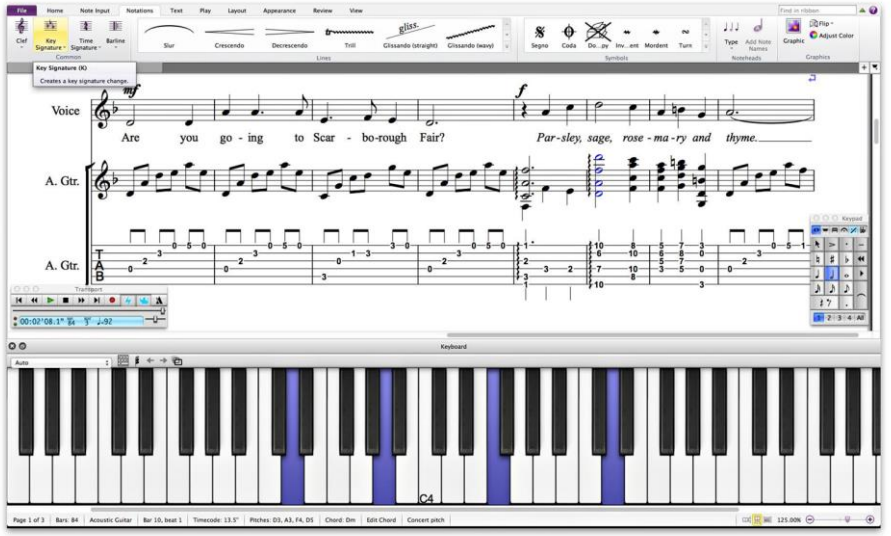
# Interactive Music Systems

Zhiyao Duan
Associate Professor of ECE and CS
University of Rochester

Presentation at WiSSAP 2023, IIT Kanpur, December 18-21, 2023

# Music Making Tools




# Music Making Tools


Credit: Alasabyss/Getty Images; Kelso Harper/Scientific American

https://www.scientificamerican.com/podcast/episode/artificial-intelligence-helped-make-the-coolest-song-youve-heard-this-week/ (March, 2023)

# The world's first AI-composed music album is here, and it sounds amazing

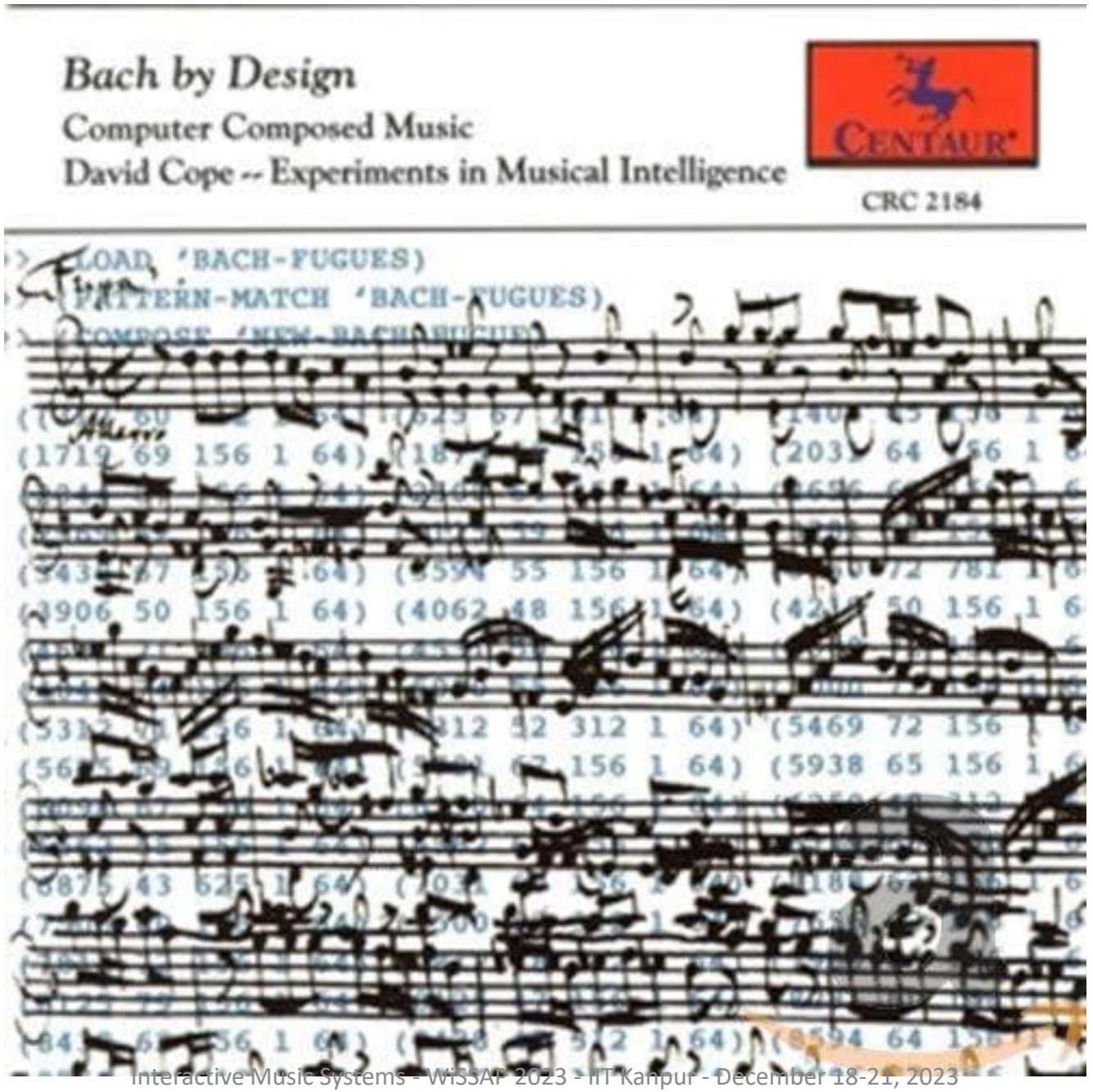
CATCH TEAM | Updated on: 22 August 2017, 19:36 IST

Follo Follow @catchnews

Taryn Southern
Electronic Dance Music



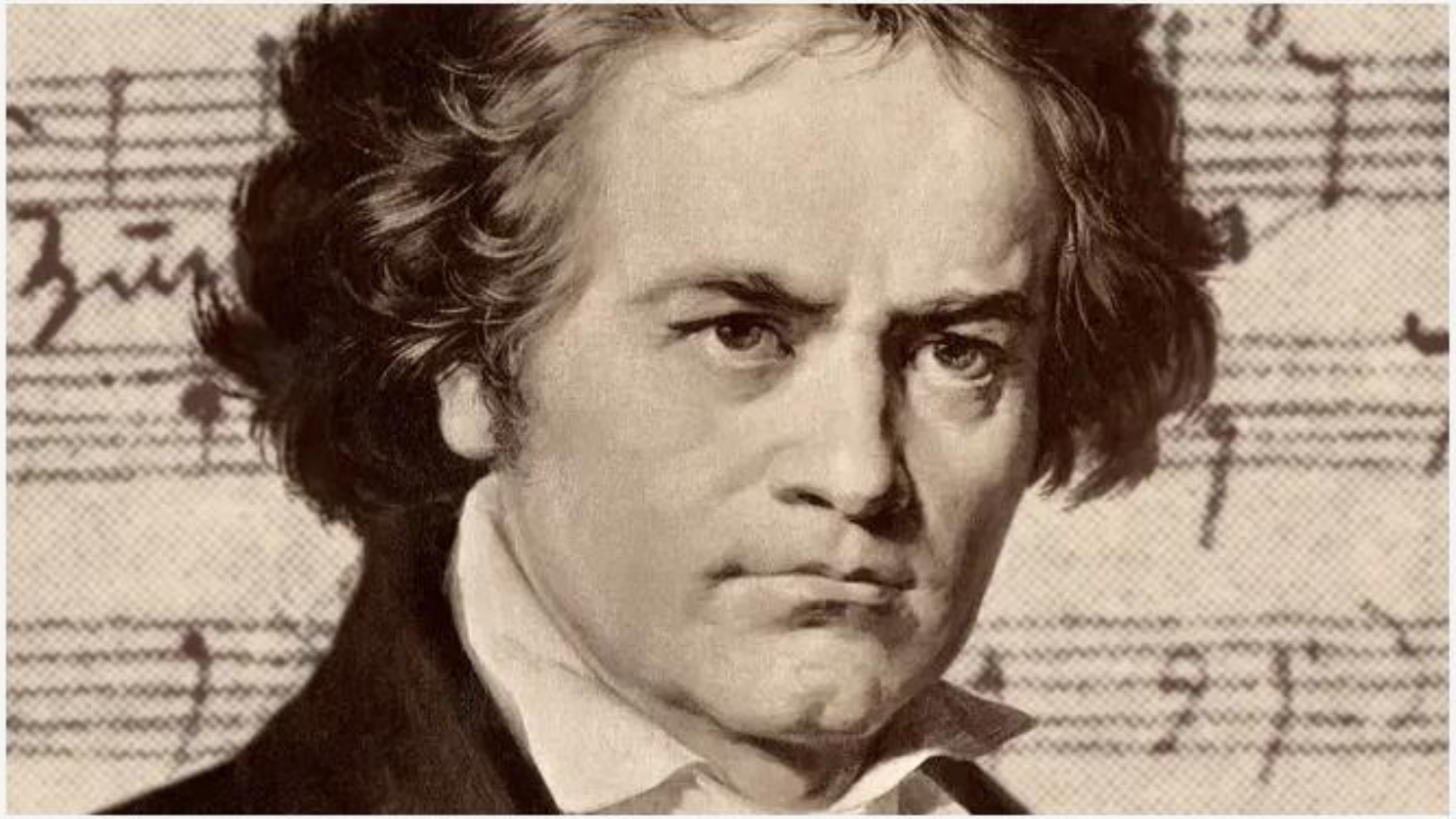
https://www.catchnews.com/entertainment-news/the-first-ai-composed-music-album-is-here-and-it-sounds-amazing-78436.html




David Cope's Experiments in Musical Intelligence (EMI)
"Bach by Design"
1993

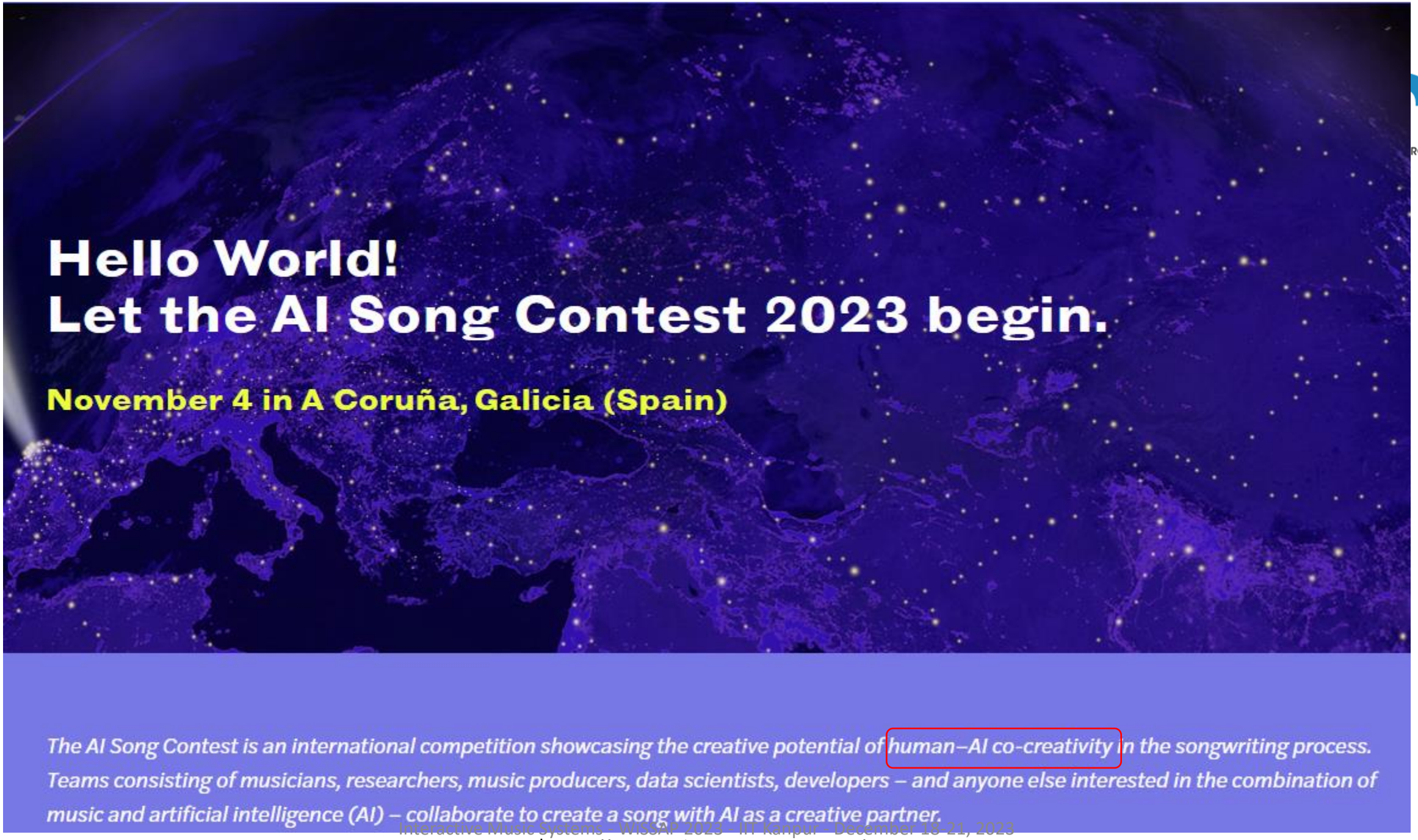
# Beethoven's unfinished Tenth Symphony completed by artificial intelligence

28 September 2021, 14:44 | Updated: 28 September 2021, 16:10



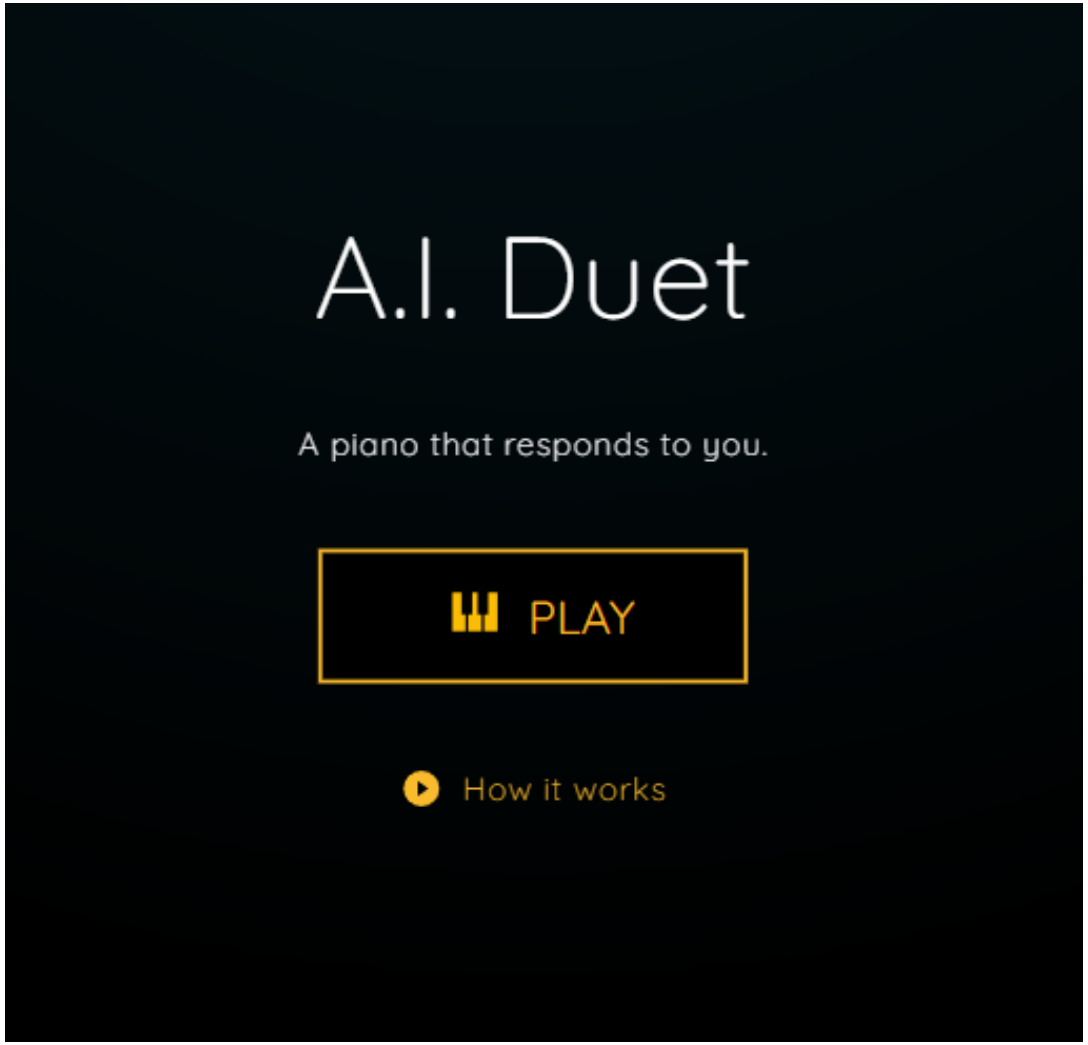
Beethoven's Tenth Symphony completed by AI. *Picture: Alamy*

https://www.classicfm.com/composers/beethoven/unfinished-tenth-symphony-completed-by-artificial-intelligence/






https://www.aisongcontest.com/

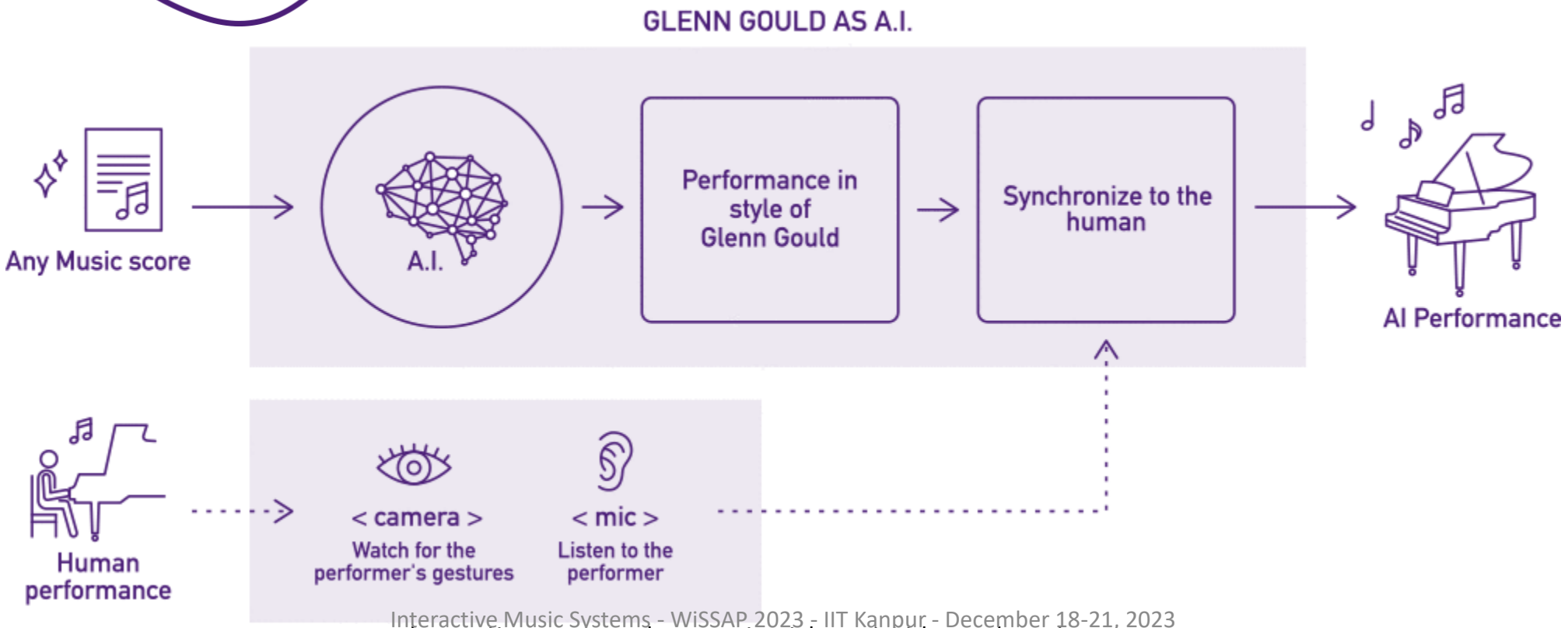



https://experiments.withgoogle.com/ai/ai-duet/view/

7 SEP. 2019

GLENN GOULD AS A.I.

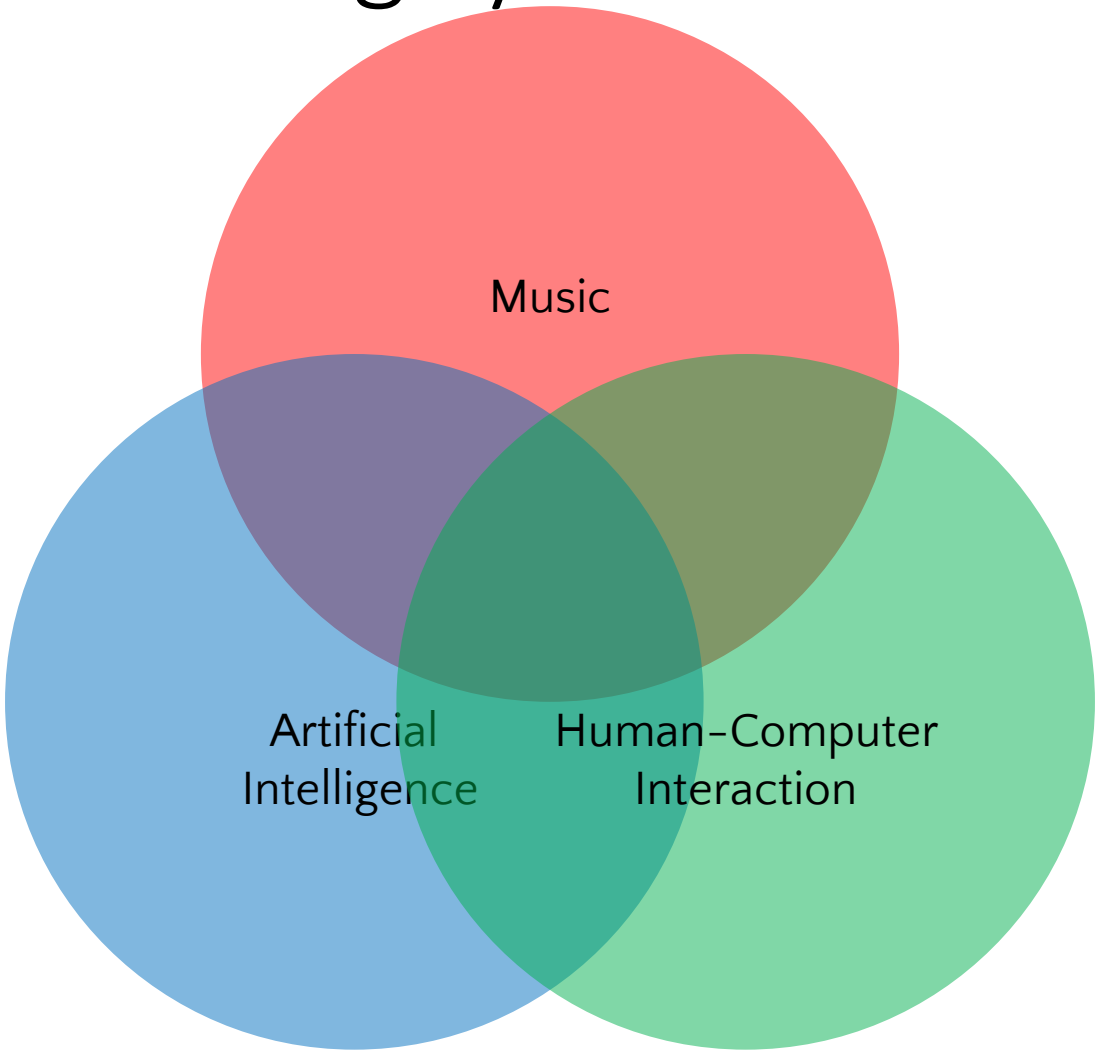
# YAMAHA's Music with AI Project






https://www.yamaha.com/en/about/ai/dear_glenn/

# Why is building computer-assisted music making systems interesting?




- Interdisciplinary
- It integrates research from multiple subareas of music information retrieval
  - Music analysis
    - Audio
    - Symbolic
    - Visual
  - Music generation
  - Music audio/visual/gesture synthesis
  - Interaction design and user Studies

# Why are computer-assisted music making systems useful?


**Music Creation**

- Frees users from tedious work
- Lowers cost
- Improves accessibility

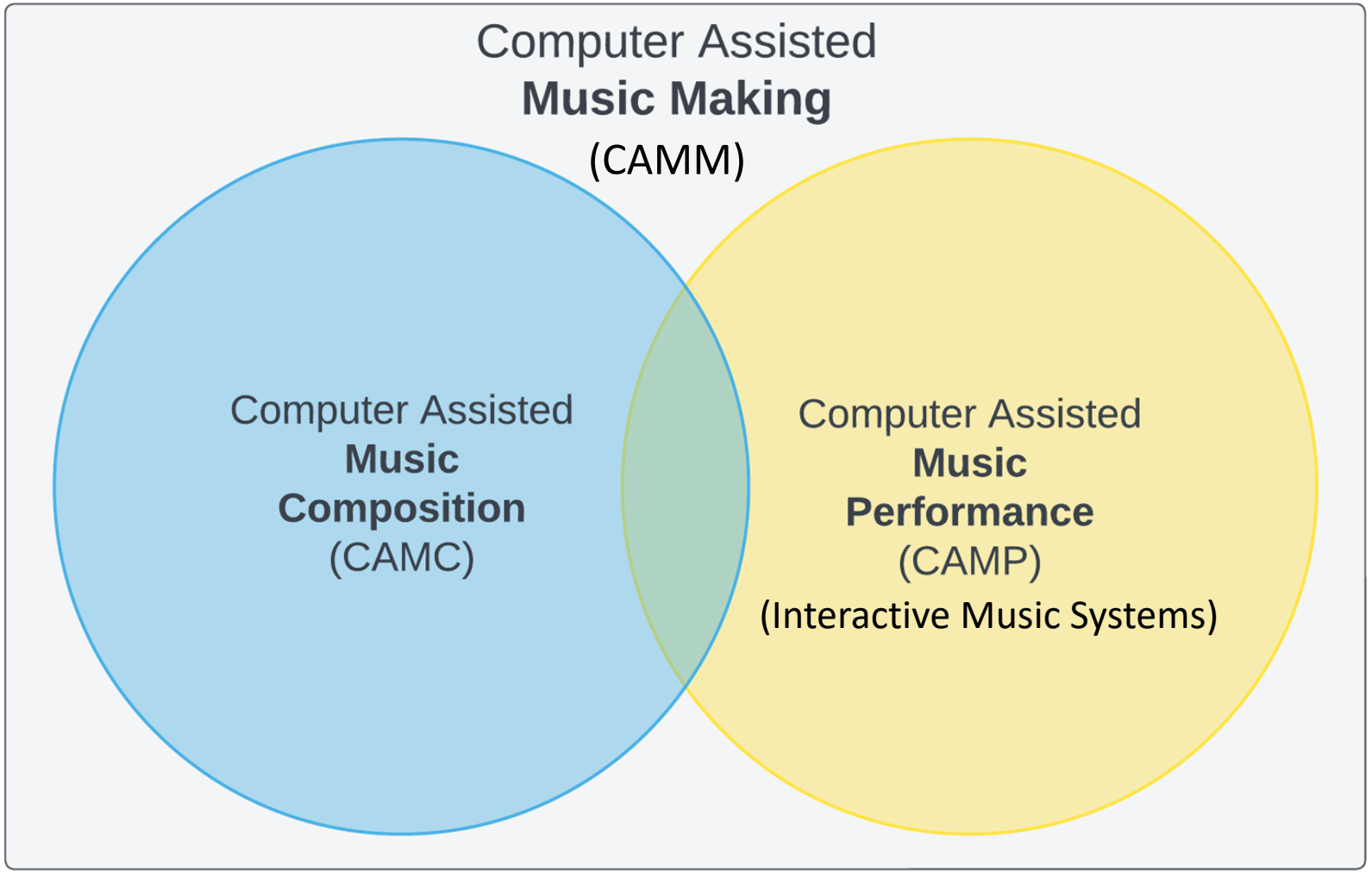
**Music Education**

- Composition
- Arrangement
- Improvisation
- Ensemble performance

**Music Entertainment**

- Novel content
- Various formats
- Anytime/anywhere

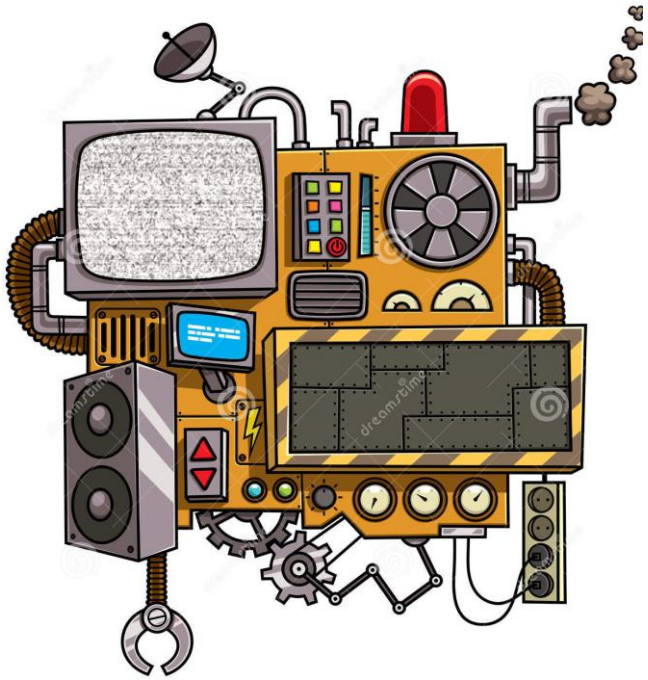
# Two Categories



# Interactive Music Systems

Exchange of musical information

through some media (e.g., audio, visual, touch, gesture, brain signals)



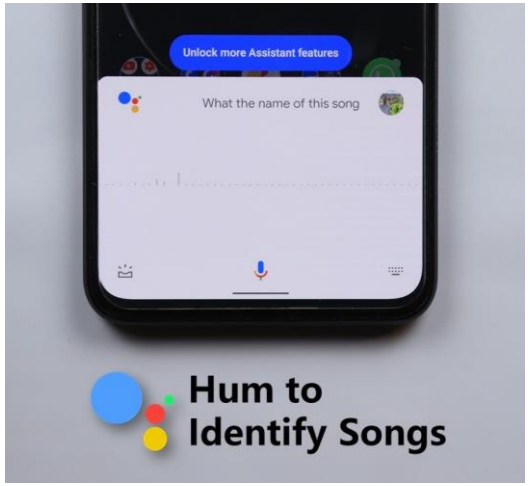
Human

System (or instrument, machine, agent, robot)

# Are they interactive music systems?

Little interactivity

No intelligence




Not for fun

# Interaction Is the Primary Goal


"One does not 'use' an instrument to accomplish some ultimate goal: one plays it, and often that is the only goal."

---- McDermott, J., Gifford, T., Bouwer, A., & Wagy, M. (2013a). Should music interaction be easy? In S. Holland, K. Wilkie, P. Mulholland, & A. Seago (Eds.), Music and human computer interaction (pp. 29–48). London: Springer.

# Automatic Music Accompaniment Systems

Plays solo and leads

Receives limited cues for coordination

Audio and (visual)

Audio

Plays back pre-recorded accompaniment

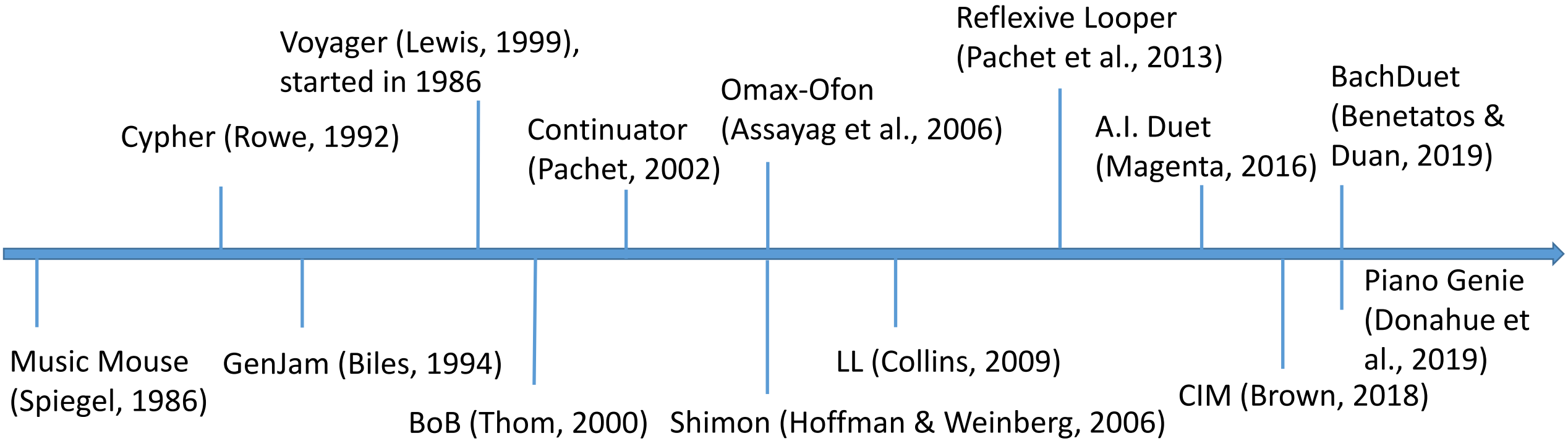
Follows human

Limited understanding of human performance

Limited expressiveness

Piano Tutor (Dannenberg et al., 1993)
Music Plus One (Raphael, 1999)
Antescofo (Cont, 2008)
Eurydice (Nakamura et al., 2015)
Humanoid Robot (Xia et al., 2016)

# Interactive Music Systems
## Beyond Automatic Accompaniment



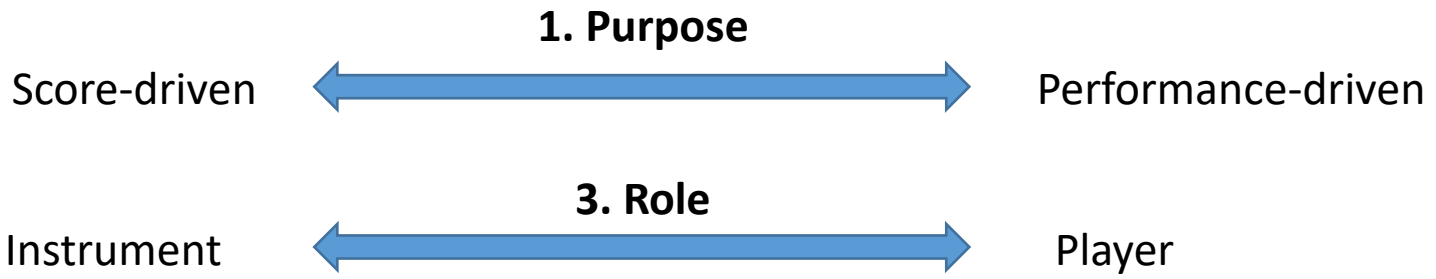

Kivanc Tatar & Philippe Pasquier, Musical agent: A typology and state of the art towards musical metacreation, *Journal of New Music Research*, 2019.

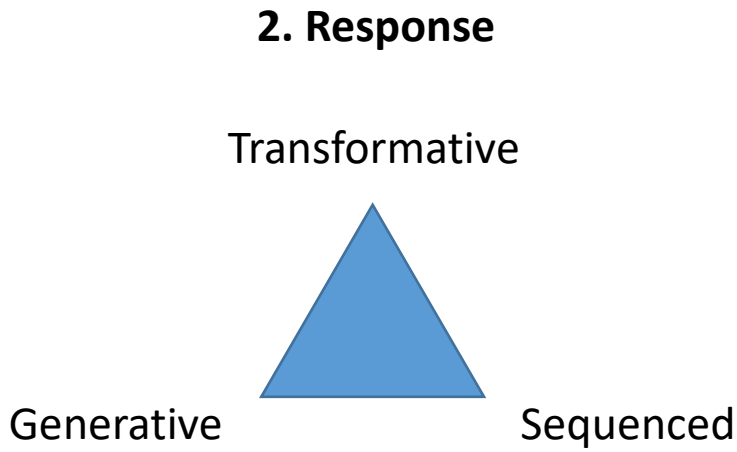
Toby Gifford, et al., Computational systems for music improvisation, *Digital Creativity*, 2018.

# Categorization

- Three dimensions classifying interactive music systems

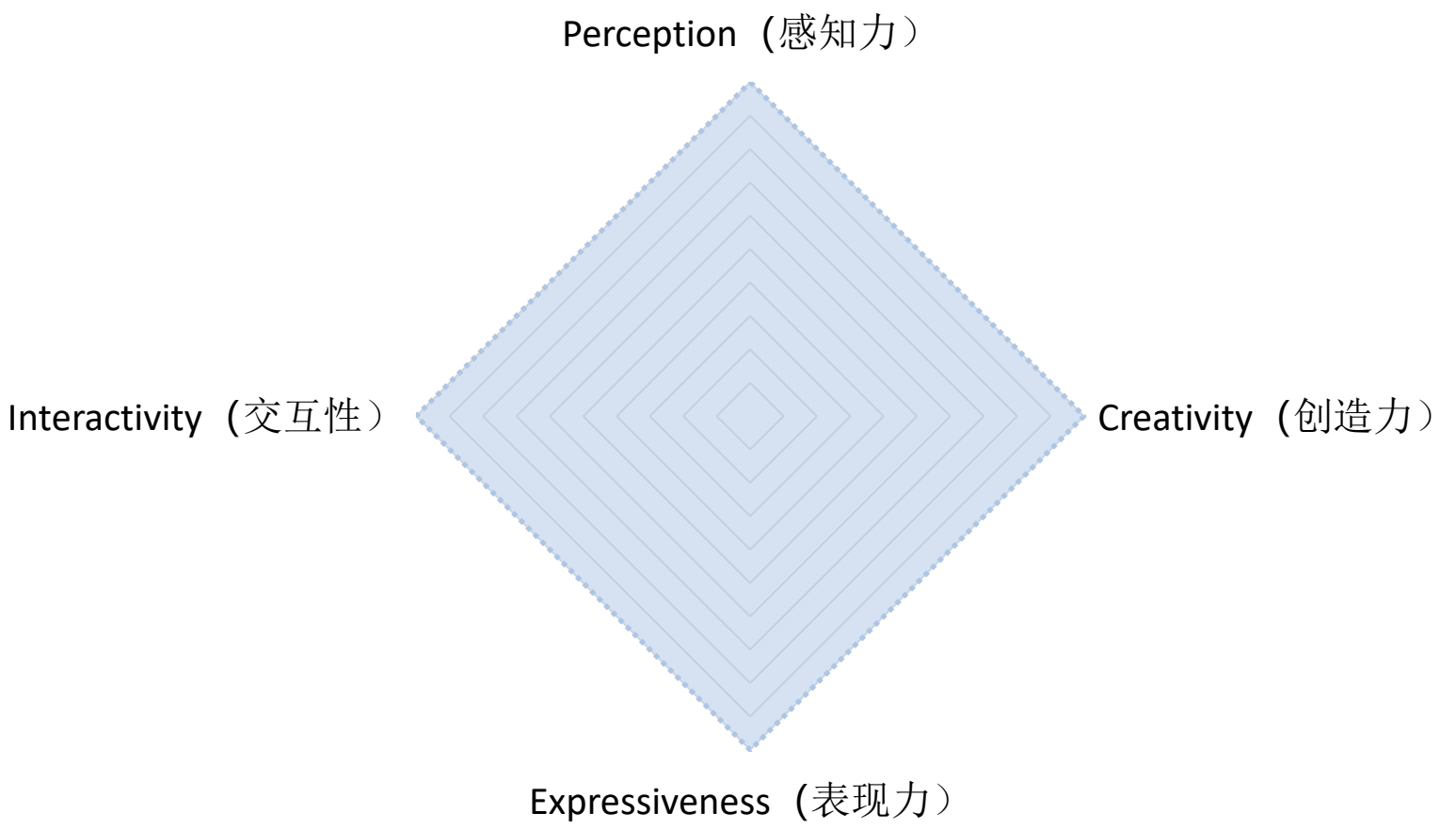
(Rowe, Interactive Music Systems, 1992)



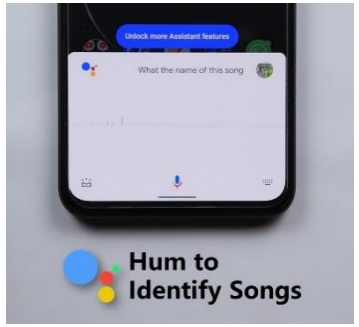



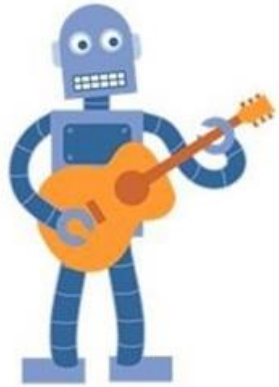

  - E.g., automatic accompaniment systems
    - Score-driven, sequenced, player
- This perspective is more on functionality design instead of capability

# Four Dimensions of Capability

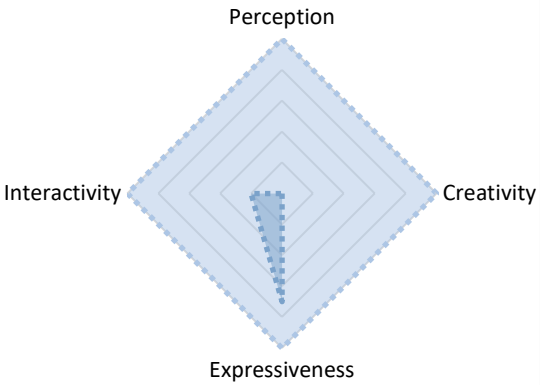


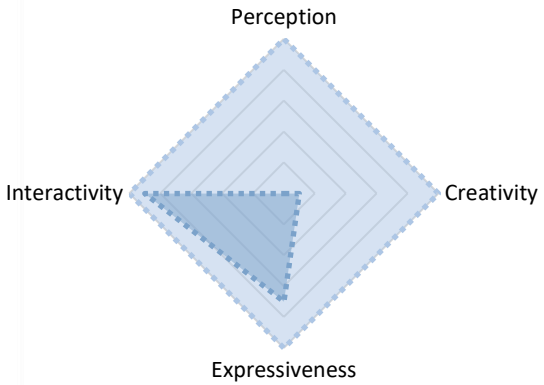
# Let’s measure some systems

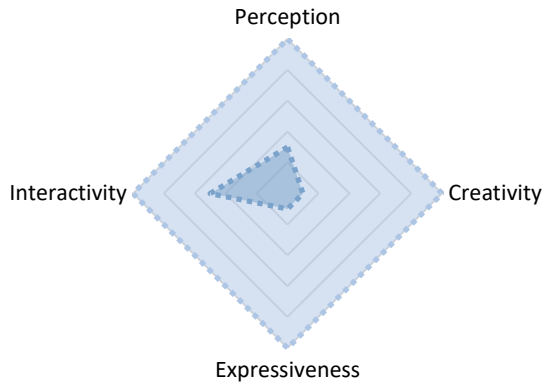


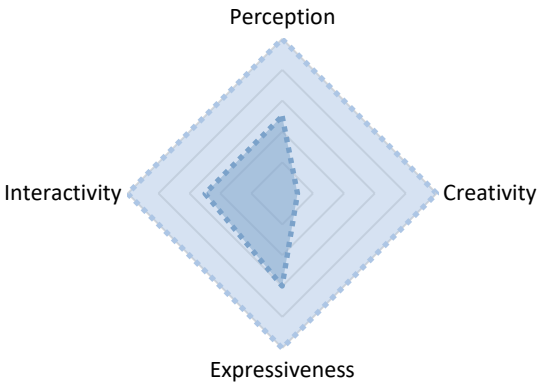



Automatic accompaniment

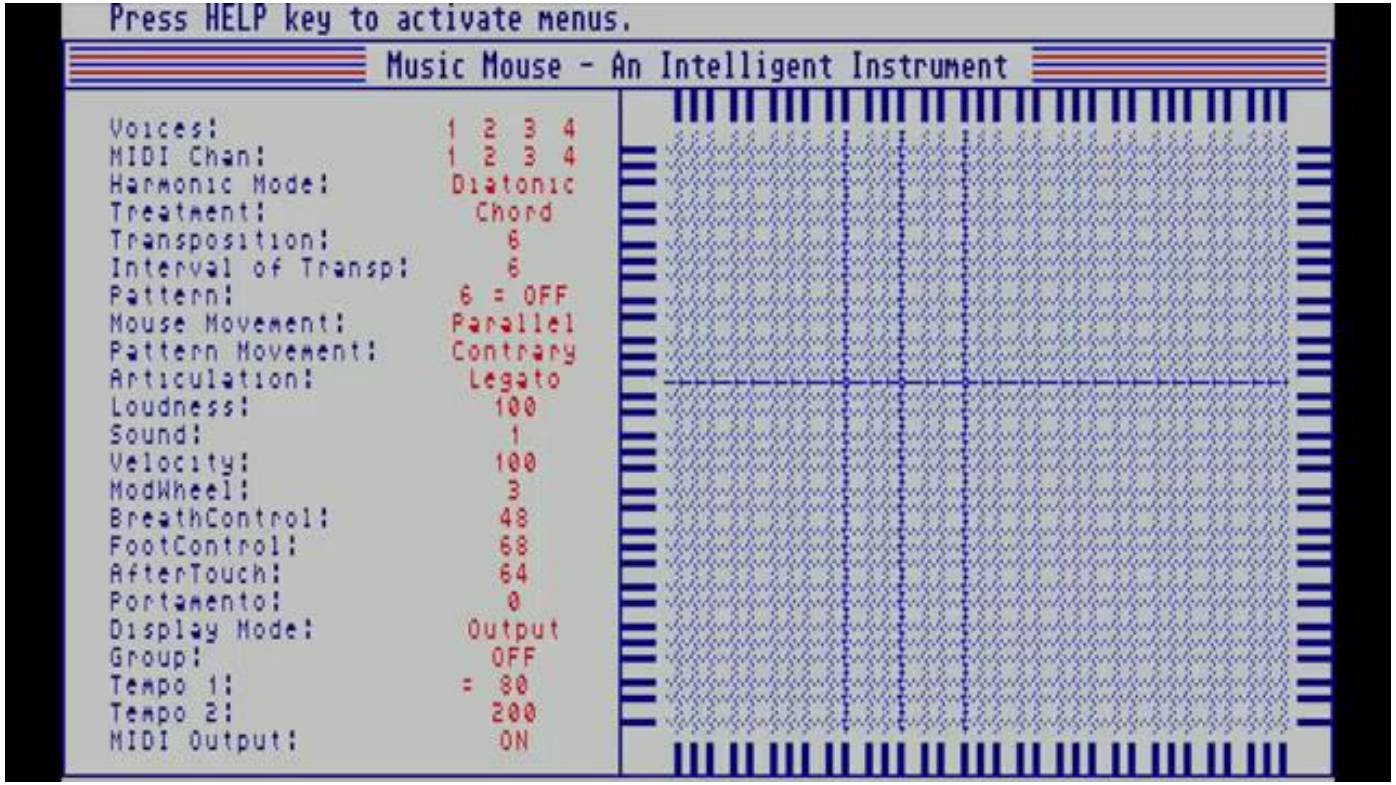


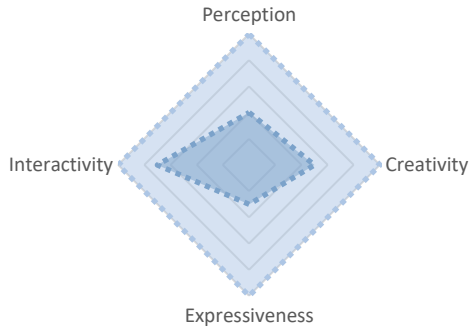



# Music Mouse (http://musicmouse.com/) (Laurie Spiegel, 1986)

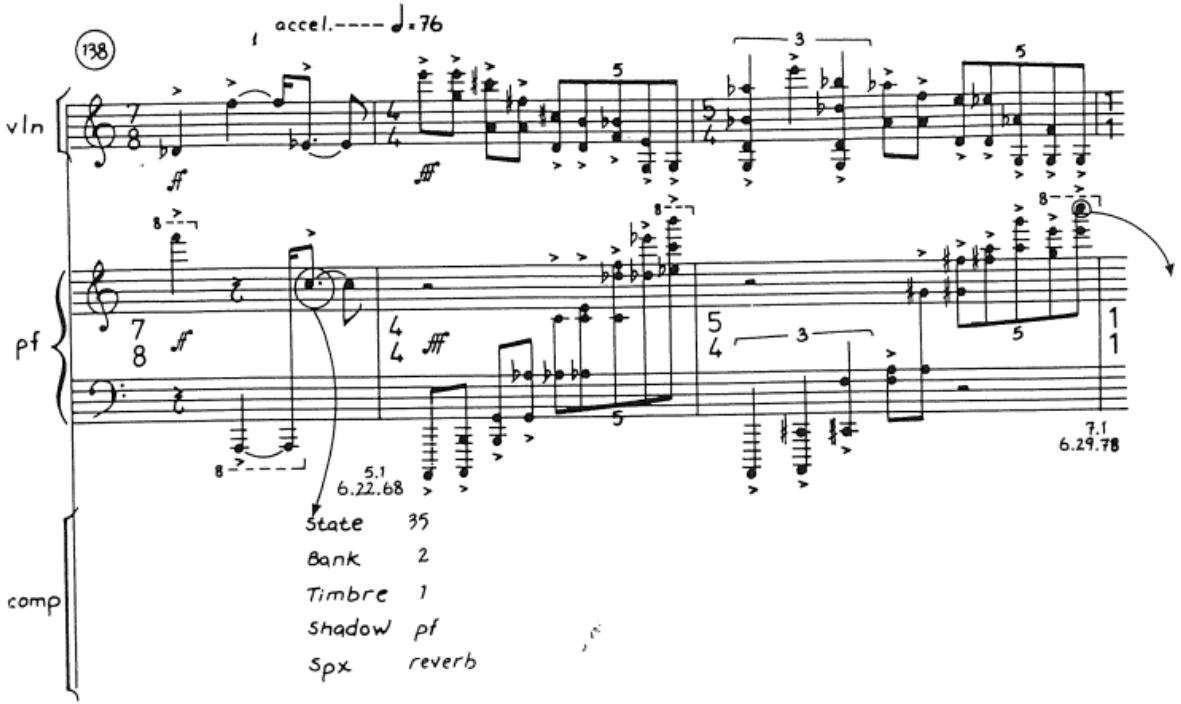



https://www.youtube.com/watch?v=D-mmEvGOopk

- Rule-based music harmonization and improvisation
  - User moves mouse in 2D space, controlling 2 voices
  - System generates the other 2 voices
  - User uses keyboard commands to control orchestration, harmonic mode, tempo, etc.

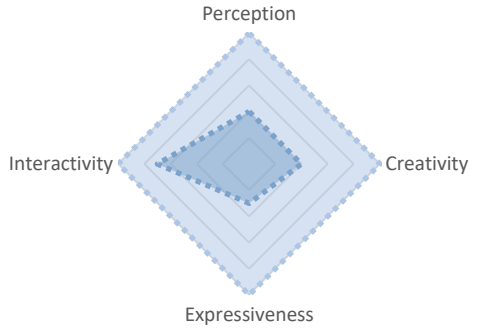


# Cypher
# (Robert Rowe, 1992)

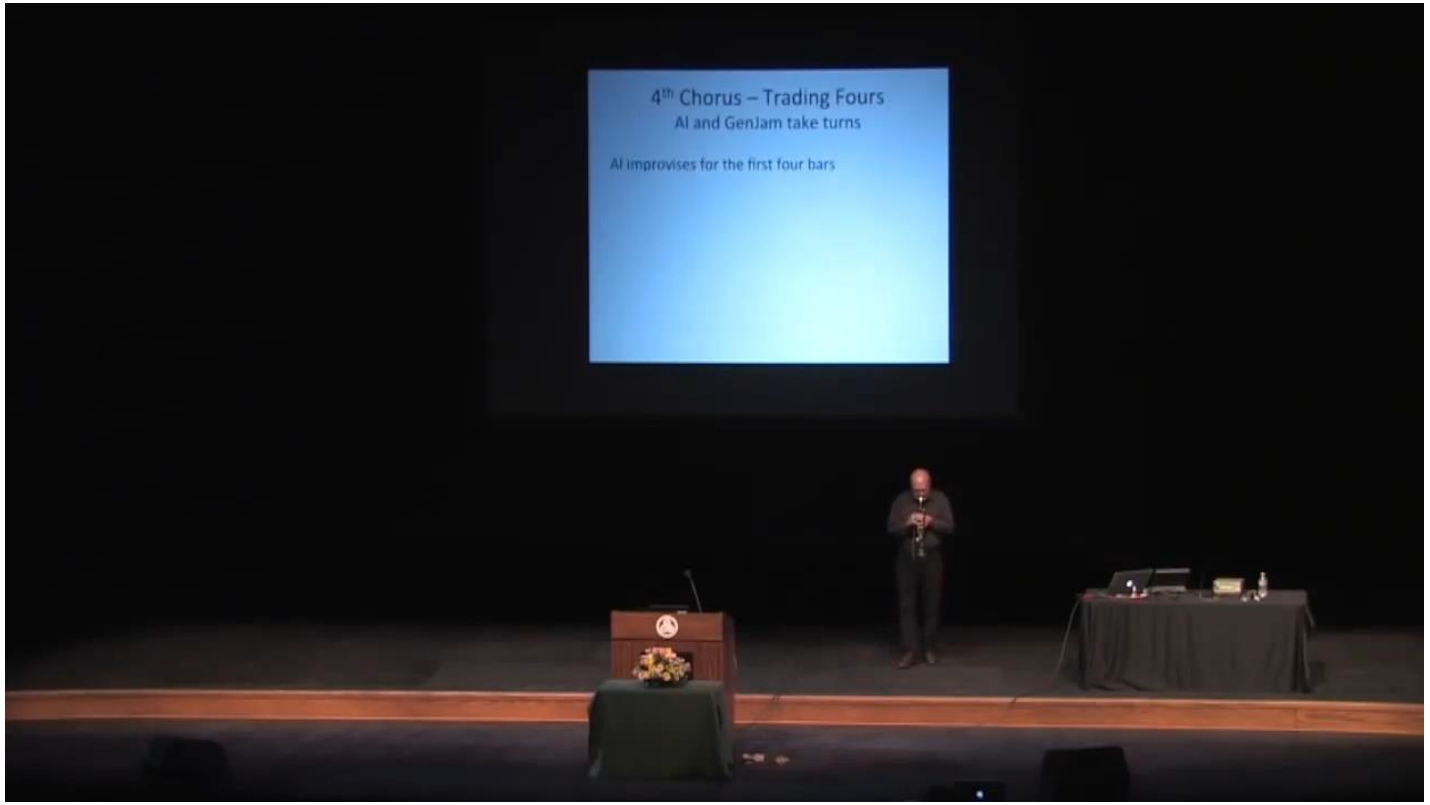



(Rowe, Interactive Music Systems, 1993)

- Multi-agent system responding to human MIDI input in real time
  - Listener analyzes MIDI input (e.g., vertical density, attack speed, loudness, register, duration and harmony, beats, tonal pivots, etc.)
  - Player produces musical output in a virtually deterministic way

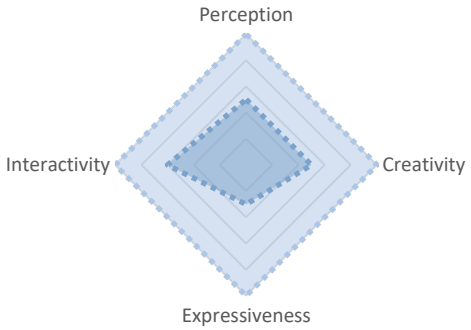


# GenJam
# (Al Biles, 1994)




https://www.youtube.com/watch?v=rFBhwQUZGxg

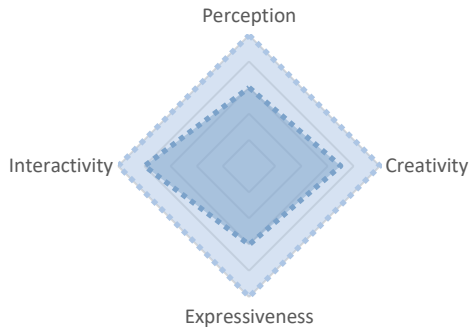
- Genetic algorithm for jazz improvisation (trade fours)
  - Listens to human's four measures
  - Maps to its chromosome representation
  - Mutates the chromosomes
  - Generates the next four measures



# Voyager
# (George Lewis, 1999)

https://www.youtube.com/watch?v=IBPJ2HAmsc8

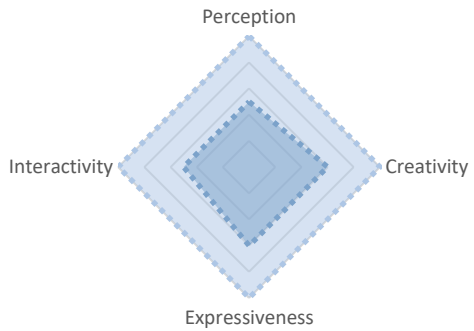
- Multi-agent system with stochastic selection of agent combinations
  - Listens to MIDI or acoustic data (e.g., tempo, note spacing, melodic interval width, primary pitch material, octave range, microtonal transposition, and volume)
  - Improvises on many musical aspects (e.g., timbre, volume, microtonal transposition, tempo, tactus, note probability distributions, pitch interval range, and inter-onset time intervals)



# Continuator (François Pachet, 2002)


https://www.youtube.com/watch?v=ynPWOMzossI

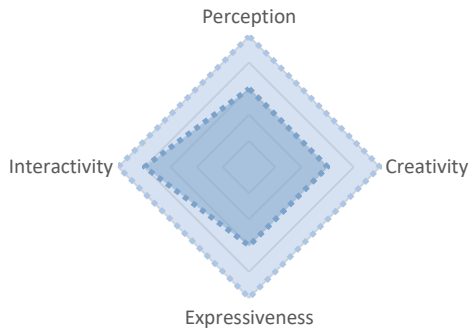
- Continuing music in the same style
  - Modeling user MIDI input sequences with a variable-order Markov model and builds pre-fix trees
  - Random traversals of trees to generate continuations



# Omax-Ofon
# (Assayag, Bloch, & Chemillier, 2006)

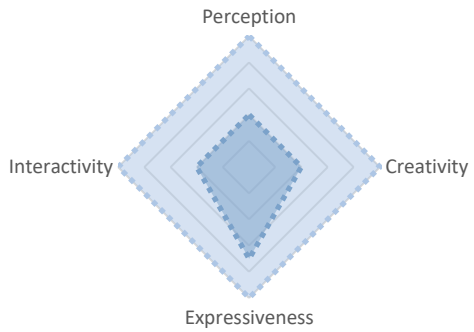

https://www.youtube.com/watch?v=2jFpGQbrcag

- Improvising based on what users just played
  - Modeling note sequences with factor oracle (a finite state automaton for efficient string matching)
  - Sampling sub-sequences to play back
  - Supports MIDI/audio input and multi-player/system settings

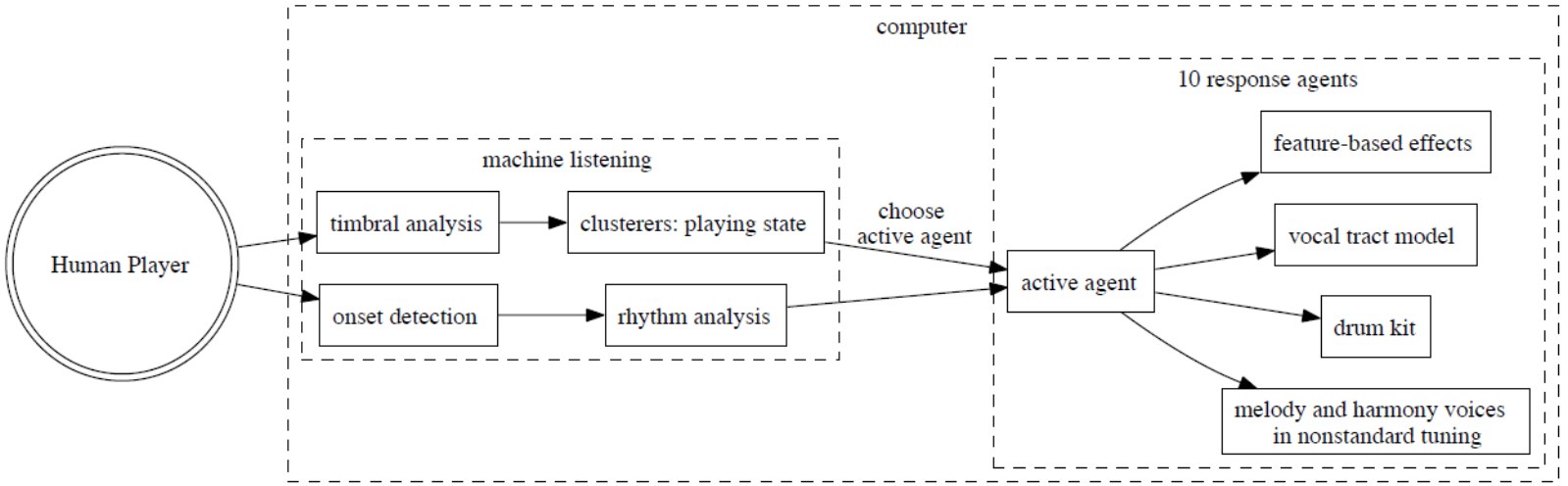


# Shimon
# (Hoffman & Weinberg, 2006)

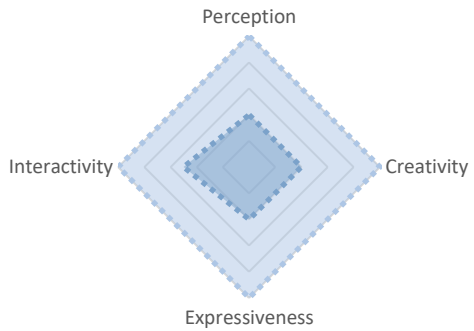
- A robotic marimba player for interactive improvisation
  - Physical embodiment greatly helps the audience to enjoy the performance
  - Beat tracking and chord matching to adapt to human's tempo variation
  - Improvisation centered around the choreographic aspect of the movement



# LL (ListeningLearning) (Nick Collins, 2009)



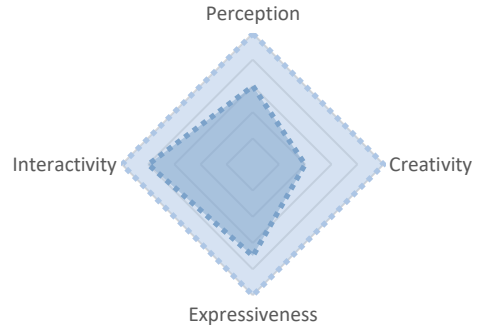

- Rule-based system for free improvisation with humans
  - Rhythm tracking: onset, inter-onset interval
  - Silence detection: perceived loudness
  - Timbral state clustering: using low-level acoustic features
  - Generation: choose among 10 agents to follow the human's timbral state



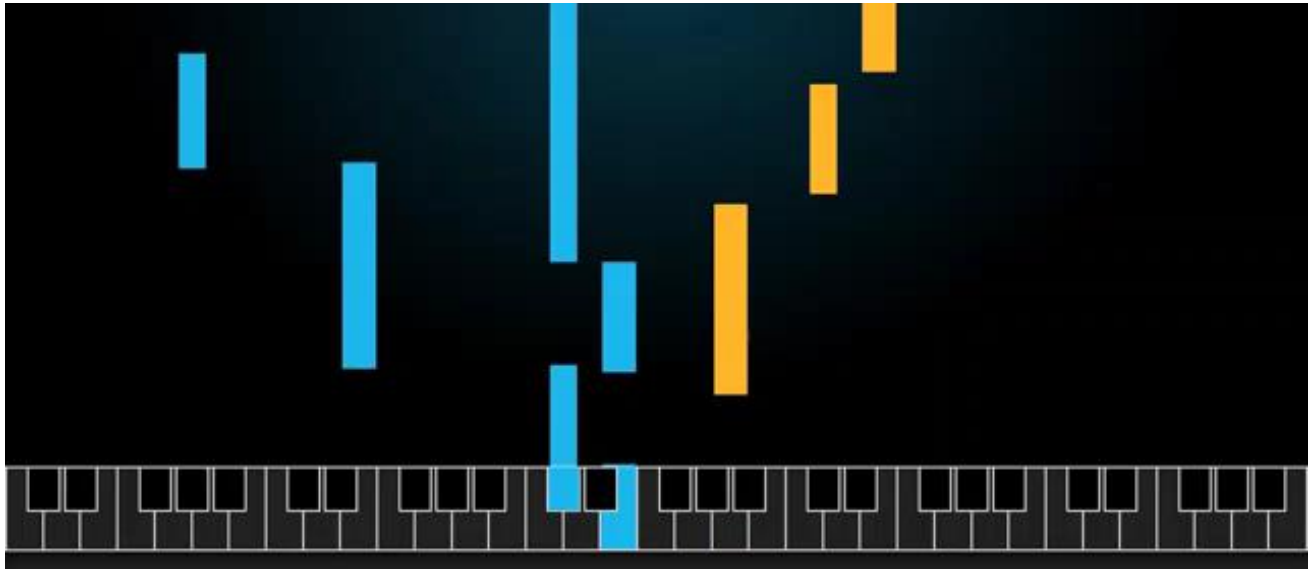
# Reflexive Looper (Pachet et al., 2013)

https://www.youtube.com/watch?v=oquvn8GybRs

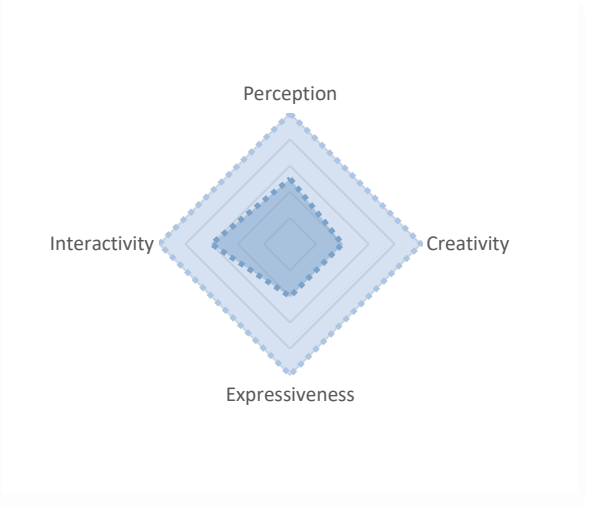
- A system allowing users to play with past virtual copies of themselves
  - Takes simultaneous MIDI and audio input: MIDI for analysis and audio for resynthesis
  - Uses an SVM classifier trained on MIDI data to classify the mode of user playing: bass, chords, and melody
  - Resynthesizes the other modes using past input audio



# A.I. Duet
# (Google Magenta, 2016)



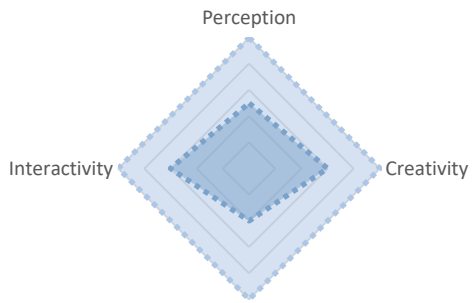
- A neural network model that responds to tunes played by the user on a MIDI keyboard using a similar style



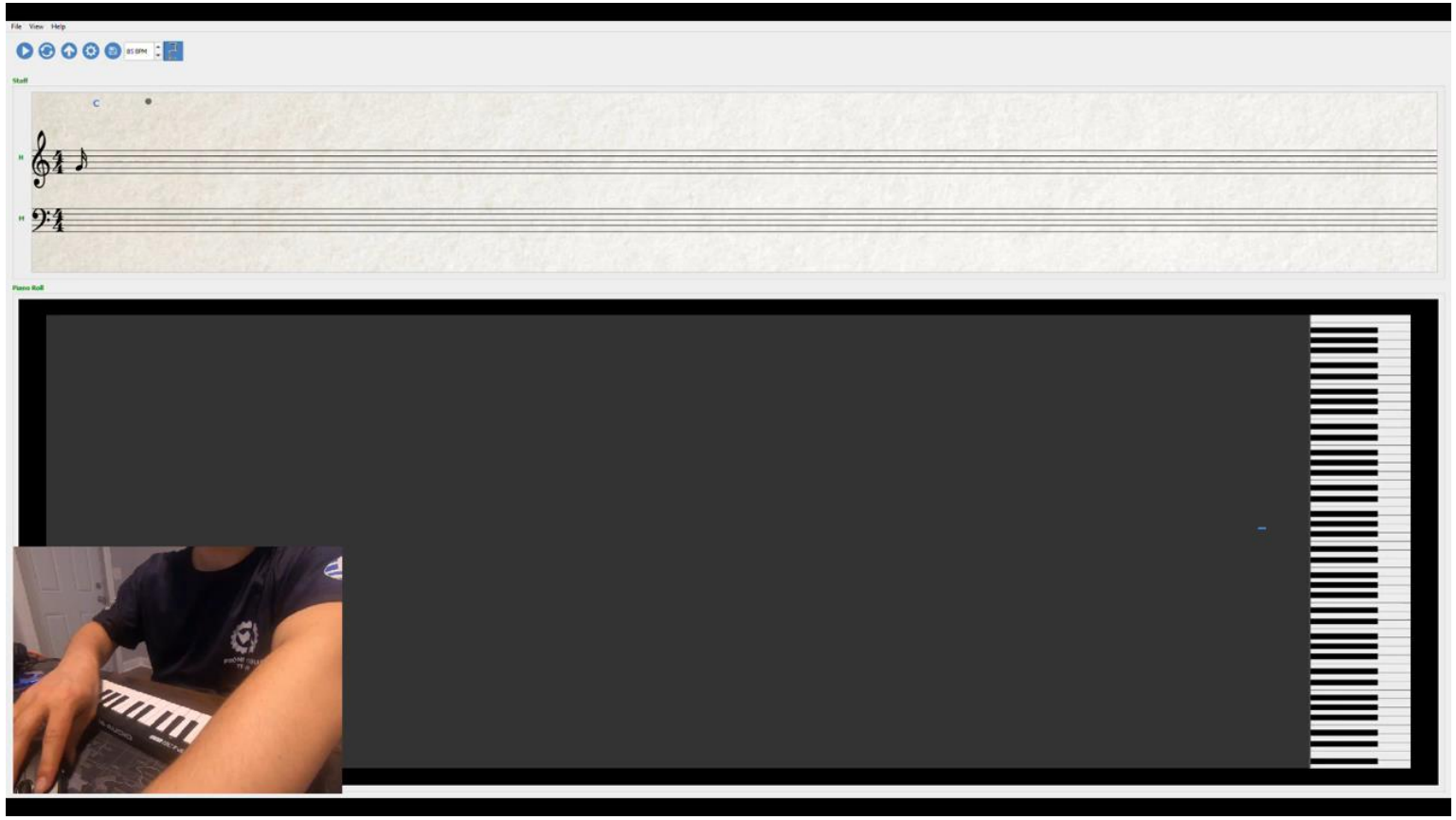
# Piano Genie (Donahue, Simon, & Dieleman, 2019)

https://www.youtube.com/watch?v=YRb0XAnUpIk

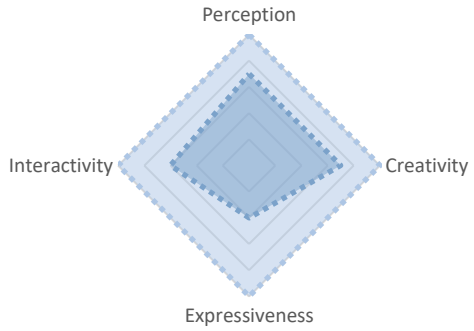
- Allowing users to improvise piano music on an 8-button controller
  - Uses an autoencoder to map note sequences in the 88-d space (corresponding to the 88 piano keys) to sequences in the 8-d space
  - Trained on 1400 piano performances by skilled pianists



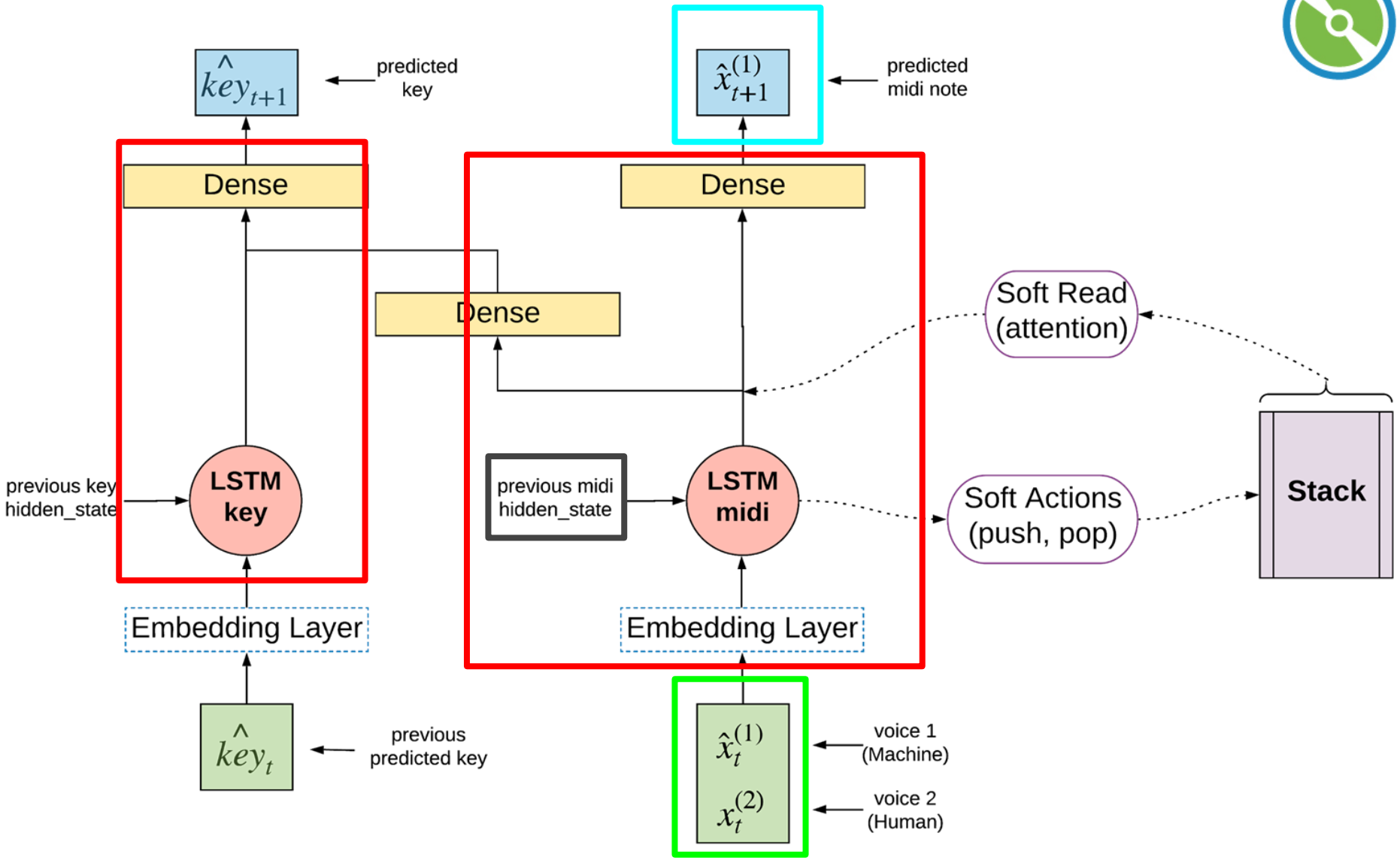
# BachDuet (https://bachduet.com/) (Benetatos & Duan, 2019)



- A neural network based system to allow human-AI duet improvisation in the style of Western counterpoint
  - Trained on outer voices of 370+ Bach chorales
  - Relatively equal role between human and AI – 6:4
  - Only supports MIDI input and fixed tempo

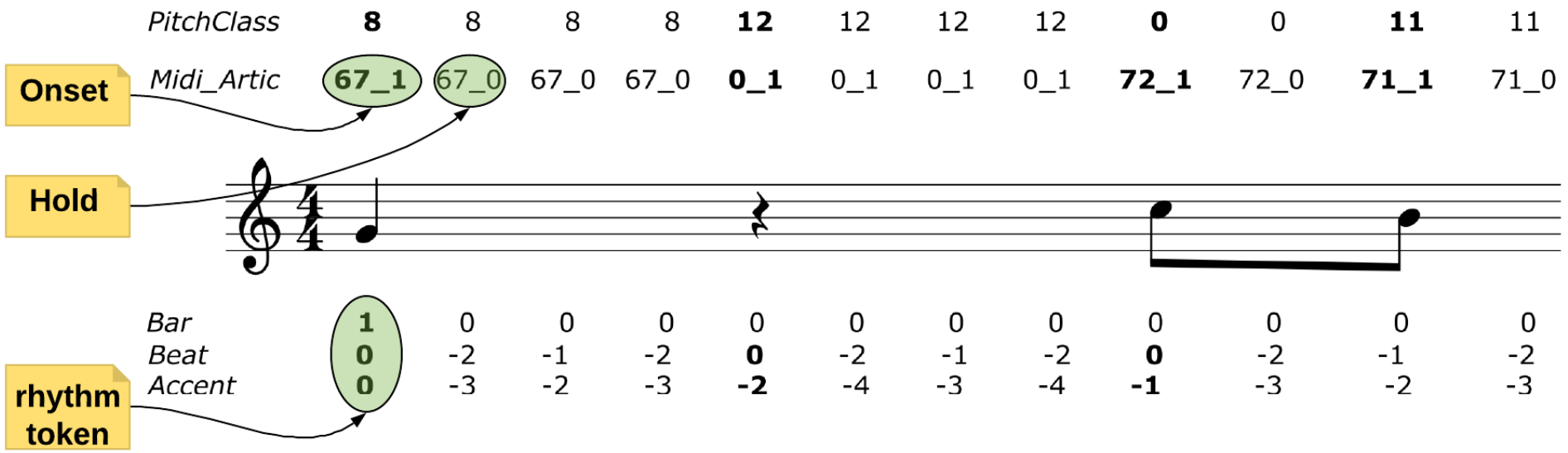


# Model Architecture

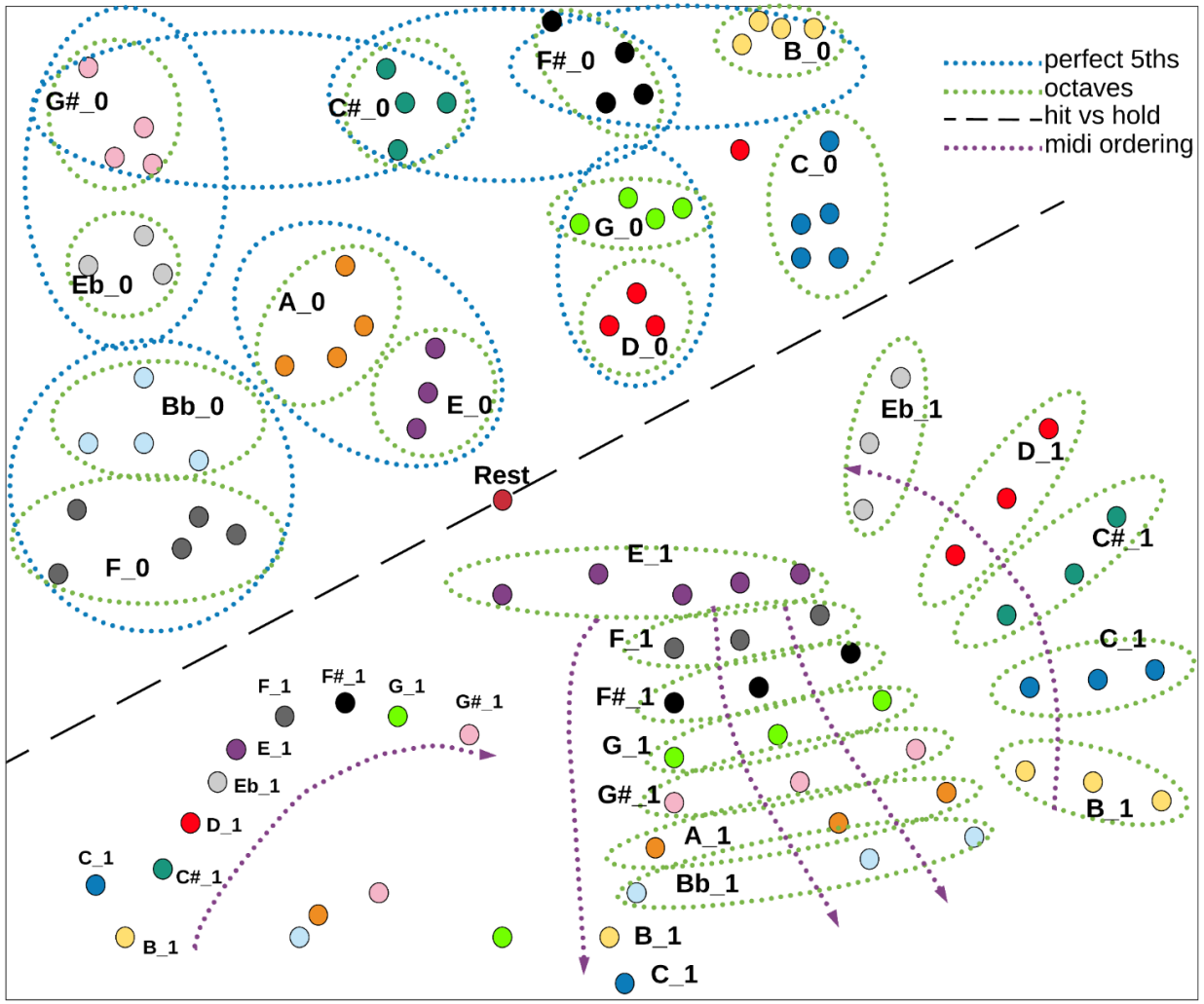


# Music Encoding

(Yan et al., ISMIR'18)

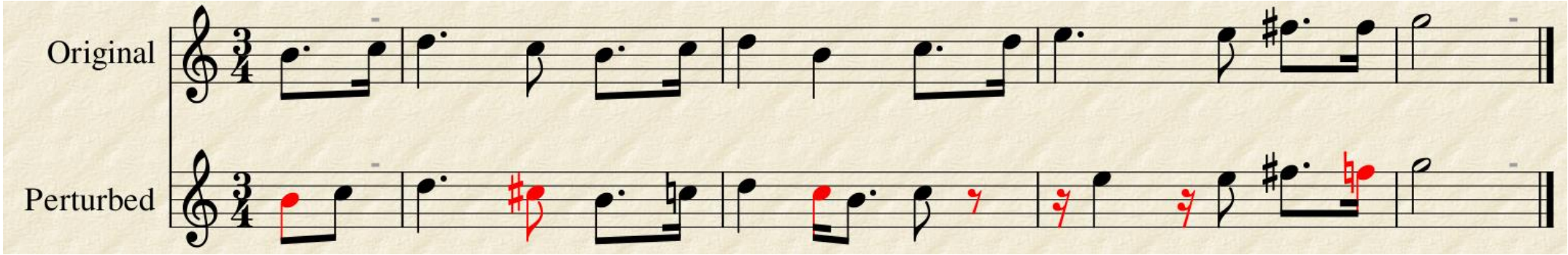


# Music Embedding



# Model Training

- Training data: soprano-bass duets from J.S. Bach chorales
- Augmented with 12 transpositions, totaling 8760 duets
- Data perturbation to improve robustness to human mistakes
  - Time shift by 16$^{th}$ note, wrong duration, pitch shift by 1 semitone, deletion



# Subjective Tests

- Group 1 (user study)
  - 13 musically trained participants
  - Above average keyboard and improvisation skills
  - Played with BachDuet (HM task) and with another human (HH task)
  - Answered a questionnaire about their experience

- Group 2 (Turing listening test)
  - 48 musically trained participants
  - Listened to a random selection of duets from Group's 1 HH and HM tasks
  - Predicted which duets were HH and which were HM
  - Rated the quality of duets on a scale from 1 (dislike) to 5 (like)

# Results

- Group 1 (user study)
  - Things they liked
    - The GUI is intuitive, and the visualization is very accurate
    - BachDuet's output is consistent with the Chorale style
    - The machine and the human have relatively equal roles in the improvisation
    - Most users think BachDuet can improve their improvisation skills
  - Things they disliked
    - The key prediction can be improved
    - Most users rated their interaction with another human (8.6/10) slightly higher than with BachDuet (8/10)

# Results

- Group 2 (Turing listening test)
  - The HM duets received better (3.64/5 vs 3.54/5) but not statistically significant rating
  - Participants could not easily differentiate between HH and HM duets

| true\predicted | HH | HM |
|---|---|---|
| HH | 45.3% | 54.7% |
| HM | 48.1% | 51.9% |

# Examples

## -- Which is HH/HM?

Human vs Machine
(BachDuet)

Human vs Human

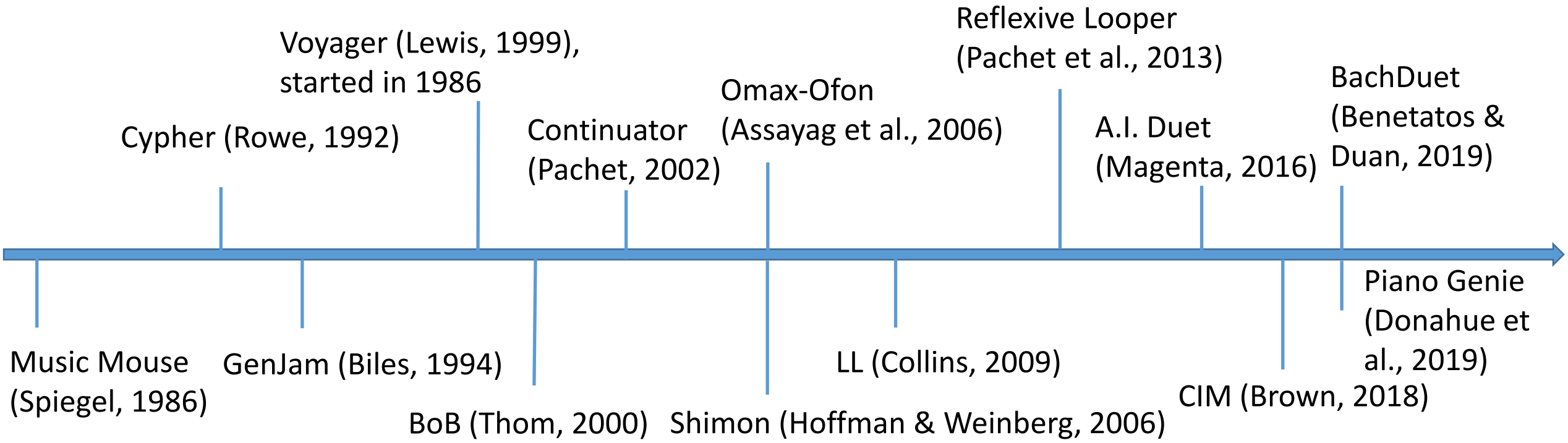
# Summary

- Duet counterpoint improvisation is a feasible but not so easy task for classical musicians
- Users find BachDuet a good partner for duet improvisation
- Listeners cannot distinguish between HH and HM duets

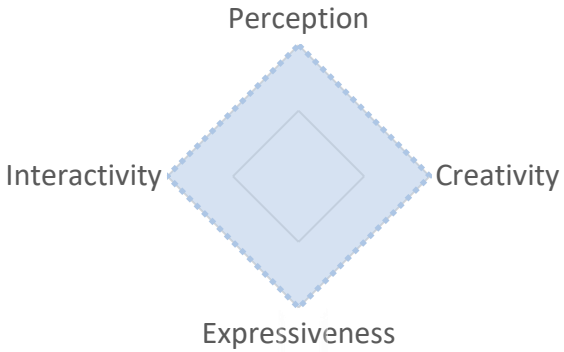
- Limitations
  - Fixed tempo
  - Trained on limited data
  - Only MIDI is supported (not Audio) so far
  - No visual interaction

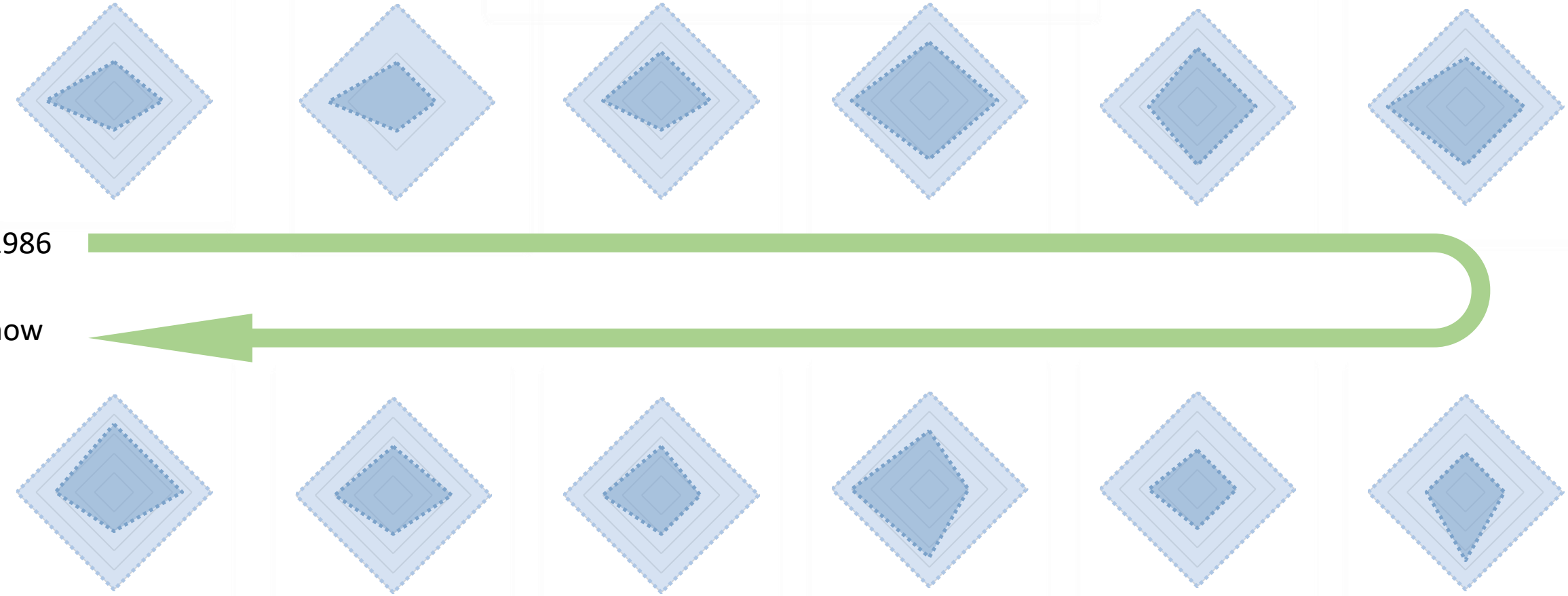
- Web version: https://bachduet.com/

# Interactive Music Systems
## Beyond Automatic Accompaniment



# The Trend?

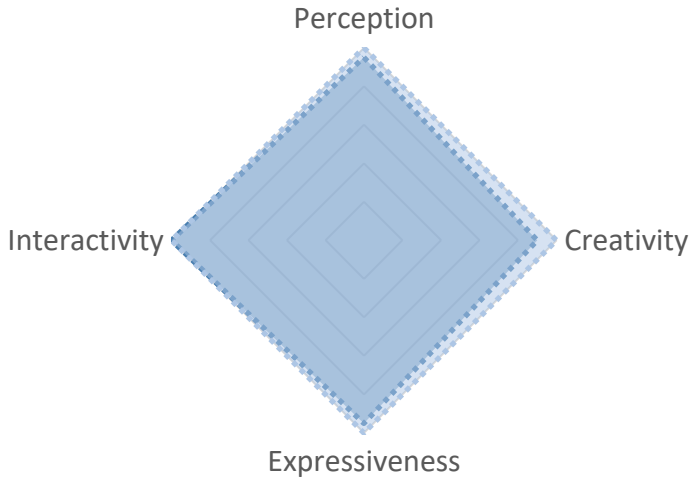

Perception
Interactivity
Creativity
Expressiveness


1986
now

# The Trend?

- Perception: was improved due to audio analysis and music language modeling techniques
- Creativity: was improved due to machine learning based music generation techniques
- Expressiveness: a few systems started to leverage physical embodiment and visual rendering techniques
- Interactivity: did not seem to be improved, yet different ways of interaction were attempted

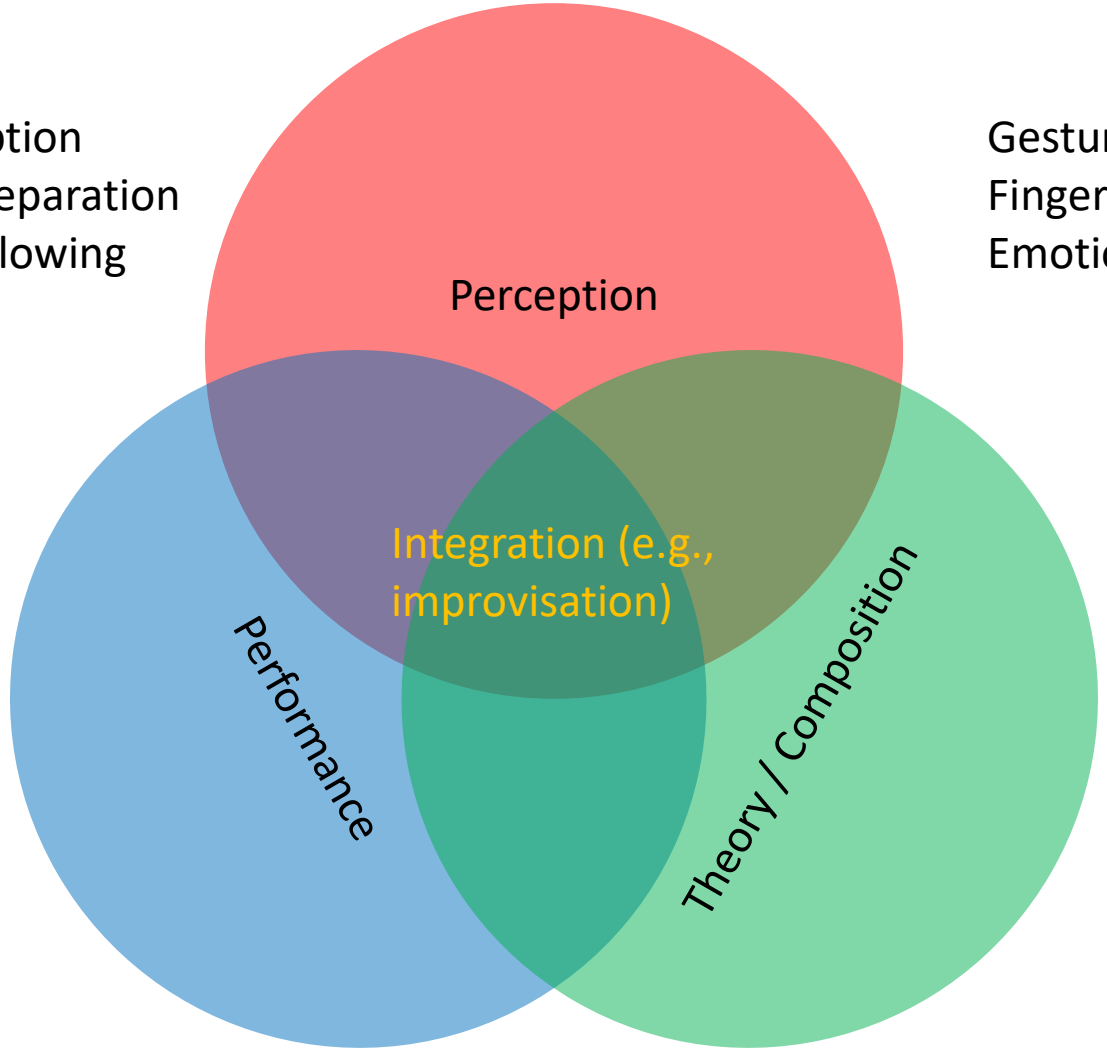
# Human Musicians in a Jazz Combo



# I think this is what we need

Transcription
Source separation
Score following

Gesture/movement analysis
Fingering analysis
Emotion/intension estimation



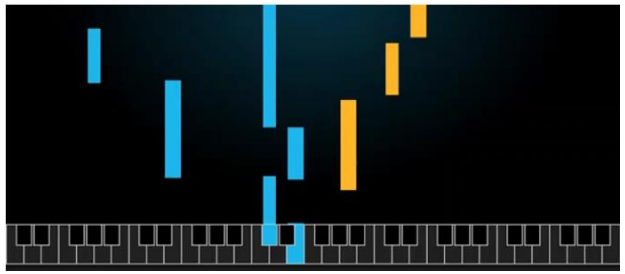

Sound synthesis
Expressiveness modeling

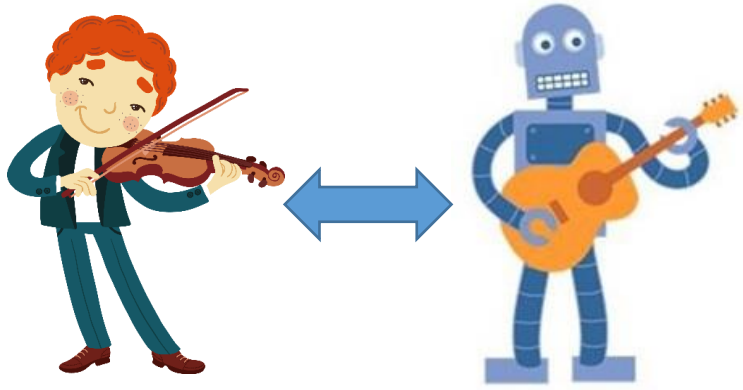
Visual rendering
Physical embodiment

Music language models
Music generation

# Moving from Symbolic to Audio

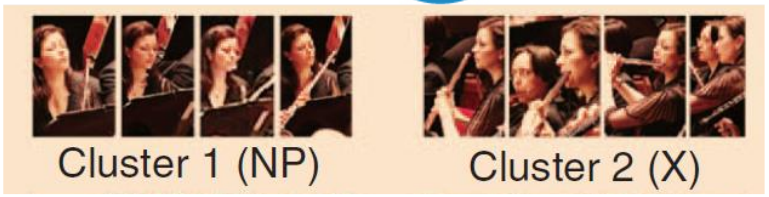
- Audio signals offer much more musical expressiveness, but
  - Most existing musical agents work in the symbolic domain
  - Agents work in the audio domain often only analyze low-level features (e.g., Voyager, LL) or monophonic audio (e.g., GenJam, Omax-Ofon)
- Need more robust music analysis algorithms
  - Real-time beat tracking and rhythm analysis
    - Beat tracking for percussion-less music input, e.g., singing voice [Heydari et al., SingNet, 2023]
  - Fine-grained polyphonic pitch tracking to analyze pitch fluctuations e.g., vibrato
  - Robust score following to performance mistakes, improvisation, and structural changes
- Need expressive audio synthesis and coordinating it with human performance on timing, dynamics, and timbre



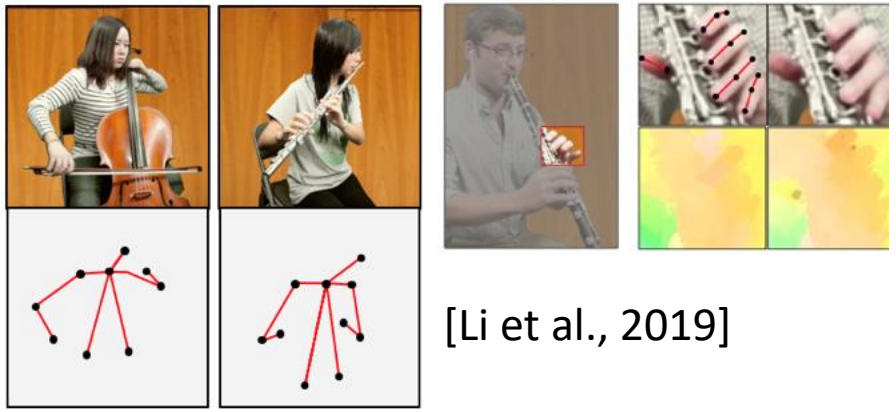


# Incorporating the Visual Modality

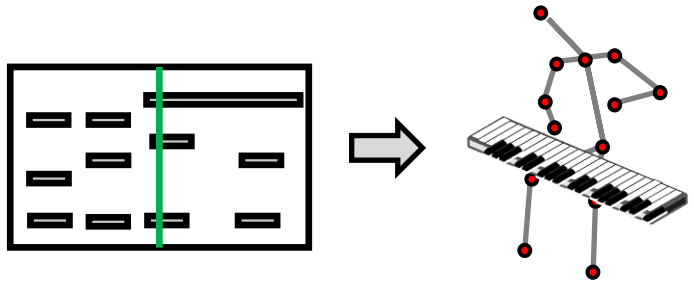
- Music performance is audiovisual in nature
  - Visual performance is important in musical expression
  - Musicians use visual cues to coordinate on timing, dynamics and intention
- Need algorithms to analyze various aspects of visual performance
  - Instrument recognition, body movement, facial expression, fingering motion
  - Audiovisual association and joint analysis
- Need real-time expressive visual rendering and coordinating it with human performance
- 2019 ISMIR Tutorial on Audiovisual Processing Processing: https://github.com/bochen1106/ISMIR2019-Tutorial3-Audiovisual-Music-Processing




[Bazzica et al., 2016]



[Li et al., 2019]



[Li et al., 2018]

# Personalization

- Existing CAMM systems are “standardized”
- Can we make them adapt to user behaviors, habits, and preferences?
  - Learning from rehearsal: Chris Raphael’s Music Plus One system [Raphael, 2002]
- Some ideas
  - Provide more options in preference settings
  - Learn from interaction history
  - Provide feedback to users after interaction
  - Make suggestions on improvisation

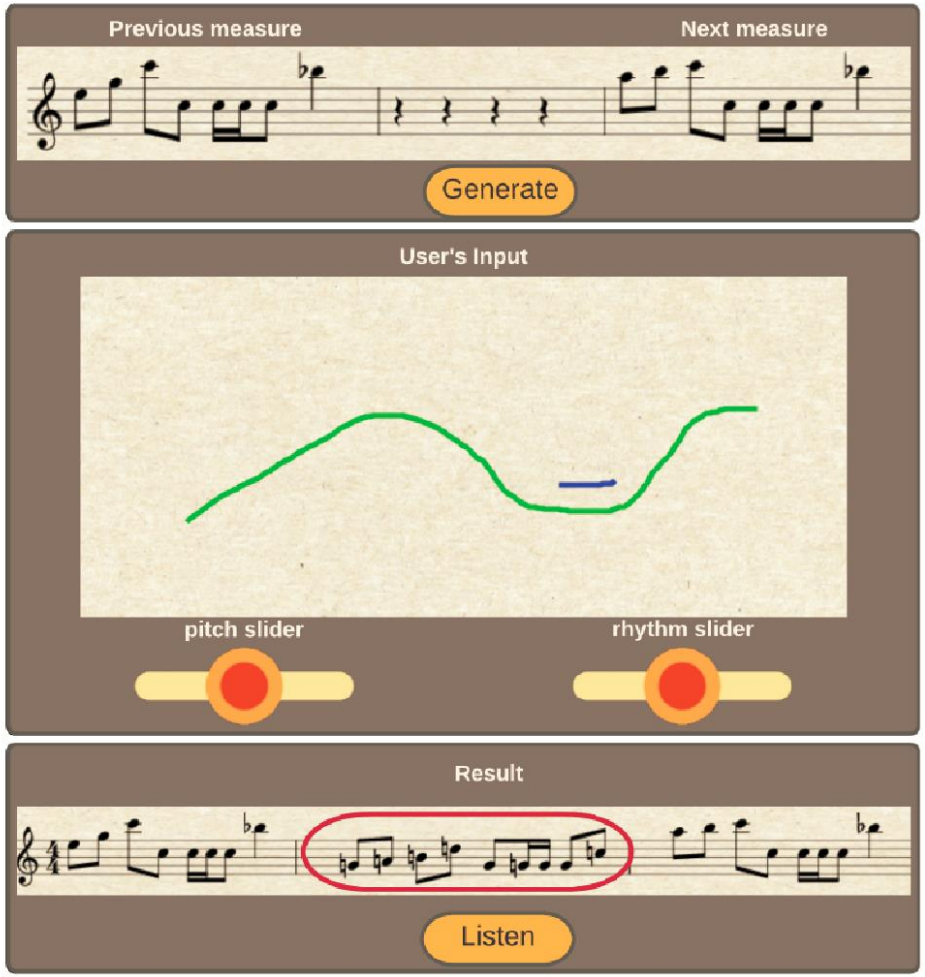
Chris Raphael’s Music Plus One system demo

# Improving Accessibility

- Music theory is hard
- Music instruments are not easy to learn

- Many existing CAMM systems seem to be even harder to interact with
  - Require music background + software literacy

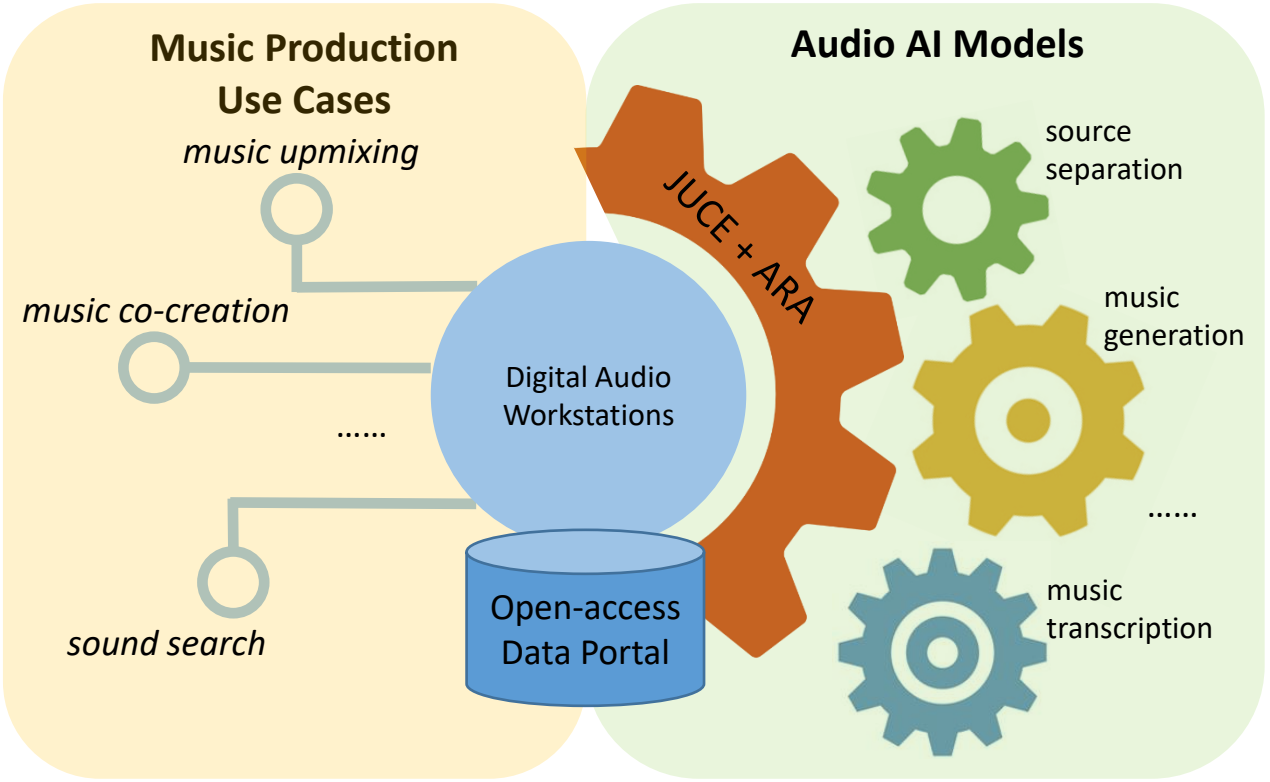
- Lower the barrier to entry?
  - Use music AI to augment users' capabilities




Draw&Listen [Benetatos & Duan, 2022]

# Bridging Music AI and Music Production

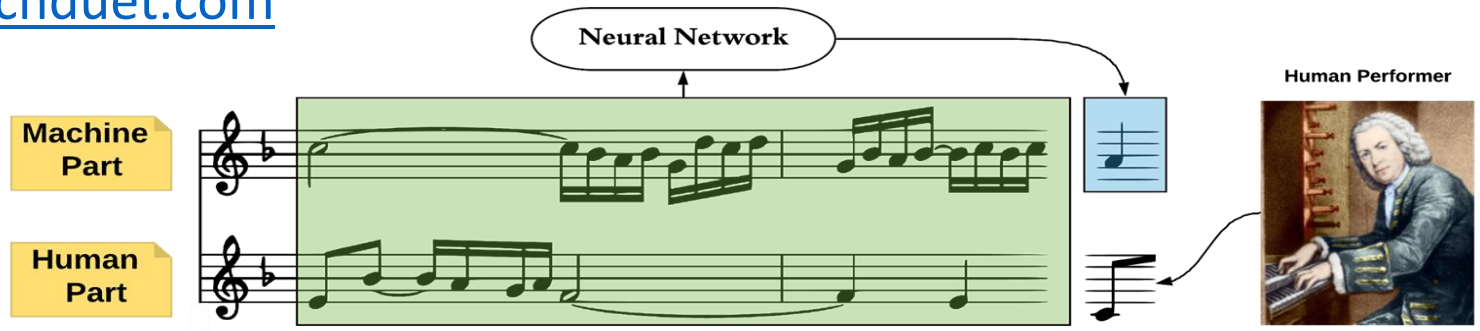
- Music AI models are released on a weekly basis, but most of them only stay in the labs
- Tools for music making (e.g., notation software, DAWs) have limited AI functionalities
- Idea: Build (open-source) tools to bridge the gap
  - Euterpe [Zang et al., 2023]
  - Hosted, Asynchronous, Remote Processing (HARP) for audio AI plug-ins [Garcia et al., 2023]
  - Commercial software: Neutone




TEAMuP NSF project (2022-2026)

# Developing CAMM for Education

- Music education
  - Practicing counterpoint improvisation with BachDuet [Benetatos et al., 2020] www.bachduet.com



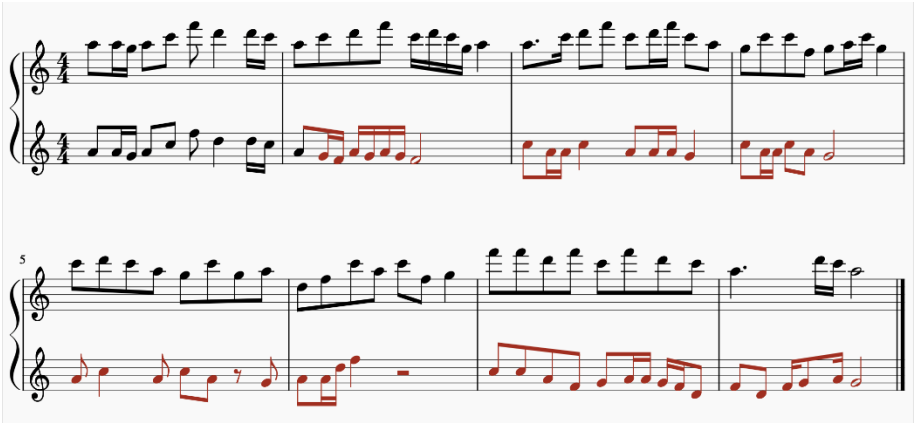

- Computing education
  - EarSketch by Jason Freeman at Georgia Tech: “EarSketch helps students learn in Python or JavaScript through manipulating loops, composing beats, remixin sounds, and applying effects to a multi-track digital audio workstation”
  - TunePad by Michael Horn at Northwestern: “TunePad is a free online platform creating music with the Python programming language.”

EarSketch

TUNEPAD

# Diversifying Music Styles

- Existing CAMM systems focus on Western music styles
- One challenge for diversifying styles is the lack of training data
- For example: Counterpoint composition in Chinese folk music style is an important direction in Chinese music composition
  - One idea for automating this task is to use inverse reinforcement learning to fuse Western counterpoint with Chinese folk styles



Demo for "When counterpoint meets Chinese folk melodies" [Jiang et al., 2020]

# Incorporating Large Language Models


- LLMs (e.g., ChatGPT) represent the most significant AI advances in recent years
- Music generation with LLMs
  - Google's MusicLM: Generating high-fidelity music from text input, [Agostinelli et al., 2023]
  - Meta's MusicGen: Controllable text-to-music generator, [Copet, et al., 2023]

  - Tutorial #3: Transformer-based Symbolic Music Generation: Fundamentals to Advanced Concepts, Stylistic Considerations, Conditioning Mechanisms and Large Language Models, by Berker Banar, Pedro Sarmento, and Sara Adkins
- These models allow people to use natural language to guide the music generation process
- Look forward to more interaction mechanisms and user control flexibilities


Thank You!

# References


- Agostinelli A, Denk TI, Borsos Z, Engel J, Verzetti M, Caillon A, Huang Q, Jansen A, Roberts A, Tagliasacchi M, Sharifi M. Musiclm: Generating music from text. arXiv preprint arXiv:2301.11325. 2023 Jan 26.
- Christodoulos Benetatos and Zhiyao Duan, Draw and listen! A sketch-based system for music inpainting, Transactions of the International Society for Music Information Retrieval, vol. 29, pp. 2288-2292, 2022.
- Christodoulos Benetatos, Joseph VanderStel, and Zhiyao Duan, BachDuet: A deep learning system for human-machine counterpoint improvisation, in Proc. International Conference on New Interfaces for Musical Expression (NIME), 2020, pp. 635-640.
- Copet J, Kreuk F, Gat I, Remez T, Kant D, Synnaeve G, Adi Y, Défossez A. Simple and Controllable Music Generation. arXiv preprint arXiv:2306.05284. 2023 Jun 8.
- Freeman J., Magerko B., Edwards D., Mcklin T., Lee T., Moore R. (2019). "EarSketch: engaging broad populations in computing through music," in Communications of the ACM.
- Hugo Flores Garcia, Christodoulos Benetatos, Patrick O'Reilly, Aldo Aguilar, Zhiyao Duan and Bryan Pardo, “HARP: Bringing deep learning to the DAW with hosted, asynchronous, remote processing, NeurIPS 2023 Workshop on Machine Learning for Creativity and Design.
- Bochen Li, Akira Maezawa, and Zhiyao Duan, Skeleton plays piano: online generation of pianist body movements from MIDI performance, in Proc. International Society for Music Information Retrieval Conference (ISMIR), 2018, pp. 218-224.

# References


- Bochen Li, Karthik Denish, Chenliang Xu, Gaurav Sharma, and Zhiyao Duan, Online audio-visual source association for chamber music performances, Transactions of the International Society for Music Information Retrieval, vol. 2, no. 1, pp. 29–42, 2019.
- McDermott, J., Gifford, T., Bouwer, A., & Wagy, M. (2013a). Should music interaction be easy? In S. Holland, K. Wilkie, P. Mulholland, & A. Seago (Eds.), Music and human computer interaction (pp. 29–48). London: Springer.
- Mojtaba Heydari, Ju-Chiang Wang, and Zhiyao Duan, SingNet: a real-time singing voice beat and downbeat tracking system, in Proc. IEEE Internationsl Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.
- Nan Jiang, Sheng Jin‡, Zhiyao Duan, and Changshui Zhang, When counterpoint meets Chinese folk melodies, accepted by The Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS), 2020.
- Christopher Raphael, C. 2002. A bayesian network for real-time musical accompaniment. In Dietterich, T.; Becker, S.; and Ghahramani, Z., eds., Advances in Neural Information Processing Systems, NIPS 14. MIT Press.
- Yongyi Zang*, Christodoulos Benetatos*, and Zhiyao Duan, Euterpe: A web framework for music interaction and creation, Journal of Audio Engineering Society, 2023. (* equal contribution)