A Brief History of Interactive Music Systems

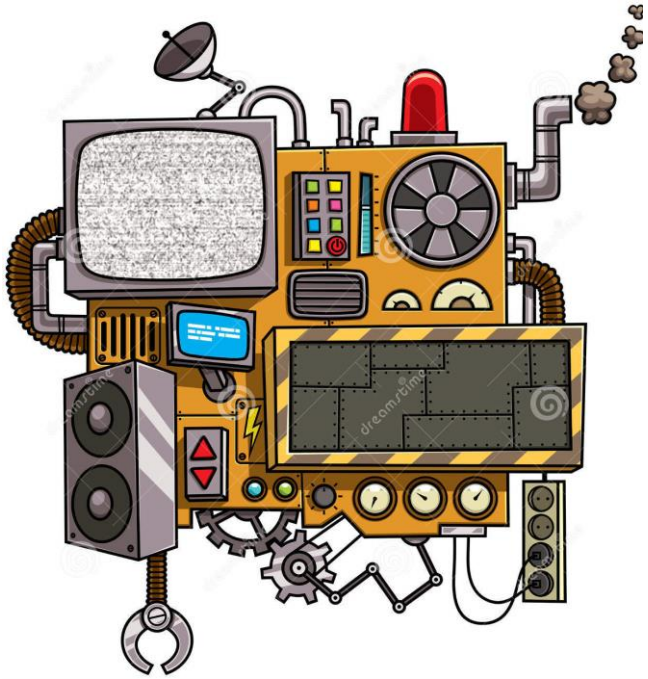
Zhiyao Duan
Audio Information Research (AIR) Lab
Department of Electrical and Computer Engineering
University of Rochester

Tutorial for the CAAI International Conference on Artificial Intelligence (CICAI)
Theme 3 – Art and Artificial Intelligence
Beijing, China, August 27, 2022

Interactive Music System

Exchange of musical information

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



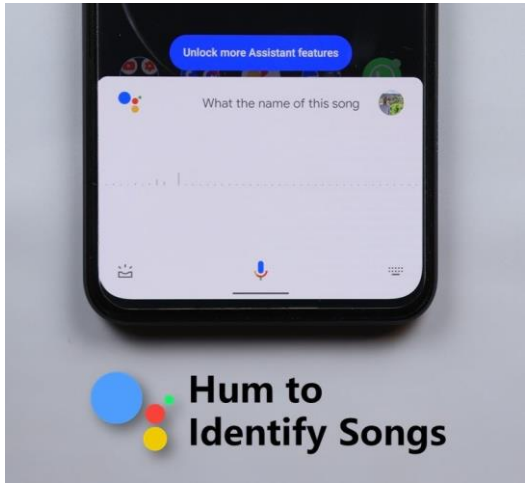
Human

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

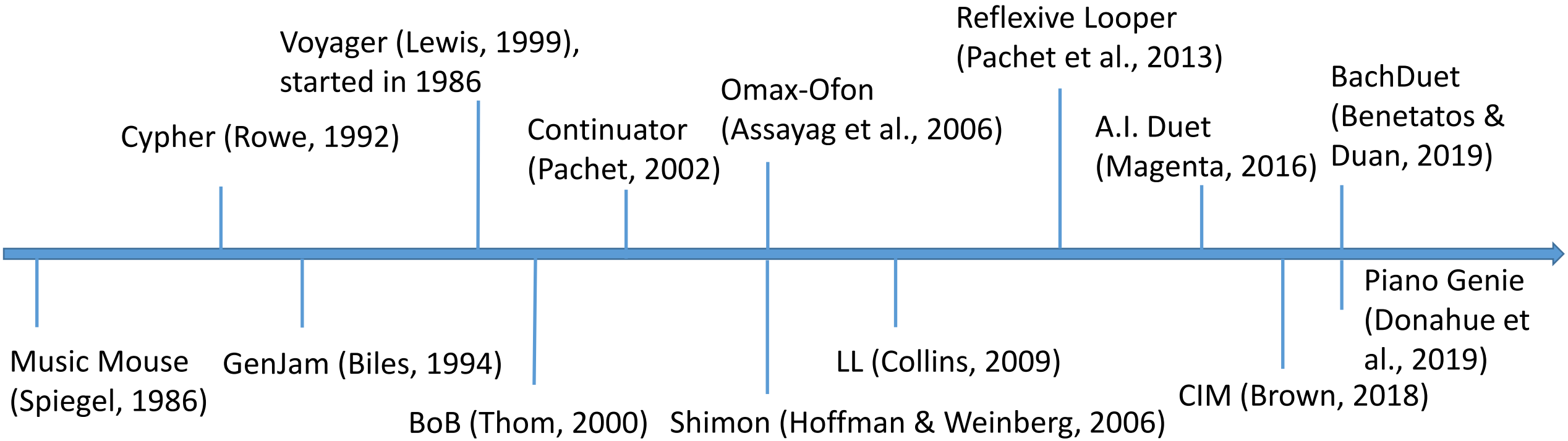
Follows human

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



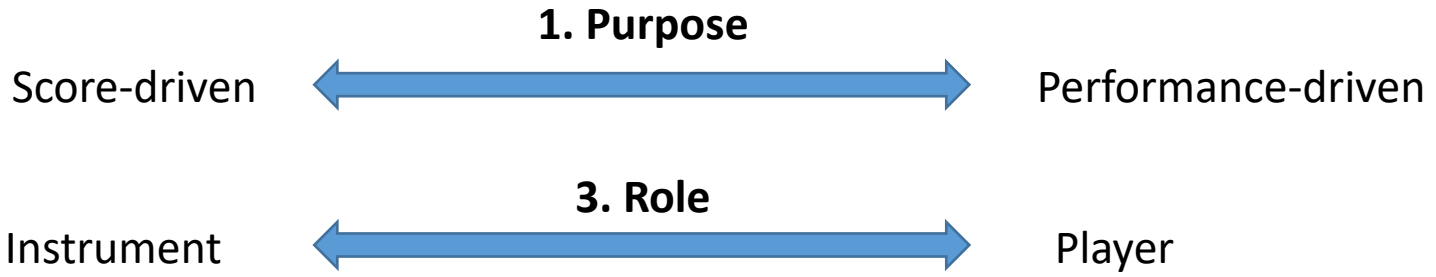
Kivanc Tatar & Philippe Pasquier, Musical agent: A typology and state of the art towards musical metacreation, *Journal of New Music Research*, 2019.

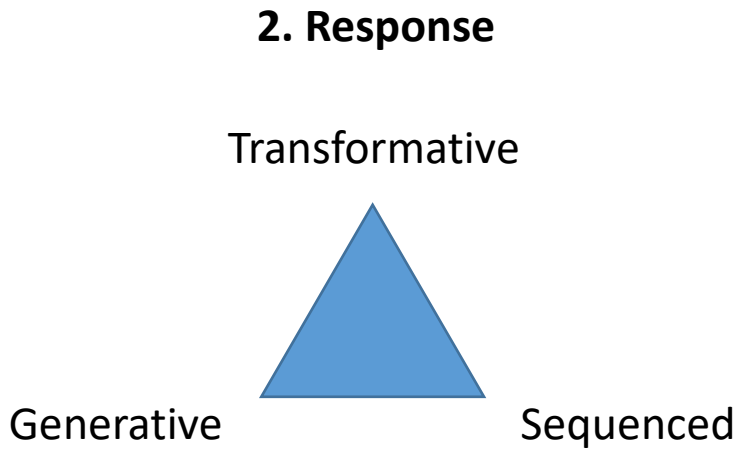
Toby Gifford, et al., Computational systems for music improvisation, *Digital Creativity*, 2018.

Categorization

- Three dimensions classifying interactive music systems

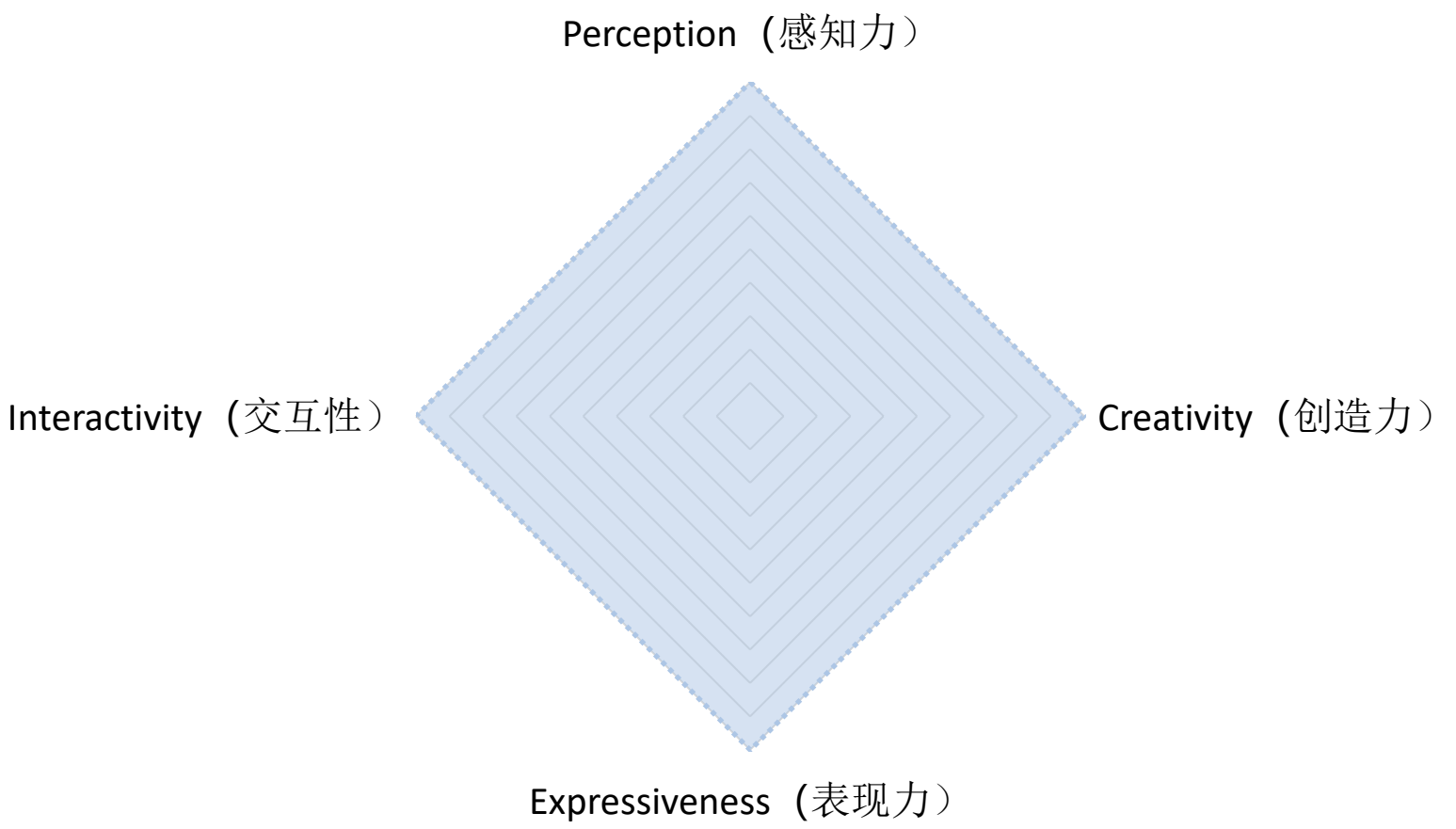
(Rowe, Interactive Music Systems, 1992)



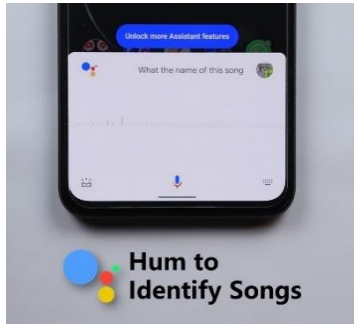


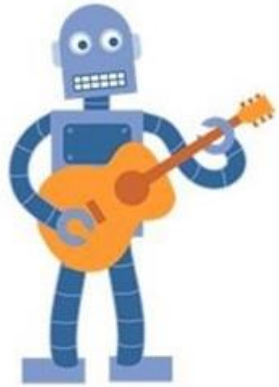
  - E.g., automatic accompaniment systems
    - Score-driven, sequenced, player
- This perspective is more on functionality design instead of capability

Four Dimensions of Capability

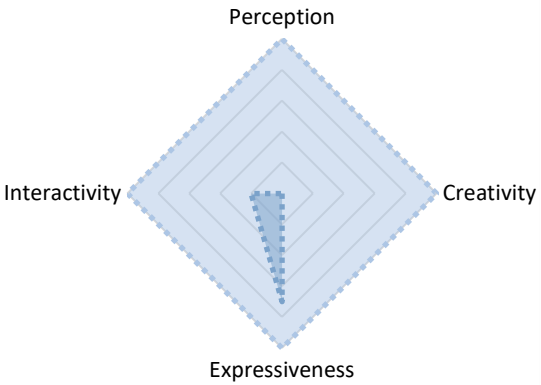


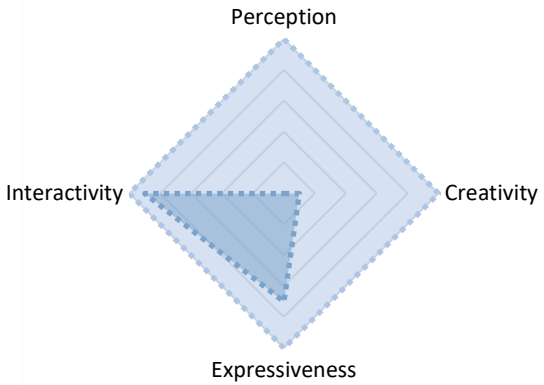
Let’s measure some systems

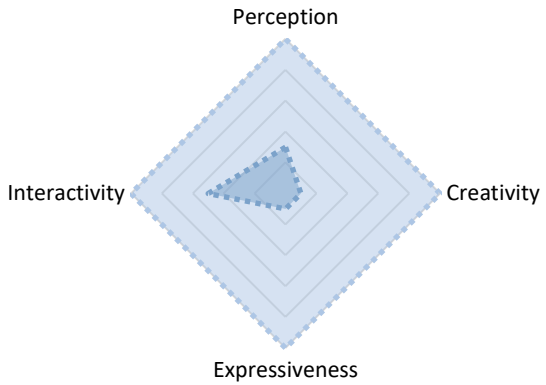


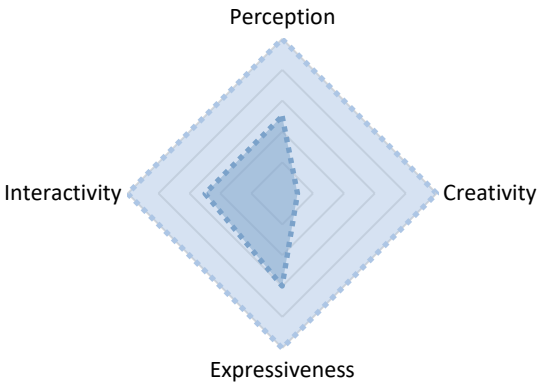


Automatic accompaniment

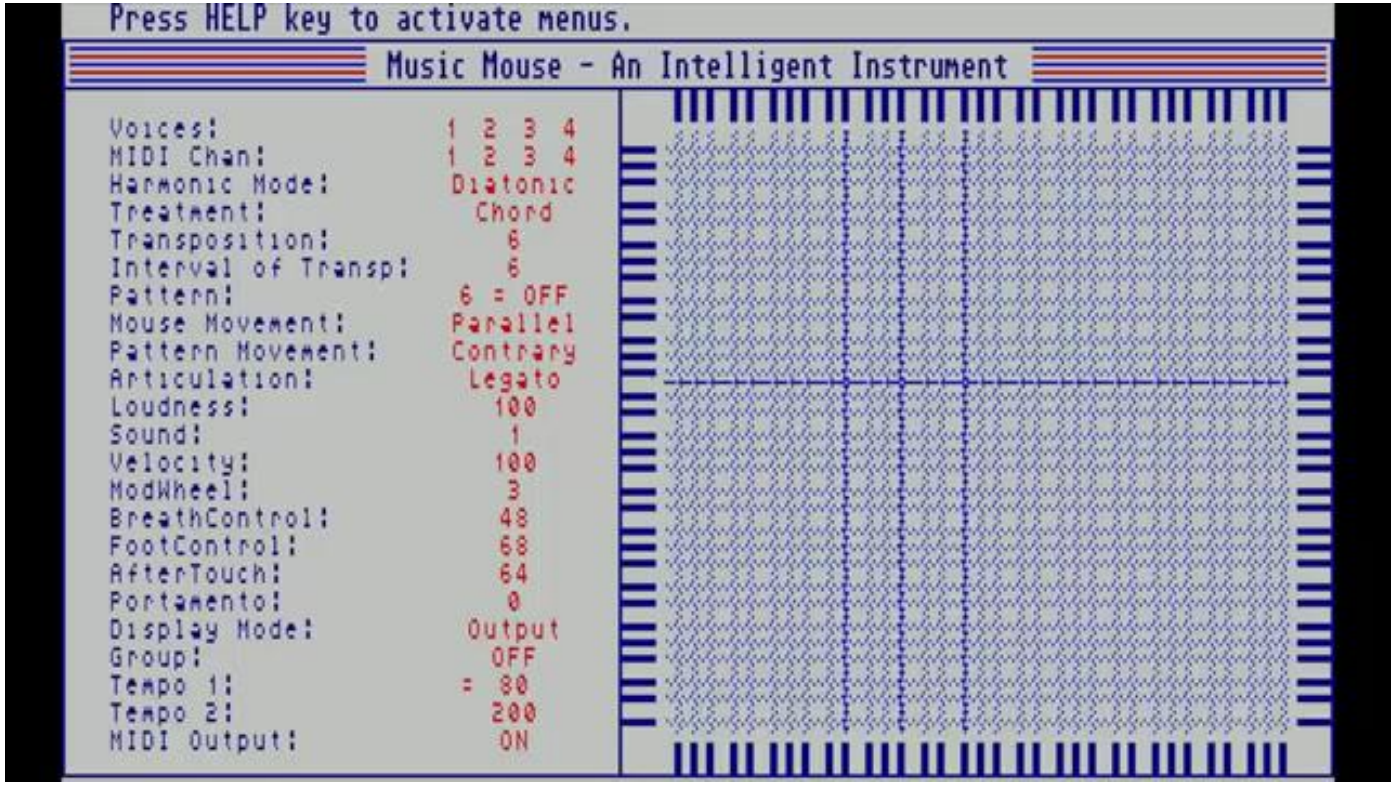






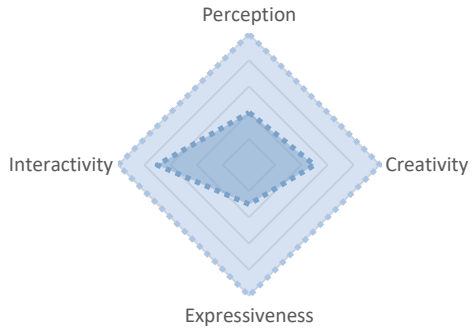


Music Mouse (http://musicmouse.com/) (Laurie Spiegel, 1986)

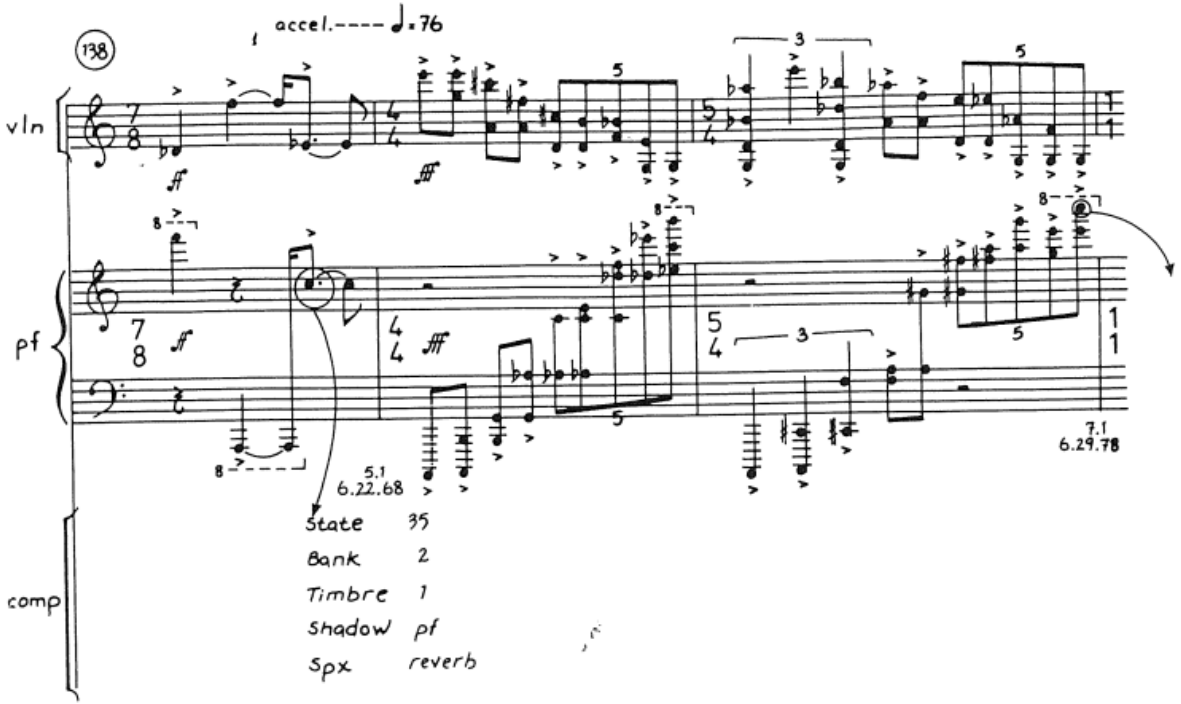


https://www.youtube.com/watch?v=D-mmEvGOopk

- 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.

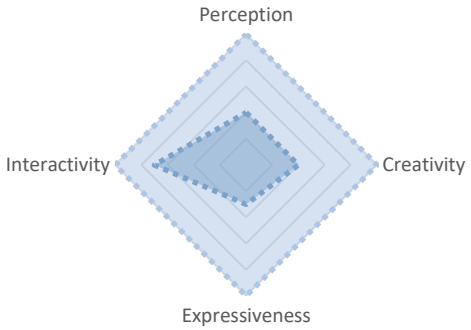


Cypher
(Robert Rowe, 1992)

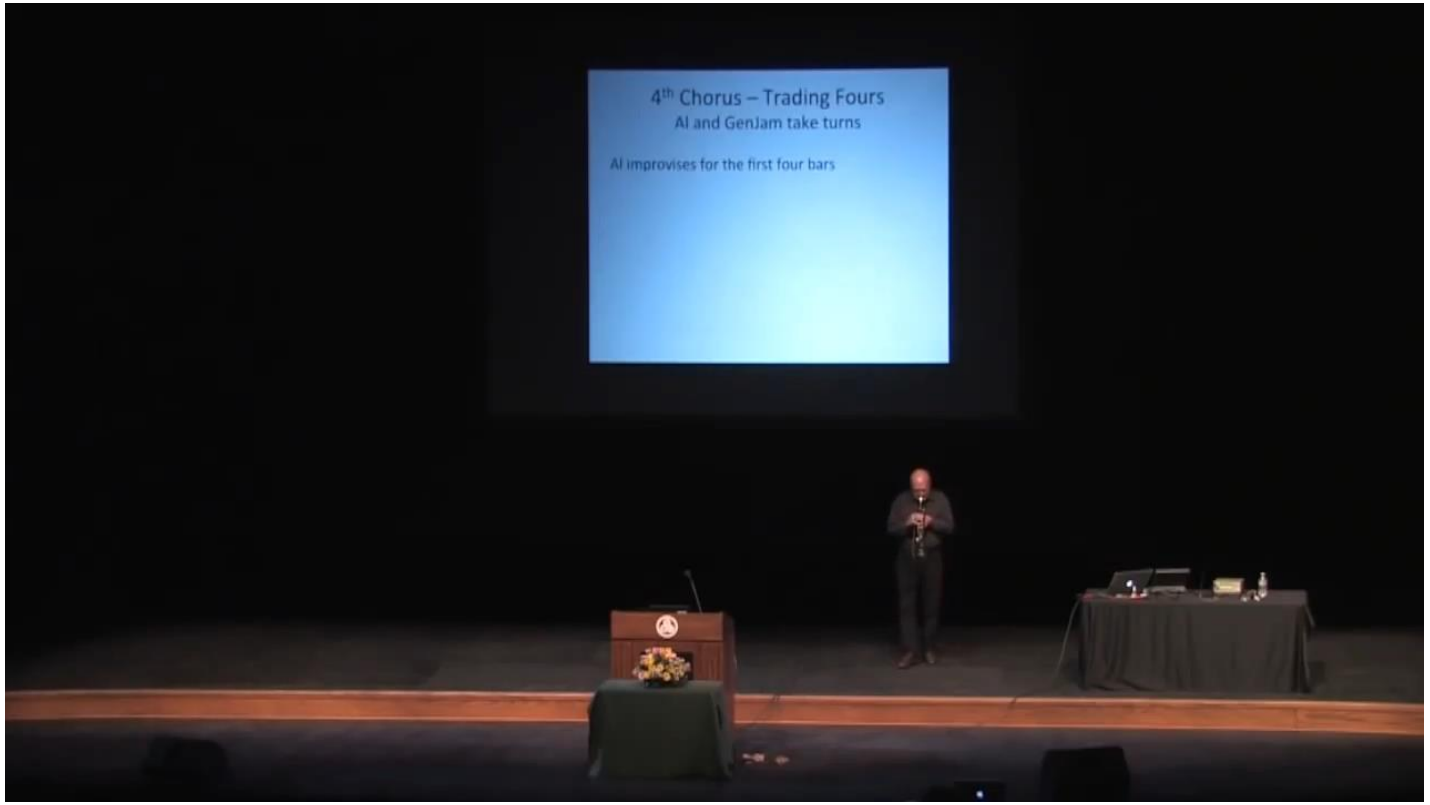


(Rowe, Interactive Music Systems, 1993)

- 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

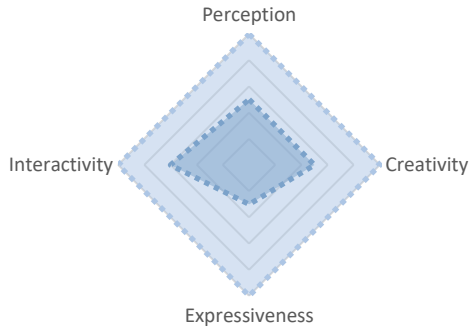


GenJam
(Al Biles, 1994)



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

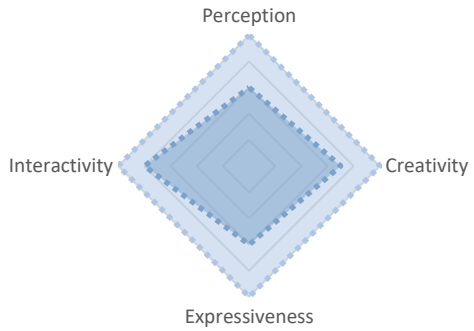
- Genetic algorithm for jazz improvisation (trade fours)
  - Listens to human's four measures
  - Maps to its chromosome representation
  - Mutates the chromosomes
  - Generates the next four measures



Voyager
(George Lewis, 1999)

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

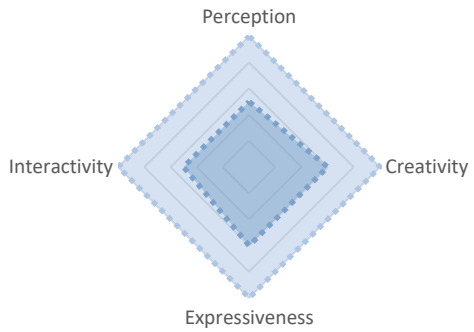
- Multi-agent system with stochastic selection of agent combinations
  - Listens to MIDI or acoustic data (e.g., tempo, note spacing, melodic interval width, primary pitch material, octave range, microtonal transposition, and volume)
  - Improvises on many musical aspects (e.g., timbre, volume, microtonal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

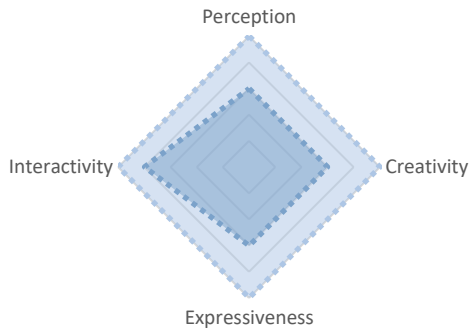
- Continuing music in the same style
  - Modeling user MIDI input sequences with a variable-order Markov model and builds pre-fix trees
  - Random traversals of trees to generate continuations



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

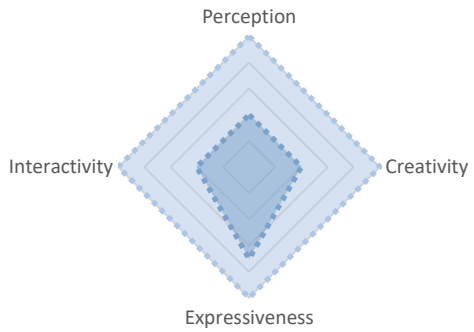
https://www.youtube.com/watch?v=2jFpGQbrcag

- 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

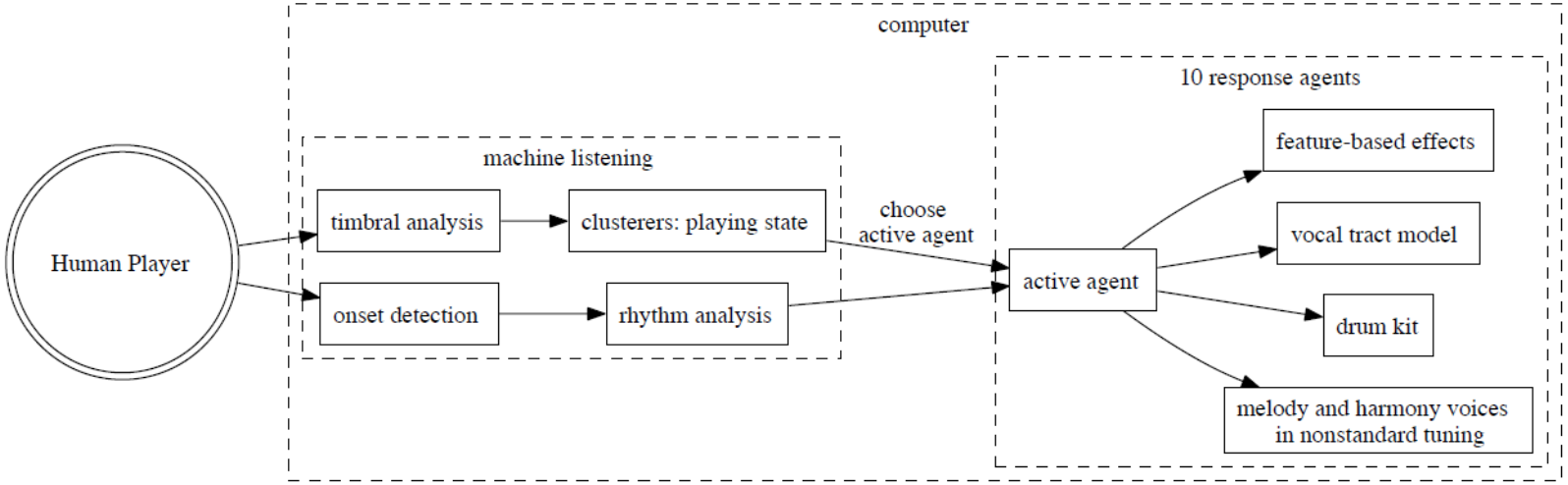


Shimon
(Hoffman & Weinberg, 2006)

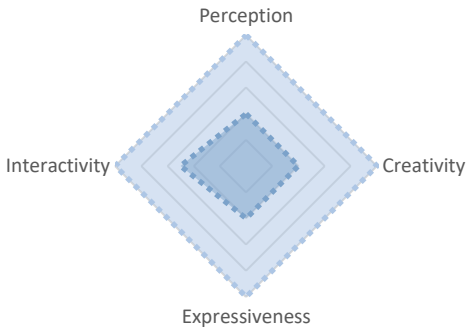
- A robotic marimba player for interactive improvisation
  - Physical embodiment greatly helps the audience to enjoy the performance
  - Beat tracking and chord matching to adapt to human's tempo variation
  - Improvisation centered around the choreographic aspect of the movement



LL (ListeningLearning) (Nick Collins, 2009)



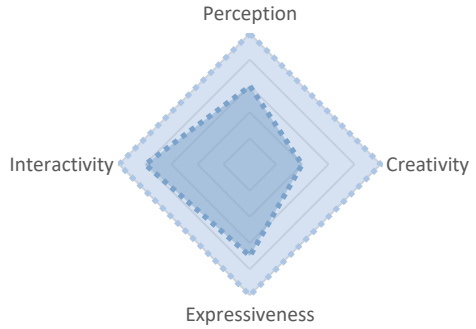
- Rule-based system for free improvisation with humans
  - Rhythm tracking: onset, inter-onset interval
  - Silence detection: perceived loudness
  - Timbral state clustering: using low-level acoustic features
  - Generation: choose among 10 agents to follow the human's timbral state



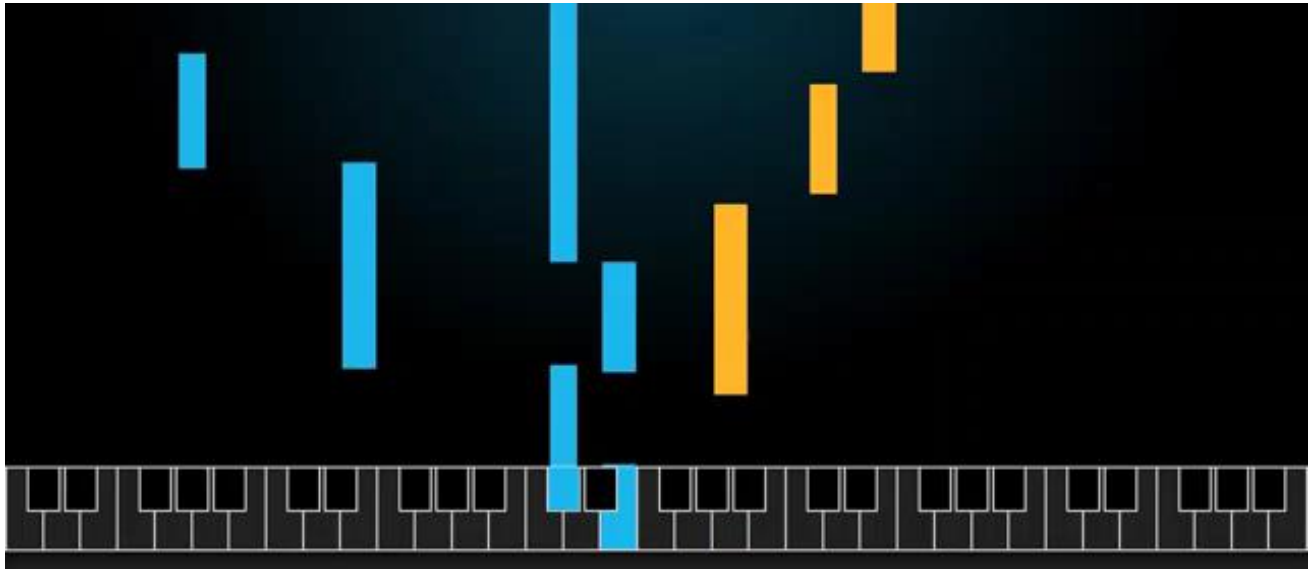
Reflexive Looper (Pachet et al., 2013)

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

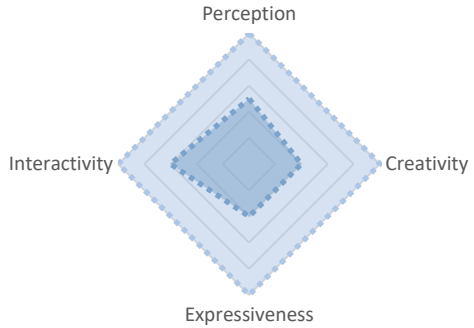
- A system allowing users to play with past virtual copies of themselves
  - Takes simultaneous MIDI and audio input: MIDI for analysis and audio for resynthesis
  - Uses an SVM classifier trained on MIDI data to classify the mode of user playing: bass, chords, and melody
  - Resynthesizes the other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



A.I. Duet
(Google Magenta, 2016)



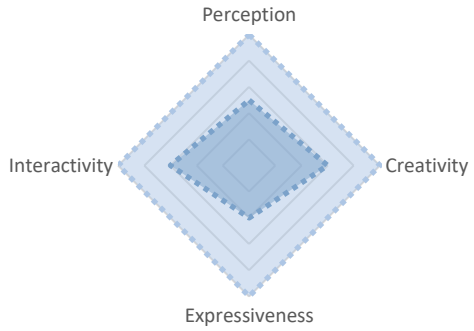
- A neural network model that responds to tunes played by the user on a MIDI keyboard using a similar style



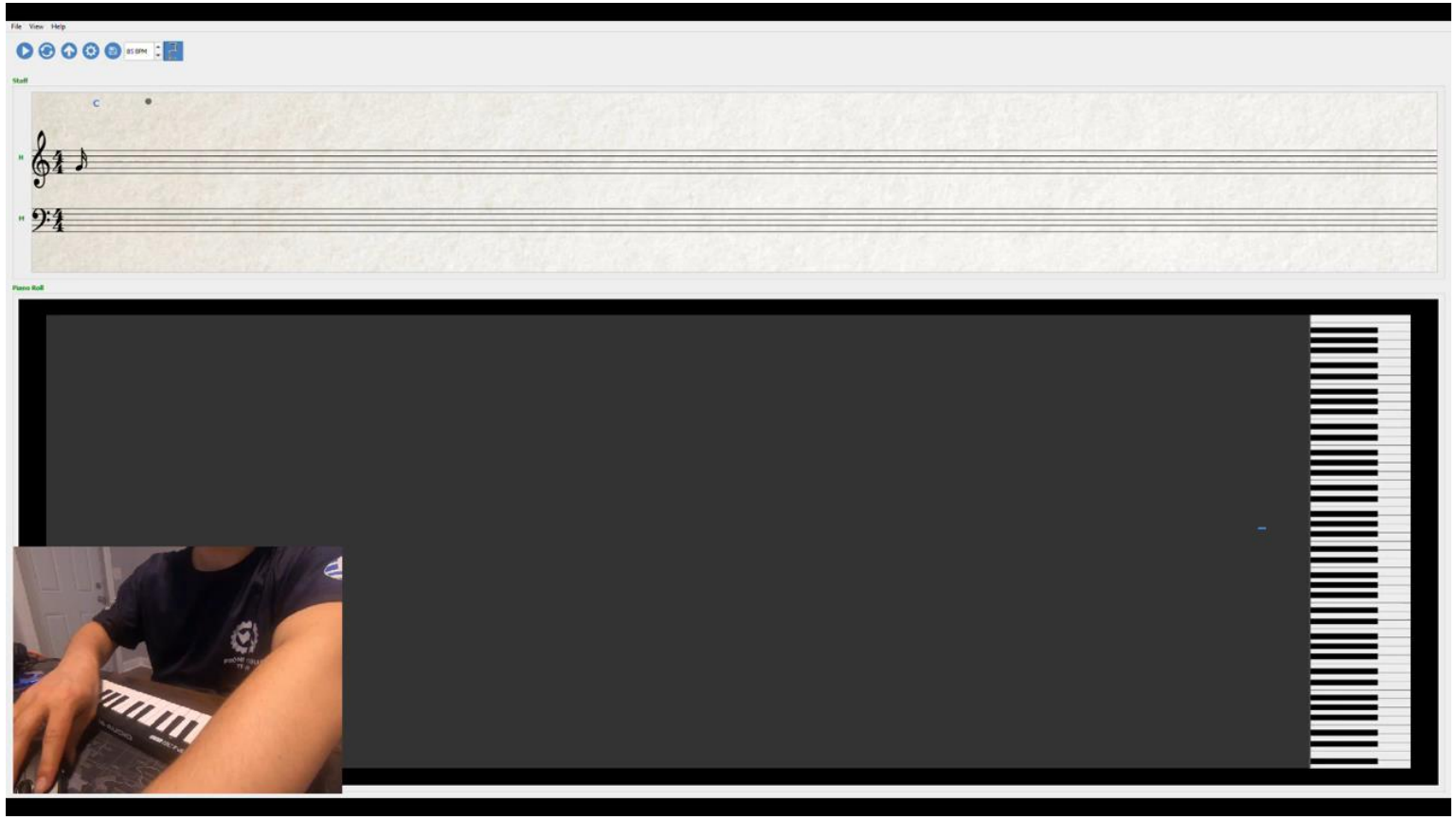
Piano Genie (Donahue, Simon, & Dieleman, 2019)

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

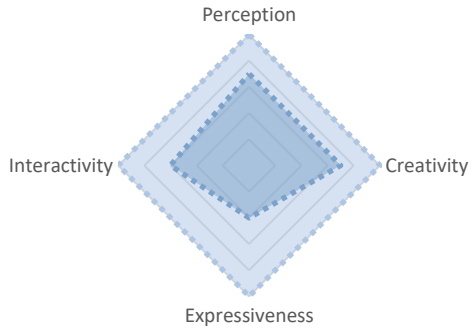
- Allowing users to improvise piano music on an 8-button controller
  - Uses an autoencoder to map note sequences in the 88-d space (corresponding to the 88 piano keys) to sequences in the 8-d space
  - Trained on 1400 piano performances by skilled pianists



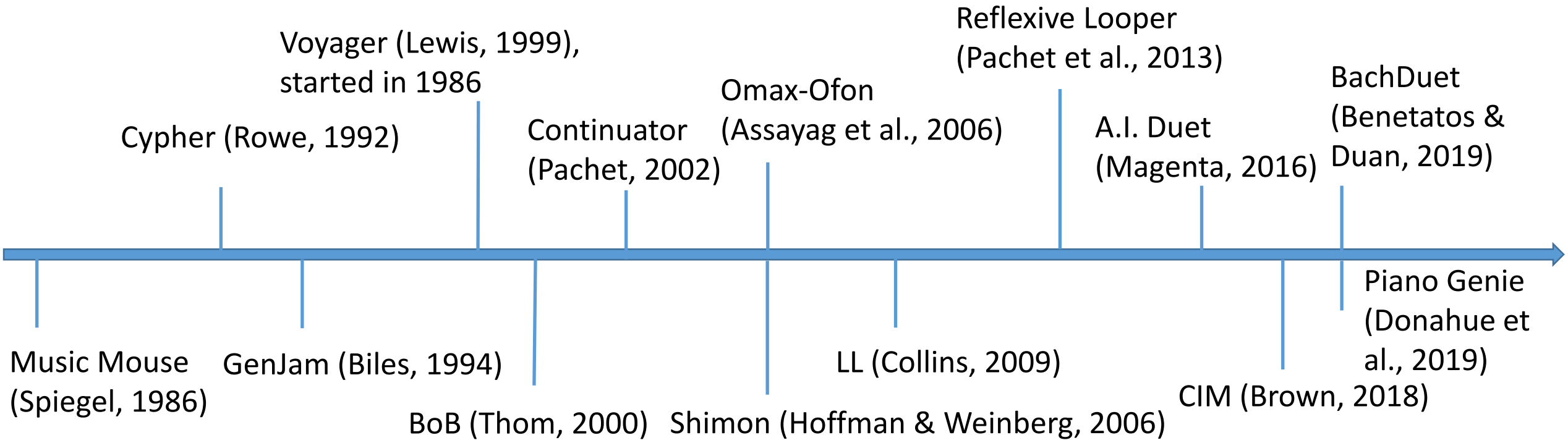
BachDuet (https://bachduet.com/) (Benetatos & Duan, 2019)



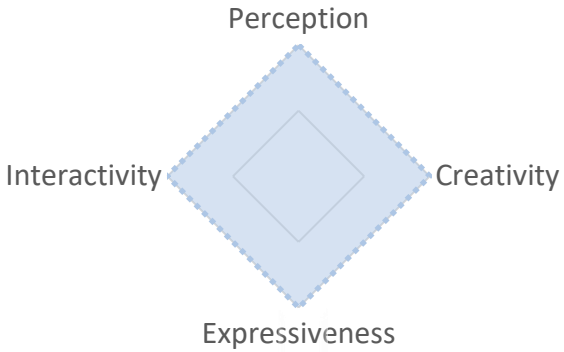
- A neural network based system to allow human-AI duet improvisation in the style of Western counterpoint
  - Trained on outer voices of 370+ Bach chorales
  - Relatively equal role between human and AI – 6:4
  - Only supports MIDI input and fixed tempo

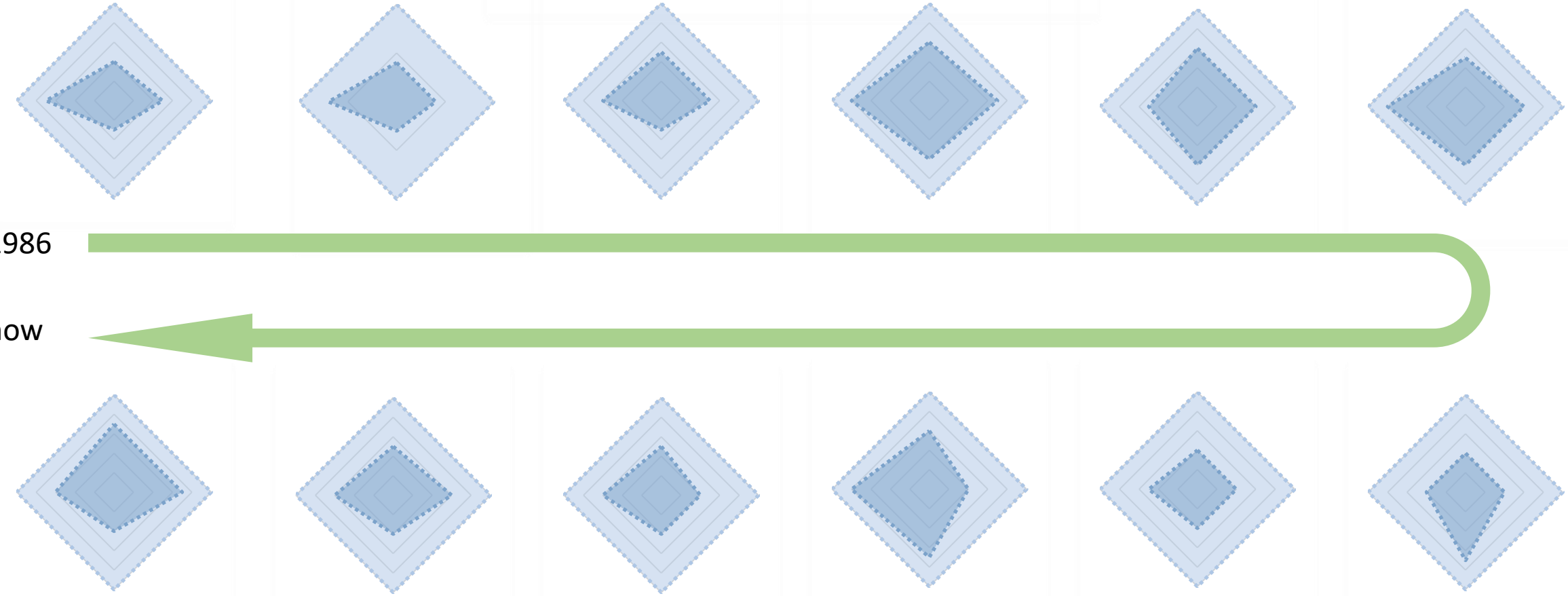


Interactive Music Systems
Beyond Automatic Accompaniment



The Trend?

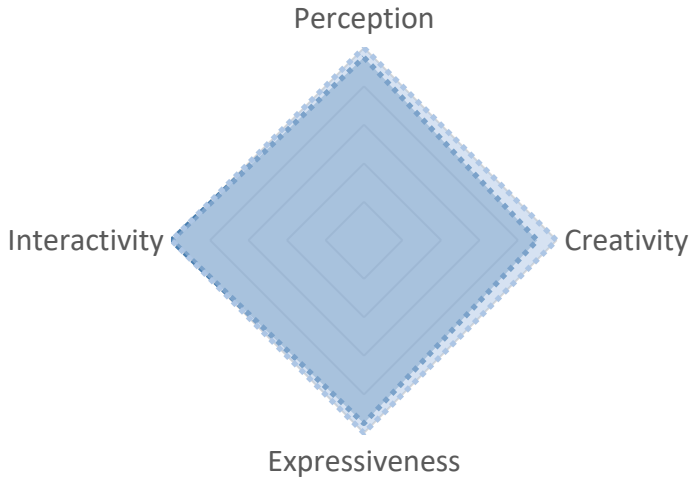

Perception
Interactivity
Creativity
Expressiveness


1986
now

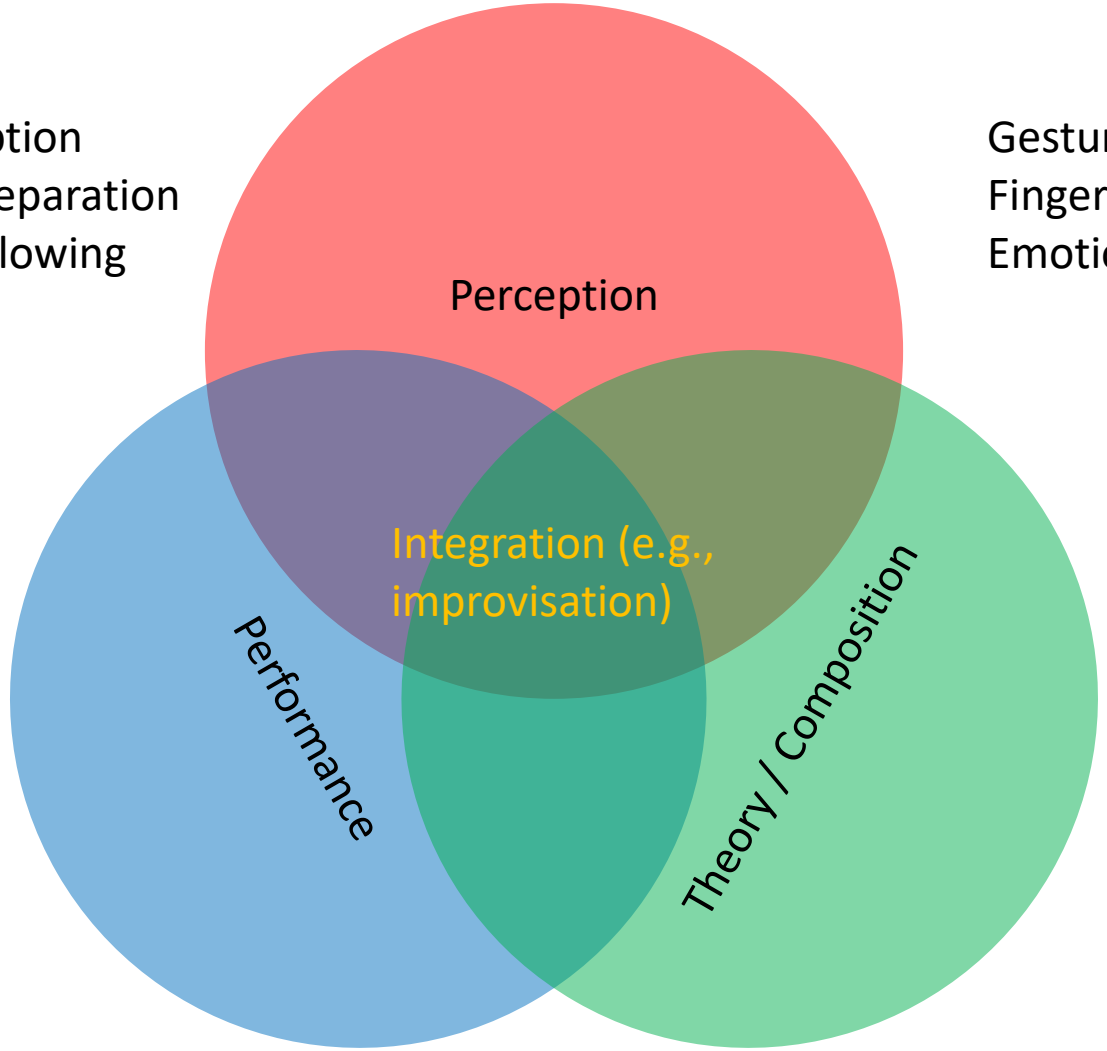
The Trend?

- 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

AUDIO INFORMATION RESEARCH