Using artificial intelligence in string performance and teaching

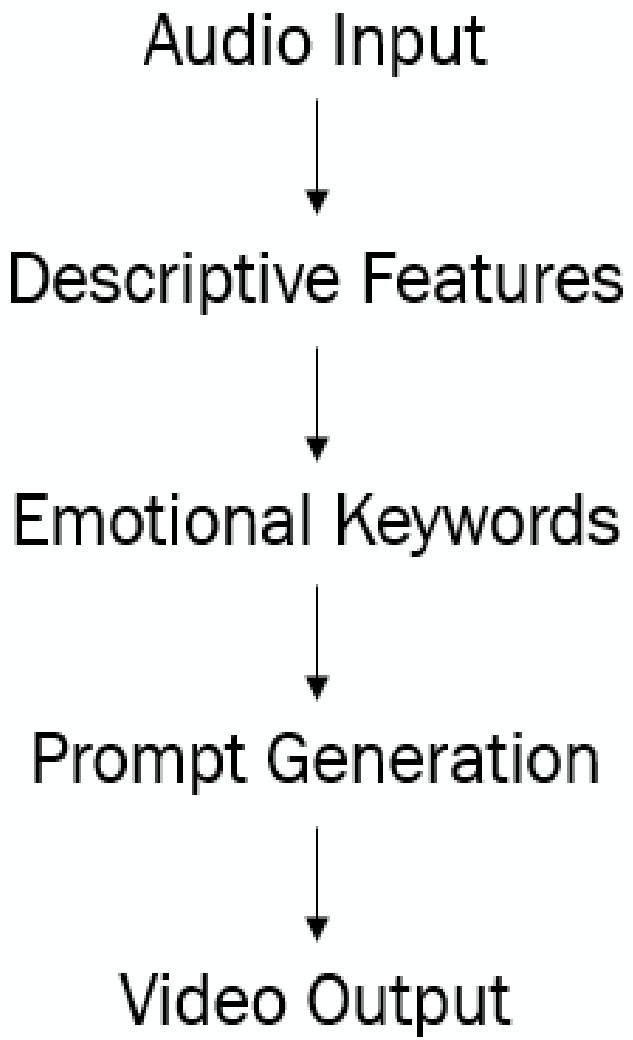
Dr. Kristen Yeon-Ji Yun
Cellist
Clinical Associate Professor
Purdue University

PURDUE UNIVERSITY | Department of Music
PURDUE UNIVERSITY | Elmore Family School of Electrical and Computer Engineering
VIP @Purdue
CREATE!
SURF SUMMER UNDERGRADUATE RESEARCH FELLOWSHIP

1. MUS2VID

- Real-Time Image Generation by AI in concert
- AI analyzes musical features and the trained AI models will generate relevant images during performance.

Performance



1. Music is analyzed and classified into multiple dimensions, such as instrumentation, emotion, tempo, number of musicians, pitch range, harmony, dynamic, etc.

2. The feature matrix becomes input to two machine learning models: a classification neural network and a support vector machine.

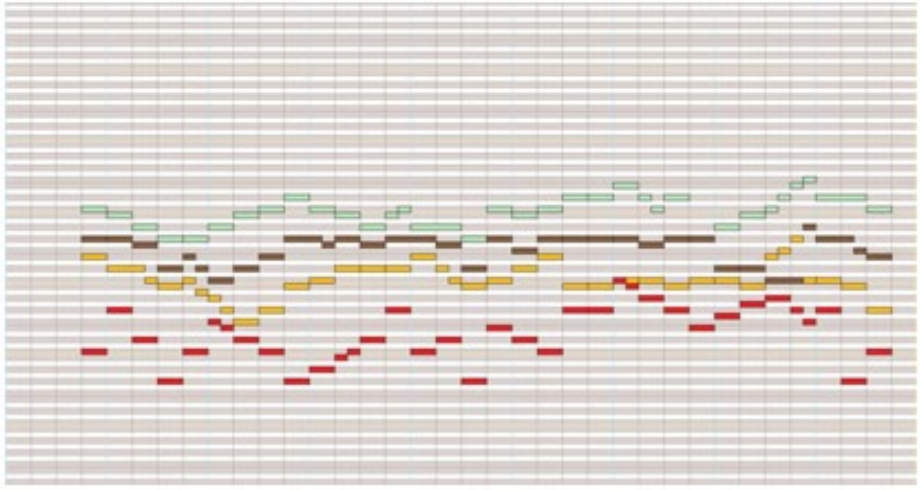
3. These models predict the genre of the input audio and theemotions, respectively. These predictions form the bas is for text prompts that describe the music.

4. The descriptions enter generative machine models (e.g., Stable Diffusion) in order to create visual representations.

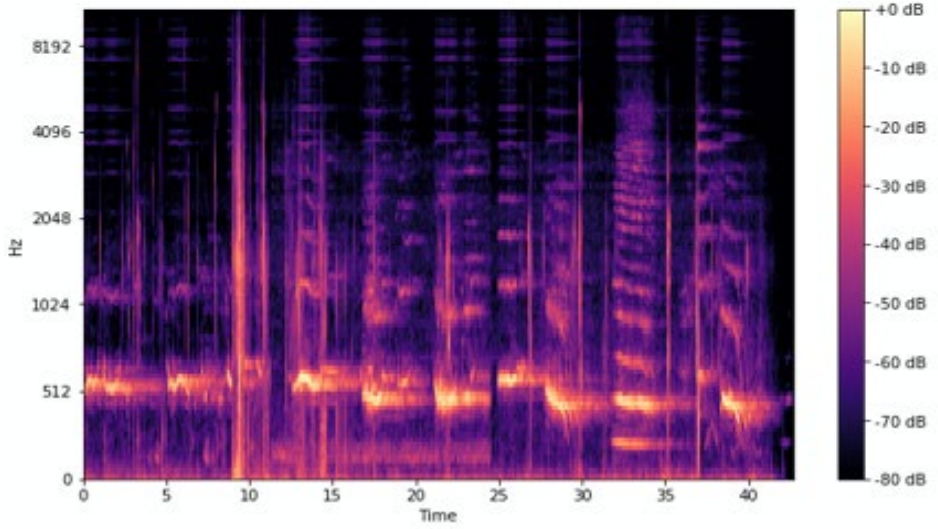
5. The visual representations are updated as the music is performed.

Feature Extraction

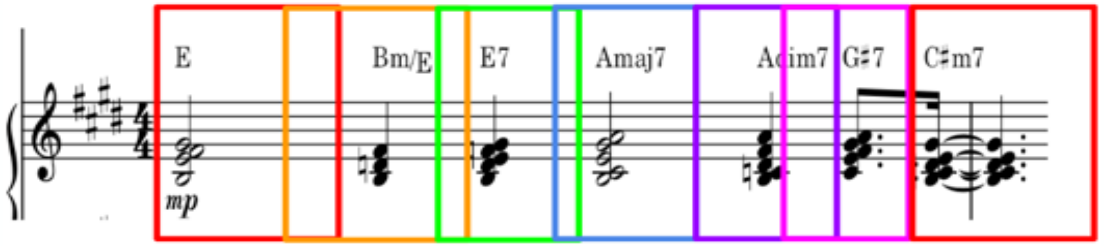
- Piano rolls represent the timing, duration, and pitch of all notes played throughout a song
- This format doesn't capture timbre or dynamics



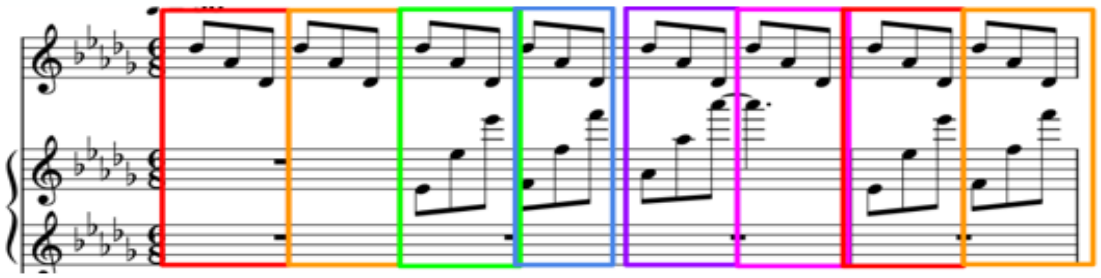
- Spectrograms represent music visually with frequency, time, and intensity.
- Captures a wider range of features but less precise with notes



Feature Extraction and Chord Detection



- Notes played within a window of time are assigned a prominence based on duration, length, octave etc.
- Most prominent notes are used to predict the chord being played
- Popular chords throughout a song can predict the key

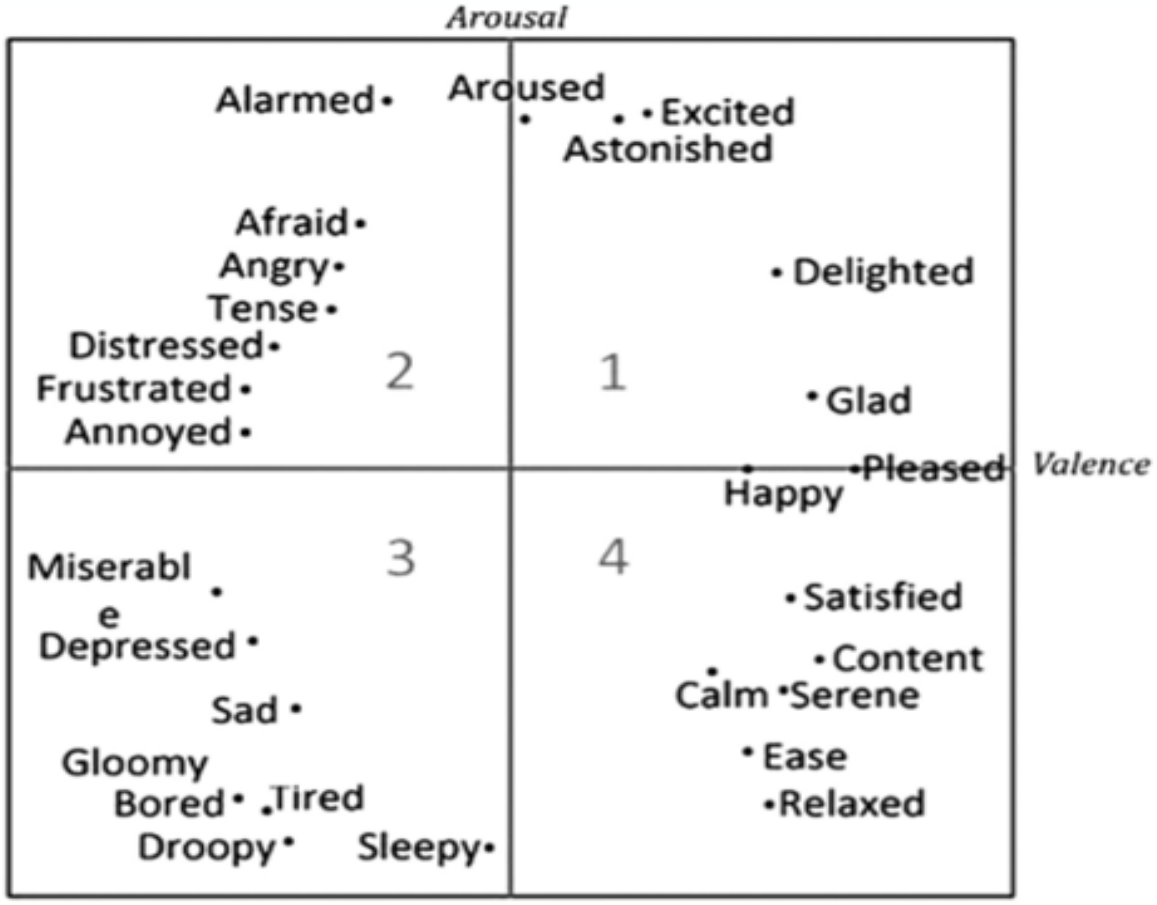


Music and Emotion

Emotion	Tempo	Volume/ Dynamics	Rhythm	Pitch Range	Harmony	Common Features
Happiness	Fast, consistent	Medium-high, small variability	Regular	High, wide	Consonant	Perfect 4th and 5th, staccato, trills
Sadness	Slow, consistent	Low, small variability	Firm	Narrow	Dissonant	Ritardando, legato, minor 2nd
Anger	Fast, consistent	High, small variability	Complex	High, narrow	Dissonant	Staccato, accents on dissonant notes
Fear/ Stress	Fast, variable	Large variability	Irregular, varied	Wide	Dissonant	Staccato, pauses,
Tenderness/ Serenity	Slow	Medium-low, low variability	Little variability	Narrow	Consonant	Legato, accents on consonant notes

- Some features are often connected to specific emotions (i.e., major keys often convey positive emotions)
- Musical pieces establish expectations
- Reinforcing or defying expectations lowers/increases tension (arousal/ intensity of the emotion)

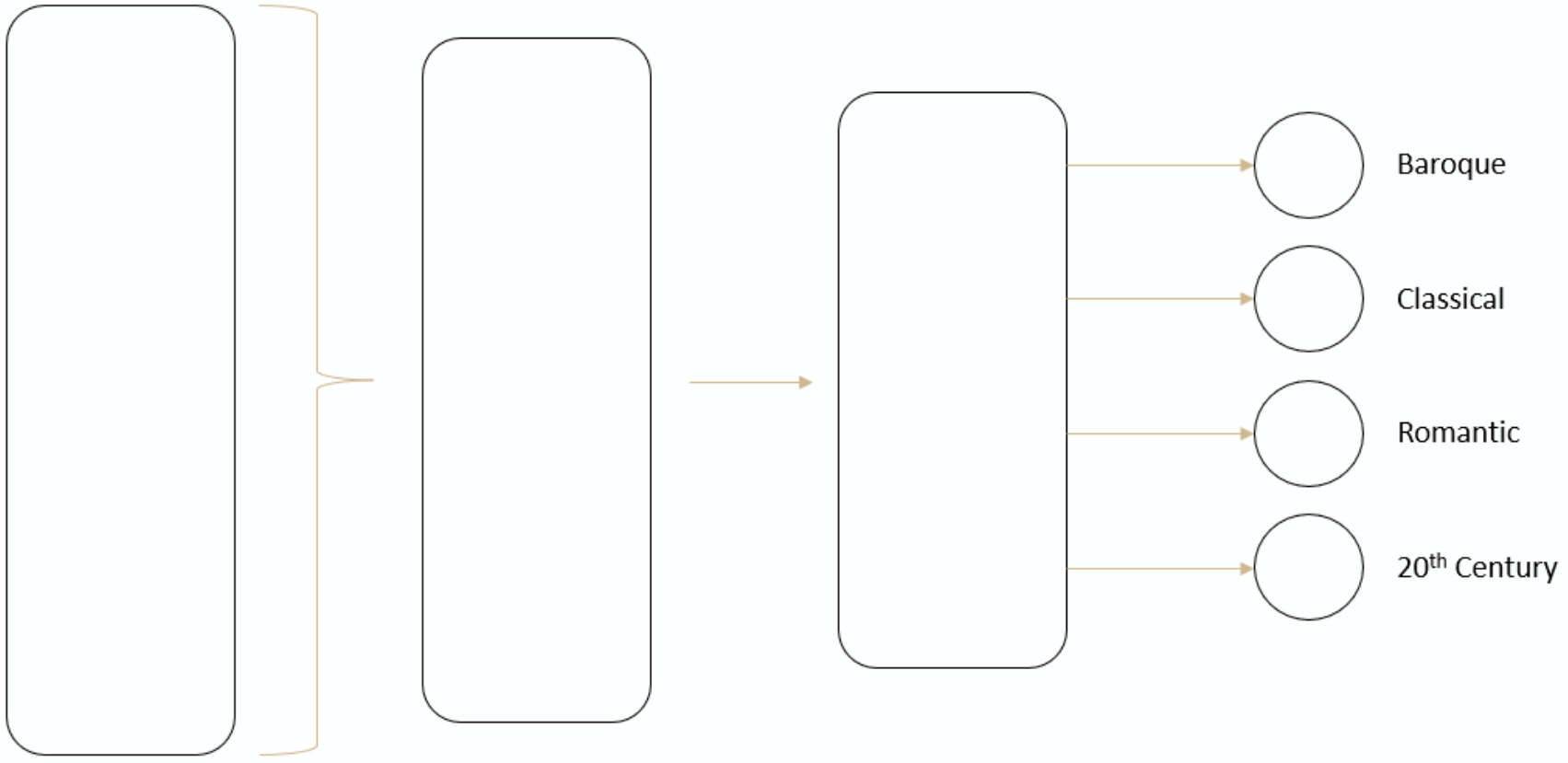
Emotional Prediction



Russel's circumplex model of emotion [1]

- Emotions are described on axes of valence and arousal, with values from 1 to 9 for each
- We predict valence and arousal for new audio clips using a machine learning model trained on a dataset of songs with emotional labels

Genre Classification



Emotional language to Art

- Warm colors
- Bright, vibrant lighting

Excitement, happiness, energy

- Pastel shades, light blues and greens
- Soft, diffuse lighting

Calmness, serenity, relaxation

- Intense, saturated colors
- Dim or harsh lighting

Anger, fear, intensity

- Cool, muted colors
- Soft or dim lighting

Sadness, tranquility, melancholy

Stable Diffusion and Prompt Engineering

"Mansion with waterfall in the woods"

"A beautiful mansion beside a waterfall in the woods, by josef thoma, matte painting, trending on artstation HQ" + negative prompt

Prompt Generation

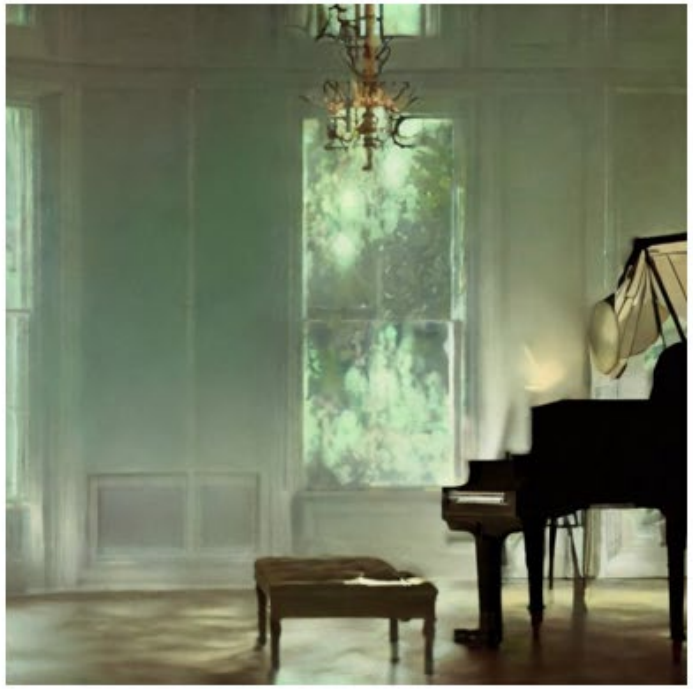
[Musical Period] [Artist(s)] [Lighting] [Colors] [Perspective] [Style]

Genre = “Romantic” Emotion = “Relaxed”

Romantic classical music in the style of Claude Monet,

soft lighting, low intensity, neutral and earthy tones,

Wide shot, matte painting } Randomly Selected



Concert Video

Future Work

- Improving accuracy of models with larger datasets and new features
- Using frame interpolation to increase animation quality
- Direct connections between music and visuals without creating text prompts in between
    - Training a model like Stable Diffusion on a set of images with audio “captions”

Broad Impact

- As audiences can enjoy informatic and artistic images throughout concerts, this technology can help outreach to broader audiences, and improve presentation, education, and promotion of any concert program.
- Younger generations can approach to Classical music better in concert with this technology.
- It can also be used to help hearing impair people to appreciate music through visual images.

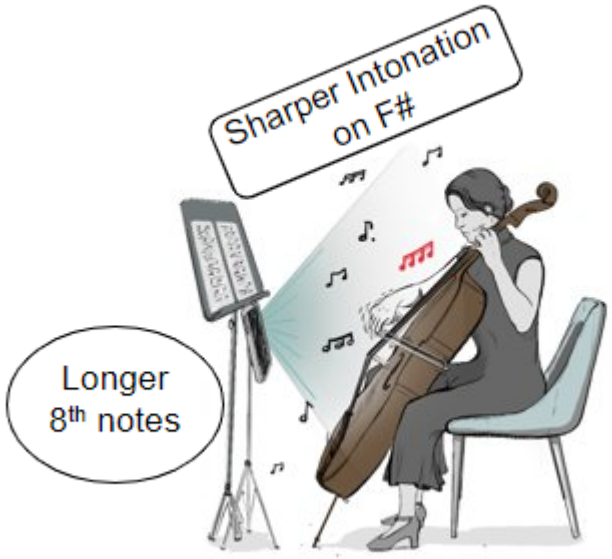
2. Evaluator and Companion

Two tools which can help musicians’ practice in solo and/or ensemble

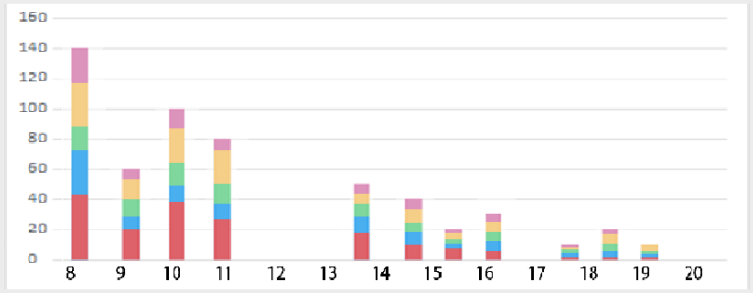
A grant received from National Science Foundation (Artificial Intelligence Technology for Future Music Performers. Award number: 2326198) announced last month https://www.nsf.gov/awardsearch/showAward?AWD_ID=2326198

What are the everyday challenges for musicians?

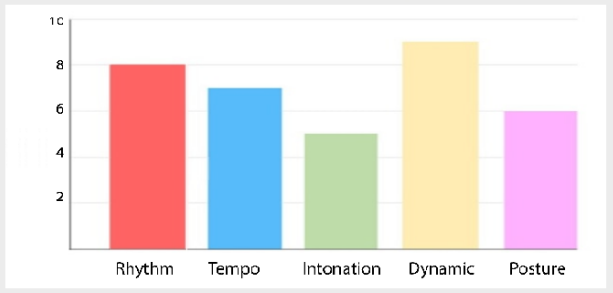
1) Evaluator



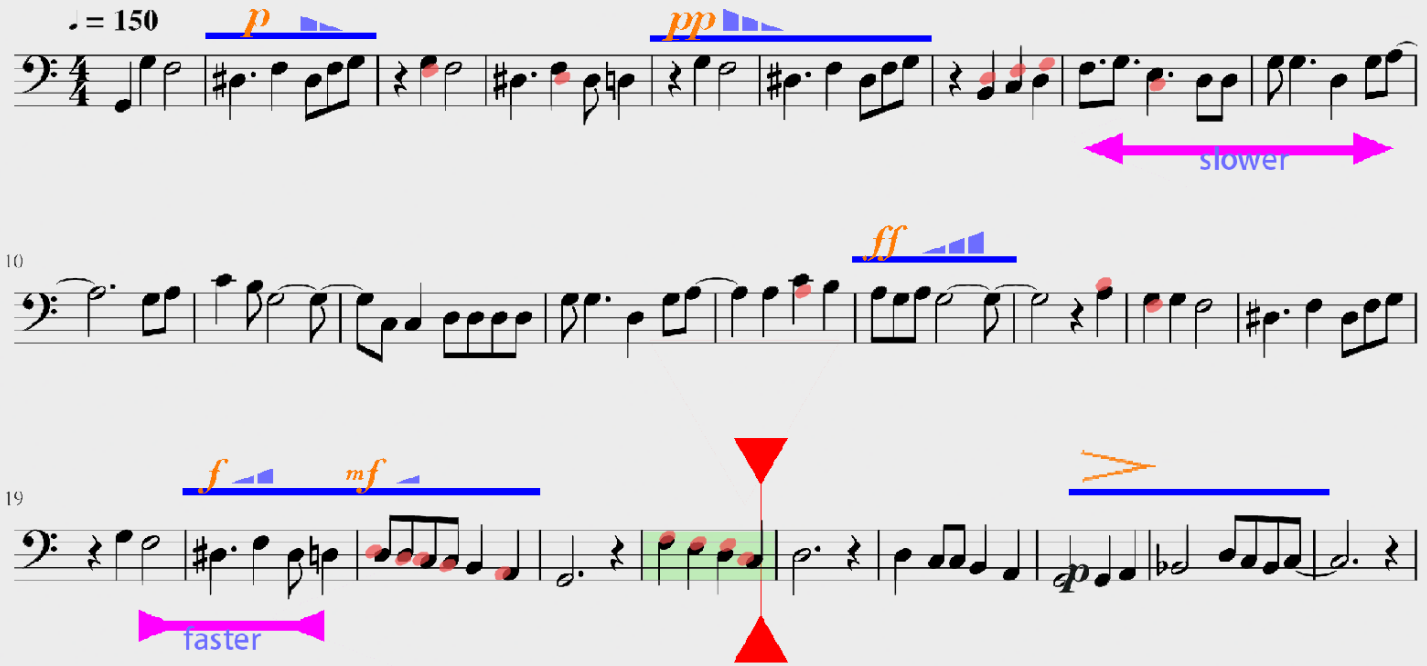
- aims to improve individual practice and performance
- analyzes a musician's sound and compares it to digitized music scores to detect deviations in intonation, rhythm, and dynamics
- computer vision is used to detect incorrect postures.

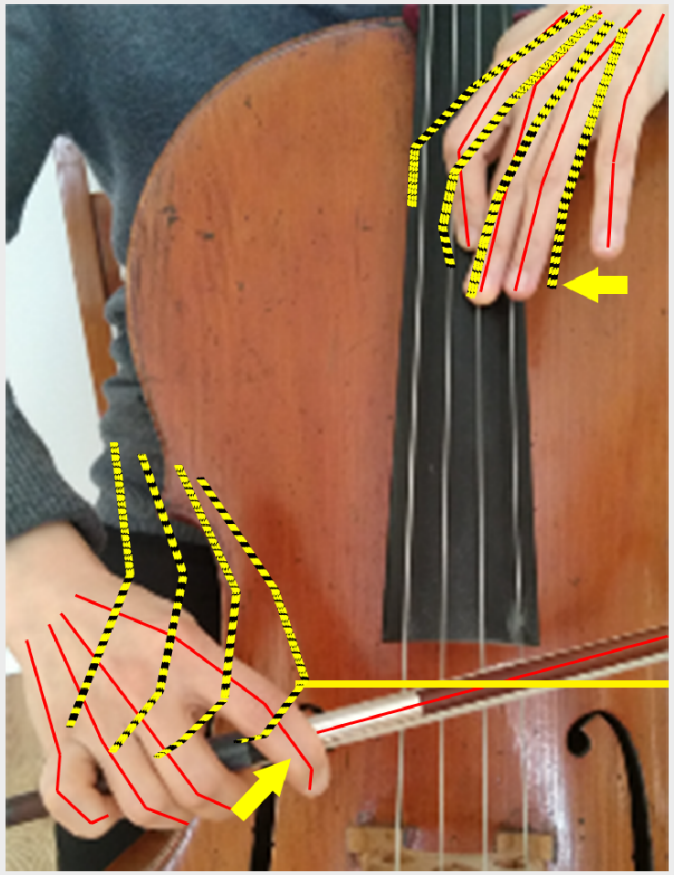


Practice History



Practice Errors





Wrong position Correct position

Posture Analysis

2) Companion

- plays the part of one or several instruments to replace absent musicians
- It can match tempo, and style of the human musicians
- also responds in real-time to verbal instructions.

VIP (Vertically Integrated Project)

- Undergraduate research team with various background
- Some grad-students will join from next semester.

Technology

- computer vision
- natural language processing
- audio analysis
- transformer

Evaluation

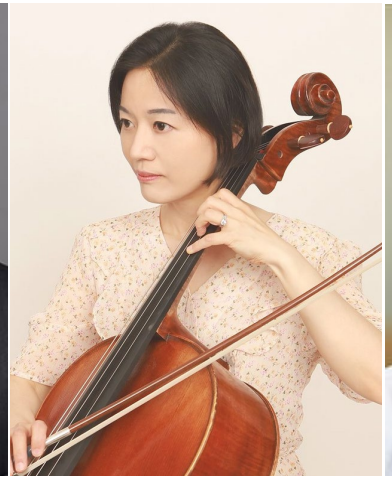
- user studies
- surveys
- focus groups
- longitudinal deployments.

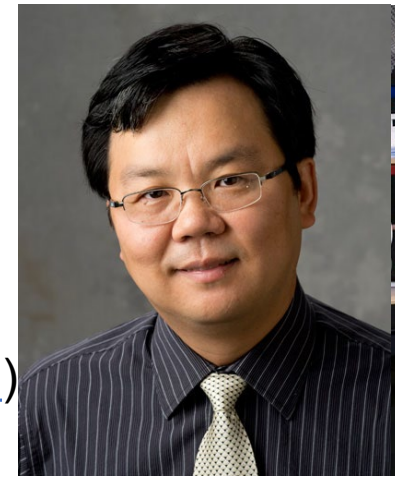
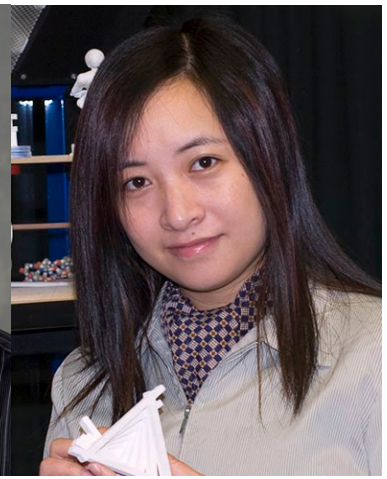
Who are involved in this research?

PI: Kristen Yeon Ji Yun (yun98@purdue.edu)
Music

- Cheryl Zhen Yu Qian (qianz@purdue.edu)
  Art & Design
- Victor Yingjie Chen (victorchen@purdue.edu)
  Computer Graphics Technology
- Yung-Hsiang Lu (yunglu@purdue.edu),
  Electrical and Computer Engineering
- Mohammad Saifur Rahman (mrahman@purdu.edu)
  School of Management

- Ka-Wai Yu (ka-wai.yu@utahtech.edu)
  Music at Utah Tech University


Department of Music | Elmore Family School of Electrical and Computer Engineering | @Purdue | CREATE! | SURF SUMMER UNDERGRADUATE RESEARCH FELLOWSHIP

References

- [1] R. Panda, R. Malheiro, and R. P. Paiva, “Novel Audio Features for Music Emotion Recognition,” *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 614–626, Oct. 2020, doi: 10.1109/TAFFC.2018.2820691.

Survey for musicians