

Integrating the Cognitive with the Physical: Musical Path Planning for an Improvising Robot

Mason Bretan
pbretan3@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Gil Weinberg
gilw@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Abstract

Embodied cognition is a theory stating that the processes and functions comprising the human mind are influenced by a person's physical body. *Embodied musical cognition* is a theory of the musical mind stating that the person's body largely influences his or her musical experiences and actions (such as performing, learning, or listening to music). In this work, a proof of concept demonstrating the utility of an embodied musical cognition for robotic musicianship is described. Though alternative theories attempting to explain human musical cognition exist (such as *cognitivism* and *connectionism*), this work contends that the integration of physical constraints and musical knowledge is vital for a robot in order to optimize note generating decisions based on limitations of sound generating motion and enable more engaging performance through increased coherence between the generated music and sound accompanying motion. Moreover, such a system allows for efficient and autonomous exploration of the relationship between music and physicality and the resulting music that is contingent on such a connection.

Introduction

Several advantages and opportunities emerge as a result of the physical embodiment of machine musicians compared to that of their pure software counterparts. These “robotic musicians” assume additional abilities to entertain, engage, socialize, and produce sound. Such desirable capacities are intrinsic to most social robotic platforms in general, as a result of physical presence and embodiment, and benefit both those directly immersed in the interaction as well as those simply witnessing it (Kidd and Breazeal 2004). Though many of these interactive assets must be explicitly designed to address human perceptual and social tendencies, in music some advantageous characteristics arise entirely as a result of the inherent coupling between the robot's spatial movements and sound generation. The byproduct of sound generating movements is increased levels of rhythmic coordination and synchronization within ensembles because interacting musicians are able to anticipate the robot's musical onsets or behaviors through visual cues (Hoffman and Weinberg 2010; Lim et al. 2010).

These benefits, however, do not come without the additional constraints that are coupled to the natural world in which we live. An artificial intelligence (AI) controlling a mechanical body that is bound to the physical laws of nature must address its own presence in 3-dimensional space in order to function properly. This includes not only knowing the location of each of its degrees-of-freedom (Dofs), but also understanding how multiple DoFs need to behave and work together in order to complete a task. Perhaps most importantly, the AI must also understand when a task is impossible (given its physicality) and anticipate failure before it causes damage to itself or corrupts certain aspects of the task making success completely unattainable even with additional help.

In the musical domain these constraints are often (if not always) compounded by issues of timing. It is not enough to simply arrange the DoFs in a desired order; the timing and sequencing of the movements must also be considered. Often, depending on the time constraint and robot's physical design, the path in which each DoF moves and relocates itself must be optimized in order to reach specified locations in a timely manner. This constraint satisfaction process of developing coordinated movements is referred to as ‘path planning’.

Here, we explore path planning and its relationship to the notion of *embodied musical cognition*, which states that an individual's body largely influences his or her understanding, experience, and decision processes pertaining to music (Godøy and Leman 2010; Leman 2008). In particular, we examine how proprioception and embodiment can (and we argue should) influence the musical decision process of an improvising robot musician. The hypothesis is that a robot that utilizes a music generation method that jointly optimizes for its physical constraints as well as its general musical knowledge will increase performative expressivity by providing increased coherence among higher level musical ideas, sound generating motions, and sound accompanying motions. Additionally, such an integrated musical decision process will result in music that is defined by the robot's physical identity, hopefully leading to interesting music and even entirely new genres.

Traditionally, machine musicianship has focused on software applications that respond to and generate music (Lewis 2000; Drummond 2009; Whalley 2010). Robotic musician-

ship has been defined to be comprised of two components (Bretan and Weinberg 2016). One of which is ‘machine musicianship’ or the development of cognitive models depicting aspects of music perception, composition, performance, and theory. The other component is ‘musical mechatronics’ or the study and construction of mechanical systems capable of sound generation.

In musical mechatronics there are several things that a researcher considers. There is no single physical design of a musical robot that is perfect and there have been no claims that one such design exists. Instead, designers make trade offs that may account for a robot’s size, mass, possible anthropomorphic design, the instrument it will play, the specific genre(s) it will play, method of sound actuation, ability to provide useful visual cues, the ability to provide social cues, energy consumption, price, and aesthetics. In other words, designers make decisions that are influenced by both music-specific ambitions and physical characteristics.

This paper focuses on the machine musicianship aspect of robotic musicianship and it presents the notion that the music making decisions of a generative algorithm must also be influenced by the robot’s own physicality. Similarly to the processes that go into formulating the physical design of a musical robot, a robotic musician should have an intelligence that integrates the cognitive and physical domains such that the musical and physical behaviors are a result of a decision process that jointly optimizes musical goals, path planning, and human perception.

This work addresses several research questions:

1. **How can autonomous decision processes based on music incorporate the physical domains?** Including additional constraints on music requires a rethinking of the traditional machine musicianship concepts. Though previous methods for generating music are still relevant and can be useful resources, an integrated approach needs to address a significant expansion of parameters and complexity. Specifically, the note-level stochastic methods that are widely used may not be feasible for an integrated optimization method.
2. **What is the best way to represent the physicality of the robot within the decision process?** Some DoFs may need to be expressed individually (such as those directly involved in sound generation), but in order to reduce complexity other DoFs (such as those involved in generating sound accompanying gestures) need to be represented as a collective or as part of a higher level physical behavior. Designing meaningful physical behaviors that incorporate multiple DoFs, and various velocities, accelerations, positions, and trajectories will be essential.
3. **What aspects of the pure software generative music algorithms will need to be modified to better suit embodied methods?** While it might be possible to create a completely real-time system in which the algorithm generates and plays notes on-line, such a method is probably not ideal. The nature of musical path planning is that an optimal sequence of moves is generated in order to achieve the musical goals. Therefore, rather than creating paths with a length of only a single note or move, the

system should generate paths with lengths that represent complete musical ideas (though it is possible a complete idea is indeed only a single note). These musical chunks may be portions of phrases, complete phrases, or even entire structured improvisations from beginning to end.

4. **Is there a single integration approach that can be useful for many robotic platforms that have vastly different designs and functionalities?** A joint optimization methodology and single algorithm may be suitable for many music generating robots, however, a single state space representing all platforms is probably not possible. The physical design of robotic musicians tend to vary significantly from platform to platform. Instead, adjustments will likely need to be made that address a specific robot’s physical characteristics and intended interactions and behaviors.

In this remainder of this paper we describe specific motivations for such a joint optimization music generation method. We discuss related work in the music and cognitive science domains and address how embodied cognition concepts have influenced our design. Finally, we provide an overview to our generative musical path planning system that creates different musical motifs in which the source of variance stems from the physical constraints.

Motivation

Though there are various theories of mind and several examples of robots being used to study natural human behavior and brain science (Atkeson et al. 2000; Cheng et al. 2007; Asada et al. 2009), our work does not attempt to prove or disprove one particular theory. Rather, we demonstrate why an algorithmic design inspired by embodied cognitive processes is more suited for robotic musicianship than disembodied cognitivist approaches. The general premise is that embodied cognitive methods would enable a robotic musician to find the most effective solution in conveying its musical goals given its physical constraints.

Additionally, qualitative and anecdotal evidence suggests that some of the higher level musical semantics that describe a person’s style emerge as a result of that person’s physical interaction with his or her environment (Gibson 2006). Gibson (2006) explores instrument specific musical tendencies in jazz improvisation and based on qualitative interviews with professional musicians speculates that some patterns or motifs partially arise due to the natural affordances of the instrument on which they’re being performed. For example, the trumpet’s nature makes it more suitable for small intervals going up and down a scale as opposed to arpeggios, something that the piano is almost perfectly designed for and may be the reason why arpeggios are so prevalent in pianists’ improvisations. One musician describes Eric Dolphy’s playing style on saxophone containing massive intervals as “bloody impossible on trumpet.” Another musician explains that there are more notes available in a single hand position on electric bass than double bass and as a result one tends to play more four note patterns on electric and more three note patterns on double bass.

The alternative to integration would be to create individual and separate modules for each function or variable (music knowledge, physicality, social gestures, etc.) However, in creating separate modules we risk failing to encapsulate the higher order dynamical system that contextualizes each function and illustrates the reason for their ontogenesis in the first place. This idea is the premise for ‘developmental robotics’ in which a robot’s higher level processes emerge as a result of its physical interaction with the environment (Asada et al. 2009; Weng 2004). A robot’s physicality bridges the gap between its internal cognition and the physical infrastructure of the environment by providing it with the ability to gain information essential to autonomous learning and decision making (Kuniyoshi et al. 2007). In order for this ability to be realized the robot must have an understanding of how to interact with the environment given the constraints defined by its physical embodiment.

Why is it important to address the physical in musical decision processes? Currently, robotic musicians employ an intelligence that supports a work flow as such:

1. Using traditional machine musicianship methodology generate a note or sequence of notes
2. Send the note(s) to a path planner that generates a sequence of movements that results in performing the generated note(s)

On the surface this may seem sufficient, but consider the differences between software based machine musicianship and robotic musicianship – Software applications aren’t bound to natural physics and as a result can be designed to play any note, combination of notes, timbre, volume, speed, and numerous other parameters. On the contrary, a robot operating in the physical world is limited in how it can move in space and, thus, limited by what it can sonically achieve.

Most composers write music capable of being performed by humans with two arms and two hands. Additionally, those studying composition learn what is and isn’t humanly possible on different musical instruments. Composers use their knowledge to make informed decisions addressing the physical world during the composition process. If people naturally had one arm and one hand these composers would write music suitable for such a physical form. They most likely wouldn’t write music for two armed people and give the one armed performer the responsibility of finding a way to play it. Likewise, the musical intelligence of a robotic musician should create music specific to its physical form that permits 100% of the notes to be performed all of the time.

The utility of coherent sound accompanying movements in music performance There is also motivation for a robot to obtain a physical-musical cognition that goes beyond the constraints and opportunities related to sound producing movements. Humans often employ multi-modal communication processes when interacting with one another. A musician’s secondary movements, or sound accompanying movements, are not only useful in making a performance more entertaining and engaging, but also serve more functional purposes by providing the ability to communicate intent and influence how observers interpret the music.

In one study, (Vines et al. 2006) show that a musician’s movements serve to both augment and reduce the perception of tension in music. It was also found, in this study and several others, that observers use the visual cues from movement to help understand the performer’s internal states, concepts of phrasing, and changes in emotional content (Wanderley et al. 2005; Dahl and Friberg 2007; Livingstone, Thompson, and Russo 2009). The differences between seeing a musical performance and simply listening to one are characterized by observers’ perceptual and even physiological responses (Chapados and Levitin 2008). One experience may not be better or worse, but the differences exist and a robotic musician should leverage the relationship between ancillary motion and musical features to enhance its ability to communicate its own musical goals and interpretations.

Understanding how physical behaviors are connected to human perception is important for effective application of a robot’s sound accompanying movements. (Nusseck and Wanderley 2009) found that the multi-modal experience from watching a human perform is more dependent on the overall movement characteristics of the whole body and relative motion of limbs to each other rather than specific arm or torso movements. This was also found to be true in a study with a small 5-DoF faceless robot in which the overall physical motion characteristics (such as velocity, acceleration, and periodicity) were much more effective for communicating sentiment compared to individual DoF positions (Bretan, Hoffman, and Weinberg 2016).

A commonly held view representing the relationship between machine musicianship and robotic musicianship is that the former encompasses all research related to machine generated music and is a component of robotic musicianship. Historically, however, machine musicianship has referred to software applications (Rowe 2004) and as such the generative methods designed for robotic musicians have been interchangeable with pure software musicians. The algorithms tend to be constructed to merely generate notes (or sounds) that are musically significant. Though machine musicianship and robotic musicianship are undoubtedly related and share many of the same objectives and challenges, physical embodiment is such a polarizing and distinctive attribute that new methods for optimally incorporating a sense of physical self into an algorithm should be developed. These methods would be truly unique to robotic musicianship. If machine musicianship sets out to algorithmically find a sequence of notes that is musically significant, robotic musicianship sets out to find a sequence of notes that is musically significant, playable by the robot, and its musical character can be enhanced through simultaneous physical behaviors.

Related Work

Musical Domain There are several examples of musical mechatronics with vastly different designs that allow performance on different musical instruments. The Logo’s Foundation has developed several non-anthropomorphic percussive, string, and wind playing robots (Maes, Raes, and Rogers 2011; Raes 2014) that are midi-controlled. Similarly, Expressive machines Musical Instruments (EMMI) and

Karmetik have developed different drum and string playing robots with impressive mechanical control (Troy Rogers and Barton 2014; Kapur et al. 2004; Kapur, Murphy, and Carnegie). At Georgia Tech, Haile and Shimon are percussive robots with anthropomorphic designs (Weinberg et al. 2008; Nikolaidis and Weinberg 2010). Shimon is additionally able to provide social cues using its head gestures (Bretan et al. 2012). Each of these robotic systems exhibits varying degrees of autonomy and intelligence. However, each similarly possesses a musical intelligence system that is independent of its physical control parameters.

One objective of algorithmic composition is to re-purpose the inherent behaviors and characteristics of computational functions such as chaos, genetic, and cellular automata for a musical context. The algorithms yield interesting behaviors that can be useful source material for generating music. In an installation by Trimpin, the acoustic behavior resulting from the interaction of sounds within the physical world was leveraged by distributing hundreds of automated guitar systems within a room (Trimpin). However, this was achieved through manual manipulation and configuration of each instrument. An example of algorithmically leveraging interesting behaviors in the physical world was done by Albin et al. in a swarming robotics system that generated music based on different swarming behaviors (Albin, Weinberg, and Egerstedt 2012). The robots' absolute physical positions and relative positions to one another were used for musical mappings. Though physical parameters were used to generate music, the specific parameters did not come from the robots' own physicality and designs, but rather their locations, therefore, nullifying the possibility for interesting music to occur that is algorithmically born out of physical identity.

Though the notion of integrating cognitive and physical parameters into a unified intelligence for music generation and performance is new, there are examples of such a unification in other robotic applications. For a robot to learn to complete different tasks involving its environment it naturally needs to apply a sense of physical self. In one system, a robot learned to adapt its motion primitives (such as reaching) to different forces applied to its motors, while still effectively achieving the task (Shadmehr and Mussa-Ivaldi 1994). Though there is an ideal or independent kinematic plan for the motion primitive, the system is able to adapt its physical trajectories and electrical current draw based on dynamical forces.

Embodied Musical Cognition The theory of embodied musical cognition is a model that illustrates the processes of the human musical mind. Unlike the dualist theory of *cognitivism*, which treats the mind as a distinct entity independent from the body, this theory states that the manner in which a person perceives, generates, and interacts with music in all domains is dependent on the properties of that person's body (Leman 2008). The cognitivist approach has been used to explain various structures in music such as tonal tension (Lerdahl and Jackendoff 1985; Lerdahl and Krumhansl 2007), melodic peaks (Eitan 1997), and melodic similarity (Hewlett and Selfridge-Field 1998). Though these models are successful in explaining musical

structure for a large body of music, because they are reliant on a serialist implementation of systematic rules they fail to explain structure for musical genres that do not adhere to the rules. These models are considered to represent human musical cognition in a narrow sense (Wilson and Foglia 2011).

Embodied musical cognition represents a theory in which thinking and acting are intertwined and complex cognitive processes result from the joint process. For example, based on this theory, a person that simply studies music theory without playing an instrument experiences music differently than an instrumentalist; both may have a profound musical knowledge, but the instrumentalist's physical interaction with music has molded his or her musical mind making the experience different. Though there is mounting evidence that the mind is indeed embodied (Leman et al. 2009; Foglia and Wilson 2013; Thompson and Luck 2008; Toivainen, Luck, and Thompson 2010), even if this is not necessarily the case, it seems not only logical, but necessary for a robot to have an embodied cognition. It has been shown that complex and adaptive behaviors can arise in an artificial agent simply as a result of including the agent's physical body as a component of the mental functioning and reasoning process (Asada et al. 2009). This seems to be a better solution than explicitly designing every rule addressing how the robot must function with its environment as per the cognitivist approach, especially considering the complex and chaotic nature of the physical world. In fact, there has been a recent trend to include embodied cognition in robotic systems (Ziemke 2005; Metta et al. 2008; Breazeal, Gray, and Berlin 2009; Trafton et al. 2013). The musical world is also complex making embodied cognition appealing for musical applications. However, the question of how to computationally design embodied musical cognition for a robotic musician remains and is described in the following sections.

Musical Path Planning

As a first step towards demonstrating the effects of integrating the cognitive and physical in music generation a general path planning method is needed. This requires a representation of the physical constraints imposed by a robots' sound producing movements and the algorithm to find the optimal sequence of movements. This work uses a Viterbi search algorithm to find such a sequence.

Viterbi relies on a state space, S , observation space, O , sequence of observations, Y , transition matrix A , and emission matrix B . As an example we consider the design of Shimon, a particular four armed marimba playing robot (Nikolaidis and Weinberg 2010). Shimon's four arm design limits its ability to play certain musical passages. Defining the physical constraints that impose these limitations in a manner a computer can understand is essential. We define each state in S as being a physical configuration of the four arms. Transition and emission matrices are typically defined as probability matrices that describe the likelihood of certain events occurring. In path planning these matrices do not describe probabilities, but rather binary descriptors of whether an event is possible.

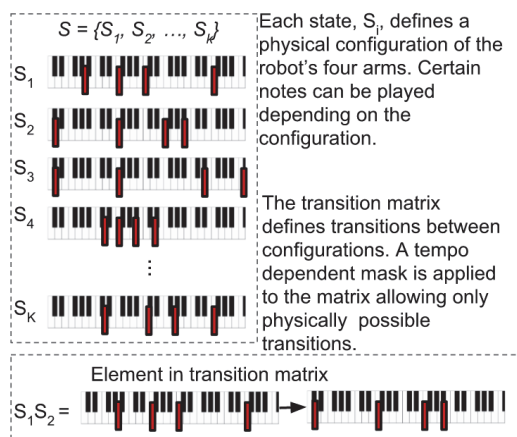


Figure 1: Physical state space of the Shimon robot.

The transition matrix is a tempo dependent matrix that defines whether it is possible to move from one physical configuration to the next. If the sequence of observations is a sequence of notes then the emission matrix simply describes whether it is possible to play the note given the state, S_i , and observation, Y_t , at time t . Figure 1 provides a visual description of the state space.

An additional cost can be applied to the Viterbi function by penalizing state transitions based on the distance they travel. This removes extraneous movements and the most efficient path can be found.

Evaluation The path planning algorithm can be objectively evaluated by comparing it to Shimon’s current on-line motion algorithm as a baseline. The baseline does not find a global optimum given a sequence of notes, but rather makes greedy decisions at the rate of an individual note. When a note is received the arm closest to the note is chosen to move and strike the note. This may seem reasonable, however, in practice many moves are not possible because of speed limits and the width of each arm being greater than the width of a bar on the marimba (meaning the note cannot be reached without collision). To prevent collision between arms a note can be transposed to a different octave and the method is repeated. In more extreme cases in which no arm can play the original or a transposed version of the note then the note is dropped completely.

Both algorithms were given one hundred pre-composed monophonic note sequences. The sequences came from a mix of classical melodies and transcribed jazz improvisations. Using the two algorithms, note sequences played by the Shimon robot were generated. The rate at which each algorithm transposed a note and dropped a note was computed. The results are shown in Table 1.

	octave transpose rate	drop rate
baseline	39.2%	14.9%
viterbi	11.1%	1.7%

Discussion The results demonstrate that movement sequences that are optimized for sequences of notes (such as a phrase) perform better. However, one may argue that it is

possible to develop a robot that is capable of playing everything and as a result doesn’t require path planning or a necessity to include its physical identity as part of its note generating algorithms. This may be especially true for some percussion robots. For example, Eric Singer’s *Orchestrion* system designed for Pat Metheny has a marimba playing robot that has a striker over each key allowing it to play any sequence of notes. Similarly, a disklavier has no need for a path planning algorithm that coordinates multiple motors.

Despite the advantages of these systems (and ignoring their more disadvantageous attributes such as a paucity of social and visual cues), it can still be contended that incorporating a sense of physical identity is essential for new music to arise. While there are no limits on what may be played and such systems can be sufficient for many musical applications, there are definitely opportunities that may be missed if any note generating system is implemented. For example, a note-to-note statistical model trained from human performances will inherently produce something playable by people. Unless the robot knows it is capable of performing outside of what is capable by humans it will never take full advantage of its physicality. Techniques such as 15 note chord harmonies or simultaneous musical lines in six different octaves may have musical utility and beauty, but will not be discovered by a musical intelligence lacking knowledge about its embodiment. Typically, composers and developers explicitly address these opportunities within the coding or composition process. In the next section we describe a generative music algorithm in which only the physical constraints are described and the intelligence discovers what is possible.

Embodied Generative Music

In this section a generative algorithm that explores embodied musical cognition within the context of jazz improvisation is discussed. The purpose of this paper is not to evaluate the quality of the generated improvisations, but rather to demonstrate the utility of integrating the physical constraints into the decision process. Figure 2 describes the general framework for the system. It follows the idea that higher-level musical semantics are necessary.

Parametrized representations of higher-level musical semantics The phrase ‘musical goals’ has been used a few times in this text. This is because the validity and efficacy of an integrated system relies on the assumption that there are multiple ways for a single higher-level musical idea or goal to be achieved. Without this assumption the methodology will fail, however, there is a lucid cogency to the assumption that presuming its truth is logically sound. For example, if a musician is asked to write a motif that ascends in pitch there are many possible motifs that can satisfy this constraint.

The same holds true for more complex semantic musical features such as tension. Within jazz improvisation (note that classical definitions of tension and resolution are somewhat different) a musician may build tension by adding notes that fall outside of the scale harmony. This technique is used by many jazz musicians, yet, each musician is able to construct unique note sequences while still attaining the higher level

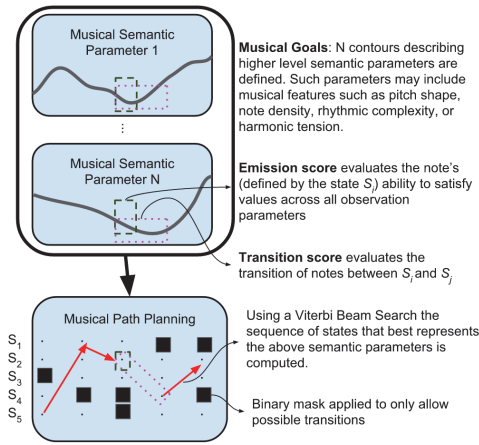


Figure 2: Framework of the generative embodied music system.

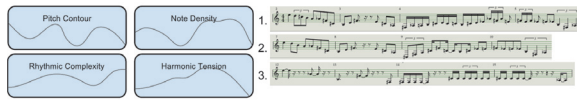


Figure 3: Embodied generation samples. Three motifs are generated to satisfy the musical parameters on the left using different physical constraints. **1.** A single arm robot that can move very fast. **2.** Robot with four larger and slower arms (Shimon setting) that must avoid collision. **3.** Very slow moving single arm robot with a fast strike rate.

goal of building tension. Furthermore, musicians are taught numerous other methods for building tension through manipulations of features such as dynamics, note density, or timbre. The premise is that a robot must rely on such musical semantics and have goals that govern individual note choices. A single note sequence may be musically meaningful, but not necessarily possible when considering the robot’s physical constraints, therefore, an alternative solution must be found. The purpose of this generative system is to find such a solution.

Joint Optimization In the previous section the Viterbi algorithm was used to find the physical movement sequence necessary to play a sequence of notes, Y , that was provided prior to the search. Here, there exists an observation sequence for each musical semantic parameter (as shown in Figure 2) and Viterbi is used to find the optimal note sequence derived from both the musical parameters and physical constraints. Therefore, the emission and transition matrices not only constrain decisions to the robot’s embodiment, but also evaluate the note-to-note decisions based on their ability to satisfy the semantic goals.

Each state in S is now defined by a configuration and a note to be played. For Shimon, the state space includes about 73,815 possible arm configurations. This state space is likely to vary considerably depending on the robotic platform. For platforms with enormous state spaces finding the global optimal path is not always feasible if the path is to be found in a timely manner. Therefore, a beam search is applied in

order to prune unreasonable branches quickly. Additionally, the Viterbi algorithm lends itself to distributed computing techniques and can be computed across multiple CPUs if the state spaces are very large.

For each musical semantic parameter, p_n (note density, rhythmic complexity, etc.) a time series of observations is provided (for the purpose of this paper assume these time series are manually provided to the system). The emission score for state, s_i , at observation time, t , describes the aggregate score across all N parameters

$$b_t = \sum_{n=0}^N \lambda_n R(p_{n_t}, s_i) \quad (1)$$

where $R(p_n, s_i)$ describes the instantaneous score of the note in s_i given the semantic parameter and λ_n is applied to each parameter to describe its weight within the overall score. The transition score is described as

$$a_t = \sum_{n=0}^N \lambda_n R(p_{n_t}, p_{n_{t-1}}, s_t, s_{t-1}) \quad (2)$$

where $R(p_{n_t}, p_{n_{t-1}}, s_t, s_{t-1})$ is the score describing how well the transition between notes in the states s_t and s_{t-1} represent the transition between p_{n_t} and $p_{n_{t-1}}$. The optimal path is computed given these emission and transition scores for the semantic parameters.

Qualitative Results Figure 3 shows musical outputs in which the semantic observation sequences are static, but the physical constraints are modified (see video for audio examples¹). Thus, the differences among the outputs are a result of physical design and unique embodiment.

Conclusion

In this paper we argue for integrative generative music system for robotic musicians. We describe a musical path planning method based on Viterbi beam search. By jointly optimizing between a system’s physical constraints and musical reasoning and decision processes we demonstrate that different types of music can emerge.

References

- Albin, A.; Weinberg, G.; and Egerstedt, M. 2012. Musical abstractions in distributed multi-robot systems. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 451–458. IEEE.
- Asada, M.; Hosoda, K.; Kuniyoshi, Y.; Ishiguro, H.; Inui, T.; Yoshikawa, Y.; Ogino, M.; and Yoshida, C. 2009. Cognitive developmental robotics: a survey. *Autonomous Mental Development, IEEE Transactions on* 1(1):12–34.
- Atkeson, C. G.; Hale, J. G.; Pollick, F. E.; Riley, M.; Kotosaka, S.; Schaul, S.; Shibata, T.; Tevatia, G.; Ude, A.; Vijayakumar, S.; et al. 2000. Using humanoid robots to study human behavior. *IEEE Intelligent Systems and their applications* 15(4):46–56.
- Breazeal, C.; Gray, J.; and Berlin, M. 2009. An embodied cognition approach to mindreading skills for socially intelligent robots. *The International Journal of Robotics Research* 28(5):656–680.

¹<https://www.youtube.com/watch?v=S2Yda1ndKVc>

- Bretan, M., and Weinberg, G. 2016. A survey of robotic musicianship. *Communications of the ACM* 59(5):100–109.
- Bretan, M.; Cicconet, M.; Nikolaidis, R.; and Weinberg, G. 2012. Developing and composing for a robotic musician. In *Proc. International Computer Music Conference on (ICMC'12)*.
- Bretan, M.; Hoffman, G.; and Weinberg, G. 2016. Emotionally expressive dynamic physical behaviors in robots. *International Journal of Human-Computer Studies*.
- Chapados, C., and Levitin, D. J. 2008. Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition* 108(3):639–651.
- Cheng, G.; Hyon, S.-H.; Morimoto, J.; Ude, A.; Hale, J. G.; Colvin, G.; Scroggin, W.; and Jacobsen, S. C. 2007. Cb: A humanoid research platform for exploring neuroscience. *Advanced Robotics* 21(10):1097–1114.
- Dahl, S., and Friberg, A. 2007. Visual perception of expressiveness in musicians' body movements.
- Drummond, J. 2009. Understanding interactive systems. *Organised Sound* 14(02):124–133.
- Eitan, Z. 1997. *Highpoints: A study of melodic peaks*. University of Pennsylvania Press Philadelphia, PA.
- Foglia, L., and Wilson, R. A. 2013. Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science* 4(3):319–325.
- Gibson, W. 2006. Material culture and embodied action: sociological notes on the examination of musical instruments in jazz improvisation. *The Sociological Review* 54(1):171–187.
- Godøy, R. I., and Leman, M. 2010. *Musical gestures: Sound, movement, and meaning*. Routledge.
- Hewlett, W. B., and Selfridge-Field, E. 1998. *Melodic similarity: Concepts, procedures, and applications*, volume 11. The MIT Press.
- Hoffman, G., and Weinberg, G. 2010. Synchronization in human-robot musicianship. In *RO-MAN, 2010 IEEE*, 718–724. IEEE.
- Kapur, A.; Lazier, A. J.; Davidson, P.; Wilson, R. S.; and Cook, P. R. 2004. The electronic sitar controller. In *Proceedings of the 2004 conference on New interfaces for musical expression*, 7–12. National University of Singapore.
- Kapur, A.; Murphy, J.; and Carnegie, D. Kritaanjli: A robotic harmonium for performance, pedagogy and research.
- Kidd, C. D., and Breazeal, C. 2004. Effect of a robot on user perceptions. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, 3559–3564. IEEE.
- Kuniyoshi, Y.; Yorozu, Y.; Suzuki, S.; Sangawa, S.; Ohmura, Y.; Terada, K.; and Nagakubo, A. 2007. Emergence and development of embodied cognition: A constructivist approach using robots. *Progress in brain research* 164:425–445.
- Leman, M.; Desmet, F.; Styns, F.; Van Noorden, L.; and Moelants, D. 2009. Sharing musical expression through embodied listening: a case study based on chinese guqin music.
- Leman, M. 2008. *Embodied music cognition and mediation technology*. Mit Press.
- Lerdahl, F., and Jackendoff, R. 1985. *A generative theory of tonal music*. MIT press.
- Lerdahl, F., and Krumhansl, C. L. 2007. Modeling tonal tension.
- Lewis, G. E. 2000. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal* 10:33–39.
- Lim, A.; Mizumoto, T.; Cahier, L.-K.; Otsuka, T.; Takahashi, T.; Komatani, K.; Ogata, T.; and Okuno, H. G. 2010. Robot musical accompaniment: integrating audio and visual cues for real-time synchronization with a human flutist. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 1964–1969. IEEE.
- Livingstone, S. R.; Thompson, W. F.; and Russo, F. A. 2009. Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography.
- Maes, L.; Raes, G.-W.; and Rogers, T. 2011. The man and machine robot orchestra at logos. *Computer Music Journal* 35(4):28–48.
- Metta, G.; Sandini, G.; Vernon, D.; Natale, L.; and Nori, F. 2008. The icub humanoid robot: an open platform for research in embodied cognition. In *Proceedings of the 8th workshop on performance metrics for intelligent systems*, 50–56. ACM.
- Nikolaidis, R., and Weinberg, G. 2010. Playing with the masters: A model for improvisatory musical interaction between robots and humans. In *RO-MAN, 2010 IEEE*, 712–717. IEEE.
- Nusseck, M., and Wanderley, M. M. 2009. Music and motion how music-related ancillary body movements contribute to the experience of music.
- Raes, G. 2014. Robots and automatons catalogue. Available online at http://logosfoundation.org/instrum_gwr/automatons.html (accessed Oct 2, 2014).
- Rowe, R. 2004. *Machine musicianship*. MIT press.
- Shadmehr, R., and Mussa-Ivaldi, F. A. 1994. Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience* 14(5):3208–3224.
- Thompson, M., and Luck, G. 2008. Effect of pianists expressive intention on amount and type of body movement. In *10th International Conference on Music Perception and Cognition, Sapporo, Japan*.
- Toiviainen, P.; Luck, G.; and Thompson, M. R. 2010. Embodied meter: hierarchical eigenmodes in music-induced movement.
- Trafton, G.; Hiatt, L.; Harrison, A.; Tamborello, F.; Khemlani, S.; and Schultz, A. 2013. Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction* 2(1):30–55.
- Trimpin. Portfolio. In *Seattle, Washington*.
- Troy Rogers, S. K., and Barton, S. 2014. Emmi: Expressive machines musical instruments. Available online at <http://expressivemachines.com/> (accessed Oct 6, 2014).
- Vines, B. W.; Krumhansl, C. L.; Wanderley, M. M.; and Levitin, D. J. 2006. Cross-modal interactions in the perception of musical performance. *Cognition* 101(1):80–113.
- Wanderley, M. M.; Vines, B. W.; Middleton, N.; McKay, C.; and Hatch, W. 2005. The musical significance of clarinetists' ancillary gestures: an exploration of the field. *Journal of New Music Research* 34(1):97–113.
- Weinberg, G.; Godfrey, M.; Rae, A.; and Rhoads, J. 2008. A real-time genetic algorithm in human-robot musical improvisation. In *Computer music modeling and retrieval. Sense of sounds*. Springer. 351–359.
- Weng, J. 2004. Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics* 1(02):199–236.
- Whalley, I. 2010. Generative improv. & interactive music project (giimp). *Proceedings of NIME 2010* 255–258.
- Wilson, R. A., and Foglia, L. 2011. Embodied cognition.
- Ziemke, T. 2005. Cybernetics and embodied cognition: on the construction of realities in organisms and robots. *Kybernetes* 34(1/2):118–128.