The Human, the Mechanical, and the Spaces in between: Explorations in Human-Robotic Musical Improvisation

Scott Barton

Worcester Polytechnic Institute, Worcester, MA
sdbarton@wpi.edu

Abstract

HARMI (Human and Robotic Musical Improvisation) is a software and hardware system that enables musical robots to improvise with human performers. The goal of the system is not to replicate human musicians, but rather to explore the novel kinds of musical expression that machines can produce. At the same time, the system seeks to create spaces where humans and robots can communicate with each other in a common language. To help achieve the former, ideas from contemporary compositional practice and music theory were used to shape the system's expressive capabilities. In regard to the latter, research from the field of cognitive psychology was incorporated to enable communication, interaction, and understanding between human and robotic performers. The system was partly developed in conjunction with a residency at High Concept Laboratories in Chicago, IL, where a group of human improvisers performed with the robotic instruments. The system represents an approach to the question of how humans and robots can interact and improvise in musical contexts. This approach purports to highlight the unique expressive spaces of humans, the unique expressive spaces of machines, and the shared spaces between the two.

Introduction

HARMI (Human and Robotic Musical Improvisation) is a software and hardware system that enables musical robots to improvise with human performers. The physical instruments were built by the Music, Perception, and Robotics Lab at WPI and EMMI (Expressive Machines Musical Instruments), and the software was programmed by the author. In order to create a robotic musician that can interact with human performers and at the same time can create music that is novel and compelling, the design of this system incorporates ideas from both cognitive psychology, contemporary compositional practice and music theory. In regard to the former, as robotic and human musical improvisers share perceptual and cognitive capabilities, certain kinds of musical communication are enabled. This does not mean that a robotic improviser has to replicate a human one, and as HARMI does not seek to replicate human improvisers, it differs from previous efforts in the field. Instead, HARMI seeks to both find common ground that facilitates communication, and to explore and exhibit the interpretive and expressive spaces that are unique to machines and humans. Ideas from contemporary compositional practice and music theory are integrated so as to create new kinds of ideas that are musically, and not just technologically, valuable.

Prior work

A number of systems have been developed that enable humans to musically improvise with machines. Efforts in this field are often included in the category "interactive music" (see Winkler 1995 and Winkler 2001) that may or may not include improvisatory capabilities. One category of interactive systems uses software to interpret the gestures of human collaborators (these gestures can be communicated via MIDI or audio data) and then produces sounds (which can be synthesized, pre-recorded and/or processed) through loudspeakers. Specific examples include Cypher (Rowe 1992), Voyager (Lewis 2000), and the Continuator (Pachet 2002). Another category uses software to interpret the gestures of human collaborators but produces sound using physical electro-mechanical instruments. Compared to loudspeakers, physical instruments produce sounds that have unique acoustical qualities, clarify the causality involved in sound production, are capable of nuanced and idiosyncratic expression as a function of their physical design, give visual cues and feedback to human collaborators, and offer visual interest to an audience in performance (some of these features are discussed in Weinberg and Driscoll 2007). Because of these capabilities, such systems offer possibilities for new kinds of musical expression and interaction. Previous work in this area includes Gil Weinberg et al.'s Haile (Weinberg and Driscoll 2006, 2007), Shimon (Hoffman and Weinberg 2011), and Kapur's MahaDeviBot (Kapur 2011). HARMI purports to further the work in this field by integrating ideas from cognitive psychology and contemporary compositional practice and music theory to allow for humans and machines to create new kinds of musical expression in improvisatory contexts.

A Creative Improviser: Musical, Perceptual and Cognitive Considerations

Listening and Responding

Musical improvisation involves action and reaction; processing, interpretation and creativity. We can conflate these ideas into two core phases that a musical improviser inhabits during an interaction: *listening* and *responding*. We can use these general categories to guide the design of a robotic musical improviser. In humans, the concept of listening involves perceptual and cognitive mechanisms. We must interpret sonic information from the world in order to understand temporal relationships, spatial location and source. We understand both individual sonic events, which we can describe using concepts such as pitch and timbre, and larger organizations that comprise those events, such as rhythms, melodies and musical pieces. After we have listened, we choose to either let our collaborators continue to speak, or we may choose to respond with ideas of our own, which are often derived from ideas that we have encoded in memory. Experience, physical capabilities, preferences, and cultural conventions shape these processes (see Lewis 1999).

In order to incorporate such functionality in an improvising machine, programmers typically design *modes* that carry out particular cognitive or perceptual functions. For example, in Rowe's *Cypher*, the composition section of the program includes three modes that transform "heard" material, generate material in a variety of styles, and play material from a stored library (Rowe 1992). The modes of Weinberg et al.'s Haile include imitation, stochastic transformation, simple accompaniment, perceptual transformation, beat detection and perceptual accompaniment (Weinberg and Driscoll 2007). HARMI was designed according to a similar paradigm of modular functionality, although it differs from previous efforts in regard to the musical motivations and aesthetic preferences that shape the kinds of functionality that have been incorporated into the software.

Ideas from Music Composition and Theory

HARMI's purpose is to make contemporary music, therefore, ideas from (contemporary) compositional practice and music theory, such as multiple simultaneous beat rates, non-integer-multiple quantization, rhythmic structures based on frequency-weighted distributions, isorhythms, gesture transformations, and reiteration (including pattern matching) were used to guide which interpretive and choice-making functionality was integrated into the software.

The Beat and Quantization

The notion of the *beat*, or the *tactus*, is considered to be a primary component of rhythmic experience (Krumhansl 2000, which nicely summarizes the work of Fraisse; Honing 2012; Kapur 2011; Janata 2012). The *tactus* is usually one or two rates that is/are described as primary, most salient, "intermediate" and "moderate" relative to other rates in a rhythmic texture (Lerdahl and Jackendoff 1983, p. 21; Parncutt 1994). This is not to say that the rate that we can tap our foot to is the only one of significance. Some theorists, such as Yeston (1976), Krebs (1987) and London (2004) highlight that we are sensitive to multiple simultaneous hierarchically-related regular rates in a rhythmic texture, particularly in the context of meter. The relationships between these rates shape our experience of rhythm. Krebs (1987) describes the relationship between rhythmic levels in terms of metrical consonance and dissonance, which correspond to the degree of alignment between rhythmic levels. Rhythmic alignment is a function of both configuration and the mathematical relationship between rates. In musical perception and production, alignment and the mathematical relationship between rates are limited by the temporal accuracy of the system interpreting and creating those rates. Computer-driven robotic performers, which are capable of high degrees of temporal accuracy, are therefore able to perceive and perform rhythms in ways that human musicians cannot. Musical robots thus open the door for new compositional and improvisational possibilities.

HARMI explores these possibilities by analyzing rhythmic textures to find multiple beat rates, not just the rate that we find most salient as human listeners. The creative process then chooses from these various rates instead of calculating multiples or divisions of a primary beat. By approaching rhythm in this way, novel kinds of rhythmic configurations and relationships can be created.

To illustrate the difference between a system that interprets multiple beat levels as compared to one that adjusts all levels relative to a tactus, consider a rhythmic texture that contains rates at 243 msec, 407 msec and 734 msec.[1] A common approach to automatic rhythmic interpretation is to quantize the temporal locations of elements in order to simplify the proportional relationships between rates.[2] This allows one rate to be specified as the tactus to which the others are related by small integer ratios. Thus, by identifying the primary beat and quantizing the other rhythmic elements relative to that beat, the relationship between 243, 407 and 734 could become 1:2:4. While such an interpretation is consonant with Western notational practice (we now have eighth, quarter and half notes) and the evidence that humans perceive durations categorically (Clarke 1987; Schulze, 1989; Sasaki et al. 1998; Hoopen et al. 2006; Desain and Honing 2003; Krumhansl 2000; London 2004), modifying durations in this way is problematic for a number of reasons. First, restricting rhythmic configurations to small integer ratios produces idealized versions of durations and temporal relationships that typically inspire notions of an undesirable kind of "mechanical" production. Second, these processes filter out (in the case of interpretation) or prevent (in the case of production) rhythmic richness. This richness is a vehicle for expressivity and "feel": some players speed up or slow down eighth notes relative to quarter notes, or play "in front of" or "behind" the beat in order to convey shifts in energy, a certain mood, or define their individual interpretive styles. More universally, this richness can help define musical genres via minute timing conventions, such as "swing" in Jazz, or rubato in Romantic Western Art Music. This richness allows complex superimposition rates and cross-rhythms. Perhaps most importantly (for those interested in making new music), this richness allows composers (and robots programmed by composers) to voice new kinds of rhythmic configurations and relationships that can lead to new kinds of musical styles, conventions and identities. A rhythmic configuration consisting of the aforementioned durations (243 msec, 407 msec, 734 msec) projects an identity distinct from that of its small-integer ratio counterpart. One imagines the diversity of rhythmic identities that are possible given the temporal capabilities of our machines, yet have been relatively unexplored in musical practice (importantly, these rhythms can still be periodic and cyclic, and thus, beat-based). Such configurations and relationships inspire us to think of rhythm in new ways: How will we, as human listeners, experience such mechanically-produced rhythms given our tendency to perceive temporal relationships categorically? How can an artificial intelligence interpret and produce such rhythms while also being sensitive to the temporal categories and expressive timing that characterize human perception? How can a robotic improviser explore new rhythmic territory while simultaneously being able to communicate with human musicians? By enhancing our vocabulary beyond that of simple grids, we open the door to these fantastic rhythmic questions and possibilities.

HARMI was designed with these ideas in mind, thus it can produce quantized rates that are related by complex proportions. This process preserves timing nuance; complex rhythmic configurations and relationships; and subtle tempi alterations made possible by multi-rate analysis, and at the same time, gives the compositional systems a limited set of values from which to find and produce rhythmic patterns.

Metric Frequencies, Transformation and Reiteration

HARMI also makes use of the idea that we are more perceptually sensitive to certain metric positions than others (Palmer and Krumhansl 1990). Some have connected this sensitivity with frequency distributions in meter-based musical canons, which both reflect and shape perceptual and production tendencies.[3] HARMI extends this idea to duration, so that temporal intervals that the system chooses depend on those that were heard in the past. In HARMI, temporal intervals are weighted based on their frequency of occurrence within a particular grouping. This frequency distribution becomes a probability distribution that governs how intervals will be chosen in the process of creating new rhythms. The order of these intervals is chosen according to transformational processes.

Transformation is a core component of musical compositional practice and is implemented in HARMI in a number of ways. When we transform a musical idea, we reiterate some components of the original gesture while at the same time adding other ideas around those original components. We can understand a statement *A* and its transformation *B* by considering the number of operations (such as additions, subtractions, and substitutions) that are required to turn *A* into *B* (which also can describe inter-entity similarity: see Hahn et al. 2003; Orpen & Huron 1992). We can use these *transformational distances* (Hahn et al. 2003) to represent and create new ideas from ones heard in an interaction. In HARMI, transformations occur via random and sequential processes. In one mode, a rhythm is transformed by substituting durations one at a time: the location of the substitution within the sequence is chosen at non-repeating random, and the duration is chosen from the probability distribution. In another, the number of alterations that is to be made is randomly chosen within a restricted range, the location of those transformations within the sequence is chosen, and then the durations to be substituted are chosen from the probability distribution.

[1] Here, a "rate" is a single numerical representation of some distribution of IOIs (inter-onset intervals).

[2] There are number of different approaches to the problem of quantization: see Desain 1993 and Desain and Honing 1989 for a discussion of the topic.

[3] David Huron discusses this idea and related research in *Sweet Anticipation*.

The number of alterations is not restricted to the length of the phrase, so that additions can be made.

Reiteration without transformation also plays an important role in musical improvisation. There are a number of reasons for this. When one musician repeats the idea of another, he shows that basic communication is occurring successfully (if you are really listening to me, you can repeat exactly what I said). It also motivates an ensemble towards a shared idea, which can then be the source of future musical explorations. Thus, the system has the ability to reiterate the pitch and durational sequences of human collaborators.

Within the category of reiteration, HARMI has the ability to match both pitch and rhythmic patterns played by performers. In one mode, when the system detects a pattern, it will then repeat that pattern a certain number of times, after which it will transform the pattern in one of the ways discussed above. The system can find and express pitch and rhythmic patterns independently. Given the temporal variability of human musical performance, multi-rate quantization, as described earlier, is used in order find patterns.

The above shows the extent to which the system is as much a composition as it is an autonomous agent, and thus it expresses the values and aesthetic preferences of the author. At the same time, it interprets and produces ideas in a way that the author didn't necessarily anticipate, thus we can say it expresses ideas in its *own* way, which provides novel and interesting ingredients to a musical mix. The design of the system therefore creates spaces for the human, for the mechanical and for the areas in between. The character and boundaries of these spaces are ready to be defined and explored by composers and performers.

HARMI in Practice; Future Directions

HARMI was partly developed in conjunction with a group of human improvisers during a residency at High Concept Laboratories in Chicago, IL in July 2013. The rehearsals during this residency allowed performers to share their thoughts about the system and how it could be improved.

These rehearsals revealed the need for feedback when the HARMI was listening. Because of the number of musicians involved in the improvisations (up to three humans and two robots), the human improvisers sometimes found it difficult to determine when and "where" the robot's attention was directed. Visual solutions to this problem include lights, projection screens, or anthropomorphic electromechanical structures (the latter is a feature of Weinberg et al.'s Shimon) that illuminate, display or move to convey the robot's attention. Alternatively (or in combination), one could utilize an auditory feedback system that produces a sound when a note is heard. An auditory system has a number of advantages over a visual one in musical contexts. First, an auditory feedback system requires that an improviser actively attend to the individual components of a musical texture in order to distinguish and interpret the auditory feedback. This is not necessarily the case in a visual system, which may cue a visually sensitive improviser whose auditory attention is not focused on the rest of the musicians, or the music as a whole. The latter is a problem: careful listening is an essential part of musical communication. Second, an auditory feedback system allows the human performers to understand how the robot "hears" in the same language (one of rhythms, pitches, phrases, etc.) and modality that they are "speaking". This understanding can motivate human musicians to play, experiment, and create in new ways.

The system was therefore modified so that auditory cues were given when HARMI heard a tone onset. It became clear how the machine interpreted some gestures as a human musician would, and some gestures in unique ways. This bit of functionality proved to be inspirational to the human musicians, who subsequently experimented with the ways that HARMI "hears". As a result, the musicians learned how to communicate with the machine, which invited the musicians to try new ways of playing, which caused the output from the robots to be unexpected and interesting. As positive surprise provides some of the greatest moments in music, particularly improvised music, these results were successful and inspirational.

Understanding how HARMI "hears" in unique ways excites ideas about other ways a robotic improviser could interpret sonic information. When presented with microtonal pitch sequences or multiphonic-rich passages, frequency relationships could be translated into temporal ones (this kind of functionality is particularly important for instruments that improvise in contemporary musical contexts but are restrained by equal temperament). A musical robot could interpret rhythmic patterns in ways that a human performer typically would not. For example, the run and gap principles (Garner and Gottwald 1968; Rosenbaum and Collyer 1998) describe how human listeners perceptually organize auditory cyclic temporal patterns. An artificial musical intelligence does not have to be governed by the same principles, and thus may choose pattern beginnings and configurations in unique ways. The combination of these interpretations with human ones could produce interesting musical textures.

The rehearsals also motivated questions about memory and musical form. While human memory privileges musical information heard in the recent past (Brower 1993; Snyder 2000), an artificial intelligence need not be governed by the same sorts of temporal constraints. A musical robot could reproduce an idea voiced at any point in an improvisation's past. These recollections could be reproduced exactly, or they could be colored and transformed in a variety of ways depending on what else had been perceived during the musical interactions, or on other kinds of knowledge encoded in the memory system. These sorts of alternative recollective capabilities could provide structure that would allow human-robot collaborators to create new improvisational forms.

These results and contemplations reflect the core approach taken in the design of HARMI, which is the attempt to find not only the shared perceptual and production spaces between human and robotic improvisers, but to also highlight those spaces that are uniquely human and uniquely robotic. As these spaces are explored in greater depth through integration of learning, sensitivity to physical design, higher-level perceptual abilities, aesthetic preferences and stylistic conventions, new kinds of music will be created.

References

Brower, C. 1993. Memory and the Perception of Rhythm. *Music Theory Spectrum*, 19–35.

Clarke, E. F. 1987. Categorical rhythm perception: An ecological perspective. *Action and perception in rhythm and music*, *55*, 19–33.

Desain, P., & Honing, H. 1989. The quantization of musical time: A connectionist approach. *Computer Music Journal*, *13*(3), 56–66.

Desain, P. 1993. A connectionist and a traditional AI quantizer, symbolic versus sub-symbolic models of rhythm perception. *Contemporary Music Review*, *9*(1-2), 239–254.

Desain, P., & Honing, H. 2003. The formation of rhythmic categories and metric priming. *Perception*, *32*(3), 341–365.

Garner, W. R., & Gottwald, R. L. 1968. The perception and learning of temporal patterns. *The Quarterly journal of experimental psychology*, *20*(2), 97–109.

Hahn, U.; Chater, N.; and Richardson, L. B. 2003. Similarity as Transformation. *Cognition* 87.1: 1–32.

Honing, H. 2012. Without It No Music: Beat Induction as a Fundamental Musical Trait. *Annals of the New York Academy of Sciences* 1252.1: 85–91.

Hoopen, G. T., Sasaki, T., and Nakajima, Y. 1998. Categorical rhythm perception as a result of unilateral assimilation in time-shrinking. *Music Perception*, 201–222.

Janata, P.; Tomic, S. T.; and Haberman, J. M. 2012. Sensorimotor Coupling in Music and the Psychology of the Groove. Journal of Experimental Psychology: General 141.1: 54–75.

Kapur, A. 2011. Multimodal Techniques for Human/Robot Interaction. Musical Robots and Interactive Multimodal Systems. Solis, J. and Ng, K eds. Springer Berlin Heidelberg. 215–232.

Krebs, H. 1987. Some extensions of the concepts of metrical consonance and dissonance. *Journal of Music Theory*, *31*(1), 99–120.

Lerdahl, F. A. and Jackendoff, R. S. 1983. A generative theory of tonal music. The MIT Press.

Lewis, G. E. 1999. Interacting with Latter-day Musical Automata. Contemporary Music Review 18.3: 99–112.

Lewis, G. E. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. Leonardo Music Journal 10: 33–39.

London, J. 2004. Hearing in Time : Psychological Aspects of Musical Meter. New York: Oxford University Press.

Orpen, K. S., and Huron, D. 1992. Measurement of Similarity in Music: A Quantitative Approach for Non-parametric Representations. Computers in music research 4: 1–44.

Pachet, F. 2003. The Continuator: Musical Interaction with Style. Journal of New Music Research 32.3: 333–341.

Palmer, C., & Krumhansl, C. L. 1990. Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728–741.

Parncutt, R. 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception, 11, 409–409.

Rosenbaum, D. A., & Collyer, C. E. 1998. Timing of Behavior: Neural, Psychological, and Computational Perspectives. The MIT Press.

Rowe, R. 1992. Machine Listening and Composing with Cypher. Computer Music Journal 16.1: 43–63.

Sasaki, T., Hoopen, G. T., & Nakajima, Y. 1998. Categorical rhythm perception as a result of unilateral assimilation in time-shrinking. *Music Perception*, 201–222.

Schulze, H.-H. 1989. Categorical perception of rhythmic patterns. *Psychological Research*, *51*(1), 10–15.

Snyder, B. 2000. *Music and memory : an introduction.* Cambridge, Mass.: MIT Press.

Weinberg, G., and Driscoll, S. 2006. Robot-human Interaction with an Anthropomorphic Percussionist. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1229–1232. New York, NY, USA: ACM.

Weinberg, G., and Driscoll, S. 2007. The Interactive Robotic Percussionist: New Developments in Form, Mechanics, Perception and Interaction Design. In Proceedings of the ACM/IEEE International Conference on Human-robot Interaction, 97–104. New York, NY, USA: ACM.

Winkler, T. 2001. *Composing Interactive Music*. The MIT Press.

Winkler, T. 1995. Strategies for Interaction: Computer Music, Performance, and Multimedia. In Proceedings of the 1995 Connecticut College Symposium on Arts and Technology.

Yeston, M. 1976. *The stratification of musical rhythm*. Yale University Press New Haven.