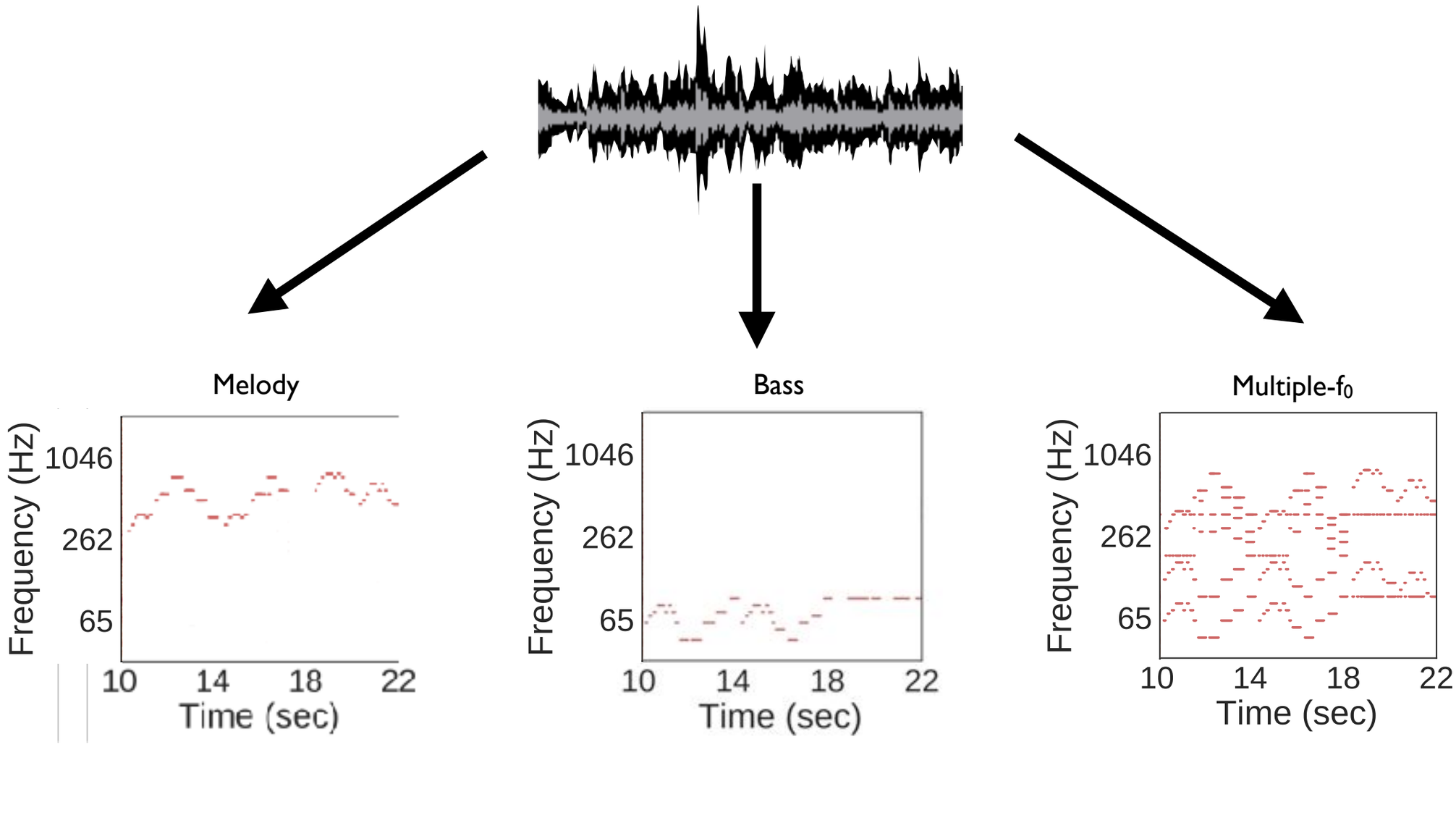
Polyphonic Fundamental Frequency Estimation

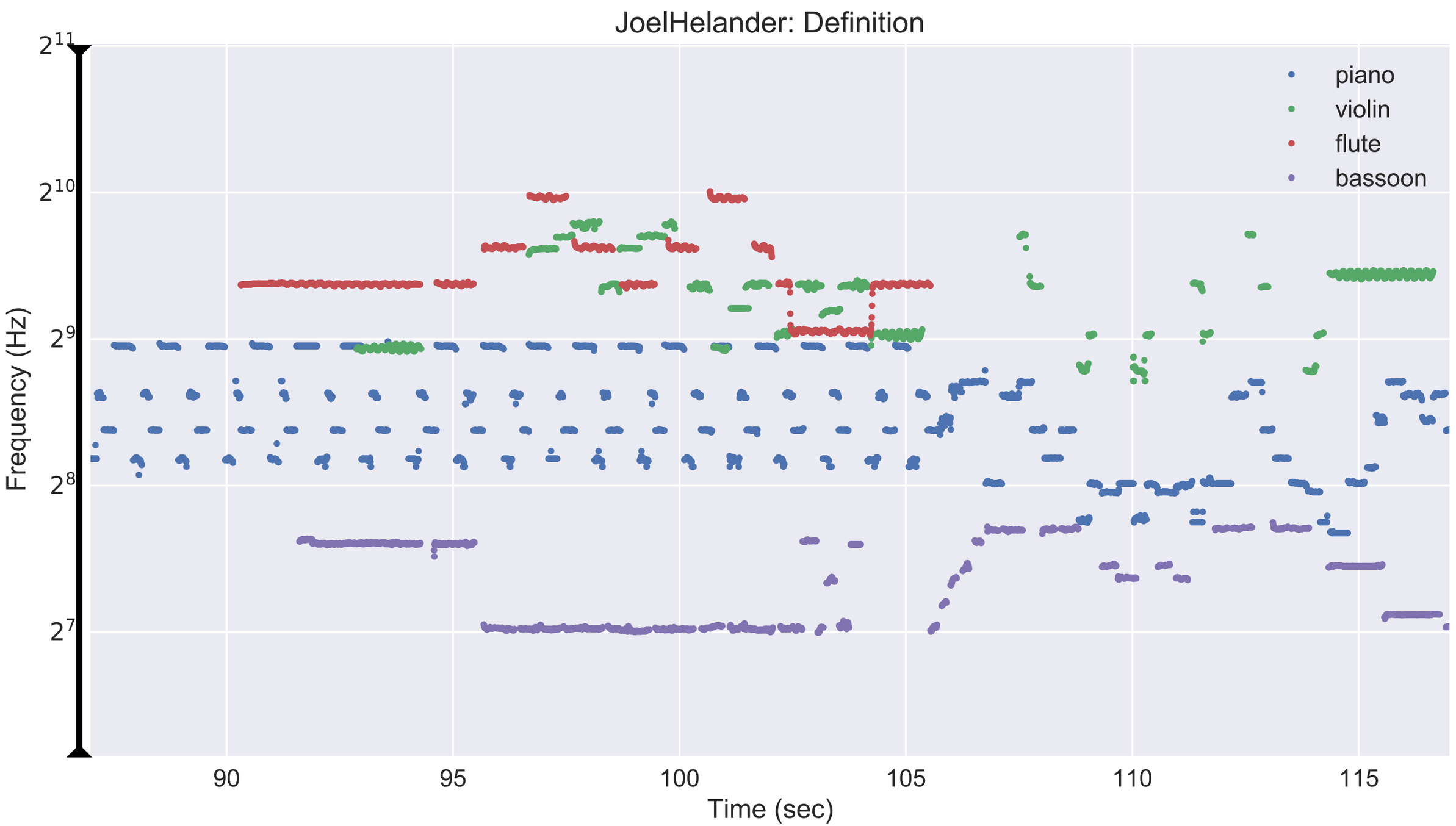
f_0 Estimation Tutorial - Part 2

Rachel Bittner @r4b1tt

Polyphonic f_0 Estimation



Polyphonic f_0 Estimation

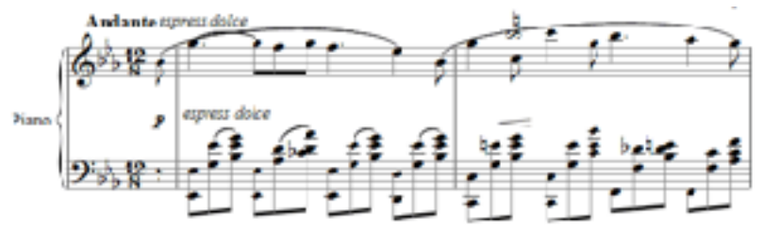


Polyphonic f_0 estimation tasks

- Multiple-f_0/multi-pitch estimation 
- Melody estimation 
- Vocal-f_0 estimation (not the same!) 
- Bass line estimation 
- (solo) polyphonic source transcription 
- source-specific f_0 estimation 

Applications

- Transcription

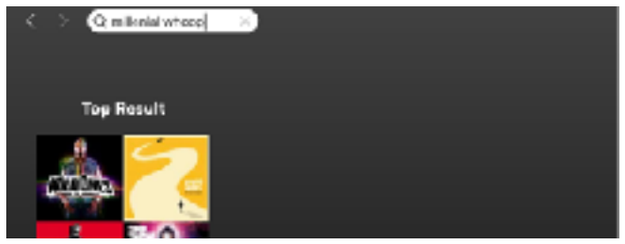


- Music Generation + Style Transfer

- Audio Effects

- Performance Analysis

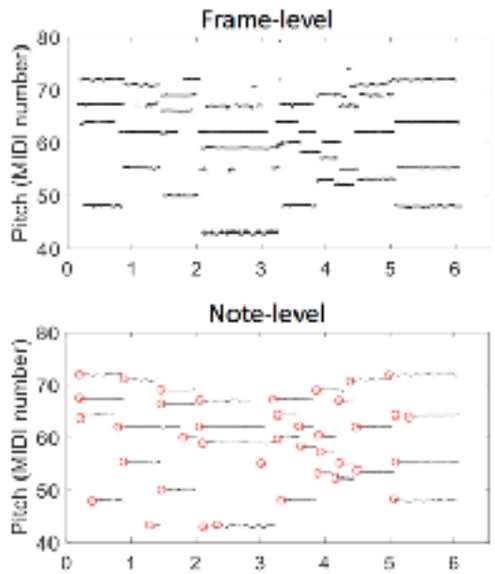
- Similarity and Retrieval

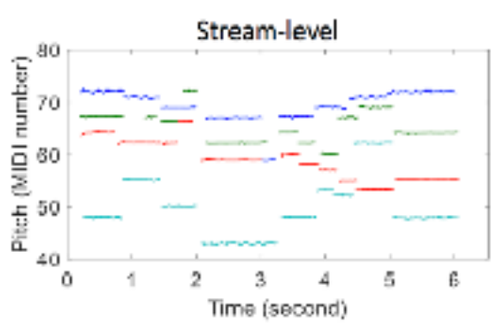


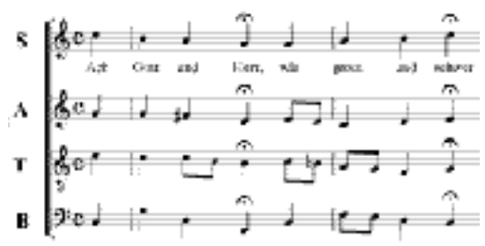
Background

Related Tasks

- frame-level f_0 estimation
- Note estimation
- Streaming
- Score transcription

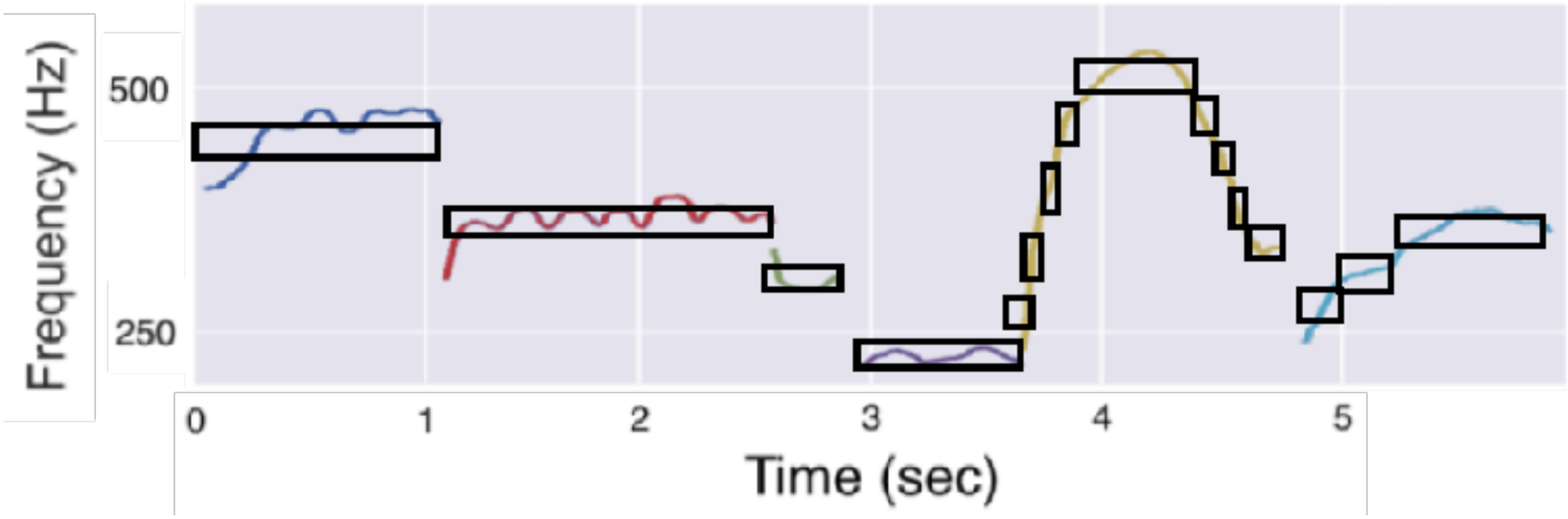






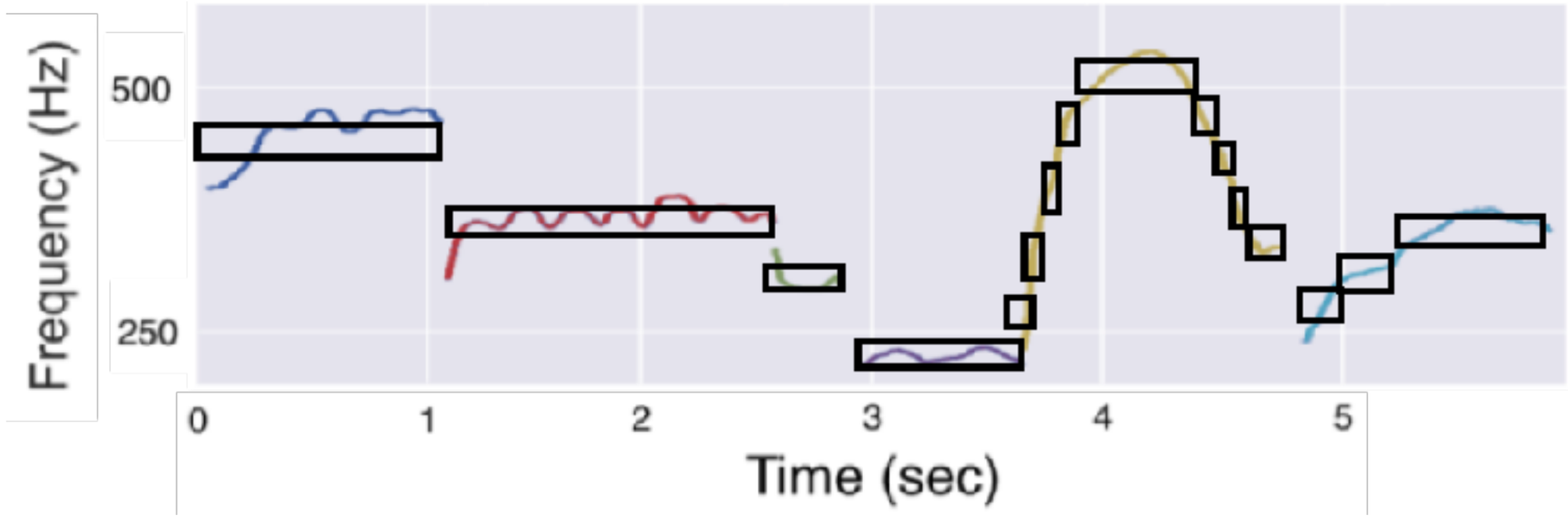
Emmanouil Benetos, Simon Dixon, Zhiyao Duan, and Sebastian Ewert
"Automatic Music Transcription: An Overview"
To appear - *IEEE Signal Processing Magazine*

Time Resolution



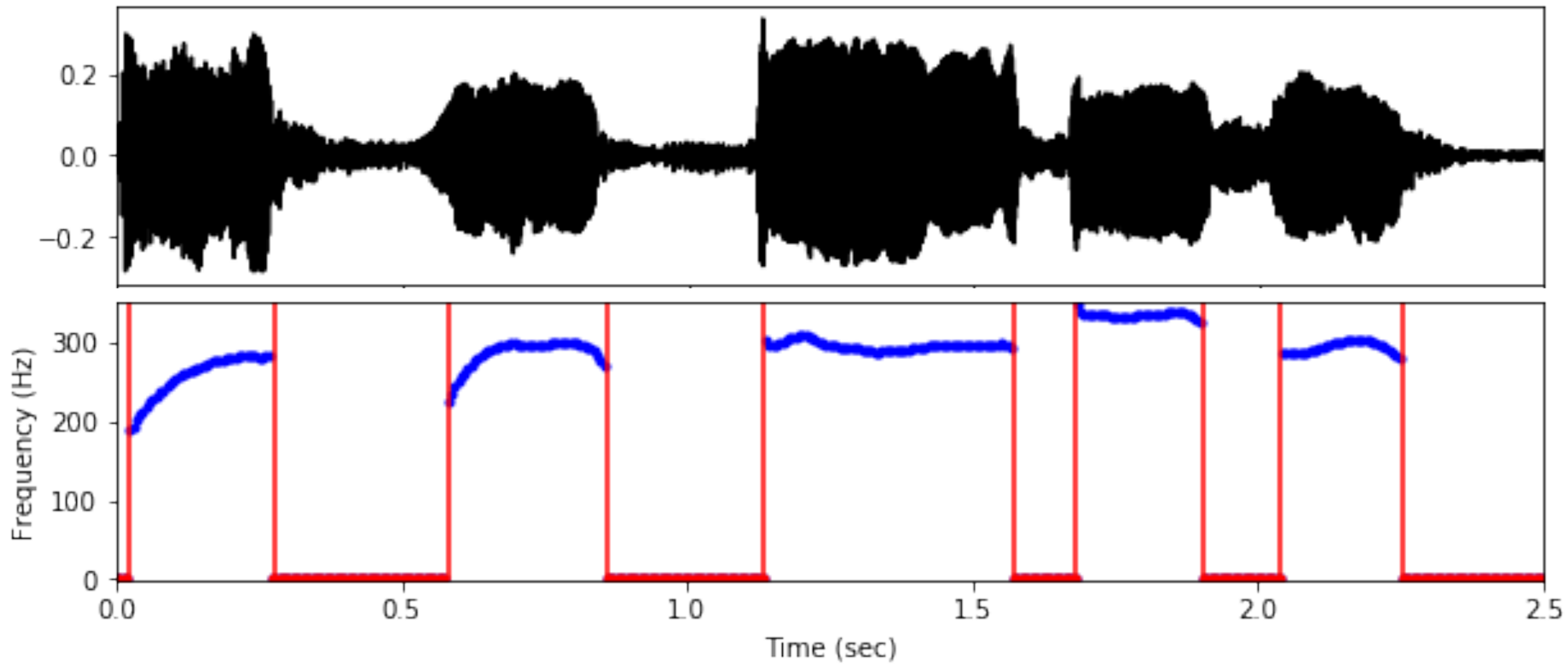
- **Fine**: f0 trajectories - (time, f0)
- **Coarse**: Notes - (start time, end time, f0)
- **Coarsest**: Quantized Notes - times are quantized to fractional beats

Frequency Resolution



- **Discrete**: Semitone resolution (integer midi notes)
- "**Continuous**": less than semitone resolution (e.g. 1/5th semitone)

Voicing



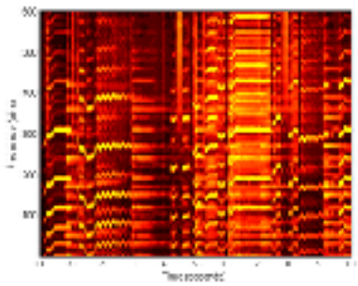
- "**Voiced**" - represented with f_0 value > 0
- "**Unvoiced**" - represented with values <= 0

Approaches to f_0 Estimation

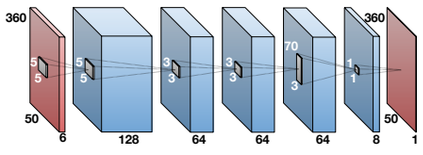
Typical Approach

Polyphonic Music

Time Frequency Transform

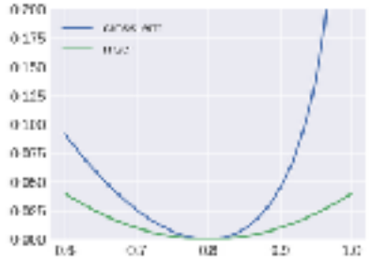


Model

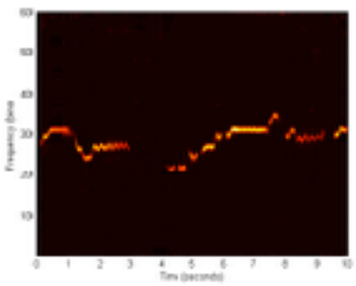


+

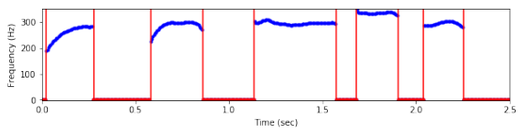
Optimization Function

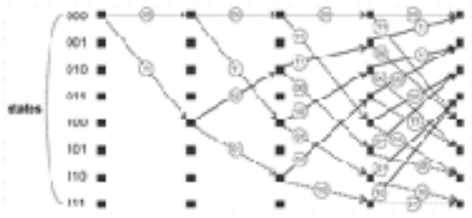


Salience Representation

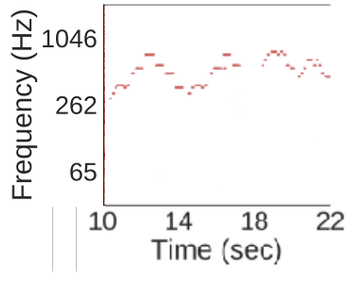


Voicing Determination + Decoding





f_0 Time Series

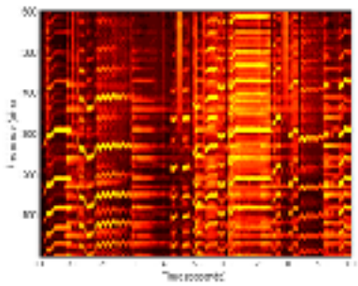


time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

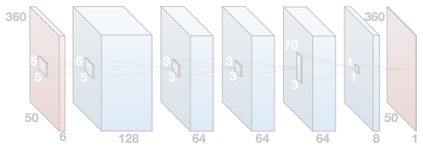
Typical Approach

Polyphonic Music

Time Frequency Transform



Model

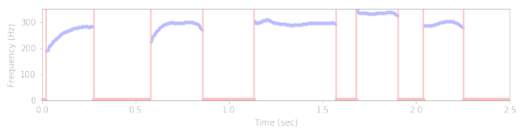


+

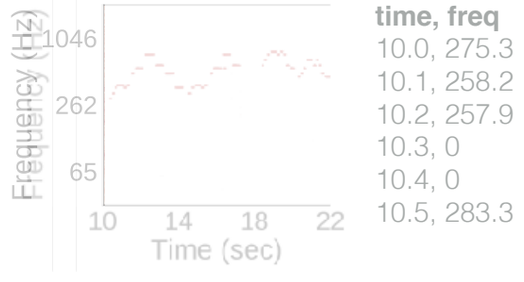
Optimization Function

Salience Representation

Voicing Determination + Decoding



f_0 Time Series



Polyphonic f_0 Estimation

3/7: Approaches

Rachel Bittner @r4b1tt

Time-Frequency Transforms

Magnitude STFT

Matija Marolt
"On Finding Melodic Lines in Audio Recordings"
DAFx (2004)

Multi-resolution FFT

Karin Dressler
"Towards Computational Auditory Scene Analysis: Melody Extraction from Polyphonic Music"
CMMR (2012)

Magnitude CQT

Benoit Fuentes, Antoine Liutkus, Roland Badeau, Gaël Richard
"Probabilistic Model for Main Melody Extraction using constant-Q transform"
ICASSP (2012)

HCQT

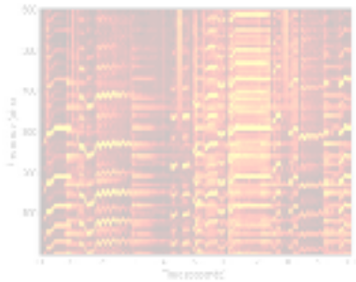
Rachel Bittner, Brian McFee, Justin Salamon, Peter Li, Juan Pablo Bello
"Deep Salience Representations for F0 Estimation in Polyphonic Music"
ISMIR (2017)

Justin Salamon, Emilia Gómez, Jordi Bonada
"Sinusoid Extraction and Salience Function Design for Predominant Melody Estimation"
DAFx (2011)

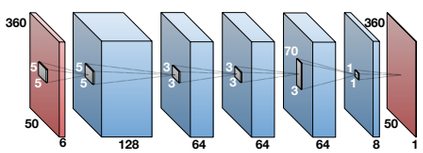
Typical Approach

Polyphonic Music

Time Frequency Transform

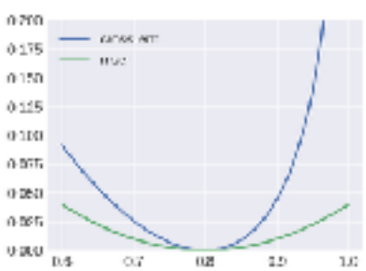


Model



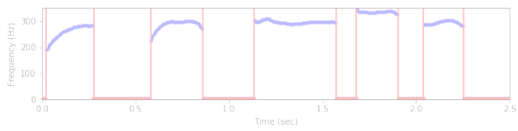
+

Optimization Function

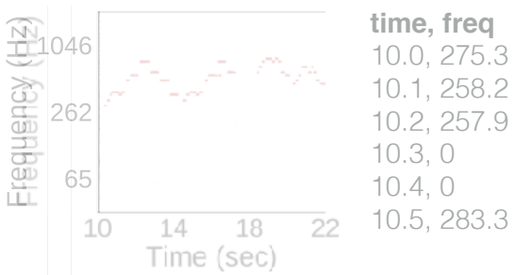


Salience Representation

Voicing Determination + Decoding



f_0 Time Series



Models + Optimization

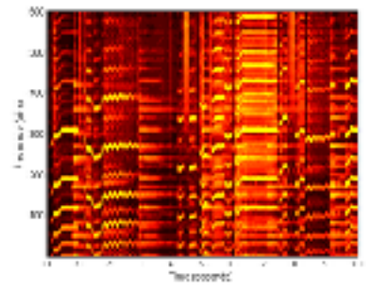
1) Salience Modeling

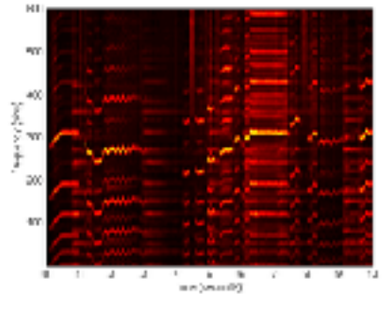
2) Nonnegative Matrix Factorization

3) Neural Networks

Models + Optimization

1) Salience Modeling





Harmonic Summation

Matti Ryynänen and Anssi Klapuri
“Automatic transcription of melody, baseline and chords in polyphonic music”
CMJ (2008)

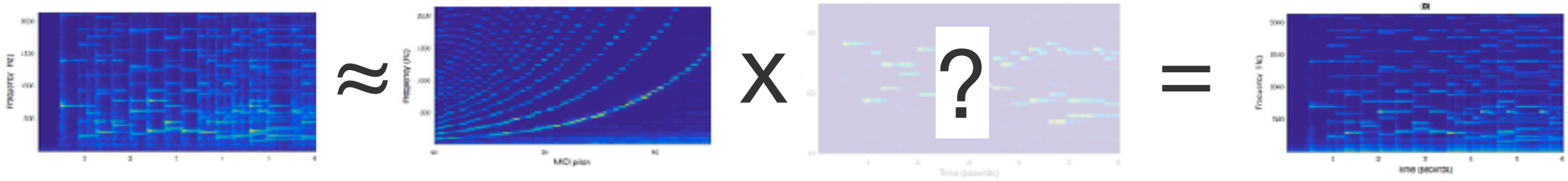
Source Filter Modeling

Jean Louis Durrieu, Gaël Richard, Bertrand David, Cédric Févotte
“Source/filter model for unsupervised main melody extraction from polyphonic audio signals.”
IEEE TASLP (2010)

Models + Optimization

2) Nonnegative Matrix Factorization

Input Templates Activations Reconstruction

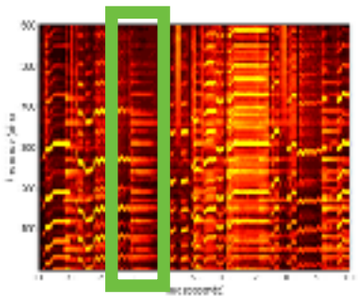


find the  Activations that minimize d(Input , Reconstruction)

Emmanouil Benetos, Simon Dixon, Zhiyao Duan, and Sebastian Ewert
"Automatic Music Transcription: An Overview"
To appear - *IEEE Signal Processing Magazine*

Models + Optimization

3) Neural Networks



DNN

CNN

RNN

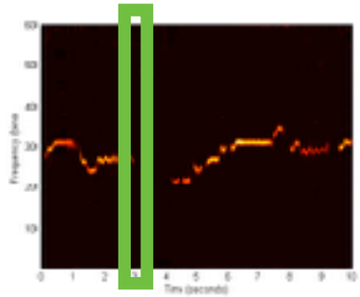
CRNN

+

Binary Classification

Multiclass Classification

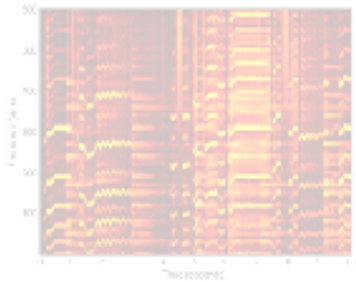
Multilabel Classification



Typical Approach

Polyphonic Music

Time Frequency Transform

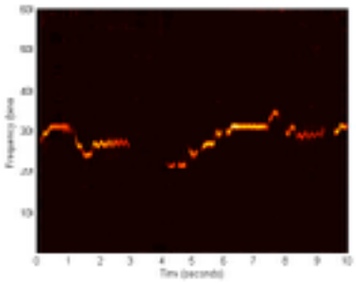


Model

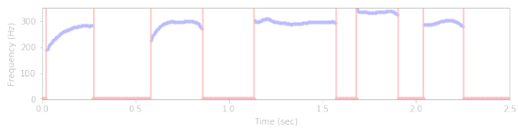
+

Optimization Function

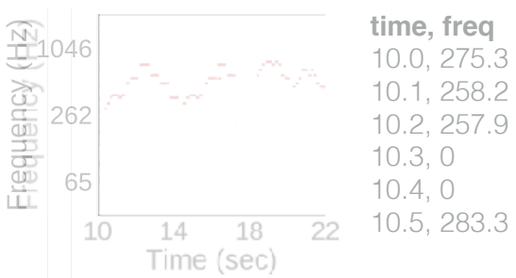
Salience Representation



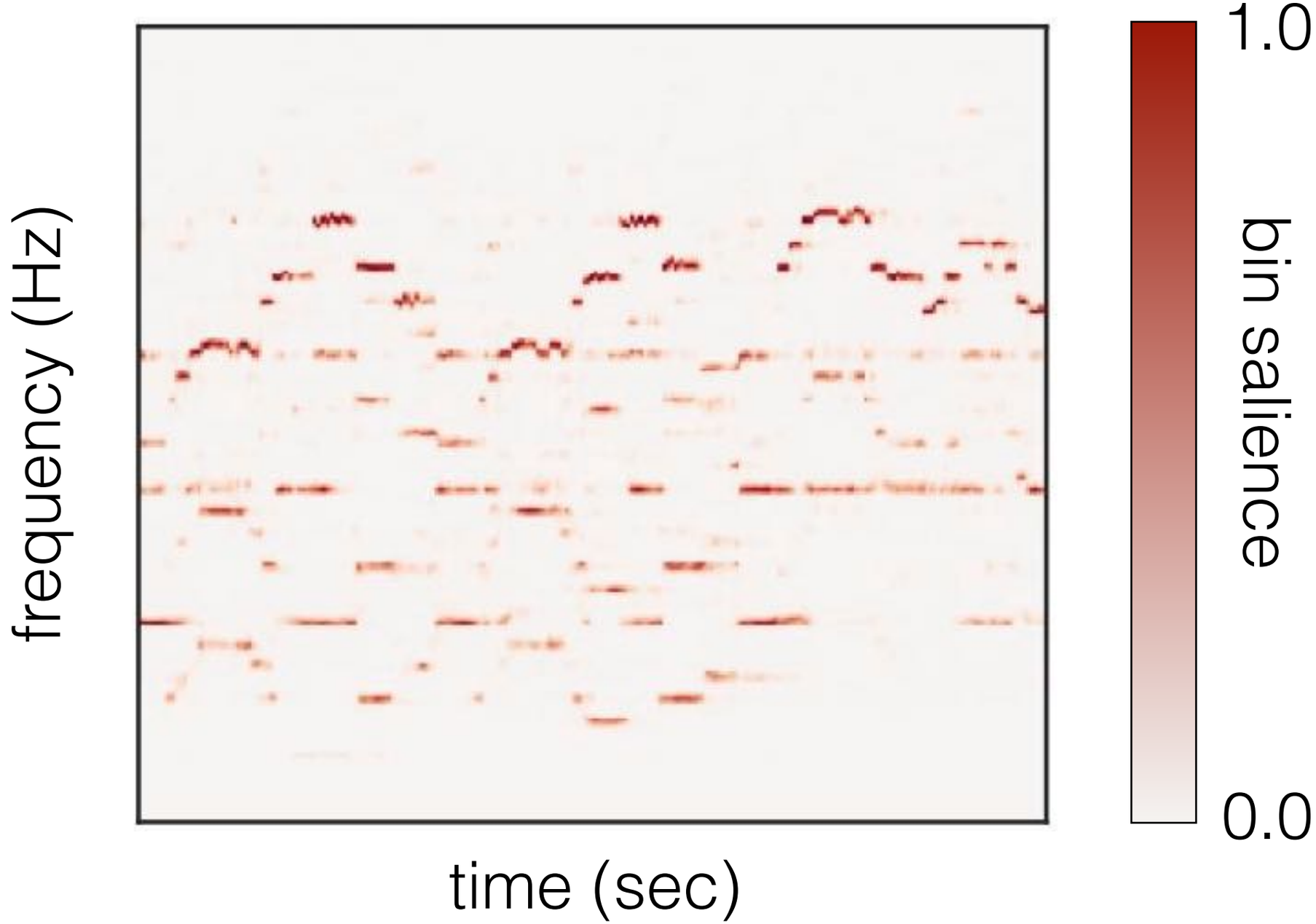
Voicing Determination + Decoding



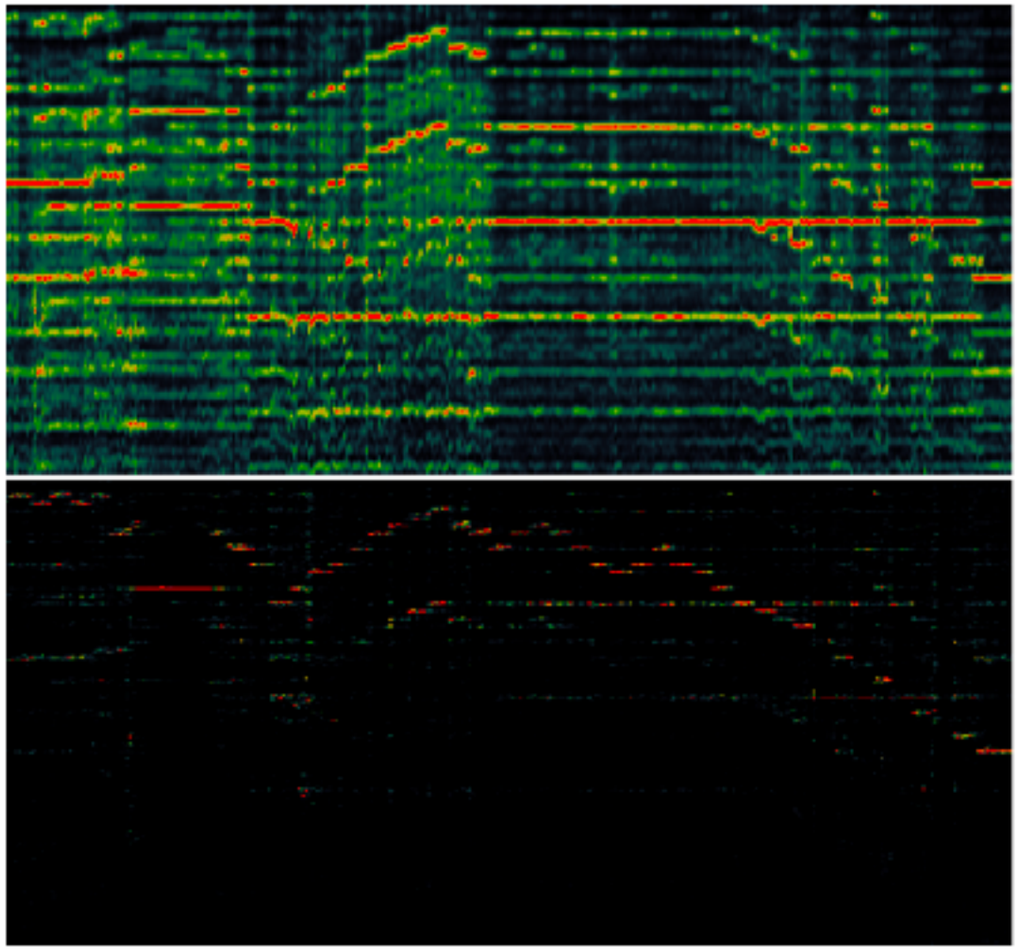
f_0 Time Series



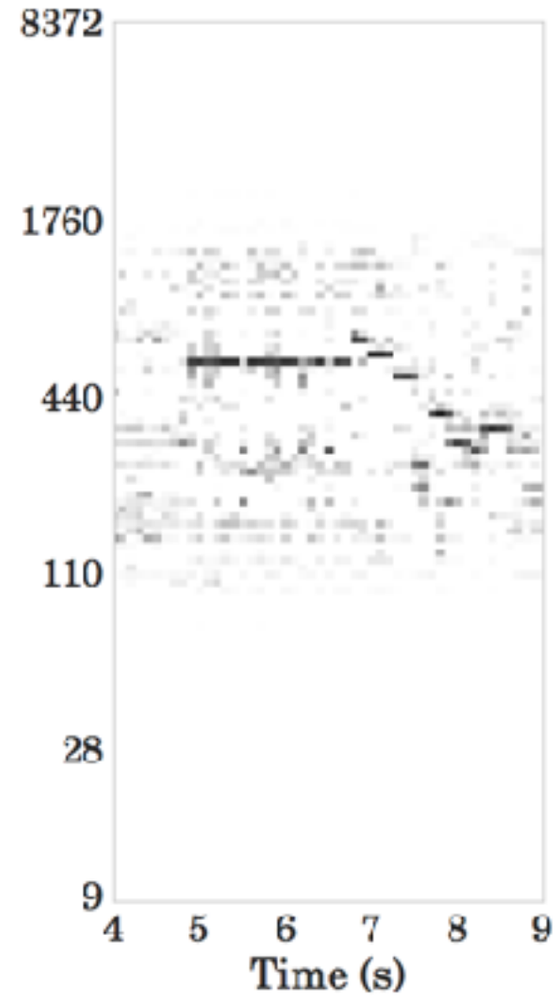
Salience Representations



Salience Representations



Juan José Bosch, Ricard Marxer, Emilia Gómez
“Evaluation and Combination of Pitch Estimation Methods for Melody Extraction in Symphonic Classical Music”
JNMR (2016)

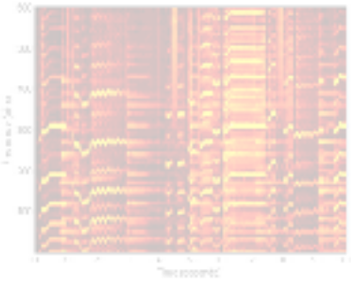


Stefan Balke, Christan Dittmar, Jakob Abeßer, Meinard Müller
“Data Driven Solo Voice Enhancement for Jazz Music Retrieval”
ICASSP (2017)

Typical Approach

Polyphonic Music

Time Frequency Transform



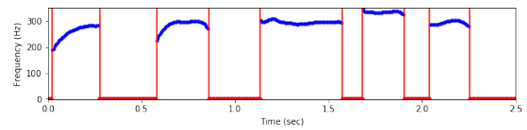
Model

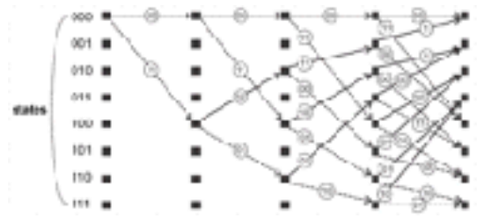
+

Optimization Function

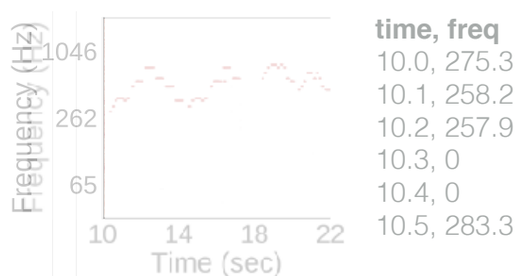
Salience Representation

Voicing Determination + Decoding





f_0 Time Series

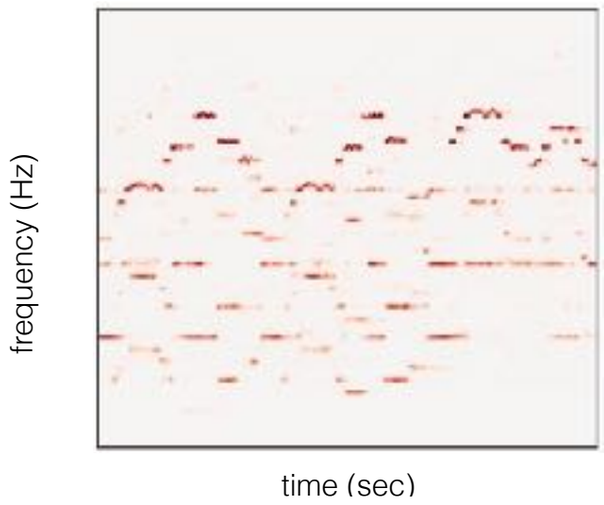


time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

Voicing and Decoding

Thresholding

e.g. Vocal Activity Detection



argmax

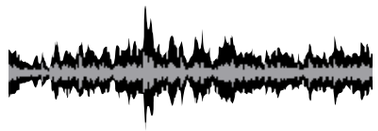
Viterbi

Contour Tracking

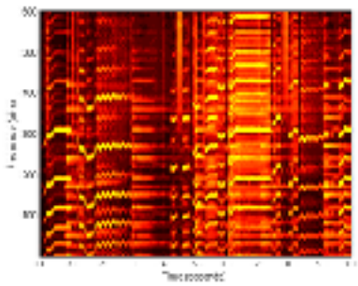
Example 1: Salience

Justin Salamon, Emilia Gómez
"Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics"
IEEE TASLP (2012)

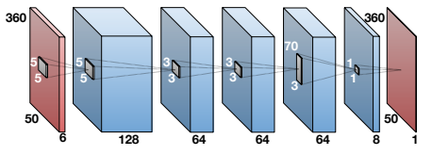
Polyphonic Music



STFT

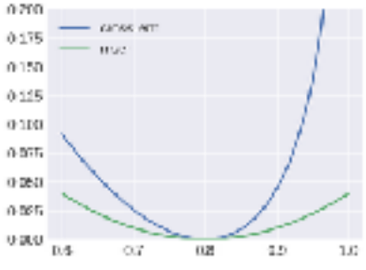


Peak Filtering + Harmonic Summation

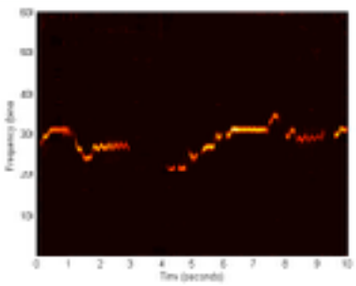


+

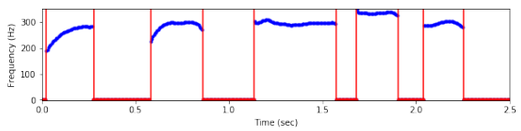
(No optimization)

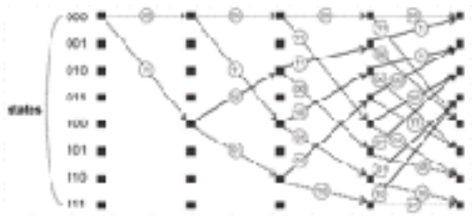


Salience Representation

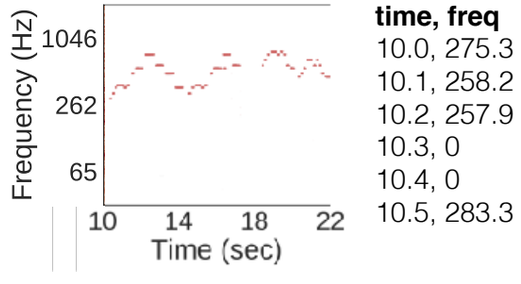


Contour Tracking + Contour Selection





f_0 Time Series



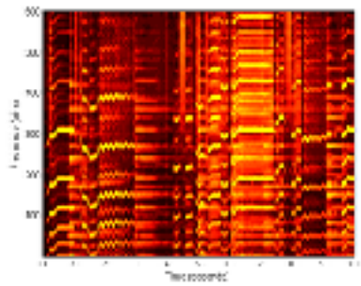
time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

Example 2: NMF

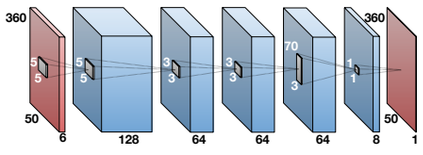
Emannouil Benetos, Simon Dixon
"Multiple-instrument polyphonic music transcription using a temporally constrained shift-invariant model"
JASA (2013)

Polyphonic Music

Log-frequency Spectrogram

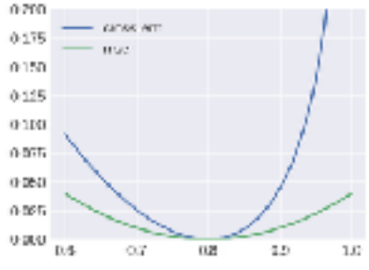


Shift Invariant PLCA

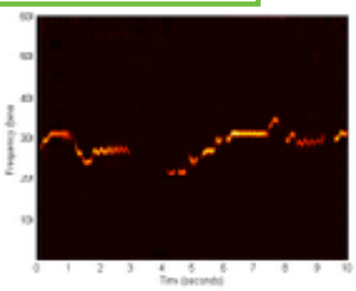


+

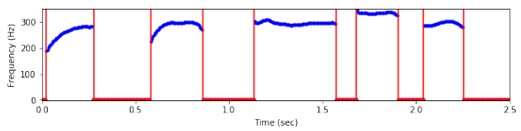
KL-Divergence

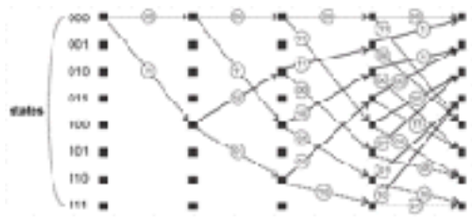


Template Activations

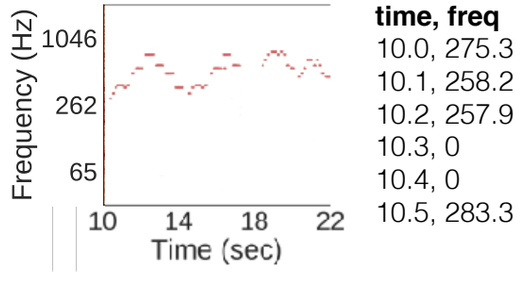


Pitch-wise HMM





f_0 Time Series

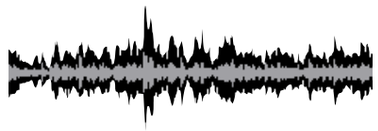


time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

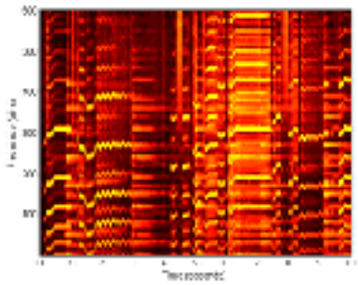
Example 3: Neural Network 1

François Rigaud, Mattieu Radenen
“Singing Voice Melody Transcription using Deep Neural Netowrks”
ISMIR (2016)

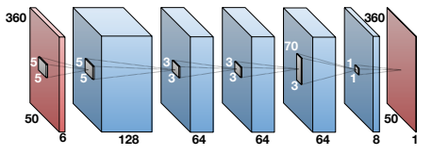
Polyphonic Music



STFT + HPSS

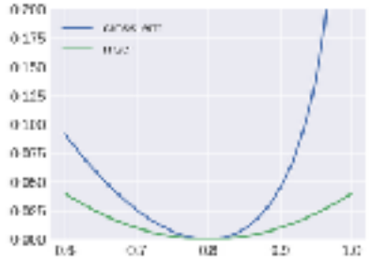


Fully Connected Network

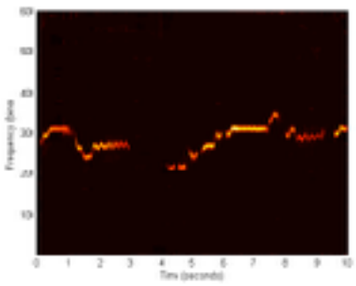


+

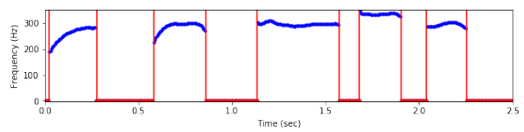
Cross Entropy

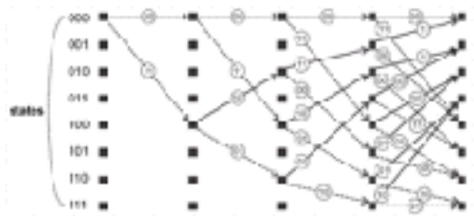


Pitch Class Probabilities

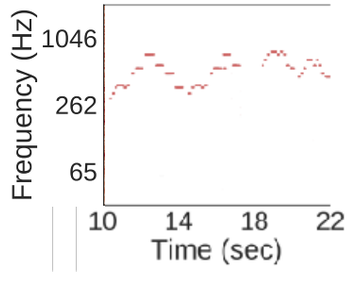


Vocal activity detection Neural Network + viterbi decoding





f_0 Time Series

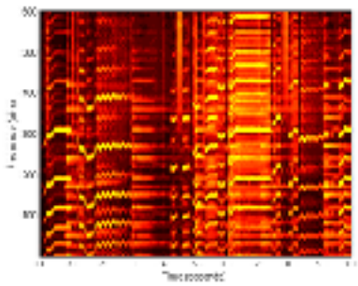


time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

Example 4: Neural Network 2

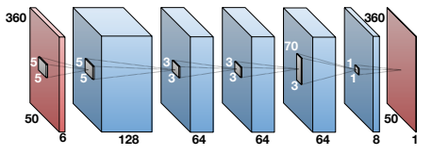
Sebastian Böck, Markus Schedl
“Polyphonic piano note transcription with recurrent neural networks”
ICASSP (2012)

Polyphonic Music



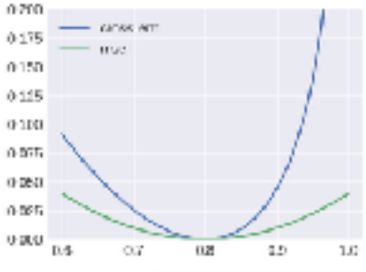
STFT

Bidirectional LSTM

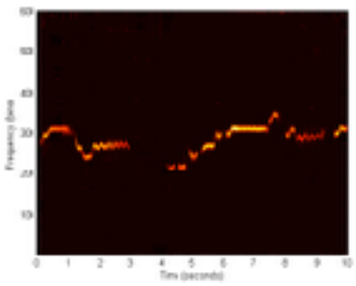


+

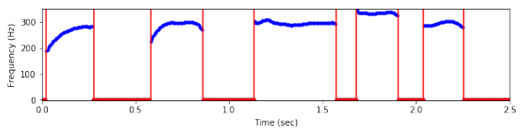
Cross Entropy

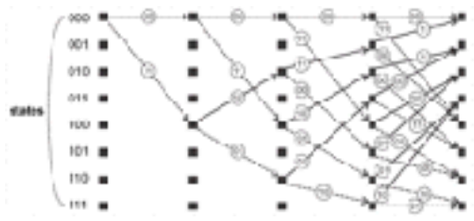


Onset Semitone Likelihoods

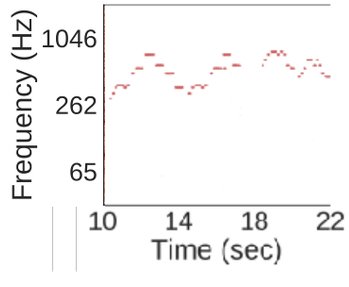


Thresholding





f_0 Onset Series



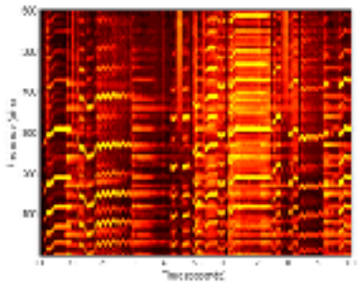
time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

Example 5: Neural Network 3

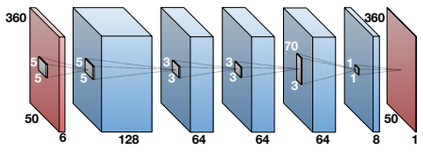
Rachel Bittner, Brian McFee, Justin Salamon, Peter Li, Juan Pablo Bello
"Deep Salience Representations for F0 Estimation in Polyphonic Music"
ISMIR (2017)

Polyphonic Music

HCQT

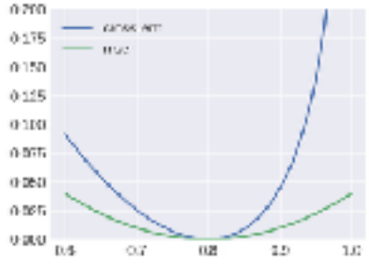


CNN

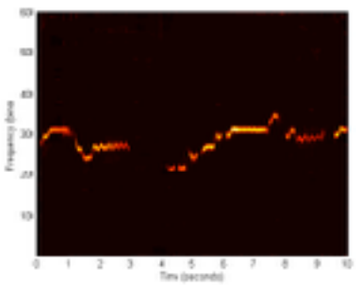


+

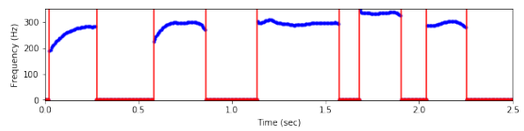
Binary Cross Entropy

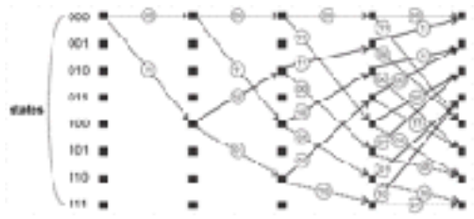


Learned Salience

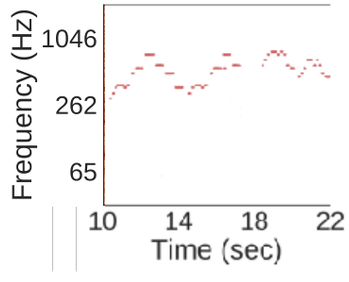


Threshold + Argmax





f_0 Time Series



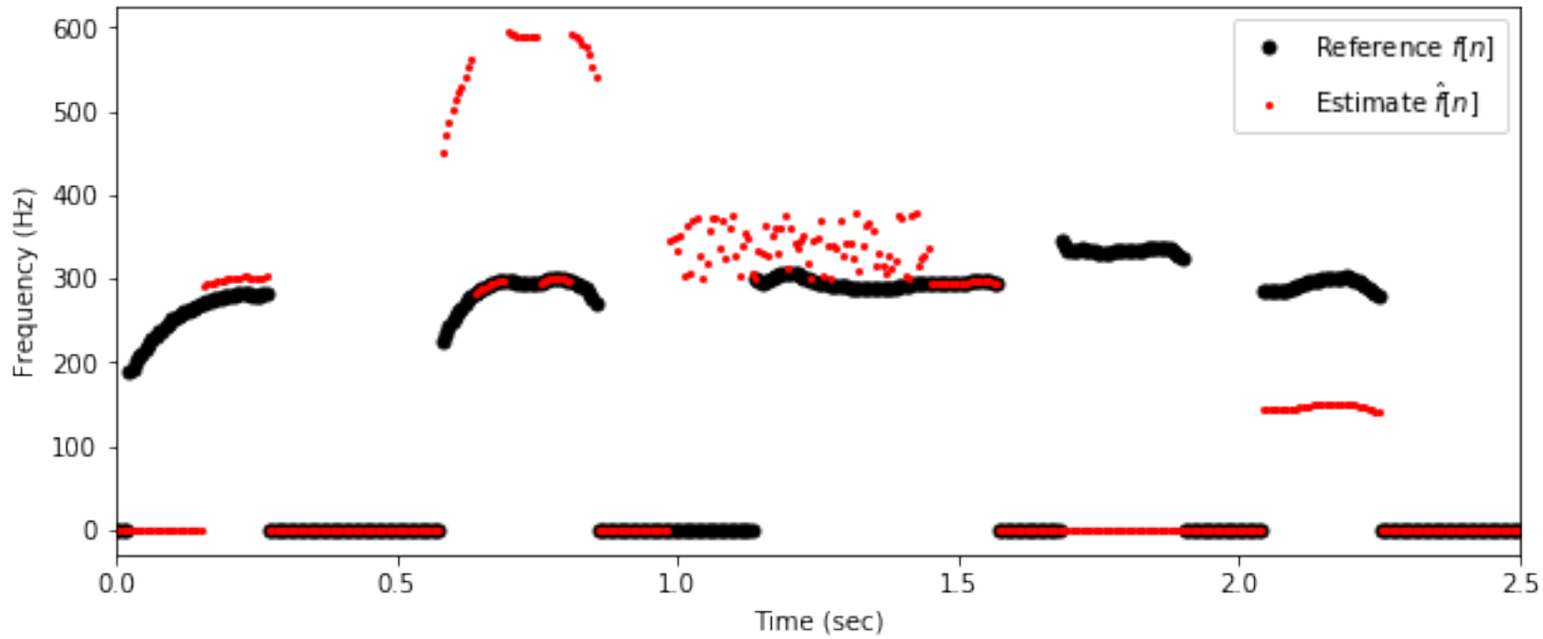
time, freq
10.0, 275.3
10.1, 258.2
10.2, 257.9
10.3, 0
10.4, 0
10.5, 283.3

What about source separation?

Evaluation

Single-f0 Qualitative Evaluation

1. Plots



2. Sonification

$$y(t) = a(t) \times \sin(2\pi \int^t f(x)\, dx)$$

> a(t) - amplitude over time
  voicing: binary
  confidence: continuous
> f(t) - frequency estimate

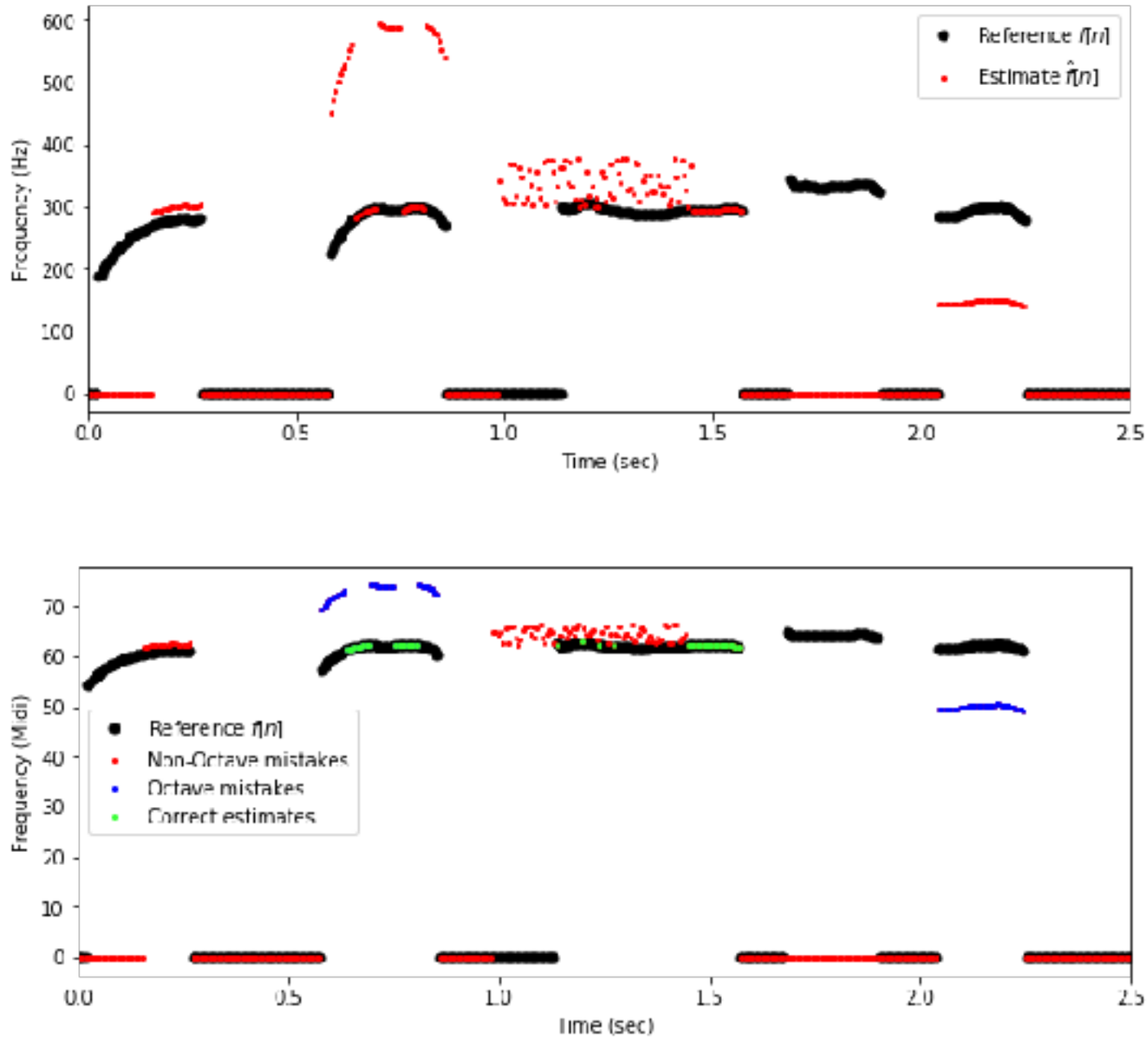
Mix Audio

Reference f0

Estimate f0

Single-f0 Quantitative Evaluation

- Gross Percentage Error
- Mean Squared Error

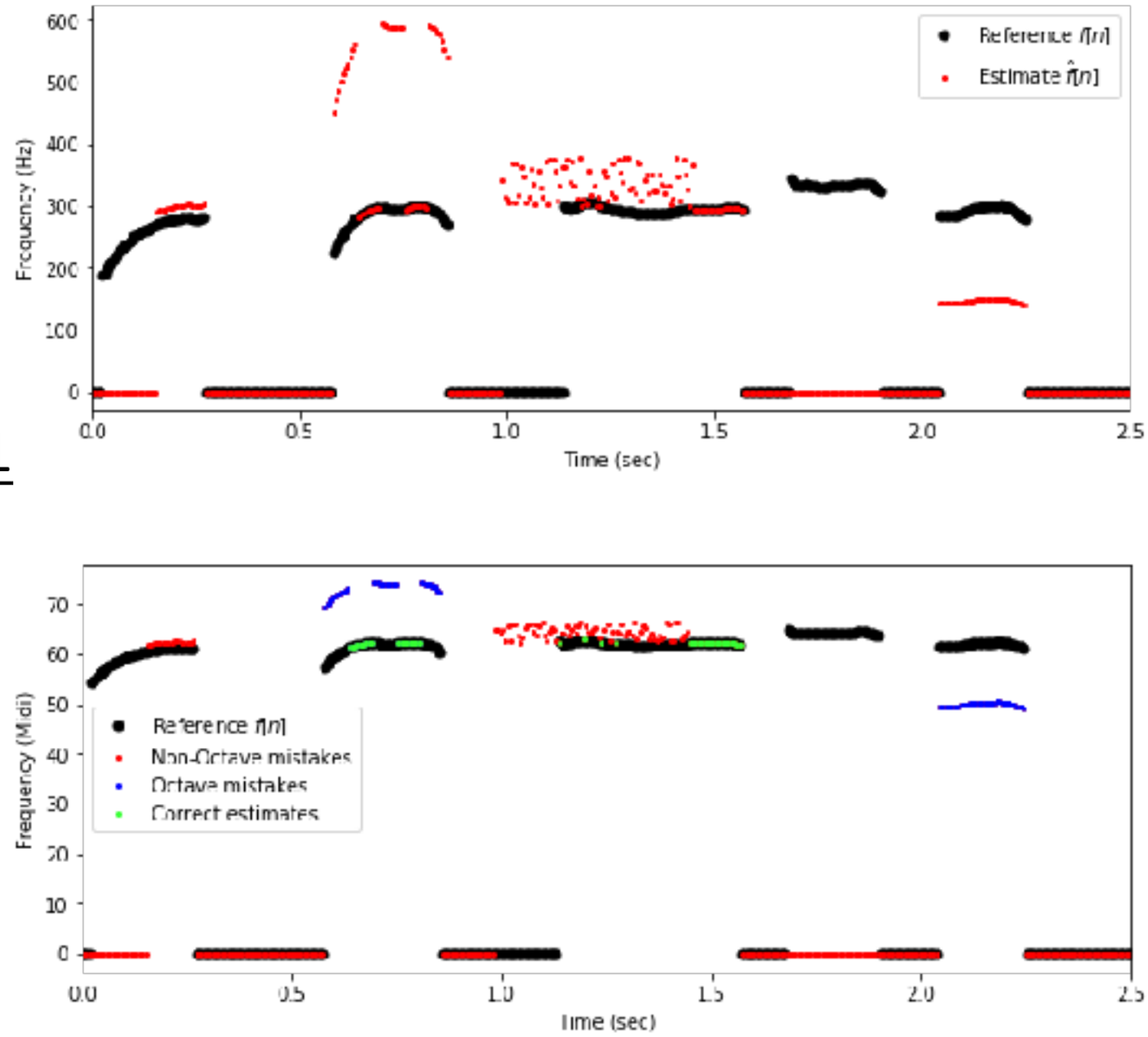


Single-f0 Quantitative Evaluation

```
('Voicing Recall', 0.746)
('Voicing False Alarm', 0.136)
('Raw Pitch Accuracy', 0.1875)
('Raw Chroma Accuracy', 0.463)
('Overall Accuracy', 0.487)
```

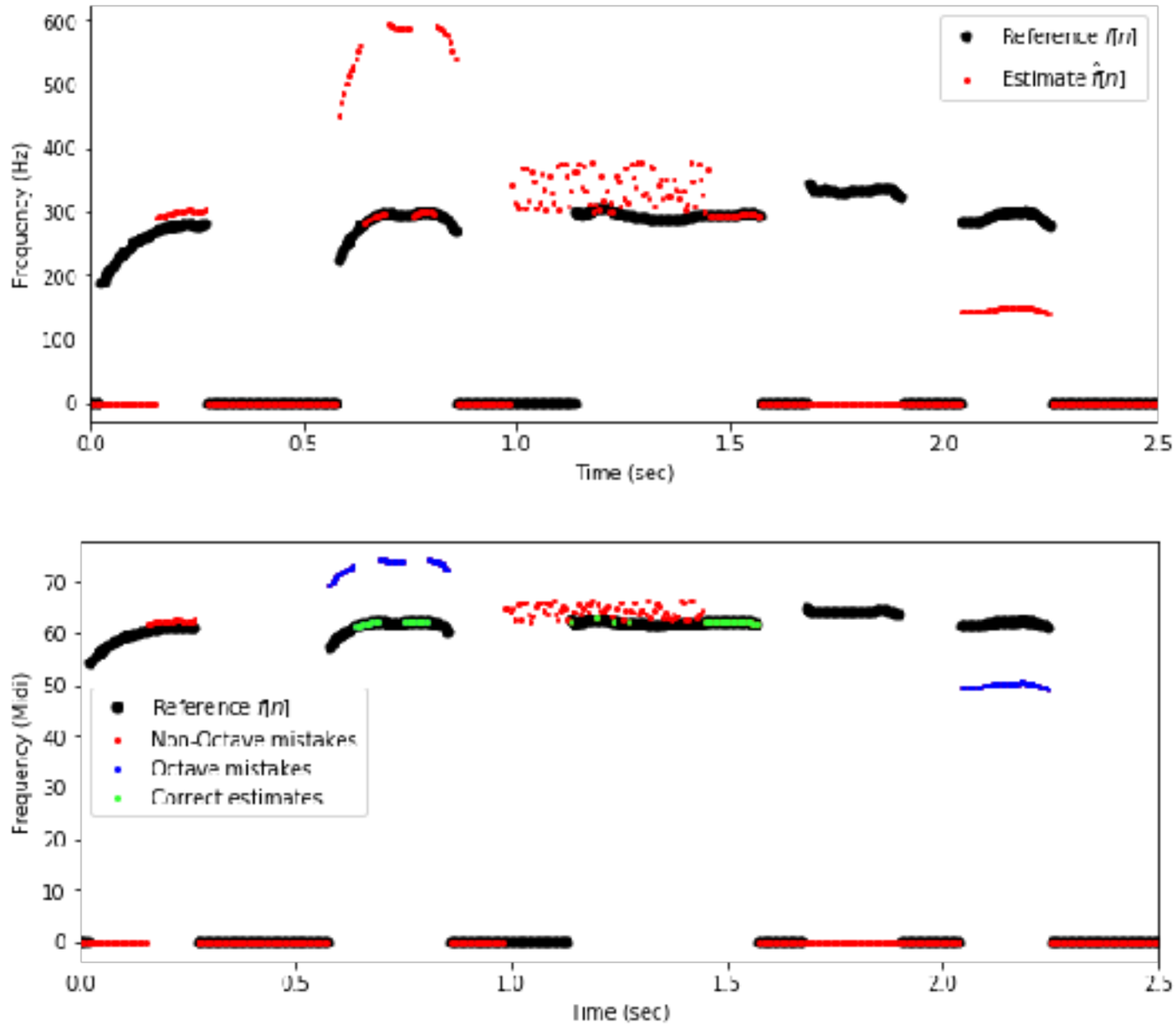
github.com/craffel/mir_eval

mir_eval.melody.evaluate

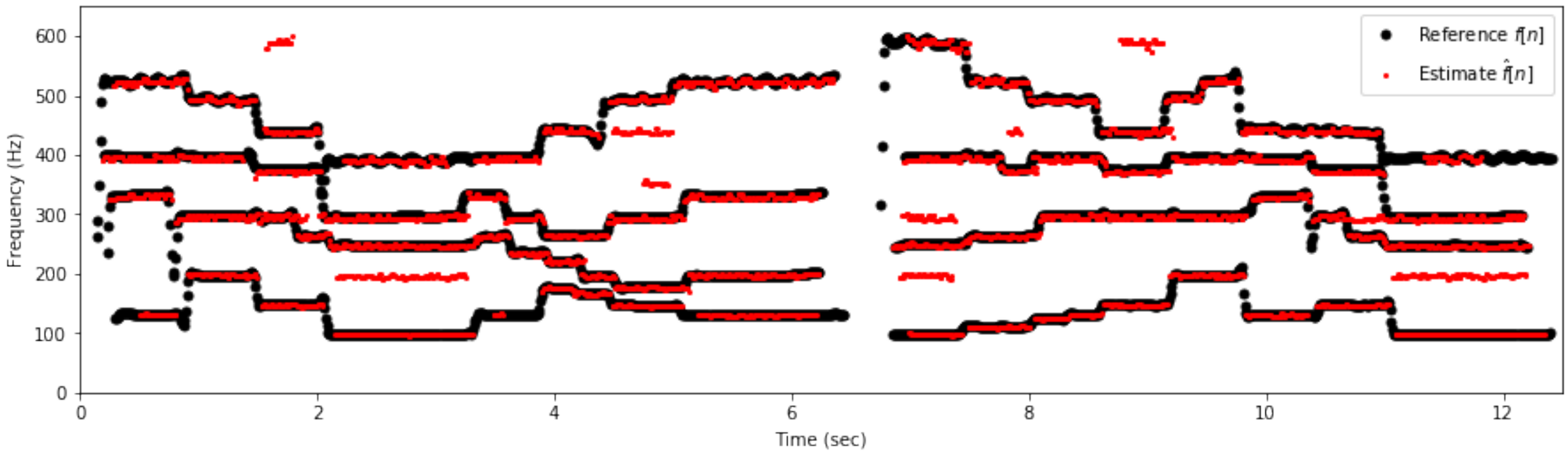


Single-f0 Quantitative Evaluation

- Octave Jump Ratio (OJ)
- Chroma Continuity (CC)



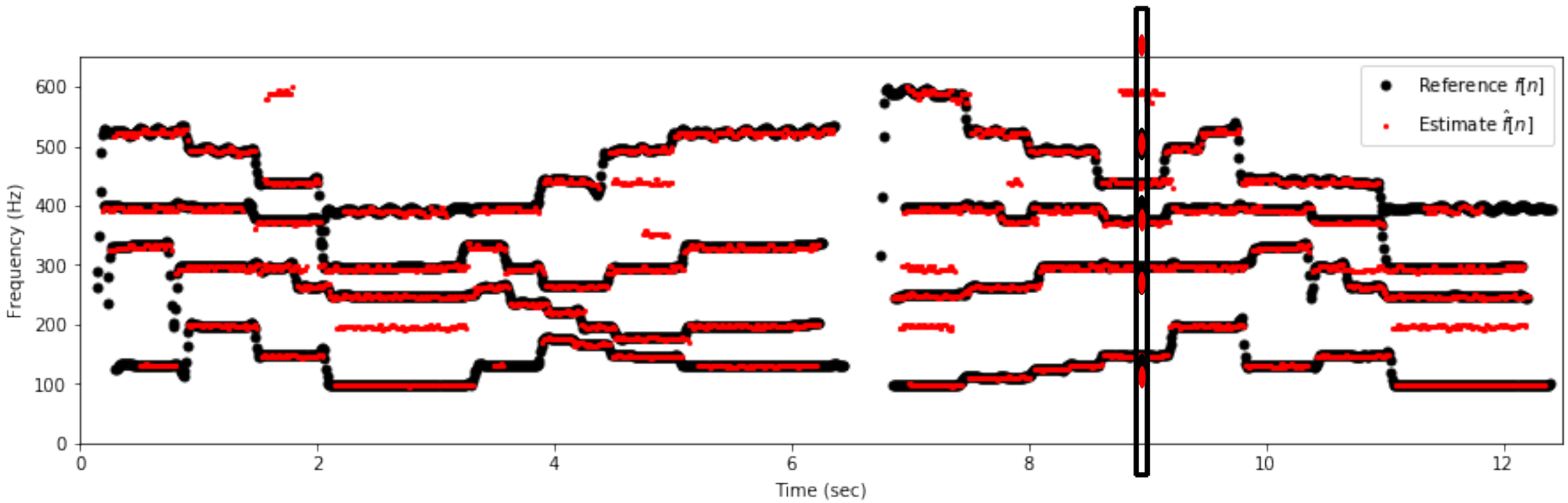
Multiple-f0 Qualitative Evaluation



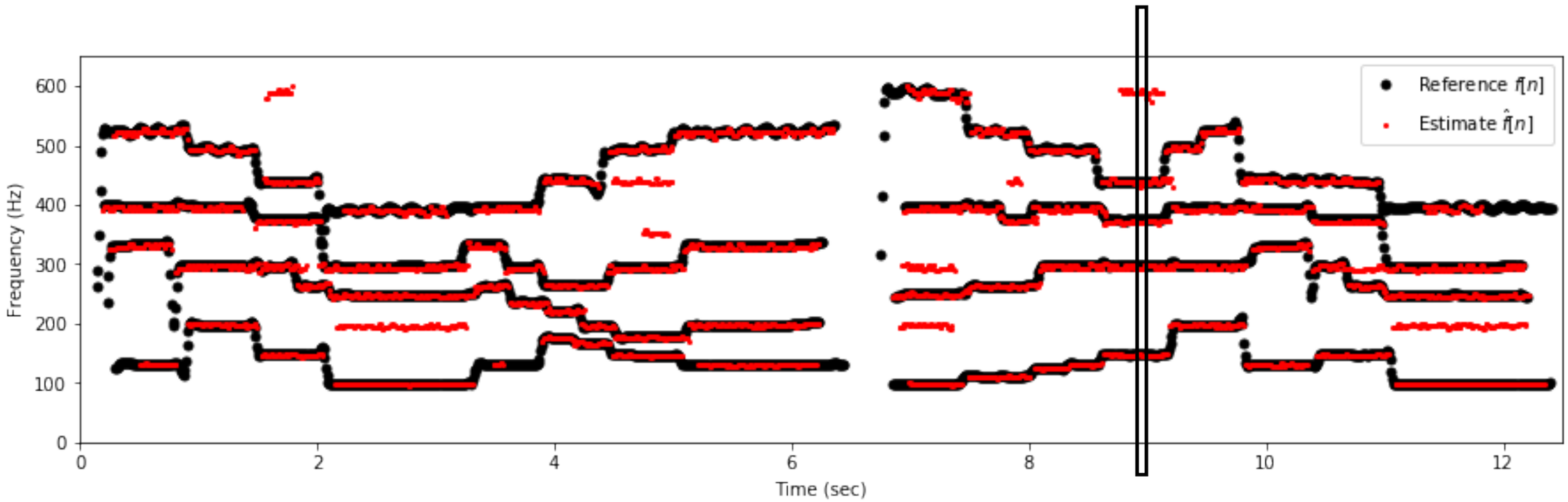
Reference f0

Estimate f0

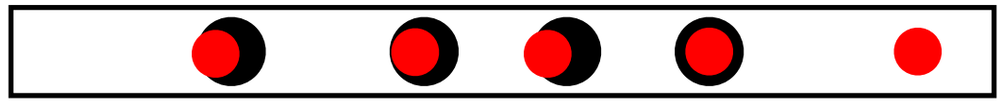
Multiple-f0 Quantitative Evaluation



Multiple-f0 Quantitative Evaluation




```
('Precision', 0.881),
('Recall', 0.886),
('Accuracy', 0.791),
('Substitution Error', 0.047),
('Miss Error', 0.067),
('False AlarmaError', 0.072),
('Total Error', 0.187),
('Chroma Precision', 0.900),
('Chroma Recall', 0.905),
('Chroma Accuracy', 0.823),
('Chroma Substitution Error', 0.028),
('Chroma Miss Error', 0.067),
('Chroma False Alarm Error', 0.072),
('Chroma Total Error', 0.167)
```

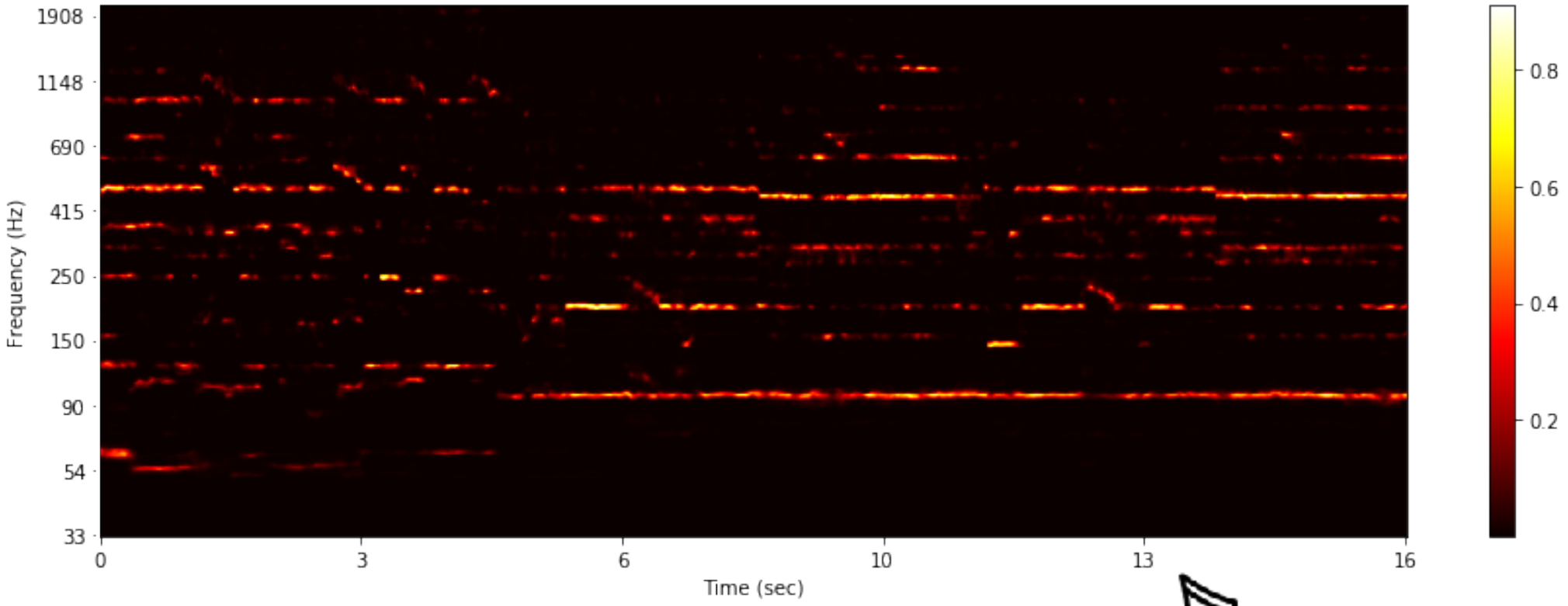


Frequency (Hz)

github.com/craffel/mir_eval

mir_eval.multipitch.evaluate

Salience Qualitative Evaluation



$y(t) = a(t) \times \sin(2\pi \int^t f(x)\, dx)$

> a(t) - “amplitude” over time
  voicing —> binary
  “confidence” —> continuous
> f(t) - frequency estimate

Mix Audio

Salience

Multi-f_0

Datasets

f_0 Datasets

	Year	Dur.	Multi-f0	Melody	Vocal	Bass	Piano	Guitar
RWC	2002	7 h		notes				
MAPS	2008	18 h	notes				notes	
MIR-1K	2009	3 h		notes				
Bach10	2012	5 m	notes					
MedleyDB	2014	7 h		f0	f0	f0 (1h)		
iKala	2015	2 h		f0	f0			
Su-AMT	2015/6	12 m	notes					
Weimar Jazz	2016	13 h		notes		notes (1h)		
OrchSet	2016	23 m		f0				
Lakh MIDI	2016	15 d	notes					
MusicNet	2018	34 h	notes					
GuitarSet	2018	3 h	notes					notes

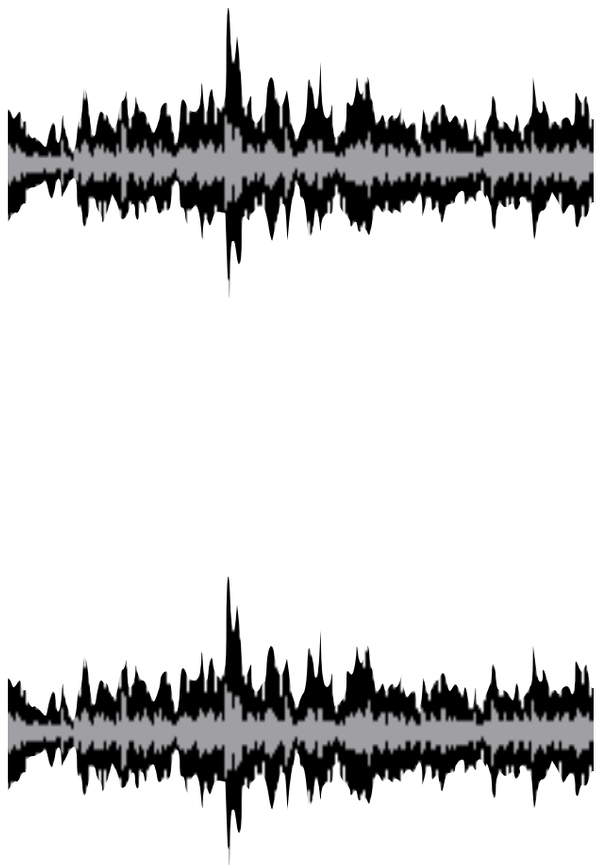
f_0 Datasets

	Link
RWC	https://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-p.html
MAPS	http://www.tsi.telecom-paristech.fr/aao/en/2010/07/08/maps-database-a-piano-database-for-multipitch-estimation-and-automatic-transcription-of-music/
MIR-1K	https://sites.google.com/site/unvoicedsoundseparation/mir-1k
Bach10	http://music.cs.northwestern.edu/data/Bach10.html
MedleyDB	http://medleydb.weebly.com/
iKala	http://mac.citi.sinica.edu.tw/ikala/
Su-AMT	https://sites.google.com/site/lisupage/research/new-methodology-of-building-polyphonic-datasets-for-amt
Weimar Jazz	https://jazzomat.hfm-weimar.de/dbformat/dboverview.html
OrchSet	https://www.upf.edu/web/mtg/orchset
Lakh MIDI	http://colinraffel.com/projects/lmd/
MusicNet	https://homes.cs.washington.edu/~thickstn/musicnet.html
GuitarSet	https://guitarset.weebly.com/

Data Generation

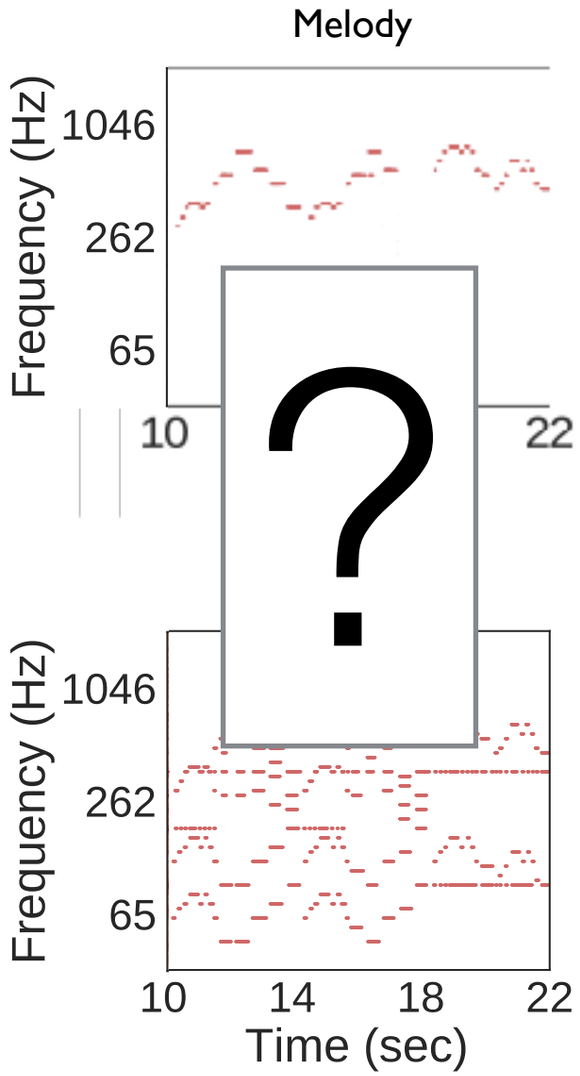
How do we obtain f_0 labels?

Polyphonic Audio

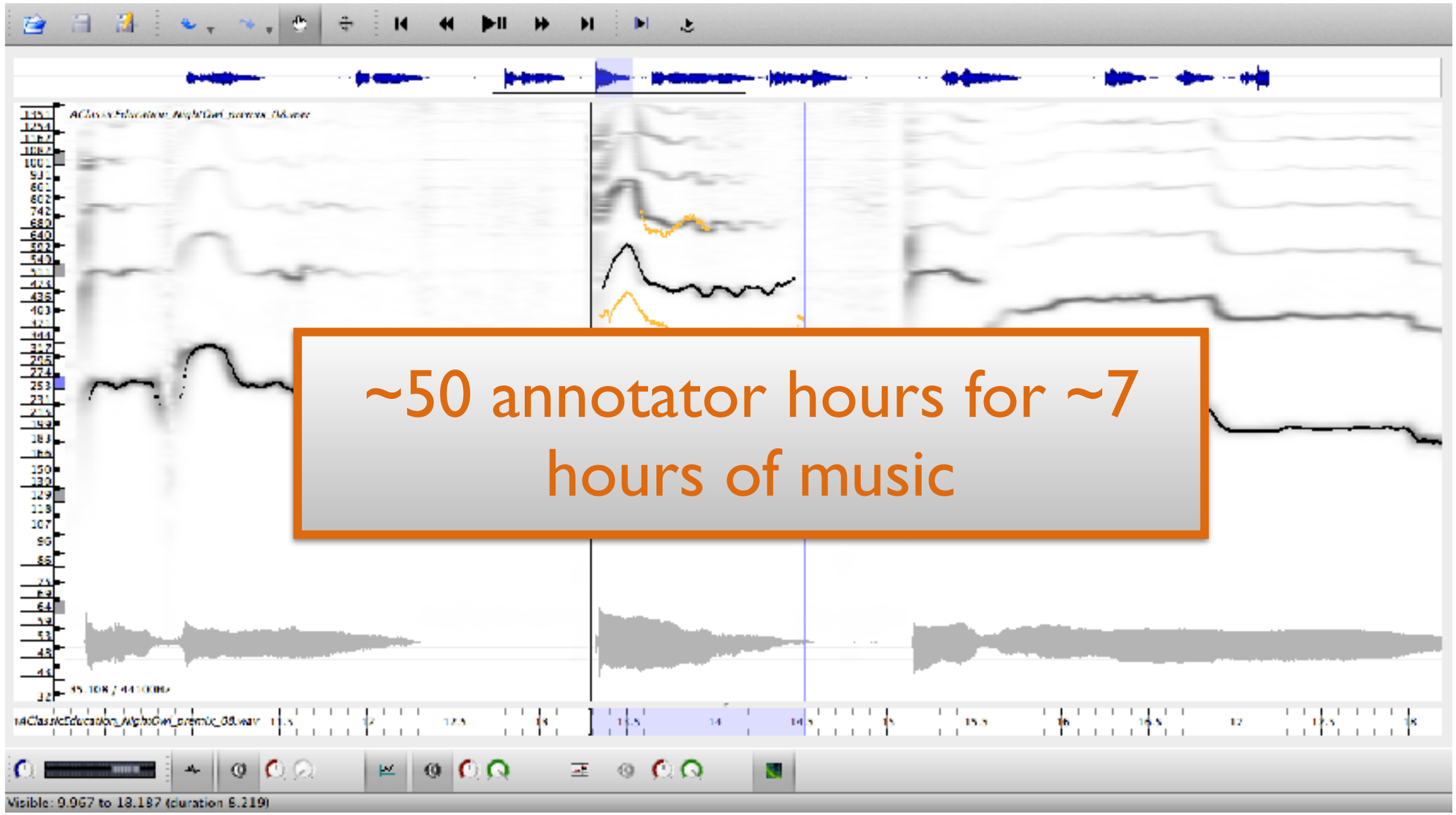


+

f_0 Annotations

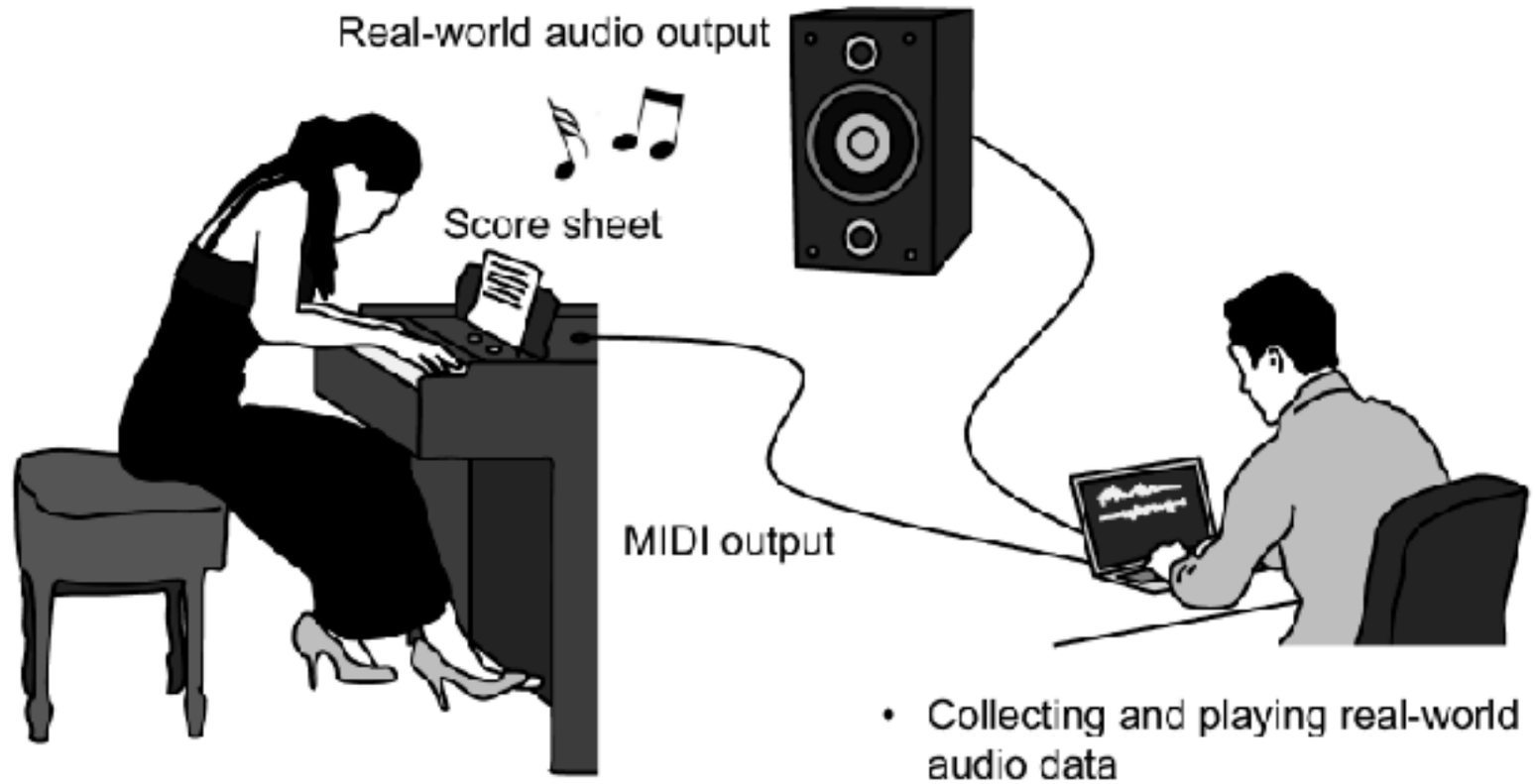


Manual f_0 correction



https://code.soundsoftware.ac.uk/projects/tony

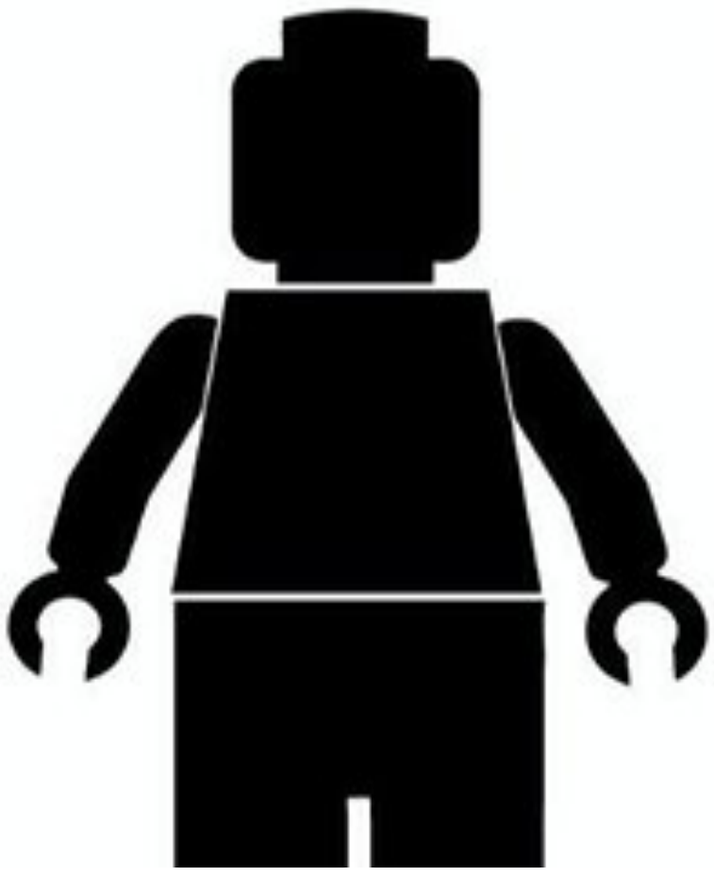
Assisted Note Annotations



- Note following = note-level annotation
- Collecting and playing real-world audio data
- Checking note-level annotation

Li Su, Yi-Hsuan Yang
“Escaping from the Abyss of Manual Annotation: New Methodology of Building Polyphonic Datasets for Automatic Music Transcription”
ICMC (2015)

Synthesized MIDI

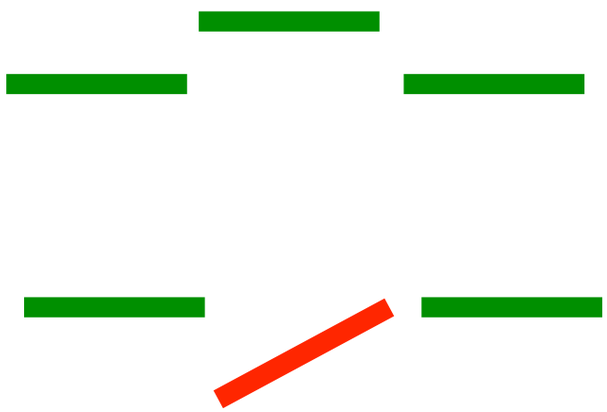


Annotation by Resynthesis

Played

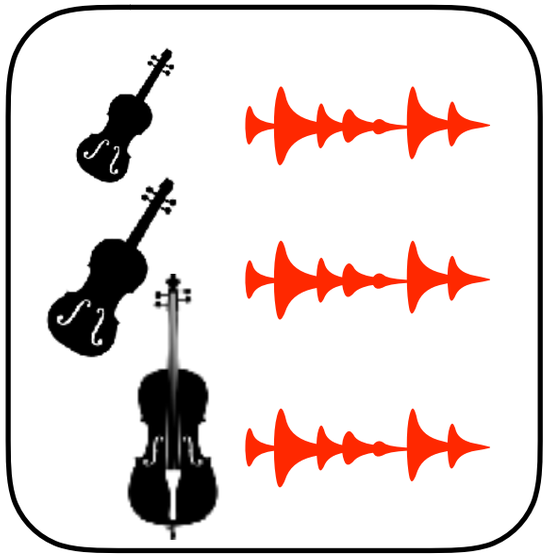
Estimated f0

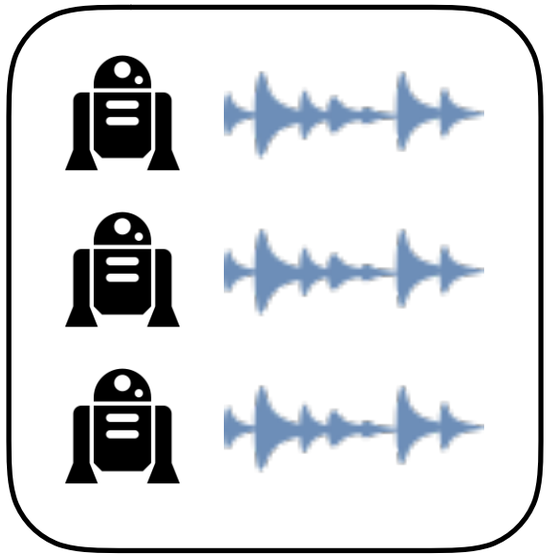
Synthesized f0



Justin Salamon, Rachel Bittner, Jordi Bonada, Juan José Bosch, Emilia Gómez, Juan Pablo Bello
"An Analysis/Synthesis Framework for Automatic f0 Annotation of Multitrack Datasets"
ISMIR (2016)

Annotation by Resynthesis

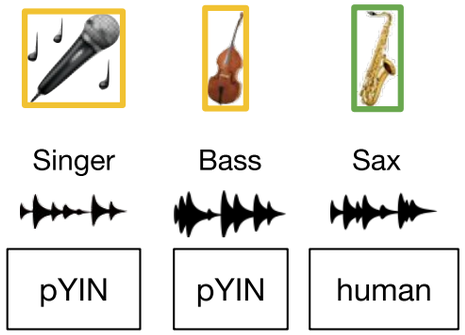




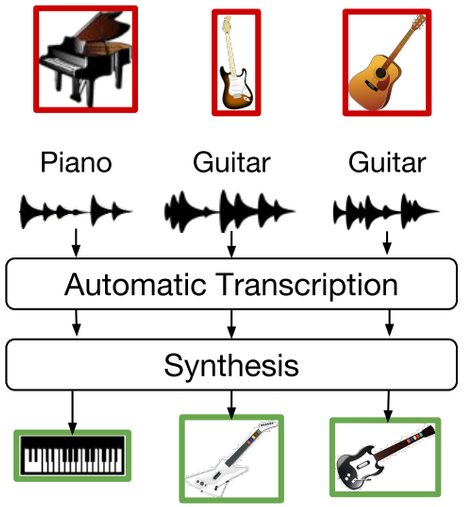
Justin Salamon, Rachel Bittner, Jordi Bonada, Juan José Bosch, Emilia Gómez, Juan Pablo Bello
“An Analysis/Synthesis Framework for Automatic f0 Annotation of Multitrack Datasets”
ISMIR (2016)

Automatic Multitrack Annotations

Available Instrument Annotations



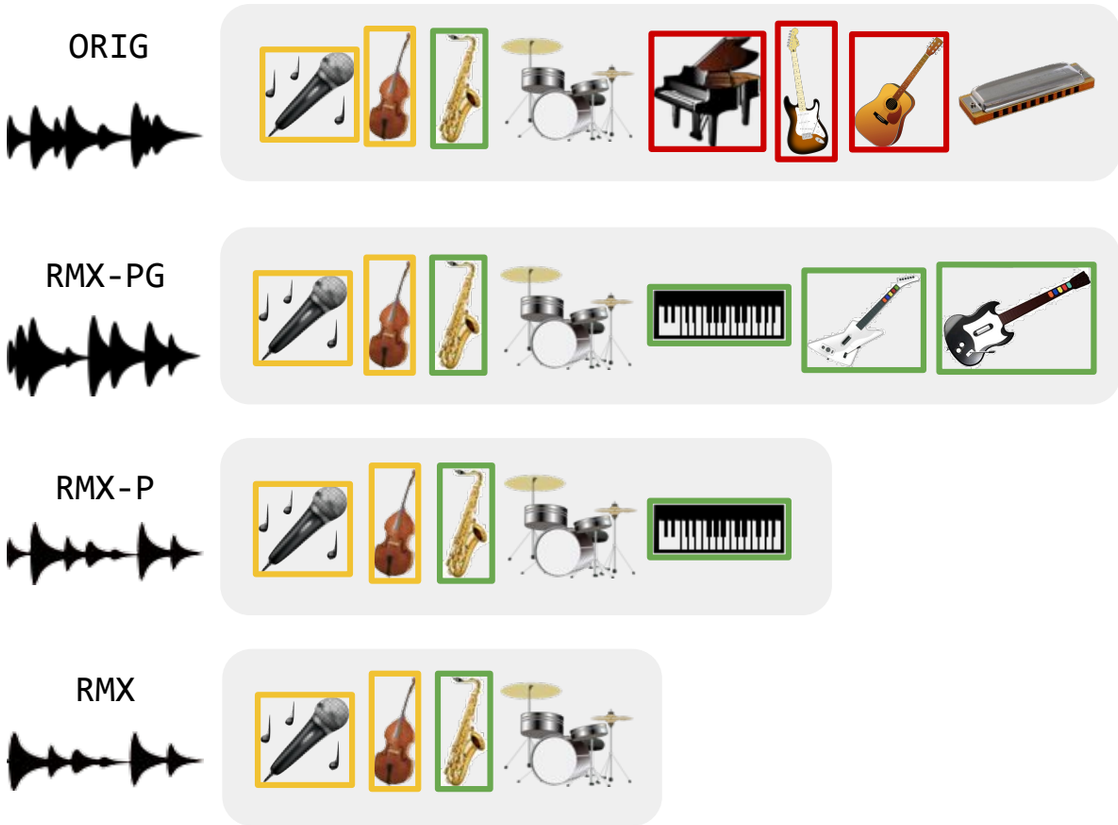
Instrument Annotations via Synthesis



Individual Instrument Annotation Quality

Poor Approximate Exact

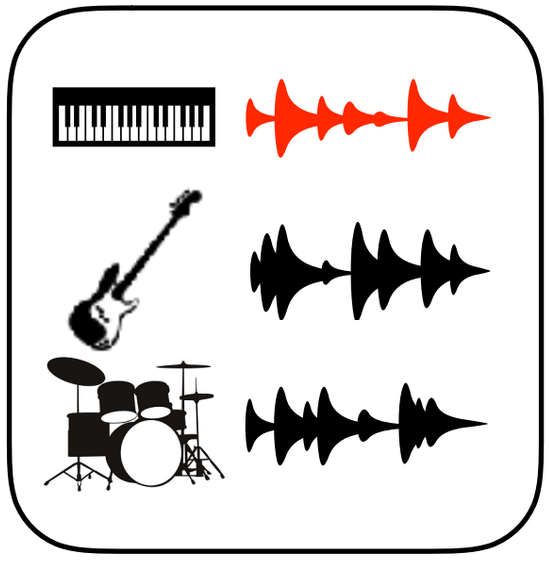
Mixture Types and Available f_0 Annotations

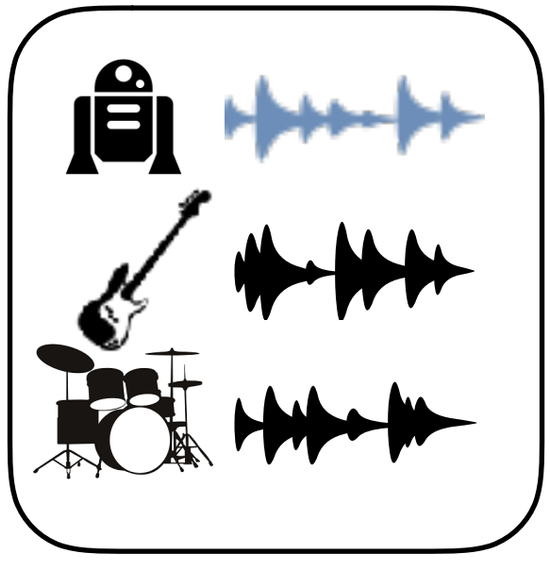


Mixture Annotation Types (this example)

	ORIG	RMX-PG	RMX-P	RMX
Multiple-f_0	✘	✓	✓	✓
Melody	✓	✓	✓	✓
Bass	✓	✓	✓	✓
Vocal	✓	✓	✓	✓
Piano	✘	✓	✓	✓=0
Guitar	✘	✓	✓=0	✓=0

Piano/Guitar Replacement Examples





Summary

- Most polyphonic f0 estimation models produce a pitch salience representation and decode to f0 or notes
  - Tasks are approached separately most of the time
  - Transcription and f0 estimation are almost the same
- Evaluation metrics are primarily frame based
- Datasets for each individual task are limited
- Multitracks can be used to generate data more easily

Open Problems

- Perceptually motivated evaluation metrics
  - go beyond frame level
  - incorporate salience
- Optimization function that is an upper bound for f0 evaluation metrics
- Directly predict outputs: f0 time series or notes
- Joint source separation + f0 estimation
- Include onset information for f0 trajectories
- Train a simple system on a ton of data and see what happens