Belinda Thom — Research Statement

My research is aimed at creating intelligent agents that collaborate—as first-class citizens—with humans in creative art-based domains. My interests are interdisciplinary. In my graduate career, I have specialized in applying Artificial Intelligence (AI) and Machine Learning (ML) techniques to the customization of real-time, interactive, computer music systems. The motivation has been to enable the computer's equal participation in live, improvised musical exchanges—to bring the computer out of the box, so to speak.

Thesis System

For my Ph.D. thesis, which I plan to defend in May 2001, I am developing Band-OUT-of-a-Box (BoB), an agent that trades improvised solos with a musician in real-time. This work is the first to introduce the domain of *improvised melodic companionship*, where the goal is to build an agent that is your personal musical friend, an agent that trades licks with you and enhances your ability to capture and experiment with transient creative processes when improvising alone.

This thesis investigates the hypothesis that an improvised musical companion can be constructed by developing methods that automatically personalize (customize) the agent's musical behavior to that of the musician. This work introduces a probabilistic computational model of improvisation whose parameters are estimated from the musician's improvised examples. Specifically, I developed an unsupervised learning procedure that algorithmically defines—i.e. *operationalizes*—the user's "musically-appropriate" behavior. This operationalization is decomposed into two core technologies:

- **Customized Perception:** Estimate the user's distinct "modes-of-playing" by partitioning the bars of their solos into distinct clusters (e.g. playing styles, scalar preferences, bags-of-licks, musical intentions, etc.). Abstract musical perception amounts to converting a bar of a solo into a data point, whose features record its average tonal, intervallic, and directional trends. BoB assigns data points to those clusters that are most likely to have generated them.
- **Customized Generation:** The per-bar sequence of abstractions that were perceived for the musician's solo call drives the generation of BoB's solo response. This capability amounts to algorithmically reversing the perceptual many-to-one mapping. Specifically, the average behavior of an entire cluster's tonal, intervallic, and directional trends are used to control a random walk through a graph of note nodes.

This research provides an important contribution: learning is used to operationally specify the musician's ill-specified artistic notions. Most importantly, this learning is coupled to the generation of new solos, and closing this loop is a crucial prerequisite to first-class improvised exchanges.

In contrast to more typical approaches, BoB focuses on learning musical concepts from *small* data sets. Rather than learning from a large corpus (e.g. the J.S. Bach chorales), BoB learns about a particular musician acting in a particular setting (e.g. learn the musician for this song). One way I have addressed the issue of over-fitting that can occur with small training sets is to use a Bayesian approach, which provides a principled way to incorporate prior beliefs. Furthermore, I use an explicit probabilistic model whose features are independent, which reduces the complexity of the learned model. A notable distinction of this approach to music is that I use a semi-parametric modeling method (the mixture model). By fitting functional forms to smaller, more localized partitions of the data, this approach has the capability to summarize the data in ways that are more salient than a global model would. For example, by providing the ability to learn different playing modes, information regarding what makes a musical experience interesting, e.g. its juxtaposition of distinct playing styles, is directly built into the model. I have demonstrated BoB's ability to both learn powerful musician-specific abstractions (with as little as 120 bars of transcription; about 5 minutes of playing), and to integrate this learning into the generation of new solos.

Guiding Research Philosophies

1) Artistic goals are at least as valuable as purely technical ones, the idea being that good art (transformed into a computational setting) breeds good science, but the converse is not necessarily true. I believe that AI technology will remain impoverished until it can deeply perceive and interact with humans in their own poorly understood creative endeavors. I also believe that in tackling this problem, we must investigate real interactive art-based systems.

2) In artistic domains, one often needs to operationalize some form of "artistically-appropriate" behavior. I believe that ML techniques offer a viable solution to this need. For example, in my thesis, clustering techniques have been successfully shown to learn important musical abstractions when trained on the improviser's own example (e.g. distributed representations of Bebop scales emerged when trained on saxophonist Charlie Parker). This operationalization is both necessary and non-trivial, for even practiced improvisers *cannot describe their behavior in technically useful term*s. However, while "musically-appropriate" behavior is something for which we have little conscious access to, a common technique used in many interactive computer music systems (and, more generally, in other forms of interactive art) is to let the composer make it "sound good" (let the artist "author" it). While it is valuable to develop compositional (authoring) tools, this focus sidesteps the important issue that I embrace: upgrading the computer into a first-class, creative citizen.

Future Research Directions

My research interests are aimed at creating collaborative agents in creative, art-based domains. An important aspect of such work involves balancing engineering and artistic needs. For example, a deep respect for, and understanding of, music and improvisation guided the high-level design of BoB, while its probabilistic foundation allowed principled perception and generation methods to be developed and evaluated. A unifying theme in my future research efforts will involve exploring and exploiting the tension between artistic (subjective, ill defined) and engineering (mathematical, algorithmic, principled) needs. I am especially excited about investigating this balance in the realm of interactive agents because it will allow me to consider the question of artistic intelligence and creativity from *within* the human context (after all, creativity does not just happen in a vacuum). My graduate work establishes a foundation for this type of research in the domain of improvised music, and motivates me to further investigate the interesting machine learning challenges that arise in this setting (e.g. feature extraction, learning from small data sets, sequence learning, real-time adaptation, etc.). My extensive experience as a musician also makes this domain an ideal one to pursue. I am also very open to collaboration with others—artists, statisticians, cognitive scientists, psychologists, experts in HCI, AI, graphics, multimedia, etc. My hope is that such collaborations will produce agents that, when interacting with the user, produce note-worthy, compelling, and beautiful results.

As an improvisational violinist, I want to hook myself up to a system like BoB, collaborating with it to produce interactive art pieces that complement me and provide insight into my own musical personality. While BoB provides an initial-framework within which to build such a system, additional functionality is needed. Real-time interaction is currently simulated with transcribed improvisations. *Live interaction*—which is a prerequisite to many really interesting experiments—requires that I develop a real-time mapping between the user's performed note-durations and an "appropriate" musical score. ML is currently being applied to the automatic transcription of classical score-based performances; I plan to conduct similar research in the highly expressive, improvised musical setting. Another critical issue that needs to be addressed essentially involves reversing transcription: providing the agent with the ability to play a mechanical score in a musically convincing manner.

A crucial aspect of my research will involve developing principled mechanisms for *evaluating interactive art-based agents.* A basic problem with art-based AI research is that, because artistic quality is ultimately based upon subjective opinions, there is not the same type of idealized success measure that one gets in engineering (compare measuring "Is this compelling?" with "How many zip codes were sorted correctly?"). As an initial step, an artistic Turing test may indicate the worth of a particular approach. With ML-based agents, principled analysis is also possible (how well does an algorithm work on simulated data; how important are various features for achieving a particular discrimination; etc). I plan to use these types of techniques to further investigate what features are most salient for building computational music models.

In a nutshell, the most exciting issue that I hope to address is this: *to what degree can an interactive system boot strap off of its user's creativity, providing an experience in which the sum of both parts (agent+musician) is better than that of either individual?* Answering this question necessarily involves having many people play with a system (the web is a great medium for such distribution). I have no doubt that invaluable lessons will be gained when such playing occurs with BoB. Such activity will spawn numerous new ideas and insights whose value cannot possibly be assessed before the playing begins.