Improvisational Media Space ::
Architecture and Strategies for Evolution

Paul Nemirovsky and Rebecca Luger-Guillaume

MIT Media Laboratory, 20 Ames St., Cambridge, MA, USA 02139
`{pauln, beccalg}@media.mit.edu`

Abstract. This paper presents the current state in an ongoing development of the Emonic Environment (EE): a real-time improvisational system employing evolutionary principles for the mutation of media space. We position the associated problems in the context of user interaction, provide eight principles essential for creating an improvisational media environment, follow with a description of how the EE implements these principles, and conclude with a description of the evolutionary algorithms' functionality.

1 Introduction

These days, a desktop PC is capable of real-time recording, processing, and playback of audiovisual media. It can exchange sounds and images with remote users, or access vast and readily available databases online. With this wealth of information around, how can a computer help us in structuring it, perusing and organizing the media into something that is novel and *uniquely ours*? Furthermore, how can that assistance be provided in real-time and with no precise planning / guidance given by the user? To address this problem, in section II, we propose integrating the power of an *editor* and the ease-of-use of a *browser* into a new media interaction paradigm, based on the principles of non-idiomatic improvisation[1]. We posit that these principles can provide us with a powerful framework for addressing the problem. In section III, we present an up-to-date version of the principles (initially proposed in [1]). In order to exemplify the new paradigm, in section IV, we present the Emonic Environment, a system that makes computer-assisted construction and navigation of media space possible. While improvisation offers a general theoretical framework, to implement this framework, a method is needed. In section V, we argue that genetic algorithms are appropriate for the task and describe the problems encountered in their implementation within the EE.

---

[1] Not following one fixed aesthetic idiom, such as a particular music style.

2 Beyond Editor and Browser

Today’s world of computer interaction largely remains a cold and impersonal affair: a participant directs input to a computer, which blindly and mechanically responds in the exact way it was programmed. The types of input and response vary dramatically, with two prevalent paradigms emerging: the editor and the browser.

The editor paradigm forces us to regard any creative exploration in terms of individual components: notes of a sonata, files of the peer-to-peer, shots of a movie, articles of a newspaper. The computer is passive; it is capable of responding to a fixed set of commands by performing a fixed set of functions, but provides no inspiration. As humans, we are expected to know precisely when and how to activate functions and produce results. That view is limiting, for it does not prompt further learning and limits us to what we already know and understand. Furthermore, we frequently struggle to describe our experience in creating and perceiving music, stories, or films in terms of individual components; a higher-level, structural view is required.

Such a view is provided by the browser paradigm, located at the opposite end of the interactive spectrum. The browser paradigm regards the computer as a black box, guided by a set of algorithms hidden from human view. The user is relegated to controlling a high-level space to the extent defined by the browser’s rules, with no ability to affect the internals of the ongoing processes or dynamically change the structures of what is being browsed.

This editor / browser dichotomy negates what is special about human interaction: the ability to dynamically switch between high- and low-level views of the process being explored; to think structurally. Looking at the creative process as a type of search, both paradigms can be said to assume that the participants have and are aware of a fixed objective, even before the search begins. As a result of non-structural thinking and an implied search objective, we are denied further learning or creativity.

As researchers further realize the value of improvisational creativity, we see a growing effort to combine evolution and improvisation. Projects such as Voyager [2], Galapagos [3], ChaOs [4], and Swarm Music [5], while not necessarily explicitly improvisational, present media spaces that incorporate evolutionary change and user feedback. We attempt to further expand this area of research by creating a playground for building non-idiomatic improvisational media environments that allow their users to move freely on the editor/browser axis while providing both computer-influenced and human-driven evolution in an improvisational context.

3 Improvisational Principles

The EE is inspired by the ideas of non-idiomatic improvisation and experimental music tradition [6]. In this section, we present our reading of these ideas formulated as 8 principles upon which the EE is built. We focus on aspects that deal with the nature of the performance, the type of audience interaction, and the role of evolution as a vehicle for exploring the media space.

1. Dynamic nature of improvisational structures. Structural representations used in the course of improvisation are incorporated, modified, and purged dynamically to

satisfy the improviser's changing focus (as affected by both aesthetic criteria and stimuli from the outside environment). We implement this principle by constantly initiating and accounting for change in the system's state through the evolutionary process that is able to adjust to the incoming new stimuli.

2. Changing, multi-leveled focus. An improviser thinks about what he is doing at different levels of abstraction simultaneously. Continuously switching between the macro- and micro-levels, he attends to the minute details at one moment, only to switch to structural development (e.g., a climax) a second later. We implement this principle by providing a visualization of the EE's underlying multi-layered architecture, allowing the participants to switch focus between particular elements and a higher-level structural view of the system's activity.
3. Diversity of types. Improvisation is a result of interrelating multiple perceptual inputs and memories; an improvisation resulting in a video 'output' is nevertheless an improvisation that includes auditory, tactile, and other formative content. We implement this principle by allowing users of the EE to interact with sampled and generated sound, video, and photographs (work on incorporating text is ongoing).
4. Relevance of context. An improviser's decision-making is rooted in the totality of his perception of the moment. In other words, improvisation is not formed in a vacuum; it incorporates and reflects the environment in which it is created. We implement this principle by providing an interface to which a variety of sensors (e.g. temperature, motion) can be connected, affecting the ongoing evolutionary process[2].
5. Process, not artifact production, as the goal. An improviser, unlike a feature-film cinematographer, a Western composer or a product designer, is not concerned with a final production – a movie, a sonata, a pop song, or a chair. While improvisation might be recorded and, as such, seen as a fixed construct, improvisation is the process of exploration, contextualizing and interrelating memories and perceptions. The improviser weaves together an array of 'sketches,' which gain their relevance and meaning only as the improvisation unfolds. We implement this principle by ensuring that each interaction with the EE is unique. The traditional notion of recording is absent in the EE. Instead, states of the system are saved and, if restored, will evolve to a different configuration given the currently active evolutionary processes.
6. Absence of a static plan. The process of improvisational creation cannot be (should not be) thought of in terms of planning – or as a result of planning. Instead, the act of improvisation might be more aptly described as one of exploration and continuous evolution of multiple planning possibilities. An improviser is expected to be far less concerned with playing perfectly to a specification than with breaking new ground and learning from unintended mistakes and unexpected successes. This principle is implemented in the EE by the mere nature of the evolutionary process, which is adaptive by definition.
7. Distributed responsibility and control. In an improvisational performance, no fixed contract specifying responsibilities of control (i.e., a balance of power) exists between the performers. The criteria that define the degree to which each party

---

[2] This aspect of the system has recently been explored at an installation of the EE at the Centre Pompidou (Paris, 2003-4) as part of the Non-Standard Architectures exhibition.

assumes creative control over different aspects of the ongoing improvisation are set dynamically, according to implicit and explicit negotiations between performers. Giving up part of the control frees the improviser from preoccupation with creating a perfect finalized product resulting in lower costs of experimentation. We implement this principle in two ways: First, the evolutionary process plays a role of a co-improviser, sharing control with the participant. Second, the networked nature of the EE allows multiple participants to share the control of the ongoing performance with the evolution as the mediator.

8. Audience as a participant. From the passive audience of the linear storytelling to the nearly equally passive audience of the multiple-choice "interactive" environments, a strict giver / taker dichotomy has been enforced between the consumer (the audience) and the producer (the performer). In the context of improvisation, however, such a distinction is obsolete; anyone can co-improvise, so long as the effect of his activity is seen or heard in some way by the other performers. Similarly, even when not actively participating in the act of media creation, the audience is not to be regarded as passive but as a part of the improvisational circle. The implementation of this principle is evident in the participant interaction within the EE; by merely browsing (observing), the participant affects the ongoing evolution, thus, blurring the distinction between an audience and a creator.

4 The Emonic Environment

Our research interest lies in exploring situations of co-improvisation: a scenario in which the participants and the system are symbiotically contributing to a shared performance. The system described in this section – the Emonic Environment (EE) – is our attempt at creating an improvisational framework that lets its participants control media experiences through both direct input and ongoing evolutionary processes. The EE integrates component-centric precision and process-centric structural control into a single environment for improvisational media exploration in time and space.

The EE consists of 3 interconnected networks: Perceptual, Structural**,** and Mediated. Perceptual and Structural networks exist as independent entities. The Perceptual network is a media-processing engine that determines how media is generated, modified, and played back (e.g., speed, volume, pattern). It is populated with *emons*[3] of various types, each having an array of properties that can be connected in numerous ways. On a higher level, the Structural network, revolving mainly around the concept of stimulation changes over time, provides a structure of control through a neural net populated with *nodes*. The Structural network is concerned solely with change of its elements' activities. It is a purely abstract system for controlling objects, without which its control is utterly meaningless. Connections between the elements in the two networks are utilized to establish chains of control throughout the EE.

---

[3] Emon: functional media-processing primitive; combined together, emons form an interconnected structure for generation, modification and presentation of media.

In order to relate the activity within the Structural network and the media processing of the Perceptual network, an auxiliary network, the Mediator, is provided. Together, the three networks form an evolving environment for improvisational interaction. Every action of a participant is based upon the following evolutionary capabilities of the environment: individual properties of the objects can be adjusted, the system can be set to evolve into another system, and the participant can choose to contribute feedback reflecting on the ongoing evolutionary process. The three networks all operate in real time and in a balance determined by a participant, making the environment 'come alive.' Let's look at the individual components of the system[4], and then proceed to a larger-frame picture of the networks and their functionality.

Master Tempo Emon In any system comprising multiple, separable parts, a uniform measurement of time must exist, acting as the beating heart by which the entire environment can live. The conductor of an orchestra, the sun, the heart, for example, all provide a basis from which the rest of the environment can determine its course of action. The Master Tempo acts as the emon that provides the pulse to be used by other elements of the Perceptual network as a synchronization reference. On each of its beats, it fires signals to all of the directly connected emons, thus, prompting a reaction within the rest of the system.

The Master Tempo can be adjusted dynamically. A slower tempo may result in a network in which numerous actions occur within each beat, while a faster tempo may contribute to a fast-paced but less intricate network. Changing the Master Tempo can have an effect on the entire attitude of an environment.

Tempo Emon Tempo emons, although dependent on a supplied beat, have the power to break up their parent tempo into smaller, more distinct bits of time. Through nesting Tempo emons, the participant becomes able to create complex polyrhythmic temporal structures without having to understand concepts of music theory and composition dealing with time. As a result, the Perceptual network becomes capable of processing events in an asynchronous manner.

Additionally, Tempo emons can introduce a time delay, propagating a beat slightly offset from the original tempo. The offset property of a Tempo emon controls the amount of time that separates notification of an incoming beat and its further propagation. Such time variance can be utilized to create an echo or stagger effect.

Filter Emon No system is complete without the ability to ignore the directives provided by others. Such 'resistance' of an element within a system can be formalized using the concept of a mask or filter. The Filter emon implements a mutable pattern of resistance, propagating events at its own discretion. This action can be thought of as denying stimulation to the emons connected to it in accordance with the given mask.

A filter consists of a binary array signaling which events the filter should and should not propagate. The filter operates in a circular fashion, looping through the array in synch with the input events it is filtering. Filters can be used to further

---

[4] Additionally, the more advanced users can utilize the provided architecture to expand the amount of available control by building new processing components.

enhance the media performance by allowing for variation in frequency of the output. For example, instead of a simple, repetitive chain of events playing an audio sample every beat, the filter allows for a structured but not monotonous repetition. Filters can be modified by changing either the entire pattern or individual bits of the mask.

Audio Playback Emon Each Audio Playback emon controls how a given audio element, be it participant-recorded or retrieved from a database, can be played back. The playback is activated by receiving a beat from a connected time source (e.g., Tempo or Filter). Audio Playback emons constitute the audible body of the environment, each having adjustable properties of volume, start and stop cue points within the sample, playback direction, time and pitch shift, audio file location, and more, all of which can be modified in real-time.

Each Audio Playback emon can store a sequence of volumes to process at each playback occasion. Subsequently, fades and crossfades can be performed, allowing different samples to become prominent at different times.

Start and stop cues serve to define additional boundaries within the audio sample. Consider an audio sample with a middle section that fits ‘well’ in a given network, but, the participant does not want to clutter the overall output with the rest of the sample. Instead of permanently changing the sample, start and stop cue points can be added to indicate which portions of the sample should be played. As the network evolves, the participant might decide that the rest of the audio sample fits the new network state better, thus, requiring a simple change of cues. Similarly, as a network evolves on its own, the portion of an audio sample being played can be changed to create a different sound without modifying the source file.

Furthermore, the participant may desire to play multiple sections of the same audio file. Two implementations of this functionality have been designed utilizing a system of multiple start and stop cues: (1) a mask that preserves the length of the sample in its playback, introducing silence in the masked-out portions of the sample, or (2) a selective mask that plays only the desired portions, skipping over the masked-out spots.

Additional variance in an audio sample’s character can be achieved by changing the playback direction or time and pitch shift. Each of these properties adds to the repertoire of the environment.

Master Output Emon One of the key improvisational principles is the importance of context (surrounding events) in the course of improvisation. The Master Output emon is one way to express this concept by controlling the spatial placement of sound. While it may be intriguing to have a network of intricately connected Tempo, Filter, and Audio Playback emons that result in a one-dimensional audio output, spatializing the participants’ audio space can bring the EE closer to the contextual world of the participant. A person standing in one part of the room may hear a certain output and choose to modify the tempo he hears, while a participant on the other side of the room may hear a combination of other outputs and change what audio samples are playing.

How is this separation of outputs achieved? The Master Output emon allows us to connect Audio Playback emons to specific audio outputs. By dynamically changing the connections between Audio Playback and Master Output emons, we control the spatial presence of audio in particular areas. Mutation of this balance results in shifting or rotating of sound around a room for a fully encompassing atmosphere.

The type of output the Master Output provides is defined automatically dependent on the type of the incoming emon; in this way, the Master Output is capable of controlling color of lighting in a room, streaming video over the net, delivering commands to cell phones, and more.

Perceptual Network All the emons interact through the Perceptual network, a network that has connections that define the character of *what* and *how* we hear (unlike the Structural network, described below, which defines *when* and *why* we hear). By combining multiple Tempo, Filter, and Audio Playback emons, a Perceptual network can be built to taste, both visually and aurally. The relations in the network are the foundation of how the system works: originating from a single, repetitive beat; propagating through tempo adjusters to become faster or off-beat; propagating again through filters which alter the overall pattern of emons' processing; and finally ending with a disjoint collage of signals being provided to audio samples and output into our physical environment.

Structural Network On a different level, unaware of emon activity, nodes of the Structural network are entwined to create a net similar to that of neurons. The nodes each possess an independent activation level, its value continuously decaying. They communicate by sending stimuli, which can originate at any point in the network. When a node is triggered, implying that its activation level is beyond a propagation threshold, it sends stimuli of proportional strength to all connected nodes. The growth and decay of nodes' activation levels are always due to outside stimulation.

Mediator How can we invoke change in the Perceptual network activity? By using the Mediator, we couple the control of the Structural network with the Perceptual network's processing. The Mediator is notified every time a Structural *node* passes a threshold and maps this information to a property of a Perceptual *emon*. The Mediator insulates each network and allows for their independence, making it possible to replace either network with a different type of controller or media system if desired. In separating these two different layers of processing, the participants are allowed to balance their focus between control levels throughout their performance.

To implement this layered construct, we designed nodal and emonic *tokens*. The nodal tokens specify two properties: the activation level at which a notification will occur and the direction of change in the activation level at the time of the token's notification (e.g., activation level rising/decreasing past 0.5). The emonic tokens correspond to specific properties of an emon and instructions for its change. Two forms of emonic tokens exist: relative and direct. A relative token can refer to toggling a value, such as a digit in a filter, or in/decreasing a value, such as a tempo. A direct token corresponds to setting a property to a set value, such as a filename or a tempo. Participants have control over which, if any, tokens are connected through the Mediator. These mediated connections can also change through mutation, leading to a complex system where transparent changes imperatively affect the overall state.

5 Evolutionary Methods for Effecting Change

From the principles stated in section III, it should be clear that the aim of the EE is not generating system actions predictable in their behavior (for no behavior is universally good or bad), but, rather making the EE capable of adaptation to unforeseen changes within the media space. GAs allow us to make the EE capable of adaptation. Furthermore, the very integration of component-centric precision and process-centric structural control exemplified in the EE is made possible by utilizing GAs. Evolution serves as our instrument for effecting change within the media space, affording flexibility in the interaction possible with the system, and potentially leading the participants to unforeseen and emergent spaces for media creation and consumption.

5.1 GAs: Implementation and its challenges

In the following sections we will address two questions: (1) what are the individuals that populate the evolutionary space, and (2) what constitutes a 'good' fitness function, given the idiomatically unconstrained media space?

The first question is the easiest to answer: the individuals are all the different system properties. For the Perceptual layer, these properties are emon-type specific (e.g., the Offset of Tempo emons). For the Structural layer, these are the properties of the nodes. For the Mediated layer, these are the token sets' mappings. There are also overall system properties as well as a possibility of creating macro-properties consisting of two or more properties from different layers.

The second question is much more involved and has no simple answer. In the EE, we think about evolution in terms of the system as a whole progressing from one state to another. Where is the evolution going at any given point? This is defined by the user, who selects a state toward which the evolution will progress. We call these attractor states *magnets*. The progress of evolution thus depends on human input and principles of evaluation rather than on survival.

In order to define a magnet, we need to define a representation of the system's state; that is, a function that defines what is relevant about a given state, so that given that representation we can decide whether two states are similar.

Each individual is mutated in order to reach a desired *magnet*. As the evolutionary process takes place, the environment learns which types of mutations are preferred for that particular evolution and combines the mutations that bring us closer to a magnet with the participants' preferences.

In order to judge which mutations are better than others and to make sure those are more likely to occur, we need to be able to evaluate the distance in parametric space of an environment to a state described by a magnet. We can do that by determining the fitness of the current state, and then proceeding to evolve randomly by a distance of one mutation at a time. This concept of trial and error makes the environment learn which mutations are more desirable. After each mutation, an evaluation will occur to determine if the current state is closer to the magnet. This process of mutation and evaluation continues until the current environment and the magnet are within a reasonable distance from each other.

The fitness of an environment is represented through a numeric representation, or a *checksum*. This checksum is reflective of the individuals present in the current state and is calculated uniformly. By weighting each individual on a different scale, we can derive a checksum that represents a given state, then used in comparing two states to each other. In addition to the basic weights of each property, a participant can weigh which mutations he personally prefers to occur more often. These combined weights shape the overall evolution so that it continuously tries to balance the participant's desires and the direction suggested by a magnet.

Mutations available within the EE are component-specific. For example for nodes there are Add-Node, Change-DecayRate and more; for Tempo emon, there is SetBeat, and SetPhase. General system properties, such as the number of mutations per minute are themselves subjects to the evolution. Currently there are about 60 different types of mutations available within the system.

In order to check the distance between two states we need to have a method for checksum comparison. There are two factors that influence a mutation's worth: (1) the overall direction of change, and (2) the rate of change. A combination of the two factors can be used to reflect the desirability of any mutation.

The direction of change is an obvious factor: if the difference between the current and the magnet checksums has decreased, the mutation is favorable. The rate of change is less obvious: consider filling an empty glass with water. If we were to begin by first adding single teaspoons at a time, it would take an absurd amount of teaspoons to fill the glass. However, if we were to start by adding a cup of water at a time until the glass is almost full, and then continue by adding a teaspoon at a time so the glass does not overflow, the method would be more efficient. In the case of the EE, performing a mutation that only moves the state slightly closer to the magnet is not considered as valuable as a mutation that moves the state closer in more appropriately-sized steps. Because of the scaling factors, each property change affects the checksum differently, thus resulting in mutation probabilities reflective of the desirability of the change. Each time a mutation occurs due to its having the best-predicted probability, its fitness is reevaluated to reflect the new resulting state. In this way, the EE learns about its progress and attempts different mutations in order to step closer to the magnet's checksum.

5.2 Multiple Magnets

No real life example exists where one influence single-handedly defines the evolution of an object or environment. In our context, we want to have the possibility of creating a divergent sequence of states. This becomes possible if we change the current state of the environment in multiple directions at once. While placing a single magnet in the EE influences, over a period of time, mutation of the current state, bringing its characteristics closer to those of the magnet, placing multiple magnets should have a similar effect, but with an inherent competition between themselves.

To do so, we employ an evolutionary process that takes into account that a given mutation will bring the environment closer to one magnet, while farther away from another. Active magnets are each assigned equivalent initial weights. After each mutation, the evolutionary process evaluates the distance between the mutation performed and the state defined by the magnet with the current most probable weight. Magnets are chosen in a roulette fashion with the probabilities of picking each magnet defined by its weight within the EE. As the evolution progresses, the participants dynamically alter the weights assigned to the magnets, thus, shaping the path taken by the evolutionary process. In doing so, we can account for multiple influences on the current state. The concept of multiple magnets enables us to determine if a given mutation contributes to obtaining a better balance between the currently active magnets. This balancing act reflects the nature of improvisation – a process of continuous flux.

6 Conclusion

This paper is a progress report. We have focused on the implementation, showing how principles of non-idiomatic improvisation help us in shaping a real-time evolutionary system for improvisation. We have shown how evolving a media space can be possible through multi-layered control architecture and magnet-based strategies for evolution.

We hope to continue exploring improvisational landscapes, bringing about new strategies for evolving media spaces and applying these to facilitate creation, modification, search, exchange, and performance in the domains of digital media.

Acknowledgements

The authors thank Glorianna Davenport, Ariadna Quattoni and Tammy Gilson for their ideas and help.

References

1. Nemirovsky, P. and Watson, R. (2003) "Genetic Improvisation Model :: a framework for real-time performance environments". In Proceedings of *EuroGP2003.* Essex, UK.
2. Lewis, G., *Interacting with latter-day musical automata.* Contemporary Music Review, 1999. **18**(3): p. 99-112.
3. Sims, K., *Galapagos*. 1993. Information online at: http://www.genarts.com/galapagos/
4. Miranda, E. R., *On the Origins and Evolution of Music in Virtual Worlds*, in Bentley, P. and D. Corne, *Creative evolutionary systems*. 2002, San Francisco, CA. Academic Press.
5. Blackwell T.M. and Bentley P.J. (2002) "Improvised Music with Swarms", 2002 Congress on Evolutionary Computation, CEC-2002.