Motivation, Microdrives and Microgoals in Mockingbird

Michael F. Lynch
Rensselaer Polytechnic Institute
lynchm2@rpi.edu

Abstract

This paper is a work-in-progress report about Mockingbird, an intelligent musical agent (IMA) based on Sun's Clarion cognitive architecture (Sun 2003). In the first part we present the Clarion architecture and the manner in which its Motivation Subsystem models drive states and goals. In the second part we propose a potential structure for modeling fine-grained secondary drives in the context of a free improvisational performance.

Introduction

The Mockingbird Project is a work-in-process in which we attempt to integrate the Clarion cognitive architecture (Sun 2003) with Van Nort's extended musical instrument FILTER (Van Nort, Braasch, and Oliveros 2012; Van Nort, Oliveros, and Braasch 2013; Van Nort, Braasch, and Oliveros 2009). Mockingbird is a musical accompanist and improviser that interacts with a live performer, capable of mapping and co-locating temporal events and building on them to manifest creative musical intuition.

In operation, Mockingbird listens to the audio output of a human performer, simultaneously recording that signal while performing extensive auditory analysis. From the analysis data, Mockingbird makes real-time decisions based on compositional- and performance-based metrics, and accompanies the live musician as a separate stand-alone performer. As befits its name, Mockingbird does the last by playing back excerpts of the performer's previous material while applying various contextually-appropriate transformations (time-stretching/compression, pitch shifting, spatialization, etc.) to that material.

System Architecture

System Components

As seen in Figure 1, Mockingbird consists primarily of five components: (1) auditory analysis module, (2) Clarion cognitive agent, (3) audio recording and playback module, (4) interface module that converts Clarion commands into performance outputs, and (5) output generation itself. Components #1, #3, #4, and #5 are constructed in Max, while #2 typically runs on a separate computer, with a fast bidirectional UDP/OSC connection between them.

The auditory analysis and output components all run on Max and are derived from and extend Van Nort's FILTER. The auditory analysis module processes the human performer's audio stream and extracts from it a number of musically salient features that form the basis for Mockingbird's subsequent musical responses. The audio recording and playback module provides the raw materials out of which Mockingbird constructs its accompaniment. The interface component maps the outputs of Clarion (so-called *action chunks*) to the performative commands that lead to its audio responses.

The Clarion Cognitive Architecture

Clarion is a modern, hybrid cognitive architecture developed by Sun and colleagues and grounded in cognitive psychology (Sun 2003; 2013) that attempts to model many aspects of human cognition, including learning, motivation, episodic memory, personality, and affective processing. The current release of Clarion is a software library suitable both for constructing cognitive models for experimentation and for constructing artificial intelligence applications.

Clarion is composed of four major systems: the Action Control System (ACS), Non-Action Control System (NACS), Meta-cognition System (MCS), and Motivation System (MS). Figure 2 shows the overall internal architecture of Clarion. All these subsystems incorporate both symbolic (i.e., localist, rule-based) and sub-symbolic (i.e., connectionist, artificial neural net) levels, arranged in an architecture that can be used to model a variety of human behaviors. Clarion has been the focus of a large number of research efforts within cognitive science, for which see (Sun 2013). The employment of Clarion in the context of musical performance is however, entirely new.

Clarion is most appropriately used for high-level learning and reasoning over a domain that is already fully featurized, that is, where lower-level processing of sensory data has already taken place. In Mockingbird the auditory analysis from FILTER acts as the auditory cortex to Clarion, which deals with learning, reasoning, and the generation of performance actions. Outputs from Clarion drive the output portion of FILTER, now in effect the motor system for Mockingbird, to create the performative responses themselves.

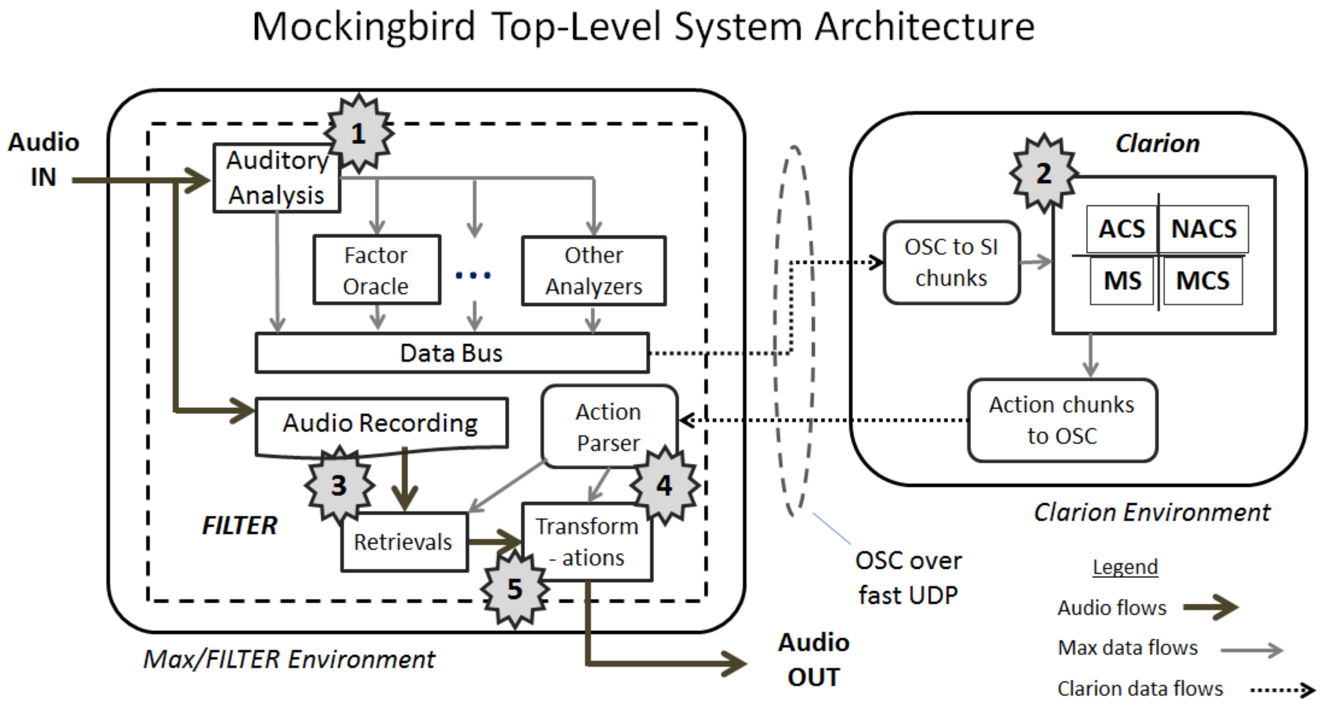

Figure 1: Overall Structure of Mockingbird

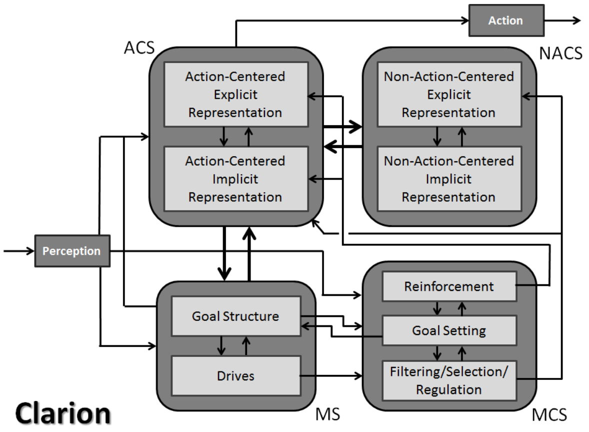
Action
ACS
NACS
Action-Centered Explicit Representation
Action-Centered Implicit Representation
Non-Action-Centered Explicit Representation
Non-Action-Centered Implicit Representation
Perception
Goal Structure
Drives
Reinforcement
Goal Setting
Filtering/Selection/ Regulation
Clarion
MS
MCS

Figure 2. Clarion Architecture

Clarion models factors of interest in the world as Dimension/Value (D/V) pairs. For example, the loudness of the incoming audio signal can be a Dimension that can assume one of these Values from the set of *(tacit, ppp, pp, p, mp, m, mf, f, ff, fff)*. Dimensions of interest in the modeling environment are "quantized" to a finite set of values which are then input to both the symbolic level of the ACS and on corresponding nodes in the neural networks in the subsymbolic level. (It is also possible to interpolate beween input node activations by proportionally firing adjacent neurons that bracket the input value.) A named collection of D/V pairs is referred to as a "chunk."

Auditory Analysis Module

FILTER (Van Nort, Braasch, and Oliveros 2012; Van Nort, Oliveros, and Braasch 2013; Van Nort, Braasch, and Oliveros 2009) is an improvisational hyperinstrument that incorporates machine listening and learns information embedded in the fine-structure features and sonic gestures of the incoming audio stream from its improvisational partners over a moving time window, yielding a number of metrics from its signal processing submodules. These then form the basis for unsupervised learning in FILTER, enabling it to accompany the performer in real time.

In Mockingbird, however, the featurized information extracted from the audio input instead become the sense data to the Clarion agent, while "action chunks" from Clarion drive the output module of FILTER to create the actual musical response. The internal linkage between FILTER's input processing and output processing is thus in effect, decoupled.

One important metric is *tension*, itself partially derived and manifested from several metrics such as physical energy (which mainfests as amplitude or "loudness") and roughness. These signal processing metrics are discussed in (Van Nort, Braasch, and Oliveros 2010). A partial list of these extracted features is shown Table 1.

Training Materials

The neural networks in the lower level of the ACS must be trained prior to the performance. For this we are using multitrack masters from the Triple Point ensemble (Pauline Oliveros, Jonas Braasch and Doug Van Nort) to be used to pre-

Feature	Meaning
Loudness	Overall signal amplitude
Pitch Class	vector of pitch activations
Pitch	loudest detected pitch
Root Pitch	lowest detected pitch
Spectral Energy	vector of spectral energy bands
Spectral Centroid	weighted midpt. of spectral energy
Spectral Deviation	spectral spread relative to centroid
Energy	RMS energy of the signal
Onset	indicates start of a sonic gesture
Offset	indicates conclusion of a gesture
Density	rate new onsets are generated
Noisiness	level of unpitched content
Tension	(see text)
Dissonance	ratio of dissonance to consonance
Roughness	level of non-harmonic content

Table 1: Partial List of Extracted Features

pare the training sets. The selfsame audio analysis module described above is used in the preparation of these training sets (essentially, large CSV files).

In this first iteration of Mockingbird, the actions generated by Clarion indicate points in the previously recorded audio stream beginning at some particular sample and having a known duration, along with any tranformations (pitch-shift, time-stretch, etc.), based on what Clarion has decided is the most satisfactory response it can make from what it has, in effect, listened to so far. This step can entail some degree of stochastic selection (as a tunable parameter) from among candidate actions as well. Mockingbird need not necessarily issue performance gestures continually; it ought to be able to remain silent if it determines silence is the best response in that moment. There may also be intentional inaction to allow a recently activated response gesture to run to completion.

We are using what we believe is a novel approach for generating the reinforcement signals for determining appropriate responses used to train the networks. While moving through the material from start to end, the reinforcement signal to apply to any given time point is obtained from a corresponding time point a short time in the future. For example, when processing the material at time 3:05 (mm:ss), the reinforcement signal to be applied is taken from the analysis at 3:07.5. That is, what the human performer actually did 2.5 seconds later in the piece is taken as the “correct” behavior to generate when the musical context resembles what it was like at 3:05. Obviously, this example time interval is only one possible value among many and we are experimenting with several different lead time values.

The Clarion Motivation Subsystem

The Motivation Subsystem (MS) in Clarion can provide intentionality by modeling the agent’s internal drive states and goals, supplying the agent with motivated behavior that goes beyond merely reactive. Motivations are not externally set but are internally generated.

The MS’s subsymbolic level essentially models *drives* within the agent. Following Reiss (2004), Sun (2009) holds that drives in humans can be factored into essentially 17 orthogonal primary drives, listed in Table 2. Drives in Clarion are modeled as neural networks and are not generally regarded as cognitively available (available to direct conscious examination). They only become so indirectly when an elevated drive level prompts the agent to switch goals, since goals reside in the top level and *are* cognitively available.

Physiological	Social
Food	Affiliation/Belongingness
Water	Dominance/Power
Sleep	Recognition/Achievement
Avoid physical danger	Autonomy
Reproduction	Deference
Avoid unpleasant stimuli	Similance
	Fairness
	Honor
	Nurturance
	Conservation
	Curiosity

Table 2: Primary Drives

Six of the 17 primary drives are physiological in nature while eleven are social in nature (Sun 2009). By default, all 17 primary drives are available in the MS, but it is up to the modeler to decide which of them will be of importance within a given simulation or application (Sun and Wilson 2011; Wilson 2012).

A given drive has associated with it a set-point, or threshold, along with its actual, temporally changing level. Drive levels can thus be considered somewhat “analog,” in that they are modeled as floating point values that vary over time in response to changes of state both in the enviroment and internal to the agent.

When a drive state reaches the point where it exceeds its threshold value it may (subject to surrounding context) trigger a change in the goal being pursued by the agent, causing the agent to engage in different behaviors able to reduce the drive state to a more tolerable level. Different goals are mapped to different sets of possible actions available for the agent to select toward satisfying that goal.

The symbolic, rule-based level models the available *goals* within the agent. A Clarion agent holds a set of all possible goal states in a so-called *goal list*. Only one goal can be in effect at a time, representing the intentional state of the agent toward its attentional target at that moment.

All the while the agent is in a given goal state, other drive levels are continually changing. When the current drive level being addressed by the current goal falls below its threshold, some other elevated drive state can then cause some new goal to be selected. It is also possible for the agent to reach a state of temporary quiescence if no drive state is elevated enough to cause the activation of a new goal.

We hope to demonstrate that the pattern of setpoints in the vector of thresholds over the drives has considerable influence on the resulting overall behavior of the agent.

Secondary Goals in a Musical Context

We can now ask, what sorts of drive and goal states do musicians move through as the piece unfolds? In particular, what possible, more fine-grained drives might bear on such performance? Certainly, different levels of primary drives can affect performance (e.g., a fatigued musician may well be motivated to perform differently from a well-rested one), but the primary drives by themselves do not get at the more purely local and transitory goals that an agent might pursue during the course of a performance, which we here are calling microdrives and microgoals. We are thus more interested here in the so-called secondary drives which are, theoretically, boundless in number.

We hypothesize that these secondary drives operate at a more fine-grained level, in this case, over the course of a performance. For example, a drive for novelty could continue to rise during long stretches of relatively unchanging material until such time as it triggers a change in the goal state, thus prompting Mockingbird to perform something novel but still in keeping with the state of the performance up to that point. Similarly, a lengthy passage of mostly upper register sounds could be arranged to cause the agent to reach a new goal of injecting a new lower-register phrase.

Complementarity of Drive States

We know of no literature that attempts to catalog possible secondary drives in quite these terms. As a starting point, we regard the auditory analysis metrics as an initial set of performance factors that can be mapped to corresponding sets of drives and goals. One eventual research objective of this work is to compile such a catalog of various drive states and goals based on experimentation and evaluation of their performative consequences.

One such drive, already hinted at, concerns the complementary notions of tension and relaxation (Braasch et al. 2012). Over the arc of a performance piece, tension and relaxation alternate in accordance with dynamics that are not always directly accessible. In free improvisation, the alternations between tension and relaxation are determined by the mutual interaction and interplay of the performers, from which the listener might discern a resulting ebb and flow within the piece. The presence of alternating phases of increasing and decreasing tensions within a free improvisation work is well-recognized and so form the first drive we are examining in this research.

Consider the drive states: *drive-to-tension* and *drive-to-relaxation*. We suggest using paired competing complementary drives, not a single drive where "low" somehow means relaxation and "high" somehow means tension. During a performance the agent will, for example, during periods of relative quiet, experience rising levels of *drive-to-tension*. Eventually, if the piece continues to be relatively quiet, the threshold for *drive-to-tension* is reached, allowing a switch in the goal state in the agent toward producing more tension-raising material, thus helping to satisfy *drive-to-tension*. At some point the level of *drive-to-tension* falls below its threshold (since its corresponding goal has been satisfied) and some other goal state can later be selected and made dominant.

The operation of any of these complementary drive pairs is as follows. Assume that threshold values and drive levels are floating point numbers clamped between 0.0 and 1.0. At the start of the performance, the tension and level are both at 0.0. Assume further that the drive threshold for *drive-to-tension* internal to the agent is set to some value, say, 0.8, and the *drive-to-relaxation* set to, say, 0.3, as shown in Figure 3.

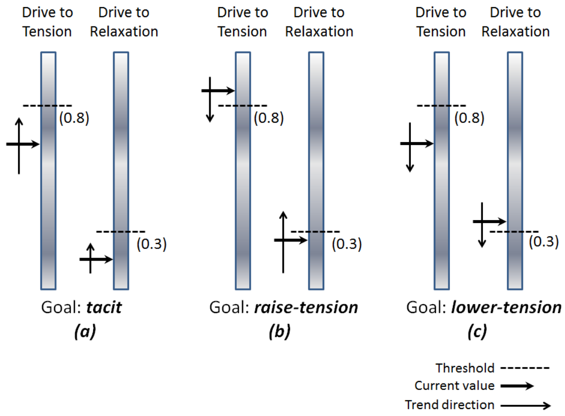

Figure 3. Changing Drives and Goals

Assuming the pieces starts off quietly such that the actual level for *drive-to-tension* soon reaches its threshold. The agent can now select a new goal, that of `raise-tension`. The agent will now behave so as to increase tension in its resulting musical output, resulting in an actual increase in the overall tension of the piece, whether or not the human performer goes along (moreso if she does). With the actual tension value now rising, the value of the *drive-to-tension* level starts to fall, since it is now being satisfied.

The tension-increasing goal is still in effect, however, until pre-empted by some other over-threshold drive state that triggers another goal switch. This does not necessarily take place immediately, and that other goal is not necessarily the goal of `lower-tension`. At some point, however, the agent will switch to a new goal and thus to a new set of behaviors that can work to satisfy whichever drive state triggered the switch.

Continuing, unless the `lower-tension` goal had been selected, the agent continues to output high-tension material. As this is occuring, the *drive-to-relaxation* continues to rise until it does go over-threshold, and, when eventually selected, now triggers the goal switch that brings about new tension-lowering behaviors.

This is illustrated in Figure 3. In *(a)*, at the start of the piece, Mockingbird's initial goal is `tacit`, that is, generate no output (it is however accumulating musical content from the human performer's output). The differential rates of tension rising faster than relaxation indicate that the human performer is currently creating relatively low-tension material and thus the *drive-to-tension* rises more rapidly. Then, as still shown in *(a)*, the set-point for *drive-to-tension* begins to approach its threshold of 0.8. In *(b)* this set-point has in fact now risen to above the threshold, and this leads to the selection of a new goal, `raise-tension`. Mockingbird now begins to create high-tension responses and the trend-line for the tension drive now heads downward. Meanwhile, the drive to relaxation, previously rising at a relatively slow rate, now rises more rapidly. In *(c)*, the value for the *drive-to-relaxtion* has now exceeded its set-point, and now Mockingbird switches to yet another goal, that of `lower-tension`.

In our formulation, it is the competition between two opposing drive states, rather than high/low values along a single drive state dimension, that accounts for the emergence of alternating patterns of tension and relaxation. Moreover, the settings of the threshold values directly determine the nature of the resulting performance. Setting the tension threshold low and relaxation high should result in a more aggressive style of performance (as with the above settings), since the agent spends more time dealing with the more easily triggered *drive-to-tension*. Also, the gap in the two thresholds will to some extent determine the rate at which the agent tends to cycle between tension-increasing and tension-reducing goals. We expect these thresholds to be an important determiner of the agent's overall performance style.

Note that the above example considers only these two drives/goals in isolation; in practice there would be a number of other complementary pairs whose values are constantly changing and whose goals are also in competition for selection. This same paired-drive approach can be used with other drive states, some examples of which include: dissonant/consonant, sparse/dense, busy/static, etc. A second set of drive states can target the insertion of material in several frequency bands (e.g., in bands roughly corresponding to conventional bass, baritone, tenor, alto, soprano registers). A third possible drive concerns where in the piece the performance is situated, such that the behaviors expected of the agent at the beginning of the piece differ from the behaviors expected of the agent as the piece is moving toward a close. This would be a representation of an overall performance "arc" present through the piece taken as a whole.

Also, we expect some of the actions issued as a result of goal selections will follow a "latching" behavior (i.e., keeping to a particular performative behavior until expliticly countermanded by its opposite) while others do not.

Thus, the vector of drive states, the current goal, and the ongoing latched and unlatched behaviors together comprise a rich context within which Mockingbird is constrained to make its gestural selections. In this fashion, through these evolving patterns of dynamical constraints, along with other decision-making actions coming out of Clarion, we hope to demonstrate that Mockingbird is able to accompany a human performer in a musically satisfying manner.

Conclusions and Future Work

We are currently developing Mockingbird with a targeted goal of demonstrating a live performance in early 2015. The initial implementation of Mockingbird will match a single agent to a single human performer. From there we intend to develop multi-agent / multi-human performers.

A second line of development is to engage the other capabilities in Clarion. In this early stage of this work, we are employing only the ACS and the MS, leaving the NACS and the MCS for later. Incorporating the NACS (in particular its General Knowledge Store (GKS)) would allow Mockingbird to retain and use much more material than is generated during a single live performance. The MCS is also of interest, since it is able to impose long-term "retunings" of the drive state threshold setpoints themselves. Finally there is Clarion's episodic memory, a future capability still under development.

Acknowledgments

Mockingbird has been funded through a Rensselaer Seed Grant in 2012. We also note the contributions of Jonas Braasch, Selmer Bringsjord, Nikhil Deshpande, Zach Layton, Pauline Oliveros, Ron Sun, Doug Van Nort.

References

Braasch, J.; Van Nort, D.; Oliveros, P.; Bringsjord, S.; Sundar Govindarajulu, N.; Kuebler, C.; and Parks, A. 2012. A creative artificially-intuitive and reasoning agent in the context of live music improvisation. In *Music, mind, and invention workshop: creativity at the intersection of music and computation*.

Reiss, S. 2004. Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology* 8(7):179–193.

Sun, R., and Wilson, N. 2011. Motivational processes within the perceptionaction cycle. In *Perception-Action Cycle*. New York: Springer. 449–472.

Sun, R. 2003. A tutorial on clarion 5.0. Technical report, Rensselaer Polytechnic Institute.

Sun, R. 2009. Motivational representations within a computational cognitive architecture. *Cognitive Computation* 1(1):91–103.

Sun, R. 2013. https://sites.google.com/site/clarioncognitivearchitecture. Technical report.

Van Nort, D.; Braasch, J.; and Oliveros, P. 2009. A system for musical improvisation combining sonic gesture recognition and genetic algorithms. In *Proceedings of the 6th Sound and Music Computing Conference*, 131–136. Sound and Music Computing Conference.

Van Nort, D.; Braasch, J.; and Oliveros, P. 2010. Sound texture analysis based on a dynamical systems model and empirical mode decomposition. In *ACM Convention 129*. Audio Engineering Society.

Van Nort, D.; Braasch, J.; and Oliveros, P. 2012. Mapping to musical actions in the filter system. In *Proc. Of International Conference on New Interfaces for Musical Expression (NIME"12)*. New Interfaces for Musical Expression.

Van Nort, D.; Oliveros, P.; and Braasch, J. 2013. Electroacoustic improvisation and deeply listening machines. *Journal of New Music Research* 42(4):303–324.

Wilson, N. 2012. *Towards a psychologically realistic comprehensive computational theory of emotion*. Ph.D. Dissertation, Rensselaer Polytechnic Institute.