Improvisational Interaction
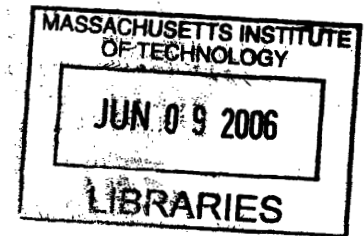## :: a Framework for Structural Exploration of Media

by

Paul Nemirovsky
M.S. Massachusetts Institute of Technology, 1999

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning,

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Media Arts and Sciences

at the Massachusetts Institute of Technology
June 2006

© Massachusetts Institute of Technology, 2006
All Rights Reserved

Signature of Author
Program in Media Arts and Sciences
February 7, 2006

Certified by
Glorianna Davenport
Principal Research Associate, Media Fabrics Group
MIT Media Arts and Sciences
Thesis Supervisor

Accepted by
Andrew B. Lippman
Chair, Departmental Committee on Graduate Studies
Program in Media Arts and Sciences

the man who cannot imagine a horse galloping on a tomato is an idiot.
-andre breton

Improvisational Interaction
:: a Framework for Structural Exploration of Media

by Paul Nemirovsky

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology,

on February 7, 2006, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Media Arts and Sciences

Abstract

Whenever we use computers to interact with media, our experience is that of *direct control*, and the goal of our interactions is either artifact-production (the editor paradigm) or passive exploration (the browser paradigm). This thesis proposes an alternative: a model of media interaction based on the ideas of non-idiomatic improvisation that encourages *active exploration* of media and its structures. We argue that in order to facilitate this kind of exploration, (1) computational tools must actively participate in the creative process and (2) the interaction framework must allow structural exploration of media. This leads to our main claim: improvisation should be considered a valid and appropriate paradigm for media interaction.

To this extent, we present a *Cognitive Model of Improvisational Action (CMIA)* that integrates element-centric and process-centric (structural) modes of control into a single framework for media exploration. This model allows participants to switch their attention between compositional elements and structural processes. The model is argued to be particularly powerful in leading us to novel spaces for media creation and consumption.

We follow by presenting the Emonic Environment (*Implementation*), an interactive system built on the principles of CMIA. We conclude by describing two studies (*Scenarios & Experiments*) that analyze the ways in which Emonic Environment affects how people interact and think about their interactions with digital media. These studies illustrate the potential of CMIA as a paradigm for interaction between humans and machines.

Thesis Supervisor: Glorianna Davenport
Title: Principal Research Associate, Media Fabrics, Program in Media Arts and Science, MIT

Thesis Committee

Thesis Supervisor:_____

Glorianna Davenport
Principal Research Associate
Director, Media Fabrics Group
MIT Media Lab

Thesis Reader:_____

Whitman Richards
Professor
MIT Computer Science and Artificial Intelligence Laboratory

Thesis Reader:_____

François Pachet
Senior Research Scientist
Sony Computer Science Laboratory, Paris

To my grandmother, Dvora Nemirovsky

Acknowledgements

Personal Experience :: Dimming the Lights

Time to perform. The piano stands still. People are whispering, shifting for a better view of the stage. We are three – the people, the piano, me. The piano – long, large, and black; me – small next to it, only expected to touch it in certain ways, through the keyboard, or maybe a sensuous touching of a few strings or the wood; the people, with the remarks they make, the clothes they wear. All of us are expecting an emergence of the unforeseen.

I look at the marks on the keyboard and little white scratches left by those who have played before. Imperfections that make this piano unique. Scratches reminding me of stick figures I made as a child when playing with bread, letting my fingers fight and merge into each other. Scratches I make with my fingernails across my belly, that take minutes to fade.

Fade away. Bummm-bommmmm.Tinn-tann-poo-too-Tim. Speaking of Tim, do I know any Tim, hell knows, T-I-M. Running through the Australian fields of grass. I wonder. Ships? Zhhhhhuuuunnnn-duuuu-leeee. Yeah Tim, Bummm-bimmm-tam-tim, tam tim. Zhuuuuuuu.

I play with fields of energy now. Unfinished thoughts, half-baked comments on thoughts relevant a second ago but no longer significant, for the energy is mutating, and, to keep it, I must follow. I refuse the comfort of my motor skills, for they work only to stifle the energy with predictability and boredom. Sometimes, I relent, only to ambush later and inconspicuously work my way out of the obvious.

I look at the people – to capture an expression, a piece of clothing, a reflection of light. These quick glances provide hints to where I can go next to stay in tune with the energy among the audience, the piano and me. The piano, as an object, is not important; it's just that the interface is flexible enough to allow many different sounds. I could equally improvise with any other object if I knew how to cause it to produce a variety of sounds. Not sounds I would be able to predict – this is not needed – but sounds I would be able to counteract with yet other sounds, dynamically juxtaposing a large variety of different colors, bringing a never-ending alternative to what was there just a moment ago. In this improvisational landscape we travel together – the audience, the piano, and I. Neither of us is in the "driver seat"; we inspire and feedback to each other.

It is never one thought, one melody, one sound that I am attending to but multiplicities of thoughts, melodies, sounds – unfinished parts with no beginning or end, spreading like an endless kaleidoscope with parts never in a describable order but always in a conflict of coexisting parallels. These parallels are perspectives, each with its own character and intention. Through following their conflict and taking sides, I am able to improvise, but unable to ever go back, because there is no unified past to return to – each of the perspectives affects and is affected as time goes on; not a linear time but one shaped by the conflict and disagreement yet always capable of finding new ways to disagree and spawn off new parallels with which to be in conflict.

It is easy to see why improvisation has been a part of our cultural landscape as long as we can remember. Think of improvising as an activity:

* Doing it does not require an immutable system of rules.

* The result is always different yet the development is not chaotic.

* You can see it across cultural, socio-economic, and religious groups.

* It allows expression of vague thoughts and feelings.

* It's OK if you make a mistake. Errors are part of the process.

* It is heterarchical (interconnected) rather than hierarchical.

* It is more socially engaging than premeditated forms of entertainment; audience can take part in shaping the improvisation.

To put it simply – in its uninhibited form, improvisation is the way to create without being limited by rules and regulations of one given system.

Road Map and Aims

This work puts forward the following argument: frameworks for human-machine interaction tend to separate between modes of production and consumption, preventing their users from simultaneously engaging in both. Furthermore, existing frameworks mainly focus on arranging individual media components rather than on creating and observing structures that bring these components together. Finally, the machine's potential as a participant in the creative process is largely unexplored. We propose to address these deficiencies by introducing *improvisational interaction*: a method for structural browsing and manipulation of digital media in real-time, in which the machine plays the role of a co-improviser.

The focus of our effort is on creative entertainment. While improvisational interaction holds promise for a variety of domains (e.g., information retrieval and learning), we believe entertainment to be an appropriate starting point due to lesser opposition to new ideas in the field of entertainment than other fields such as learning. In this work, improvisational interaction is shown as a vehicle for exploring *digital media spaces* (DMS) – networks of structural elements controlling the behavior of media content over time.

This work aims to contribute to the advancement of the field of human computer interaction by: (1) defining a cognitive model of improvisational action in the context of interacting with digital media; (2) developing a framework for exploring DMS in an improvisational fashion, and (3) showing the applicability of genetic algorithms to construction and navigation of DMS.

We start by defining the problem space: observing that computer-based creative action today is limited not by lack of computational power but rather by frameworks and paradigms inherited from traditional notions of *machines* and *users*, implicitly manifested in the roles humans and computers are expected to fulfill in interacting with each other.

We go on to argue for the importance of introducing improvisation as an interactive method in the context of digital media, and the appropriateness of genetic algorithms as a method for DMS exploration.

We follow by introducing an interactive paradigm inspired by non-idiomatic improvisation and formalized as *CMIA* – a Cognitive Model of Improvisational Action, and describe the Emonic Environment, a framework for audiovisual manipulation based on the CMIA.

Finally, we describe two studies exploring how a wide variety of people used the Emonic Environment and present quantitative evidence to support the validity of the CMIA.

We conclude by describing existing usages and potential applications for the Emonic Environment system in particular and CMIA in general.

Problem & Motivation

Digital interactive environments of today primarily rely on and assume a final purpose in every interaction between a human and a computer. These environments are being designed with the objective that "users" fulfill such purpose with minimum time and effort. In other words, what matters is that you arrive at the goal in the shortest possible amount of time; the path by which you get there is largely irrelevant. This type of interaction (manifested in the entire spectrum of our encounters with digital media) denies its participants the full power of serendipitous exploration and discovery in the name of a higher efficiency in achieving some final result.

As we show below, in many activities such efficiency is ineffective and limiting, and furthermore, the final product may be irrelevant. For a lack of better denominator, that class of activities can be described as process-centric – situations where the process of exploration is more important than a particular final result. Many of our interactions with digital media, creative browsing in particular, fall within that class.

Interactive paradigms manifested in most available tools for machine-based media creation are in conflict with the immediacy of human creative desire to express: the tools are either (1) unable to afford such expression, let alone expand on it, without their users exhaustingly specifying exactly what it is that they want the tool to do (a demand unrealistic in a real-time creative setting), or (2) limit the expression to what can be called 'toy interaction' – narrowing the choices by channeling users' input into predefined templates, thus preventing anything the tool's author would consider to be a mistake or not within the confines of a particular style or form.

Consider a simple creative scenario: let's say that you would like to create a 'beat' – a non-monotonous rhythm consisting of a few audio samples. Currently, the only option available for this task is a deterministic tool. We consider a tool to be deterministic when for each input there is one defined output, and when the tool design aims at minimizing the amount of exploration necessary to discover the correspondences of inputs and output. An example of such a tool is an editor. Another option is a 'wizard tool' (a concept popularized lately by many musical toys) – a type of interface where the user can control a very limited array of properties of the media space being created, and have his input 'fixed up' by the tool itself to 'sound better'. Wizard tools might not be immediately deterministic in their users' eyes, but become such as soon as they discover the 'trick' of how the tool operates.

Back to the beat. If you were to opt to create your beat in a conventional editor, you would need to know what constitutes a measure of rhythm in the editor, how to define a polyrhythmic structure, how to synchronize events and arrange them on a timeline or in a codependency chart; these are just a few of the things one must know before actually doing anything. Even for professional musicians, who already possess such knowledge, it is hard if not outright impossible to effortlessly transform a creative thought into an actual implementation, without having to decide at each step precisely how their original thoughts fit the particularities of the tool in question. A user of such tool is forced to act in a planned manner – deciding on a course of action, learning how to do the actions planned, and then fighting the tool till the final outcome is achieved.

Alternatively, opting for a wizard tool requires very little. In the case of making a beat, you would have to (1) choose some appropriate sounds from a preset category, and (2) choose a function according to which the rhythm will be achieved. In fact, even these two may not be available, with the computer picking the sounds, deciding on the function, and leaving you, the artist, completely at the mercy of the tool creator's algorithms. Since the inner workings of these algorithms are hidden (and are frequently hard to understand), the artist remains with no ability to affect the internals of their functionality during the performance. Furthermore, most of such tools aim at 'fixing up' what their creators consider as mistakes, promulgating the idea that no matter what you do, the computer will 'improve' (or one could say mediocre-ize) your results. The artists are droids, merely there to provide the algorithm with some input upon which the wizard-tool can operate.

Neither of the two interfaces is flexible enough to enable a creative process. The editor leaves you with all the control; the wizard takes all the control away. Neither tool asks you – how much control do *you* want?

All you might have wanted to do was to make *some* rhythm playing in a way that is *somehow* interesting, yet be able, whenever you so desire, to tweak and break the rhythm and the algorithms that make it up – but without having to think of and/or be responsible for a final 'product'. This ability to be vague about your desires at one point yet very specific the next is crucial for our creative drive – yet virtually impossible to achieve in today's interactive frameworks.

Even professional improvisers, if asked to do so in the middle of the improvisation, would be hard-pressed to specify the next steps they are about to take. Editing and browsing tools for the manipulation of digital media however require just that: knowing precisely what it is we want to do, how we want to do it, and when. Each of our actions is usually followed by some reaction from the machine (nearly always directly and linearly correspondent to the input provided), leaving us with the obligation to decide on the next step. To put it simply, in interactive frameworks of today people (as

manifested even by their accepted moniker, 'users') are slaves to the machine, serving a computational paradigm that has nothing in common with the casual, intimate, vague way in which human creative process happens outside the computer domain.

One may ask: how would this be different were improvisational interaction an accepted framework for manipulating digital media? Let's take the example of browsing and arranging one's media collection: an editor, such as Adobe Photoshop or Premiere, enables its users to edit each individual picture (or video) in a variety of ways; a browser, such as Picasa, allows you to easily sift through a large collection of images. Photoshop provides no structure at all: each picture is an independent artifact, with no connection to any other. Premiere provides a timeline but no way for elements to arrange themselves on the timeline in different manner depending on the context (has to be specified by the human). Picasa provides some sorting structure (organizing pictures into albums automatically based on the source image's date) but the structure, once created, is fixed – the machine makes no attempt to change it or suggest new structural configurations of the media. In other words, the machine makes no effort to interest or intrigue the human. This is not a question of a lack of features; even if Picasa were to allow advanced editing, or Photoshop would add automatic album-categorization, the machine would retain its passive role, remaining unfit for people who desire serendipitous exploration and play.

Tools such as Flickr, a web-based tool for media sharing and organization, represent a step in the right direction. In Flickr people annotate parts of images provided by other users with their own descriptions. By contributing the descriptions, they include the picture in their own narratives, modifying their own role from browsing to a limited authorship. At the same time the tags' structure is static, each tag tightly bound to a given image. The structure has no behavior or any temporal characteristics of its own. In other words, Flickr albums carry no structural information that would allow them to be experienced and manipulated according to a beat or a pattern – it is impossible to request actions such as "apply same set of tagging principles to two different picture sets" or "substitute another chair in this kitchen picture every three seconds". Furthermore, there is no temporal view of the changes in structure; put simply, in the tag framework, content and structure are not completely separated, preventing us from thinking about the two independently.

In addition to enabling structural treatment of time within the piece (ability to create beats and patterns of actions and observe their behavior over time), and enabling seamless integration of editing and browsing, a framework based on improvisational interaction would elevate the machine to the role of a co-improviser, capable of independent decisions yet also capable of receiving feedback from the human. The human in turn would no longer be a *user*, rising to become a member *participating audience*, a fluid collective of people interested in exploring and manipulating structure that governs the

media broadcast for everybody's consumption. In other words, in an interactive framework based on improvisational interaction, people would finally get a chance to manipulate digital media by exploring structures and processes governing that media rather than focusing on the content-based control of the individual artifacts that comprise the structure being explored.

Incorporating improvisation into human-computer interaction means that both our methods for interacting with the machine and our expectations in regards to the results of such interaction are to undergo a change. To mention just a few, we can expect:

- Flexible participation length due to process-centric nature of interaction (start or stop participating whenever you feel like)

- Lower price of content-centric mistakes (don't think about whether what you do is right or wrong for this image or sound but rather whether the overall structural development is going in the way you want)

- Absence of fixed participatory roles (pick what kind of things you want to focus on; don't worry about the rest, it will either be done by others or not done which is also all right)

- No ultimate goal (you are not composing or performing a song; you are engaging in a process, the goal of which is the fun of discovery along the way and learning from it – finding the interesting combinations of sounds, images, and texts that result from actions by you, other participants, and the machine.

Traditionally, from the passive audience of linear storytelling systems to the nearly equally passive audience of multiple-choice interactive environments, a strict giver / taker dichotomy has been enforced between the producer (the performer) and the consumer (the audience). In improvisation 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. Any audience becomes a pool of potential participants which, even when not actively participating in the act of media creation, are regarded as a part of the improvisational circle.

Bringing improvisational characteristics into our interactions with digital media is likely to lower the entry barrier for the computer users who wish they could act as creators yet are afraid of 'getting it wrong'. The only option available today for those who desire creativity in the digital domain yet aren't willing to invest significant time in mastering the intricacies of editor tools are the wizard toys (discussed in more detail elsewhere in this thesis).

Wizard tools, usually limited to one medium (primarily sound) allow easy creation and manipulation of a given media type. The ease of use stems from the templates built into the tool that implicitly define the kinds of manipulation that are permissible and not. In our view wizard tools lead us down the wrong path. On one hand, wizard tools do not encourage true exploration of the media space – the users are at a whim of the machine, which quantizes and polishes their input. This polishing ends up promoting mediocrity, while giving people a false sense of creative input. On the other hand, once the 'trick' is learned, it is hard to remain satisfied with the child-like quality of the majority of the wizard-tools. As a result a negative result is achieved – the user is likely to find out that being a creator is not all it is hyped to be – after all, everything he or she was able to do sounded and looked about the same.

In true improvisational framework, central oversight (e.g., quantizing all audio inputs) is unnecessary, due to the self-regulating nature of improvisational collaboration. To put it simply – bringing improvisational interaction to the hands people who in the past have thought of themselves as a passive audience empowers them to become full-fledged contributors to a creative process without incurring the associated 'costs' of authorship.

To sum it up, both editors and wizard tools are deterministic; the interaction is typically linear (attending to one element at a time) and the cost of mistakes is high (requiring interruption of the creative process by undoing what is done in the case of an editor or hiding the errors in the case of a wizard). As a result, the machine misses on a number of potential roles: a learning (rather than pre-programmed) accomplice, an exploratory vehicle for unknown (e.g., dynamically generated) media terrains, and an idea generator.

What needs to happen for the machine to assume these roles? First, the machine needs to presume that there will be no human evaluation of its individual actions. Second, it needs to provide the human with a structural view of the media space being explored. Finally, it needs to be able to search the media space for possibly interesting solutions with only vague criteria in hand.

How did we come to interact with machines the way we do today? The roots of the problem are twofold: on one hand all our interactions with the machine are informed by the assumptions of the 'appropriate' roles of computational devices and us as their users – assumptions stemming from the original role of computers as computing machines. Equally significantly, the Western world, where interaction frameworks in question were largely developed, has long had a predisposition for compositional and planned as opposed to instantaneous and imprecise.

How do we reform the currently prevalent ideas about human interaction and reinvent browsing, advancing past the limitations present in current interactive systems? Our attempt involves (1) formalizing what constitutes improvisational behavior in the context of digital media and (2) creating a framework that facilitates such behavior. Before doing that, let's define a number of concepts used throughout this thesis and consider theories and systems that informed this work.

Definitions

Digital Media Space (DMS) A network space, populated with audio / video / text content as well as structural elements controlling the behavior of the content. The space is explored utilizing the Emonic Environment, the project described in *System Architecture & Implementation*.

Improvising / Improvisation (1) Ability to generate, manipulate, and re-contextualize media content in real-time, dynamically synthesizing the previously processed and the new stimuli (events) into an aesthetically meaningful action of media production and/or control; (2) ability to relate to concurrent and interconnected timelines (rather than a singular sequence-time) while focusing on the process of creation rather than an outcome (a produced artifact); (3) ability to process data utilizing multiple representations (abstraction levels) at once.

Compositional paradigm A creative paradigm that focuses on producing and/or exploring final artifacts (sonatas, films, chairs, programs) by placing fixed elements (notes, bytes) in a static (or based on a rigid/centralized system of rules) structure according to a predetermined plan and/or objective. This paradigm is manifested in human-computer interaction as the editor..

Editor Framework The editor framework forces us to approach any creative exploration in terms of compositional elements: notes of a sonata, shots of a movie, articles of a newspaper, files of an MP3 collection. The computer is viewed as passive, capable of responding to a fixed set of commands by performing a strictly correspondent and fixed set of functions. The human is expected to know precisely how, when, and in what way to activate the functions to produce the results. The drawback of the editor framework is clear: if we reflect on our experience in creating music, stories, or films, we will find it hard, if not impossible, to describe it in terms of compositional elements; a higher-level, structural view will be required. In other words, editor framework is helpful in realizing a creative action, but it is not a part of the creative process itself – it doesn't help you to come up with ideas.

Browser Framework The browser framework regards the computer as a black box, guided by a set of algorithms hidden from human view. The human 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. In other words, browser is part of the creative process (it may proactively suggest new paths through media space), yet misses the realization aspect of the creative process offered by the editor. Furthermore, process-centric approach is not core to the framework. Frequently, a browser may be

just as component-centric as an editor, with the difference that the components being manipulated will be more complex (e.g., melodies vs. individual notes). Structural exploration may however still be lacking.

Improvisational interactive paradigm A creative paradigm that brings improvisational features into human-computer interaction. No predefined rules, plans, or objectives are built-in; instead, the space being explored is viewed as an evolving structure, its configuration guided by genetic algorithms and human feedback to continuously restructure itself. Its characteristics are detailed in *Cognitive Model of Improvisational Action*.

Inspirations

My favorite music is the music I haven't yet heard. –John Cage

While performing or browsing a photo album, our ideas with respect to the results of the interaction often undergo a significant change. This change in the course of the interaction process is natural – we rarely have a clear idea of looking at particular pictures in a fixed order – rather, as we explore the elements that comprise our experience we change what and how we would like to see (or hear) next. Our desired role changes as well: at one point we might want to feel like an author creating a unique piece of art; at other, like an explorer discovering and observing actions and artifacts made by others. Sometimes you may want to be in control, at other times relinquishing the control in favor of serendipity or curiosity about the actions of other participants and of the machine itself. These changing creative thoughts can be described as a desire for improvisational dialogue, where roles and results are vague and bound to change, whether our initial intent was searching for a new danceable song or creating a visual collage. We may not know all the combinations of audiovisual content that we are going to like, yet we enjoy the serendipity of looking. Such improvisational search has the potential to become not only a tool for achieving a final result, but an objective on its own.

Theory and practice are often inseparable in the case of improvisation, where some of the leading theorists are also its leading practitioners (e.g., Anthony Braxton). Frequently, theory articulated within one domain is applied to other ones, and it is not unusual to find practitioners who transcend artificial boundaries of particular media and schools of thought in their work. Distinctions of style and material are illusory – an influential visual artist may come from a music background while a media artist or a computer scientist may release music albums to a critical acclaim.

Currently, interactive tools for creative expression don't allow role- and purpose-switching – facilitating improvisational behavior is not one of their goals. Why is it so? One of the reasons might be that designers of interactive environments are unclear in regards to the benefits of improvisational action; another reason might be that they are simply unaware of such an alternative. In the discussion that follows, we attempt to address this by presenting a number of theories and practices that highlight the reasons for adding improvisational interaction to the arsenal of interactive paradigms.

Philosophy and Cognitive Science

In improvisational setting, any media artifact can be taken out of its context and re-appropriated given the improviser's desires. This re-contextualizing of an artifact depending on the perceiving agents and their environment is a concept heavily inspired by Derrida's notion of reading (Derrida, 1976). According to Derrida, reading is an active act in which meaning emerges from a tandem of writer's intention and reader's perception of the text. To him, the importance of context (or in other words, the networked structure of our perception of an artifact) is integral to understanding what we see, hear, and read. Derrida is most famous for having introduced the concept of 'deconstruction'. The main idea behind deconstruction is that a text can be opened to multiple meanings and interpretations. Meaning itself is not a fixed construct but rather a fluid divergent process. Deconstructing a 'text' involves showing inherent oppositional or dialectic contradictions in the text's internal logic. Following Derrida in his search for a post-structuralist theory of meaning, Deleuze and Guattari introduced (in The Thousand Plateaus) the concept of the rhizome (Deleuze & Guattari, 1987), a construct that allows for multiple, non-hierarchical entry and exit points in data representation and interpretation. Such a construct is reminiscent of our idea of improvisation, which is defined as a process with no clear structural form.

Among the few **cognitive theories of improvisation** that have been proposed, Pressing's (in Sloboda (1988)) is particularly interesting. For him, improvisation is association-based and consists of sequential choice of elements which either *continue* or *interrupt* some aspect of the immediately preceding context. This step-by-step (and sometimes literally note-by-note) explanation of improvisation is in sharp contrast to the work of Sagi and Vitanyi (in the same volume) which attempts to escape the cognitivist approach and focus instead on global features of structure and style. Another, more phenomenological approach that is indirectly applicable to the question of improvisational action is the description of the problem of cognition proposed by Varela (Varela, 1991).

Learning

The other day I was browsing a list of recently released music-making software and came upon the following add-on to the well-known Finale music program: the "Finale Performance Assessment" (FinaleMusic, 2006). One sentence in the description particularly stood out: "FPA will even grade the performance — indicating how many correct notes out of all possible notes and the percentage of correct notes played." Short of hitting the students'

hands, I would be hard-pressed to find a worse idea for mastering a creative process (playing music in this case). Instead of a creative exploration, learning becomes a 'shoot-em-up' game – with notes as enemies. No structural comprehension, no questioning why a child (toward whom this program is allegedly oriented) picked a wrong note – simply count the number of good shots vs. the bad ones and arrive at the 'best' final result: zero "wrong" notes.

Programs like this aren't conceived in a vacuum – they are informed by our culture of learning, one that presumes that we, as customers, need results not explorations, that minimum of mistakes are made and everything happens precisely according to some overall plan.

Improvisation presents a learning alternative that is unconcerned with imitation, unlike the example above; it progresses despite (and often due to) the lack of any clear command-and-control hierarchy. Such an idea of learning has been articulated by Seymour Papert, in his constructivist approach (Papert, 1980). Papert's main concept is that people learn by making and that learning is process-centric, making his idea of learning compatible with the improvisational interaction paradigm.

Art & Music

Non-idiomatic improvisation (a term coined by Derek Bailey (Bailey, 1993)) can be described as a step in the evolution of jazz music (from traditional jazz to free jazz to non-idiomatic improvisation). As evident from the name, its underlying principle is making music that doesn't fall under the definitions of any established musical genre. While the practitioners come primarily from the music domain (Keith Rowe, the aforementioned Derek Bailey, Eddie Prévost, some of the work of Cecil Taylor, Slava Ganelin, and many others), its insistence on freedom from predetermined structure can equally be applied to other media. Among its other principles: external sources are incorporated in the creative act, the performance is never preplanned, and the role of individual improvisers is obscured as much as possible. For example, Keith Rowe, a member of a group called AMM (Prévost, 1995) is known to incorporate radio broadcasts in his performances while the performances themselves are spontaneous. Proponents of the free improvisation movement argue against compositional practices and application of those to understanding improvisation. Their philosophy plays a significant role in the conception of improvisational interaction articulated in this thesis.

A similar school of thought comes from the **experimental music movement**. Started in the 1960s and deeply influenced by the ideas of John Cage (1961), it was popularized by Steve Reich, Philip Glass, Gavin Bryars, Terry Riley, and others (Nyman, 1999). Its goal was to change our approach to music, by getting rid of the idea that for something to be considered musical it has to obey a given set of rules of sound organization (i.e., any combination of sounds can be musical if you think of them as falling within a particular structure such as 'noises of passing cars'). In other words, similarly to the improvisational objective, experimental music's aim is to think of composition as a way of thinking, or, to quote Brian Eno, 'a process of apprehending that we, as listeners, could choose to conduct.'

Improvising could be seen as an activity in which a structure is created and manipulated in real-time, according to a fixed or evolving strategy. The focus is nearly always on the process rather than content. In order to further encourage creativity, both of their own and of the other improvisers, some artists have come up with frameworks of structural limitations. Four brief examples to illustrate the point:

1. In Brian Eno's oblique strategies (Eno, 2005) a set of cards is made available to performers. On each card a phrase or a cryptic remark can be found. That phrase can subsequently be used to solve a dilemma or break a deadlock between conflicting desires. Examples of cards:

   o State the problem in words as clearly as possible.

   o Only one element of each kind.

   o What would your closest friend do?

2. John Zorn (Zorn, 2005) came up with a somewhat similar system of his own, called game pieces. Game pieces were a necessary tool for Zorn who, in the score of the Pool, notes that he is concerned "not so much with how things SOUND as with how things WORK".

3. Chadabe, one of the pioneers of electronic music, does not propose a system like that of Zorn or Eno, but makes his structural thinking clear: "When people ask me what I do as a composer, I explain that I do not compose pieces, I compose activities. A 'piece', whatever its content, is a construction with a beginning and end that exists independent of its listeners and within its own boundaries of time. An 'activity' unfolds because of the way people perform; and consequently, an activity happens in the time of living; and art comes closer to life." (Chadabe, 2001)

4. Finally, a quote from Misha Mengelberg (Warburton, 1996), a well known free jazz practitioner:

    "*Q*: When you start a composition, do you have a sound in mind, or a structure, or an idea? Or a particular musician or ensemble?

    A: I don't know what comes first. Sometimes when I have a sound, it can't be used for anything. It's part of an idea of something that is not yet music, somehow--or maybe never becomes music.

    Q: Has the piece started before you come on stage?

    A: Well, in a way, yes, but on the other hand, I never know what's going to happen at all. I come on stage blank. (Pause.) No ideas. (Pause.) As though I am seeing the piano for the first time in my life. (Long pause.) I don't know how I do that. I seem to forget and learn easily."

To sum it up: be the strategies of improvisation based on chance (Cage's chance operations), a set of cue cards (Eno's oblique strategies), or scenarios (Zorn's game pieces), a clear separation can usually be observed. On one side, there are the creators of strategies. On the other side there are the performers (the two sometimes being the same person) who now have to follow a strategy (i.e., a fixed system of rules). It therefore seems beneficial, in the context of interaction with digital media, to merge structure building and exploration into one activity accessible to every participant of a given improvisational activity.

These developments in music were interdependent with developments in visual arts. **Futurist** and the **Dada** movements, in particular works such as Russolo's essay on noise (Russolo, 1913) and art of Marcel Duchamp are among the broad artistic inspirations – as is John Cage, who, inspired by both Dada and Futurism, incorporated their ideas of found objects and noise-as-music in his concepts of *chance operations* and *found media*. Cage's thinking affected how we go about producing and experiencing art. In his own words, "what is more musical, a truck driving by a factory or a truck driving by a music school?"

Fluxus, inspired by Dada, paralleled the experimental music movement in time, and focused on the personality and actions of the performers. The performers were frequently non-professionals, with the performance events taking place in public settings. These events, or Happenings as they are popularly known, served as an expansion of Cage's ideas of chance operations, standing in opposition to the dominant idea of art and media-making as a highly trained act. Happenings were improvisational in nature: they involved audience and encouraged the incorporation of ongoing events (context) into the performance. The 'score' of the performance often consisted of only a couple of simple instructions (e.g., 'raise one of your hands every time a

red car passes') based on which performers would develop their individual actions, with the resulting performances often highly complex due to the interactions between the performers. The idea of Happenings is what inspired our view of audience's role in interactive improvisational environments.

In order to create an audio or visual piece, the content must be generated or sampled. The idea of re-contextualizing existing media artifacts has been enormously influential - creating original work by means of sampling and scratching media samples appropriated from others has been an accepted instrument in music- and video- making of the last fifty years and constitutes a major inspiration for this work. This approach has been expressed through the techniques of superimposition, montage, and collage. One of the earliest documented studies of the effect of differing context in modern media is the Kuleshov experiment (Kuleshov, 1918), in which viewers were presented with three sequences, each consisting of a shot of an actor's face and an unrelated scene. The actor shots were identical in all three shots, yet the audience saw him act differently in each of the three, depending on the adjacent shot.

The tradition of sample-based manipulation of audio is typically traced to Schaeffer's Musique concrète (1952), continuing with Stockhausen (1956), Varese (1958), Zappa (1967), Oswald's Plunderphonics (1985), DJ Danger Mouse's Grey Album (2004), Avalanches' remixes (The Avalanches, 2000), and numerous other sample-based musical creations. In the visual domain, the tradition is even longer, going back to Duchamp's readymades (Duchamp, 1913), Warhol's soup cans (Warhol, 1964), Nam Jun Paik's video-art (Paik et al, 1993) and many others.

In the domain of interactive digital media-making, two schools of thought seem to exist in regards to contextualizing the artifact. The first school regards each media sample it uses independently of the context in which it was originally created, disregarding the networked nature of media and its behavior, and opting instead for linear or tree-like representation structures. This approach is exemplified in nearly every mainstream media editing tool (e.g., Adobe Premiere). The second school considers contextual information to be of a paramount importance (Davenport et al, 1996; Barry, 2005), enabling far more powerful storytelling. At the same time, stronger yet fixed context entails stronger ties between particular media samples. As a result, the chance of serendipity occurring in the course of the artifact's manipulation is lessened. Improvisational interaction aims at bridging the two schools by leaving the strength of the contextual links up to the participants, instead of embedding contextual rules within the underlying representation.

A well-known recent example of the technique is the work of the Emergency Broadcast Network (EBN), an artistic collective that utilizes sampled videos to create their art (EBN, 2006). While EBN's work consisted of painstaking editing, real-time **video manipulation** is becoming increasingly common. Artists who engage in it are looking to encourage experimentation and the ability to 'look under the hood' – the type of interaction all but

impossible with the 'wizard tools' described in *Problem & Motivation*. To quote Benton-C Bainbridge, an experimental video-maker: "...so even now I think it helps to be equal parts artist and renegade engineer to be a VJ. ... [I] think we need more simple and intuitive tools for media manipulation, but I'll always have a fondness for that fearless and wacky mad scientist spirit that Nam June Paik gave to start the live video art movement" (Bainbridge, 2004). Furthermore, it seems clear that artists are in a need of interactive environments that would allow both detailed and high-level media manipulation as well as group interaction. From the same source as above: "...we would usually do a collaborative jam, composed on the spot through multiple takes or completely improvised ... we cross-modulated dozens of oscillators to synthesize audiovisuals. After wiring together a tangle of cables the three of us would each grab a bank of knobs and tweak away to build up or destroy each others' oscillating groove."

An increasing number of programs aiming to facilitate improvisation are released every year (even if most of them adhere to a narrower media-specific view of improvisation than the one outlined above). In this section we will consider a number of such programs, focusing on the ones that (1) allow manipulation of audio, video, and text and (2) facilitate one or more of the characteristics outlined in *Cognitive Model of Improvisational Action*. As such, we leave out a discussion of purely procedural environments for audiovisual manipulation, most notably that of Max/MSP and its clones – that is, programs where the interaction is centered on designing interconnected manipulators of audio or video signals. One may object – don't people improvise with Max/MSP and programs alike? The answer is, of course – but so do they with a piano and nobody would consider a piano as an improvisational system; rather, it is thought of as a performance instrument that can be used for improvisation. Similarly, while wizard-tool toys and idiomatic improvisation software (such as Band-In-The-Box (PG Music, 2006)) have a wide variety of uses, they aren't described here due to a significantly different conception of improvisation.

The effort to use the machine as an artificial creator dates back more than 40 years. Hundreds of attempts at viewing the **machine as a co-creator** have been made, using rule-based, stochastic, and AI approaches. Over the last decade, algorithmic composition using **genetic algorithms** became a popular means to create new musical pieces, while requiring human input ranging from limited (Moroni, 1999) to full interaction (Lewis, 1999). Most of such systems focus on generating individual sounds or melodic snippets. Nearly all either use a linear timeline or employ a Max/MSP-like hierarchical network without providing a higher-level structural control.

Some particularly interesting developments are the Continuator (Pachet, 2002) and the Mosaicing (Zils & Pachet, 2001) applications as well as the KeyWorx (Waag Society, 2006). Continuator is a musical co-improviser that learns the style of music improvisation you perform and plays along,

keeping with the conventions of a given style. The system's output is truly impressive – in some of the examples it is nearly impossible to tell an improvising human from the machine. At the same time, the system presumes that its user is well-versed in idiomatic improvisation. Musical Mosaicing system combines unrelated audio segments to generate a musical piece. Users do not have to specify the piece they want but instead specify constraints about the whole sequence, leaving it up to the system to find the pieces that satisfy those constraints. Finally, KeyWorx is a multi-user platform for interacting with multimedia content in a synchronous (real-time) fashion. Various clients, developed on the platform enable cross-media synthesis, video conferencing and online gaming.

Voyager, a machine performer designed by George Lewis (Lewis, 1999), reflects the idea of the machine as a partner rather than a mere instrument under the human's command. As Lewis succinctly puts it, "the discourse of computer music is really shot through with prosthetic conceptions. I'm not dealing on a prosthetic basis. So when people talk about instruments as 'controllers', the language of mastery and power, or where a musical instrument becomes just a kind of 'user interface', I start to get a little nervous. That's not animistic enough for me" (Lewis, 1997). On the downside, getting Voyager to produce an impressive output requires an impressive input (provided by Lewis, one of the best living jazz trombone players). On the upside, Voyager comes closest to a true performance in tandem with the machine.

Among tools utilizing evolutionary algorithms for improvisational purposes is Tim Blackwell's Swam Music (Blackwell, 2005) which produces musical improvisations by building upon the self-organizational properties of swarms (e.g., "avoid collisions"). These properties are used to generate musical behaviors such as melodies, harmonies, and rhythms. Similarly to the Emonic Environment, the program responds to live audio input, while the output is done using sound synthesis rather than audio samples.

Another example of a music generator is CAMUS (Miranda, 2001), a cellular automata-based system based on two of the existing cell automata algorithms. It utilizes a Cartesian model with each dimension (3D) being a note. Each of the 3-note groups is evaluated by each of the two algorithms in parallel, modifying the group based on the state of the cells around it and in this way determining the duration and rhythm of each note.

VoxPopuli (Moroni et al, 2002) is a system for evolutionary music creation, trying to mimic the evolution of a living system. Each of the genetic populations is divided into fixed groups (soprano, bass, etc) and a genetic cycle attempts to continuously choose the best of each group. It uses the computer and mouse as real-time controllers, differing from other evolutionary environments that might require explicit user feedback.

In the visual domain, video DJ-ing, or VJing as it is popularly known, has quickly progressed from museum installations of curious-and-bizarre (e.g., Nam Jun Paik) to mainstream culture (VJs traveling with acts such as Coldcut, or appearing on MTV). Most of the available VJ-ing tools employ concepts from the world of traditional sound sequencing software, with the video footage presented as continuous storylines (Sony's Vegas), procedural blocks (Cycling '74's Jitter) or trigger sequencing (VidVox's GRID).

Visual artists have long explored improvisation. Among most influential are Woody and Steina Vasulka, who have worked on the collaborative relationship with the machine in the context of video. To mention just one of their works, Artifacts (Vasulkas, 1998) explores the capacities of Digital Image Articulator, a device designed by Woody Vasulka. In his own words: "I have to share creative process with the machine – it is responsible for many elements in this work. Images came to you as they come to me – in a spirit of exploration."

Another intriguing early example of real-time video improvisation is the RGB Piece (Davenport, 1981). In it, three cameras, each with a different color filter, were connected to a video mixer. As the cameramen move around, each independent of others, an unpredictable and often surprising overall scene is being captured.

Karl Sims' Galapagos (Sims, 1993) is a genetic algorithms driven system for growing a population of abstract visual forms. Users provide their feedback (interest) in a given form by standing in front of one of the twelve screens that show the elements as they undergo the process of mutation, mating, and reproduction.

Visual improvisation draws its roots from various avant-garde traditions. More recently, however, improvisation collectives have started bringing visual (and at times audiovisual) improvisation to the wide audience. Cambridge-based sosolimited (soso, 2006) and 242.pilots (242.pilots, 2006) are just two examples – collaborating in real-time to produce visual (242.pilots) or audiovisual (sosolimited) improvisations. The results are visually and conceptually stunning. The view of the performance setting itself is quite conservative, with the audience playing a traditional, passive role, and the machine acting as merely a tool satisfying improvisers' demands.

Cognitive Model of Improvisational Action (CMIA)

What is improvisational interaction and why is it necessary or beneficial? What are its measurable characteristics, particularly in the context of exploring a digital media space? How would interacting with a system that facilitates improvisational interaction be different from interacting with a digital media editor or browser?

In this chapter we introduce CMIA; a model that represents improvisational action as a set of behavioral characteristics. A participant must exhibit these characteristics for the behavior to be considered improvisational[1]. The objective of introducing this model is to provide a measurable way to evaluate the degree to which actions of both a computer system and of a human participant within it can be considered improvisational.

After presenting CMIA, we discuss the rationale for applying the model to our interactions with digital media. We suggest *genetic algorithms* as a computational method for achieving a behavior that would be regarded as improvisational by CMIA and present a formalized representation of the machine's envisaged functionality (implemented as part of the Emonic Environment, a CMIA-based system described in *System Architecture & Implementation*).

What is improvisation, when viewed in the context of media exploration? It is a process of making, sharing, and manipulating structures according to some set of strategies. In other words, improvisation, just like composition, is a process of making choices and taking actions. The difference between the two is temporal: in improvisation, the guiding strategies may be preplanned but are more likely to be emergent from the ongoing interactions, with choices (resultant from these strategies) all made in real-time, and in parallel. In other words, the activity of improvisation, unlike composition, does not demand a fixed set of rules or a defined resulting artifact: strategies that guide action are defined in the course of improvisation; the process of exploration is more important than any artifact that may result. As a result, the scope of possible actions is virtually limitless, and depends on the improviser's perception of the moment (context, stylistic constraints, feedback from other improvisers) and of the past (prior actions taken).

---

[1] Our discussion of improvisational action is limited to the context of interacting with digital media.

Defining CMIA

In what follows, we introduce six characteristics that comprise Cognitive Model of Improvisational Action (CMIA). The characteristics are:

1. Focus: Multilayered. Whether deciding what to do next, or reflecting on past actions, improvisers switch between multiple representations, focusing on details of a given media element at one moment, only to shift to structuring an overall development (e.g., a climax) a second later.

2. Content and Structure: Constantly Changing. Instead of first creating a structure and then filling in content, in improvisation both are generated and evaluated on the spot. The path through the improvisational landscape – content elements' interrelationships in terms of time, pitch, brightness, and endless other characteristics – is created dynamically. For instance, an improviser may spontaneously decide to repeat a given motif every few seconds, or increasingly desynchronize two ongoing motifs, whatever the actual motifs may be at that given moment.

3. Authoritative Score: Absent; Price of Mistakes: Low. Improvisation, unlike a performance of a composed piece, cannot be quantitatively compared with a preexisting deterministic score. As a result, noncompliance with predetermined solutions is no longer seen as a mistake. As a result, experimentation becomes less threatening, risk-taking is encouraged and generating new actions becomes easier than following predetermined ones. In other words, following through an existing structure in a "perfect" manner is no longer seen as important; instead, improvisers focus on designing new explorative paths and learning from unintended mistakes and unexpected successes.

4. Goal: Process, not Artifact Production. An improviser, unlike a feature-film cinematographer, a Western composer, or a product designer, has little concern with producing a fixed final artifact – a movie, a sonata, a pop song, or a chair. While improvisation might be recorded and, as such, seen as a fixed construct, it is primarily a process of exploring, contextualizing, and interrelating memories, perceptions, and actions, with its value being in the process of making itself. Unlike a structured narrative, which assumes convergence of sequences into a single overarching structure, improvisation is a process of dynamically exploring interrelationships within the media space. Improvisers weave together an array of "sketches" which gain relevance and meaning only as the improvisation unfolds. The importance of individual elements lessens in favor of the paths along which these elements appear, become significant, and disappear – that is, the strategies by which the overarching structure is generated and explored. Improvisers employ these strategies to find structure within chaos – only to break it again a moment later, and start looking anew.

5. Participation Responsibility: Distributed. In the compositional paradigm, performer's participation is to last as long as is required for the performance of the piece to be completed. Walking out in the middle of the creative action means that the action is stopped until the participant's return. With time, social norms have emerged that prevent such walkouts, as manifested in events such as concerts of classical music: events that force the participants into a highly ritualized act of performance with no escape until the end of the action is reached. Improvisational setting, on the other hand, allows for a more relaxed participatory mode: there is no deterministic score and the length of participation is left up to the improvisers, who are free to commence / end their participation whenever desired. As a result, no fixed contract specifying responsibilities of control (i.e., a balance of power) exists between improvisers. Improvisers set their degree of participation dynamically, following implicit and explicit negotiations and are free to renegotiate their roles in the course of improvisation. Such a dynamic framework frees them from being preoccupied with every aspect of the creative action, facilitates experimentation, and lowers the entry barrier to taking part in the creative process.

6. Context and Disruptions: Relevant and Incorporated. In a traditional Western performance, the ideal audience and environment are those that happen, for all purposes, to appear invisible. They may applaud at strictly defined times, but otherwise any action by them or ambient action within the environment (e.g., industrial noise), is seen as an obstacle to having a proper performance. Classical musicians are taught to ignore any disruptions as much as possible, performing in a complete disregard of what goes on around them. Improvisation, on the other hand, calls for incorporating as much of the context and disruptions as possible into the ongoing performance. People's voices, street noise, the temperature, the size and the reaction of the crowd, all shape the improvisation, and contribute to its uniqueness.

Case for Machine Participation

How do improvisers manage to do all that, continuously and in real-time? Don't they feel overwhelmed, having to come up with both strategies and actions in the course of improvisation, all while interpreting and reacting to changing strategies and actions of other participants and / or the audience? One way in which improvisers cope with the complexity is by performing as a group, with each participant responsible for a given aspect of the ongoing improvisation. While often appropriate, this solution has two drawbacks: first, in order to engage in an improvisational task, a sufficient number of participants are required, thus making the solution inapplicable to the scenarios of solo improvisation. More importantly, in the context of interaction

with digital media, this solution does nothing to further the role of the machine, leaving it as a command-executing box and missing out on its potential as a co-improviser.

Why trust a machine with what essentially is a creative choice? Modern computers possess a great power, that of *recombination*. Capable of searching through billions of media artifacts in a reasonable amount of time, they are also increasingly capable of putting the found artifacts together, as well as generating multiple ways in which the found artifacts may be manipulated.

Today it is possible to share / store / rate / label digital media, both your own and that of others. When interacting with digital media, we no longer think only in terms of individual artifacts – increasingly we focus on collections of media; collections that attempt to mimic the traditional photo albums and record sets (even if often missing out on the storied qualities of their analog counterparts). And yet, the role of the computer remains that of a tool waiting for our commands. Machines do not engage us in a process of media exploration; it remains up to us to decide what and how we'd like to explore. If you happen to be interested in wildlife, your computer provides you with means to search for pictures of giraffes, gazelles and polar bears, but it is up to you to initiate the search, and more importantly, decide what you do once the pictures are found. Constructs like albums are static – opening an album I created years ago shows me the same album I last saw, reflecting the limitations of our means of organizing and presenting digital media.

It is now time for the next step; we have to take our interaction with media beyond mimicking physical interfaces of yesterday such as albums. That means getting rid of static representations of individual pieces of media, and focusing on manipulating and trading networks of processing structures – structures that determine how, where, and when a given media content is seen, heard, shared, traded, or changed. Creating and manipulating such structural networks is a task perfectly suited to being performed in tandem with the machines, since computational power of recombination can be utilized to generate, suggest, and evolve the networks, as well as apply them to ever-changing sets of digital content.

Editing, Browsing, and the Benefits of Improvisation

Let's consider the two currently prevalent interactive frameworks: editing and browsing. Whenever you want to create an original work using a computer, be it a sonata or a work resume, you have a single and only option: using some type of an *editor*. The activity of editing, with its precise

control mechanisms, provides a feeling of authorship. However, the editor limits your creative power by constraining the type of media exploration that can be done in three ways: (1) the interface is passive, and fully controlled by the human; (2) the only creative action allowed is that of defining individual parameters of media artifacts (typing the letters, applying the formatting, setting the colors, etc), leading to (3) lack of encouragement for higher-level structural and temporal objectives and mappings. What are those structural objectives? Here are a few examples –

- How should rhythmic tension increase over time?

- How should the amount of red color in the images being explored affect the echo of the sounds being played?

- What should be the balance in volume between the frequently-played melodies and motives that are only played once?

In an editor, a myriad of low-level decisions mask and distract from navigating the media space structurally, as a whole, and asking questions such as above. Most importantly, the focus is exclusively on the final product, not on the process by which you explore, learn, and express.

In difference from the editor, the *browsing* framework provides a powerful feeling of traveling through information. Some browsers may expose structural characteristics of the media space being explored (e.g., change in rhythmical tension over time). At the same time, structural features typically come at the cost of a significantly reduced precision; the activity in which one engages within the browser context is one of rearranging at best, often consisting of blindly following links and templates set by the designers of the browser or of the content being explored. Lacking the control precision found in editors, users of browser-type media systems rarely think of what they do as a creative process, or of themselves as authors.

The approach proposed in this thesis – that of improvisational interaction – aims to introduce a hybrid type of experience, one in which the participants get a feeling of authorship while being able to explore the media space in a higher-level, structural fashion. Improvisational interaction, however, is not merely a mix between editing and browsing. Rather, by combining aspects of the two in the digital context, we aim to bring computer users closer to having an improvisational experience while creating, editing, and browsing media. The benefits of such an experience are clear: for improvisers, the distance between the creative thought and its resulting implementation is typically very short – the immediate, real-time nature of improvisation enables people to simultaneously be authors and navigators. Equally important, the ability to manipulate structures and content at the same time means that people are likely to focus on manipulating structures, departing from rearranging and exploring content which is their focus in the browsers and editors of today.

Defining Semantics of Media Space

Audio, video, and text media are more accessible today than ever before: billions of samples have been digitized and made publicly available. Similarly, creating media content from scratch and making it publicly accessible has also been made incomparably easier in the last few years. Hundreds of media-oriented semantic systems (i.e., grammars that define media interrelationships) have been developed – hyperlinks being the most widespread if not the most intricate example. Why then would yet another semantic representation of digital media be needed? The answer is threefold:

1. It is presumed that people know both what media content they want to find and how it should be presented. We believe that this assumption is wrong, and people often have no idea what they are looking for aside from the most general idea (e.g., 'something funny'), and they have even less of an idea about how to make anything new and exciting out of the found media, mostly resorting to simply going through the found media song by song, image by image. It is clear that if we are to reinvent browsing as a creative activity, representation of digital media must include structural elements capable of (a) proactively generating media structures (contexts) within which the media is placed / experienced and (b) responding to participants' feedback by further manipulating structural and perceptual properties of the media space being explored.

2. The meaning of a given media can be completely different depending on the context in which it is experienced. Montage has been a cinematic tool since the days of Eisenstein and Griffith, as exemplified by the Kuleshov experiment; it is now time to make montage an integral part of any experience involving digital media. It can be said that in an improvisational scenario no single context (media structure) is sufficient for fully exploring a given media artifact. In order to allow people to explore media space in an improvisational manner, the system should enable placing available media artifacts in as many contexts as possible. To provide such contexts, the structural constructs dynamically generated by the semantic representation need to be accessible, tradable, modifiable, and augmentable independently, just like the media artifacts they operate on.

3. Functionally, the structural components need to (a) utilize properties of a particular media, (b) facilitate mappings between multiple media types, and (c) be configurable by the genetic algorithms, to allow improvisation-like media exploration in tandem with the machine.

With these requirements in mind, we set out to develop a semantic representation of improvisational action. We found a *directed graph* to be well suited to our needs. The rationale for picking directed graph was rooted in the following considerations: improvisation is an activity of making and

reusing patterns over time. In the DMS context, a pattern is constituted from behaviors of interdependent media manipulators over a given timeframe. At the same time, any pattern can be broken by the improviser. We use a graph to represent individual actions as driven by the nodes, with patterns being clusters of nodes connected to each other. The recurrent qualification of the network stems from the need to acknowledge the characteristic behavior of an improviser where action A leads to action B which leads to action C which in turn leads back to action A, forming a cycle. Finally, we used a directed graph representation (one in which any pair of nodes can only be connected in one direction) to minimize situations of network oscillation (two actions causing each other ad infinitum).

Nodes (called *emons* in our network) serve as triggers for one or more media manipulations (called simply *Actions*). Edges between the nodes are paths (conditions) that determine whether and when the Actions follow each other. The weights on each edge represent the probability of transitioning from one node to another (i.e. the probability of choosing a particular action) as well as the response time.

An essential characteristic of improvisational thinking is that the perception of the space being explored changes within the course of improvisation. This change can be represented as resetting the weights on the graph according to a new set of rules. Improvising can then be seen as traversing the graph in a non-deterministic manner.

Before we consider the implications of this design, a bit more detail on the network's operation. An emon (a node which triggers media-manipulating Actions) does two things: it fires (communicating its charge to any connected emons), and it executes its assigned Actions. The actions are executed every time the emon is stimulated. The firing, on the other hand, is what gives the network its temporal flexibility. Every emon operates in a range from 0 (inactive) to 100 (fully active). Firing is based on the emon's current activation level and happens (1) periodically (every x %), (2) at a given value(s), or (3) as long as the emon in question has a value above or below a set threshold. The resulting stimuli is carried along an edge, with the individual speed and strength set for each edge (with time offset = 0 for concurrent stimulation of connected emons, or >0 for execution with a temporal delay).

What does the progression of events within a typical improvisation looks like when represented in this networked view?

-   At t_0, improviser picks a random point within the graph to start – or multiple points since improvisers usually think of more than one action at any given time. In an interactive system built on this formalization, his initial location is determined by observing the active cluster of emons.

- At t_1 and thereafter, the improviser is free to follow any edge, leading to any other emon(s) within the graph. The decision on which edge to follow is based on an evaluation of edges' weight (the likelihood of choice B following choice A).

Once we start thinking of media manipulation in a structural fashion, we can make observations in regards to behaviors of both individual structural components (emons) and of the network as a whole. The table below shows the following (assuming an ongoing evolution):

- Connection density: the ratio of the number of emons to the number of edges (connections between emons).

- Firing pattern: the amount of activation needed to make the emon fire. A fire pattern tells us the intensity at which different emons fire.

- Action density: ratio of the number of actions to the number of emons.

Connection Density	Firing Pattern	Action Density	Resulting net
High	Medium		We are likely to see an oscillatory recurrent pattern, where different areas of the network get activated in turn leading to the repeated execution of the associated Actions.
Low (or none)	Any apart from very low		We are likely to see a non-recurrent sparser activation of Actions throughout the network.
High		Low	We are likely to see recurrent but sparse activation in the network (i.e. the network largely passing the signal but not triggering any Actions).
High	Very high / very low		The resulting pattern of Actions' execution is not significantly affected by the structure of the network.

Earlier we mentioned the power of recombination. How can that power be achieved, that is how can the machine help us explore digital media space in improvisational manner? We believe one possible solution is using *genetic algorithms* (GAs) as our vehicle for media exploration. In the following section we discuss the rationale behind our decision.

Why GAs?

If machines are to become our accomplices in improvising, they need to be capable of making decisions in regards to the structure of media networks. These decisions (behaviors) need to be recognizable by us as parts of the ongoing improvisation, yet not be deterministic reactions to our actions. By making the machines capable of observing and initiating such behaviors we hope to encourage people to regard the machine as an independent creative force rather than a command-executing box with no initiative. How can we enable such behaviors in a machine? To do so, we should consider the behavior of human improvisers. As mentioned before, when improvising people often have only a vague idea of what's to come; a master plan is nonexistent. They do however have the capacity to:

- Recognize changes in the environment (be aware of the context).

- Remember and reflect on prior and current actions, both their own and those of other improvisers.

- Remember structures they found interesting or learned something from.

A machine co-improviser must then be able, to some degree, to recognize, remember, and respond to events taking place within the media space. How can it do so? We argue that the process of improvising is similar to that of evolution, making genetic algorithms (GAs) appropriate as a framework for recognizing, evaluating, and responding to improvisational actions of human participants.

The question can then be asked: is there evidence of similarity between evolutionary and improvisational processes? To answer, let's consider behavior of the GAs and compare it with that of improvisers.

Genetic algorithm (GA) is defined as "…an iterative procedure maintaining a population of structures that are candidate solutions to specific domain challenges. During each generation the structures in the current population are rated for their effectiveness as solutions, and on the basis of these evaluations, a new population of candidate structures is formed using specific 'genetic operators' such as reproduction, crossover, and mutation"[2]. GAs are particularly applicable in situations of non-linear and/or discontinuous relationships between elements (like those within a digital media space).

---

[2] From http://www.stanford.edu/~buc/SPHINcsX/

37

GAs are commonly used to evolve neural networks, paralleling our view of improvisation as a process of exploring networks of interrelated perceptual and structural data.

A number of parallels can be drawn between the operation of the GAs and behaviors exhibited by improvisers:

- Both improvisation and evolution can be seen as continuous series of mutations of an existing population (genes in the case of the GAs and media artifacts/structures in the case of improvisation). Each mutation carries a consequence even if it may not be immediately apparent in the behavior of the organism (media appearance in our case).

- The mutations result in the emergence of new individuals and behaviors.

- The number of individuals (or participants) is not predetermined.

- Mutation sequence (or performers' actions in the case of improvisation) can be described as a constrained randomality.

To see the similarities in more detail, let us consider the following comparison (the Improvisational Action column on the left parallels the characteristics of CMIA described above):

	Improvisational Action	Evolutionary Process
Focus: Multilayered	Improvisers operate on multiple abstraction levels, dealing with the minute and the overall at once.	Evolution effects changes on both individual and, indirectly, group levels.
Content & Structure: Constantly Changing	Improvisers are responsible for generating both the content and the structure that guides elements' real time interaction.	Mutation strategies unfold in real time. The direction of the mutation is determined by evaluating the last mutation and modifying the fitness function that guides the evaluation accordingly.
Authoritative Score: Absent;	In improvisation, there is no overall score to follow; improvisers have no interest in conventions of a given	Evolution can be biased toward a particular solution. However, 'wrong' mutations do not stop the evolutionary process; in fact, they

Price of Mistakes: Low	style, focusing instead on generating original paths through the DMS as well as repeatedly perusing those of the generated paths that they regard as interesting.	may lead to potentially interesting solutions. Mutations are picked dynamically, according to the fitness function, and incoming stimuli. No fixed mutation sequence is followed.
Goal: Process, not Artifact Production	Improviser is unconcerned with producing a final artifact such as a movie or a pop song. Instead, he focuses on the process by which individual media elements are brought together into an overarching structure.	Evolution is a goalless process – its purpose is in the process itself, an endless sequence of mutation and crossover operators performed upon the existing population.
Participation Responsibility: Distributed	Decisions that shape an improvisation come as a result of implicit and explicit negotiations between the participants. Improvisers are free to choose when and how much they participate. The decentralized nature of improvisation makes experimentation easier and potentially leads to the audience playing part in the creative process, affecting the improviser's actions, and turning into improvisers themselves.	A mutation that is 'popular' gets picked often as the mutation to perform. Subsequent interactions between the resulting individual and other members of the population bias the ongoing evolutionary process, affecting the fitness function that evaluates the 'goodness' of the mutation in question and thus changing the probability of that mutation happening again.
Context: Relevant and Incorporated	Given the absence of a prewritten score, actions of other improvisers as well as events in the surrounding physical space often become part of the improvisation.	Mutations are environment-dependent; changes in the environment lead to change of the fitness function which in turn affects the probability of subsequent mutations.

Given the parallels between evolutionary and improvisational processes, we argue that GAs, being a computational model of evolution might also be appropriate in enabling improvisation by the machine. In order for a GA-based system to be able to co-improvise the following needs to be done:

1. Define the types of elements that can be affected by the evolution and their possible mutations. By looking at the change in the encoded representations of the individual components, the system is then able to learn about and modify its actions. This is described in both the preceding and the next chapters.

2. Learn how to incorporate human input in the evolutionary process; it has to be slow enough for the human to see and reflect, and fast enough to effect real change within the timeframe of a performance. This is described in the following chapter.

To see these concepts in practice, a system called the Emonic Environment (or EE for short) has been designed, to serve as an implementation of CMIA and enabling improvisational interaction with digital media by bringing together structure- and element-centric approaches to media exploration.

System Architecture & Implementation

Figure 1: Samples of the Emonic Environment's visual output

Rationale for the Emonic Environment

We believe that the idea of creatively structuring a media stream from personal media collections will ultimately replace browsing as we practice it today. The Emonic Environment (EE) is a first step in this direction - it allows us to roam around our media collections, compose by browsing, transforming the act of browsing into a creative process. The objective behind the creation of the EE is to encourage people to act in an improvisational manner in the context of their interaction with media. In the *Cognitive Model of Improvisational Action* we have defined what constitutes improvisational behavior in the context of interaction with digital media. Here is a summary of how each CMIA characteristic is manifested within the EE –

1. Focus: Multilayered. EE's multilayered architecture allows participants to choose whether to focus on controlling properties of specific elements, higher-level structural constructs, evolutionary process, or one of the intermediate-level representations.

2. Content and Structure: Constantly Changing. The EE lets the participants create, employ, and exchange structural rules and content in real-time.

3. Authoritative Score: Absent; Price of Mistakes: Low. The EE does not aim to figure out aesthetic expectations of its users. Instead, by utilizing built-in genetic algorithms (GAs), the EE aims to work in tandem with the participant in generating and exploring paths within the DMS, thus making experimentation easier. The participants, in turn, provide feedback to the GAs by placing constraints on the types and nature of properties upon which the GAs operate, thus affecting the subsequent GA operation. As a result, no authoritative score is required and no action is considered to be a mistake, with the participants and the machine co-improvising by manipulating structural elements and evolutionary constraints.

4. Goal: Process, not Artifact Production. In the EE, the traditional notion of recording and reproduction is absent. Instead, participants, in tandem with the ongoing evolutionary process, explore the paths between the current and other digital media spaces. An EE network can be saved, but, upon subsequent restoration, its evolution will always be unique, depending on the current and target media spaces and active evolutionary constraints.

5. Participation Responsibility: Distributed. Participants are free to pick the high- or low-level aspects of the creative action they want to control at any given point. They are also free to step back and listen and watch other participants' and genetic algorithms' actions. However active they choose to be, the performance continues with no interruptions.

6. Context and Disruptions: Relevant and Incorporated. The EE strives to incorporate as many context-dependent inputs as possible. An array of mechanisms for doing so has been designed: sensors that perceive the outside environment; microphones that monitor the input; cameras that trigger on movement, etc. Any outside disruption to the parameters of the evolutionary process is similarly adapted to in real time. The EE is capable of interpreting input from a variety of sensors (e.g., temperature, motion), and letting this input affect both low-level individual elements and higher-level properties of the ongoing evolutionary process.

Defining the elements of the media space
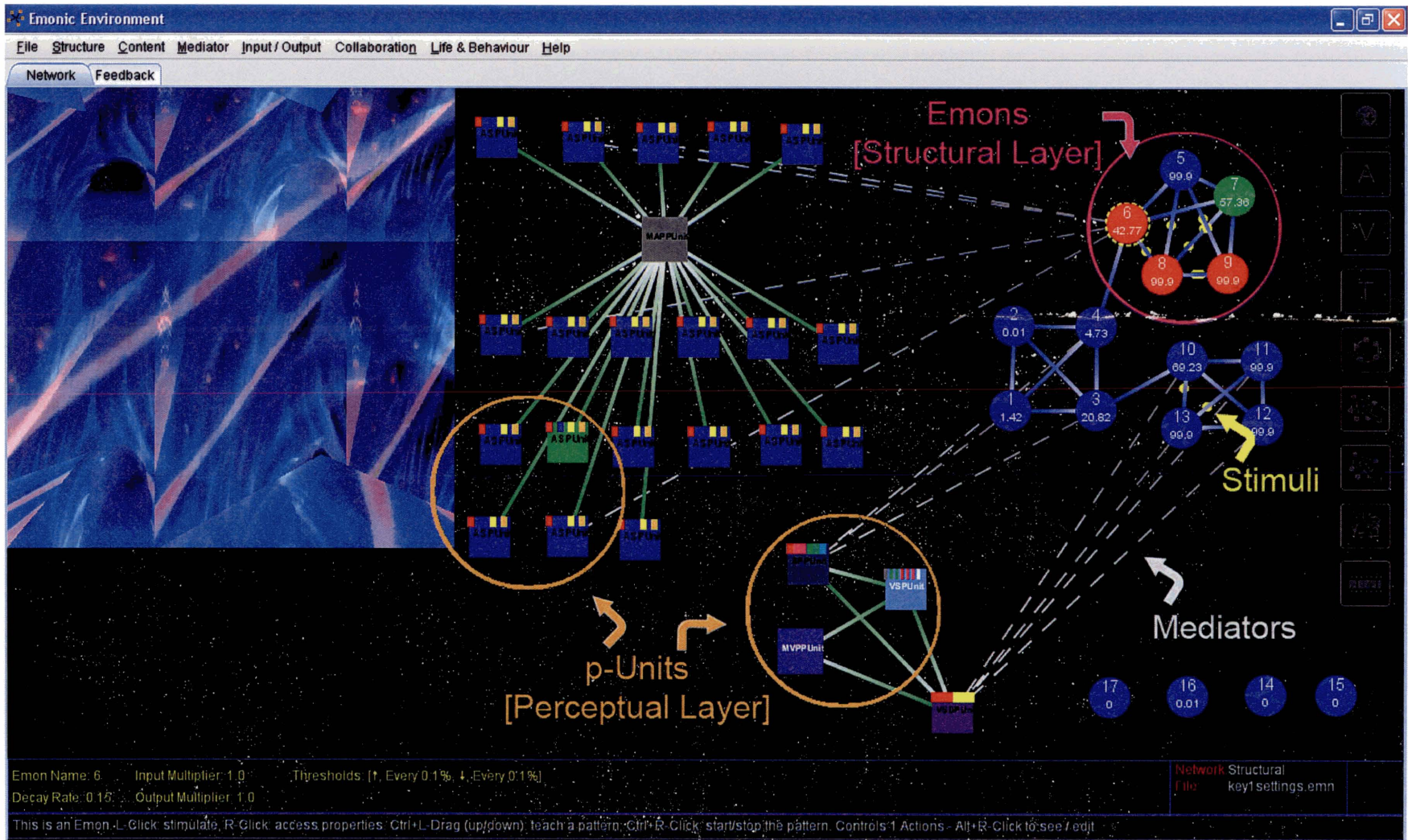
Figure 2:   Emonic Environment's user interface

Our objective in building the EE was to explore situations of co-improvising: a human participant and its machine counterpart symbiotically contributing to each other's actions. To achieve such interaction, we integrated component-centric precision (e.g., through direct input) and process-centric structural control (e.g., through evolutionary processes).

The EE, written in Java, manipulates audio, video, and text in shared and individual contexts, and utilizes sensors, microphones, cameras, and cell phones as input devices. Features described below represent only a fraction of the system's capabilities, chosen as relevant for understanding the scenarios of use described in the *Scenarios* section. We start by describing our encoding of the digital media space into a set of network elements that allow both structural control and content abstraction, describe the resulting architecture, and follow by describing a few of the elements in more detail.

Three classes of constructs for the improvisational space are defined: Two main constructs are the *Emons* and the *p-Units*. The emons act as indicators/controllers of structural activity and the p-Units as manipulators of the source media content. The two main constructs are complimented by an auxiliary one, called *Mediator*. The flow of actions in the EE is as following: an emon, when active, sends stimuli around the network. The stimuli trigger Mediator(s) associated with a given emon, which in turn leads to Mediators instructing the appropriate p-Units to change their state. That state defines what you hear and see. In other words: you play with the EE by adding, connecting, and modifying three types of elements: **Emons, p-Units, and Mediators**.

 p-Unit stands for 'processing unit'. It controls a given subset of your media's properties, its looks and sound. Multiple p-Unit types exist, each responsible for a particular type of media manipulation. P-Units can be interconnected to supplement each other's manipulation capabilities.

 Mediator is a function that lets one or more *emons* control one or more *p-Units*. Each Mediator controls a single p-Unit property, and has the capacity to change it in many ways. You can create Mediators manually, let the EE create them for you, or get them from other users of the EE online.

Emon is a neuron unit of a neural network. By triggering Mediators, it controls how p-Units' behavior changes over time. By interconnecting multiple emons, and associating some of them with Mediators, you can design an organism that controls processing of media on a higher level than merely interacting with p-Units, allowing you to think about interaction with your media in a way similar to that of a practicing improviser.

All 3 element classes (emons, p-Units, and Mediators) can be manipulated both individually and as a group, as well as exchanged with other users. Three ways to explore the EE network exist:

- Direct manipulation of p-Units (editor-like control of their individual properties). This is similar to using a media editor.

- Direct manipulation of emons and mediators. This allows defining polyrhythmic relationships between different p-Unit manipulations.

- Manipulation of emons, mediators, and p-Units by the means of an evolutionary process. This allows co-performance with the machine.

The three element classes form the three *layers* of the Emonic network. Emons and their interconnections form the *Structural* layer. Interconnected p-Units form the *Perceptual* layer. Finally, emons connected to p-Units by the means of the Mediators form the *Mediated* layer. Simply put, the Perceptual layer defines *what* and *how* we hear, see, and read, while the Structural layer defines *when* and *why*. The operation of the system is purely real-time, with no offline processing.

Perceptual layer controls parameters of audio, video, and text. It is populated with p-Units, which provide precise and immediate control of media. P-Units come in a variety of types. For instance, a Visual Sample p-Unit knows how to play a video, and the Red Scale property of that p-Unit controls the amount of red color in the image. The Structural layer, represented as a recurrent neural network is an abstracted control mechanism that defines how p-Units behave over time. The two layers communicate by the means of Mediators which, when triggered by the emons, define p-Units' change (similar in a way to how a traditional linear sequencer helps in arranging and manipulating media samples by providing a timeline and a set of possible effects). A simple EE network consisting of the Structural, Perceptual and Mediated layers is shown in fig. 2.

Interconnected emons communicate by firing *stimuli*. Every emon connection is unidirectional. In other words, if emon A fires stimuli to emon B, emon B cannot fire stimuli to emon A. The emons fire whenever they are *stimulated* (i.e., there is a significant change in their activation level). Once

stimulated, emons trigger the associated Mediator(s). At that point, the triggered Mediators determine, using the emon's activation level as a parameter, what kind of change should be effected upon the p-Units. That change ultimately defines what we hear and see.

It is up to the participants to decide whether to interact with each of the three layers directly or engage the machine as a co-improviser. Overall, three types of action are available to the human participant: (1) controlling individual properties of the objects within the three layers (editing-like activity), (2) utilizing genetic algorithms to evolve between networks, while providing feedback reflecting on the ongoing evolutionary process (browsing-like activity), and (3) contributing and exploring media and its control structures (a mix of browsing and editing).

Perceptual Layer

Overview

The Perceptual layer is populated with *p-Units*[3] of various types, each with its own set of features. The p-Units control how the media is generated, modified, and played back (e.g., speed or volume of a sound, rotation angle of a video frame, semantic relationships of a piece of text). The modular architecture of the p-Units (their ability to form connection with other p-Unit types) allows for building complex processing structures. Overall, p-Units can be thought of as building blocks of a media-processing engine. As a result, Perceptual layer can be built to taste and reconfigured in real time by combining multiple types of p-Units.

There is no single starting point for the Perceptual Layer – it has no temporal characteristics of its own. Rather, each Audio, Visual and Text Sample p-Unit can be seen as an origin of yet another processing sequence. This patchable modular synthesizer metaphor may be familiar to anyone who has played with programs such as Max/MSP (Cycling '74, 2006). Unlike the latter, the EE limits the patch metaphor to the bare functional necessity – if in Max/MSP putting processing units together is seen as the way to create pieces (or 'patches' as they called), the only purpose of exposing these in the EE to the end user is to make precise control of the ongoing perceptual change possible if desired. The p-Unit types currently available are:

---

[3] p-Unit: a media-processing functional primitive; put together, p-Units form interconnected structures that generate, modify and present media.
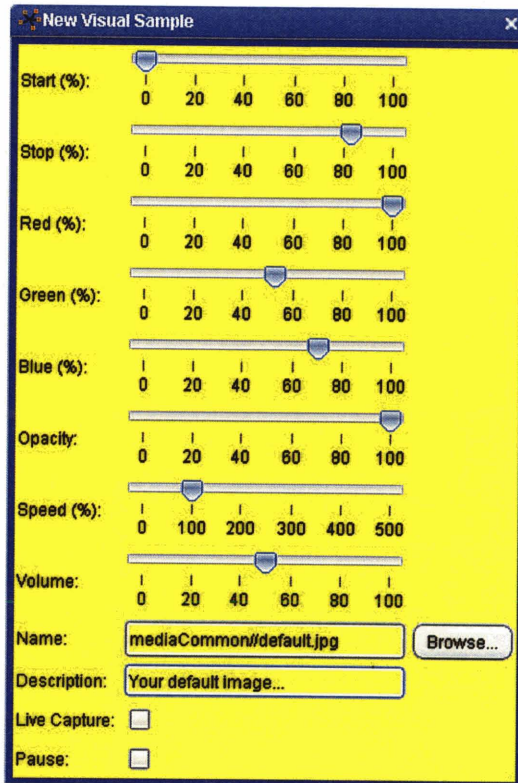
p-Unit :: Audio Sample (AS)
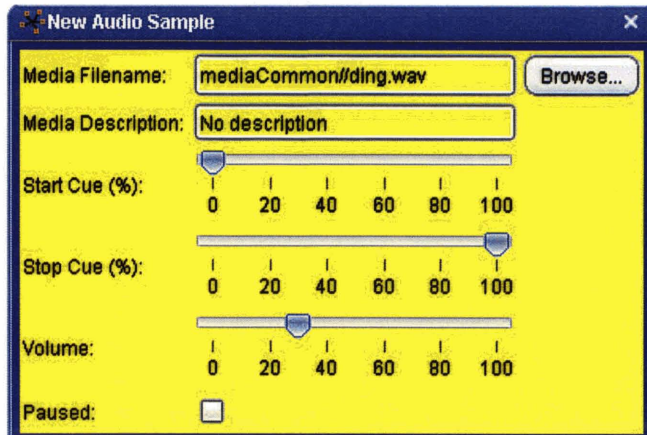
AudioSample p-Unit controls how a given audio sample, be it recorded by you, or retrieved from a database, is played back. AudioSample p-Units allow manipulation of sample-based audio; properties such as volume, start / stop cue points, playback direction, time and pitch shift, audio file location (locally or on the net), and other properties, are all individually adjustable.

To understand the usefulness of the individual properties, remember that any property repeats over time and that treating properties structurally (i.e., focusing on the nature of change rather than the particular content they might control) may lead to interesting effects over time. For example, take a look at the start and stop cues: let's say you have an audio file where the middle part fits well in a given context, but you don't feel like cluttering the overall output with the rest of the sample. Instead of permanently changing the sample, you can just change the start and stop cues, to indicate which portion of the sample should be played. Now, as you go on, you might decide that the rest of the audio sample fits the new conditions better and get a result by simply changing the cues. After all, any editor can do that, but now consider that you can set up an action that will each time change one or both of the cues, thus changing which part of the audio sample you hear based on other actions. In other words, as the network evolves, the participant might decide that the rest of the audio sample should also be explored. By adding start and/or stop cues to the list of properties being evolved, we can now explore different parts of the audio sample, creating a different sound without modifying the source file or manually changing any parameters.

p-Unit :: VisualSample (VS).

VS allows playback of video and still images. As such, it parallels the functionality of AudioSample and TextSample p-Units. To play a picture or a video, participants associate the p-Unit's Name property with a video or an image (local or remote) they would like to manipulate. VisualSample recognizes any file format recognized by Apple Quicktime. Using Apple Quicktime, individual frames of the video are extracted and passed to JOGL, an OpenGL library used for video manipulation. VisualSample's properties include:

- Start & Stop Cue, defining the limits (relative to the sample's length) within which the playback loops. Reversing the start & stop cues leads to the video playing backwards.

- Individually addressable prominence of red, blue, and green channels within the visual sample.

- Opacity, or prominence of this video source within the Master Visual Player p-Unit's display space – its visibility is calculated given the total amount of connected video sources and their respective opacity. The setting of 0 results in a completely transparent (invisible) image, 100 in a fully visible one.

- Speed of video playback, ranging between 0% and 500%, where 100% is the natural play speed.

- Volume: defines the loudness of the audio track within the video.

- Name of the visual sample to manipulate.

- Description: Similar to the Audio Sample p-Unit's Description property, it is a free form description shared and edited by all the participants of the EE. At each session, an updated list of descriptions is uploaded to the common server, where it is disseminated to the available EE clients.

- Live Capture mode, which, if activated, acquires video image from a connected video camera rather than an on-disk or online prerecorded source.

- Pause, to temporally control video playback according to Structural layer's input.

p-Unit :: Spatial Divide (SD).

Visual Subdivide ... ✕

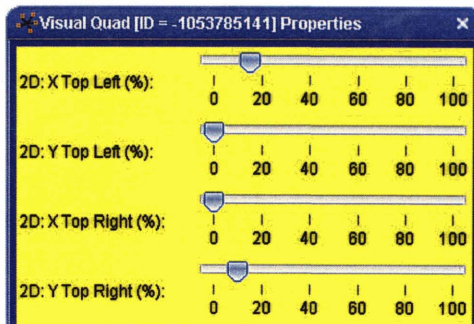|	
Horizontal Ratio:	1,2,1
Vertical Ratio:	2,2,3
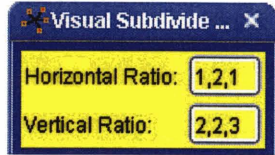
Spatial Divide p-Unit divides and replicates its visual input (a Visual Sample or another Spatial Divide p-Unit). Its properties are Horizontal and Vertical ratios. The two arrays of ratios define how the incoming signal is divided, individually for the X and Y axes, with each division becoming a clone of the original image. For example, setting the horizontal property to [1, 2, 3] will result in three horizontal copies of the source signal, arranged side by side, with the ratio of their relative sizes being 1:2:3.

p-Unit :: Spatial Position

New Spatial Position ✕

Horizontal Position: 0 20 40 60 80 100

Vertical Position: 0 20 40 60 80 100

Depth: 0 20 40 60 80 100

Horizontal Angle: 0 71 142 213 284 355

Vertical Angle: 0 71 142 213 284 355

Depth Angle: 0 71 142 213 284 355
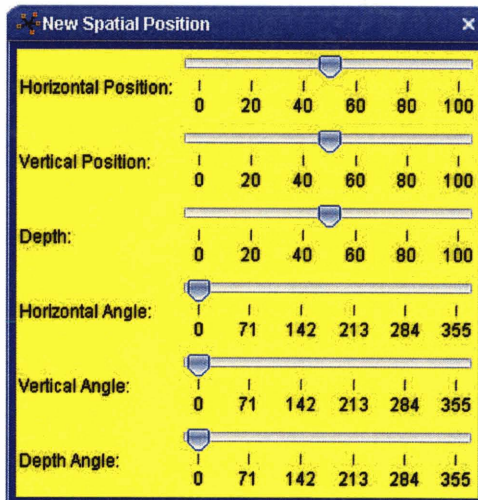
Spatial Position p-Unit defines the position and rotation of the image in 3D. The first 3 properties control the position of the associated image(s) along the 3 axes within the overall image; the latter 3 properties control the angle of rotation along each of the three axes.

p-Unit :: Visual Quad

Visual Quad [ID = -1053785141] Properties ✕

2D: X Top Left (%): 0 20 40 60 80 100

2D: Y Top Left (%): 0 20 40 60 80 100

2D: X Top Right (%): 0 20 40 60 80 100

2D: Y Top Right (%): 0 20 40 60 80 100

Visual Quad p-Unit controls the quadrature of the associated visual source(s). There are 16 properties, 8 for 2D and 8 for 3D, with each XY pair defining the transformation degree at one of 4 corners of the image.

50
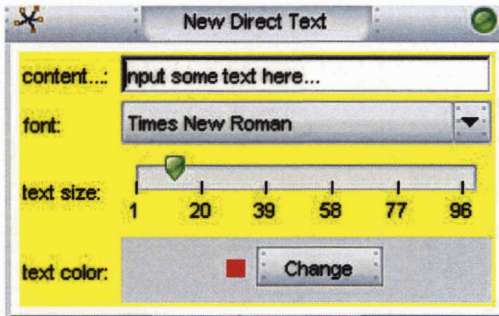
p-Unit :: Text sample

Text Sample p-Unit allows interactive input of text to be displayed in real time. Its properties are:

- Content, which defines the text to be displayed.

- Font Name, Size, and Color, defining the core properties of the output (unless overwritten by the Master Text Player).

p-Unit :: Text Related
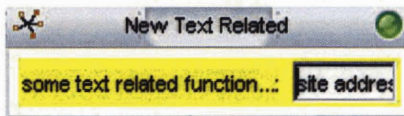
Text-Related p-Unit finds online text that is related to the text entered in the Text Sample p-Unit's input window. The resulting found text is displayed in the existing Master Text Player p-Unit's output window.

p-Unit :: Text Rhyme

Text Rhyme p-Unit finds online text that rhymes with the text entered in the Text Sample p-Unit's input window. The resulting found text is displayed in the existing Master Text Player p-Unit's output window.

p-Units :: Master Audio/Video/Text Players

One of the key improvisational principles is the continuous incorporation of ongoing contextual events into the improvisation. The Master Player p-Units express this concept by controlling the way the results of sound, video, and text manipulation are presented to the outside world. For instance, Master Audio Player can be used to spatialize the audio, useful in the context of a multi-participant improvisational action; Master Visual Player enables multi-display scenarios and layering of multiple videos. A new set of behaviors emerges as a result: one can control the spatial presence of audio, video, and text in particular areas, with the mutation of the relevant properties resulting in sound and video moving around the performance space.

p-Unit :: Master Visual Player
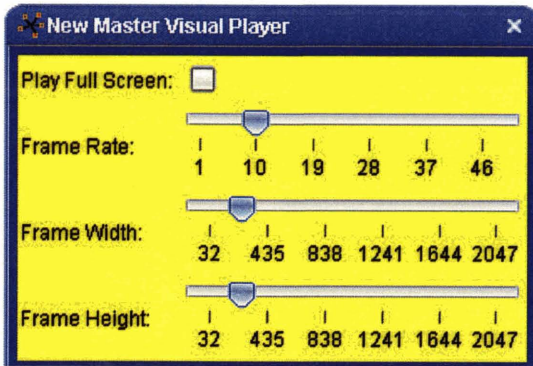
Master Visual Player p-Unit is responsible for the overall video playback, resulting from the sum of all the ongoing visual manipulations. Its properties are:

- Full Screen, defining whether to reserve one of the available screens for video playback;

- Window Width & Height, defining the size of the playback window, stretching or contracting the visual sources as necessary.

- Location, defining the video window position on screen at a given time.

p-Unit :: Master Audio Player

Master Audio Player p-Unit is responsible for the overall audio playback – a sum of all audio p-Units connected to it. Its only property is the Audio Channel # which defines the output channel of the audio card that will be used to play the resulting audio.

Master Text Player (can be laid over master visual player)
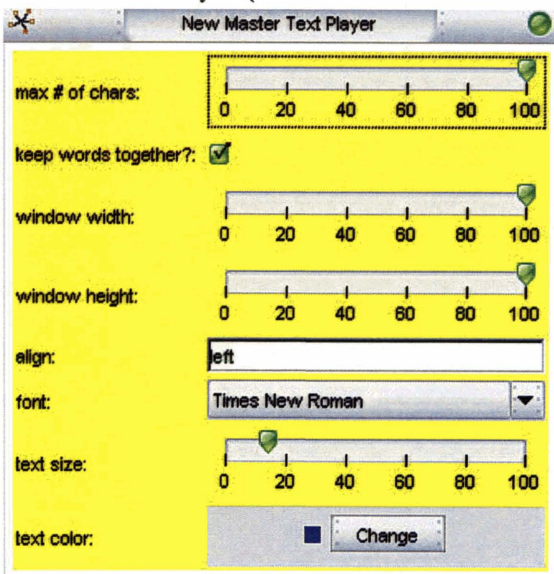
Master Text Player p-Unit is responsible for the overall text display, resulting from the sum of all the ongoing text manipulations. Its properties are:

- Maximum number of characters, which determine the size of the text buffer to appear.

- Words Together, which determines whether words wrap around.

- Text Window width and height, as well as the Align position.

- Font Name, Size, and Color, which, if set to non-default values, override custom settings made in other text p-Units.

52

Overview

The Structural layer is populated with *emons*, structural constructs entwined to create a network providing the participant with a higher-level abstracted instrument for observing and influencing the ongoing activity within the EE. The layer is modeled as a recurrent neural network, allowing evaluation of events' (stimuli) propagation and node cluster formation over time and concerned solely with the state of its elements (i.e., independent of any actual media processing that takes place as a result of its activity).

As a purely abstract system for controlling objects, it plays no role unless connected to the Perceptual layer. Each emon has its own activation level, its value continuously decaying. The emons communicate by sending stimuli, which can originate at any point in the network. When an emon is triggered, implying that its activation level is beyond a propagation threshold, it sends stimuli of proportional strength to all the connected emons. The growth and decay in the emons' activation levels can be controlled individually or en masse, by participants, other emons, MIDI controllers, sensors, or the ongoing evolutionary process.

Each emon can be seen to produce a repetitive beat; propagating through emon connections to become faster, stronger, or off-beat; propagating again through the Mediators which alter the processing of p-Units; and finally ending with a collage of temporal signals interpreted by effect producers, retrieval mechanisms, and audio/video/text players, with the result being output into one or more physical environments.

Figure 3:   Emon states. Emon's name on the top, activation level on the bottom.

Each emon has two visualized properties (fig. 2). The important one is the *activation level* which tells you how "active" a certain emon is (the second property – emon's name – can be set to whatever you want it to be). As mentioned above, emons communicate by sending stimuli to each other. When an emon receives a stimulus, its activation level goes up. Stimuli are sent to all the emons connected to the emon that is being stimulated, provided that the activation level of the emon in question is higher (or lower) than a given threshold (explained below). All operations can be performed on either a single emon or on a group of emons. An emon can be stimulated:

- Manually, by clicking on it and watching its activation level change.

- By stimuli: create a connection from emon *A* to emon *B*, then stimulate *A*, and see the activation propagate to *B*.

- By teaching it an activity pattern: see *Pattern Learner* section below.

- Periodically: by using the *Activity Agent*'s Automatic Activation feature which probabilistically triggers the least active emons.

- By outside source: by using input from sensors (serial / USB), MIDI controllers, and cell phones to stimulate an emon (or a group of emons).
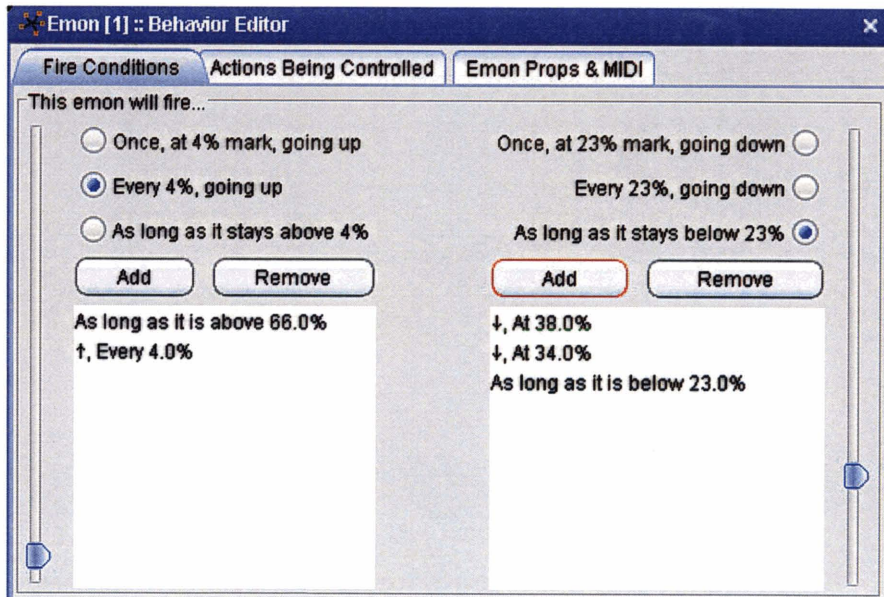
Figure 4:   Behavior Editor :: Firing Conditions menu.

unlimited number of associated thresholds.

Thresholds

You can set the thresholds upon crossing which a particular emon will fire, or send stimuli out (fig. 4). The panel shown allows you to add new and edit existing thresholds by moving sliders located on the panel's sides. The left half of the panel sets firing conditions for when emon's activation level is going up; similarly, the right half defines the threshold action for when the emon's activation level is decreasing.

Three types of thresholds exist: The "Once…" type is set so that firing will occur only at one specific activation level. The "Every…" type is set so that firing occurs every time the emon's activation level increases by a specified constant percentage. The "As Long…" type is set so that firing will occur as long as the emon's activation level is above (or below) the activation level of your choosing. Each emon can have an

54

Setting emon properties
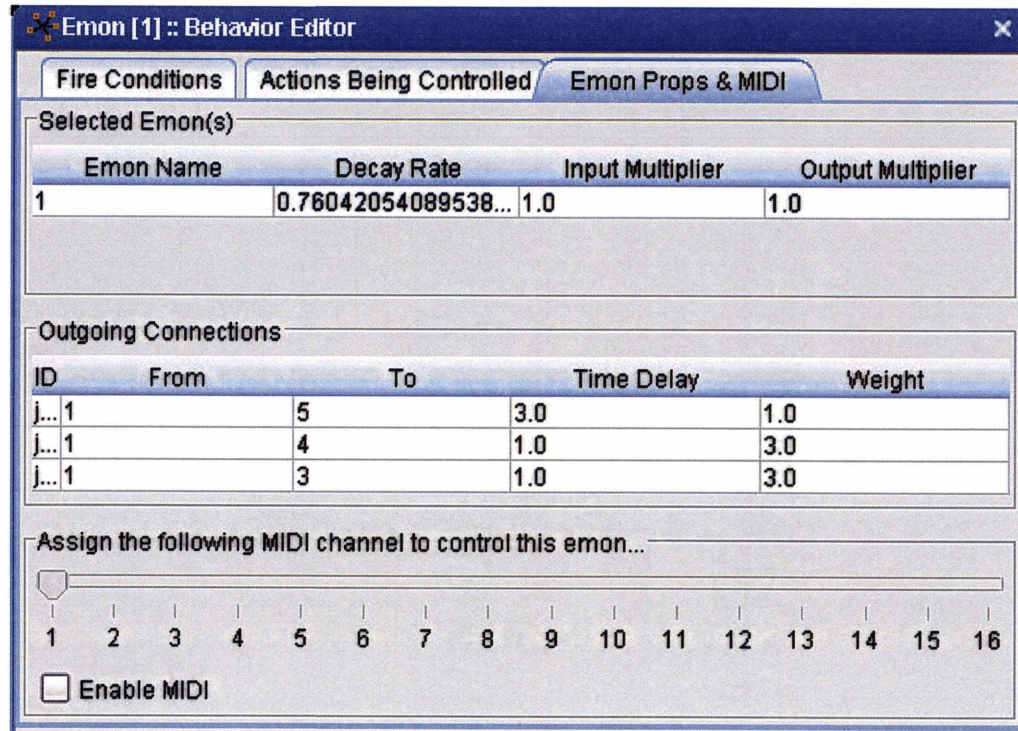
Figure 5: Behavior Editor :: Emon Properties menu.

Using the Behavior Editor Properties panel (fig. 5) it is possible to set the core properties of an emon such as the decay rate. The panel shown allows you to name the emon, see the names of emons it is connected to (if any) and define the emon's behavior when activated. The core emon properties are:

- *Decay Rate*, which defines the decrease per second of emon's activation level. Setting too high of a decay rate might result in a network with no continuous activity. Conversely, setting it too low might result in an oscillating network, where stimuli are generated faster than they are decayed.

- *Weight*, which is the strength of each stimulus released by that emon on a given connection.

- *Time Delay*, the time a stimulus takes to reach destination.

- *Input Multiplier*, which determines the ratio by which the weight of an incoming stimulus is multiplied before calculating the effect of the stimulus on the activation level of the receiving emon.

- *Output Multiplier*, which determines the ratio by which the weight of a stimulus is multiplied before that stimulus is fired.

Pattern Learner

Sometimes a repeatable actions' pattern – that is a temporal construct – might be desired. A simple example would be an image fading in and out or a sound panning to play from a different speaker every couple of seconds. Essentially, any repetitive, continuous action can be expressed as a pattern learnable by an EE's module called Pattern Learner. Each emon is capable of learning a unique pattern and executing it at any specified point in time. A given pattern defines the activation level of the emon over time. Given the emon's control of the perceptual characteristics, this in effect means that any type of modification can be made according to this pattern. Instead of specifying a line of opacity events, the opacity can be directly mapped to the emon's value, and an up-and-down pattern can be defined. As a result the opacity will be going up and down with the specified speed and pauses. You can also teach an emon to change its activation level at certain times. E.g., you could say, go up 40%, wait 2 seconds, go down to 10%, wait 3 seconds, etc. The pattern is executed in a loop and as a result, whenever the emon gets activated it fires.

Mediator

The Mediated layer maps the neural activity of the Structural layer onto the media processing activity of the Perceptual layer. The mapping is reconfigurable in real time and unifies the layers into a framework for improvisational action. Each emon can have an unlimited amount of thresholds; crossing a threshold triggers a Mediator, indicating that an action needs to be taken. Mediators are constructs that translate emon activity into data that can be processed by the p-Units. Each Mediator modifies a single p-Unit property; its flexibility comes from the multiple ways in which it can

Figure 6:   Mediator (Action) Designer.

map incoming emon activation data onto one of the properties of the p-Unit in question. Mediators, both individually and as bundles, are also tradable with other participants within the EE. Simply put, the Mediators insulate the Structural layer from the Perceptual one, allowing for decoupling of their processing power. This makes it possible, if desired, to replace or augment Perceptual layer with a different type of processing engine (e.g., using the Structural layer to drive Max/MSP patches).

Our method for navigating (evolving) media networks utilizes the Mediator to access the properties of components located at both Structural and Perceptual layers. The genetic algorithms that facilitate the navigation mutate the properties, biased by the participants' feedback and currently active constraints on the evolutionary process.

Evolutionary Process and Feedback

When improvising, we think about the structure – establishing connections between different events, evaluating what-if scenarios, observing, creating, and responding to focal points of action. In other words, the media landscape can be seen as a network of interconnected elements, with improvising being the task of changing the structural characteristics of that space. It is therefore clear that evolution of digital media space by a machine should equally focus on the structural layer of the network – the layer that guides when and upon which conditions parameters and interconnections of the structural elements are modified, affecting the underlying media content. The machine needs to be able to:

- Understand the state of structural elements at any given time and within the changing context, be it the core structural elements (emons), their interconnections, and general network dynamics (e.g., formation of clusters).

- Learn in real-time – there is no use of learned data if it cannot be applied at the moment it is needed. Pre-training the system before actually playing with it makes little sense, since the desired actions are unique to each improvisational session.

- Be capable of changing the elements to express the learning it has done. A self-evident point that nevertheless merits mentioning is that learning, in a situation of co-improvisation between the human and the machine necessitates that the machine has a way to show to the human what its actions are rather than simply performing these in the background.

- Provide the participants with the possibility to constrain the evolutionary action to a given subset of elements. This, however, should not be expected or demanded.

By incorporating GAs, the Emonic Environment is able to:

- Learn what kinds of mutations are popular with its human partners.

- Step away from the black-box algorithmic approach to system operation, showing the internals of GA's action, and making the participants realize how their input affects the ongoing evolution.

- Enter a dialogue by reacting to the provided feedback and giving human participants a chance to control the evolution's intensity and scope.

EE's genetic algorithms operate in the following fashion: at each mutation cycle (a period of which is defined by the user), the GA picks an emon to mutate. The emon is picked according to its score. The score is increased by 1 each time an emon is hit by a stimulus. The probability of picking a given emon is $1/S$, where S is the score (the value is normalized to 1). In other words, the probability of picking a given emon is inversely proportional to that emon's score – the more active an emon is, the less is the likelihood it will be picked as a subject for evolution. The frequency of the evolutionary cycle is defined according to the top slider in fig. 7. The rationale for the way in which we choose the emons to be activated is threefold:

- It provides a clear metric observable by the human, between the amount of explicit human input and the machine taking action.

- It is nondeterministic yet not arbitrary.

- It seems to be appropriate as a method for instituting intelligent automatic activity – improvisation is a dynamic process where people contribute as little or as much as they desire at any point in time. Mutating the less active emons allows the machine to take over the parts the human appears to have neglected. This picking method also helps in making the evolutionary process appear as having a clear connection to the human input and therefore more appealing when the human provides feedback to the machine.

Once an emon has been picked, the GA has to decide which mutation to pick. First we randomly pick one of the categories below and then, a mutation choice is made within the selected category based on principle similar to that described for choosing emon. That is, it is more likely that the properties mutated will be the ones user interacted with the least. The available mutations are:

1. Emon-centric (internal properties of individual emons, as explained in the *Structural Layer* section):

   o Decay rate

   o Input & Output multipliers

2. Network-centric – properties of edges (the connections between the emons) that ultimately define how dense the events within the network are:

   o Time delay

   o Connection weight

3. Mediator-centric (controlling how many Mediators a given emon controls as well as the parameters of each Mediator).

4. Threshold-centric – those are the conditions upon which the associated Mediator(s) will be triggered.

   o Described in the Thresholds section above.

Evolution and the human participant differ in their capabilities. There are two ways in which the human is 'stronger' than the machine:

- Participants can add and delete emons, unlike the machine which is limited to operating on the already existent emons.

- It is up to the participants to control the rate of the evolution (or disable it altogether).

It is possible to evolve a Mediator with variable strength. Four such strength (or constraint) levels are available:

- Unconstrained. Any parameters, controlled p-Unit types, etc are fair game for the mutation.

- Somewhat constrained: you get to choose the p-Unit type to be evolved, but the particular property to be controlled is up to the machine

- Constrained: You pick both the type and the property to be evolved, but the type of change is up to the machine.

- Tightly constrained: You get to pick the type, the property, and the type of change and the machine evolves the values and the boundaries.

The resulting Mediators, just like the manually designed ones, can always be edited, evolved, and traded in the Emonic network.

The constraints are probabilistic, that is, they increase or decrease the probabilities of possible mutations. Each action within the EE (e.g., 'increase-tempo-of-audio-element-1') is formalized as a possible mutation. The likelihood (probability P) of that mutation being executed depends on the context – i.e., P=1 in the case where the change is initiated explicitly by a participant, and 0<P<1 for an action initiated in any other context. For every case where 0<P<1, it is determined as a weighted sum of the constraint values set in real time by the participant and the computed distance to the target system state (called *magnet* and described in the next section). Evolutionary constraints fit well the notion of improvisational action – placing the constraints results in dynamic adaptation of system's behavior thus leading to generation of potentially relevant media paths.

In the EE, two evolution possibilities are available. The first bases its fitness function solely on the feedback received from the participants in real time. The second utilizes a pre-saved state of the network and progresses by continuously comparing the current state with that preset. As a result, to determine the direction of the evolutionary process, the participant may choose to (1) place constraints (described in the previous section) and/or (2) place a convergent attractor (or detractor) that defines the overall target state (multiple simultaneous attractors are also possible and are described below). We summarily refer to these attractors and detractors as *magnets* (described later in this chapter).

Activity & Evolution Agent

Figure 7:   Activity & Evolution Agent.
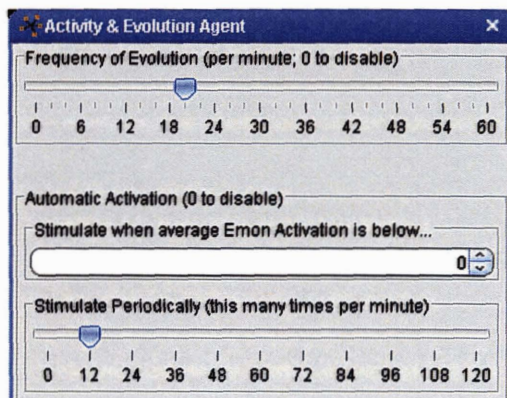
Evolution & Activity Agent provides a human-controlled interface to the ongoing evolutionary process. This agent allows the user to specify the frequency of the evolution (with a single emon picked for mutation at each cycle) and the periodicity of automatic activation (with the system able to automatically activate one or more emons either periodically or when the overall activity in the structural network is below a certain level).

Magnets

Magnets are a high-level instrument for exploring the DMS. As such, they necessitate having a method for generating representation for any arbitrary state of the DMS, in real time. That representation must contain essential information about a given DMS, so that, given two DMS states' representations, the magnet may determine the distance between the two.

In order to move from DMS A to DMS B, each individual is mutated (while weighing in the previously placed evolutionary constraints). As the magnet operates, it learns which types of mutations are preferred for reaching a particular state of the DMS and combines the mutations that lead closer to that state.

In order to learn which mutations are better than others for a given movement, and to make sure those are more likely to occur, a method has been designed to evaluate the distance in parametric space between DMS A (current state) and DMS B (target state). We do this by determining the fitness of the current state, and then proceeding to evolve randomly by a distance of one mutation at a time. After each mutation, an evaluation occurs to determine if the current state is closer to the magnet. Such trial and error makes the environment learn which mutations are more desirable. The 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 numerically, as a checksum. The checksum is reflective of the individuals present in the current state and is calculated uniformly. By weighting each individual on a different scale, the derived checksum attempts to more closely represent the particularities and importance of certain individuals in a given state, taking into account not only the basic weights of each property but also the constraints introduced by the participant (as described in the section above). As a result of this ongoing mix between the learned weights and the participant-set constraints, the combined weights shape the overall evolution to continuously try and balance the participant's desires and the direction suggested by a magnet.

Multiple Magnets

When improvising, we rarely have a single state in mind that we would like to reach. No real life example exists where one influence single-handedly defines the evolution of an object or an environment. In the context of DMS, we wanted to have the possibility of converging to a set of states. This becomes possible if we evolve from the current DMS state in multiple directions at the same time. While placing a single magnet in the EE mutates the

current DMS's individuals closer to those of the target DMS, placing multiple magnets should have a similar effect, but with an inherent competition between the magnets. To do so, we employ an evolutionary process that takes into account the simple fact that a given mutation is likely to bring the environment closer to one magnet while farther away from another. We start by placing >1 magnets, each assigned equal initial weights. After each mutation, the evolutionary process evaluates the distance between the mutation performed and the DMS targeted by the magnet with the currently highest (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. The concept of multiple magnets reflects the nature of improvisation – a process of continuous flux.

Input & Output Methods

Being limited to keyboard and mouse for interactive purposes is limiting. In order to enable a more flexible exploration of the media space, the EE integrates tools for sampling media as well as interpreting and being guided by control information from a number of external interfaces. In particular, real-time media acquisition includes sampling audio from connected microphones, and sampling audio from directly connected or remote webcams. Output capabilities include spatialization of sound to multiple speakers, visualization to one more projection screens, and control of RGB lighting systems such as ColorKinetics RGB grid (ColorKinetics, 2006). Control capabilities include MIDI controllers, virtually any sensor with serial port connectivity, OSC network control (Wright et al, 1997) and both sampling of media and interface control utilizing Nokia smart phones.

Applications

The enjoyment of browsing a personal collection of images, videos, and sounds is hampered by the hunt and peck approach to browsing that currently pervades most computer storage and browsing systems (folder by folder, title by title, author by author, day by day, style by style). Most people are comfortable browsing their MP3 and image collections, but they do not extract as much pleasure out of browsing as they might under a system where the activity of browsing becomes a creative act, instantly turning them into co-authors rather than merely followers of material created by others.

As personal media collections grow larger, it is less and less likely that the hunt and peck mode of browsing will result in a pleasant journey through a personal collection – while it might result in an efficient retrieval, it does not serve up serendipity or satisfy the desire to structure meaning. The way information is stored and the way it is explored are still far too closely related – it is possible to arrange materials by date, name, etc., but the browsing environment does not usually generate "paths" for exploring the collection.

Generating such paths – or in other words, actively applying structure to a collection – is often thought to be an exciting, "creative" experience; activities such as editing, composing (a montage or a piece of music), and improvising all provide different approaches to finding meaning through structural manipulation. Of these only improvisation merges the activity of creating and performing into a single real-time action.

In the following section we describe the EE's ability to facilitate improvisation by browsing – generating real-time performances in tandem with the machine by browsing audiovisual media collections. We contrast the ways of achieving the described tasks using the EE vs. doing the same with a prototypical media editor. We argue that in the scenarios presented, the EE requires less low-level effort, while enabling the kind of structural control of media that would not be possible with traditional tools.

Personal Media Collection Explorer

Anyone can imagine a setting in which they might want to create or browse digital media. It could be a party in your house, where you would like to play the role of a video DJ, an after-dinner activity where you and your friends browse media together, or a personal browsing experience where you play with photos, videos, and sounds you have accumulated on your computer. At some point in this process you might feel like contributing more to

the collection of samples – in that case you might opt for using a cell phone, video camera, a microphone, or sensors and MIDI controllers as your sampling and controlling instrument of choice.

We describe a few such sample scenarios, where unique audiovisual experiences are created by browsing in a structural, networked manner[4]. In these scenarios, one gets to create a performance and get the feeling of authorship, while benefiting from serendipity associated with browsing – all by building and controlling an Emonic network.

Solo Browsing and Performance

Michael, an artist, wants to browse through the collection of images and sounds he accumulated on his computer. He doesn't care much for any end result in particular, but wants the experience of browsing to be fun and unlike the 100s of previous searches and browsing experiences he had. He decides that he would like to explore his media using the following exploratory structure: each time a new image is loaded, it will rotate in an oscillating pattern and the accompanying sound will also switch. To do that he adds three emons, assigning emon #1 to randomly iterate through the images on his hard drive, emon #2 to manipulate rotation of the available image, and emon #3 to switch the playback of a sound. He then teaches the emon that controls the rotation of the video a pattern according to which it is to rotate, and connects the three emons.

After watching a few such iterations, Michael gets bored. He wants something new, but isn't sure what it is supposed to be. He turns off the pattern he taught and sets to explore the structure of the network manually. He adds a few more emons, assigns them random Mediators and tries stimulating the network, trying to figure out emons' actions in a game-like manner. He then finds one emon that performs transformations that are clearly not to his taste. He tells the system to evolve the emon, and continues to play with the network, exploring the behaviors of its various components.

At some point, he runs out of ideas for what to do. That's when he turns on the evolution. He sits back, only stimulating the emons when he sees or hears something he likes, and otherwise just watching the way events unfold. He starts with a very fast rate of the evolution, in an attempt to introduce a lot of new ideas into the network. After a couple of minutes, he sees that it has evolved to do some pretty interesting things. He slows the

---

[4] We presume that users of the EE have access to a collection of media (either on their hard drives or online).

rate of the evolutionary process, and begins yet again to take a more proactive role in the experience, stimulating the emons, controlling properties of the p-Units, and contributing new media.

Shared Browsing and Performance

First shared scenario takes place in an airport hall. Shaheri is a schoolgirl who travels a lot. A significant portion of her time is spent in the airports, waiting for the departing plane. Her only companion on the trips is her cell phone. She loves taking blurry pictures of funny strangers and signs and sending these to her friends around the globe. She wishes those around her could see her pictures too, but, being shy, she never shares her found-moments with strangers.

Utilizing the cell phone component of the Emonic Environment, she logs into the airport-based EE, displayed on a large screen in the waiting hall. She is able to contribute her pictures and videos and manipulate those contributed by others, all while maintaining her anonymity yet seeing that others see and interact with the pictures she provides.

The second scenario takes place in Laura's living room. Using the TViews interaction table (Mazalek, 2005), a surface that can be navigated by simultaneous movement of an unlimited number of tangible pucks, she and her friends are able to improvise their media together. Each one of them picks a puck assigned to a certain kind of structural manipulation in the EE. They play with the pucks, changing their position, functionality, and ownership (passing those to each other), engaging in a process that closely approximates that of an artistic improvisational performance.

Experiments and Evaluation

Like any creative activity, improvising is a learning process. Unlike other, more structured activities, one can start improvising and get immediate results without learning the rules of what's expected and what's appropriate. At the same time, in improvisation, just like in other creative activities, it is possible to get better with time. Getting better at improvising means to progress to seeing structure where non-improvisers see only individual sounds or images.

How do we know if a person is thinking improvisationally? The studies presented below analyze the question of whether people who used the Emonic Environment did exhibit an improvisational behavior as defined in *Cognitive Model of Improvisational Action*. Two studies utilizing the EE are described. The first was conducted in a public venue (at Centre Pompidou in Paris) and involved exhibition goers over the course of 3 months; the second was conducted in a controlled setting (at the MIT Media Lab) over a period of a month, and included musicians, artists, engineers, and amateurs.

Early Feedback

A couple of years back, when the EE was at its initial development stage, we invited two people to come and use the system. Both subjects were given ~10-minute long introductions and were asked to create networks. This was their feedback:

Participant A: "When I design my network and I want things to play together, I'll connect them [the emons] closely together; when I don't, I don't connect them closely [in space]. No, I didn't think about sequence and how things are one after another, I thought about time when they co-occur and so if I didn't want them [to play] together I'd put them in different clusters. Many times the system is stubborn and it wants to play the same thing over and over, so I fight it so it doesn't do the same thing. [It] is especially in this context that I think about connections and not the elements it connects because I am concerned with breaking its stubbornness, its loops. There is a trade-off between me controlling it and the system surprising me."

Participant B: "Your system reminded me of the tapes I used to make when I was little, making music by gluing parts of different tapes together. It's kind of like that only with lots of things going on, so I have to tell them "shut up!" all the time, they go out of control sometimes but sometimes they don't want to wake up [talking about the self-activation of emons], it confused the hell out of me in the beginning but it was fun I didn't know

what I was doing, it really didn't let me plan. Yes I would consider what I made music [the tester is a professional (idiomatic) jazz musician], it has tension and it made me think differently. Actually no, I take that [first] one back it doesn't always have tension, the beginnings and ends are bad, if I don't all the time do stuff with it, it doesn't do anything interesting [system didn't have proactive evolution at this point]. Yes I would spend more time with it; I would really want to have more [media] in the system to begin with."

Interestingly enough, the first user (a non-musician) ended up deleting the proposed default network of 15 emons and created one all her own with >60 nodes by recording various combinations of voice, guitar, and captured radio broadcasts. The second user (a professional musician) did find the recording feature neat, but only recorded a few rhythms with it, preferring to look for more materials among the ones already found in the media directory (the ability to stream media from the net was not yet implemented). Both seem to have enjoyed playing with the system and liked the initial feeling of unpredictability of the system, even though the musician had a much harder time adapting to the idea of 'creative chaos'.

These mini-experiments were our first real indication that people adapt fairly easily to the idea of sequence as an unplanned music creation with no well-defined goals. They also showed the premise of a proactive audience – the two participants were never told to build new networks, they just did, with primarily custom and ready-made elements respectively. To sum it up – these informal results showed us that using the EE might let us learn aspects of people's perception of improvisational process that would be hard to discover otherwise.

Study I: Centre Pompidou

The Emonic Environment has been presented as part of the 'Non-Standard Architectures' exhibition that took place at Centre Pompidou in Paris between December 2003 and March 2004 (Nemirovsky et al, 2003). This exhibition presented an interesting challenge for creating an improvisational experience in public space: exhibition space goers, with very limited patience and likely only a few minutes to spare per exhibit, yet at the same time clearly interested in observing a creative statement and thus potentially open to making a creative statement of their own. Taking part in the exhibition also meant proving that the Emonic is capable of transforming a public-space into a performance, and the flow of visitors into performers.

An additional challenge was presented by the exhibition's focus: exploration of new forms of architecture. In order to avoid distracting from the architecture presented, we wanted to create a setup where no obvious computer elements, such as keyboards, displays, and alike would be present. To enable this transparent experience, we chose audio and lights as our outputs, both controlled purely by environment-driven sensors with no visible UI.

The inputs of the system consisted of four directed microphones and a range of sensors, interpreting motion, pressure, and sound levels. The output consisted of eight speakers distributed around our area of the exhibit, as well as a bed of LED lights integrated within the Thermocline, an architectural component designed by SERVO (SERVO, 2002).

The interaction with the installation was envisaged as follows: visitors would come in, hear some sound coming from some of the speakers, pass by or sit on the Thermocline (a plastic couch with lights in fig. 7), speak and hear their voices reflected and twisted in the multiple speakers and see change in lighting. The evolution, driven by peoples movements and actions on and around the couch, would control the nature and spatial placement of sound, as well as intensity and shape of the emanating light.
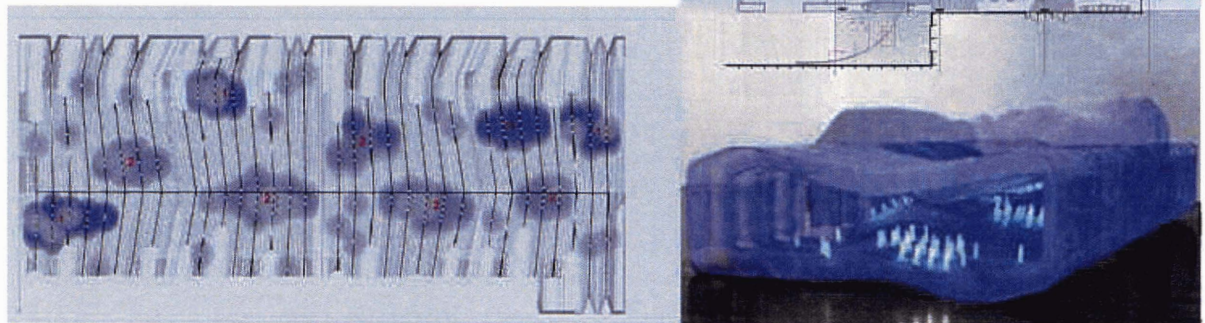
Figure 8: Thermocline with lights turned on.

The actual behavior of the visitors fell into three distinct categories:

1. Didn't care at all, passing right through. In our observation about 20% of visitors fell into that category.

2. Spent a longer time looking at the display, and listened to the sounds but made no apparent connection between their actions and those of the installation. That accounted for about 40% of the visitors.

3. Recognized the fact they could control what is going on. People in this group kept around for awhile, exploring the mappings, moving back and forth on and off the Thermocline, observing the effect they have on the sound and the lights, and then trying some more. About 40% of visitors belonged to that group, with the observed length of interaction ranging from a couple of minutes to over an hour.

The reactions to the EE were promising. At the same time, it was clear that a more formal understanding of people's behavior and interaction was needed to fully explore EE's potential. With this in mind, we embarked on a third iteration of our system, one described in this thesis.

Study II: Media Lab

Objective

As we reflected on the exhibition at Centre Pompidou, it was clear that a more controlled study was needed to evaluate the validity of our improvisational approach. In designing the next study we were interested to answer the following questions:

-   When given a choice, are there patterns in the way people choose to shift between structural and direct control of their media? Do these patterns support the assertions made in CMIA, or does the direct control remain the preferred way of interaction with media?

-   If movement between the two control modes is observed, what is its path over time?

-   Do the actions taken by people using the EE fall in line with the improvisational behavior characteristics formalized in the Cognitive Model of Improvisational Action? In other words, do people appear to act improvisationally?

To answer these questions, a new version of the system has been designed (the one described in this thesis), including a data acquisition module (described below). The quantitative data obtained in the course of the study was mapped to the characteristics of CMIA, in an attempt to observe whether the patterns in users' manipulation of both structures and lower-level perceptual elements of the Emonic Environment resemble those outlined in our model.

Data Acquisition

People who use the EE may do so solo or in group; online or offline; using mouse & keyboard, or MIDI controllers, cell phones, and sensors to control their input. They can record new sounds or videos, sample from DVDs and MP3s, or use net-based sources. They may focus on controlling every little detail of each active media sample, or turn the evolution and automatic activation on and sit back and watch.

Given the diversity of these usage scenarios, we felt that questionnaires would not be the appropriate method for acquiring knowledge on the precise manner in which people operate the EE. A way to evaluate the data statistically seemed necessary, in order to discriminate usage patterns and use those as a validation of the CMIA.

To do so, the EE has been supplemented with a logging module, with each action made by a participant or the system itself being recorded. The format of the logging data is as follows:

- First line of each log file specifies the User and Session IDs. Upon starting the EE for the first time on a given machine, each participant was assigned a random User ID (generated when the system is run for the first time), unique to this particular copy of the EE, and kept throughout the subsequent uses in order to avoid multiple submissions by the same participant under a different ID. Session ID on the other hand was randomly generated at the beginning of each session. In other words, it is possible to say that two or more sessions came from the same participant. Privacy note: User ID, randomly generated upon the first run of the EE on a given machine, was the only information recorded on an individual participant, with no possibility of tracking the true participant identity. No personally identifiable data of any type was gathered.

- First element on each line is a timestamp with milliseconds precision, with zero being the start of a given session.

- Second element on each line is the Event Source. Each recorded action was catalogued as being initiated by either a human (User) or the machine (System). For instance, any manual operations by the human (e.g., adding a new emon) were marked as User; any activities that were result of the GAs' operation were tagged as System.

- The rest of the elements were:

    o Input Interface: The interface used as the input control – keyboard, mouse, MIDI controller, cell phone, etc.

o Event Component Source: the UI component where the event originated – a menu, popup, etc.

o Object Type: the type of the object that was changed in some way. Types of objects currently available within the EE include Emons, P-Units, and Connections.

o Object SubType: The subclass of the object type – e.g., for p-Unit it could be an AudioSamplePUnit, for a Connection it could be a StructuralConnection, etc.

o Object ID: Each object has a unique ID, generated as a random long at object's creation, and kept constant throughout.

o Object Property: Property (of an object) that is being changed – e.g., DecayRate for an Emon or SampleRate for VisualSamplePUnit.

o Object State: Reflecting major changes in object state, such as adding/deleting/disabling/etc.

o Initial/Final Values: If the object changes from one value to another (Object Property changes in particular), we log the original value as the Initial value and the new value as the Final value.

o System Function: Name of the system function that is being (de)activated. System Functions are any elements of the system other than those defined as Object Types, e.g., AudioTracker, MUPEControl, etc.

o System Function State: Similar to Object State, reflects the state of the function in question such as shown / hidden / activated / etc.

o Change Description: Any change that doesn't fall into one of the categories above is logged as a Change Description. This is largely reserved for features that are not used in the statistical analysis and are logged for informative purposes only.

At the completion of each session (i.e. when the program was shut down), the resulting log was sent to a server located in the Media Lab, where it was made available for analysis (described in the next section).

12632: User; INPUT_INTERFACE: Mouse; OBJ_TYPE: Emon; OBJ_ID: 1346235236; OBJ_PROP: ActivationLevel; OBJ_STATE: Quieted

14327: System; OBJ_TYPE: PUnit; OBJ_SUBTYPE: Visual_Sample; OBJ_ID 426236262; OBJ_PROP: RedScale; INIT_VALUE: 25.0; FINAL_VALUE: 20.0

22197: User; INPUT_INTERFACE: Mouse; OBJ_TYPE: Connection; OBJ_SUBTYPE: PerceptualConnection; OBJ_ID: [426236262, 2980823537]; OBJ_STATE: Added

---

Figure 9: A log data sample.

71

Design & Execution

Twelve subjects took part in individual controlled studies conducted at the MIT Media Lab. Each study began with a 10 minute long tutorial on the use of the EE, followed by a 10 minute long hands-on training. At the end of the training, the actual study was started and lasted for another 20+ minutes[5]. At the conclusion of the study, each participant filled a survey questionnaire (see the end of this section). Each participant took part in a single study.

The participants belonged to 3 groups: participants 3, 9, and 12 were musicians; 1, 5, 6, 7, 8, and 11 were visual artists; 2, 4, and 10 were computer scientists. The age ranged between 16 and 36. Their skills' self evaluation was as following:

General Proficiency (1-10 scale)

User ID:	1	2	3	4	5	6	7	8	9	10	11	12
Music-making	3	3	10	3	3	5	1	5	10	2	1	8
Music literacy	7	2	8	4	3	6	8	4	10	1	4	6
Video-making	6	0	2	4	2	9	9	6	0	5	8	5
Arts/Cinema literacy	8	3	2	5	8	9	5	7	3	5	5	6
General comp. skills	7	6	10	7	6	7	7	7	9	8	3	8

Figure 10: Subjects' proficiency table.

The study began with an empty network. The participants were free to use a core set of media provided by us (30 audio samples, 20 video samples, and 20 image stills) as well as, if they desired, capture new audio from the attached microphone. It was up to the subject to decide whether or not to use evolution, when to turn it on, for how long, or what the frequency of evolutionary changes should be (if at all).

---

[5] Some seemed to be too engaged to stop them immediately as the time expired, therefore sessions length ranges between 21 and 30 minutes.

Asking for assistance was allowed in case the system stopped responding or crashed or if there was a very particular technical question (e.g. 'what is the shortcut to add a new emon'). The system remained stable throughout all studies; 2 requests for technical assistance were made per session on average.

Analysis

In order to find out whether there is a correlation between subjects' behavior and CMIA's characteristics, we took a set of quantitative measurements in regards to the participants' actions within the system. We then analyzed the data gathered with respect to these measurements. Our objective was to analyze their actions while detecting patterns of interaction between the human participant and the machine. Each session consists of thousands of actions, performed by the human and the machine. For the needs of the analysis, each session was divided into intervals. In the graphs that follow, the length of an interval within each session is ten seconds. Why this particular length? It is obvious that the graphs would look different were we to pick a different time scale for evaluation and segmentation of the study's results. On one extreme, were we to pick a too-large of a segment as our unit of measurement, we would end up with each time interval containing all the possible actions – a subject is likely to employ all of the system's functionality given a long enough period. Such long periods would hide the stages and the patterns of use the subject was going through (getting comfortable with using the tool, exploring structure vs. directly controlling, etc.) On the other hand, picking a time interval that is too short would be just as meaningless – each time interval would contain just a single event, with no ability to evaluate clusters of decisions. Ten seconds interval length allowed the right balance – short enough precision-wise, yet long enough to observe patterns of use.

Our analysis involved six questions, five of which drew their information from the graphs, and one from a previous study. Each graph represents data from a single session. In what follows we show that when the participants use the EE, their behaviors are in line with the assertions made in *Cognitive Model of Improvisational Action*.

1. Focus: Multilayered.

One of the characteristics of improvisation as defined in CMIA is that when improvising, we alternate between manipulating structure and content, with our overall focus heavily biased toward manipulating the structure (unlike the process of composition, comprised largely of manipulating the content).

In analyzing the data from the subject studies, we wanted to see whether (1) both structure and content manipulation occurred and (2) the interaction had improvisational (structure-centric) or compositional (content-centric) character. To find out, we looked at the ratio of structural (emons + interconnections) and content (p-Units + interconnections) manipulations. The manipulations presented are initiated by the participant either directly (explicitly interacting with the EE's components) or indirectly (by activating the evolutionary process).

First, we notice that for all subjects structural manipulations are predominant. Second, we can distinguish between 3 response pattern groups (figs. 11-13). Groups *A* & *B*, which account for ¾ of the study participants, exhibit both structural and content manipulation. Content manipulation largely occurs in bursts and groups *A* & *B* (with the exception of two sessions) exhibit bursts of content manipulation separated by periods of longer periods of structural manipulation. The only difference between groups *A* & *B* is in the initial focus – group *A* starts off with content manipulation, progressing to structure, while group *B* presents an inverse pattern. Finally, group *C* has very little content manipulation, focusing on structural interaction.
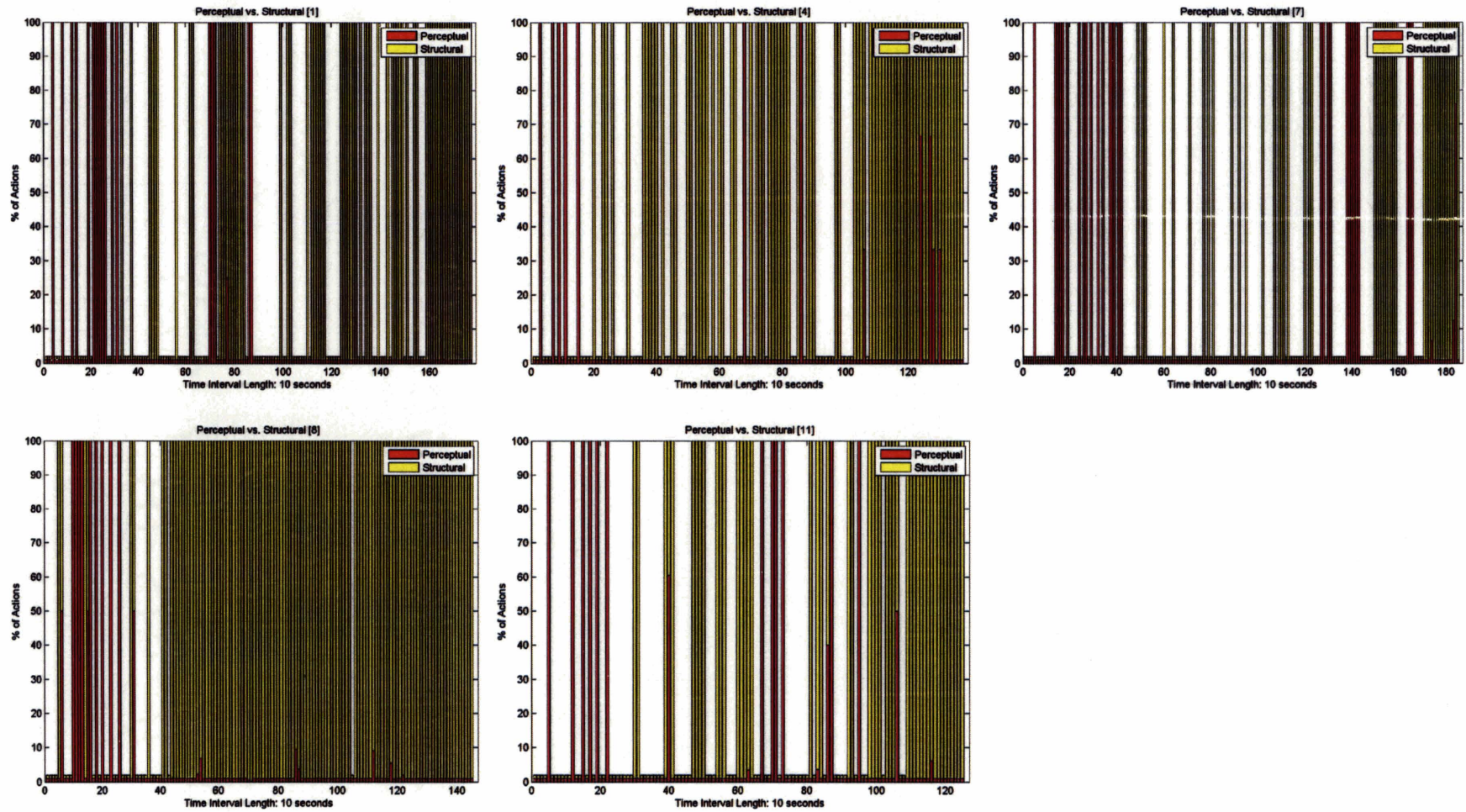
Figure 11: Ratio of Structural (yellow) vs. Perceptual (red) actions. Group A: content-centric start, followed by structural manipulation.

75
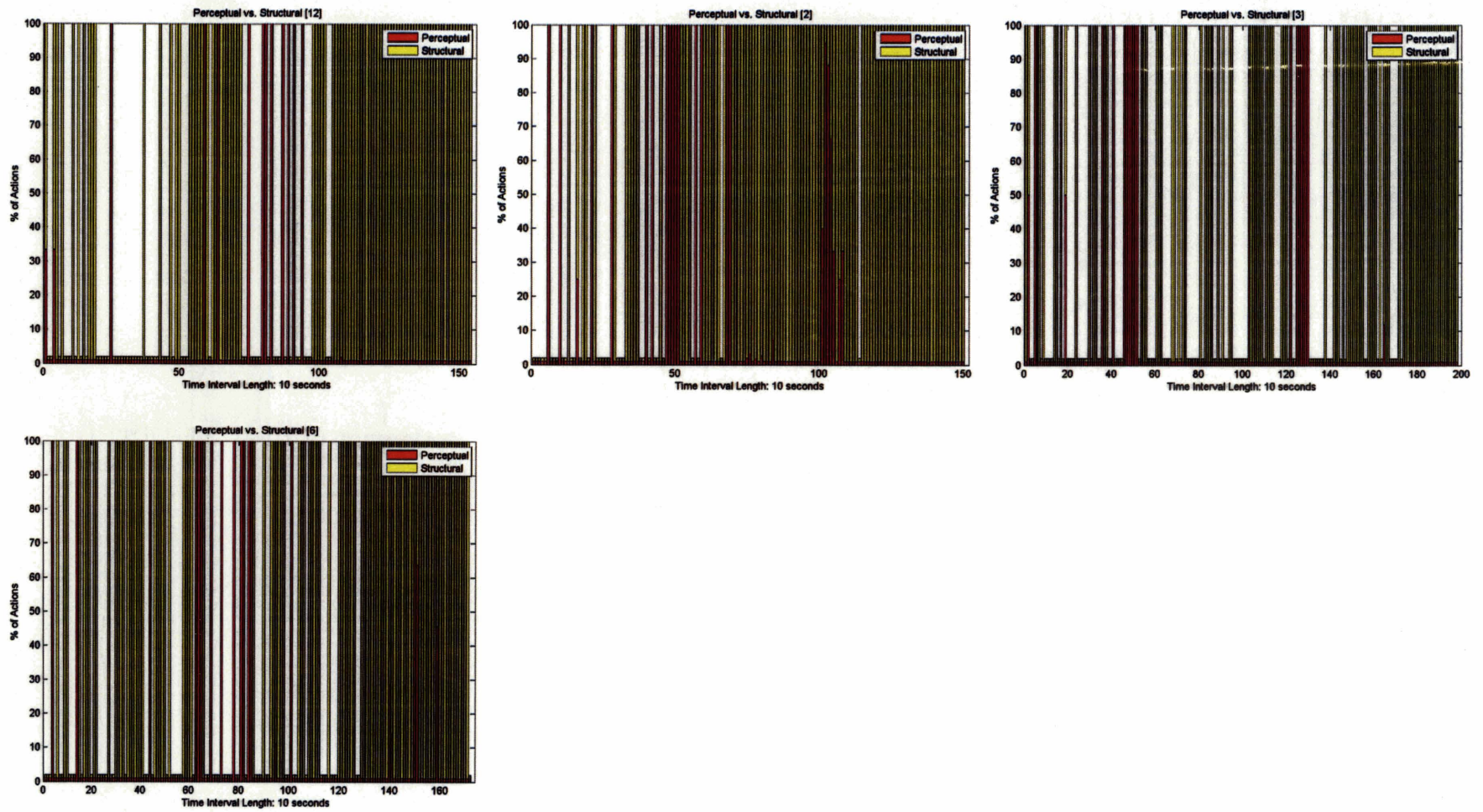
Figure 12: Ratio of Structural (yellow) vs. Perceptual (red) actions. Group B: structure-centric start, switch of focus to content and back.
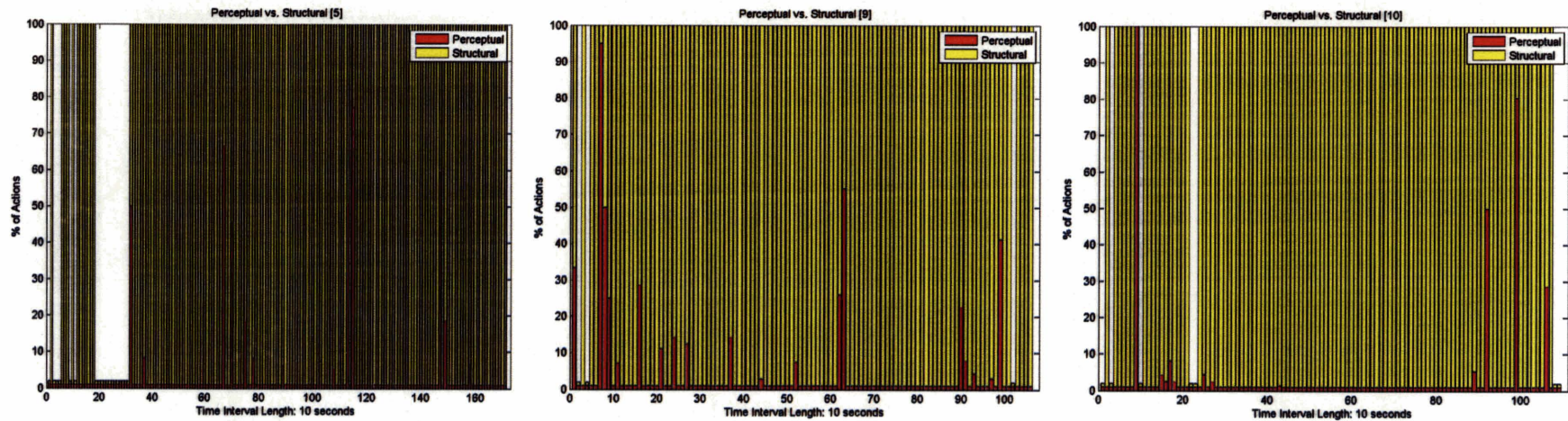
Figure 13: Ratio of Structural (yellow) vs. Perceptual (red) actions. Group C: Prevalence of structural manipulations, concurrent with content ones.

In order to further understand the character of structural activity in the sessions, we graphed (figs. 14-15) the absolute cumulative emon activation level over time. These graphs, unlike the ones above, include only explicit stimuli-generating actions by the participants of our study. It also allows us to see how much and when the abstract interconnectivity feature of the structural layer has been directly manipulated (emons that are connected to each other forward stimuli, resulting in higher cumulative activation level).

Do people directly manipulate the structural interconnectivity of the network? The answer varies – as can be seen in the graphs below, there is a spectrum of interaction with the structural layer. Two groups emerge: those that have only utilized the structural function a few times, and those who have used it throughout the session. For the latter group, structural activity appears as an integral part of the interaction cycle, occurring periodically throughout the session.
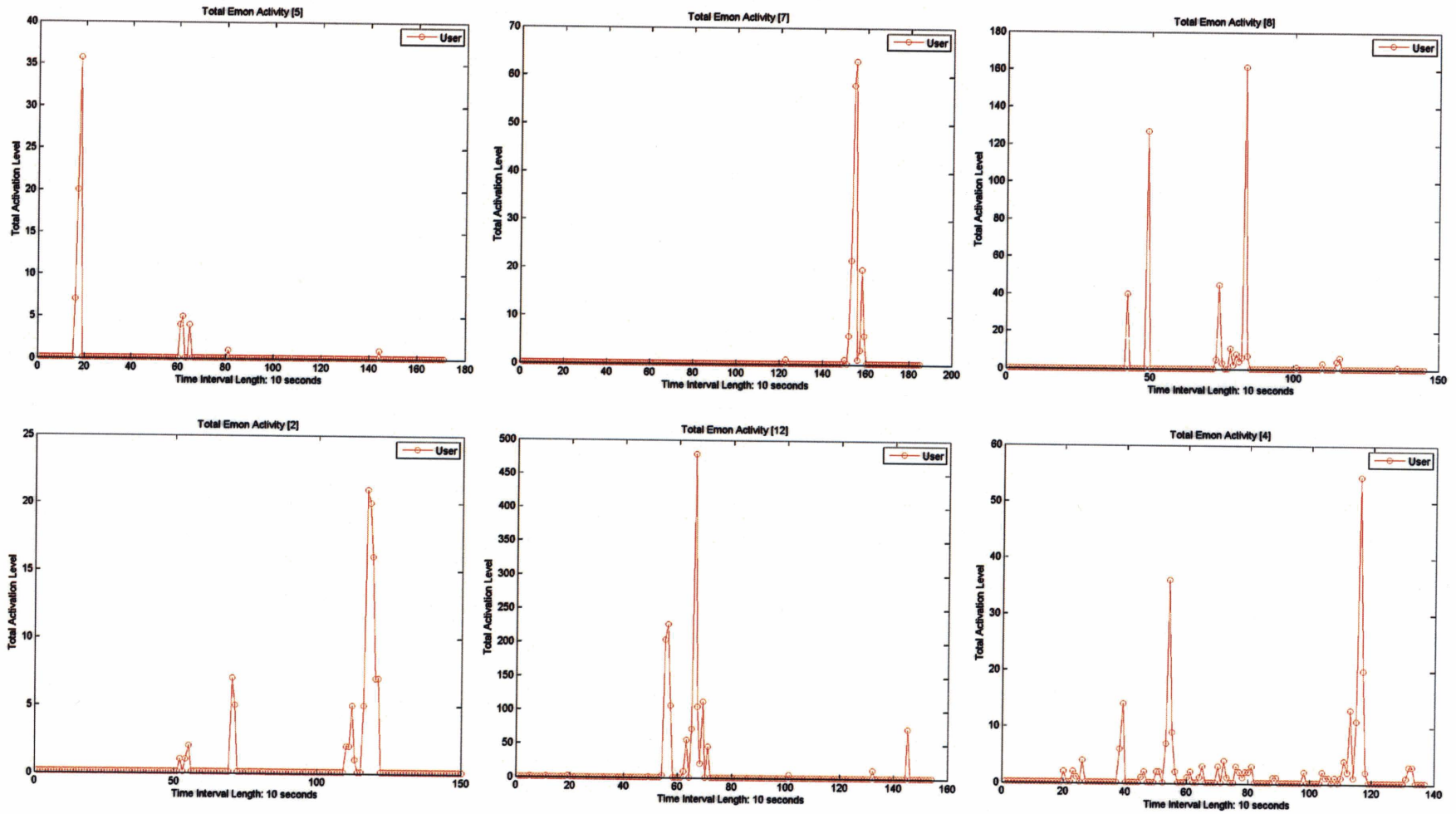
Figure 14: Cumulative structural (emon) activity by the user (excluding evolutionary action). Group A: low activity, mostly concentrated in one area.
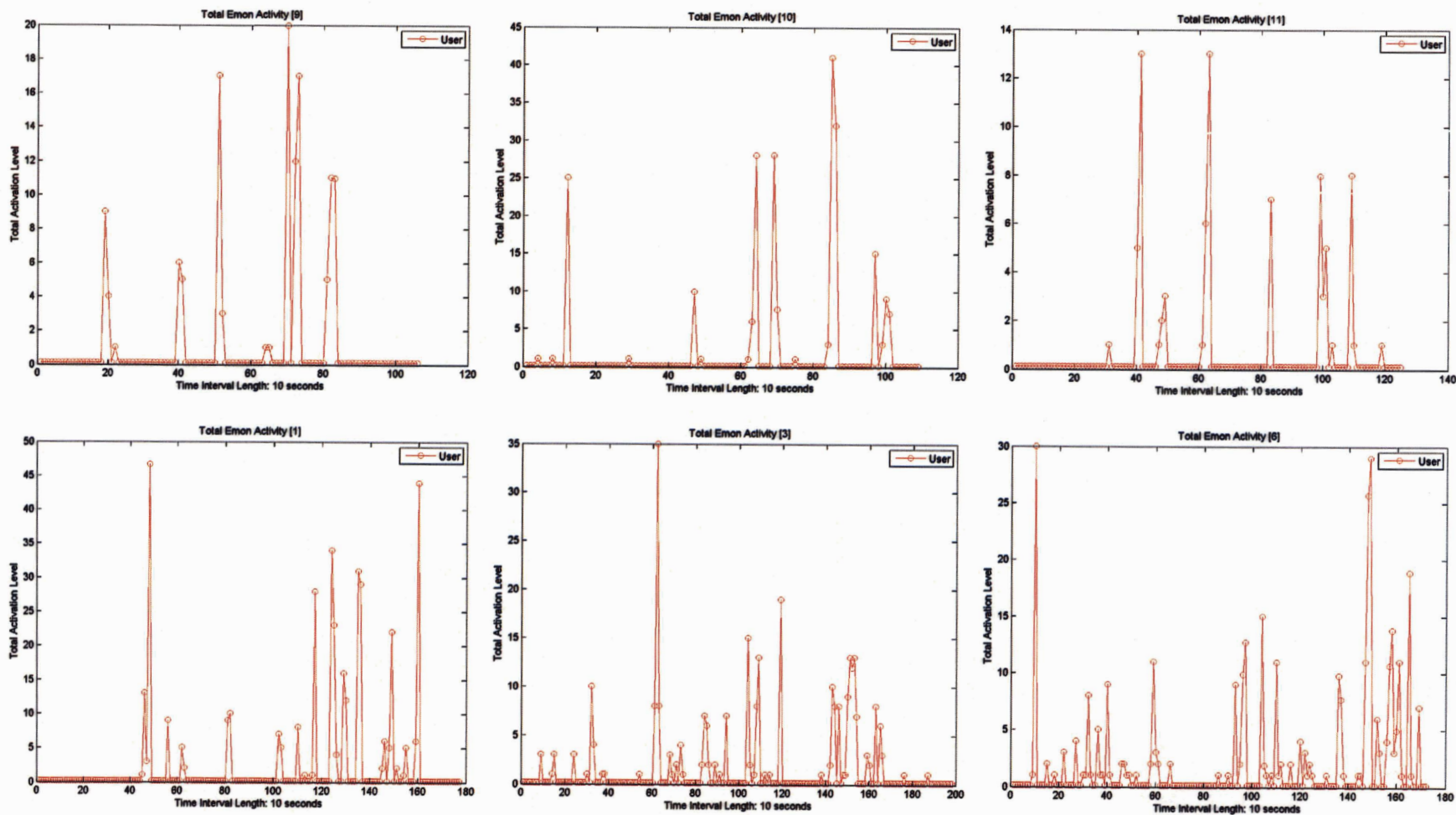
Figure 15: Cumulative structural (emon) activity excluding evolutionary action. Group B: activity throughout the session.

2. Content and Structure: Constantly Changing.

The second question concerns the ratio and timing of interacting with core structural and perceptual constructs vs. creating and exploring the interconnections between them. According to CMIA's 2nd characteristic, the two activities are to happen simultaneously. To see whether this is the behavior exhibited in using the EE, we compared the ratio of time participants spent engaged in component-centric interaction (creating and manipulating the EE's core elements, emons and p-Units) with the time spent on network-centric interaction (connecting emons to emons, p-Units to p-Units, and emons to p-Units as well as manipulating properties of these connections).

All participants appeared to spend considerable time in both component-centric and network-centric modes. The difference came in the dynamic of interaction with the two types (fig. 16-17): participants in group A switched between creating elements and exploring their connectivity from the onset of the interaction. On the other hand, group B started with manipulating the core elements and only then progressed to manipulating both as they continued interacting with the system.
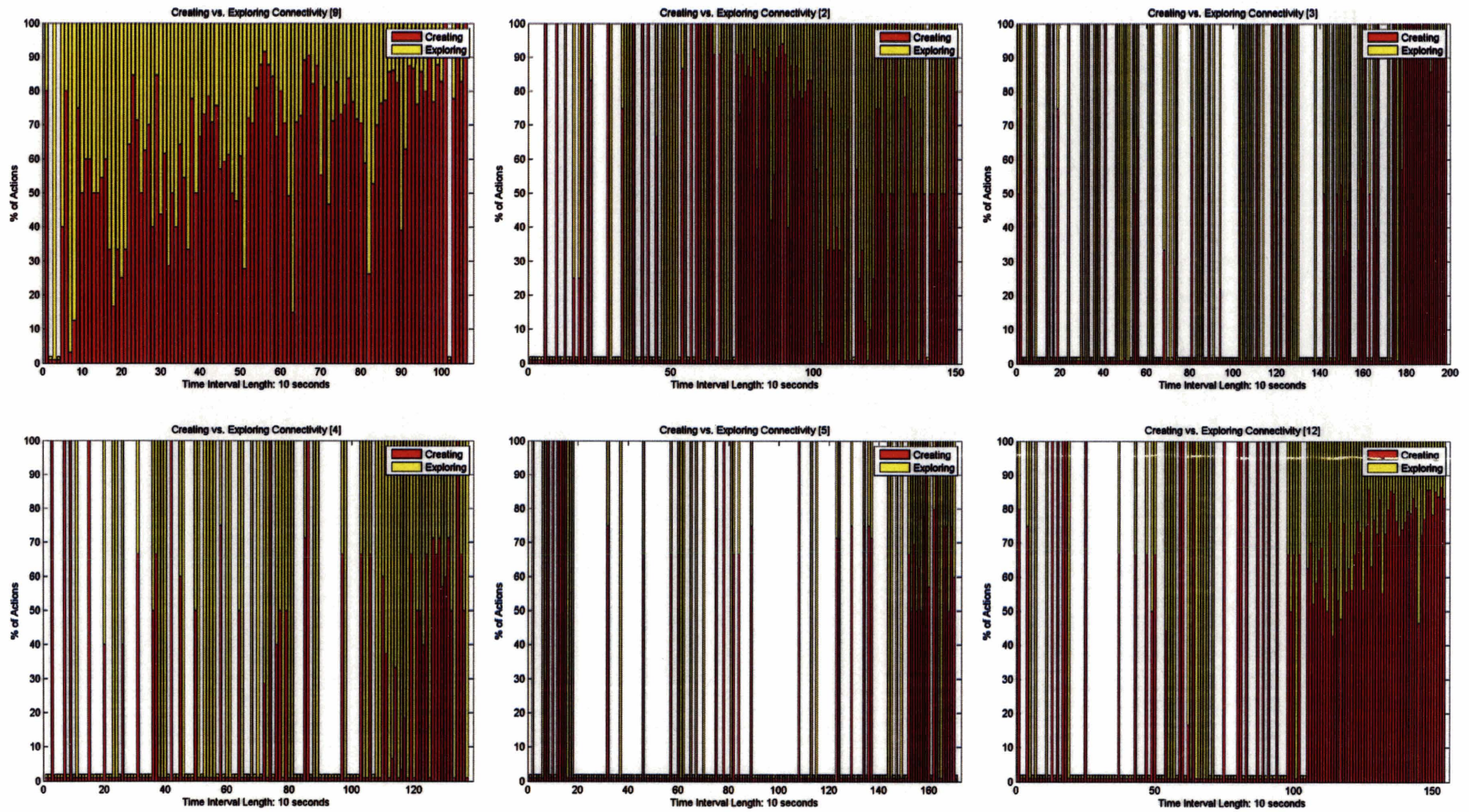
Figure 16: Core elements (red) vs. connections (yellow). Group A: elements and interconnections are manipulated simultaneously.
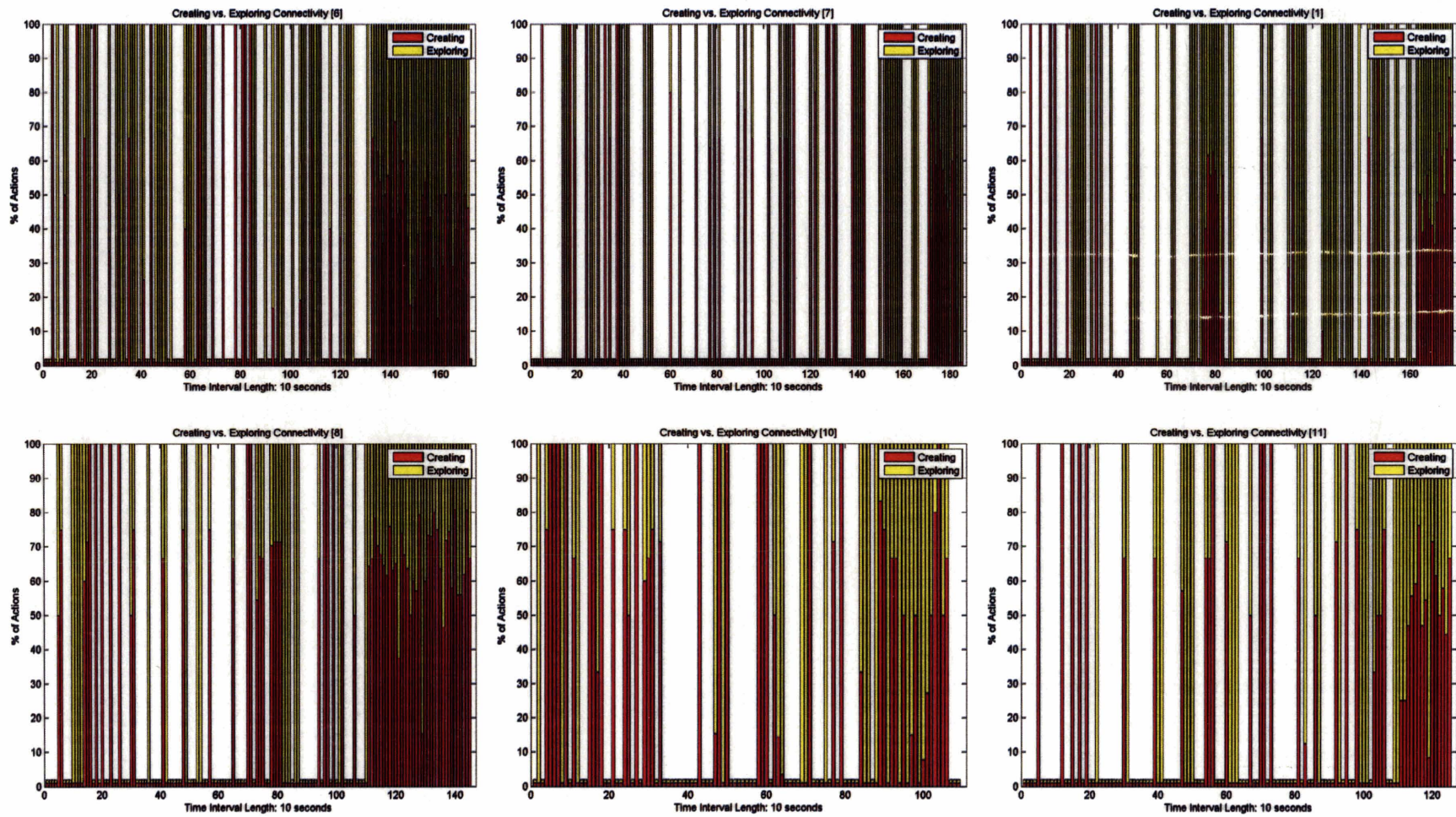
Figure 17: Core elements (red) vs. connections (yellow). Group B: separate manipulation of core elements and connections at the start; mixed end.

3. Authoritative Score: Absent; Price of Mistakes: Low.

The EE has no score and no separate modes for practicing vs. performing - the practice is also the performance. Participants' actions are sometimes a result of their initiative, at other times, a response to machine-driven changes of the network's content and structure. An interaction can only be considered improvisational if participants see it as a combination of familiar activities of browsing, editing, composing, and improvising, without artificial boundaries between the activities, all taking place in real-time and with no iterative editing outside of the improvisation's scope.

Our assertion that the EE encourages such a behavior is based on three sources of evidence. The first is the Emonic/Thermocline installation exhibited by us at Centre Pompidou in 2004. There, utilizing an array of sensors and microphones, exhibition goers were given a chance to improvise simply by walking around, positioning themselves on and around architectural pieces, and speaking/singing. No separate editing was done; rather, visitors' actions as a whole comprised the improvisation.

The second source comes from the graphs below (fig. 18-19) in which we evaluated the change in the number of structural actions over time in each of the controlled study's sessions. In group A, the number of structural actions initiated by each subject is increased towards the end – despite the unfamiliar interface and unknown media terrain to explore, as time passes, participants appear to be more open to taking action, and less afraid of making mistakes, substantiating the hypothesis that improvisational interaction facilitates experimentation. It also means that this group dedicated time to playing with both structure and content, with no long bursts of exclusively structural manipulation. Participants in group B exhibit a pattern of interaction that is less oscillatory, dedicating themselves to playing with just the structure components for extended periods of time.
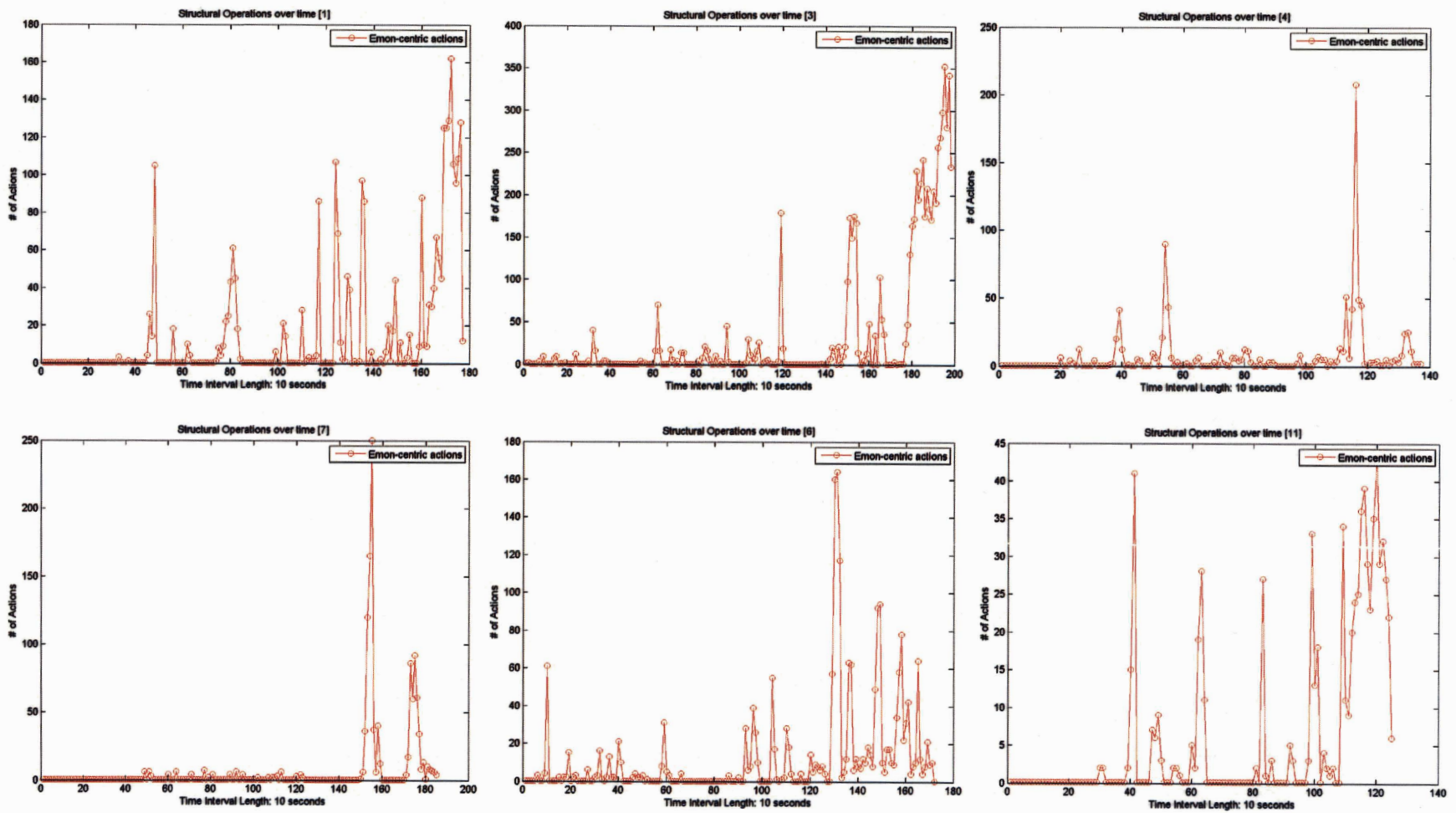
Figure 18: Structure-centric actions over time (excluding machine-driven actions). Group A: sporadic activity, increasing towards the end.
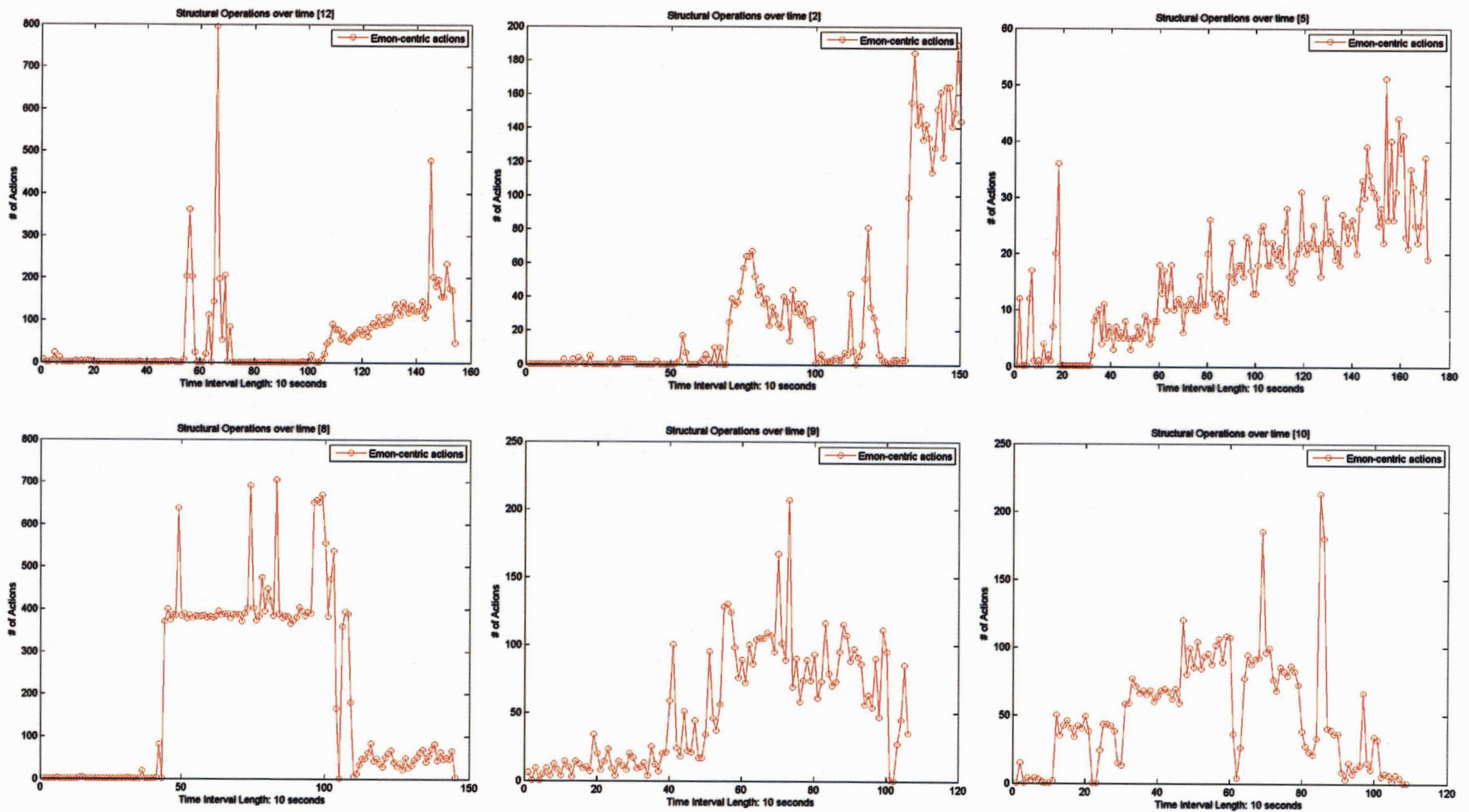
Figure 19: Structure-centric actions over time (excluding machine-driven actions). Group B: clearer structural development.

Finally, our third source is the qualitative questionnaire provided to the study subjects at the completion of their study. Their self-rating, when asked "How would you describe what you were doing (summed to 100%)?" was:

User ID:	1	2	3	4	5	6	7	8	9	10	11	12	Mean	Variance
Editing	50	0	35	50	60	50	5	20	5	30	5	30	28.33	413.89
Browsing	15	60	35	0	10	20	10	5	60	10	15	10	20.83	374.31
Composing	10	10	15	30	20	10	30	25	5	10	0	30	16.25	100.52
Improvising	25	30	15	20	0	10	20	50	30	50	80	30	30	420.83
Other					10[6]	10[7]	35[8]						4.6	

Figure 20: Activity self-evaluation by subjects.

The spread of opinions in regards to the perceived activity type serves as a proof that the subjects felt empowered to pursue different paths in exploring the media space. The lack of consensus, coupled with the fact that most participants marked all four options available, aptly illustrate our point – while traditional interactive frameworks provide one type of activity, the EE sends a mixed message, appropriate for those who wish to simultaneously pursue multiple interaction objectives.

---

[6] "Structuring a network with a structure that has a clear relation to the rhythm it produces."

[7] "Letting the system play."

[8] "Facing the creative uncertainty and the black-hole."

4. Goal: Process, not Artifact Production.

As part of the tutorial, the subjects were repeatedly told that they are free to save and restore Emonic networks whenever they wish. Icons for doing so are a prominent part of the Emonic UI. A high amount of networks saved and restored would mean that the subject attempted to create an artifact, reverting to a state that is somehow preferable compared to the current state. On the other hand, a low number of restored networks would point to users unconcerned with creating a particular piece, and focusing on the process of creation itself. In other words, given the context of the study – a participant aware that he is only taking part in a single session – continuous saving does not signify a desire to resume the activity at a different time. Rather, it points to the participant's fear of his creation becoming invalidated, by his own actions or interference from the machine. Such fear is a characteristic of compositional (rather than improvisational) thinking, since switching between snapshots would be contradictory to the improvisational approach that sees all actions taken as part of one continuous process. Similarly, the number of networks saved would point to either becoming more interested in creating an artifact (an increasing number of save operations) or increasingly focusing on the process (a save operations decrease).

As can be easily seen from the graphs below (fig. 21), all the participants seemed more interested in continuing their exploration than reverting to an earlier network state.
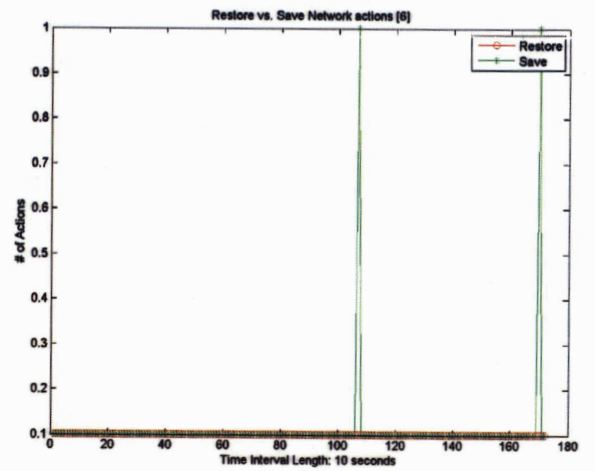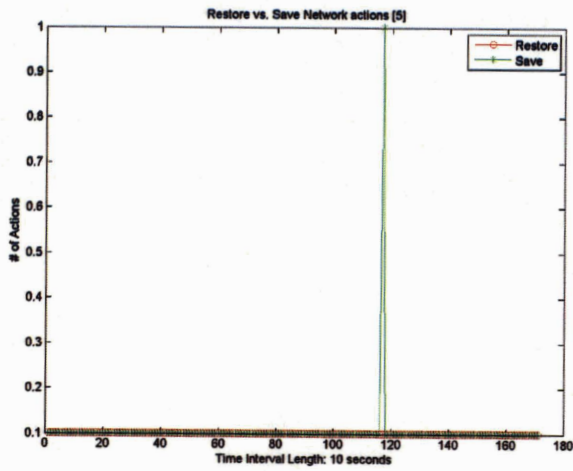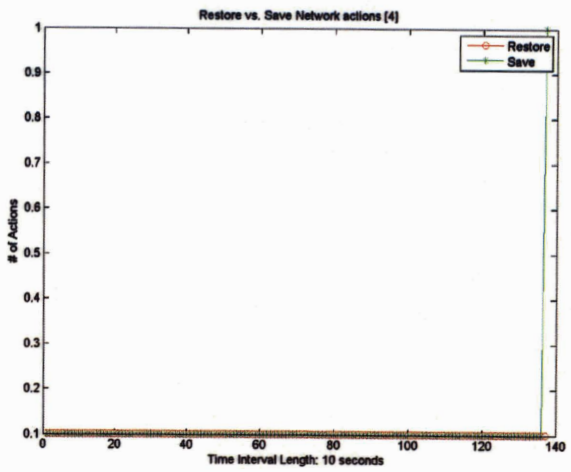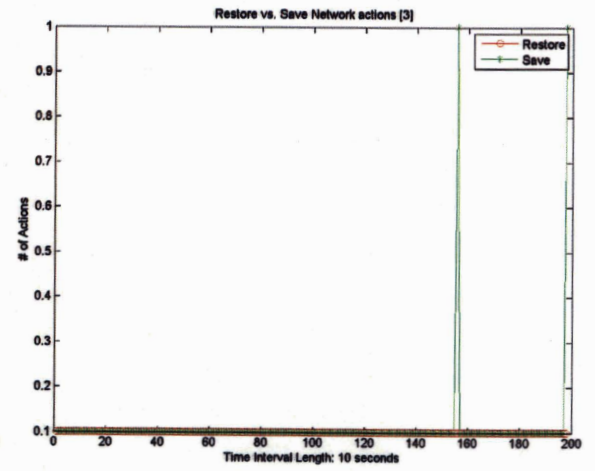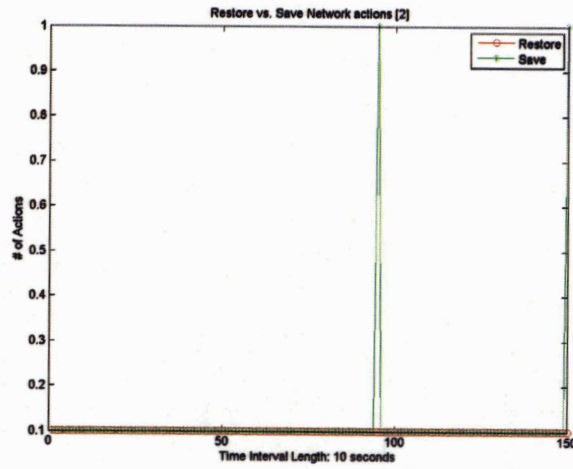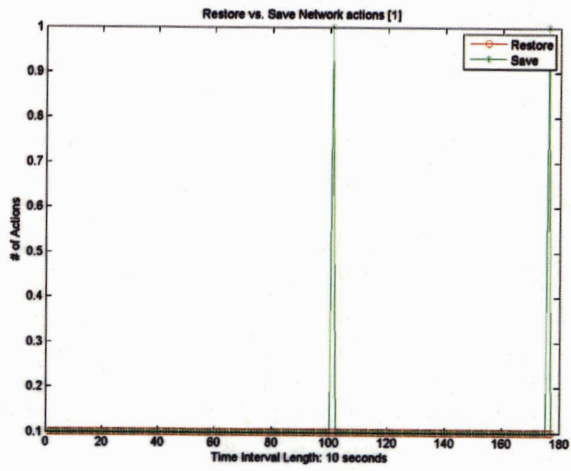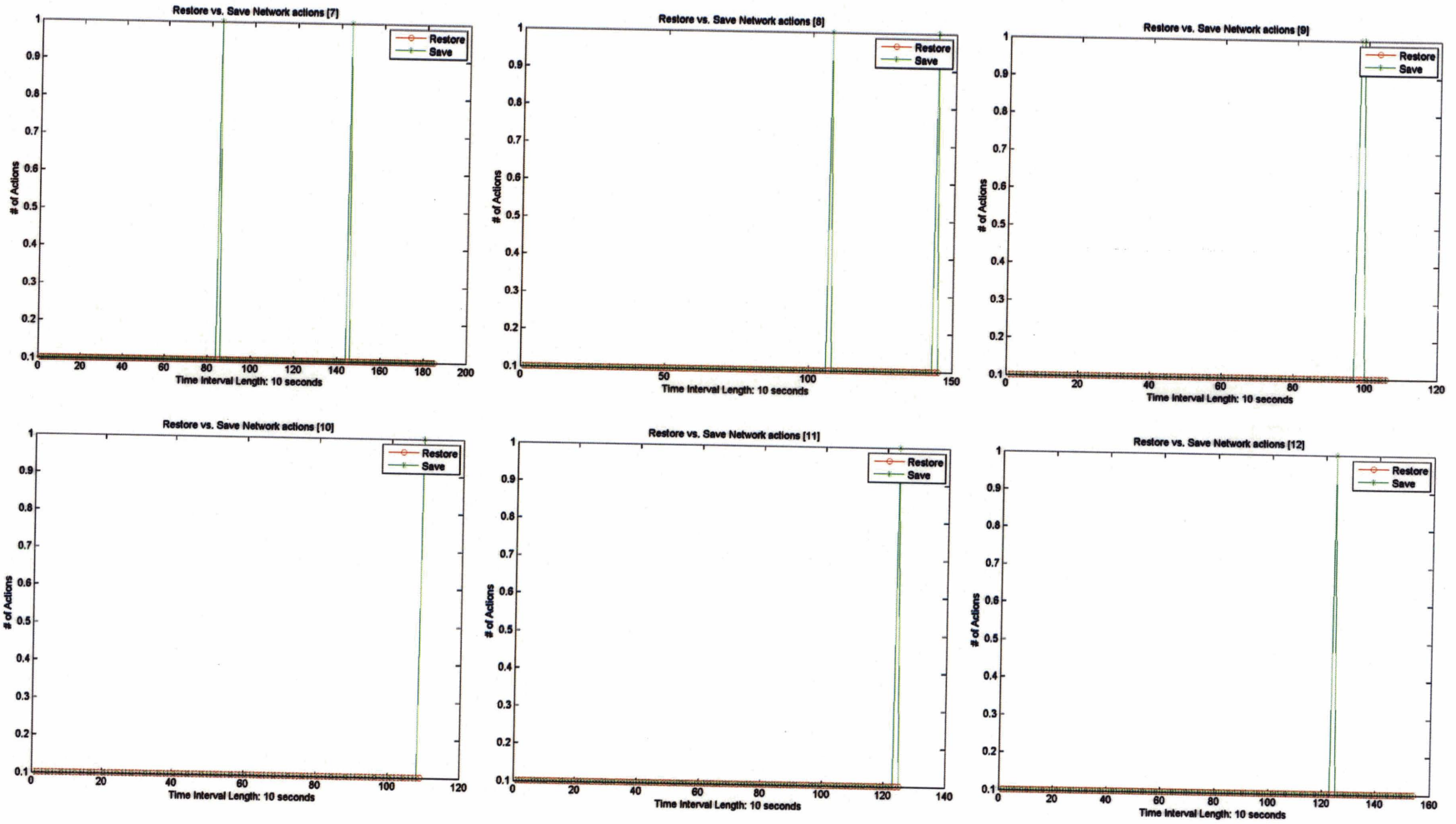
Figure 21: Save vs. the (inexistent) Restore. Everybody saves in the end, nobody bothers restoring.

5. Participation Responsibility: Distributed.

Does the balance of control between study participants and the machine resemble the balance present in a group improvisation? In order to find out, we looked at the ratio of actions initiated by a given study participant vs. actions initiated by the genetic algorithms. CMIA states that in a computer-based improvisational interaction, just like in its analogue counterpart, there must be an ongoing dialogue between the performers.

To find out whether there is such a dialogue, we looked for correlation between the intensity of improviser's actions and those of the machine. As can be seen in the graphs below (figs. 22-23), two types of behaviors seem apparent: group A (3 out of 12) has given up the control of the system early on and assumed a largely passive interaction. Group B (the remaining ¾ of the participants) exhibited a more dynamic interplay between their own and the machine's actions.
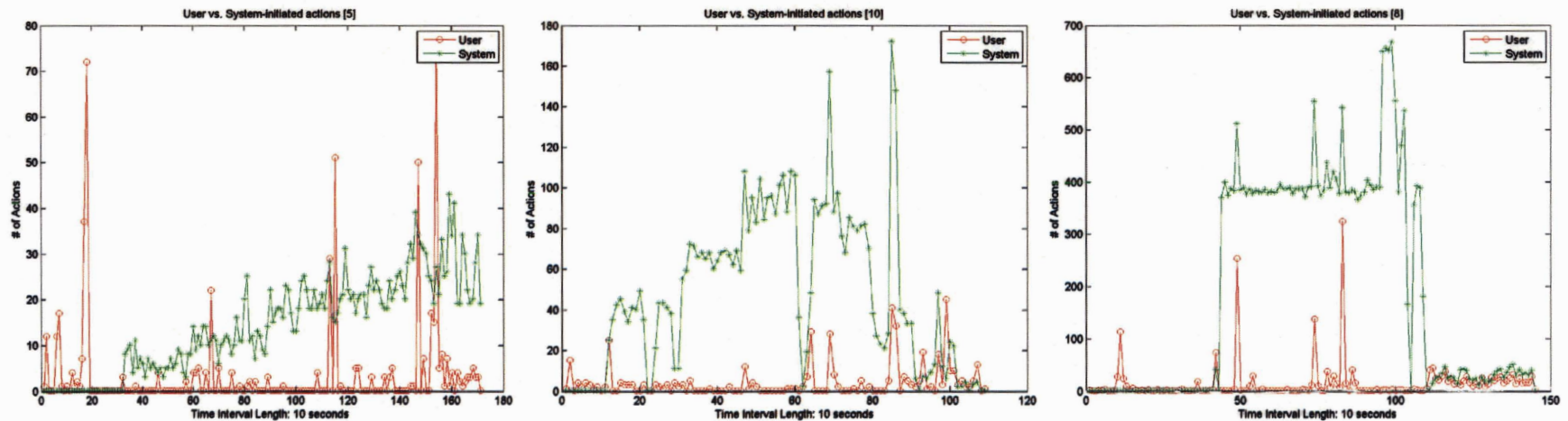
Figure 22: Amount of actions by user vs. machine. Group A: machine is in control from early on, while the user remains largely passive.
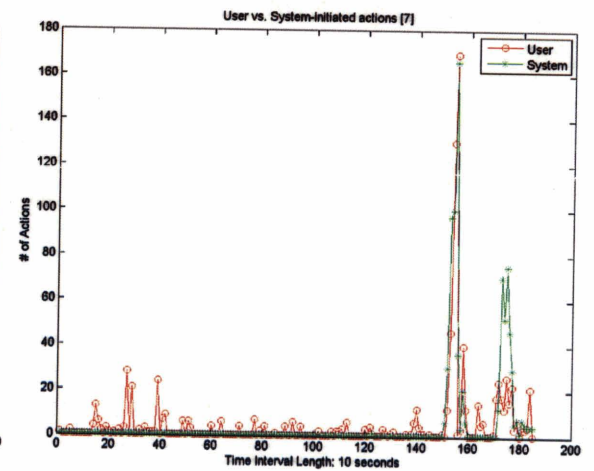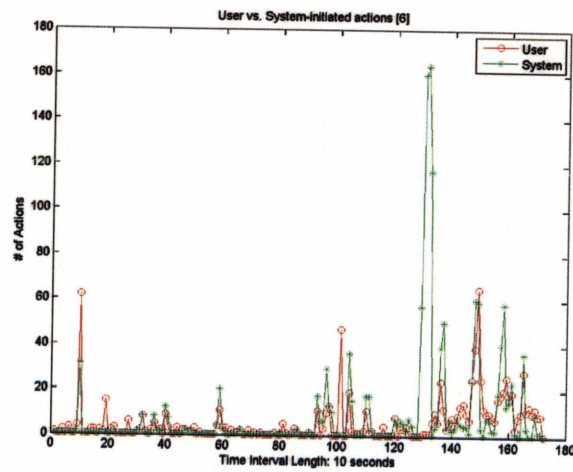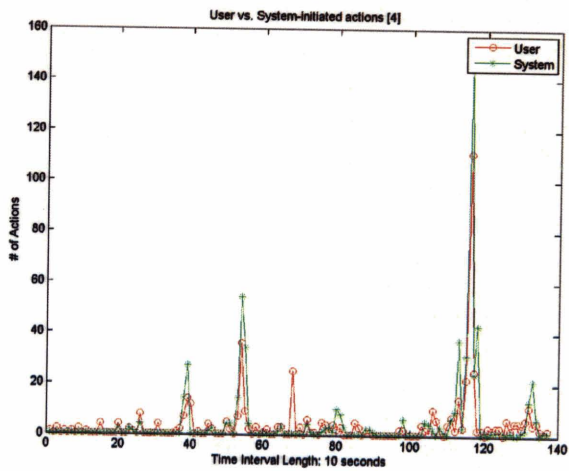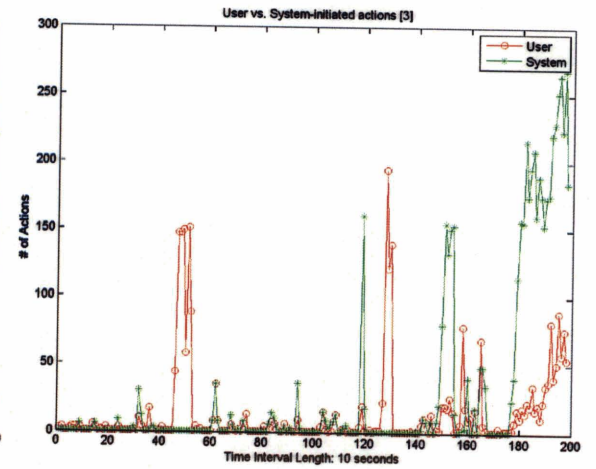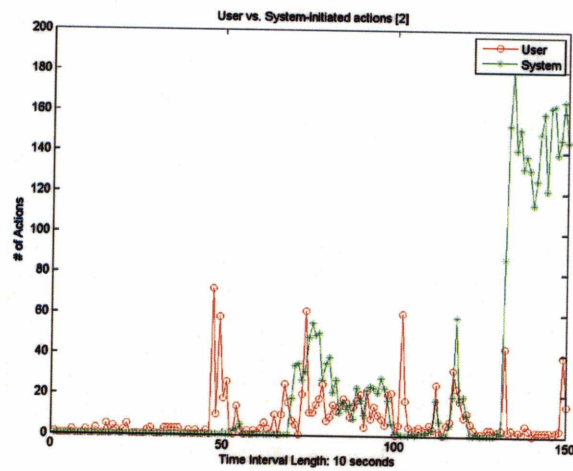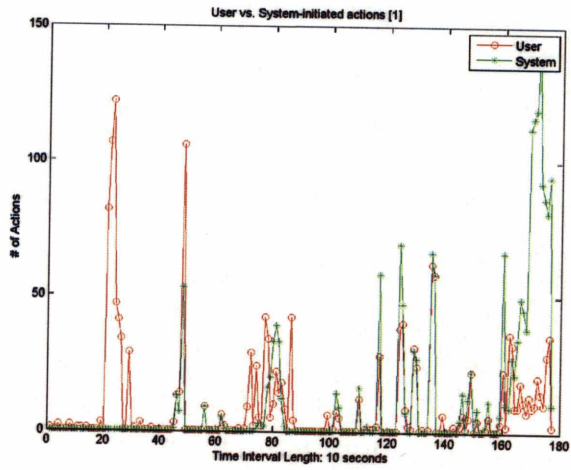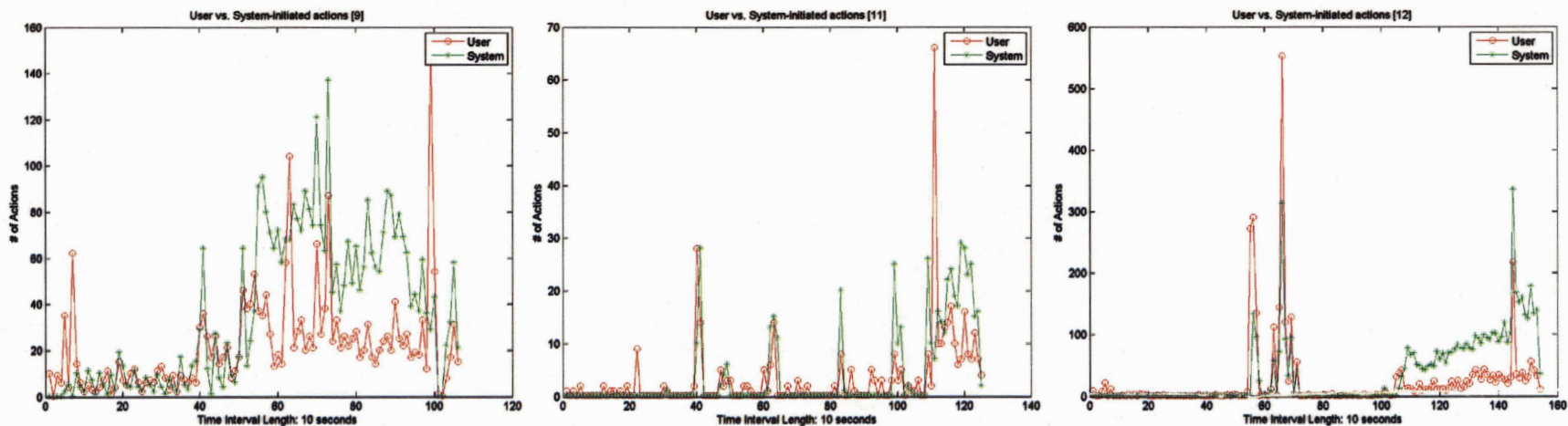
Figure 23: Amount of actions by user vs. machine. Group B: the two are in tandem, with evolution taking a more prominent role towards the end.

A different way of looking at the same data is to consider the ratio of the two at any given point in time (which allows us to see the pauses as well as the details of human-machine interplay more easily). From the graphs below (fig. 24) we see that all the sessions start with the participants directly controlling their networks for a few minutes, then, for a period of time in the middle of the session they share the control with the machine. Finally, in the later part of the session (apart from one participant session) the machine seems to play the predominant role. It is also easy to see that the second part of the session is almost invariably denser (in 10 out of 12 cases).

The results appear to support our assertion: the machine and the participant seem to be in a continuous dialogue – a spike in the actions of the subject is reflected in actions of the machine. At the same time, the two are not mere followers of each other – a more complex dynamic can be observed, where a raise in the actions of the machine many times leads the human to assume the role of an observer.
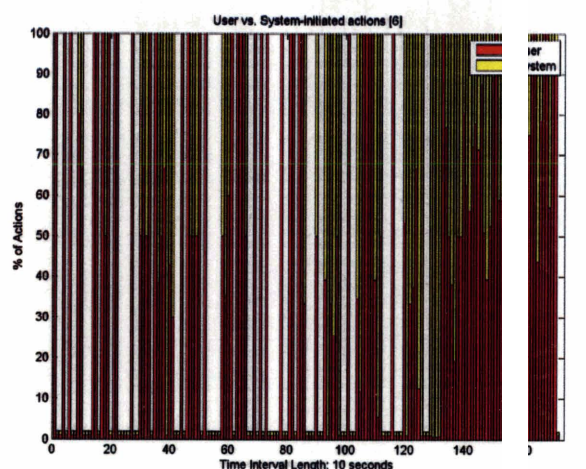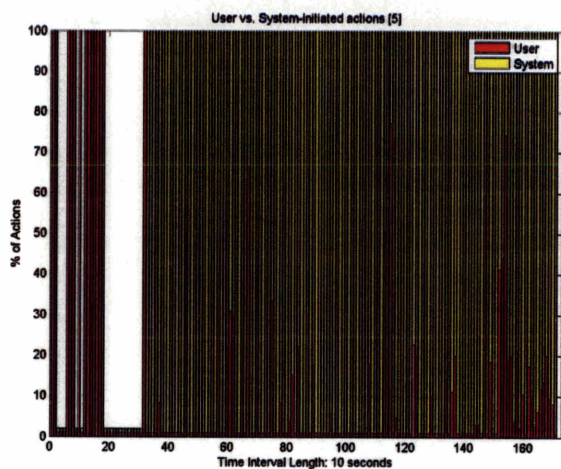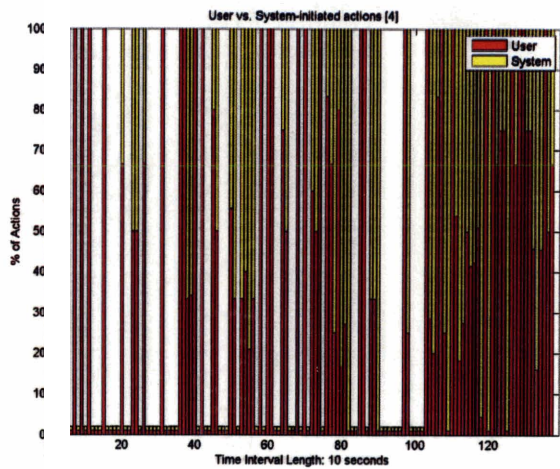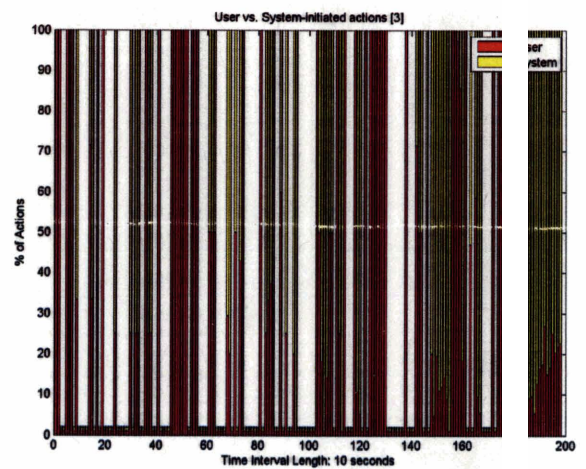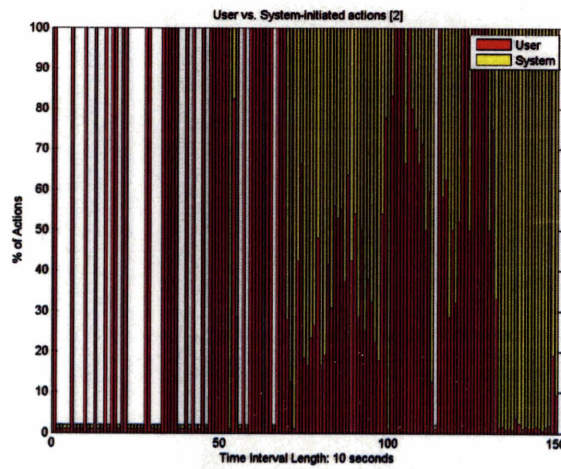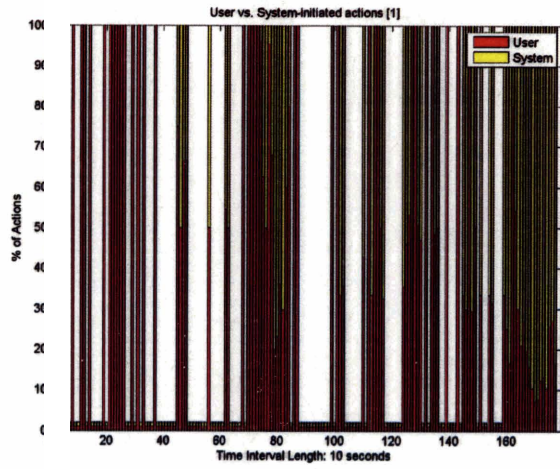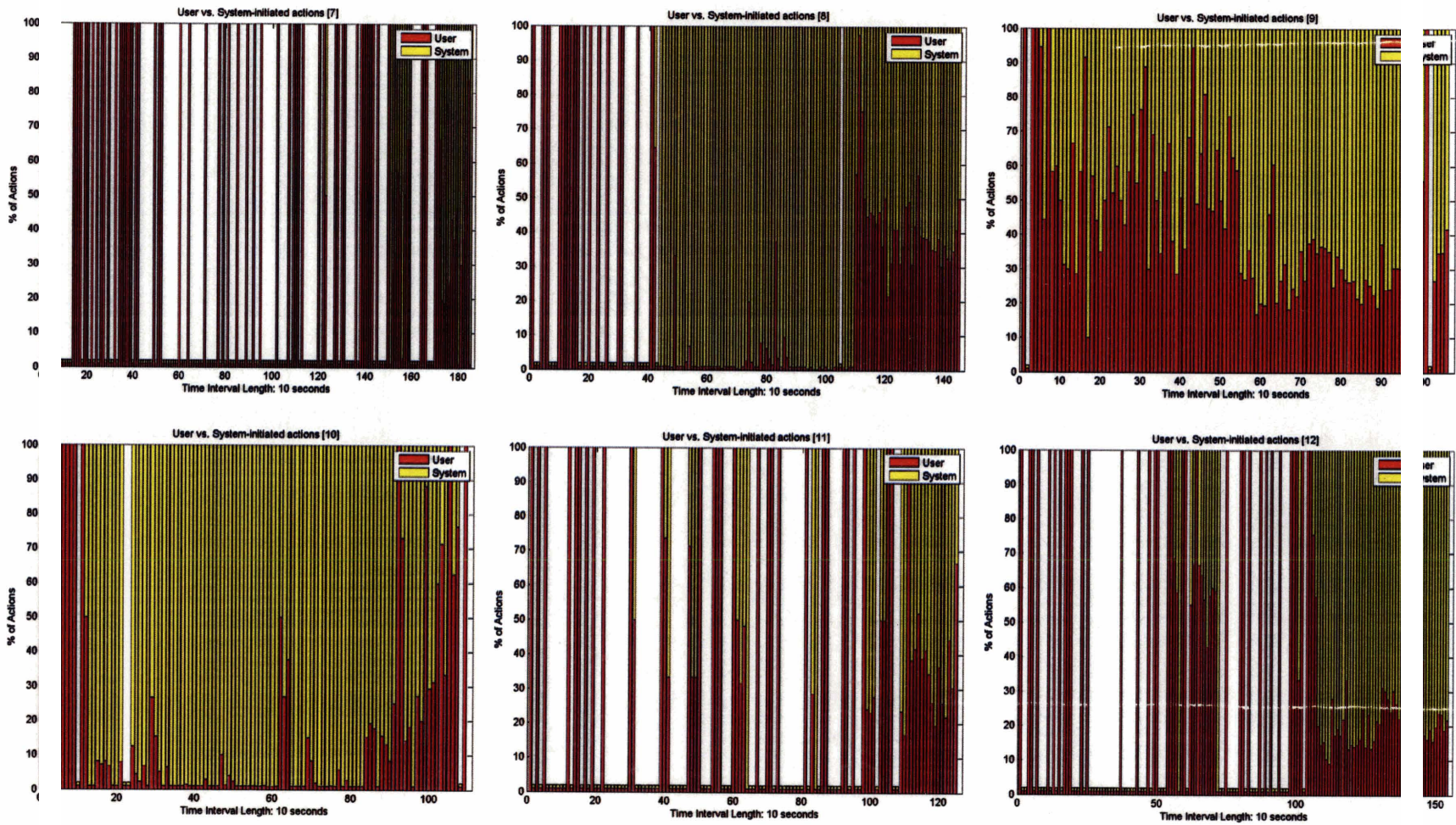
Figure 24: Ratio of actions initiated by user vs. machine.

is interesting to compare the results of this question with the questionnaire. In it, each participant was asked what he or she thought the role of the machine was: a partner in the creative process or a tool following commands. The responses are:

User ID:	1	2	3	4	5	6	7	8	9	10	11	12	Mean	Variance
A Partner in the creative Process	30	70	40	40	10	98[9]	33[10]	20	50	60	20	60	44.25	566.35
A tool following your commands	70	30	60	60	90	2	66[11]	80	50	40	80	40	55.67	564.56

Figure 25: Evaluation of machine action by study participants.

6. Context and Disruptions: Relevant and Incorporated.

In the Centre Pompidou study, an array of sensors has been installed around our part of the exhibition space. The sensors, which monitored pressure, noise, movement, and position allowed for incorporating the conditions of the exhibition space and the actions of visitors into the ongoing performance. For instance, motion sensors, located opposite of the central element of our exhibition, measured the speed and the frequency of visitors passing in front. This information was mapped to performance parameters such as spatialization and volume of sound (when the speed of visitors would cross a varying threshold, the sound would appear to be running after the visitors; at a slower speed, the sound layers would be more complex than at high speed). Pressure sensors, installed under the seat of the architectural component, measured the posture of the visitor (sitting, moving, or laying down) and controlled the evolution of the audio space. Any sounds or words uttered by the visitors were incorporated into the audio stream, and spatialized throughout the audio space depending on sensor data, either immediately or at a delay.

---

[9] "I thought the system was not a partner per se but more a contradictory partner."

[10] "Because there was no mutual understanding and I felt the evolution agent had more control."

[11] "Could have been higher percentage but the end composition is something different that I couldn't have imagined and didn't know what effects were really based on my commands."

Grouping the results

The table below presents subjects' responses sorted into groups.

User ID / Question	1a	1b	2	3	5
1 (V)	A	B	B	A	B
2 (CS)	B	A	A	B	B
3 (M)	B	B	A	A	B
4 (CS)	A	A	A	A	B
5 (M, V)	C	A	A	B	A
6 (V)	B	B	B	A	B
7 (V)	A	A	B	A	B
8 (V)	A	A	B	B	A
9 (M, CS)	C	B	A	B	B
10 (CS)	C	B	B	B	A
11 (V)	A	B	B	A	B
12 (M, CS)	B	A	A	B	B

Figure 26: Subjects' behavior sorted into groups. CS – computer scientist, M – musician, V – visual artist.

Questionnaire Review

At the conclusion of each session, in addition to the quantitative questions mentioned above, the participants also had to answer 10 freeform questions. The responses are in the ID order, and each question has a summary overview. Here are their replies:

1. What functionality did you like the most, and why?

Overview: evolution is #1 winner, gathering 8 votes; structural layer's interconnectivity is a close second with 6 votes.

Responses:

1. Interrelationships between different emons – the capacity of one to trigger another in an ad-hoc network.
2. I like the evolution, in particular the part that evolves the levels of activation across the network. I like that feature because I really felt like I was playing with the computer, she will activate something and I will deactivate it or activate something else. I like it because unlike other features it felt a little more transparent.
3. The connections between emons and their resulting behavior are very much like what happens in the real world, which gives unlimited possibilities to create different scenarios.
4. Evolution: Because you can explore something that you don't think of. Further, learning your likes/dislikes makes things easier.
5. Structural layer as a control of various actions – it enables to visualize the relationship between actions and therefore ease (improve?) the composition and editing process. Evolution – brings delicate changes to the sensory output while keeping its original state apparent.
6. The emons that connect to each other and the fact that the system goes on its own and changes your stuff.
7. I liked the ability to freely put it together (i.e. juxtapose and overlay) and take it apart (i.e. make it isolated). Why – because I don't know exactly what I was composing but that in itself was a big constraint and needed physical freedom manipulation and reaction of end-composition.
8. I like that it can "evolve," because having total control in a digital environment is not really very nice for creative practice – one desires an element of chance! I also like the ability to record one's own new sample because, in contrast to the "evolution" as complemented upon above, and in contrast to browsing and editing, sometimes you may have a feeling of just [what] exactly you want to put in there and then "poof boom" you can do it with the record-your-own feature.
9. The evolutionary feature – it was very easy to get cool, complex textures to work and experiment with.
10. Making connections, exchanging media, seeing what the network will do.
11. Evolution – did better work than I did.
12. The automated evolution; surprisingly cool results with no effort.

2. What functionality did you like the least, and why?

Overview: No single thing, but the complaints include feeling not enough control (generally and of audio in particular), minor UI issues, and evolution (2 votes) because it messes up a "master plan" or doesn't give enough control to the user and "messes things up".

Responses:

1. It takes some time to get used to thinking of how it is not the object that is the focus, but rather spatial transformations and time emons. This is not criticism, but just a note that it requires different thinking.
2. At the beginning I really didn't like that by default when you want to add a connection between an emon and a p-Unit by default it randomly connects it to a bunch of p-Units instead of connecting it to the p-Unit that I want it [to connect]. It seemed like the system wants to force you to give up control over the creation and it was a little annoying. After a while though I started to realize that I have to think about creating in this environment in a different manner. I still have mixed feelings although because I like to have more control when I create something, otherwise I feel like I am being lazy.
3. I liked everything.
4. Perhaps most dominating mode is audio and it was hard for me to control that the most.
5. N/A.
6. The Shift + Right-click stuff: not intuitive enough.
7. I disliked the evolution agent. It surprised me with elements I was not thinking about which wanted me to reveal more of that and try out some other interesting effects, but at the end it messed up my initial master plan - a master plan that I didn't have but emotionally I felt out of control.
8. I did not like that every new sample started at the same position and volume, thought I suppose I could get used to that. I did not like it because they invaded sometimes, and other times they were not really present but I was not in touch with them enough to know that they needed to be brought in. Hmmm, perhaps you have to deliberately place and add volume to each sample, but then I dunno – not so important. It all was cool. :p
9. I think the neural network was a bit confusing musically to me. I could understand what it was doing cerebrally, but not musically. I think seeing it work automatically was very interesting though – almost mesmerizing.
10. Evolutionary steps, they messed up things too much, need more control while doing evolutionary steps (blocking some properties form evolving).
11. Nothing specific, but it was hard for me to create with the logic of the emons and samples.
12. The editing of specific properties; precise controls take time to reach.

3. What functionality you lacked / wished for?

Overview: UI issues, API for object creation, more control of audio (2 votes) and of evolution, a tangible interface for interaction with the system (2 votes, however this functionality does exist, although was not used in the study), an undo feature (from the person who wanted more control generally).

Responses:

1. That I couldn't attach the exact same trigger values to different emons by manually typing the numbers in.
2. I can't think of anything in particular.
3. To create your own objects and properties, not necessary audio and video.
4. More versatile control of audio. Improvisation can happen once you are comfortable with the interface. Currently, as a beginner it was a little tough to play with things. But once learnt, it seems like you can improvise.
5. None.
6. Tangible means of interaction with the system and screen is too small.
7. The 'undo' feature and ability to segment a space in time. Things kept evolving, moving forward, and I am wondering how some of the past can be kept.
8. Bigger screen and interface – more physical, and that fights back in a sense, I mean so that I have to work to reign it in as it evolves, etc... really comes to life!
9. Hmm... aside from fixes to some interface problems I would have liked some more templates to start with as well as more functions to morph the audio (video was fine).
10. More control of the evolution change; more properties for video manipulation (like effects).
11. –.
12. Synchronization of media; [ability to introduce] dramatic change.

4. Which program do you think is similar to this one, and in which way?

Overview: Max/MSP gathers 5 votes, but only for the graphical aspect (the layout of components).

Responses:

1. Generative components (parametric architectural modeling software). EMI automatic music re-composer [David Cope's Experiments in Music Intelligence].
2. I can't think of any.
3. It's like various audio and video editing programs mixed together.
4. I have only used Premiere and CoolEdit that seem very similar. Other one that comes to my mind is the Hyperscore.
5. I don't know any programs for media.
6. None, there is only a similarity with Max/MSP in term of linking.
7. Similar is any media composition tool like adobe Photoshop, After Effects, Premiere – because you can compose things freely. However, Emonic Environment is different because the useful thing is that it revealed things that I wasn't thinking about and has its complexity of time that a single brain cannot comprehensibly think of immediately.
8. Max/MSP and Jitter are the programs I know of that are sort of close, but they are not as nice, Paul.
9. None I can think of.

10. Max/MSP, but only the layout language. Anything with random choice (nothing I know of).
11. Flash, Director, Max/MSP, Jitter, they are all about editing media in a sequential manner.
12. Max/MSP, PureData for the graphical aspect of programming actions.

5. Who do you think would like to use this software? Why?

Overview: DJs, VJs, architects, children, interactive designers, partygoers, people with creative block, participants in social networks.

Responses:

1. Artists interested in computer aided design.
2. People who like to browse in a creative manner, like DJs.
3. Creative people who interested to explore their imagination.
4. Music composers. Video Artists and DJs who can now control both audio and video.
5. Architects, kids, anyone who's interested in making collage.
6. It can be fun over parties with friends to create patterns and let the system go on its own. Anyone from the party could join, interact with the system and add to it. It could be a fantastic VJ and DJ project to be made on a larger scale though and more tangibility would be needed to make it more intuitive and spontaneous.
7. I think it would be nice to somehow parallel (apply) this to social situations. Then it would give equal opportunities to people. It would reveal hidden talent that we are not aware of.
8. If the interface were more or a partner you collaborated and struggled with, I think in performance it would be great! Also, could be a tool for people with "writer's block" or "videographer's block" or "musician's block" as it adds elements of chance and allows for quick unexpected juxtapositions... Like a deck of cards, leaves room for lots of imagination, too.
9. I think this is a medium for artists who want to express themselves using a different form of media.
10. Experts – might want more control; average users – would like it with better balance of control / evolution; novices – might be good as creative play, not planning too much.
11. VJs – can be a nice real-time editing tool.
12. VJs – DJs – interactive designers; because it gives quick and customizable results with little effort and the possibility to interact with a given network.

6. If you'd have a professional version of this software, with nicer design, more features, and bug-free, would you use it? If yes, what would you use it for? If no, why?

Overview: 10 yes, 2 no. To create videos, animations, music, browse media collections, explore dynamic media interrelationship, to communicate to others for whom you have no words, make media installations and performances.

Responses:

1. Yes, composing video animations.
2. I might use it to browse a collection in unpredictable manners, or maybe to create my own performances.
3. I would like to have it very much, but I don't know what I will use it for, maybe to compose things maybe to understand how media elements interact in time and space.
4. Yes. To create music videos and maybe for live performance if I get any good.
5. Yes, for video editing purposes.
6. I would use it as a collaborative tool. I don't think I would use it for a personal use (but maybe...)
7. Yes I would use it because the features allow me to manipulate audio, video, etc., not as static entities but as content media. I would like to use it possibly for new audiovisual expressions that can be used for communication in for example ads, etc.
8. Yes, I'd use it. I'd make networks and send them to people that I want to say something to, but don't have words for. Hehehe.
9. I don't think I would use it because the way I think is too linear. It's hard for me to feel as involved in creating say a composition when I feel I don't have enough control XXXX, it's fun to just experiment with – maybe with someone else (although that's another feature) i.e., networked.
10. Music composition / DJ-ing, maybe on top of my own playlists, same for videos, on top of my own video content.
11. Probably not, cause I am not in the need of video-audio mixing and special effects.
12. Yes, to make fun and rich interactive media installations for lots of money and no time.

7. Is there some feature of this software that is *different* from other programs you know? If so, what is it?

Overview: Temporal characteristics of structural layer (6), pseudo-random effects, learning, evolution (7).

Responses:

1. Time is a separate object.
2. I think the whole thing is different, I don't know of any software that provides soo much unpredictable inputs, and is partially 'uncontrollable', I like that, though [it] is something you need to get use to.
3. Mix all the things that we usually use separately together and of course the result of this.
4. Learning is different, the machine also suggests improvements which is good I think.
5. I don't know any media related software, I assume the evolution is unique and [so is] the multi-user aspect of the software.
6. Well it is such a different program on its own, but having developed VJ and DJ software applications for real time editing, there are some overlaps in the idea of changing colors in real time, adding video effects, and so on. However the rest is very different and seems to combine powerful aspects from different environment: the link between events (max), the real-time video editing (some research-based projects), and the AI side.
7. Difference in ability for time manipulation and evolutionary methods.
8. Paul's soul is in it. :) That makes it funny and smart and nice to play with. Okay, also I think the evolution and choosing the different behaviors to be different kinds of pseudo-randomness controlled is great to make the whole experience more lively and less under direction of a dictator.

9. Actually there seem to be a lot of different features – the mediums I use are completely different (i.e., notation programs, and editing programs).
10. Evolution / network control.
11. The evolution – is a killer!
12. The evolution layer is new and missing to regular programs usually too static. I think that's the real contribution of this model. And in the end nobody really cares about the static part, as long as there are lots of visuals and sounds, and the system reacts to inputs.

8. Did you find evolution useful? If yes, in what way? If no, why?

Overview: Yes – 10, No – 5 (some participants answered both yes and no). Yes, stimulating and an ideas/possibilities generator. No, confusing / not controllable rigidly enough.

Responses:

1. Didn't quite understand what it did. Somehow recorded my pattern and recomposed in similar ways, but a bit too randomly. Segmenting and recomposing might have to be improved
2. Definitely, look at the question about what features I liked most.
3. I found it useful as an ideas generator and to simulate some real world processes of how things happen.
4. Yes, it refined whatever I created. It can show you if you lack imagination.
5. Yes, in that it let the user explore slight modifications to his design, with somewhat of a graphic representation. Users could then learn about new possibilities and also experience the sensory outcome of not being in full control of what they make.
6. I did not find it useful per se as I did not feel it was following what I wanted to do, however I felt it was motivating to see an evolution on what I was doing, slightly frustrating when I wanted to control better the content, but original for more interactive space application, e.g., parties.
7. No – and possibly I have control issues. But a surprising yes, because it presented new elements and types of compositions I couldn't have thought of.
8. Yes, see V.
9. Yes (see my favorite feature question).
10. Not, but I wish. It was too messy. Needs control to exclude properties from evolving.
11. Yes. It worked better than I did.
12. Yes in assisting transformations; no in getting me confused at times; more practice might help in controlling what's happening.

9. Do you feel the Emonic helped you think about creating and browsing media in a way different from how you did it before?

Overview: Yes – 11, No – 2 (one participant has both).

Responses:

1. Yes.
2. Certainly, and in an almost violent way, at some points I would be annoyed that the system wouldn't do what I want it, and I guess that's the game, really getting to feel that you are with an active partner, [it] is just that we are soo used to using the computer as a passive tool that it is hard at the beginning.
3. Yes.
4. Perhaps, yes. Though I'll have to use it long enough to really justify my answer.
5. Yes
6. Yes definitely.
7. Well it really got me thinking about the partnership between me the creator and what the machine can do. Since there are so many things out there and new ways people think, those types of suggestions are helpful. I think that evolution agents go beyond the secretary agents that were popular back a decade ago. It really isn't assisting me but I think if I gave it some time I would gain some trust and take on partnership and not just suggestions.
8. Yes.
9. Yes – that was one of the most enjoyable aspects of it.
10. Yes.
11. No, I didn't really fully understand its use, except for being a sequencer.
12. Not really in the browsing aspect. But in the mixing media part, it's been pretty interesting. It's manipulating the medium at a level that's not tractable by hand, and which gives good results.

10. Any other comments?

1. ...
2. ...
3. I think this system could be used in different areas of life not only media.
4. Learning the interface might take a while...
5. ...
6. This is crazy work, and a fantastic combination of video making, DJ, and artificial intelligence. A lot of potential in the entertainment world...
7. Great job, you have changed my view and make me desire to want more of media applications. I will also take a different light to machines.
8. Was fun! Also nice to do with another person. Sometimes I wanted to be alone with it to not feel self-conscious, but other times I was glad you ([study conductor]) were here because I could pull you into it, or the machine could do something evocative and pull you into it, and then it was like playing in a group, not alone, perhaps more so than if it had just been me and machine. Oh, also, thank you! :)
9. ...
10. I like the networks thinking for timeline, the evolution has potential but made too much changes/damage...
11. ...
12. I want to see it in action for real with users playing it together.

Conclusion

Contributions

This thesis' contributions are as following:

1. Presenting an alternative model of media interaction based on the ideas of non-idiomatic improvisation that encourages *active exploration* of media and its structures.

2. Developing a framework in which computational tools actively participate in the creative process and the interaction encourages structural exploration of media.

3. Formalizing the concept of improvisational interaction as *Cognitive Model of Improvisational Action (CMIA)*, a model of interaction that allows participants to switch their attention between compositional elements and structural processes.

4. Presenting an implementation that reflects the characteristics of CMIA, the Emonic Environment.

5. Conducting two studies that analyzed how users of the Emonic Environment interact with digital media.

Future Work

The model of improvisational behavior presented in this work can be extended by further analyzing the characteristics of improvisation and their applicability to the domain of human-computer interaction. It would be worthwhile to extend our work by conducting a comparative study of the behavior of people using our and other systems. In addition, we would like to continue our online studies, to bring in more people from varying backgrounds and skill levels. An additional study utilizing the group improvisation component of the Emonic Environment would also be valuable.

The system is currently audio and image-centric. We would like to further explore its potential in manipulating text so it can be used to browse news collections and text databases. In terms of the setting, the system has been used at a museum exhibition at Centre Pompidou. We would like to continue studies in public venues, such as airports, parks, subways, etc, encouraging collective improvisational media exploration.

Finally, a more immediate work includes an ongoing addition of new processing capabilities, as well as making it possible for the EE to act as an improvisational hub, used by third party systems to provide an improvisational experience to their users, without having to directly interact with the EE.

Final Remarks

Many of the features of existing interactive frameworks are over 20 years old, and do not reflect the current creative desires of computer users. The improvisational interaction may provide a much needed alternative, however our understanding of improvisation is still limited and there is a clear need for more research concerning improvisational interaction theory. An essential component of that research involves creating applications that utilize the improvisational approach. It is our hope that this work will encourage such efforts.

The aim of this work is showing improvisation to be a way of thinking appropriate for human-machine interaction rather than just a performance technique. The fluid nature of improvisation, with meaning created in a feedback loop between the participant and the machine (rather than in some pre-structured fashion) allows us to be creative not through planning but through action, observation, and reflection.

Artists of early 20th century took upon themselves to reconsider the role art had in the public life. It is now time for us to similarly expand the use of computational devices (including ordinary computers as well as auditory, visual, olfactory, and other sensory interfaces), taking our interaction principles away from fixed to an exchangeable (that is, multi-user / interconnected / peer-to-peer / distributed) environment with no unified plan of action yet a multitude of opportunities for context incorporation, multiple coexisting temporal structures and most importantly audience being the creators rather than listeners or viewers. It is our hope that developing improvisational approaches further will open a wealth of new opportunities for human creativity in entertainment and learning.

References

242.pilots, 2006. http://www.miau-miau.com/242pilots

The Avalanches, 2000. Since I Left You. http://en.wikipedia.org/wiki/Since_I_Left_You

Bailey D. 1993. *Improvisation : its nature and practice in music.* New York: Da Capo Press. xiii.

Bainbridge, B., 2004. Featured Artist of Vidvox. http://vidvox.net/artist-bcb.php

Barry, B., 2005. Mindful Documentary. Massachusetts Institute of Technology, Ph.D. thesis.

Blackwell, T., 2005. Swarm Music. http://www.timblackwell.com

Chadabe, J., 2001. http://www.chadabe.com/statement.html

ColorKinetics, 2006. http://www.colorkinetics.com/products/pro/index.htm#rgb

Cycling '74, 2006. Max/MSP. http://www.cycling74.com/products/maxmsp

Davenport, G., Morse, C., Murtaugh, M., Baird, F., Lachman, R., Cho, P., Tiongson, P., Stanley, L., 1996. "Jerome B. Wiesner, 1915-1994: A Random Walk through the 20th Century". http://ic.media.mit.edu/projects/JBW/

Davenport, G., 1981. RGB Piece. http://ic.media.mit.edu/people/gid/

Deleuze, G., Guattari, F., 1987. *A thousand plateaus : capitalism and schizophrenia.* Minneapolis: University of Minnesota Press.

Derrida J. 1976. *Of grammatology.* Baltimore: Johns Hopkins University Press.

DJ Danger Mouse, 2004. The Grey Album. http://www.illegal-art.org/audio/grey.html

Duchamp, M., 1913. Bicycle Wheel. Sculpture.

Emergency Broadcast Network, 2006. Website. http://fusionanomaly.net/ebn.html

Eno, B., 2005. http://music.hyperreal.org/artists/brian_eno/obliques.html

FinaleMusic, 2006. http://www.finalemusic.com/finale/features/educatortools/fpa.aspx

Kuleshov Experiment (1918). Wikipedia article, 2006. http://en.wikipedia.org/wiki/Kuleshov_Experiment

Lewis, G., 1997. Resonance magazine, Vol 6, #1. "Person to... Person?". Interview by Lawrence Casserley. http://www.l-m-c.org.uk/texts/lewis.html

Mazalek, A., 2005. Media Tables: An extensible method for developing multi-user media interaction platforms for shared spaces. Ph.D. Thesis, Massachusetts Institute of Technology.

Miranda ER. 2001. *Composing music with computers*. Oxford ; Boston: Focal Press. xviii.

Moroni, A., Manzolli, J., Von Zuben, F.,, Gudwin, R., 2002. *Vox Populi: Evolutionary Computation for Music Evolution* in Creative Evolutionary Systems ed. P. Bentley. San Francisco, USA: Morgan Kaufmann, pp. 205 – 221.

Nemirovsky, P., Burleson, W., Luger-Guillaume, B., Biderman, A., 2003-2004. Non-Standard Architectures. Exhibition in collaboration with the SERVO architectural collective. Centre Pompidou, Paris.

Nyman M. 1999. *Experimental music: Cage and beyond*. Cambridge; New York: Cambridge University Press. xx.

Oswald, J., 1985. Plunderphonics: audio piracy as a compositional prerogative. http://www.plunderphonics.com.

Pachet, F., 2002. The continuator: Musical interaction with style. In Proceedings of International Computer Music Conference, Goteborg, Sweden.

Paik NJ, Stooss T, Kellein T. 1993. *Nam June Paik : video time, video space*. New York: H.N. Abrams.

Papert, S. (1980). *Mindstorms : children, computers, and powerful ideas*. New York : Basic Books.

PG Music, 2006. Band-in-a-Box. http://www.pgmusic.com/bandbox.htm

Prévost, E. (1995). *No sound is innocent: AMM and the practice of self-invention, meta-musical narratives, essays*. Matching Tye, near Harlow, Essex, UK: Copula.

Russolo, L., "The Art of Noises", 1913. http://www.unknown.nu/futurism/noises.html

Schaeffer, P., 1952. A la recherche d'une musique concrète. Paris: Le Seuil..

SERVO, 2002. Thermocline. http://www.s-e-r-v-o.com/

Sims K. 1993. Galápagos. Interactive installation. http://www.genarts.com/galapagos/index.html

Sloboda JA. 2000. *Generative processes in music: the psychology of performance, improvisation, and composition.* Oxford [England] New York: Clarendon Press; Oxford University Press. xiii.

Sosolimited audiovisual collective. Botez, T.,Gunther, E., Manor, J., Rothenberg, J., 2006. http://www.sosolimited.com/

Stockhausen, K., 1956. Gesang der Jünglinge. A tape-loop piece.

Varela, F. J. (1991). *The embodied mind: cognitive science and human experience.* MIT Press, Cambridge, Mass.

Varèse, E., 1958. Poème électronique. An audio piece.

Vasulka W, Gogota H, NTT, Video Gallery SCAN. 1998. *The brotherhood.* Tokyo, Japan: NTT InterCommunication Center.

Waag Society, 2006. The KeyWorx platform. http://www.keyworx.org/

Warburton, D., 1996. Interview with Misha Mengelberg. http://www.paristransatlantic.com/magazine/interviews/mengelberg.html

Warhol, A., 1964. Campbell's Soup Can. Silkscreen on canvas.

Wright, M., and Freed, A., 1997. "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers." Proceedings of the 1997 International Computer Music Conference. San Francisco: International Computer Music Association, pp. 101-104.

Yzerbyt, V.Y., Lories, G., Dardenne, B. (Eds.) (1998). *Metacognition : cognitive and social dimensions.* Sage Pub: London ; Thousand Oaks, Calif.

Zappa, F., 1967. Lumpy Gravy. Audio Recording.

Zils, A., Pachet, F., 2001. Musical mosaicing. In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-01), Limerick, Ireland.

Zorn, J., 2005 - http://www.omnology.com/zorn05.html

List of Figures