

# Intelligent Learning Agents for Music-Based Interaction and Analysis

## (Doctoral Consortium)

Elad Liebman  
Supervisor: Prof. Peter Stone  
University of Texas at Austin  
Computer Science Department  
Austin, TX  
eladlieb@cs.utexas.edu

### ABSTRACT

My main research motivation is to develop complete autonomous agents that interact with people socially. For an agent to be social with respect to humans, it needs to be able to parse and process the multitude of aspects that comprise the human cultural experience. That in itself gives rise to many fascinating learning problems. I am interested in tackling these fundamental problems from an empirical as well as a theoretical perspective. Music, as a general target domain, serves as an excellent testbed for these research ideas. Musical skills - playing music (alone or in a group), analyzing music or composing it - all involve extremely advanced knowledge representation and problem solving tools. Creating “musical agents” - agents that can interact richly with people in the music domain - is a challenge that holds the potential of advancing social agents research, and contributing important and broadly applicable AI knowledge. This belief is fueled not just by my background in computer science and artificial intelligence, but also by my deep passion for music as well as my extensive musical training.

### 1. BACKGROUND

In recent years, the application of algorithmic tools in cultural domains has become increasingly prevalent. An interesting example is Sisley the Abstract Painter [11], a project designed to algorithmically emulate modern paintings of varying abstraction levels, given an input photograph. Another recent example uses visual recognition tools to study what makes the architectural styles common to different cities distinctive [4]. Focusing on music, computers have accompanied both the creation and the analysis of music almost since they first came into existence. The famous composer Iannis Xenakis used computers in the early 1960s to generate numerical patterns, which he later transcribed into sheet music. A decade later, the renowned composer and conductor Pierre Boulez founded IRCAM (Institut de Recherche et Coordination Acoustic/Musique), where composers, computer scientists, and engineers study music and sound and new tools for creating electro-acoustic

**Appears in:** *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*, Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.  
Copyright © 2015, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

and computer-generated music. Two decades after, David Cope would use statistical models to try and emulate classical composers [3]. Another important body of work in the application of learning and computational tools in music involves the music information retrieval (or MIR) community. Over the last decade, many researchers have applied computational tools to carry out tasks such as genre identification [2, 9], music summarization [5], music database querying [8] and melodic segmentation [10].

### 2. PROPOSED RESEARCH

My main long-term goal is the creation of large-scale autonomous social agents. Specifically, I aim to create musical agents, i.e., computational entities capable of high level musical reasoning. To achieve this goal, several crucial sub-problems must be dealt with. Such sub-problems include (but are not limited to) a) “parsing” music intelligently (segmenting, annotating and representing musical knowledge); b) using parsed musical segments to analyze given music and to identify previously observed traits in new examples; c) learning to identify similarities and disparities between two given pieces; d) being able to analyze structure and thematic material organization in given pieces; e) combining observations to identify periods, genres and personal styles; e) modeling and predicting aesthetic preferences; f) Modeling/quantifying aesthetic value and or information rate (or surprise); g) being able to perform music intelligently.

Thus far, most examples in the literature only tackled one or two of these issues at a time, and indeed it’s unlikely that all of these sub-problems could be fully resolved within the scope of a single Ph.D. thesis. It is my goal however to address all of them to a level that’s sufficient for the generation of a complete, functioning agent, and delve more deeply into the problems that are more crucial for the establishment of an agent framework.

Finally, such an agent should be tested on their ability to interact meaningfully with humans and other agents in a musical setting. Examples of such meaningful interaction include intelligently assigning significance to musical cues, being capable of identifying the taste of an interacting person and respond appropriately, or the ability to cooperate with people and other agents in musical improvisation and the creation of new music altogether. If successful, my research will contribute to various applications, such as improving adaptive music recommendation, intelligent computerized

musical accompaniment, and natural human-computer interaction.

### 3. ONGOING WORK

My main long-term goal is the creation of large-scale autonomous social agents. As part of my current research toward this goal, I have tackled several related issues. I will now briefly survey three of them:

#### 3.1 Learning Representative Sets

One of the most basic issues one comes across when trying to analyze large quantities of musical data is the intractability of comparing large music corpora. If one wishes to compare and analyze two bodies of work, one needs to first find an efficient way to represent them. Such representation has to be sparse enough to minimize redundancy, but also rich enough to capture nuances. For these purposes, clustering musical segments by similarity and selecting a representative from each cluster would be a natural approach. However, existing clustering techniques present various challenges in the context of the representative selection task. These challenges are even greater when dealing with non-metric data, such as musical segments, where only a pairwise similarity measure exists. In collaboration with Prof. Peter Stone and Prof. Benny Chor, I've compared various approaches for this problem and analyzed it theoretically. In a forthcoming submission, we discuss some of these results and present  $\delta$ -medoids, a novel algorithm for the representative selection. The key insight of  $\delta$ -medoids is that it first adaptively selects a representative set that sufficiently covers the data, and then iteratively attempts to improve this set until it reaches convergence. We show that the algorithm performs well on musical data. We also show it works well in another empirical domain - motion analysis for simulated robotics.

#### 3.2 Autonomous Playlist Generation through Reinforcement Learning

A specific domain in which intelligent agents, music analysis, and preference elicitation connect is in the construction of musical recommender systems. To the best of my knowledge, most of these systems focus on predicting the preference of individual songs independently based on a learned model of a listener (see [6, 1, 7], for example). However, a relatively well known fact in the subfield of cognitive neuroscience which studies music cognition is that music is experienced in temporal context and in sequence. I therefore focus on the issue of learning to adaptively build music playlists based on user interaction. In a work done in collaboration with Prof. Peter Stone and Prof. Maytal Saars-Tschansky, we employ a reinforcement learning approach to music recommendation. The proposed agent learns a preference model for both individual songs and song transitions, and recommends complete song sequences rather than just individual songs. To reduce exploration time, the algorithm actively initializes its model based on user feedback. The resulting agent has been tested on human participants and been shown to increase the expected enjoyment of listeners from the selected song sequence, compared to a greedy approach that only tries to select individual songs.

#### 3.3 Learning to Model Responses to Musical Stimuli and Musical Preferences

One of the key issues in building a successful social agent

with respect to music is its ability to model and predict human responses to musical cues. This gives rise to two interesting questions: a) Can an agent successfully model how people's behaviors are modified given different musical cues? and b) Can we successfully learn musical preferences, either over a population or through interaction with a single person? These issues are in central to a current research effort, done in collaboration with Prof. Corey White from Syracuse University. In this project we focus on modeling how different music stimuli affect human decision making, and whether it is possible for opposite agents to predict and utilize this knowledge when performing tasks jointly with human participants.

### 4. FINAL NOTES

If we envision a future where intelligent artificial agents interact with humans, we would like to make this interaction as natural as possible. We would therefore like to give AI the ability to understand and communicate within cultural settings. Such progress would have many real world practical benefits, from recommender systems and business intelligence to negotiations and personalized human-computer interaction. It will also have direct ramifications for Facebook, as a significant part of the services Facebook provides are pinned on learning the underlying interests and needs of people in cultural settings, to enable new and more interesting ways of knowledge sharing and communication. But beyond its practical usefulness, having AI tackle complex cultural domains, which require advanced cognitive skills, would signify a meaningful breakthrough for AI research in general. I believe my research towards learning social agents in the music domain will make significant progress towards such a breakthrough.

### REFERENCES

- [1] L. Barrington, R. Oda, and G. R. Lanckriet. Smarter than genius? human evaluation of music recommender systems. In *ISMIR*, volume 9, pages 357–362, 2009.
- [2] J. J. Burred and A. Lerch. A hierarchical approach to automatic musical genre classification. In *Proc. 6th Int. Conf. on Digital Audio Effects* 03, 2003.
- [3] D. Cope. *Experiments in musical intelligence*, volume 12. AR Editions Madison, WI, 1996.
- [4] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. A. Efros. What makes paris look like paris? *ACM Transactions on Graphics (TOG)*, 31(4):101, 2012.
- [5] S. Dubnov, S. McAdams, and R. Reynolds. 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.
- [6] J. John. Pandora and the music genome project. *Scientific Computing*, 23(10):40–41, 2006.
- [7] G. Kreitz and F. Niemela. Spotify—large scale, low latency, p2p music-on-demand streaming. In *Peer-to-Peer Computing (P2P)*, 2010 *IEEE Tenth International Conference on*, pages 1–10. IEEE, 2010.
- [8] L. Lu, H. You, H. Zhang, et al. A new approach to query by humming in music retrieval. In *ICME*, pages 22–25, 2001.
- [9] C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. In *ISMIR*, 2004.
- [10] M. Pearce, D. Müllensiefen, and G. A. Wiggins. A comparison of statistical and rule-based models of melodic segmentation. In *ISMIR*, pages 89–94, 2008.
- [11] M. Zhao and S.-C. Zhu. Sisley the abstract painter. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*, pages 99–107. ACM, 2010.