A creative artificially-intuitive and reasoning agent in the context of live music improvisation

Jonas Braasch[1], Doug Van Nort[1,2], Pauline Oliveros[2], Selmer Bringsjord[3], Naveen Sundar Govindarajulu[3], Colin Kuebler[3], Anthony Parks[1]

[1]School of Architecture, [2]Arts Department, [3]Department of Cognitive Science
Rensselaer Polytechnic Institute, 110 8th Street
Troy, NY 12180, USA

***Abstract*—This paper reports on the architecture and performance of a creative artificially-intuitive and reasoning agent (CAIRA) as an improviser and conductor for improvised avant-garde music. The agent's listening skills are based on a music recognition system that simulates the human auditory periphery to perform an Auditory Scene Analysis (ASA). Its simulation of cognitive processes includes a cognitive calculus for reasoning and decision-making using logic based-reasoning. The agent is evaluated in live sessions with music ensembles.**

Keywords—automated music improvisation systems, informative feedback models, artificial creativity, cognitive modeling, free music, auditory scene analysis

I. INTRODUCTION

Numerous attempts have been made to design machine improvisation/composition algorithms to generate music material in the context of various musical styles [Cop87, Fri91, Wid92, Jac96]. While Oliveros' *Expanded Instrument System* (EIS) acts on audio signals [Oli91, Gam98], in most cases these algorithms use a symbolic language, such as the Musical Instrument Digital Interface (MIDI) format, to code various music parameters. For example, Lewis' *Voyager* system [Lew00] and Pachet's *Continuator* [Pac04] work with MIDI data in order to interact with an individual performer. The system transforms and enhances the material of the human performer by generating new material from the received MIDI code, which may be derived from an acoustical sound source using an audio-to-MIDI converter (Typically these systems fail if more than one musical instrument is included in the acoustic signal.). In the case of the *Continuator*, a learning algorithm based on a Hidden Markov Model (HMM) helps the system to copy the musical style of the human performer.

Following these traditions, this paper describes an intelligent agent that was developed to perform music improvisations in the context of free music. In order to cope with free music, the agent simulates human listening using standard techniques of Computational Auditory Scene Analysis (CASA) including pitch perception, tracking of rhythmical structures, and timbre and texture recognition (see Fig. 1). It uses a Hidden Markov Model (HMM) to recognize musical gestures and Evolutionary Algorithms to create new material. Recently, the authors have begun to integrate a logic-based reasoning system into the overall architecture for a hypothesis-driven approach (see top-down processes in Fig. 1). The current musical output module of the system consists of presenting audio material that is a processed version of input sound which the agent picks up during a given session, or from audio material that has been presented to the agent in a prior live session. The material is analyzed using the HMM machine listening tools and CASA modules, restructured through the evolutionary algorithms and then presented in the context of what is being played live by the other musicians. Alternatively, the agent can also conduct a small ensemble using a graphic score and instructions that are updated live. In the following three sections, the basic architecture of CAIRA will be described, followed by a concrete performance example in Section V.

II. MICROPHONE-AIDED COMPUTATIONAL AUDITORY SCENE ANALYSIS (MACASA)

Stemming from the seminal work of Albert Bregman on the perceptual organization and grouping of sounds [Bre90], a body of work has arisen whose primary goal is the computational modeling of the mechanisms by which humans parse audio streams, grouping percepts into identifiable sound objects. This field of Computational Auditory Scene Analysis (CASA) [Ell90, Ros98] shares similar goals with the analysis stage of our project for several reasons. Primary among these is that contemporary improvised music often does not structure itself by classical paradigms of musical structure – key, meter, melodic progressions, etc. – but rather by working on the level of sound in a more direct, low-level (from a signal processing point of view) and visceral way.

One of the unsolved challenges in CASA is the robust separation of auditory streams from a complex sound mixture. Unfortunately, sound mixtures of music performances are among the most complex cases, and the typically long reverberation times in concert venues are an additional obstacle for robust CASA performance. To circumvent this problem, we separate the individual instruments electro-acoustically using closely positioned microphones for each participating musician. Additional room microphones can be utilized to automatically calibrate the individual microphone signal levels [e.g., see Bra11b], which is important when the inter-musician relationships need to be determined from these data.

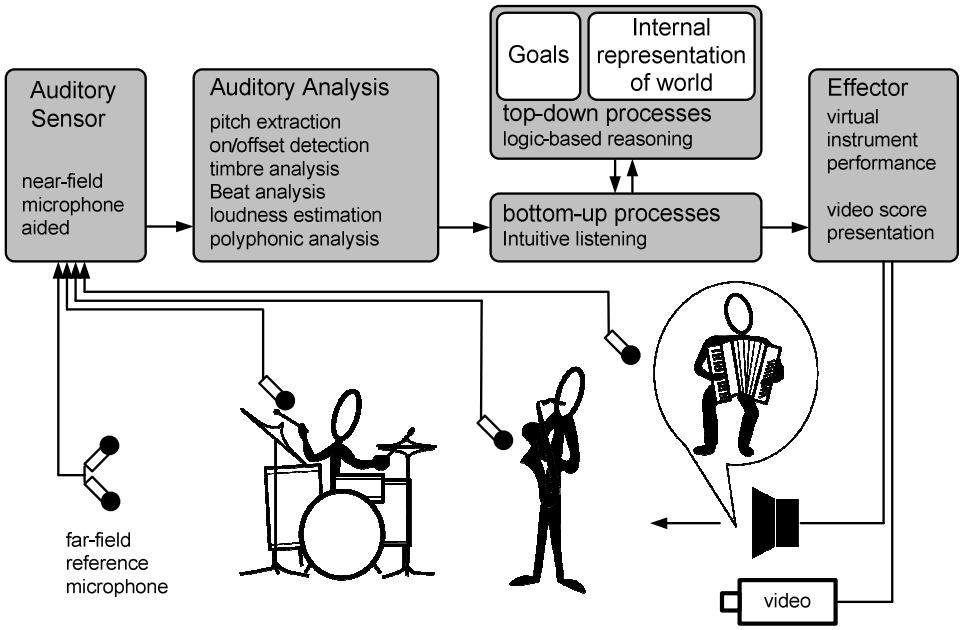

Figure 1. Schematic of the creative artificially-intuitive and reasoning agent CAIRA

III. GESTALT-BASED IMPROVISATION MODEL BASED ON INTUITIVE LISTENING

The artificially-intuitive listening and music performance processes are simulated using a Hidden Markov Model (HMM) for sonic gesture recognition and Genetic Algorithms (GA) for the creation of new sonic material [Nor09, Nor10]. In the first step of this stage, the system continually extracts spectral and temporal sound features. At the same time, onsets and offsets are tracked on a filtered version of the signal, which act as discrete cues for the system to begin recognizing sonic gestures. When such a cue is received, a set of parallel Hidden Markov Model (HMM) based gesture recognizers follow the audio, with the specific number of these being chosen as a product of needed resolution as well as processing power. The recognition continually provides a vector of probabilities relative to a "dictionary" of reference gestures. Processing on this vector extracts features related to maximum likelihood and confidence, and this information drives the fitness, crossover, mutation and evolution rate of a GA process acting on the parameter output space [Van09].

IV. LOGIC-BASED REASONING DRIVEN WORLD MODEL

A. Overview

In order to better understand the relationship between bottom-up and top-down mechanisms of creativity, a knowledge-based top-down model complements the bottom-up stages that were described in the previous two sections. CAIRA's knowledge-based system is described using first-order logic notation (for a detailed description of CAIRA's ontology see [Bra11]). For example CAIRA knows that every musician has an associated time-varying dynamic level in seven ascending values from *tacit* to *ff*. The agent also possesses some fundamental knowledge of music structure recognition based on jazz music practice. It knows what a solo is and understands that musicians take turns in playing solos, while being accompanied by the remaining ensemble. The agent also has a set of beliefs. For example it could be instructed to believe that every soloist should perform exactly one solo per piece.

One of the key analysis parameters for CAIRA is the estimation of the tension arc, which describes the current perceived tension of an improvisation. In this context, the term 'arc' is derived from common practice of gradually increasing the tension until the climax of a performance part is reached and then gradually decreasing tension to end it. Thus, tension often has the shape of an arc over time, but it can also have different time courses. It is noteworthy that we are not focusing here on tonal tension curves that are typically only a few bars long (i.e. demonstrating low tension whenever the tonal structure is resolved and the tonic appears). Instead, we are interested in longer structures, describing a parameter that is also related to *Emotional Force* [McA02].

Using the individual microphone signals, the agent tracks the running loudness of each musical instrument using the Dynamic Loudness Model of [Cha02]. The Dynamic Loudness Model is based on a fairly complex simulation of the auditory periphery including the simulation of auditory filters and masking effects. In addition, the psychoacoustic parameters of roughness and sharpness are calculated according to [Dan97] and [Zwi99]. In the current implementation, CAIRA estimates tension arcs for each musician from simulated psychophysical parameters. Based on these perceptual parameters and through its logic capabilities,

the system recognizes different configurations for various patterns, e.g., it realizes that one of the musicians is performing an accompanied solo, by noticing that the performer is louder and has a denser texture than the remaining performers. The system can also notice that the tension arc is reaching a climax when all musicians perform denser ensemble textures. CAIRA takes action by either adapting her music performance to the analysis results, or by presenting a dynamic visual score as described in more detail in the next section. CAIRA can, for example, suggest that a performer should end his or her solo, because it is becoming too long or it can encourage another musician to take more initiative. It can guide endings and help an ensemble to fuse its sounds together.

B. Tension Arc Calculation

In a previous study, we decided to calculate the tension arcs T from a combination of loudness L and roughness data R [Bra11a]:

$$T=L^4+a\cdot R^3,$$

with an adjusting factor a. In this paper, we also suggested to include *information rate* (e.g., as defined by [Dub03, Dub06]) as an additional parameter for the tension arc calculation. A real-time capable solution was developed measuring the rate and range of notes per 2-second time interval. To achieve this, pitch is measured using the YIN algorithm and converted to MIDI note numbers. Next, the number of notes is counted within a 2-second interval discounting the repetition of identical notes. The standard deviation of the note sequence is then determined from the list of midi note numbers. Finally, the information rate is determined from the product of *number of notes* and *standard deviation of MIDI note numbers*. Practically, we measured values between 0 and 100.

In addition, we measured the number of note onsets, by applying an envelope follower, calculating the rate of change of its output signal and then counting the incidents above a given positive threshold. A refractory period of 20 ms was applied, before the next onset is counted to avoid counting the same onsets multiple times. The tension curve is calculated using the following equation:

$$T=L+0.5\cdot((1-b)\cdot R+b\cdot I+O),$$

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. The parameter b is the quality factor from the YIN pitch algorithm. A value of one indicates a very tonal signal with a strong strength of pitch, while a value of zero indicates a noisy signal without defined pitch. The parameter is used to trade off roughness and information rate between tonal and noise-like signals.

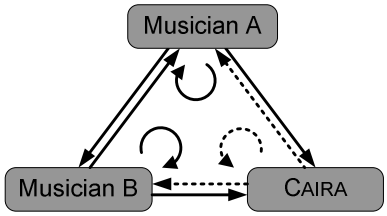

Figure 2. Schematic communication scheme for a free music performance. Each musician has to establish individual communication channels to all other musicians and also observe oneself. Dashed lines symbolize MaCASA enabled machine listening.

C. CAIRA's Self-Observation

Based on the tension arc data, CAIRA assesses the current state of the improvisatory ensemble addressing questions of who is playing a solo or whether it is likely that the improvisation will come to an end soon. For this purpose, the agent analyses the relationships between the tension arcs of each musician, including the tension arc measured from CAIRA'S own acoustic performance (see Figure 2). The robust measurement of individual tension arcs is possible, because each musician is captured with a separate microphone.

Figure 3. Video Still from *Configured Night*

V. CONFIGURED NIGHT

An example of the visual score produced by CAIRA was adapted from an audio-visual work titled *Configured Night*. The core idea of this piece is based on video footage of night scenery recorded from train rides. The material serves both as visual art work and visual score. A catalog of clips was created for the piece. Each clip starts and ends with a dark sequence, which regularly occurs when filming from a train at night, so that the clips can be arranged seamlessly in any order. For the piece, the various clips are categorized according to visual density, rate of change, object sizes, among others features. Figure 3 shows a few stills from the footage. The top-left figure is a very sparse scene from an Amtrak train ride along the Hudson river, the top-center still is taken from a train ride in Germany with camera focusing on the raindrop-sparkled window. The right image is taken from a train ride in Sendai, and characterized by numerous lights in very symmetrical arrangement. The footage can also be used to blend between different levels of concrete vs. abstract. The piece starts in a randomly selected train station and ends in another one in a different continent.

In our concrete example, CAIRA performs with Braasch (soprano saxophone) and Van Nort (GREIS [Nor10]) using sound material from Oliveros (Roland V-Accordion). Prior to the performance, CAIRA'S HMM module was trained on Oliveros' performance with the trio *Triple Point* (Braasch, Oliveros, Van Nort). For this purpose, acoustically isolated accordion tracks from a 14-minute clip of a trio session were used to feed the machine learning algorithm. During the performance, this material is transformed and played back based on a dialog with the two live musicians.

Visual leitmotivs exist for each ensemble scenario (e.g., improvisation start, low tension group performance, high-tension group performance, CAIRA'S solo, laptop solo, saxophone solo, improvisation end). Within the high-tension group performance mode, CAIRA also arranges rapid moving video fragments rhythmically to the music performance. While the score is not binding for the live musicians in this piece, it gives insight into the operation of CAIRA, and also provides useful feedback to understand the "intentions" and internal state of the intelligent agent.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1002851. The real-time implementation of the CAIRA system was written in Max/MSP utilizing various custom externals and abstractions as well as the FTM, Gabor and MnM packages from IRCAM, externals from CNMAT and Tristan Jehan's toolboxes (also using their loudness and roughness algorithms for a single-machine, stand-alone version of CAIRA).

REFERENCES

[Bra11a] Braasch, J. , Bringsjord, S., Kuebler, C., Oliveros, P., Parks, A., Van Nort, D. (2011) Caira – a Creative Artificially-Intuitive and Reasoning Agent as conductor of telematic music improvisations, Proc. 131th Audio Engineering Society Convention, Oct. 20-23, 2011, New York, NY, Paper Number 8546.

[Bra11b] Braasch, J., Peters, N., Van Nort, D., Oliveros, P., Chafe, C. (2011) *A Spatial Display for Telematic Music Performances*, in: Principles and Applications of Spatial Hearing: Proceedings of the First International Workshop on IWPASH (Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, H. Kato (eds.) World Scientific Pub Co Inc, ISBN: 9814313874, 436–451.

[Cha02] Chalupper, J., Fastl, H. (2002) Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acustica united with Acustica **88**, 378–386.

[Cop87] Cope, D. (1987). An expert system for computer-assisted composition, Computer Music Journal 11(4), 30–46.

[Dub03] Dubnov, S., Non-gaussian source-filter and independent components generalizations of spectral flatness measure. In Proceedings of the International Conference on Independent Components Analysis (ICA2003), 143–148, Porto, Portugal, 2003.

[Dub06] Dubnov, S., McAdams, S., Reynolds, R., Structural and affective aspects of music from statistical audio signal analysis. Journal of the American Society for Information Science and Technology, 57(11):1526–1536, 2006.

[Ell96] Ellis, D.P.W. (1996) Prediction-driven computational auditory scene analysis, Doctoral Dissertation, Massachusetts Institute of Technology.

[Fri91] Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system, Computer Music Journal 15(2), 56–71.

[Gam98] Gamper, D., Oliveros, P., "A Performer-Controlled Live Sound- Processing System: New Developments and Implementations of the *Expanded Instrument System*," Leonardo Music Journal, vol. 8, pp.33–38, 1998.

[Jac96] Jacob, B. (1996), Algorithmic composition as a model of creativity, Organised Sound 1(3), 157–165.

[Lew00] Lewis, G.E. (2000) Too Many Notes: Computers, Complexity and Culture in Voyager, Leonardo Music Journal 10, 33–39.

[Rus02] Russell, S., Norvig, P. (2002) Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River, NJ.

[Nor09] D. Van Nort, J. Braasch, P. Oliveros (2009) A system for musical improvisation combining sonic gesture recognition and genetic algorithms, in: Proceedings of the SMC 2009-6th Sound and Music Computing Conference, 23-25 July 2009, Porto, Portugal, 131–136.

[Nor10] Van Nort, D., Oliveros, P., Braasch, J. (2010) Developing Systems for Improvisation based on Listening, in Proc. of the 2010 International Computer Music Conference (ICMC 2010), New York, NY, June 1–5, 2010.

[Oli91] Oliveros, P., Panaiotis, "Expanded instrument system (EIS)," inProc. of the 1991 International Computer Music Conference (ICMC91), Montreal, QC, Canada, 1991, pp. 404–407.

[Pac04] Pachet, F. (2004) Beyond the Cybernetic Jam Fantasy: The Continuator, IEEE Computer Graphics and Applications 24(1), 31–35.

[Wid94] Widmer, G. (1994). The synergy of music theory and AI: Learning multi-level expressive interpretation, Technical Report Technical Report OEFAI-94-06, Austrian Research Institute for Artificial Intelligence.