

# Pitch Analysis for Active Music Discovery

Justin Salamon<sup>1,2</sup>

 @justin\_salamon  justin.salamon@nyu.edu  www.justinsalamon.com

<sup>1</sup>Music and Audio Research Laboratory  
<sup>2</sup>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



## In A Nutshell

## In A Nutshell

- ▶ GOAL: active music discovery

## In A Nutshell

- ▶ GOAL: active music discovery
- ▶ MEANS: pitch content analysis

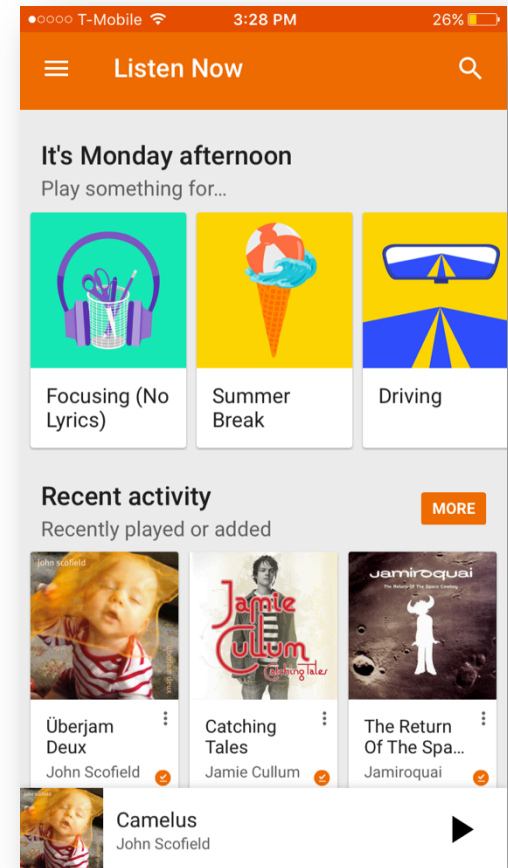
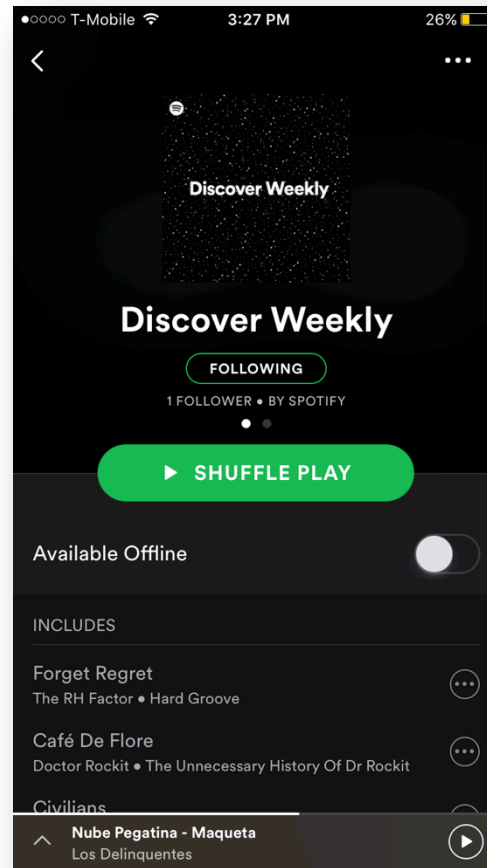
## In A Nutshell

- ▶ GOAL: active music discovery
- ▶ MEANS: pitch content analysis
- ▶ CHALLENGE: data scarcity (and solution!)

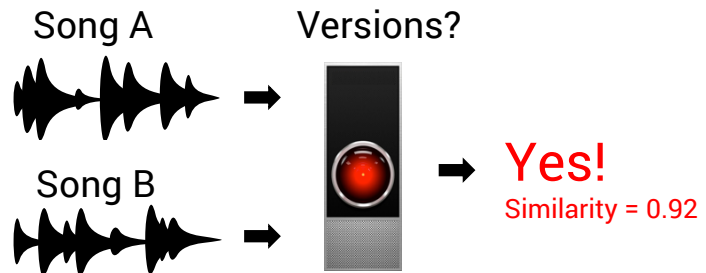
# Active Music Discovery

# Music Recommendation

# Music Recommendation

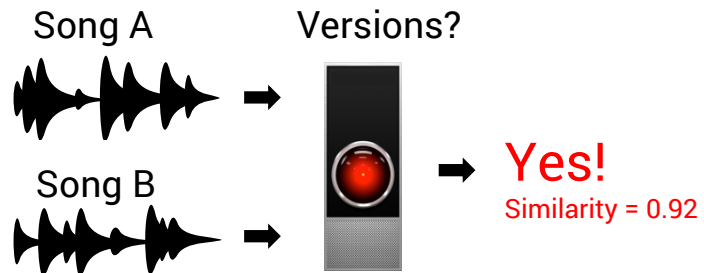




**COVER SONG ID**

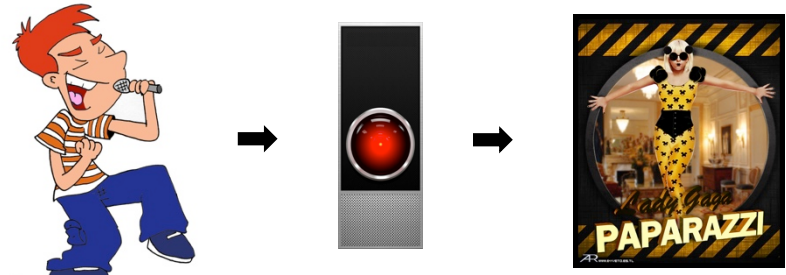
(Salamon, Serrà & Gómez, 2013)

## COVER SONG ID



(Salamon, Serrà & Gómez, 2013)

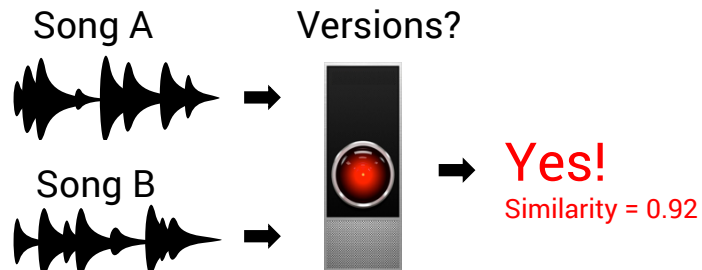
## QUERY-BY-HUMMING



(Salamon, Serrà & Gómez, 2013)

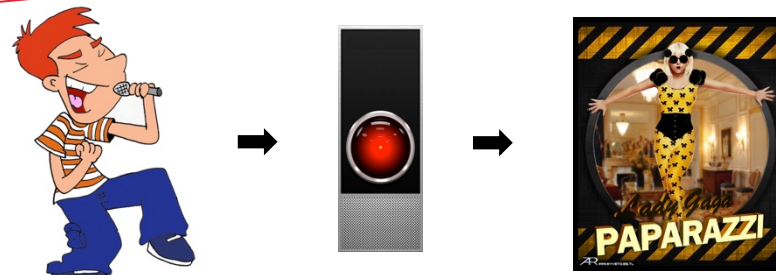
# Active Search & Discovery

## COVER SONG ID



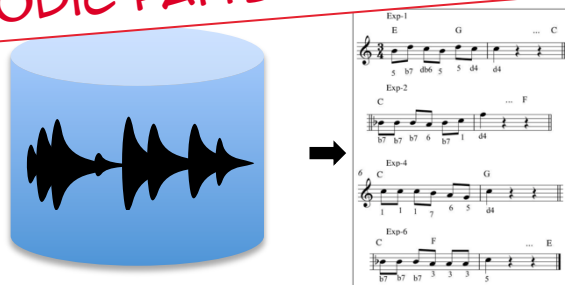
(Salamon, Serrà & Gómez, 2013)

## QUERY-BY-HUMMING



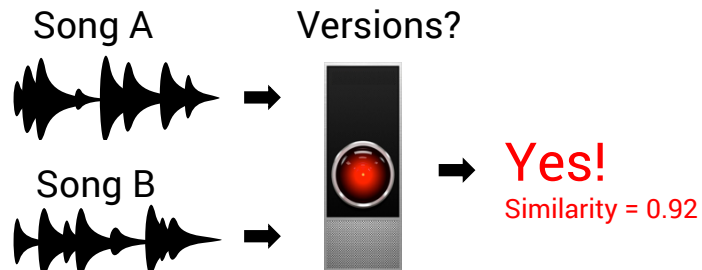
(Salamon, Serrà & Gómez, 2013)

## MELODIC PATTERN DISCOVERY



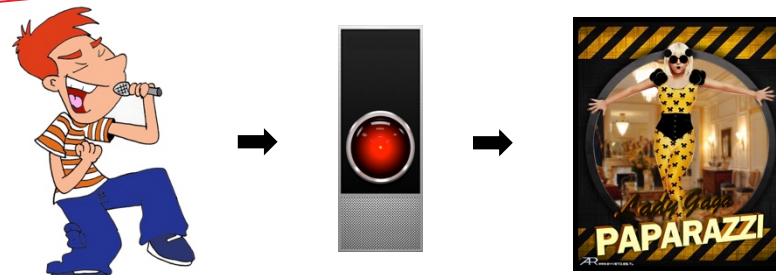
(Pikrakis et al., 2012)

## COVER SONG ID



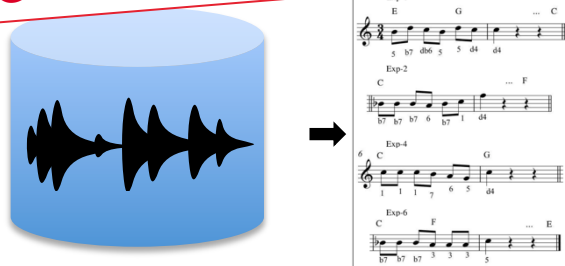
(Salamon, Serrà &amp; Gómez, 2013)

## QUERY-BY-HUMMING



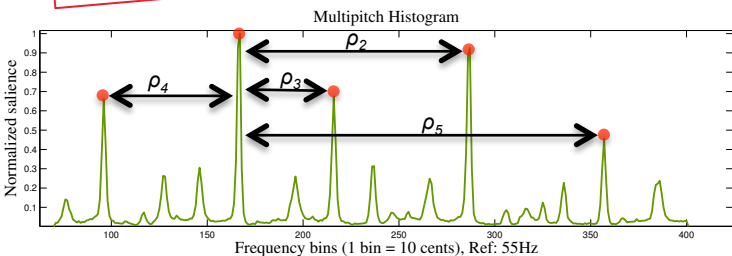
(Salamon, Serrà &amp; Gómez, 2013)

## MELODIC PATTERN DISCOVERY



(Pikrakis et al., 2012)

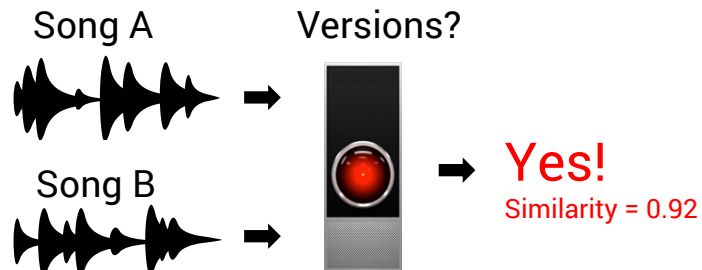
## TONIC ID



(Salamon, Gulati &amp; Serra, 2012)

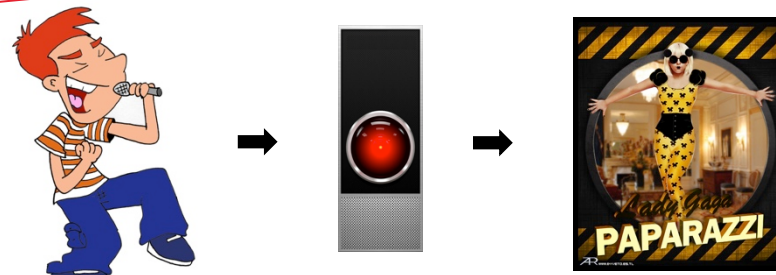
# Active Search & Discovery

## COVER SONG ID



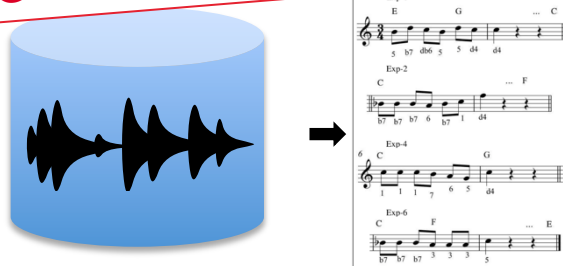
(Salamon, Serrà & Gómez, 2013)

## QUERY-BY-HUMMING



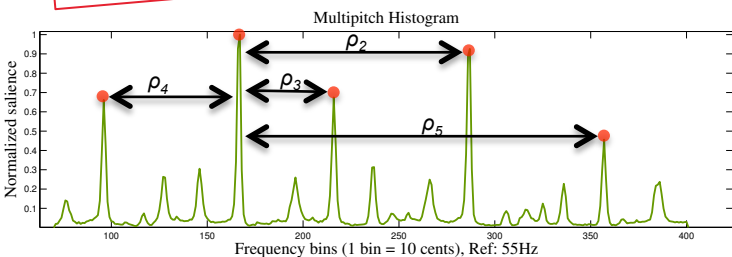
(Salamon, Serrà & Gómez, 2013)

## MELODIC PATTERN DISCOVERY



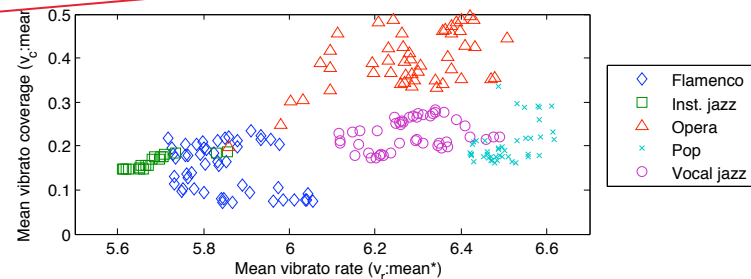
(Pikrakis et al., 2012)

## TONIC ID



(Salamon, Gulati & Serra, 2012)

## SINGING STYLE CLASSIFICATION



(Salamon, Rocha & Gómez, 2012)

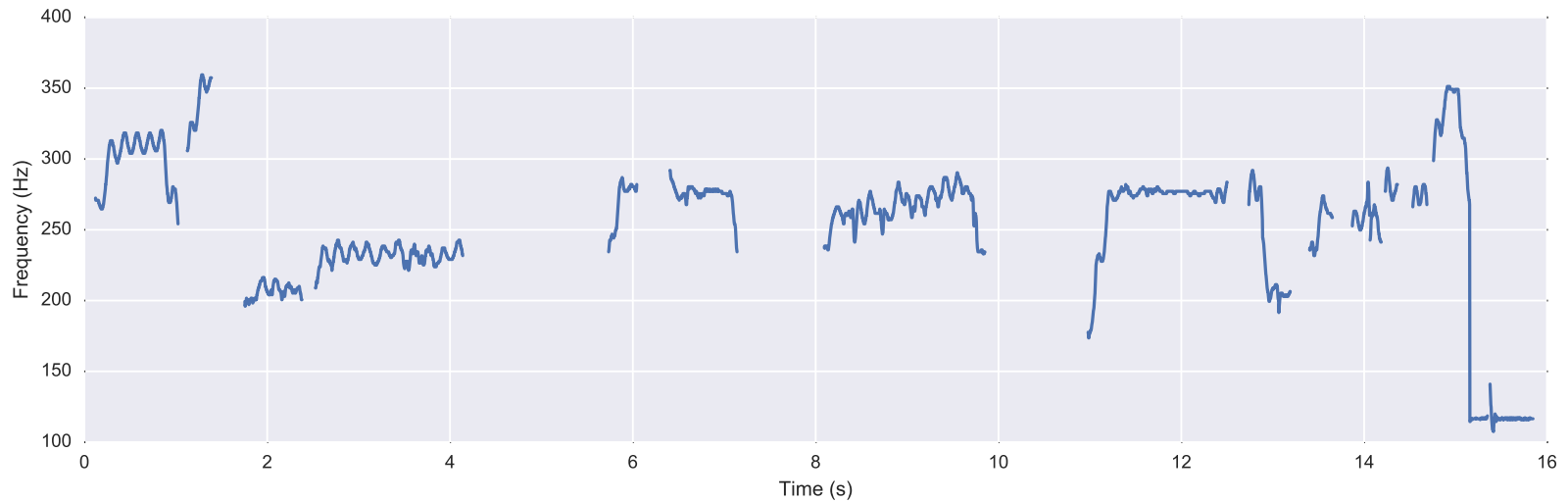
# Machine Learning for Pitch Analysis

# Melody Extraction

# Melody Extraction



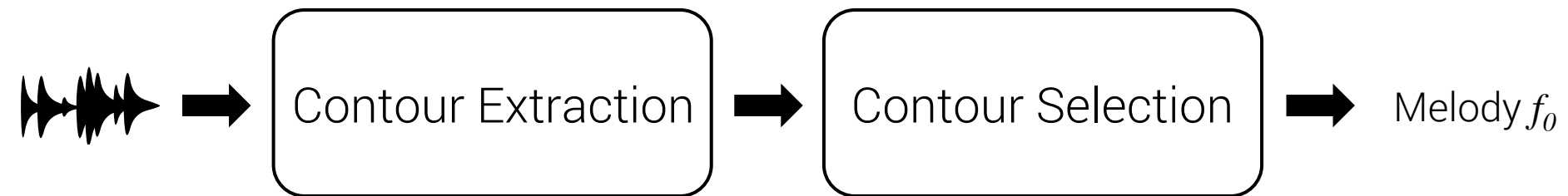
# Melody Extraction



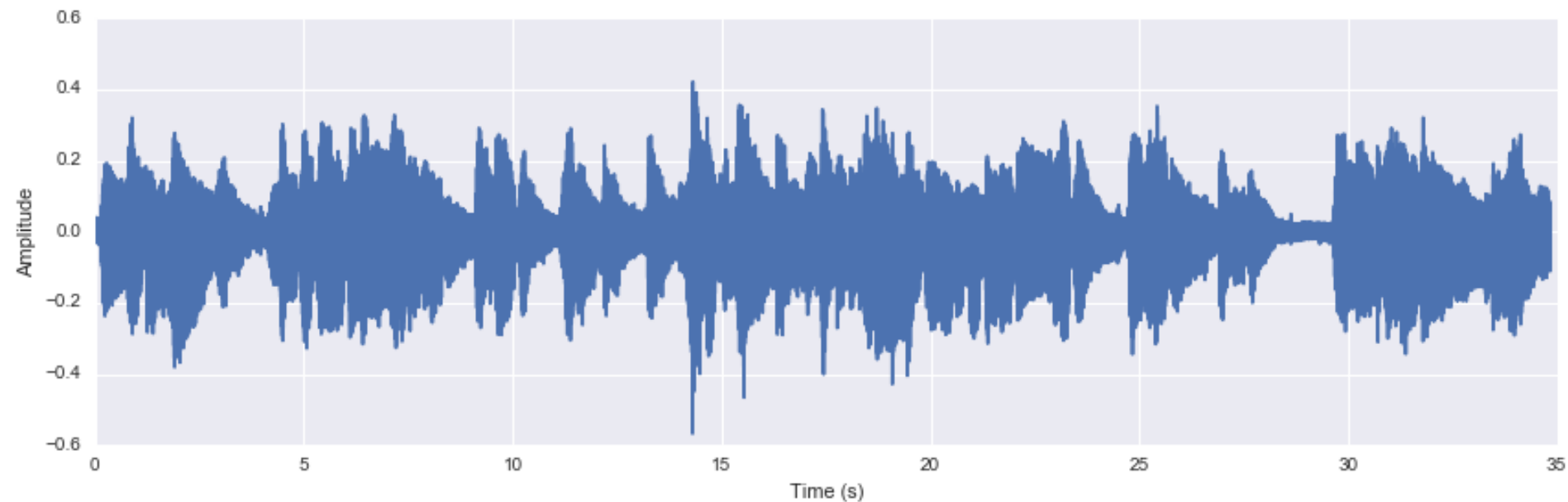
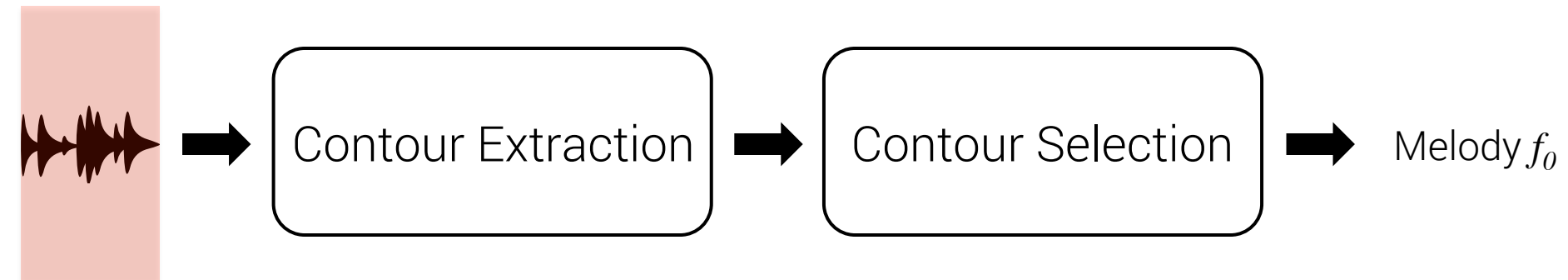
## Melodia (Salamon & Gómez, 2012)



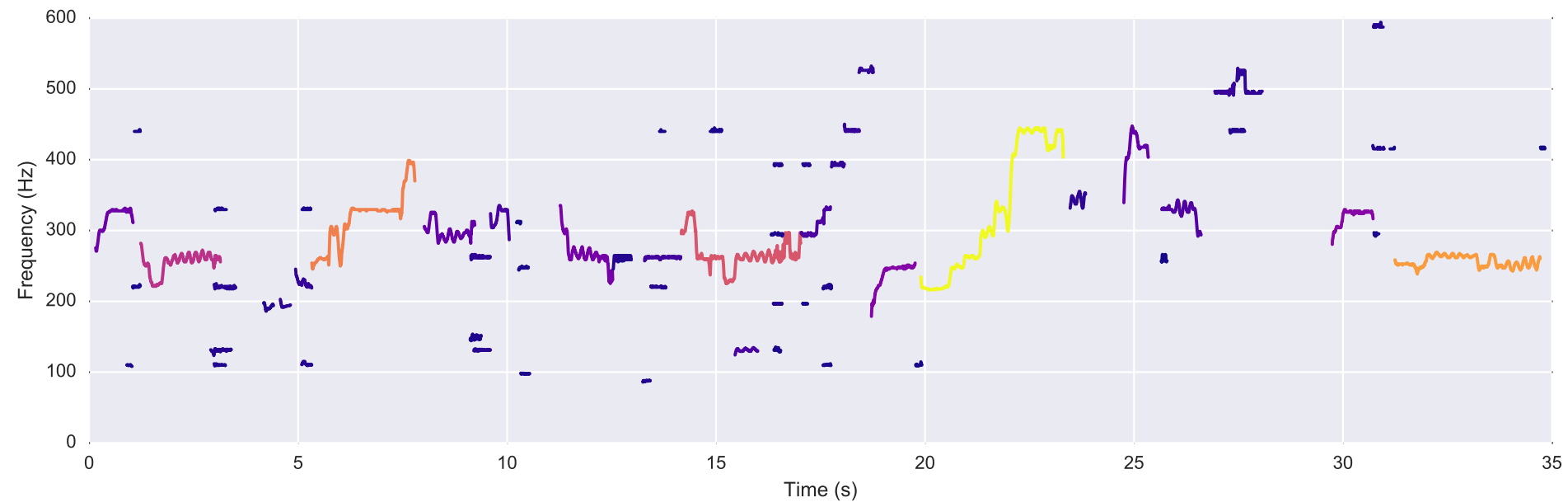
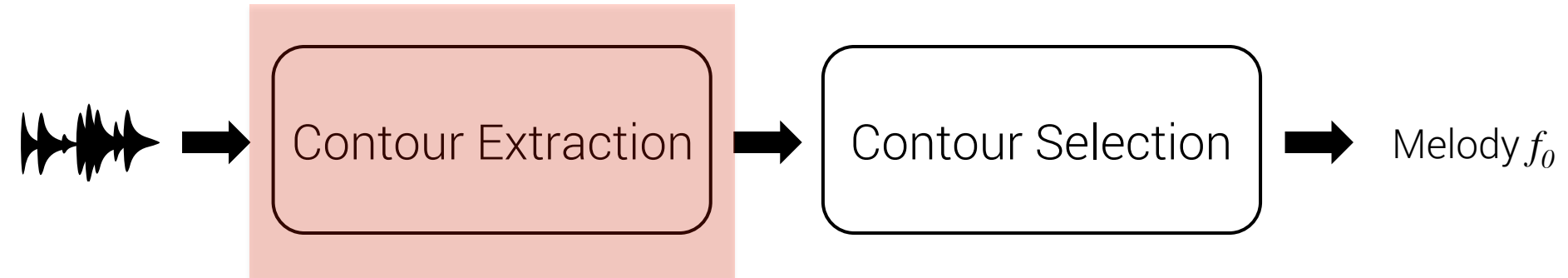
## Melodia (Salamon & Gómez, 2012)



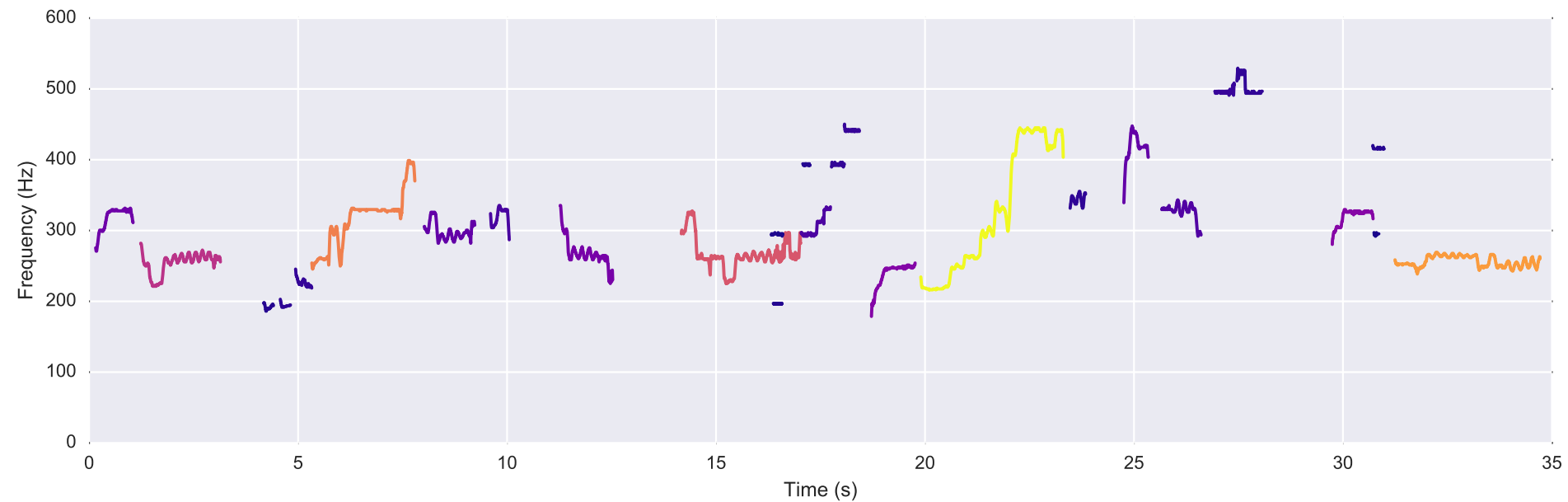
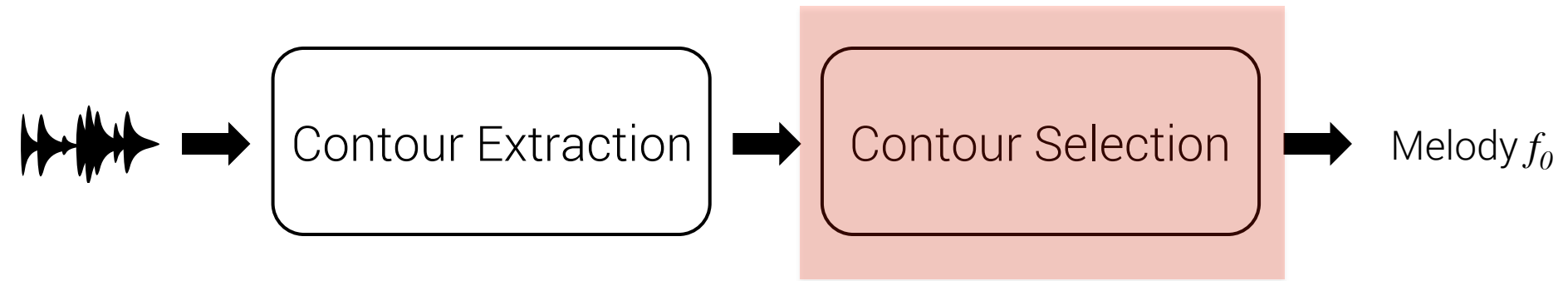
## Melodia (Salamon & Gómez, 2012)



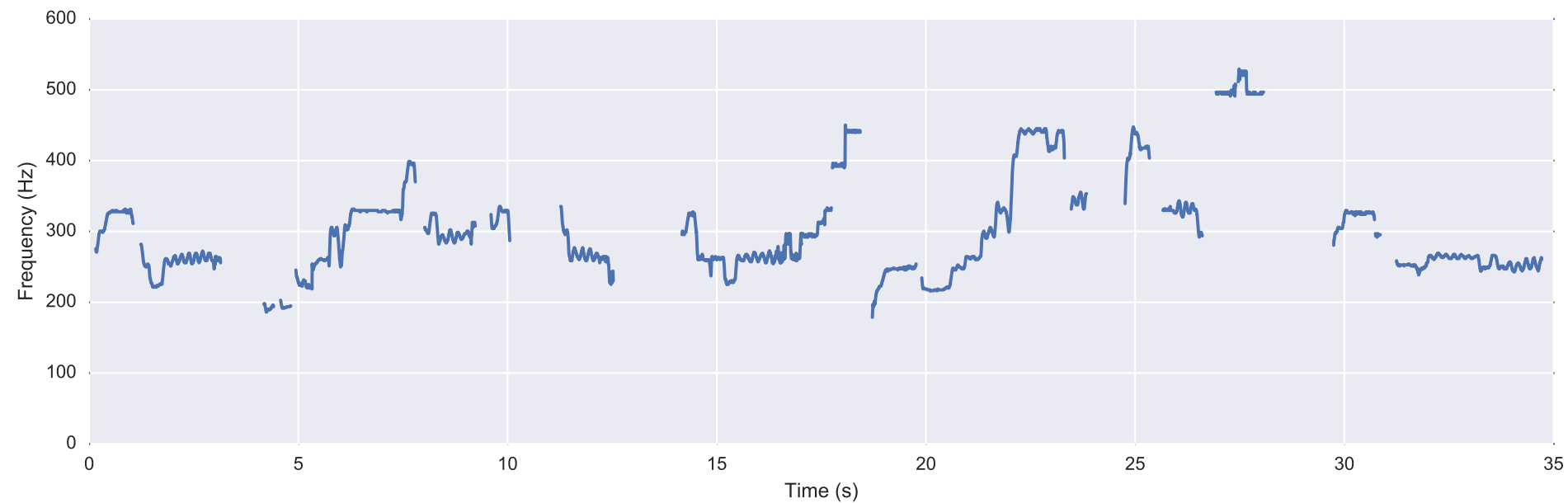
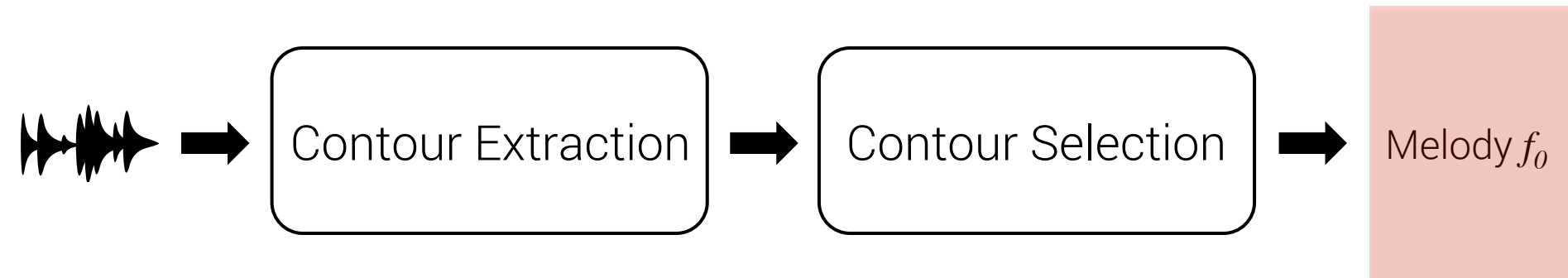
## Melodia (Salamon & Gómez, 2012)



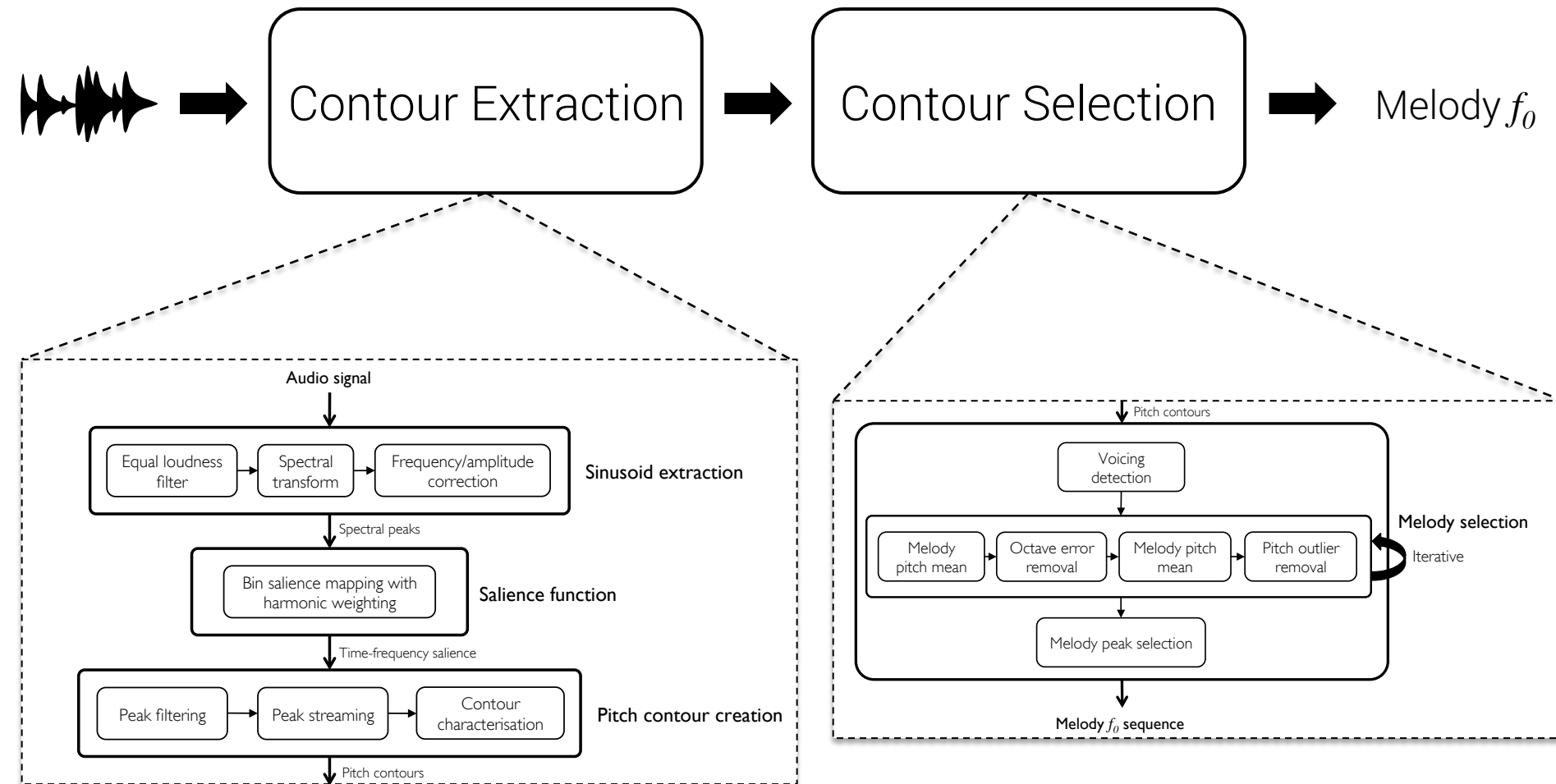
## Melodia (Salamon & Gómez, 2012)



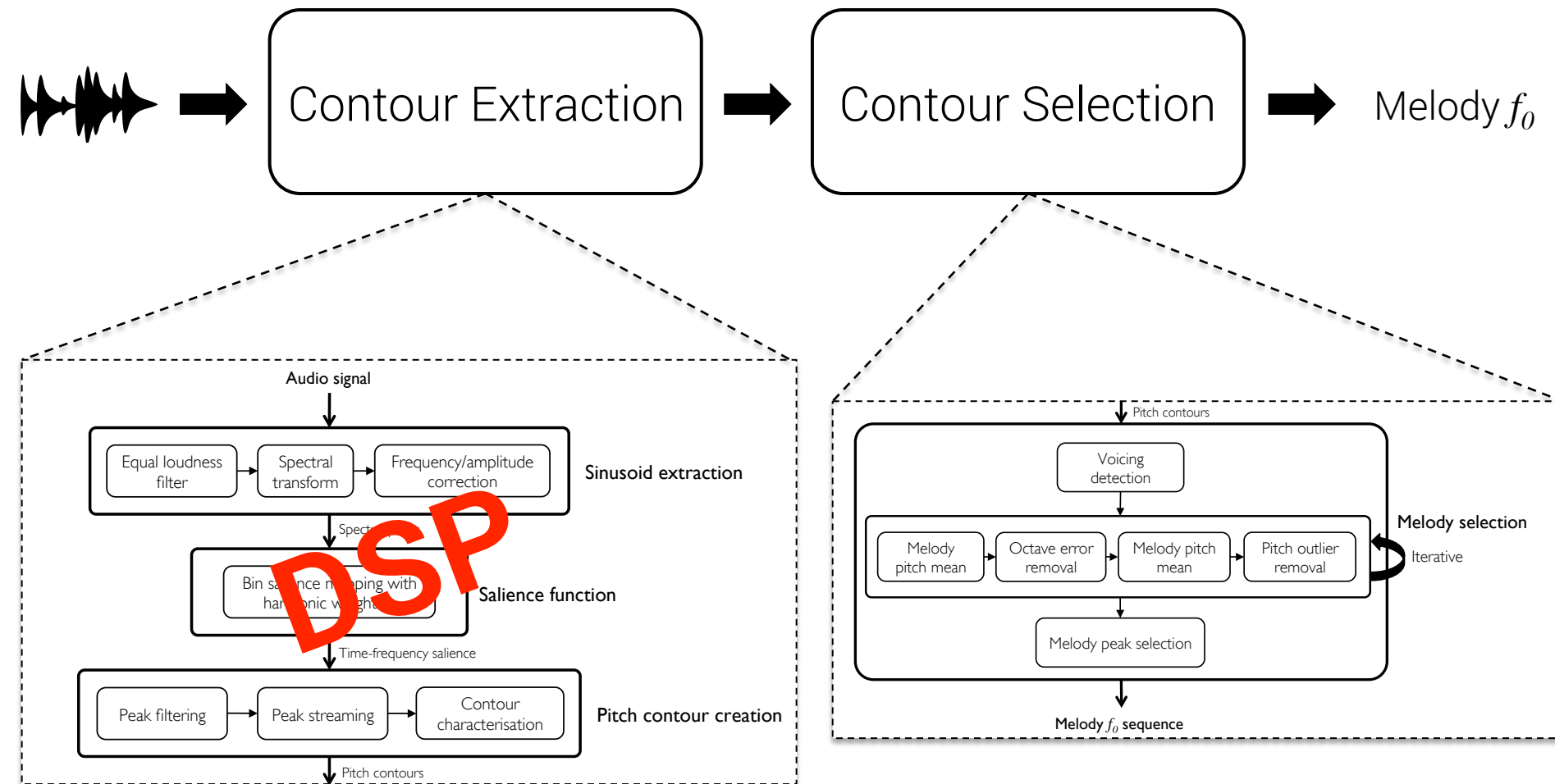
## Melodia (Salamon & Gómez, 2012)



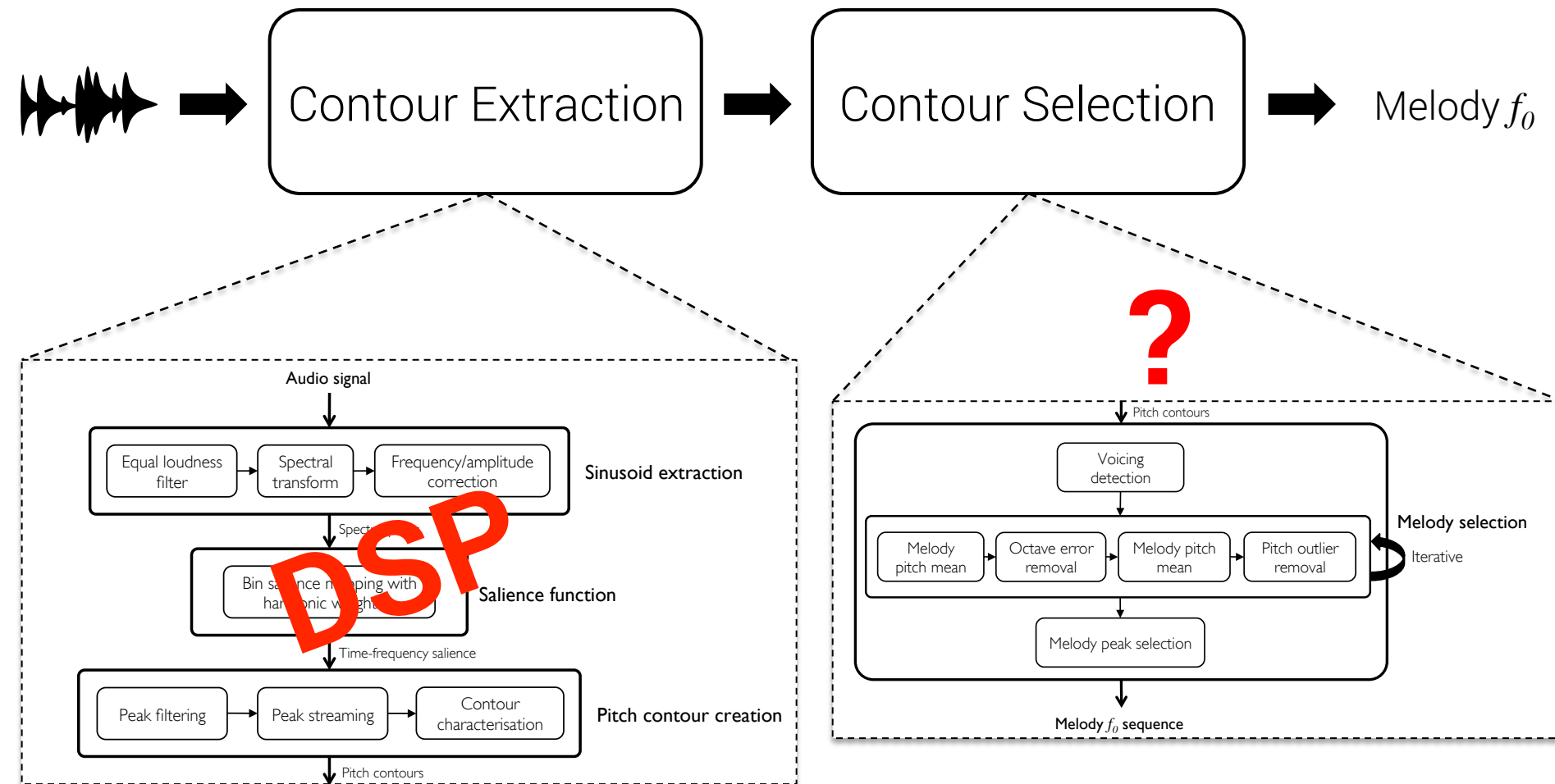
# Melodia (Salamon & Gómez, 2012)



# Melodia (Salamon & Gómez, 2012)



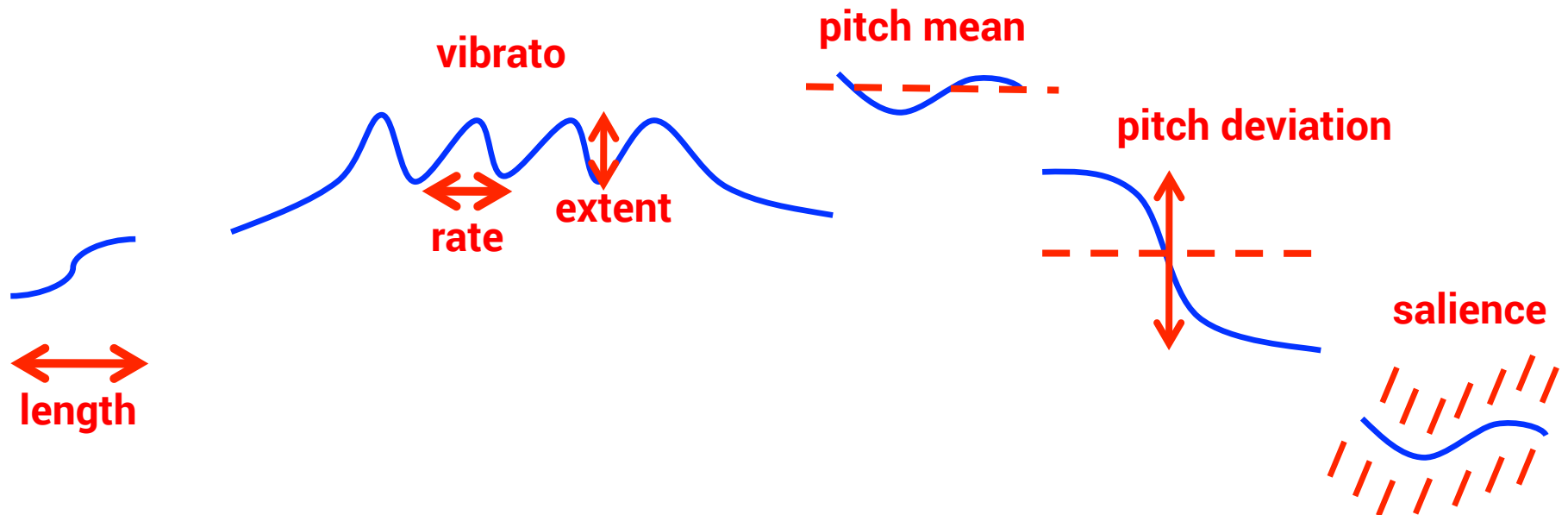
# Melodia (Salamon & Gómez, 2012)



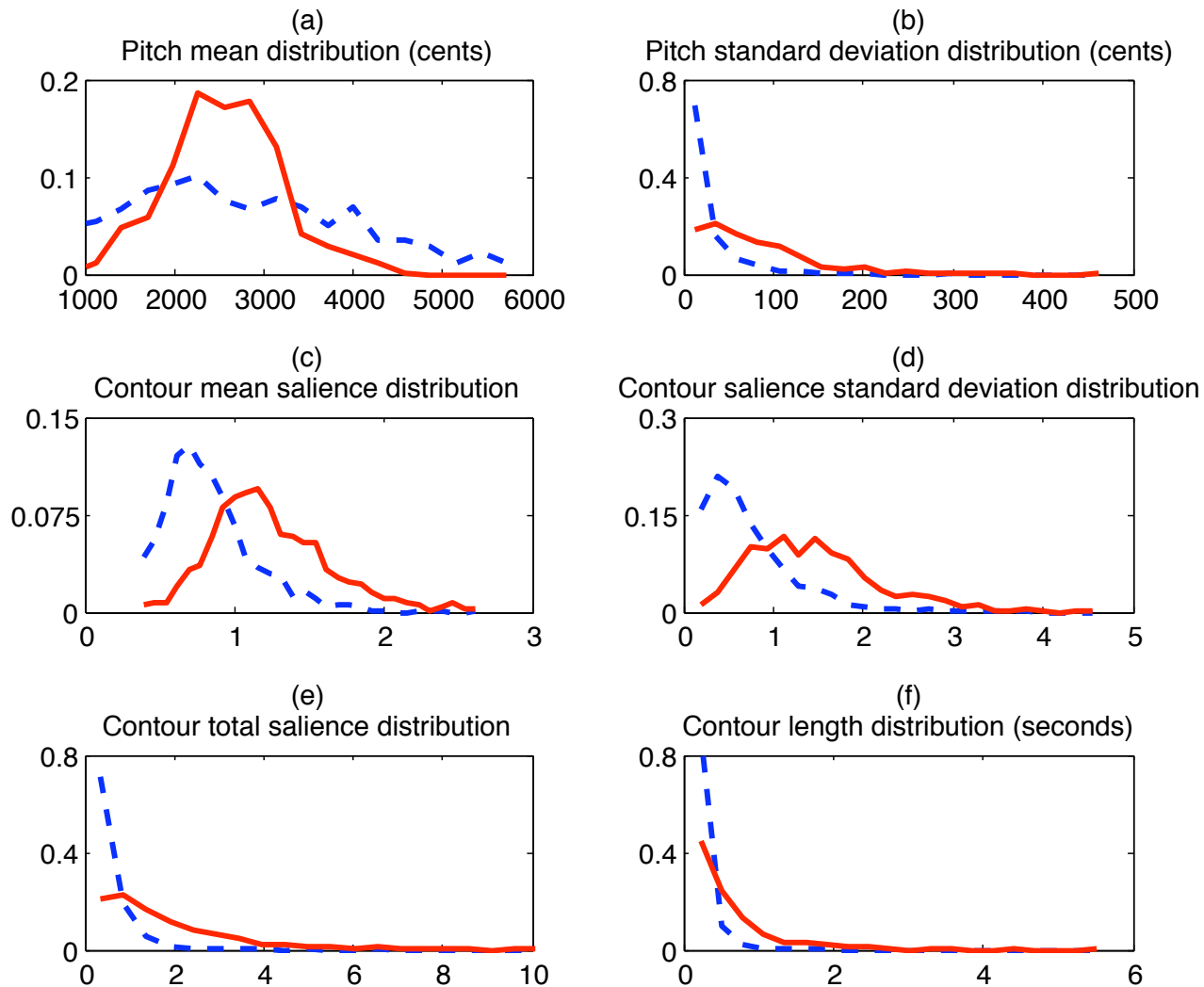
## Melodia (Salamon & Gómez, 2012)



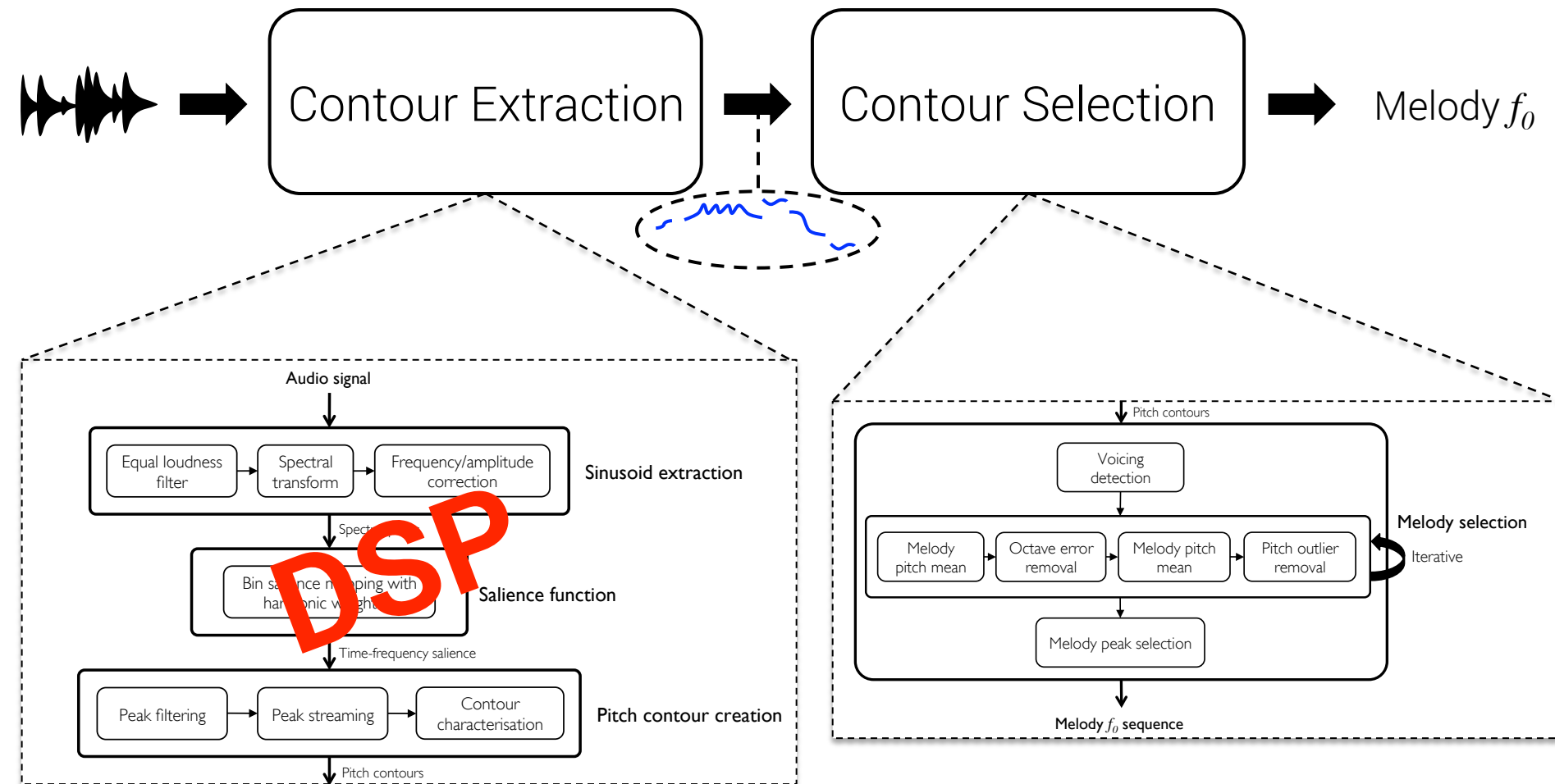
# Melodia (Salamon & Gómez, 2012)



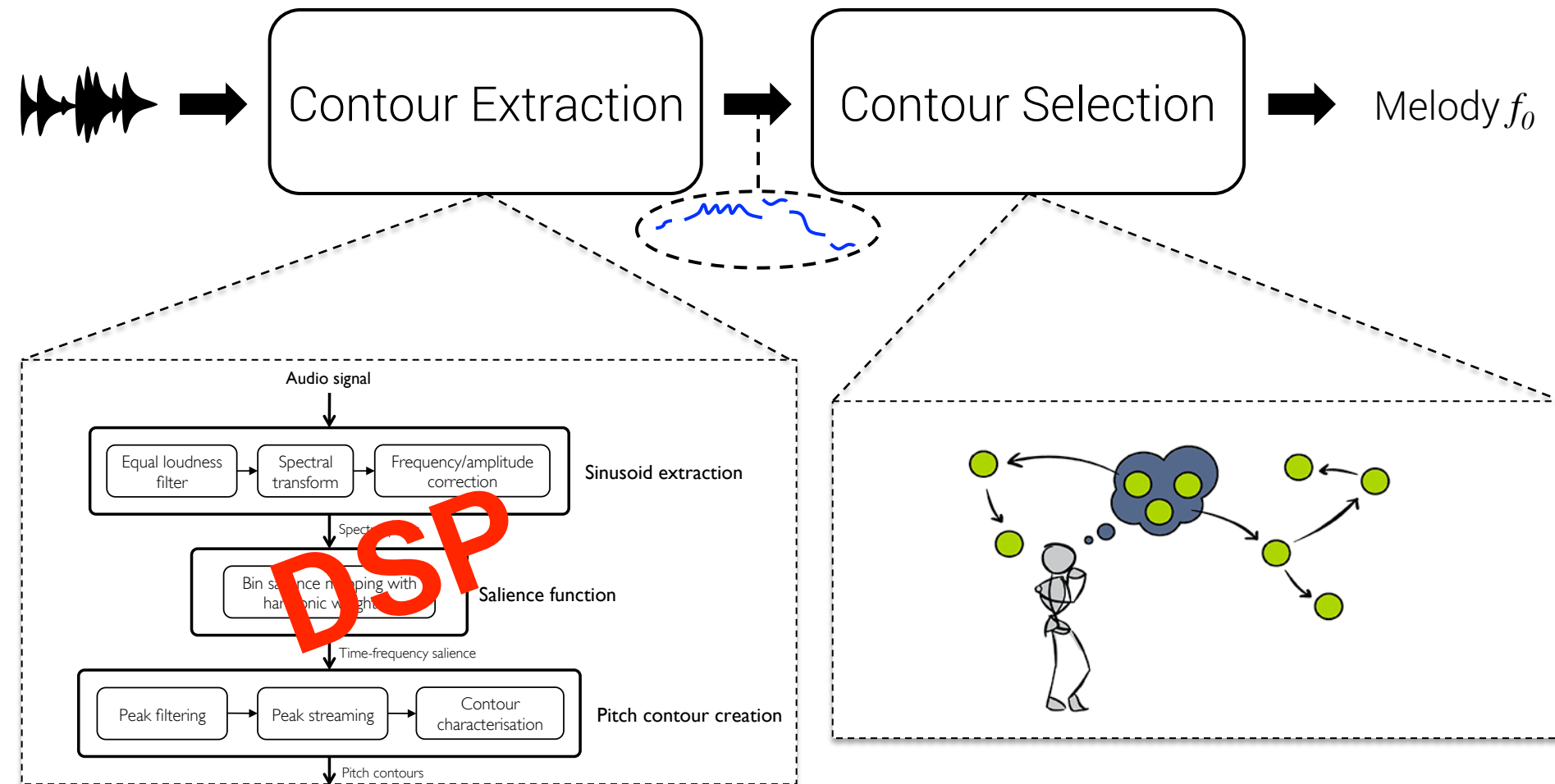
# Melodia (Salamon & Gómez, 2012)



# Melodia (Salamon & Gómez, 2012)

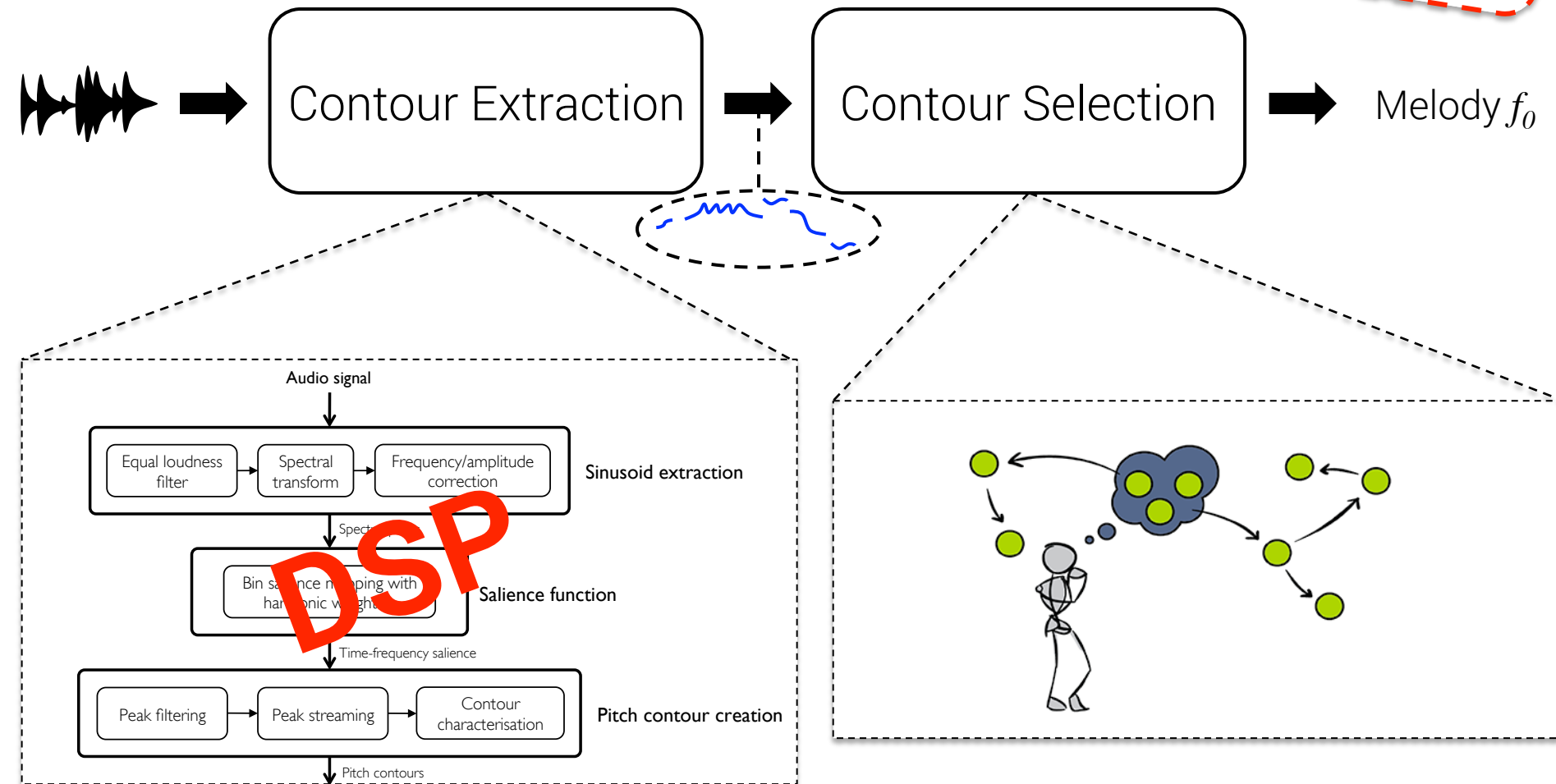


# Melodia (Salamon & Gómez, 2012)

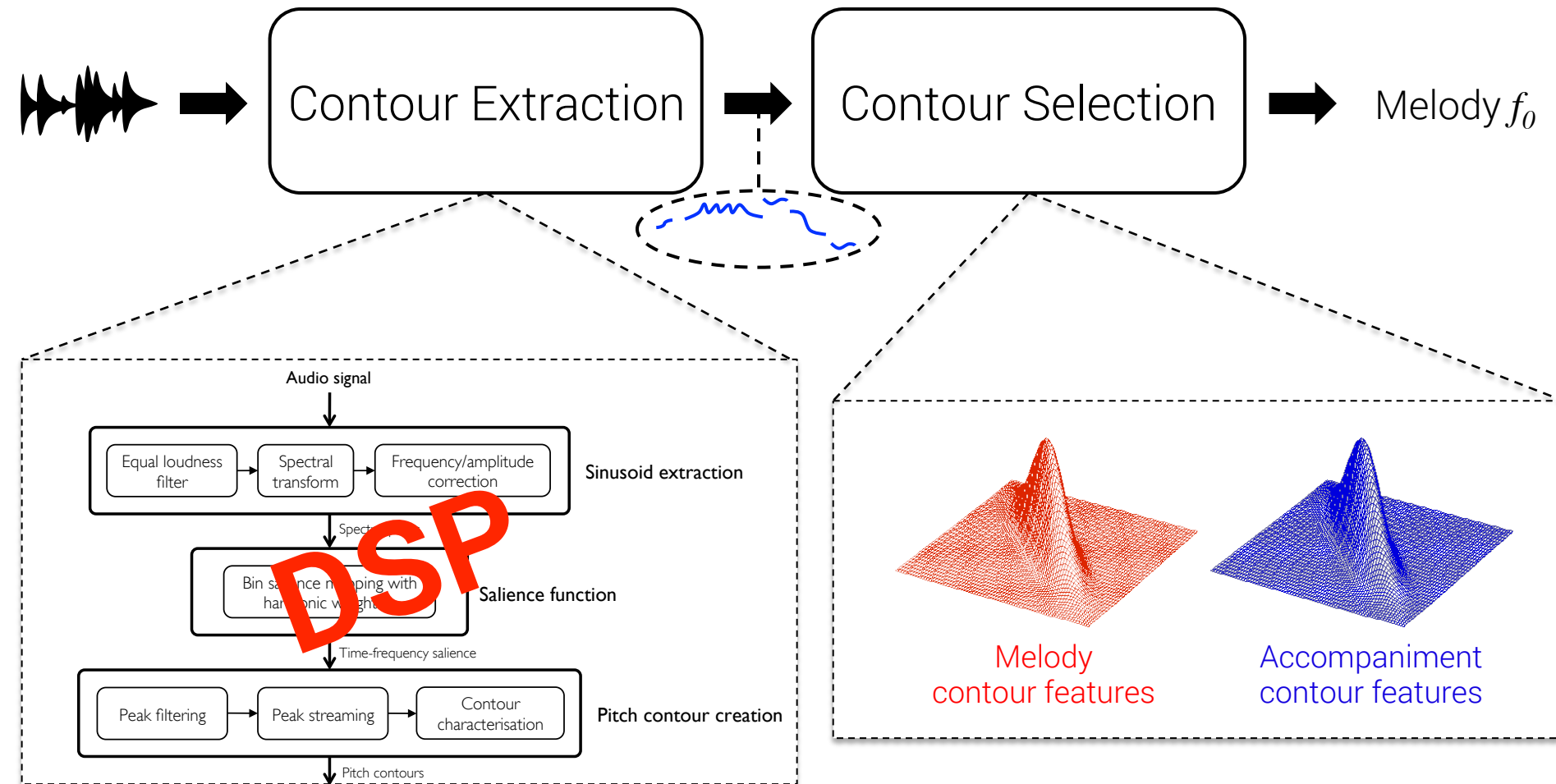


# Melodia (Salamon & Gómez, 2012)

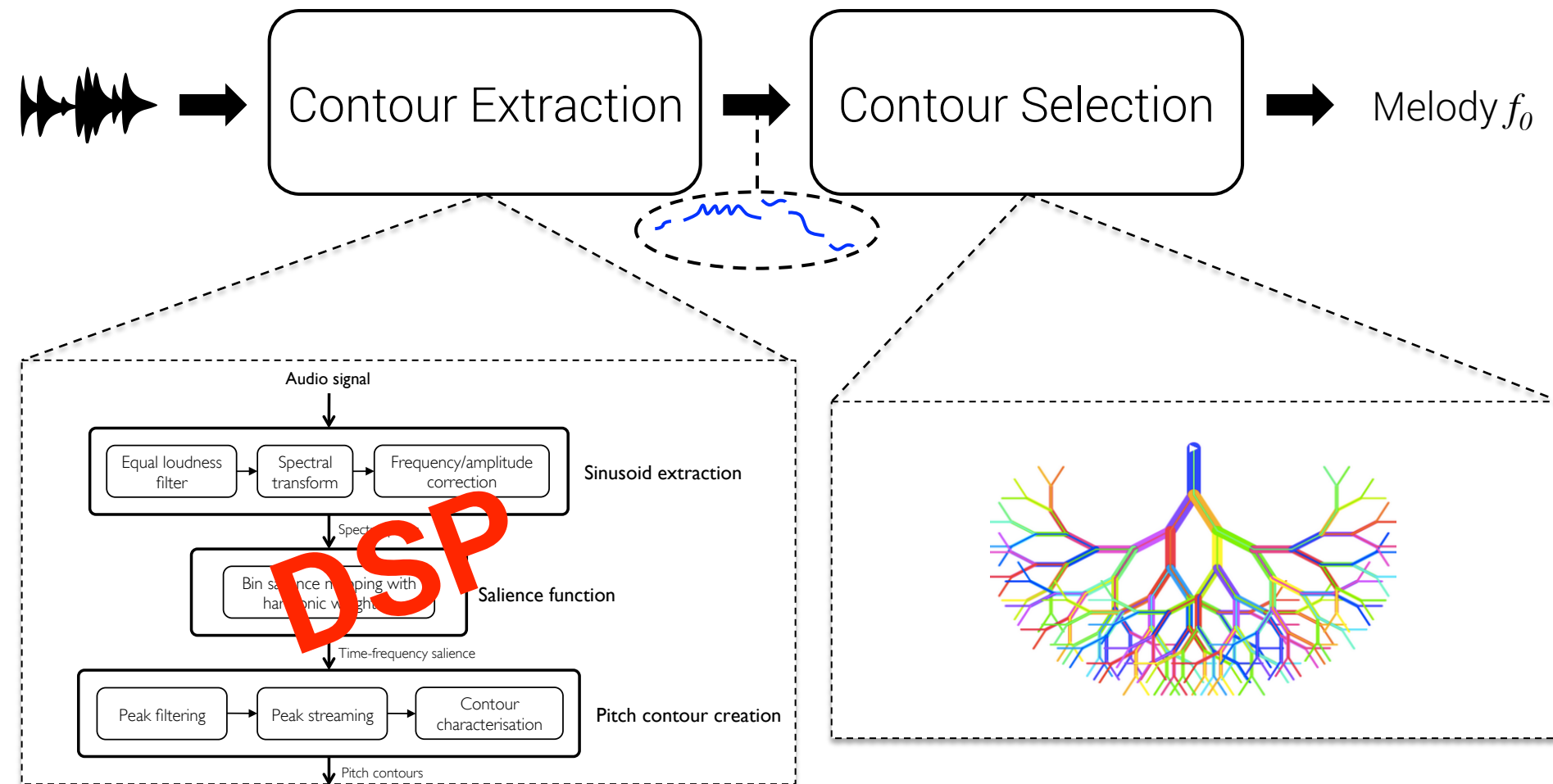
MIREX: 75%  
MedleyDB: 57%



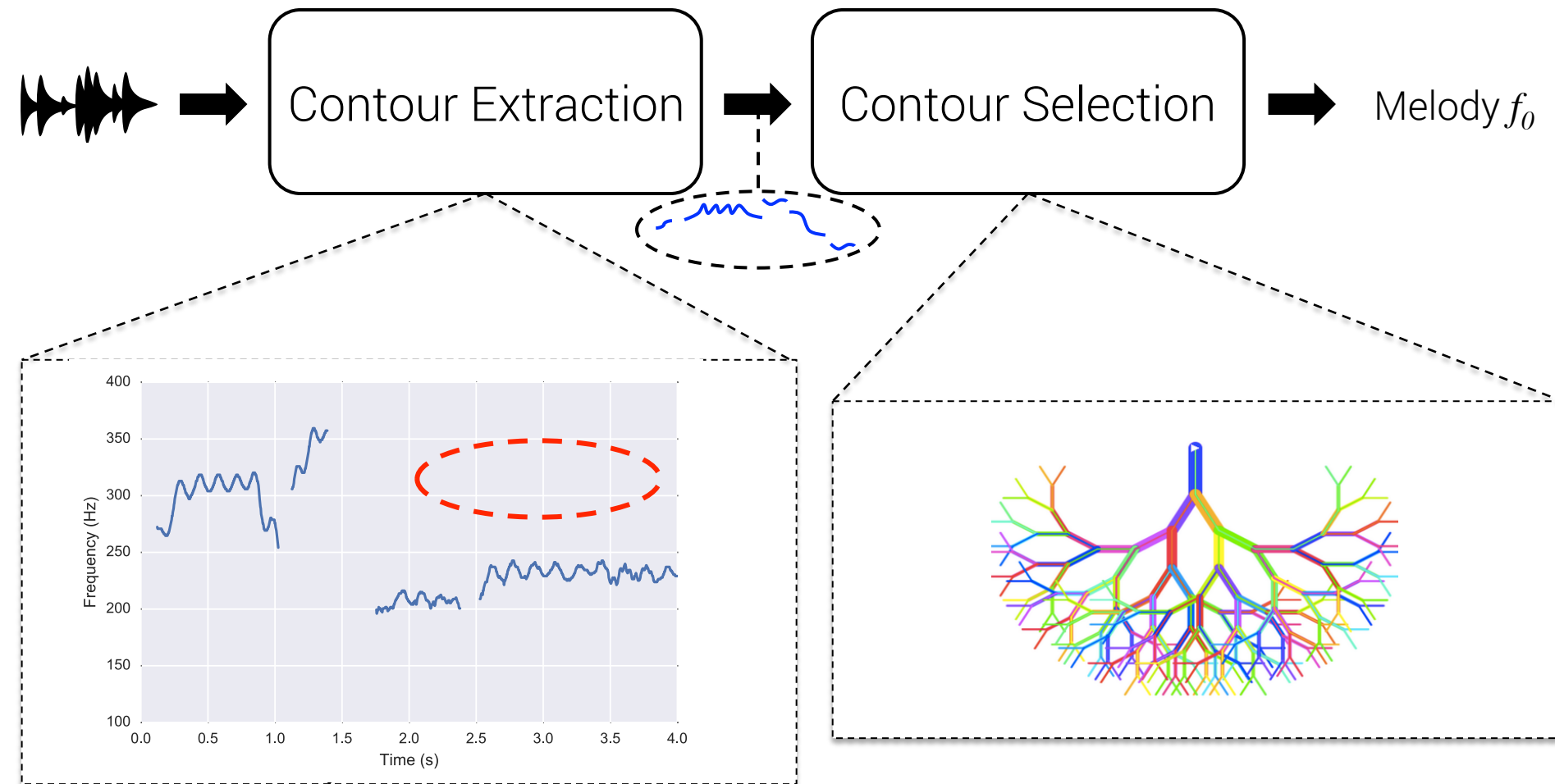
# Contour Classification (Salamon, Peeters & Röbel, 2012)



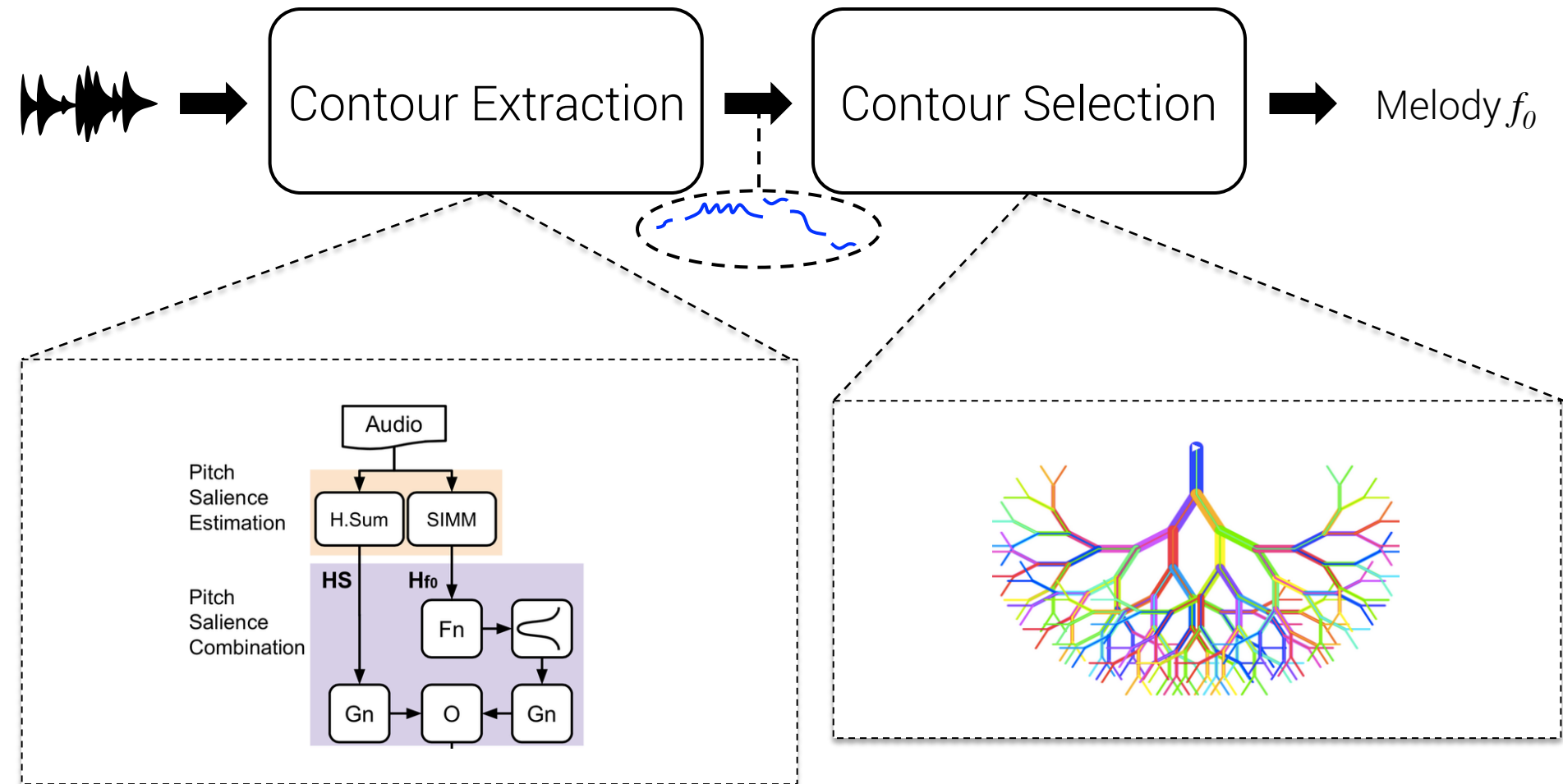
# Contour Classification (Bittner et al., 2015)



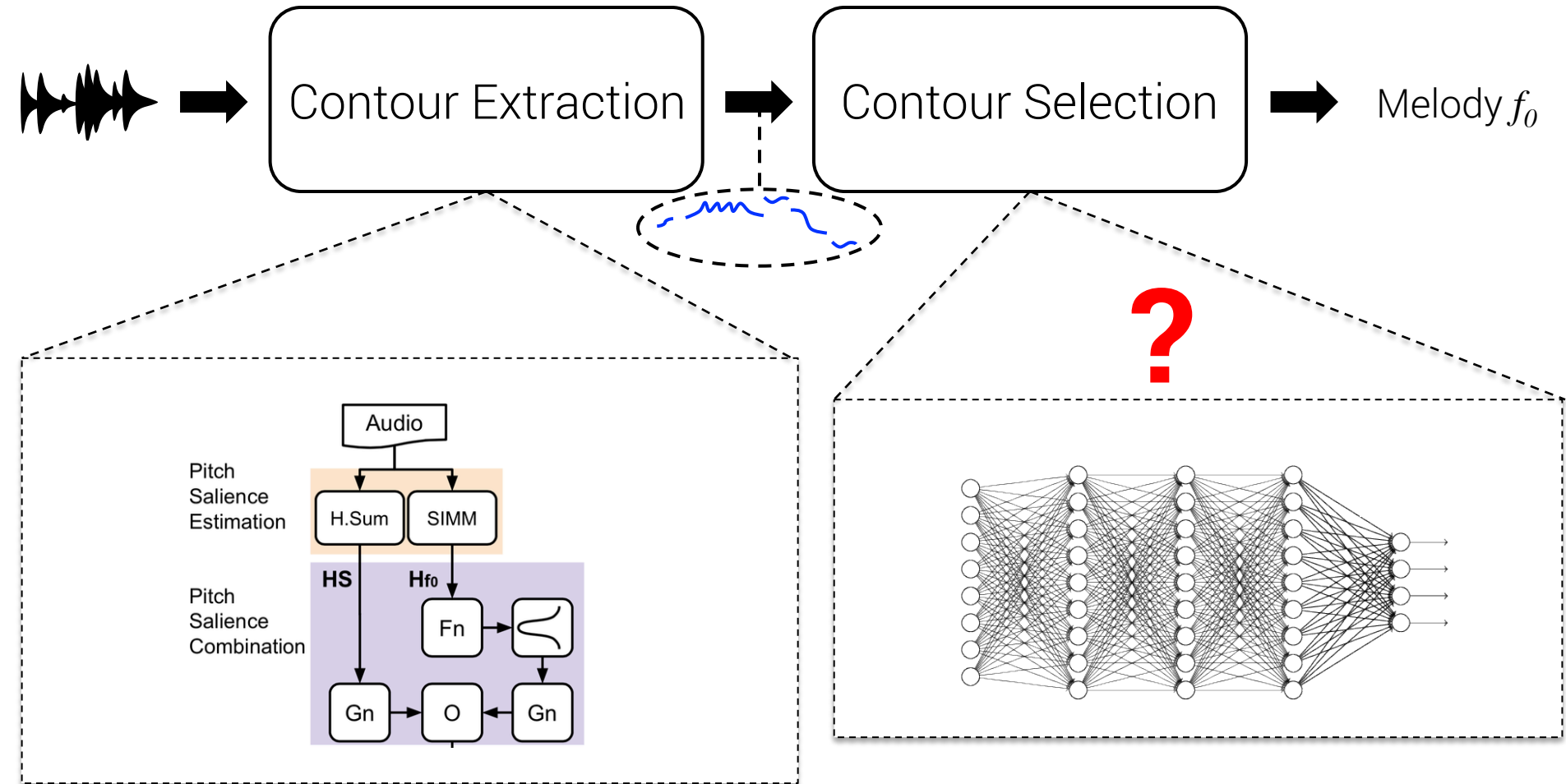
# Contour Classification (Bittner et al., 2015)



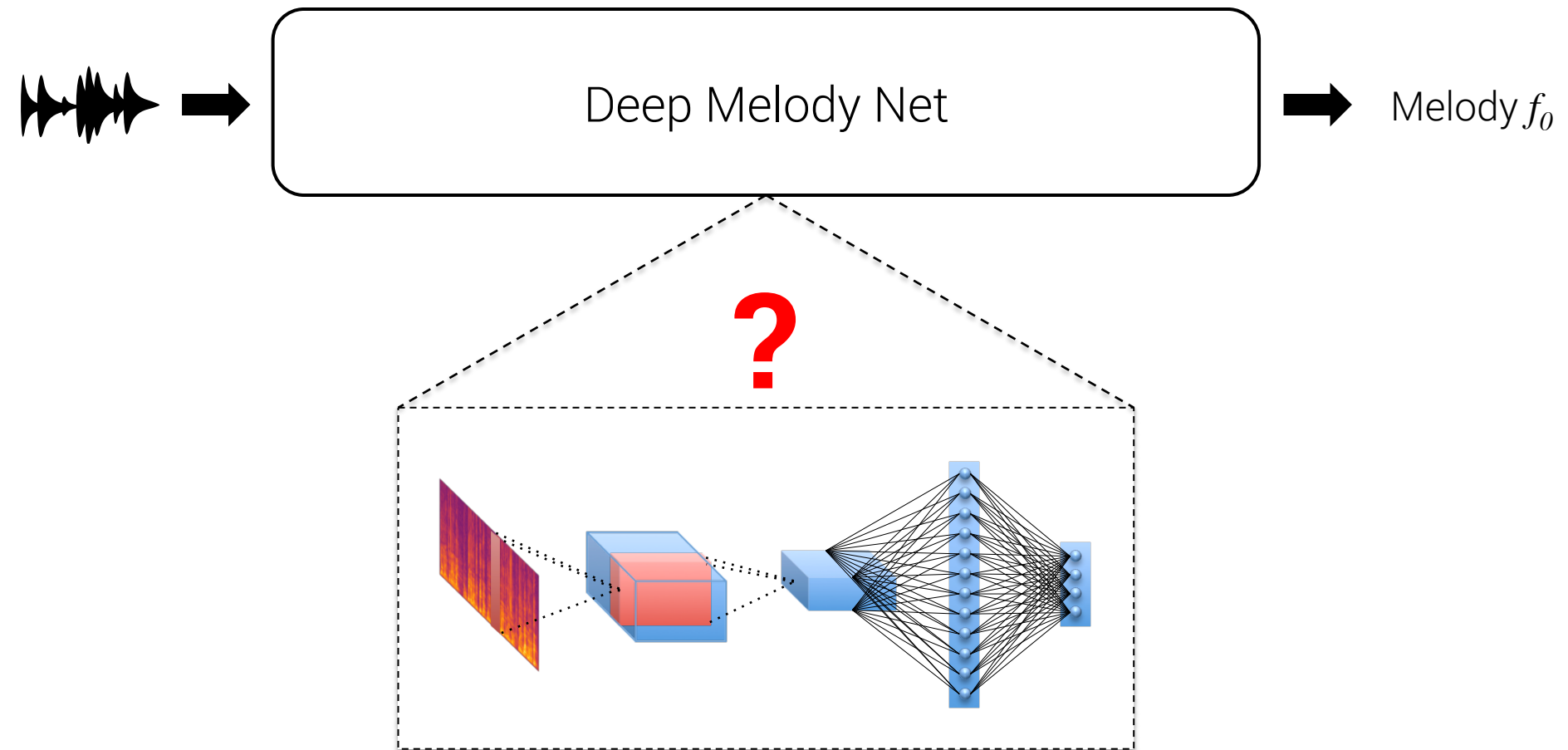
# Contour Classification (Bosch et al., 2016)



# Contour Classification (???, 2017)

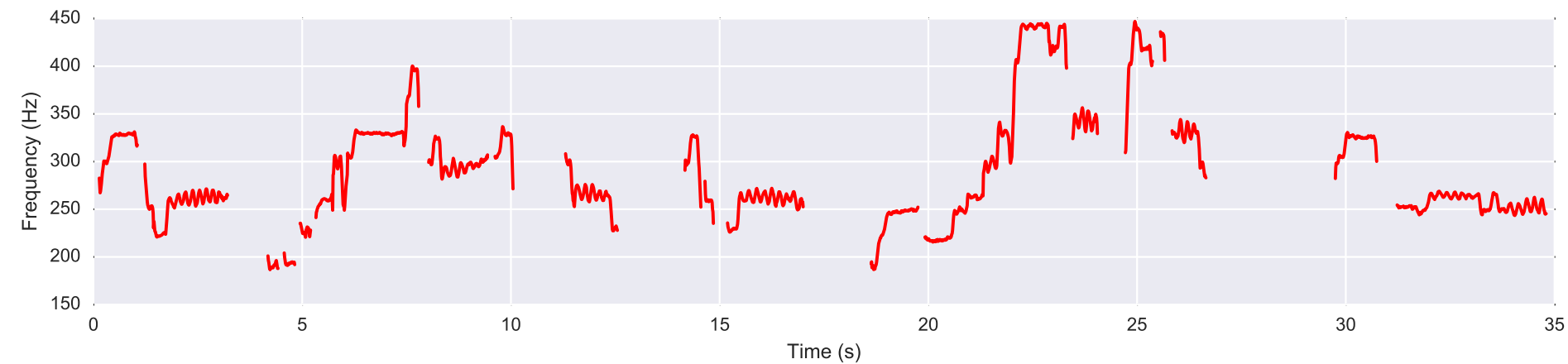


# Contour Classification (???, 2017)

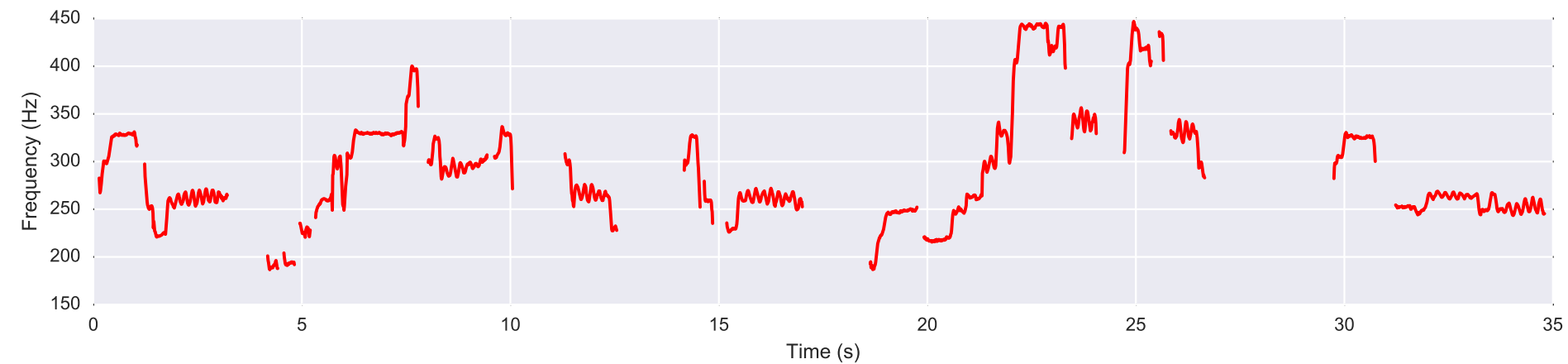
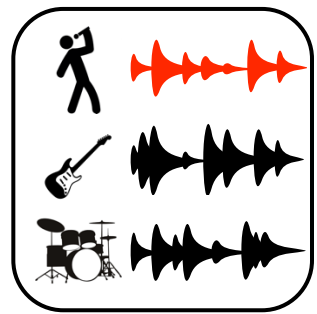


# Data Scarcity

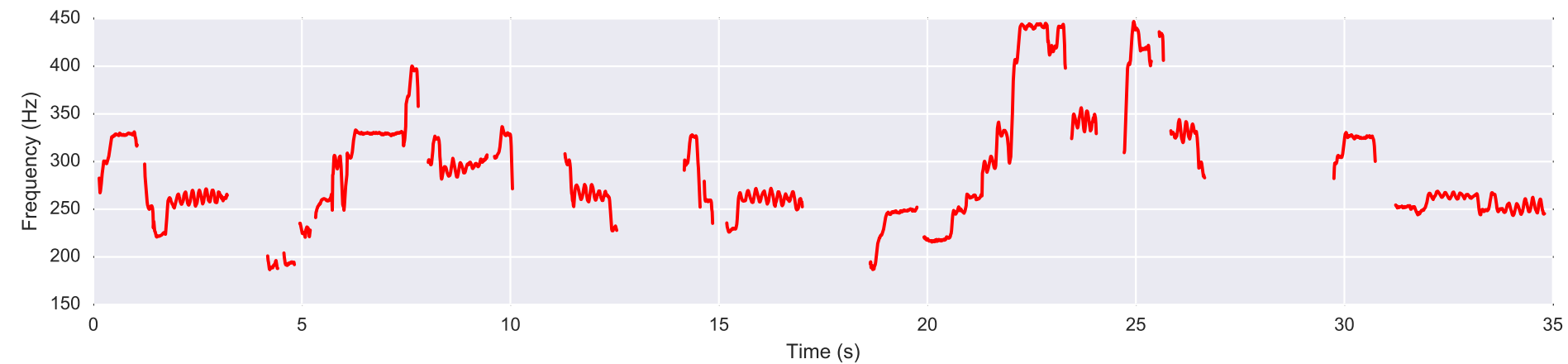
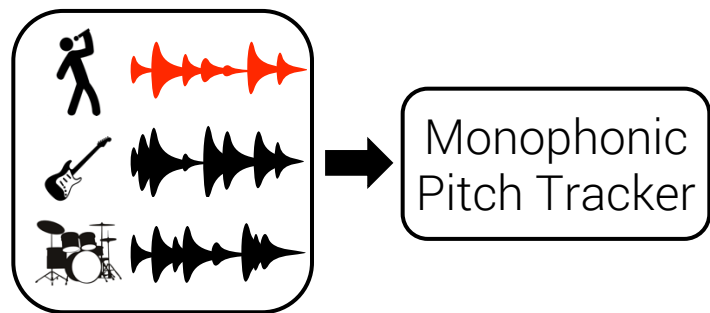
# Continuous Melody $f_0$ Annotation



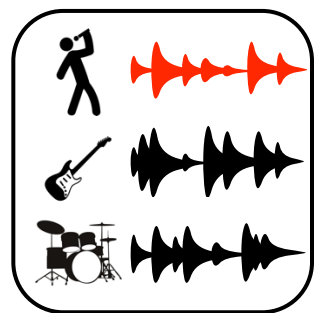
# Continuous Melody $f_0$ Annotation



# Continuous Melody $f_0$ Annotation

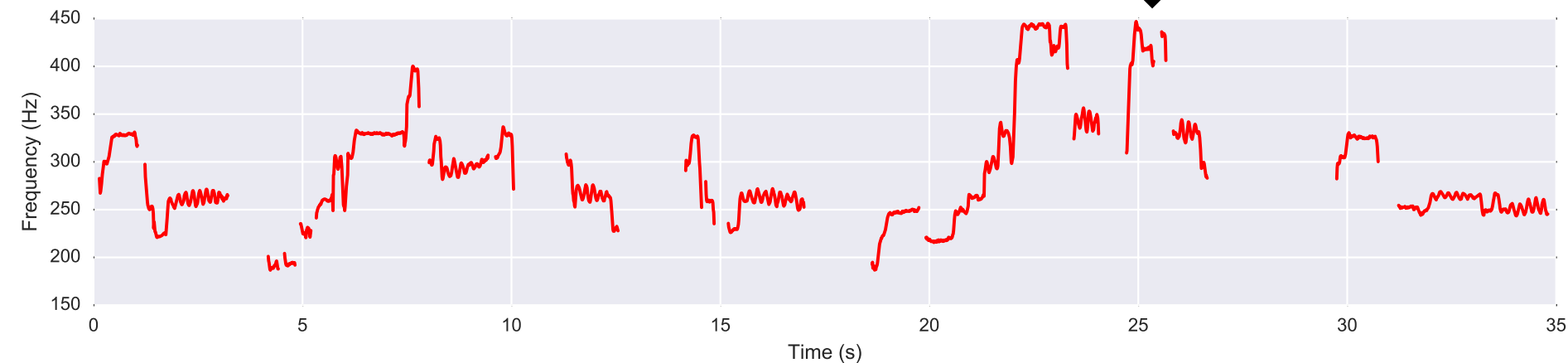
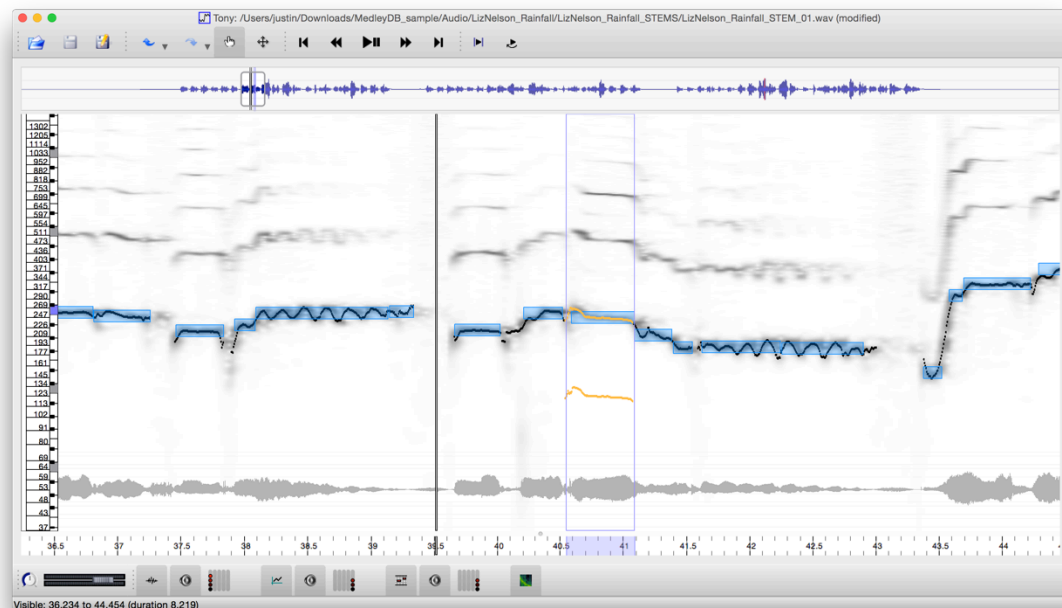


# Continuous Melody $f_0$ Annotation



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

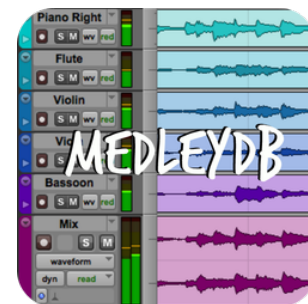
- ▶ MIREX
  - ▶ ADC2004: 6 minutes (20 excerpts)
  - ▶ MIREX05: 12 minutes (25 excerpts)
- ▶ MIR1K: 2.2 hr (1000 excerpts), C-Pop

## 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

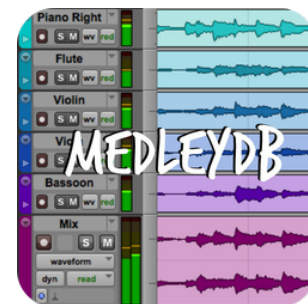
## Datasets for Melody Extraction Eval

- ▶ 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

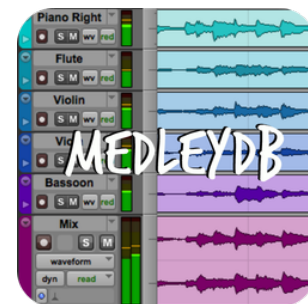


108 songs

~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

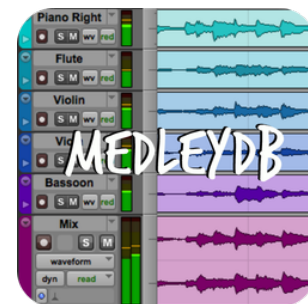


108 songs  
~50 annotator-hours

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

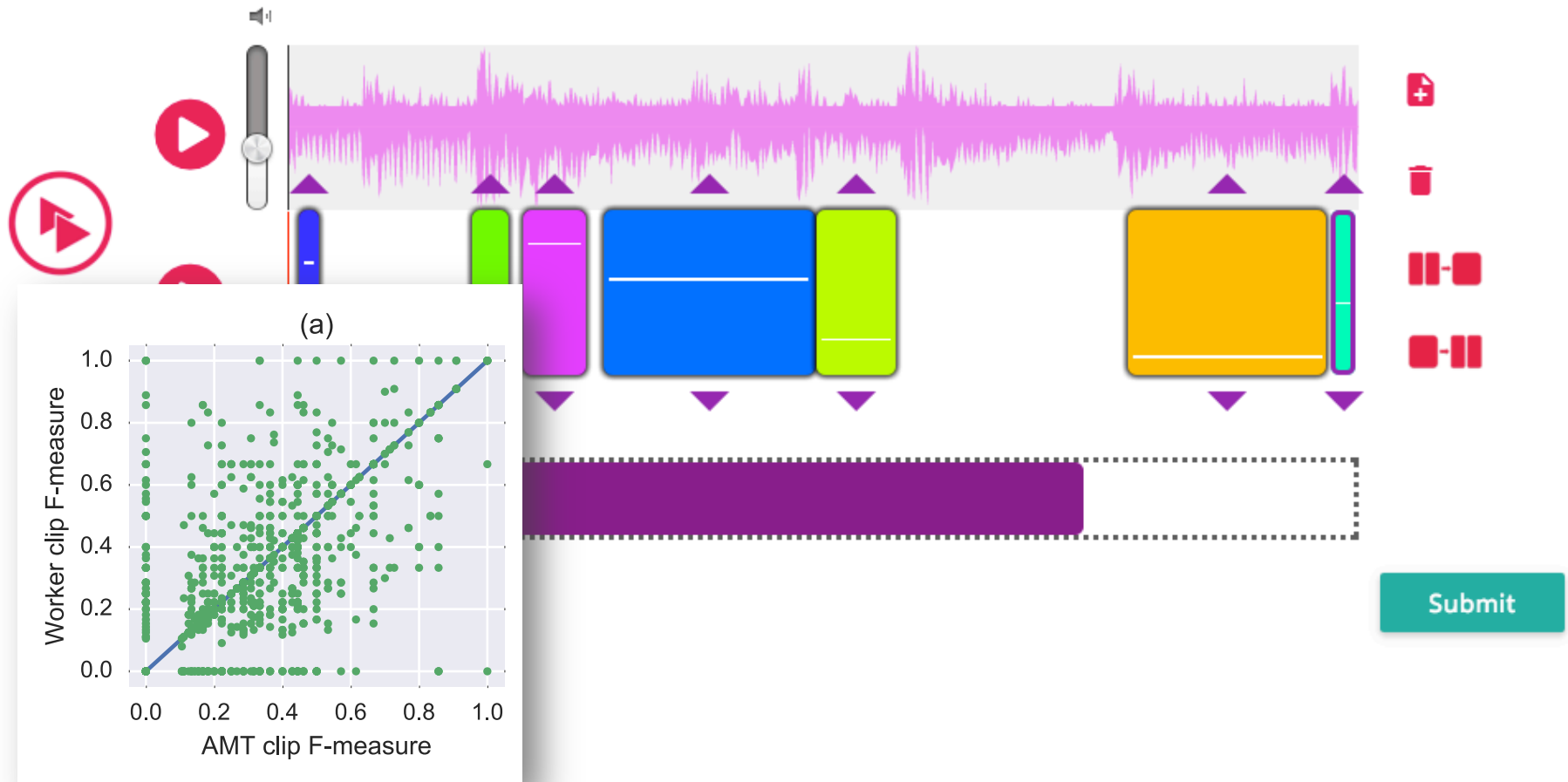
# Crowdsourcing Melody Note Annotations

Ensemble (Tse et al., 2016)

The interface displays a music annotation tool. At the top, a pink waveform represents the audio input. Below it, a series of colored rectangular blocks (blue, green, magenta, blue, yellow-green, and orange) represent annotated notes. Each note block has a white horizontal line indicating its pitch and a vertical line indicating its duration. Purple arrows point from the waveform to the notes, and purple triangles point from the notes to the waveform. On the left side, there are three red circular buttons: a play button, a volume slider, and a red button with a white arrow pointing right. On the right side, there are four red icons: a plus sign, a trash can, a pause button, and a play button. At the bottom, there is a dashed rectangular box containing a solid purple rectangle. In the bottom right corner, there is a teal button labeled "Submit".

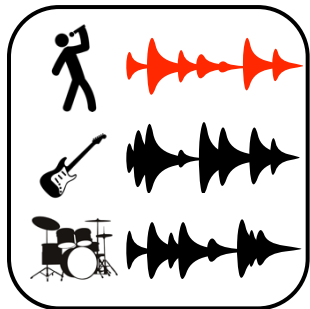
# Crowdsourcing Melody Note Annotations

Ensemble (Tse et al., 2016)

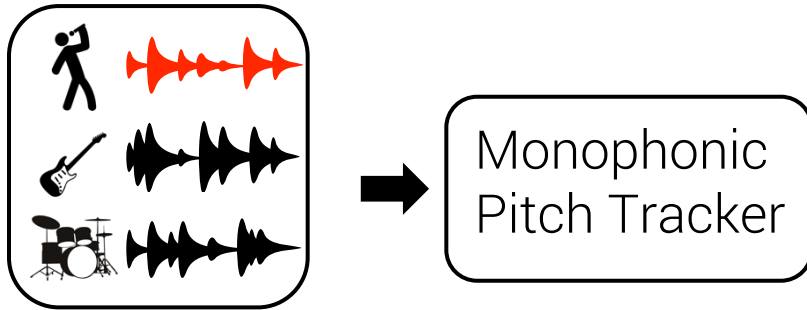


# Data Augmentation: $f_0$ Annotation-by-Synthesis

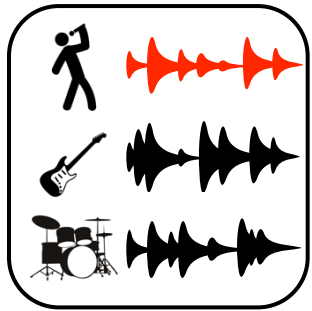
# Data Augmentation: $f_0$ Annotation-by-Synthesis



## Data Augmentation: $f_0$ Annotation-by-Synthesis



## Data Augmentation: $f_0$ Annotation-by-Synthesis



Monophonic  
Pitch Tracker



Cleaning +  
Smoothing

## Data Augmentation: $f_0$ Annotation-by-Synthesis



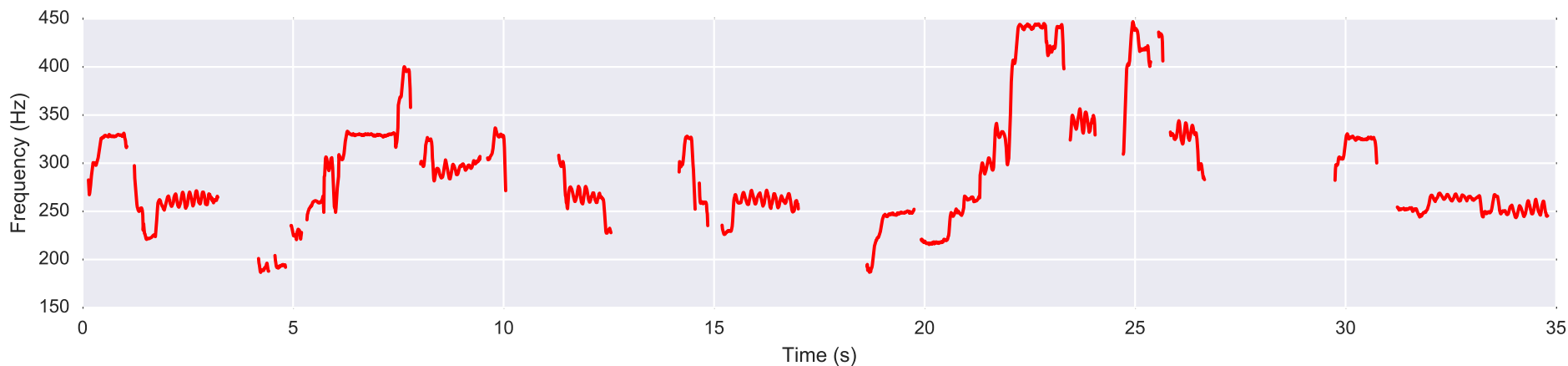
Monophonic  
Pitch Tracker



Cleaning +  
Smoothing



Sinusoidal  
Modelling



## Data Augmentation: $f_0$ Annotation-by-Synthesis



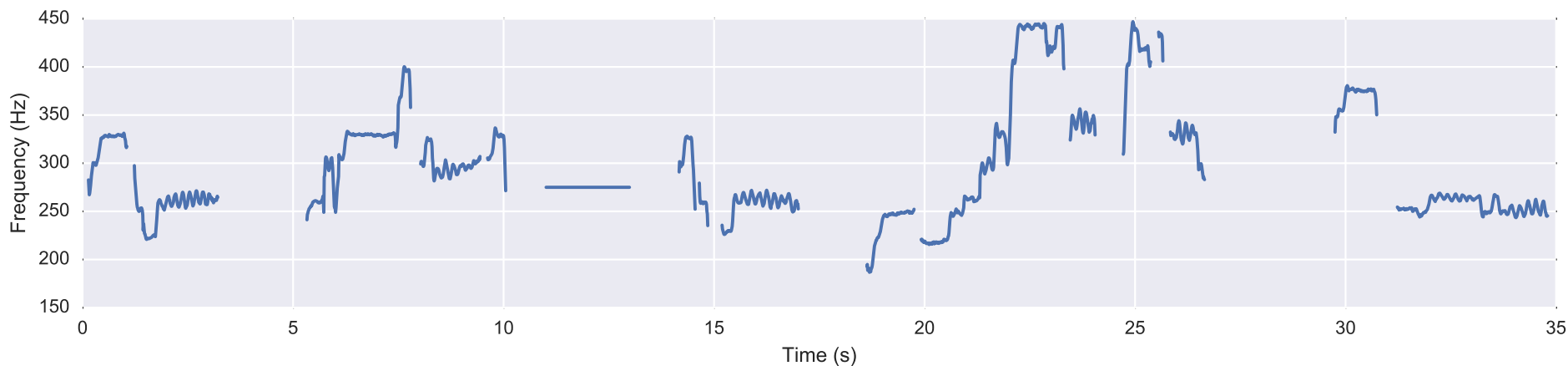
Monophonic  
Pitch Tracker



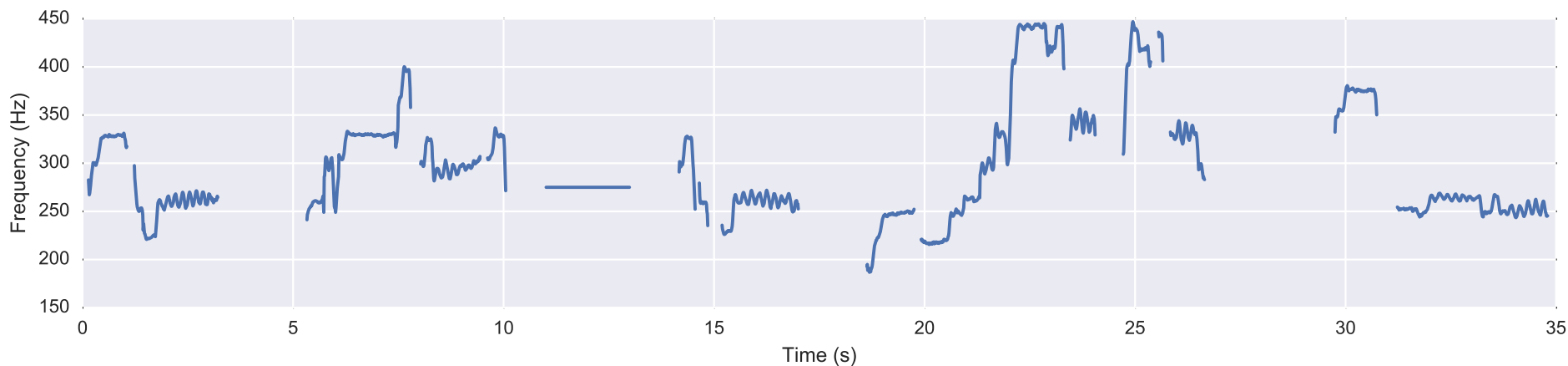
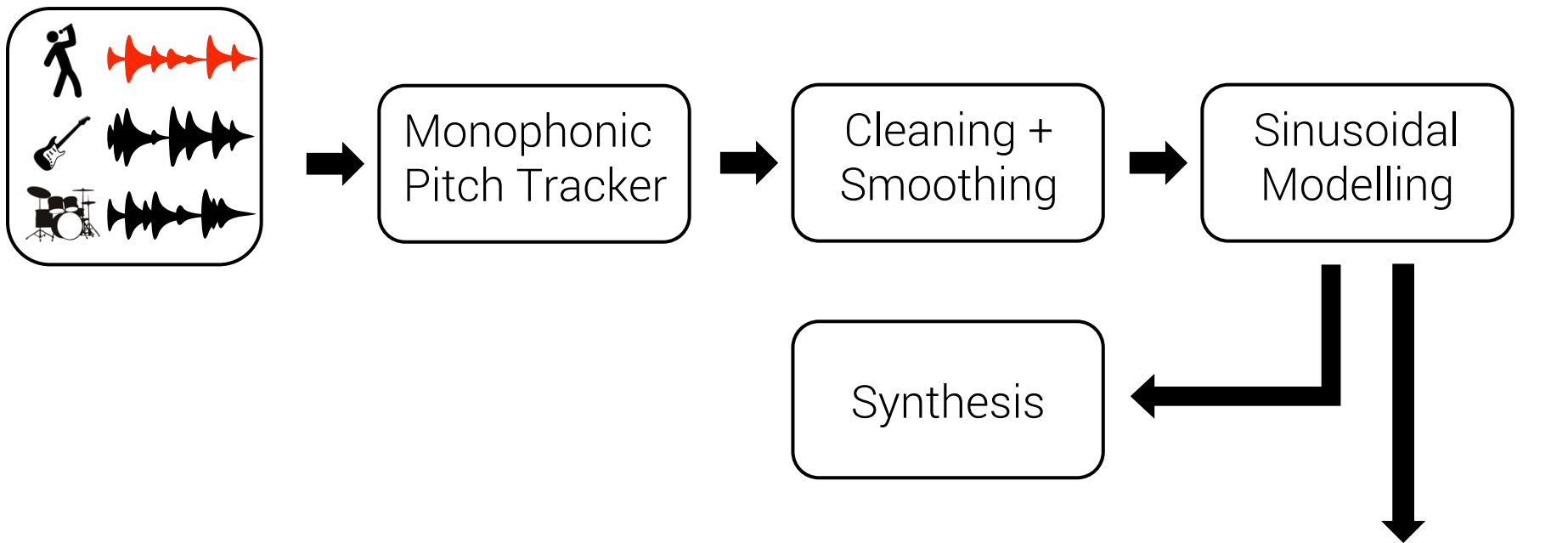
Cleaning +  
Smoothing



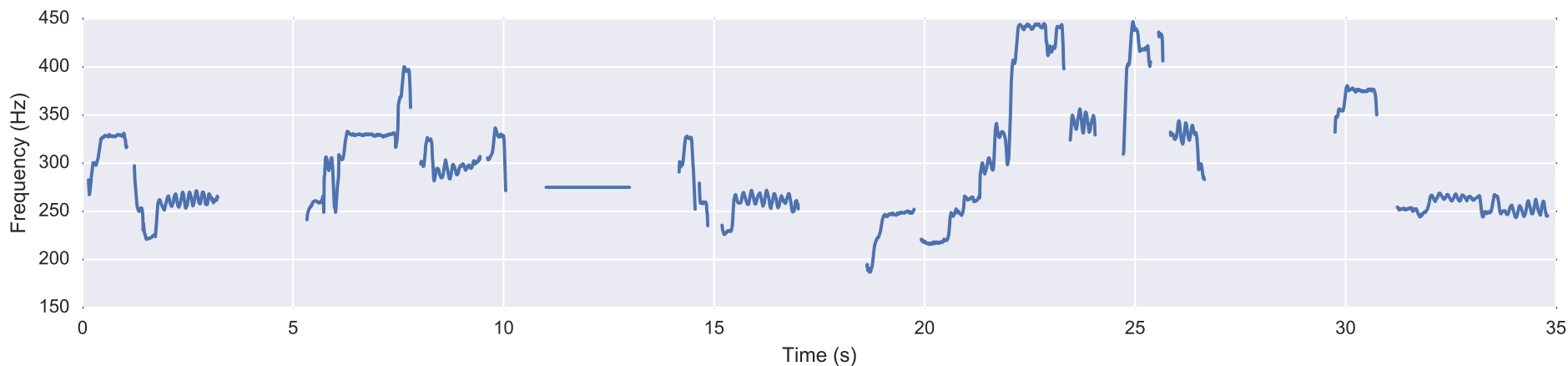
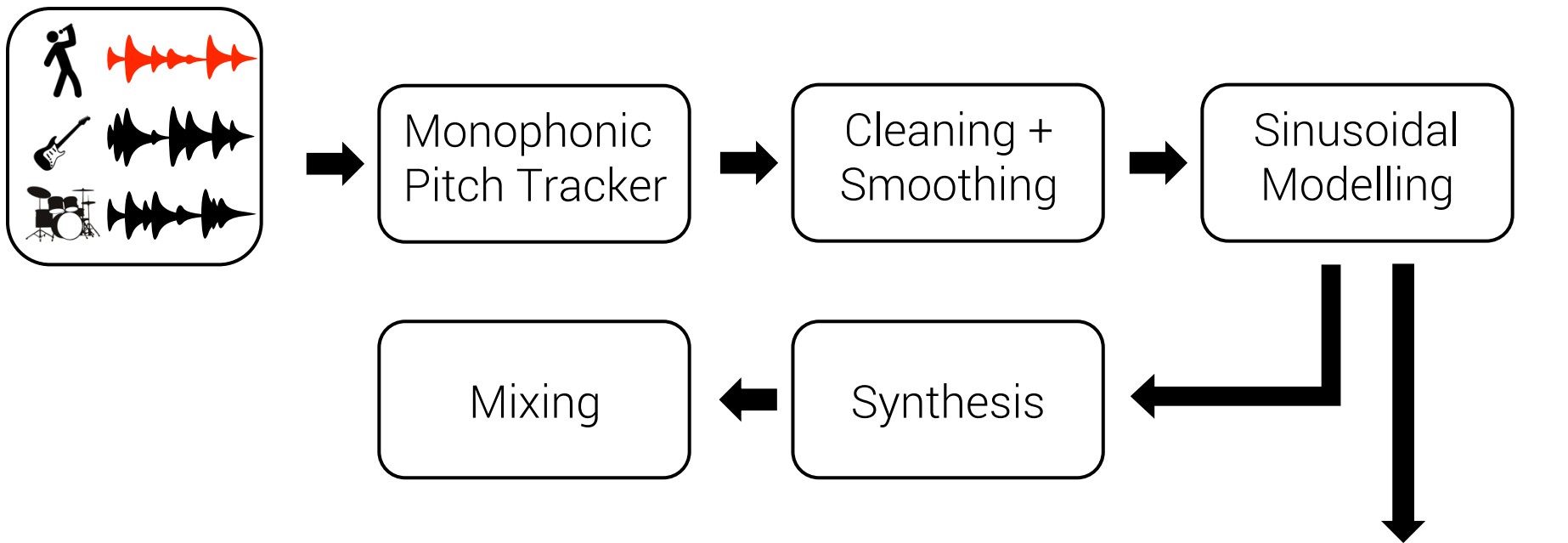
Sinusoidal  
Modelling



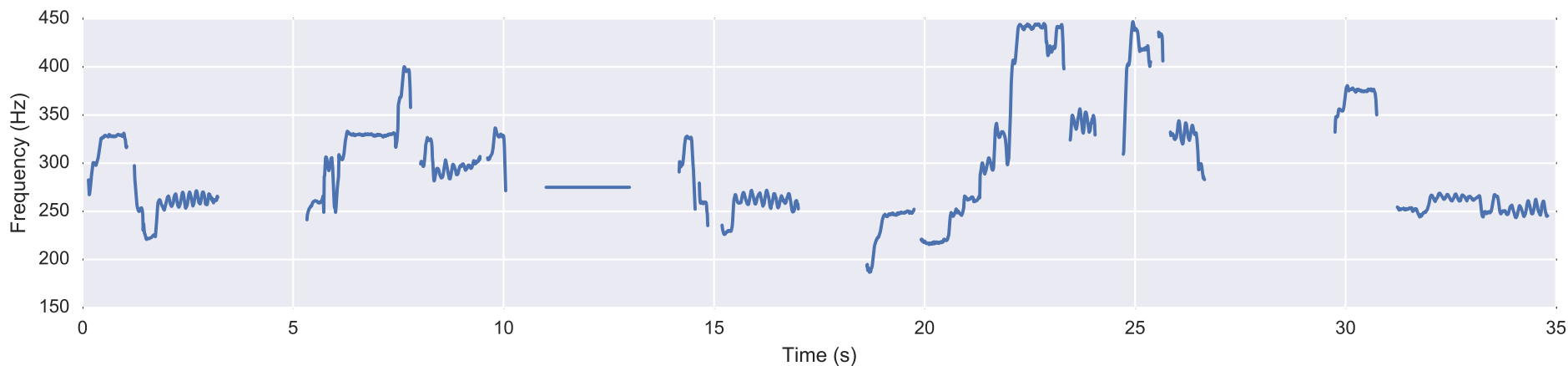
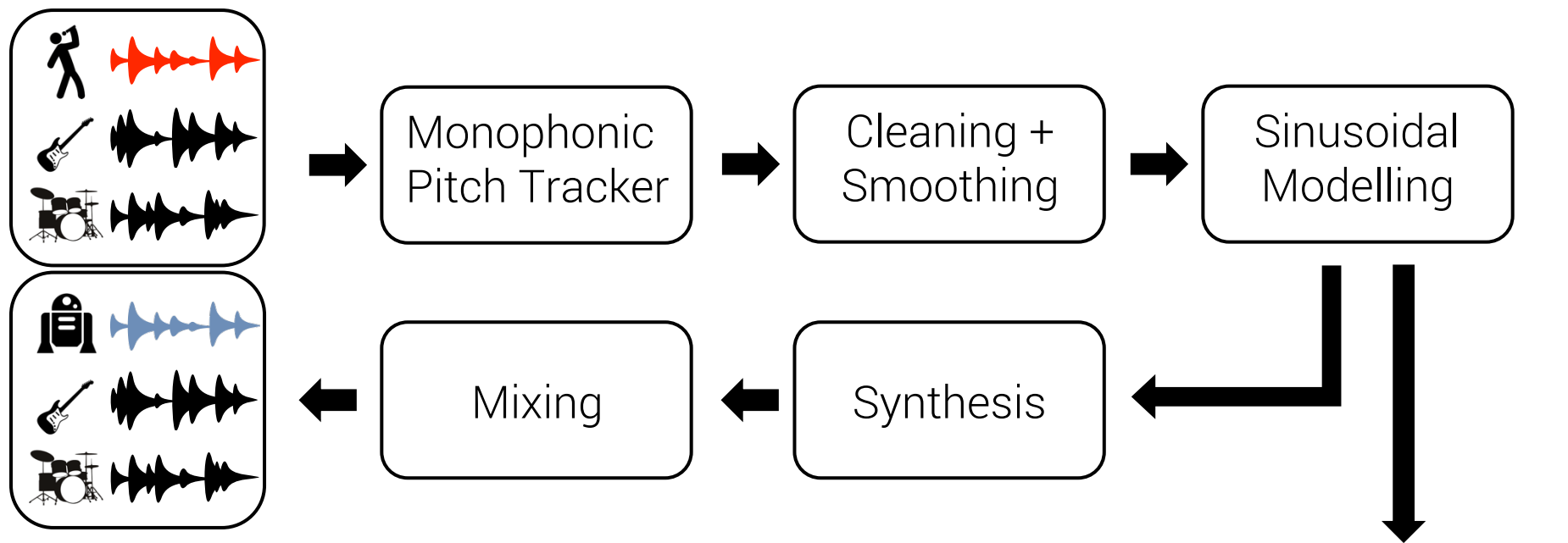
## Data Augmentation: $f_0$ Annotation-by-Synthesis



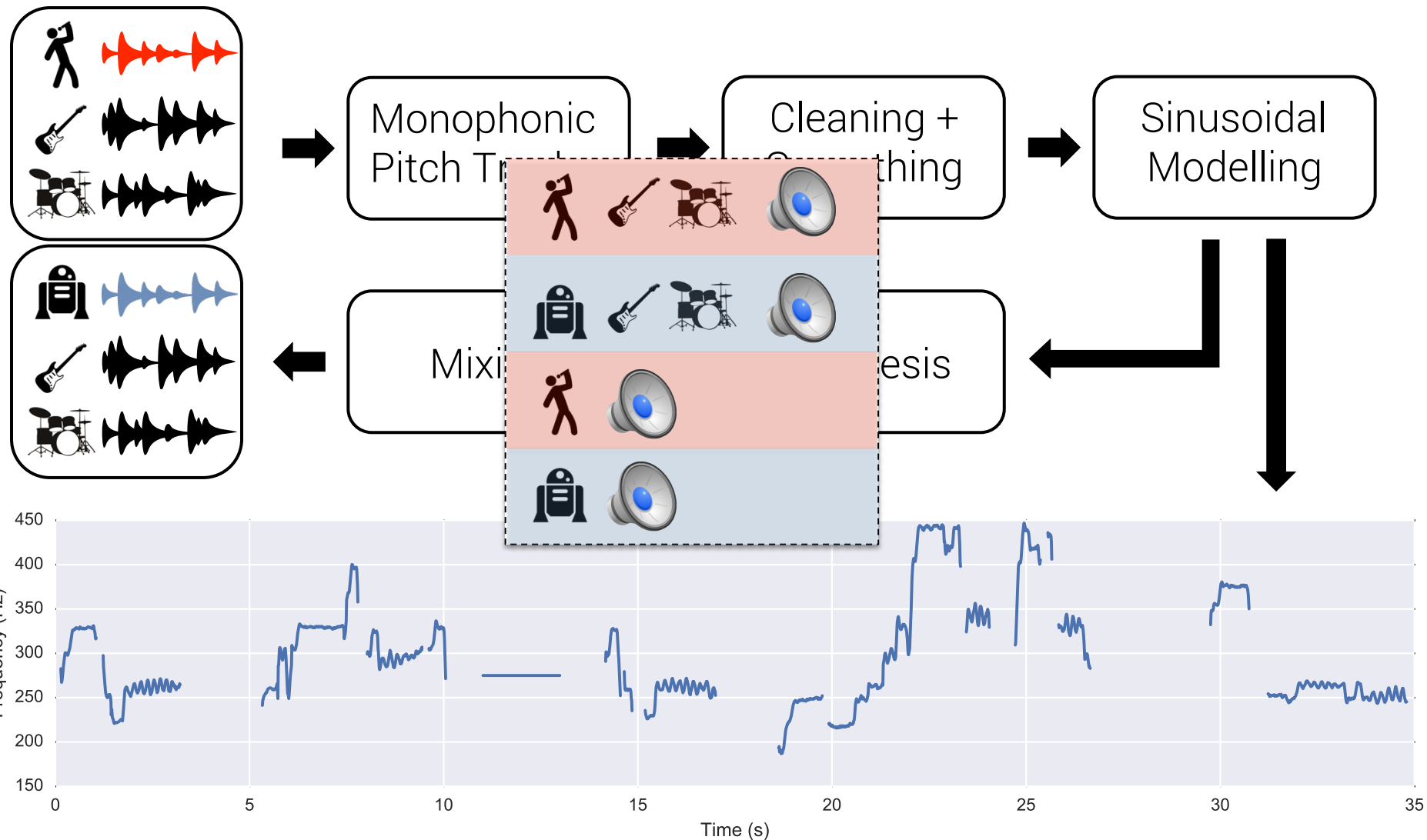
## Data Augmentation: $f_0$ Annotation-by-Synthesis



## Data Augmentation: $f_0$ Annotation-by-Synthesis

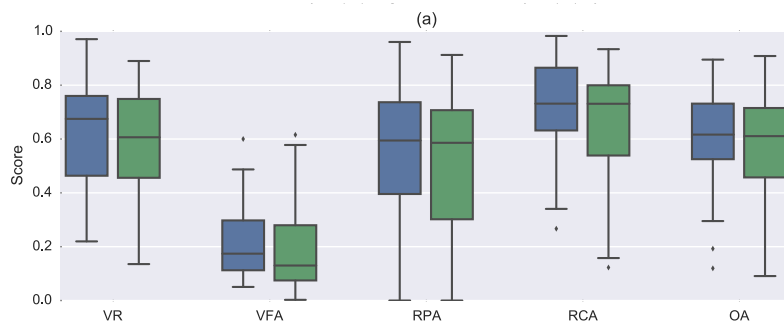


# Data Augmentation: $f_0$ Annotation-by-Synthesis

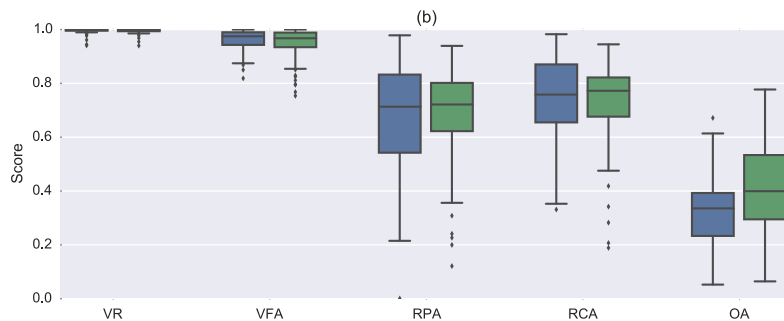


# Data Augmentation: $f_0$ Annotation-by-Synthesis

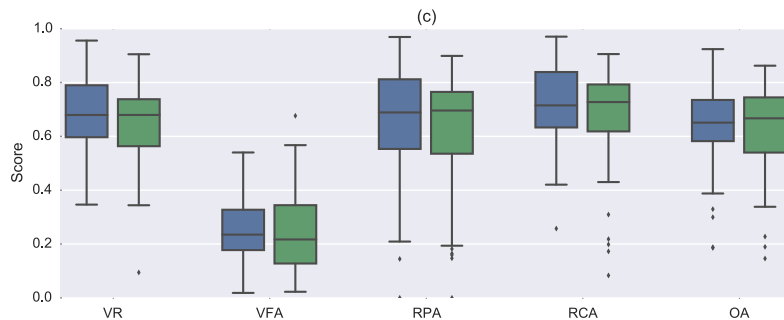
Algorithm 1





Algorithm 2



Algorithm 3



 Original data + manual annotations  
 Synthesized data + automatic annotations

Metrics: A B C D E

# 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!**

