

Maxine's Turing Test – A Player-Program as Co-Ethnographer of Socio-Aesthetic Interaction in Improvised Music

Ritwik Banerji

Center for New Music and Audio Technologies, 1750 Arch Street, Berkeley, CA 94720
Department of Music, University of California Berkeley
ritwikb@berkeley.edu

Abstract

Beyond the goal of refining system design to the needs and tastes of users, user evaluation of interactive music systems offers a method of examining the nature of musical creativity as understood by its human practitioners. In the case of improvising music systems, user study and evaluation of a system's ability to improvise may be useful in the ethnomusicological study of musical interaction in contemporary improvised music. A survey of preliminary findings based on the interactions of an improvising system, Maxine, with several improvisers is discussed, with results suggesting methodological reconfigurations of the purpose and goals of evaluating of interactive musical metacreatations.

Introduction

Maxine is an interactive musical agent based in Max/MSP, designed create music in real-time based on audio input in dialogue with a human performer, in a manner reflective of freer practices of improvisation (Banerji 2010). Within the taxonomy of interactive music systems, Maxine can be classified as either a "player program" or "Live Algorithm for Music" (Young 2008), interacting with audio input from the environment and producing her own contributions in the form of live audio processing and control of MIDI-based musical events. On one level, the goal of this research is the development of a machine capable of real-time improvisative creativity reflective of practices developed by improvisers over the past 50 years. Ideally, Maxine should improvise with and against human musical improvisers in a manner which "passes for" a human being engaging in contemporary practices of post-jazz-, or amending Derek Bailey's term, less-than-idiomatic-improvisation (Bailey 1993). On another level, to pursue this first goal, I frame this project as an effort to prepare Maxine to pass a kind of musical improviser's Turing test.

Amending the goal of fooling a human interlocutor, however, design of Maxine also seeks to produce interactive musical behavior sufficiently interesting and inspiring to a human performer such that her position in relation to a man-machine binary is made irrelevant.

With these goals in mind, it has been essential to invite human improvisers to play with Maxine and solicit their critique of her skills in improvising. As improvisers are asked to assess her ability to improvise and offer directions for improvement, the conversation consistently leads improvisers to address a more basic question of "what is improvisation?" Specifically, these conversations reveal the ways in which aesthetic evaluation of improvisation is at once an evaluation of social skills of interacting through sound, pointing to the fundamentally social nature of collaborative improvisations. Formally, these conversations are part of a user study a designer uses to more accurately match design to habits and culture of users. However, speaking to the issues of evaluation methodology raised in literature on musical metacreation (Eigenfeldt et. al. 2012), I suggest that user-studies with Maxine, and perhaps other player programs, can serve the anthropologist of music's goals of probing the nebulous ambiguity of social interaction among improvisers, through their critique of an improvising system that models and reproduces their socio-aesthetic practices. In this paper, I offer a summary of initial findings from improvisers' critiques of Maxine, outlining some ways that user study of improvising systems may help address the social-scientific and aesthetic problem of understanding how, why, and what improvisers expect of one another in real-time musical interaction.

Evaluation Methodology of Musical Metacreations and the Impetus for this Study

Over the past few decades, the theory of interactive music systems has reached several levels of new classification. Rowe introduced a useful binary between player and instrument programs, drawing this distinction between systems treated as semi-autonomous agents in performance versus those which remain the object of control of a human performer (Rowe 1993). Building on the player program concept, Blackwell and Young develop the concept of live algorithm, and discuss the ideals for how such systems should be designed (Young 2008). As noted by several researchers (Eigenfeldt et. al. 2012, Hsu 2009), while such advances in artificial creativity are key steps towards the challenges of AI research as originally framed, few researchers have documented their efforts to engage humans in evaluating such systems.

In the context of musical metacreation, evaluation of improvising music systems poses two specific challenges. First, while it is possible to evaluate some musically metacreative systems on the basis of audience, domain-expert, and academic perspectives (Eigenfeldt et. al. 2012), in the case of improvising music systems, performer perspective must be key. Placing interactive system evaluation in the context of the ongoing debate on the efficacy of the Turing test, Ariza makes the important observation that such studies must prioritize the perspective of the human subject who interacts with such systems, emphasizing that this perspective differs greatly from the removed vantage point of the audience, academic, or domain-expert (Ariza 2009). However, performers of improvised music are by no means homogenous in style, approach, or sound palette. Similarly, their expectations of such systems may diverge widely, as shall be discussed below in an overview of initial findings. Second, as Derek Bailey has noted in his influential treatise on the subject, “improvisation enjoys the curious distinction of being both the most widely practiced of all musical activities and the least understood.” (Bailey 1993) From my initial inquiries as an ethnographer of communities of improvised music, Bailey’s provocative characterization retains its accuracy. Considering the evaluation of a system aiming to improvise “freely”, like Maxine, Voyager, OMax, London/ARHS, it becomes clear that the task of such designers is radically ambiguous. This necessarily complicates any simple understanding of user feedback as a clear indication of future directions for software development. However, as a social scientist of music, I interpret these methodological quandaries not as an impasse for research, but as indication of a possible method to probe the nebulous questions of how performers perform their personal identities through their strategy of interacting itself, as well as a method of engaging the

project implied in Bailey’s bold remark. The perspective of human performers of improvised music on such systems does as much to illustrate directions for system development as it does to answer questions of how collective improvisations inscribe theories of social interaction, an issue of key interest for ethnomusicologists engaging this research area. Each critique of Maxine indicates the expectations of the particular subject, first with regard to aesthetics, but more crucially, their expectations and understandings of improvisation as a social interaction.

From the beginning of my development of Maxine, I have prioritized human improvisers’ critiques of Maxine’s ability to improvise. Pursuit of their perspective has come from two basic motivations: due diligence in systems design and a desire as a younger improviser to understand the perspective of older and more experienced players. Since my goal is for Maxine to improvise, it seemed essential to consult improvisers who both have more experience and whose performance styles I respect and admire. As an aspiring improviser, I was always curious to know how other improvisers understood their practice. Additionally, initial experiences in developing an interactive system had already begun to challenge my definition of improvisation, reveal how subjective one’s definition of improvisation can be, and highlight the deep relationship between improvising and the concept of interactivity (Banerji 2010). Seeing early stages of Maxine as already a kind of performing model of improvisation, I was eager to understand how other improvisers might critique this representation, and how this conversation might indicate how each of them as individuals understood improvisation for themselves.

From initial interactions between Maxine and human improvisers, it became clear that the due diligence-oriented frame question of “how well does Maxine improvise?” nearly always led to answers indicative of the values and expectations improvisers hold of each other in real-time musical interaction. Additionally, because Maxine is a machine, several improvisers have found themselves more comfortable critiquing her behavior than they would a human being. Thus, as a field method in the ethnography of improvisation, the critique of a software-based representation of a musical practice by its human practitioners may provide insights beyond the typical approach of ethnomusicologists interested in musical practices, in which a scholar takes the position of student of a master of a particular musical tradition as the vantage point of their research.

Method and Preliminary Results

Again, methods for evaluating systems of this kind have yet to be fully elaborated in literature. In soliciting the

feedback and commentary of improvisers on Maxine's abilities as an improviser, I have experimented with two methods: a conversational approach derived from practices of ethnography and a more quantitatively driven approach derived from cognitive science. The more ethnographic method involves an improviser playing in a non-performance setting with Maxine. As in informal playing sessions, improvisers played short pieces with Maxine, interspersed with commentary on Maxine's abilities to improvise and questions about how Maxine works. Open-endedness was the principal advantage of this method, allowing improvisers to discuss the behavior issue of Maxine they thought was most pressing or striking in their initial interactions. This approach roved around a wide set of questions, whereas the more controlled, formal method derived from experimental psychology allowed for pointed exploration of improvisers' evaluations of Maxine around specific criteria. In the end, given the range of factors affecting the course of improvisations between humans, it is not clear if one method will prove more useful over time. With their relative strengths and weaknesses, both ethnographic and quantitative methods will likely be useful in future studies.

Just as when they evaluate the improvisation of a human collaborator, many considerations affect responses to the simple question of how well an improvisation went. These factors include sound quality/timbre, interaction, musical tension and release, and other loosely defined parameters varying in importance between individuals. Specifically, an improviser's critique of Maxine's ability to improvise may be a reaction to the sound of the system itself, rather than the way the system processes sound information. An improviser may find themselves inspired by the sounds played by a system, even though they are not as pleased by the way it has been designed to deal with audio input from the performance environment and interact with the human partner. Evaluations of the skill of "improvisation" may be as dependent on how an improviser, human or machine, responds to other players as it would be on the quality or range of timbres he or she produces in response.

An ethnographic method where the human subject determines the direction of the conversation more effectively allows for communication of what subjects find most important in improvisation. This may be the key advantage of an ethnographic approach to user study over quantitative and controlled methods, in which the subject's input is framed around pre-determined criteria and points for examination. Since the design goal specifies that human-machine interaction in these situations should feel improvised, it is entirely likely, if not at least desired by researchers and performers in this subfield, that the results of the interaction will be a surprise. Given the range of musical gestures and behaviors which constitute a given improviser's criteria for evaluating the skill of

"improvisation," an open-ended method of evaluation may prove most fruitful in seeking the critiques most important to an individual improviser.

Furthermore, improvisers demonstrate a profound resistance to quantitative and normative evaluations of their practice. Banal as it may be to repeat the refrain that there is no objective measure for creativity, this fact must be considered in interpreting how improvisers respond to a quantitatively driven query. As one influential Chicago improviser claimed to me, "there is no such thing as a 'good' improviser." From his, and other improvisers', choices of collaborators to their aesthetic judgments of other artists, it is clear that his maxim is more discursive than a real reflection of what improvisers think. Nevertheless, it is crucial to consider how the verbal performance of this anti-normativity may distort the results of quantitatively driven user studies of such systems.

1st Study – Ethnographic Method

The first informal study was conducted in the summer of 2009, at a point when Maxine had reached sufficient functionality to begin asking for critique from improvisers. From experiences improvising with other players in Chicago, I found that post-playing conversations were more often free-form than pointed examinations of the playing which had just transpired. Consequently, I chose to mirror this situation by choosing a more informal approach at this stage of user study. Sessions consisted of a series of short duo improvisations between Maxine and the guest improviser, followed by the improviser's comments and questions about how Maxine processes audio input to make performance decisions. Results of this informal user study indicated a range of features improvisers would expect from an improvising system like Maxine, assuming that Maxine is held to the same standard that would be used to judge a human co-improviser. Full analysis of these results will not be discussed here, but several trends are salient enough for remark. Features or skills desired of Maxine include:

- Emulation of human improvisers personally familiar to the guest improviser
- Independence of Maxine's behavior and non-reliance on the input of a human performer
- Non-direct relationship between volume and event density of Maxine's playing as compared to guest improviser
- Capacity to shift between leader and follower roles at musically appropriate moments
- Unpredictability of Maxine's behavior and cognition
- Relevance of Maxine's sonic responses (as subjectively defined by the human interlocutor)

- Recognition of musical features, such as meter, pitch-relationships/harmony, phrasing, etc.
- Creation/simulation of “form” in real-time

Reflective of the heterogeneity of tastes and approaches of improvisers alluded to before, the capacities desired of Maxine were inconsistent across the group of improvisers surveyed. While several individuals desired each of the skills above, an opposing trait would be desired nearly as often, even by the same improviser. For example, as important as it was for Maxine to emulate the response protocol of improvisers the human most admired or was most familiar with, at other moments some individuals expressed disappointment that Maxine had too closely mimicked the improvisers they tended to collaborate with. Indeed, my primary inspirations for Maxine’s interactional principles were based on my observations of Chicago improvisers, and so these critiques were understandable, even as they signified the achievement of some measure of my original goals. Regardless of the feature or skill specified above, no consistent rules can be derived from improvisers’ evaluations of Maxine. Again, these conversations did more to reveal the variety of impressions of the concept of improvisation than they did directions for system development.

Overall observations that improvisation is a practice most closely resembling nonlinear systems (Borgo 2006) are confirmed in the interactions of improvisers with Maxine. While some or all of the qualities noted above may be intuitively theorized as good or successful improvising strategies, the key advantage of a conversational method is its provision of the opportunity for a researcher to situate the features improvisative competence listed above in specific contexts, with a variety highlighting the nonlinearity of improvisative aesthetics through numerous counterexample moments. Indeed, one improviser pointed to nonlinearity as a key feature in his evaluation of the notion of competence as an improviser. Maxine reminded him of young improvisers in the way that she, in this particular interaction, tended to follow him upwards in volume and density, while becoming “shy” if he was quiet for a moment.

2nd Study – Greater Control, Less Flexibility

In the subsequent study taking a more controlled experimental approach, many of the same trends are evident. Eight improvisers active in the San Francisco Bay Area improvising scene were invited to perform 10 short improvisations with Maxine in an experimental, non-performance setting at CNMAT in November 2010. Each improviser played 10 short improvisations with Maxine. As a control, in a handful of randomly selected takes, Maxine was set to not listen to live audio input, but process and react to pre-recorded material, without the subject

being informed of this switch. This was intended to examine whether an improviser would detect Maxine’s active listening through her performance behavior and how this would affect their satisfaction with the interaction. Improvisers were asked to quantitatively grade Maxine in the following areas after each take: relevance of Maxine’s playing, degree to which Maxine inspired the subject to respond differently, satisfaction with Maxine’s responses to musical input, and overall meaningfulness of the interaction. Additionally, improvisers were also given a qualitative prompt to describe whatever they felt was “missing” from each interaction. Results of this study will not be discussed in full here, but some methodological issues of evaluating such systems are clarified, issues which themselves indicate certain realities of collaborative improvising between humans.

Repurposing User-Study for Music Anthropology

Most crucially, asking improvisers to critique Maxine overlooks their own role in how Maxine behaves. Their evaluation of Maxine cannot be separated from an evaluation of their own performance in terms of how they interacted with Maxine. In both the holistic approach of the first study and the four evaluation categories of the second, improvisers were asked to evaluate Maxine, but not themselves. Given that these interactions were arranged in the context of improving a system, this approach is intuitive from a design perspective. However, holding that ensemble improvisation is a collaborative process, evaluations of Maxine’s abilities to improvise ought not to be considered without also paying attention to the abilities of the human subject to do the same, at least insofar as each subject demonstrated this capacity in the experiment. In other words, an evaluation of Maxine’s ability to improvise assumes that the quality of the improvisation is solely determined by the design of this system. It is more likely that the quality of an improvisation is mediated between the skills of all agents participating. Methods focusing on the evaluation of the system, and not soliciting self-critique of the human participant, will lead performers to comment as if the system is primarily responsible for the quality of the improvisation.

Despite my focus on questions oriented in this direction, some improvisers indicated that the overall quality of the improvisation, and thus Maxine’s interaction with them, was partly related to their own playing and relative familiarity with the system. In the most striking example, a participant of the second study, when asked “what was missing?” at the end of each take, listed several desired musical features and skills similar to those enumerated above. However, by take 7, this participant claimed that

what was missing from the interaction was “more rehearsals with Maxine.” Takes 8, 9, and 10 were all given the response “see #7.” Through this telling and witty response, he indicated his feeling that the quality of the improvisation was not something he could judge simply based on how Maxine played. In his view, the quality of each take was as much a result of his own playing as it was of Maxine’s design.

From these initial inquiries into the user evaluation of a system like Maxine, question-framing which focuses on the quality of the system itself may overlook a reality of human improvising, in which no one individual can be said to be solely responsible for the ultimate progress of the musical event. Likewise, the more truly interactive a system is, the more dependent its behavior should be on how a human being chooses to interact with it. In the case of human collective improvisation, the relative success of the performance cannot be traced to the actions of just one agent. These initial studies also illustrate that future research into the evaluation of such systems may ask human participants to critique themselves and their own engagement in the situation. All the same, the artificiality of focusing critique on the machine agent in a collaborative improvisation still allows for examination of the complexity and variability of what improvisers expect of one another in performance.

Results indicating the response to experimental condition itself are also seen in the fact that several improvisers exhibited what I describe as “testing behaviors.” Principal examples include the repeated use of similar gestures to see how Maxine would respond to such sonic input. Some improvisers indicated that their deployment of testing behaviors was borne of the artificiality of improvising with a machine. Nevertheless, if an improviser encountering a new agent like Maxine engages such testing behaviors, this raises the question of whether improvisers may exhibit similar kinds of “testing behavior” in what has been termed “naive rehearsal,” (Hsu 2008) or “music meeting”, (Lewis 2004) situations of improvisers meeting for the first time in performance or rehearsal. In some communities of improvised music, these kinds of first-time interactions are a regular part of social practice and this kind of user study may also prove quite useful in probing the group psychology of how these musical meetings occur and what they mean for different participants.

Conclusions and Future Directions

Though interpretations of these data are offered preliminarily, several trends are indicative of future directions of development for Maxine as well as methodological reconfigurations of evaluating player

programs or live algorithms for music. While the interpretation of this data is only preliminary in the absence of a full manifest of how Maxine, as a system, is designed, I interpret these data from a perspective informed by George Lewis’ observations about his own work, *Voyager*. Positioning his work in the context of 20th century African-American expressive arts, Lewis claims his work to be reflective of his membership of the Association for the Advancement of Creative Musicians and experiences with African-American art and social life in post-war Chicago (Lewis 2000). Similarly, Maxine’s design is a reflection of my performance experience as an improviser, observation of Chicago improvisers in performance, and my engagement with discourses about freedom, improvisation, and interactivity. Thus, if Maxine’s approach to real-time interaction is reflective of a particular cultural location, as Lewis claims of *Voyager*, then reactions of improvisers in Chicago to Maxine indicates their relative distance from her on what constitutes a kind of cultural map of preferred modes of interaction. On this virtual map, Maxine’s location can be assessed based upon how she is designed to react to information in real-time, as well as triangulating this location from the aesthetic and social similarity of improvisers (in terms of frequency of collaboration in performance and offline, embodied social networks, for example). Considering future directions for research in the evaluation of interactive music systems, I propose a collaboration which may address some of the problems of determining cultural location based on improvisers’ distances to Maxine alone. An improviser’s feeling of socio-aesthetic distance from one system, like Maxine, only puts participants at varying radii in this proposed socio-aesthetic cultural map. In future experiments, researchers in this field may learn a great deal from experiments in which the same set of human improvisers is invited to critique a set of extant improvising systems (such as IRCAM’s OMax system, David Plans Casal’s Frank, William Hsu’s London and ARHS, to name a few). Based on the critiques of improvisers, a multidimensional scaling model, as used in David Wessel’s study of timbre (Wessel 1979), may allow for normalization of the virtual cultural locations of these systems, based on analysis of their design, comments of improvisers, and analysis of their individual playing styles.

Ultimately, research in the field of improvising music systems, when coupled with diligent attention to the critique of such systems from improvisers themselves, will do much to yield systems which improvise, interact, and come closer to achieving forms of artificial creativity. However, alongside the goal of artificial creativity, such projects offer a space for technologists, humanists, and social scientists to probe the nature of human creativity, as understood by its human practitioners. Software models of

human creativity, when critiqued by the human practitioners they represent, create a venue and method for the explicitation of processes at work in real-time interactive musical creativity, a subject which has rather consistently defied and thwarted clear examination in much previous research.

References

- Ariza, Christopher. 2009. "The Interrogator as Critic: The Questionable Relevance of Turing Tests and Aesthetic Tests in the Evaluation of Generative Music Systems." in *Computer Music Journal* 33(1): 48-70.
- Assayag, G, and S Dubnov. 2004. "Using Factor Oracles for Machine Improvisation." *Soft Computing* 8: 1-7.
- Bailey, Derek. 1993. *Improvisation: It's Nature and Practice in Music*. London, England: Da Capo Press.
- Banerji, Ritwik. 2010. Maxine Banerji: The Mutually Beneficial Practices of Youth Development and Interactive Music System Development. *eContact: Journal of the Canadian Electroacoustic Community* 12 (3).
- Borgo, David. 2005. *Sync or Swarm: Improvising Music in a Complex Age*. New York: Continuum Books.
- Casal, David Plans, and Davide Morelli. 2007. "Remembering the Future: An Overview of Co-Evolution in Musical Improvisation." In *Proceedings of International Computer Music Conference*. Copenhagen, Denmark
- Eigenfeldt, Arne, Adam Burnett, and Philippe Pasquier. 2012. Evaluating Musical Metacreation in a Live Performance Context. In *Proceedings of the Third International Conference on Computational Creativity*, 140-144. Dublin, Ireland.
- Hsu, William. 2008. Two Approaches For Interaction Management in Timbre-Aware Improvisation Systems. In *Proceedings of the International Computer Music Conference*. Belfast, Ireland.
- Hsu, William, and Marc Sosnick. 2009. Evaluating Interactive Music Systems : An HCI Approach. In *Proceedings of Ninth International Conference on New Interfaces in Musical Expression*, 25-28.
- Lewis, George E. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal* 10: 33-39.
- Lewis, George E. 2004. Gittin' To Know Y'all: Improvised Music, Interculturalism, and the Racial Imagination. *Critical Studies in Improvisation* 1(1).
- Rowe, Robert. 1993. *Interactive Music Systems: Machine Listening and Composing*. Cambridge, Mass.: MIT Press.
- Wessel, David L. 1979. Timbre Space as a Musical Control Structure. *Computer Music Journal* 3 (2): 45-52.
- Young, Michael, and Tim Blackwell. 2004. Self-Organised Music. *Organised Sound* 9(2): 123-136.
- Young, Michael. 2008. NN Music: Improvising with a Living Computer. In *Computer Music Modeling and Retrieval*, 337-350. Berlin: Springer Verlag.