

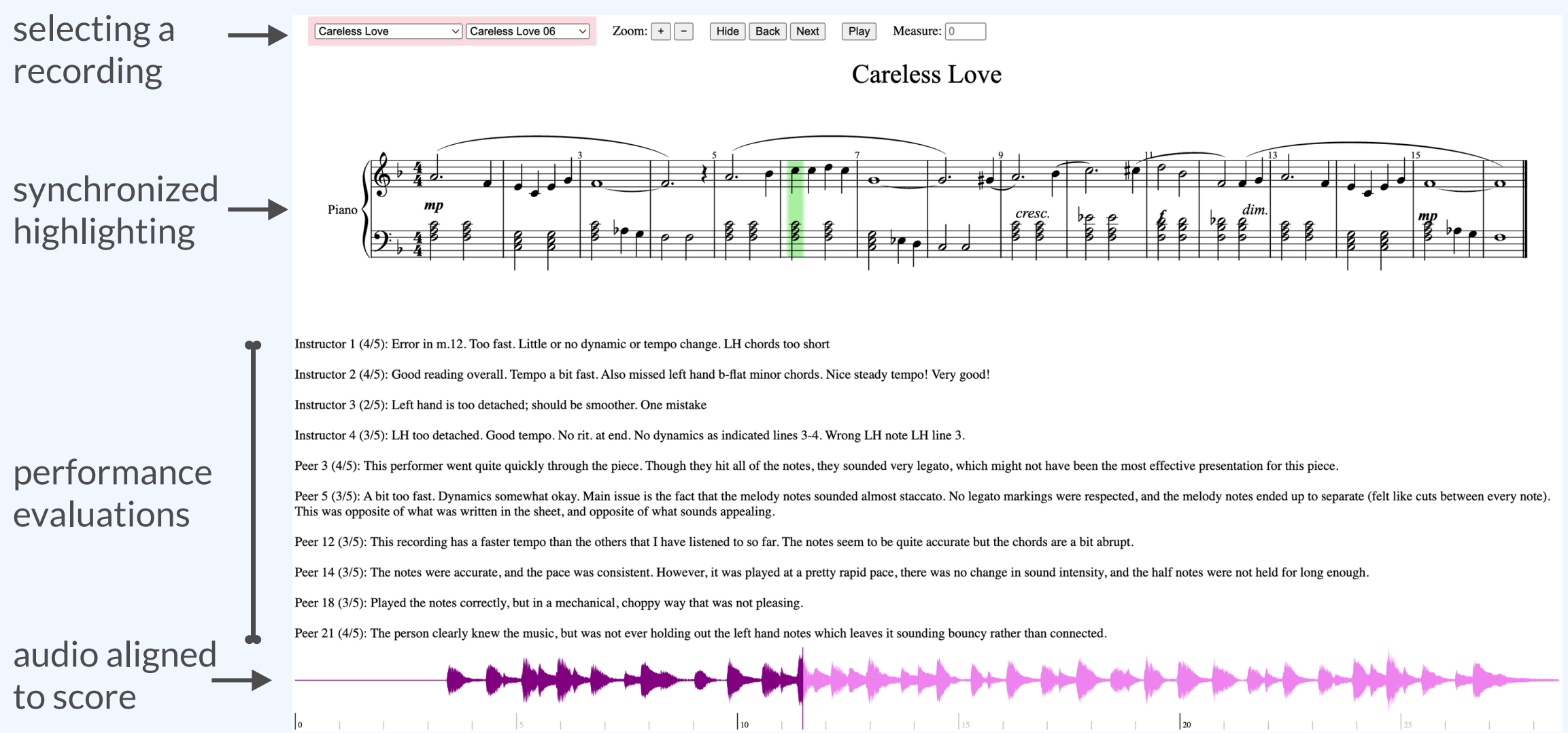
Motivation In building software that can automatically generate feedback to beginner piano players, what kinds of feedback are especially helpful?

Investigation We built and analyzed a dataset to compare how experts and novices assess piano performances and what criteria they rely on.

- Questions**
- Whether experts and novices evaluate performances differently;
 - Whether better players are also better evaluators;
 - What objective indicators can be detected and measured by computers that would reflect comparable evaluation criteria.

How to Access the Dataset

Download dataset published with DOI: 10.5281/zenodo.8392772
Explore dataset on web-based interface: mathcs.richmond.edu/~yjiang/papers/ismir23/



Data Gathering

1. Performance Recordings

[21 Players]

each plays at least one from:

self-identified as:

- 4 beginner
- 8 intermediate
- 9 advanced
- 0 professional

Piece Name	#Measures	#Recordings
Careless Love	16	11
Cielito Lindo	16	6
Lavender's Blue	16	17
Over the Waves	32	11
She Wore a Yellow Ribbon	34	13
The Blues	16	17
The Entertainer	40	8

from J. W. Bastien, *The older beginner piano course*. Kjos West, 1977.

[83 Performance Recordings In Total]

2. Performance Evaluations

[4 Piano Instructors]

each evaluates all 83 performances

[17 Peer Evaluators]

(all from the 21 players)

each evaluates 1/3 of the performances

[803 Evaluations In Total]

each performance evaluated by all 4 instructors, and 5 or 6 randomly chosen peers

Cielito Lindo 03:

Rate this performance: *

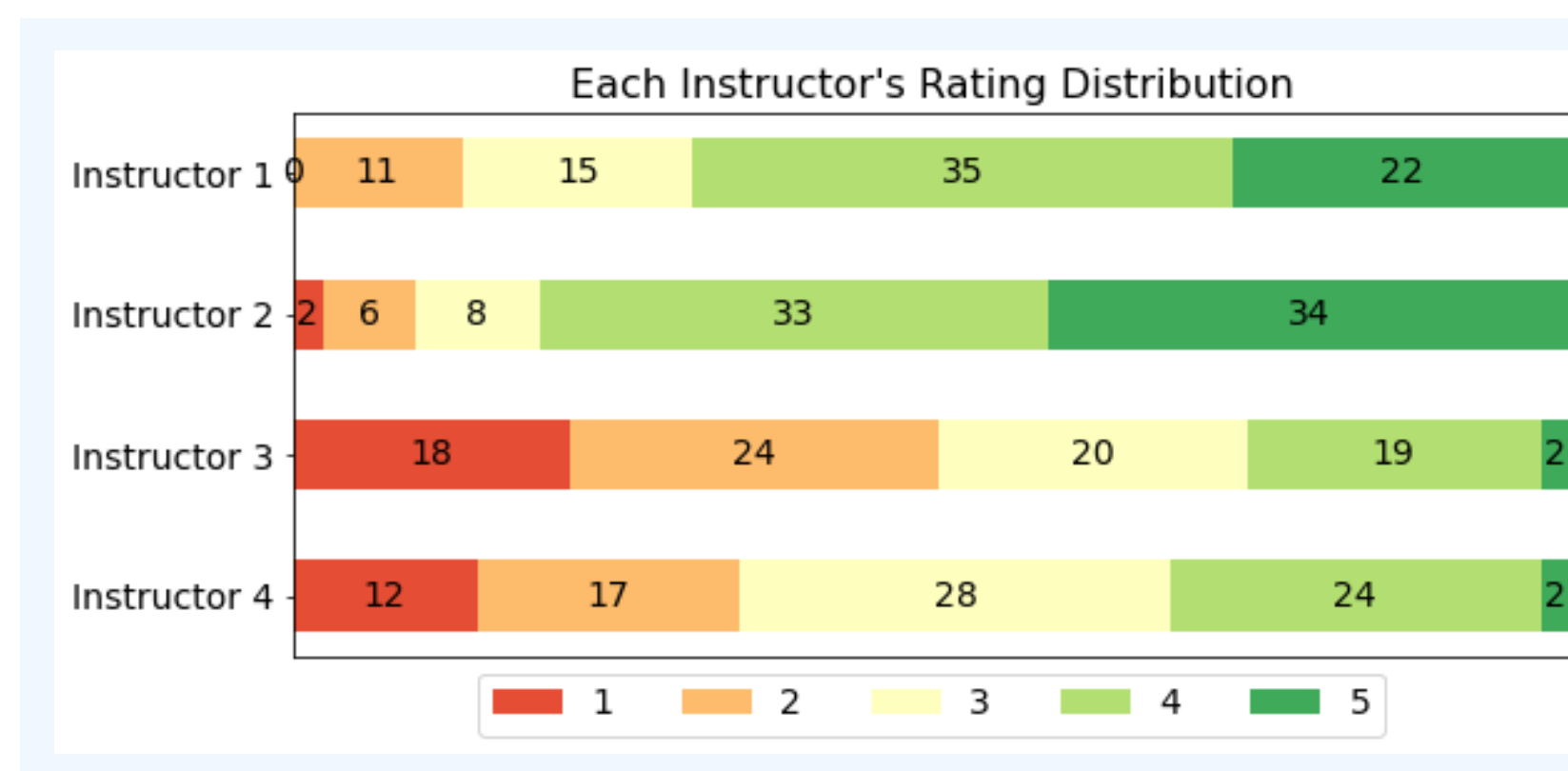
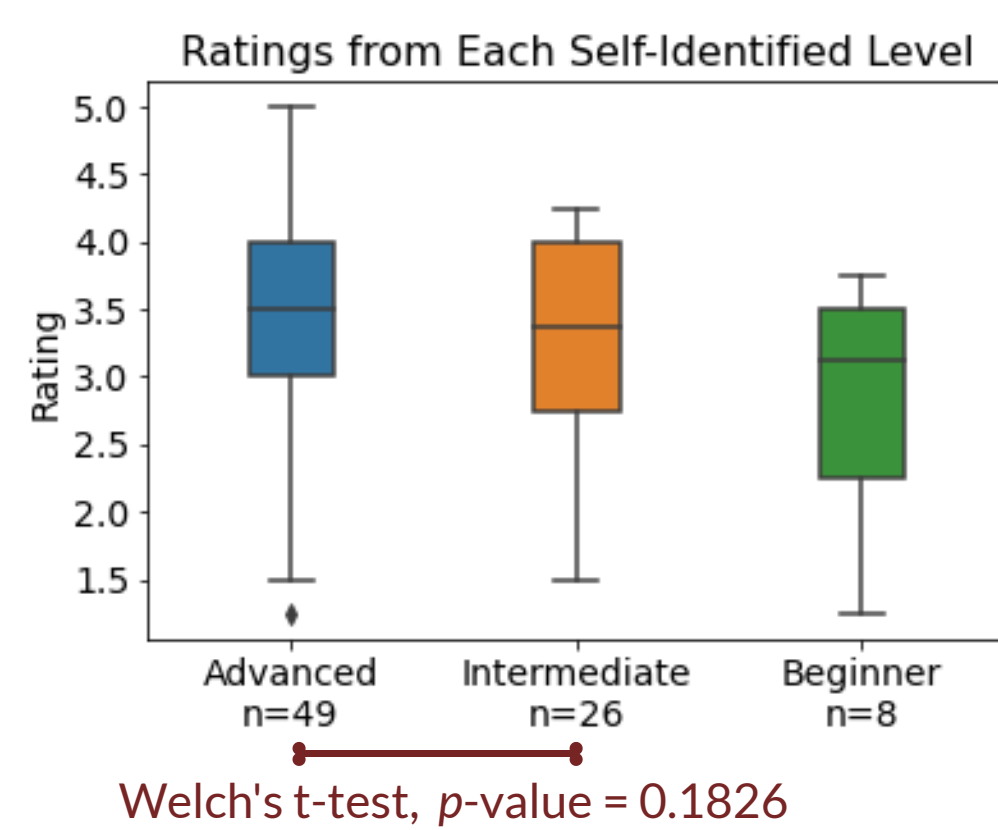
Poor Excellent

1 2 3 4 5

Please elaborate on your rating. *

Generally very accurate. RH should be much more legato

Quantitative Analysis



	I2	I3	I4
I1	.595	.514	.521
I2		.508	.563
I3			.806

Kendall's τ correlation
 p -value ≈ 0

Are better players also better evaluators?

r : the average rating received by a player from the instructors

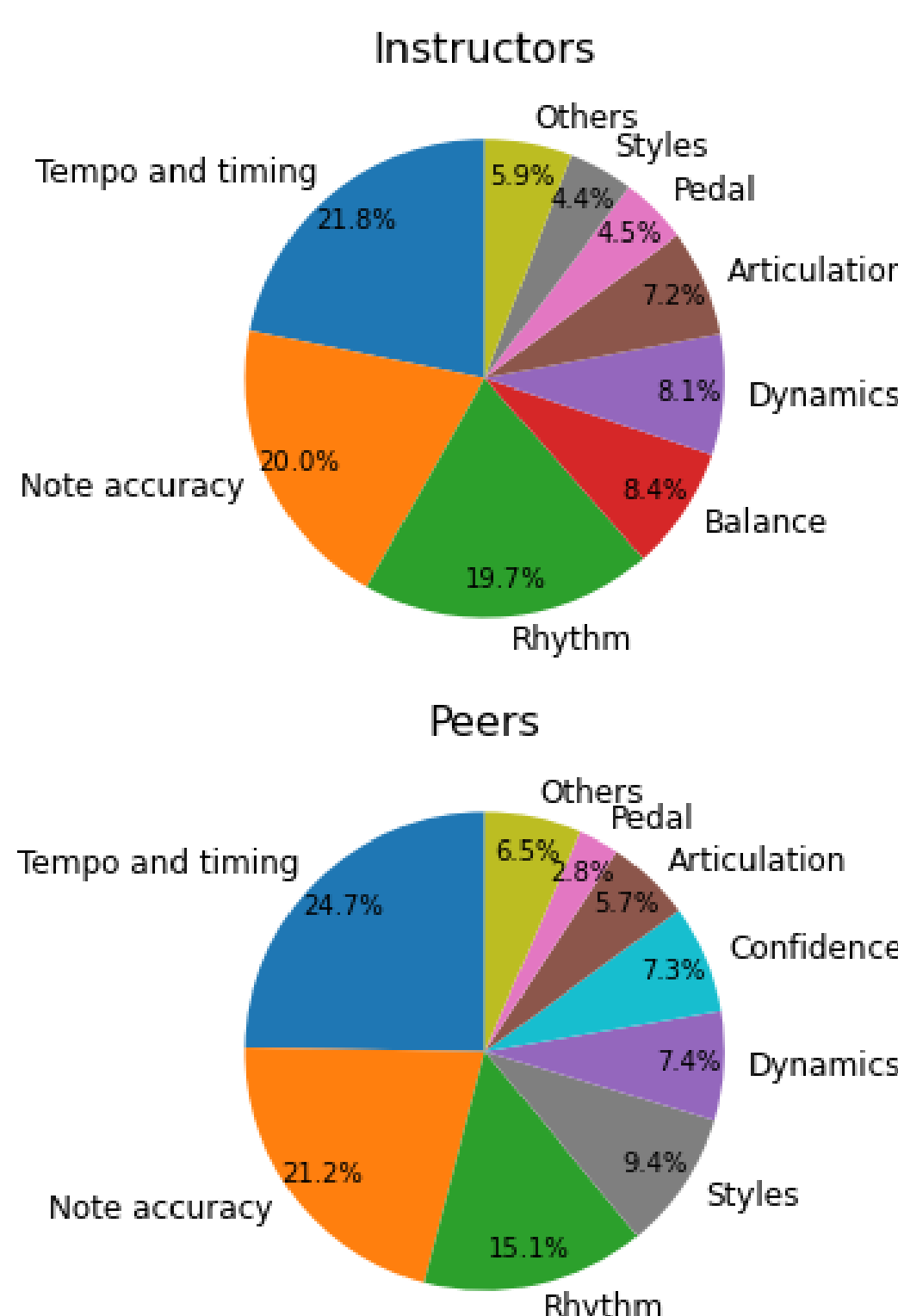
k : the Kendall's τ correlation between ratings provided by a player and by the instructors.

Spearman's ρ between r and k is:
 $\rho = 0.152$, with p -value = 0.56

Qualitative Analysis

Evaluation criteria emerged from content analysis:

Tempo and timing	Dynamics	Styles
<ul style="list-style-type: none">- inaccurate tempo- good tempo- tempo steadiness- tempo contrast- ritardando- rubato- pause	<ul style="list-style-type: none">- accurate dynamics- inaccurate dynamics- dynamic contrast- dynamic shaping	<ul style="list-style-type: none">- smooth- heavy- light- abrupt- crisp- character- lively- flow- lyrical- mechanical- style- bland- with emotion
Note accuracy	Balance (between hands)	Rhythm
<ul style="list-style-type: none">- correct note- wrong note- missed note- wrong octave	<ul style="list-style-type: none">- balance in general- left hand loudness- right hand loudness	<ul style="list-style-type: none">- correct rhythm- incorrect rhythm- Notes too short<ul style="list-style-type: none">- shortened note- left hand too short- Notes too long<ul style="list-style-type: none">- note too long
Phrasings	Articulation	Confidence
<ul style="list-style-type: none">- phrasing- melodic shaping	<ul style="list-style-type: none">- articulation in general- legato- staccato- accent	<ul style="list-style-type: none">- confident or hesitant
Pedal	Note connection	
<ul style="list-style-type: none">- inaccurate pedal- good pedal	<ul style="list-style-type: none">- choppy- connectedness	



Do instructors and players evaluate performances differently?

Abstract Criteria vs. Piano Technique Criteria

Styles	Balance	Dynamics
Confidence	Articulation	Pedal

Selected Criterion Usage Among Four Groups of Evaluators

