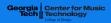


### Introduction to Audio Content Analysis

module 7.3.5: fundamental frequency detection — evaluation

alexander lerch



## introduction overview



#### corresponding textbook section

section 7.3.5

#### lecture content

- evaluation of pitch tracking systems
- challenges in annotation
- metrics

### learning objectives

• successfully plan a systematic evaluation procedure for a pitch detection system



### introduction overview



#### corresponding textbook section

section 7.3.5

#### lecture content

- evaluation of pitch tracking systems
- challenges in annotation
- metrics

### learning objectives

• successfully plan a systematic evaluation procedure for a pitch detection system



## pitch evaluation



#### goal: compare predicted pitch and ground truth pitch

- differentiate various 'pitch tracking' tasks
  - pitch of individual notes
  - pitch of monophonic melody
  - pitch of **pre-dominant melody** in polyphonic mixture
  - pitches in multi-timbral polyphonic mixture

### pitch evaluation tasks



goal: compare predicted pitch and ground truth pitch

- differentiate various 'pitch tracking' tasks
  - pitch of individual notes
  - pitch of monophonic melody
  - pitch of **pre-dominant melody** in polyphonic mixture
  - pitches in multi-timbral polyphonic mixture

# pitch evaluation annotation challenges



- pitch discretization
  - (MIDI/score) pitch of individual notes
  - F0
- time discretization
  - start and stop time of note
  - equidistant time stamps

#### how to annotate F0



# pitch evaluation annotation challenges



- pitch discretization
  - (MIDI/score) pitch of individual notes
  - F0
- time discretization
  - start and stop time of note
  - equidistant time stamps

#### how to annotate F0



- all metrics should be computed in the pitch domain, not the frequency domain
- metrics measure a match between ground truth and predicted pitch (⇒ tolerance)
- Raw Pitch Accuracy:

$$RPA = \frac{\sum\limits_{\forall n} TP_n}{\mathcal{N}}$$

$$TP_n = \left\{ \begin{array}{ll} 0, & \text{if } |\mathfrak{p}_{\mathrm{GT}}(n) - \hat{\mathfrak{p}}(n)| \geq 0.5 \\ 1, & \text{otherwise} \end{array} \right.$$

■ Raw Chroma Accuracy:

$$RCA = rac{\sum\limits_{orall n} TP_{ ext{chroma},n}}{\mathcal{N}}$$
 $TP_{ ext{chroma},n} = \left\{egin{array}{c} 0, & ext{if} & ext{mod} \left(|\mathfrak{p}_{ ext{GT}}(n) - \hat{\mathfrak{p}}(n)|, 12
ight) \geq 0.5 \\ 1, & ext{otherwise} \end{array}
ight.$ 

- all metrics should be computed in the pitch domain, not the frequency domain
- metrics measure a match between ground truth and predicted pitch (⇒ tolerance)
- Raw Pitch Accuracy:

$$RPA = rac{\sum\limits_{orall n} TP_n}{\mathcal{N}}$$
 $TP_n = \left\{ egin{array}{ll} 0, & ext{if } |\mathfrak{p}_{\mathrm{GT}}(n) - \hat{\mathfrak{p}}(n)| \geq 0.5 \\ 1, & ext{otherwise} \end{array} 
ight.$ 

■ Raw Chroma Accuracy:

$$RCA = rac{\sum\limits_{orall n} TP_{\mathrm{chroma},n}}{\mathcal{N}}$$
  $TP_{\mathrm{chroma},n} = \left\{ egin{array}{l} 0, & \mathrm{if} \mod \left(|\mathfrak{p}_{\mathrm{GT}}(n) - \hat{\mathfrak{p}}(n)|, 12
ight) \geq 0.5 \\ 1, & \mathrm{otherwise} \end{array} 
ight.$ 

- all metrics should be computed in the pitch domain, not the frequency domain
- metrics measure deviation between ground truth and predicted pitch
- MSE, MAE, standard deviation from the ground truth

- all metrics should be computed in the pitch domain, not the frequency domain
- metrics measure deviation between ground truth and predicted pitch
- MSE, MAE, standard deviation from the ground truth

# pitch evaluation result aggregation



- aggregate per datapoint (frame/note)
- aggregate per file



#### potential data problems

- pitch and time quantization
- reliability of ground truth
- time resolution mismatch of ground truth and system

#### metrics

- score pitch match (chroma match)
- measures of deviation

### factor impacting metrics

voicing detection

