017)

Contour Extraction → Contour Selection → Melody $f_0$

Audio

Pitch Salience Estimation

H.Sum SIMM

Pitch Salience Combination

HS $H_f_0$

Fn

Gn

O

Gn

?
Contour Classification (???, 2017)
Data Scarcity
Continuous Melody $f_0$ Annotation
Continuous Melody $f_0$ Annotation
Continuous Melody $f_0$ Annotation

Monophonic Pitch Tracker
Continuous Melody $f_0$ Annotation

Monophonic Pitch Tracker

Tony (Mauch et al., 2015)
Datasets for Melody Extraction Eval
Datasets for Melody Extraction Eval

- MIREX
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
Datasets for Melody Extraction Eval

- MIREX
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- MIR1K: 2.2 hr (1000 excerpts), C-Pop
Datasets for Melody Extraction Eval

- MIREX
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- MIR1K: 2.2 hr (1000 excerpts), C-Pop
- RWC-Pop: 6.8 hr (100 songs), J-Pop
Datasets for Melody Extraction Eval

- **MIREX**
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- **MIR1K**: 2.2 hr (1000 excerpts), C-Pop
- **RWC-Pop**: 6.8 hr (100 songs), J-Pop
- **MedleyDB** (Bittner et al., 2014):
  - 7.5 hr (108 songs), varied
Datasets for Melody Extraction Eval

- **MIREX**
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- **MIR1K**: 2.2 hr (1000 excerpts), C-Pop
- **RWC-Pop**: 6.8 hr (100 songs), J-Pop
- **MedleyDB (Bittner et al., 2014)**:
  - 7.5 hr (108 songs), varied
- **MedleyDBv2 (coming soon)**:
  - ~17hr (~240 songs), varied
Datasets for Melody Extraction Eval

- **MIREX**
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- **MIR1K**: 2.2 hr (1000 excerpts), C-Pop
- **RWC-Pop**: 6.8 hr (100 songs), J-Pop
- **MedleyDB (Bittner et al., 2014)**:
  - 7.5 hr (108 songs), varied
- **MedleyDBv2 (coming soon)**:
  - ~17hr (~240 songs), varied

108 songs

~50 annotator-hours
Datasets for Melody Extraction Eval

- **MIREX**
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- **MIR1K**: 2.2 hr (1000 excerpts), C-Pop
- **RWC-Pop**: 6.8 hr (100 songs), J-Pop
- **MedleyDB (Bittner et al., 2014)**:
  - 7.5 hr (108 songs), varied
- **MedleyDBv2 (coming soon)**:
  - ~17hr (~240 songs), varied

108 songs

~50 annotator-hours

20 million songs
Datasets for Melody Extraction Eval

- **MIREX**
  - ADC2004: 6 minutes (20 excerpts)
  - MIREX05: 12 minutes (25 excerpts)
- **MIR1K**: 2.2 hr (1000 excerpts), C-Pop
- **RWC-Pop**: 6.8 hr (100 songs), J-Pop
- **MedleyDB (Bittner et al., 2014)**:
  - 7.5 hr (108 songs), varied
- **MedleyDBv2 (coming soon)**:
  - ~17hr (~240 songs), varied

108 songs  
~50 annotator-hours  
20 million songs  
~1057 annotator-years
Data Scarcity
Overcoming Data Scarcity
Crowdsourcing Melody Note Annotations
Crowdsourcing Melody Note Annotations

Ensemble (Tse et al., 2016)
Crowdsourcing Melody Note Annotations

Ensemble (Tse et al., 2016)
Data Augmentation: $f_0$ Annotation-by-Synthesis
Data Augmentation: $f_0$ Annotation-by-Synthesis
Data Augmentation: $f_0$ Annotation-by-Synthesis

Monophonic Pitch Tracker
Data Augmentation: $f_0$ Annotation-by-Synthesis
Data Augmentation: $f_0$ Annotation-by-Synthesis

Monophonic Pitch Tracker → Cleaning + Smoothing → Sinusoidal Modelling
Data Augmentation: $f_0$ Annotation-by-Synthesis

- Monophonic Pitch Tracker
- Cleaning + Smoothing
- Sinusoidal Modelling

Graph showing frequency (Hz) over time (s) with peaks and valleys indicating pitch variations.
Data Augmentation: $f_0$ Annotation-by-Synthesis

1. Monophonic Pitch Tracker
2. Cleaning + Smoothing
3. Sinusoidal Modelling
4. Synthesis
Data Augmentation: $f_0$ Annotation-by-Synthesis

Monophonic Pitch Tracker → Cleaning + Smoothing → Sinusoidal Modelling

Mixing ← Synthesis

Frequency (Hz)

Time (s)
Data Augmentation: $f_0$ Annotation-by-Synthesis
Data Augmentation: $f_0$ Annotation-by-Synthesis
Data Augmentation: $f_0$ Annotation-by-Synthesis

Algorithm 1

Algorithm 2

Algorithm 3

Metrics:  A  B  C  D  E

Original data + manual annotations

Synthesized data + automatic annotations
Summary
Summary

- Active music discovery: user plays active role in retrieval
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  - Require (auto) extraction of pitch content: melody extraction
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  - Require (auto) extraction of pitch content: melody extraction
- Melody extraction: classification at contour level
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  - Require (auto) extraction of pitch content: melody extraction
- Melody extraction: classification at contour level
- Data scarcity: can’t explore high-capacity (and data hungry) models
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  - Require (auto) extraction of pitch content: melody extraction
- Melody extraction: classification at contour level
- Data scarcity: can’t explore high-capacity (and data hungry) models
- Solutions:
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  - Require (auto) extraction of pitch content: melody extraction
- Melody extraction: classification at contour level
- Data scarcity: can’t explore high-capacity (and data hungry) models
- Solutions:
  - Crowdsourcing
Summary

Active music discovery: user plays active role in retrieval
  Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  Require (auto) extraction of pitch content: melody extraction
Melody extraction: classification at contour level
Data scarcity: can’t explore high-capacity (and data hungry) models

Solutions:
  Crowdsourcing
  Data augmentation: annotation-by-synthesis
Summary

- Active music discovery: user plays active role in retrieval
  - Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  - Require (auto) extraction of pitch content: melody extraction

- Melody extraction: classification at contour level
- Data scarcity: can’t explore high-capacity (and data hungry) models

- Solutions:
  - Crowdsourcing
  - Data augmentation: annotation-by-synthesis

- Thanks!
Summary

‣ Active music discovery: user plays active role in retrieval
  ‣ Examples: QBH, search-by-singing-style, computational (ethno)musicology...
  ‣ Require (auto) extraction of pitch content: melody extraction
‣ Melody extraction: classification at contour level
‣ Data scarcity: can’t explore high-capacity (and data hungry) models
‣ Solutions:
  ‣ Crowdsourcing
  ‣ Data augmentation: annotation-by-synthesis
‣ Thanks!

@justin_salamon  justin.salamon@nyu.edu  www.justinsalamon.com