CS280r Final Project Report

Αρμονία (Harmonia): A System for Collaborative Music Composition

Mark Goldstein, David Wihl

{*markgoldstein,davidwihl*}*@g.harvard.edu*

Abstract

Increasing productivity of music composition has many positive benefits. Listeners would appreciate individually tailored music to their emotional needs and context. Composers would be facilitated by greater and more diverse cooperation yielding more innovative music. Composition agents could assist in the generation of repetitive or experimental musical forms. Therapists can use music as part of a treatment plan for autism and many other disorders. The system we propose attempts to address these myriad needs by offering two key innovations: a SharedPlan with collaborative versioning to mediate the workflow of a composition, an algorithmic evaluation of a composition against the intention of the SharedPlan to provide guidance to both human and agent composers.

1. Introduction

TODO Should contain an overview of the problem to be addressed, the approach taken to address that problem, and the results of that approach. Should provide the reader with a road map for how your argument will be developed in the other sections of the paper.

RESEARCH TODO

- Mark: talk to composers and incorporate their feedback and UI suggestions
- David: speak to David Greenberg to incorporate feedback and UI

2. Related Work

Our work builds on several areas of research related to music, computer science, and creative cognition. First we discuss related work in collaborative ideation, both in general and specifically in music. Next we discuss intelligent music systems that facilitate human composition and improvisation, and related work in Music Information Retrieval (MIR). Finally we describe previous work that applies information theory to the analysis of musical structure.

2.1. Collaborative Ideation

Collaborative Ideation (CI) seeks to improve the productivity of individuals and groups in generating ideas (ideation) through collaboration. The people involved are interested in creating related objects (we all want to write birthday cards, come up with solutions to social problems, create music) and seek either feedback or examples of others' work to enhance their individual process. The collaboration is usually centered around a shared "workspace" that allows for communication and sharing of ideas. The space may be physical or electronic. The produced ideas may be for ultimately individual use, or there can be a shared artifact such as an essay or a piece of art attributed to several ideators. The dynamics of sharing ideas through the shared place may be real-time or not, though increasingly, today's settings are real-time and virtual.

The simplest forms of CI help each participant increase their productivity in brainstorming solutions about a specified topic by showing all participants each other's ideas. While this approach has naturally been explored in human culture for millennia, the design of intelligent computer systems today aims to facilitate these activities to allow for increased creativity and productivity. Considering a setting where it is overwhelming for each ideator to view all participants' ideas, a computer system may sort and prune the idea space to show each participant only the most relevant and inspiring ideas. IdeaHound by Siangliulie et al. addresses at-scale collaboration in this setting. The system creates a semantic map of all ideators' generated ideas that allows each to easily view their ideas in the context of the entire solution space. The map is automatically generated by prompting users to interact with a personal "whiteboard" where they can cluster their ideas and separate them by semantic distance, and then by computing a global map from the collection of whiteboards. This approach bypasses the need for external workers to power semantic analysis. Using this map, IdeaHound can recommend diverse ideas that allow each ideator to span a wider part of the solution space, without needing to devote cognitive effort to search large lists of ideas. Our system takes a lot of influence from this archetypal example. However, some questions that arise when moving from this archetypal example to collaborative music composition specifically are: What happens when the generated objects are structured? What happens when ideators need to explicitly build on each other's ideas rather than only seek inspiration from them?

One setting where CI has surfaced is the space of online blogs and other services designed to share visual art and music. Ideas range from small ideas seeking suggested directions to finished pieces seeking feedback. Artists then improve upon their ideas using large-scale feedback. SoundCloud is an example of a hybrid music streaming service and collaborative ideation platform. While much of the work is presented in finished form, people also post incomplete projects. Even with finished pieces, artists sometimes share "stems" to their tracks which are individual such as "just drums", "just vocals" so that others searching for inspiration can remix their ideas to make their own new pieces. A newer platform called Blend makes this explicit, where the idea is not just to share audio recordings, but by default to share the music production software source files that were used to create the piece. This allows someone to take someone's piece-so-far not just by sound, but to actually see all of the written material, and start working with it in software right away. This setting is closer to our area of application and supports building on one another's ideas. However, what happens when the several ideators intend to create a shared piece instead of just building on each other's work in their own directions? How can a shared plan be specified, and how can a system support group work toward the goal? In our case we are interested in collaborative music composition, where the shared musical artifact has a particular goal associated with it through the duration of its existence. With SoundCloud and Blend, one may take someone's piece in a totally different direction. In our case, we would like to facilitate the creation of shared pieces (we allow for inspired individuals to take a piece in their own direction as a separate project if they would like).

https://blend.io/about

2.2. Computer Facilitated Composition and Improvisation

Computer agents with the ability to facilitate and take part in music composition and improvisation are on the spectrum of a wide number of intelligent musical systems under question in the field of Music Information Retrieval. Many of these approaches have in common a requirement to "understand" music at multiple levels such as low-level acoustic signal, mid-level musical constructs such as harmony and rhythm, and high-level level aspects such as mood and genre. A music recommendation system such as Spotify make seek to analyze music and extract a measure of relevance for a function such as "study music". In the case of systems that actually create music, people are interested not in replacing, but in interacting with human composers. Perhaps the composer has good idea "seeds", but the system may recommend variations of ideas, or re-orderings of ideas, to make them more conveying. This system knowledge may come from large-scale corpus analysis.

ChordRipple [Huang 2016] is a recent system that takes as input a progression of musical chords for a composer, and suggests substitutions of intermediate chords that preserve the original semantics of the input while serving to replace conventional choices with more interesting ones. If the composer agrees to make one of the recommended changes, the system then assists the composer in integrating interpolating between original and substituted material before and after the original change.

While our current work seeks primarily to assist teams of human composers to enrich and organize their work, we intend to design the system such that intelligent computer agent composers may be in the loop. The Google Brian team has recently launched the Magenta Project for exploring machine intelligence in music, an integrated environment of software tools and music-related datasets. Recently, Magenta released AI Duet, a computer system that reacts to human improvised gestures. Improvisation is an important part of composition. Even in steps where a human is composing, it may be beneficial to have an agent for the human to go back-and-forth on ideas with.

In settings where a piece enough specification of purpose, such as in the therapy case study where a listener may need music at a certain tempo and with a simple beat, powerful information retrieval systems make possible effective machine composition agents. Human composers may be placed at later steps of collaboration to ensure that the piece meets requirements in a humanly perceptible way.

2.3. Information Theory and Music Analysis

[THIS WILL BE CHANGED FROM LESS OF A LAUNDRY LIST TO MORE OF A SPECIFIC RELATED WORK SECTION. I WILL GET RID OF MUSIC INFO RETRIEVAL THINGS IN GENERAL AND ONLY KEEP THOSE SPECIFIC TO MUSIC ANALYSIS]

Understanding musical structure and its impact on listeners is a cogent goal to music theorists, cognitive scientists, and machine learning researchers. Music theorists such as Meyer [1956], Narmour [1992], and Huron [2006] have pointed to the dynamics of anticipation, fulfillment, and denial in the listener as a major aesthetic and stylistic characteristic of music. Cognitive scientists such as Dubnov [2006] have correlated prediction and emotional response in listeners. Info. Dynamics of Music [Pierce and Wiggins, 2012] ties evolutionarily-developed expectation tendencies to perception and cognition in music listening. These listener-centric works borrow tools of statistics and info. theory in their models. We elaborate on some historical work, and then current. In ?Mathematical Basis of the Arts?, Schillinger [1948] describes early corpus-based experiments in learning composer harmonic style. Shannon and Pierce [1956] applies work from Shannon?s seminal ?Mathematical Theory for Communication? [1948]: they experiment on games in which participants add to each other?s compositions while only seeing a few final measures of the preceding participant?s music. Analysis of listener uncertainty is made by measure of entropy within and across composer boundaries. Meyer [1956] defines musical style as ?complex systems of probability? in which meaning depends upon relationships among other terms within the style system. Youngblood?s ?Style as Information? [1958] is the first music theory work to build on Shannon. Cohen [1962] surveys preceding works using info. to describe the stochastic process of music. He analyzes the assumptions in prior work and gives departure points for future work: 1. ?the basic [assumption] is that statistical probability... corresponds to the listener's expectations... the average surprisal value... represents the listener's state of uncertainty? 2. ?Another... is that one portion of the cultural sign system can be legitimately abstracted from the whole, and that values based on this abstraction will have the same worth as when the portion is a part of the whole.? 3. ?A further assumption... a sequence of musical events is experienced on only one architectonic level: in melodic analyses, on the level of notes or intervals; in rhythmic analyses, on the level of the pulse pattern... theory will have to take account of the interaction among levels.? We need to consider several streams of info. and their interactions: rhythm, pitch, harmony, timbre. Even within one type of information stream, say pitch, we need to consider several hierarchical levels at once: rhythms constitute a local time feel but also accelerate a piece toward new sections. Works 2000-present have focused on machine learning for recommendation systems and machine composition without much use of information. Widmer?s Manifesto [2016] poses the challenges that intelligent musical systems research will face in the coming decade. He restates Cohen?s call, 54 years later, for a theory of interactions of multiple streams of musical info, as well as several rates of info. in one stream contributing to multiple levels of hierarchy. Recent info. theory challenges include generalizing mutual information to several variables. Multivariate mutual info. Is used to describe the extent to which the knowledge of one event can reduce the entropy in several other variables [V.d. Cruys 2011]. Abdallah [2012] uses predictive info. rate to measure how a listener?s mid-piece distribution over future events is revised as new info. is presented. He models the relationship of surprise and form.

3. Workflow Overview

git + intention + algorithmic eval
use of genre + mood
Other metadata
incorporate SharedPlans
Current git / music solutions
How do composers or users express intentions?
What sharing of information occurs between users and composers? Decoupling (be explicit)
How to avoid revision wars

4. Algorithmic Evaluator

Current design
MIDI
entropy discussion
KL Divergence

5. Use Cases

5.1. Individual User, Individual Composer

5.2. Multiple Composers

TODO: include failure modes

5.3. Therapist with Agent - Human Composition Team

high volume necessity

6. User Interface

TODO

7. Discussion

7.1. Enhancing or Stifling Creativity

Notes: evaluation is optional. Can be ignored by committer.

7.2. Limitations

Collaboration is offline, not real-time

Current music representation is discrete MIDI, not audio. Limits for vocals, ocean sounds

Presume that reliable corpus-based genre and mood classification solutions exist, particularly information retrieval procedures

8. Conclusion

Two Novel Contributions:

- Collaborative music composition system Intentionality, SharedPlan and Agents
- Algorithmic evaluation of composition against intention

9. Future work

- Improved agent composition
- Intelligent ad hoc composition
- Facilitator of scalable music composition
- improved evaluator, possibly RNN based

10. References

Meyer, L.B., 1956. Emotion and Meaning in Music. Chicago University Press, Chicago, IL.

Narmour, E.. 1992. The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model.

D. Huron. 2006. Sweet Anticipation: Music and the Psychology of Expectation. MIT Press, Cambridge, MA.

Abdallah, Cognitive Music Modelling: An Information Dynamics Approach. 2012 3rd International Workshop on Cognitive Information Processing (CIP)

Widmer, Gerhard. 2016. Getting closer to the essence of music: The Con Espressione manifesto. ACM Transactions on Intelligent Systems and Technology

Engel et al., Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders, 2017

Wiggins, Auditory Expectation: The Information Dynamics of Music Perception and Cognition. 2012 Topics in Cognitive Science

Moles, A.. 1966. Information Theory and Aesthetic Perception. University of Illinois Press, Urbana, IL

Two Multivariate generalizations of Pointwise Mutual Information Tim Van de Cruys, Association for Computational Linguistics 2011

Cohen, Joel E., Information Theory and Music , Behavioral Science, 7:2 (1962:Apr.) p.137

Schillinger, Joseph The mathematical basis of the arts 1948

Pierce, Electronics, waves, and messages. 1956

Pierce, Letter Scientific American 1956

Youngblood, Style as information 1958 Journal of Music Theory

The mathematical theory of communication. Shannon, Claude Elwood 1948. Bell Tel Labs Monograph