Joint bottom-up/top-down machine learning structures to simulate human audition and musical creativity

Jonas Braasch

Director of Operations, Cognitive and Immersive Systems Laboratory
Professor, School of Architecture
Rensselaer Polytechnic Institute, Troy, NY
CAIRA
THE CREATIVE ARTIFICIALLY INTUITIVE AND REASONING AGENT

Pauline Oliveros
Selmer Bringsjord

Cognition

Rational Thinking

goal driven
Top down
deductive

Intuitive Thinking

process driven
Bottom up
inductive

Sensation

Jonas Braasch
Hyperspecialization in Saxophone
CAIRA (started 2008)
THE CREATIVE ARTIFICIALLY INTUITIVE AND REASONING AGENT

Agent uses
- **Auditory Scene Analysis algorithms** to extract low-level acoustic features
- **HMM-based machine listening** tools for texture analysis
- **Genetic Algorithms** for the creation of new material
- **Logic-Based Reasoning** to make cognitive decisions
Cognitive and Immersive Systems Laboratory

- **Low Level**
  - Sensation
  - Auditory Sensing
  - Auditory Signal Processing

- **Mid Level**
  - Top down
  - HMM, EMD
  - Neural Networks

- **High Level**
  - Bottom up
  - 1st and higher order logic
  - Deep Neural Network

Diagram: Cognition, High Level, Mid Level, Low Level, Sensation, 1st and higher order logic, Deep Neural Network, HMM, EMD, Neural Networks, Auditory Sensing, Auditory Signal Processing.
Cognitive and Immersive Systems Laboratory

Low Level

High Level

Mid Level

Sensation

Auditory Sensing
Auditory Signal Processing

HMM, EMD
Neural Networks

Top down

1st and higher order logic
Deep Neural Network

Bottom up

Cognition
Duplex Pitch Perception Model

from Zigmond et al., 1999

Major chord with 1/f tone complexes (+880 Hz tone)

Braasch et al. Springer 2017
Binaural Activity Pattern

Step 1: Calculate 2 Autocorrelation functions

Step 2: Calculate Cross-Correlation function

Step 3: Cross-Correlate both functions (window) move by $k_d$

Step 4: Replace CC function with AC function


Motivation

Architecture

BICAM Localization Results using Deep Neural Networks

N. Deshpande
J. Braasch

Binaural Activity Pattern

Binaural Activity Map

<table>
<thead>
<tr>
<th>Task</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dir Lat (only)</td>
<td>98.2%</td>
<td>98.98%</td>
<td>98.6%</td>
</tr>
<tr>
<td>Dir Lat (w/refl)</td>
<td>94.0%</td>
<td>92.9%</td>
<td>92.4%</td>
</tr>
<tr>
<td>Time Delay</td>
<td>98.6%</td>
<td>98.4%</td>
<td>98.0%</td>
</tr>
<tr>
<td>Ref Lat</td>
<td>85.9%</td>
<td>85.9%</td>
<td>84.6%</td>
</tr>
</tbody>
</table>

Methods
Data Set: Reverberated Speech
BICAM output
Apple TuriCreate API
5 Lateral positions, 4 Delay conditions
Cognitive and Immersive Systems Laboratory

High Level

1\textsuperscript{st} and higher order logic
Deep Neural Network

Mid Level

HMM, EMD
Neural Networks

Low Level

Auditory Sensing
Auditory Signal Processing

Sensation

Bottom up

Top down
Sonic Gesture Recognizer

FILTER System Overview for audio including feature extraction, recognition and mapping to GA process.

Sampling of State Sequences by HMM Recognizer: all eight features are combined to form one state.

D. Van Nort, J. Braasch, P. Oliveros (2009), SMC
Cognition

High Level

Top down

Mid Level

HMM, EMD
Neural Networks

Low Level

Auditory Sensing
Auditory Signal Processing

Bottom up

1\textsuperscript{st} and higher order logic
Deep Neural Network

Sensation
**Computation of a tension arc**

**Task:**
- Agent will determine the solo/tutti Constellation in a Free-Music Ensemble
- Focuses on Tension data, which correlates strongly with Loudness

\[
T = L + 0.5 \cdot ((1-b) \cdot R + b \cdot I + O)
\]

with \(I\) the information rate, and \(O\) the onset rate. Note that all parameters, \(L, R, I, O\), are normalized between 0 and 1 and the exponential relationships between the input parameters and \(T\) are also factored into these variables.

Loudness (0.906 correlation)
Loudness & Roughness (0.915 correlation)

Braasch et al., Springer 2017
The agent needs to know about dynamic levels

We need at least two musicians to form an ensemble

At a given time slot, we have either a solo or an ensemble part

If we have only one musician, he/she must play a solo

\[ \forall x(Mx \rightarrow \forall t(Tt \rightarrow (d-l(x,t) = tacit \lor d-l(x,t) = pp \lor d-l(x,t) = p \lor d-l(x,t) = mp \lor d-l(x,t) = mf \lor d-l(x,t) = f \lor d-l(x,t) = ff])) \land (tacit < pp \land pp < p \land p < mp \land mp < mf \land mf < f \land f < ff)). \]

\[ \forall x \forall y((Mx \land My \land x \neq y) \rightarrow \exists z(Ez \land Ixz \land Iyz \land (\forall y(Ey \rightarrow y = z))). \]

\[ \neg \exists z Ez. \]

\[ \forall S \neg \exists x \text{Musician}(x) \land \text{PlaySolo}(x, S) \]

\[ \Leftrightarrow (\forall x \text{Musician}(x) \land \text{PlayEnsemble}(x, S)). \]

\[ \forall x((\text{Musician}(x) \land \forall y(\text{Musician}(y) \Rightarrow x = y)) \Rightarrow \text{PlaySolo}(x)). \]
MIKA Agent (RPI/IBM AIRC project)  


Cognitive and Immersive Systems Laboratory

Input

Model Analysis

Cognitive Functions
Top-down logic-reasoning using rule-based internal world representation

First-order logical reasoning based on Music Theory
Manual and semi-automated ontology creation

Mid-level Learning Function
Bottom-up driven statistical learning approaches

Auditory Periphery and Mid-Brain
Acoustic Sensing
Binaural manikin, Nearfield microphone

Pitch extraction
On-offset detection
Timbre analysis

Beat analysis
Loudness estimation
Polyphonic analysis

Output (knowledge acquisition and creation)

Understanding development of Theory and Performance Practice over time - Past Analysis

Analysis of theoretical works
Analysis of interpretation of selected jazz standard over time

Back Home Again in Indiana
Stella by Starlight
Summertime

Analysis of leadsheets and transcription – from MIDI file database

Recording analysis from database

1900
Blues
1925
Ragtime
Swing
Bebop
Cool
1950
Hardbop
1975
Fusion
Avantgarde
2000
Future Jazz

Future Projection
Extension of theoretical corpus
New forms of compositions and performance practice
New music generation

First NN results

Piano Roll:
Green: in scale notes
Red: out of scale notes

12-bar blues

Time

Pitch
NN:
- Multilayer Perceptron (MLP)

CNN:
- Convolutional Neural Networks (CNN)
- TensorFlow
- Inception-v3 pretrained ImageNet
- 4000 training steps
Training Data Collection

Synthetic Data Generation using Synthetic Sound Fields and/or MIDI orchestration

Interactive User Tracking In virtual Reality @ CRAIVE-Lab

24/7 Spatial AV/ Recordings AV recording at Lake George Jefferson Project

CISL infrastructure
J. Braasch
N. Keil
AIRCC project Team

J. Braasch, B. Chang, J. Goebel, R. Radke, Q. Ji
CISL project Team

Mallory Morgan, RPI
Vincent W Moriarty, IBM
Jonas Braasch, RPI
CISL and JP teams
Next Directions

• Combination of
  1. signal processing
  2. statistical models
  3. neural networks & logic
• Automatic creation of ontologies
• Automatic creation of acoustic scene databases
• Rule discovery in systems where *rule violations* are (i) allowed, (ii) strategically used, and (iii) changed over time.
• Multi-modal analysis