MEDIA SCORES
A Framework for Composing the Modern-Day Gesamtkunstwerk

Peter Alexander Torpey

Bachelor of the Arts in Media Arts
University of Arizona, 2003

Master of Sciences in Media Arts and Sciences
Massachusetts Institute of Technology, 2009

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
September 2013

Author: Peter Alexander Torpey
Program in Media Arts and Sciences
9 August 2013

Certified by: Tod Machover
Muriel R. Cooper Professor of Music and Media
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by: Patricia Maes
Alexander W. Dreyfoos Professor of Media Technology
Associate Academic Head, Program in Media Arts and Sciences

MEDIA SCORES
A Framework for Composing the Modern-Day Gesamtkunstwerk

Peter Alexander Torpey

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

ABSTRACT

Media Scores provide a theoretical and technical means to orchestrate multiple modalities in the creation of expressive works of art and performance. New technologies afford numerous opportunities to tell stories and create expressive artworks through a variety of media. However, the tools of planning, composition, design, and creation of these works remain disjoint with respect to the artwork's constituent disciplines and from the final experience. Media Scores extend the metaphor of a musical score to other modalities in order to facilitate the process of authoring and performing multimedia compositions, providing a medium through which to realize a modern-day Gesamtkunstwerk. Research into the representation and encoding of expressive intent provides the conceptual underpinnings for the development of novel interfaces for composing and performing with Media Scores. Using such a tool, the composer is able to shape an artistic work that may be performed through human and technological means in a variety of media and utilizing various modalities of expression. Media Scores offer the possibility of authoring content that incorporates live performance data and the potential for audience participation and interaction. This paradigm bridges the extremes of the continuum from composition to performance, allowing for improvisatory compositional acts at performance-time. The Media Score also provides a common point of reference in collaborative productions, as well as the infrastructure for the real-time show control of any technologies used during a live performance.

Thesis Supervisor: Tod Machover
Muriel R. Cooper Professor of Music and Media
Program in Media Arts and Sciences
Massachusetts Institute of Technology

MEDIA SCORES

A Framework for Composing the Modern-Day Gesamtkunstwerk

Peter Alexander Torpey

The following person served as a reader for this thesis:

Thesis Reader Ken Perlin
Professor of Computer Science
Media Research Lab
New York University

MEDIA SCORES
A Framework for Composing the Modern-Day Gesamtkunstwerk

Peter Alexander Torpey

The following person served as a reader for this thesis:

Thesis Reader — Alex McDowell, rdi

Associate Professor of Cinematic Arts
Interactive Media Division
University of Southern California

ACKNOWLEDGMENTS

Tod Machover

Thank you for welcoming me along on this remarkable six-year portion of the journey. I am ever-grateful for your guidance and the opportunities you have afforded to me to participate in groundbreaking projects, to leverage my skills, expand my horizons, to grow, and to create.

Alex McDowell
Ken Perlin

I am most grateful to my dissertation committee for your time, expertise, and wisdom. Our numerous discussions over the years have been exceedingly helpful in refining and broadening my ideas about the future of storytelling and the arts. Much of that has found its way into Media Scores and this thesis and will surely continue to inspire my future work.

Neri Oxman

Thank you for having been a part of this process. Your unique and varied insights and suggestions have proven to be an important part of the ideas explored in this thesis.

Mark & Naomi Torpey

Nothing can parallel the love, devotion, and unwavering support that you have always given me. By your example and because of your encouragement, I am empowered to reach for my dreams.

Elena Jessop

Through all of the projects, the furies of creation, the obstacles, and the celebrations of accomplishments that we have shared over the past several years, I am very much indebted to you, your friendship, and support.

Ben Bloomberg
Simone Ovsey
Akito van Troyer
Eyal Shahar
Janice Wang
Rébecca Kleinberger

I have had the great privilege to work alongside such a talented and creative group. You have encouraged and inspired me with your and ideas and creations. More importantly, I've been most fortunate to have you as my friends.

Eleanor Olvey — Your constant friendship throughout the years has been a comfort and reassurance.

Catherine Redfield
Sharon Chou
Noah Feehan
Cubie King, jr.
Andrea Dubin
Daniel Kohn — My dear friends beyond the Lab have provided essential companionship and diversions over the years.

Noa Machover
Edan Krolewicz
Helena Zay
Jason Haas
Emily Su
Danielle Man — *Figments* would not have been possible without the additional contributions of several people. My talented rehearsal cast helped develop the piece and the performances that would ultimately be staged. Others helped me to shape the story, the technologies, and construct the performance infrastructure. It was rewarding to have such a team to bring my ideas before an audience.

Punchdrunk
Toronto Symphony Orchestra
MIT Music and Theater Arts
MIT Musical Theatre Guild
ArtEZ Institute — Some of the work presented in this thesis was conducted in collaboration with organizations worthy of special acknowledgment. From material support to unique opportunities to explore innovative ideas, I am grateful to have worked with each.

CONTENTS

LIST OF FIGURES

All figures were created by the author unless otherwise noted in captions.

1 INTRODUCTION

> *In the impossibility of replacing the essential element of color by words or other means lies the possibility of a monumental art. Here, amidst extremely rich and different combinations, there remains to be discovered one that is based upon the principle just established. I.e., the same inner sound can be rendered at the same moment by different arts. But apart from this general sound, each art will display that extra element which is essential and peculiar to itself, thereby adding to that inner sound which they have in common a richness and power that cannot be attained by one art alone.*
>
> —Wassily Kandinsky

The theater, the concert hall, the busker on the street corner all draw us in. These spaces are transformed through the magic of performance and we, the privileged audience, are granted a few moments to lose ourselves in another world. We are amazed, saddened, empowered, and freed by the moving portrayals, the soaring melodies, and the unexpected taking place right before our eyes. Works of art move us, no doubt. We adorn our walls with imagery and color. We trek for miles to sit before a renowned statue seen only in photos since childhood. We stand in awe at the threshold into great works of architecture.

We revel in the value of the talent, the message, the intention that goes into creating a work of art. We trust the blend of intuition and skill that an artist applies to his craft in that mysterious way he can transform us and make us feel something particular that transcends mere words. The works of art that draw on more modes of expression, that enlist more of our senses, that give us the illusion of stepping into someone's dreamed world bring together the talents of many and a broad range of abilities and ways of creating meaning. These complex works are not the results of a single hand, the way a sculptor or novelist can build their world for us in isolation, but of many constructing and performing an artist's idea. As audiences and spectators, we play our role in the construction of meaning in art and, as new technological practices transform and comingle the artistic disciplines, our role grows along with the capabilities of the artist to shape new worlds, new messages, and to invite us in.

1.1 The Art-Work of the Future

When Richard Wagner introduced his bold notion of the *Gesamtkunstwerk* in 1849, an operatic art form that subsumes all other art forms in the service of telling a story [112], it paved the way for the continued growth of opera as a pinnacle of human expression. Still young, when compared to the over 4000-year history of theater, it can be argued that opera is still finding its footing and exploring the frontiers of theatrical practice. Wagner's later operas, or "dramas" as he came to prefer calling them, strove to realize his ideal of Gesamtkunstwerk, as have the stagings of his work in the many years since their premier. For example, director and theatrical innovator Robert Lepage has garnered much attention for his technically elaborate productions of Wagner's *Ring Cycle* at the Metropolitan Opera in New York City, pushing the boundaries of stage machinery, automation, and interactive visual design in an attempt to realize the fantastical and cinematic scope of Wagner's scores [105]. Indeed, the score for these works supplies not only the music, the backbone of the operatic form in which they are set, but also the design and staging [110]. Wagner sought to create not only the music and the text of his dramatic masterpieces, but also to define the entirety of the world in which these stories unfold to be presented before the audience.

Figure 1: Page from the score for *Götterdämmerung*
Wagner's scores combine music notation with descriptions of setting and staging (detail below). (Reprinted from [110])

For Wagner, the Gesamtkunstwerk is the union of Hegel's three humanist arts—dance (movement), tone (music), poetry—and the three plastic arts—architecture, sculpture, painting [40]. In more recent times, the named artistic disciplines have been amended to include photography, comics, gastronomy, cinema, and theater. The latter two appear by virtue of Ricciotto Canudo's argument for the convergence of the plastic and time-based arts as descriptive of new forms [20]. It is in this synergistic spirit that

Gesamtkunstwerk would be the art form that draws from all of these modes of expression and, indeed, the whole of the numerous human faculties, mathematical, mechanical, and technological arts.

Today, Gesamtkunstwerk most readily describes the media of musical theater, cinema, and video games. In Wagner's vision, all of the constituent arts would serve in perfect unison, each keeping the others in harmonious balance and none overbearing. However, in each of these media, there is an imbalance that lets one or more arts or modes dominate over the others. The lasting convergence of these media is imminent, perhaps then realizing a pure form of Gesamtkunstwerk. Nevertheless, they are all essentially multidisciplinary art forms that strive to tell a story, in some cases interactively, in a rich and immersive way. They employ a synaesthetic blend of the humanistic and plastic arts in the service of relating a story and at a complexity such that a team of collaborators is necessary to make the telling of the story a reality. Even with the massive collaboration required, there is an author, a director, a composer that sets forth the vision for that telling, as Wagner did for his operas.

Technology introduces new modes for storytelling, but is also the connective tissue among the arts that constitute a Gesamtkunstwerk. All of the modalities of representation must work in concert and communication is critical among the designers each responsible for their own discipline to ensure that the resulting work of art represents a unified whole that is faithful to the composer's intent and the artist's and director's visions. By linking the different aspects of a production, as well as introducing new ones, it underscores the importance of communication and the causal relationship of changes in one aspect impacting others in the course of developing a work.

As these art forms coalesce, as we consume art and story in a multiplicity of media, if we define the message to be the creator's intention, it follows that the modern-day Gesamtkunstwerk need not be tied to a single medium or type of experience. Can a work of art exist independent of its form? Can a composer's intent be written in such a way that it can be performed or appreciated in any number of different media? How can new works of art satisfy our natural craving to assume, from time to time, an active role in our experience of art? How does one create a composed work of art that can also be personal, interactive, and responsive?

1.2 Objectives for Media Scores

Although various disciplines have evolved methodologies to create or shape individual parts of a production, none of them serve as a unified, expressive method to allow an individual creator, group of collaborators, or a creative team to compose at once the content and design of a Gesamtkunstwerk. Additionally, these methods serve to describe the formal properties of their particular medium, envisioning a singular final output. There has been no established way in these practices to permit and describe the intention of the work, regardless of the medium, or the audience's potential involvement in experience of the work.

The work investigated in this thesis aims to provide authors, directors, designers, and composers with a vital tool that enables them to create and shape the telling of their story. I leverage research in the areas of perception and cognition in the design of an environment that can translate expressive representations across modalities and to provide an environment that can afford creators with the necessary power to sculpt the many facets of storytelling media. This tool will take the form of a *Media Score*, a visual and semantic notation that draws on the metaphor of a musical score and that aims to encode expressive artistic intention of the composer of a Gesamtkunstwerk.

If we consider the Media Score as a representation of the composer's artistic intent, then the Media Score should be an aesthetic artifact like any other possible representation. The Media Score is simply one representation of the same intent that a full instantiation of the Gesamtkunstwerk depicts. Practically, it is a sketch of expressive parameters with a semantic relation that must be crafted to encode the intent of the artwork. In the score, the composer can shape these parameters, define how they interact, and annotate them with events and assets that inform the performance of the score.

I propose several additional objectives for Media Scores that position this form of notation and creation to facilitate the increasingly complex task of realizing multidisciplinary works of art in time-less and performance contexts. The score language itself should be medium-agnostic with respect to its final performance. Media Scores should make no implicit assumptions about how a work of art will be presented and, in doing so, opens the possibility for many possible presentations of a work—as decided by the artist, producers of the work, or even the audience—in terms of medium of experience and duration of experience. Media Scores should also be collaborative references for works that have multiple authors or instantiations of the work that require multiple individuals to produce and perform. The computational nature of an implementation of a Media Score, as presented in this thesis, means that the Media Score not only serves as a common language for describing the work itself, but as the language or heart of any performance technologies that are used in the production of the score. With such capabilities, the score should also provide composers with techniques for incorporating a degree of interactivity in their work and ensuring that interactive variability is consistent with the authorial intent. With the breadth of applicability, Media Scores becomes an essential tool during all phases of the production of a work. These objectives will be revisited in greater detail in Section 4.2.

1.3 Reflections on a Journey

This thesis, like any, stands as a milestone in one's academic and professional journey. The work presented herein is a culmination of my six-year tenure as a research assistant in the Opera of the Future group at the MIT Media Lab. More than that, it is a waypoint on a meandering path that began as soon as I was able to hold a crayon and that has taken me through a youth exploring visual arts, traditional media, the stage, photography, and music. A fascination with the sciences was fused with the act of artistic creation, which led me to study film, video, and new media. Eventually making my way to the Media Lab, I now find myself firmly entrenched in the realm of theater, live performance, and unique multimedia experiences, all the while drawing from the myriad of skills that I have cultivated in all of these disciplines.

Consequently, the work summarized in this thesis documents particular aspects of a selection the many projects I have had the good fortune to be a part of during the past six years. In particular, I will describe the creative challenges encountered and techniques developed that have evolved into the formulation of Media Scores as a practical approach to address the needs of contemporary innovative performance works.

1.4 Culmination of Idea and Practice

The vision of what would become Media Scores began to crystalize during conversations I had with production designer Alex McDowell during our work on the opera *Death and the Powers*. We identified a need for a visual language to represent the complex and interrelated design considerations during the development of the project. McDowell noted that no existing software tools begin to address all the needs of design, planning, and collaboration across multiple departments. From his experience in the film industry, he noted some of the deficiencies in workflow and conventional practices that he has encountered and that could be resolved by new software tools. From these conversations, I began to experiment with representations that would address these concerns for communicating during the design process for *Death and the Powers*. This and other projects that lay the conceptual groundwork for Media Scores are presented in Section 4.1.

In designing the implementation of Media Scores as discussed in the remainder of Chapter 4, I considered my own creative workflow when designing interactive visuals or theatrical lighting, for example. I tried to imagine a tool that I would use to facilitate my creative processes in these diverse scenarios. Along the way, this consideration merged with ideas like Disembodied Performance (see Section 3.1.1), the desire for a formalization of mapping expression across modalities, and a personal fascination with notation systems. As the concept took shape, the breadth of capabilities that such a tool could serve during the creation and production of a Gesamtkunstwerk became apparent and these additional directions were incorporated into the system's design, so that it functions beyond the role of simply a reference document for production.

1.5 Terminology

Before we commence in earnest with a discussion of Media Scores and the research at hand, it is apropos to clarify the sense of several terms I will use regularly throughout this thesis. There is great flexibility in the way these terms may be defined, so I present operational definitions for the purpose of this document in order to minimize ambiguity.

Modality and Medium

I use the term *modality* to refer to a particular manner of sensation or perception or a process that produces information content for perception in a particular sense. As examples, the visual, auditory, or tactile domains constitute different modalities as their corresponding stimuli are sensed and interpreted through the senses of vision, hearing, and touch, respectively. However, I would also consider music or the written word as independent modalities. Even though these two examples rely on the sensory mechanisms of hearing and vision, they become meaningful at higher perceptual levels, as distinct from simply sounds and shapes. Other distinctions of modalities might include still versus moving images, representational versus abstract images, and so on.

In contrast to modality, I use *medium* to suggest an established artistic context that is comprised of one or more modalities and particular circumstances of consumption. The medium of literature relies on the modality of the written word and a prosaic form of discourse. The medium of theatre would minimally employ, per traditional definitions, the modalities of spoken text, gestural movement in a space, the visual design and composition of the space and actors in the space. By these definitions, *multimedia* and *multimodal* would have comparable meanings.

Parameters

I shall use the term *parameters* specifically to mean values that may vary over time and have well-defined semantics. It should be noted that this usage does not reflect the typical definition in mathematical contexts. Usually, parameters are values that configure a function or a mathematical model to produce a result given some set of variable arguments. In most cases, my usage of parameters will refer to data sets or streams of continuous input—such as live performance data sensed from performers or audiences—that could represent the output of a generating process or input to a function or system. This usage more closely aligns with the arguments to parametric equations in mathematics or parameter arguments in computer science.

In the discussion of artistic modes of representation, I will distinguish between *formal parameters* and *expressive parameters*. Formal parameters represent some feature specific to a medium of representation. For example, formal parameters of music include pitch, tempo, and amplitude, among others. In the visual domain, some formal parameters are hue, length, and thickness. Expressive parameters represent more abstract concepts that are perceived at a higher level. They are not morphological features of a medium, but may be aggregates of formal parameters or directly correlated to a formal parameter. They are related to the intention of a communicative gesture in some medium or the experience of that gesture, but are not specific to any particular medium. Expressive parameters can be defined at various levels of cognition or understanding, from simple concepts such as *complexity* or *intensity* that are closely related to the formal parameters in a specific medium to high-level concepts such as *doubt* and *fear*.

Parameters play an important role in the conceptualization of Media Scores and other systems. Affective parametric spaces are discussed in Section 2.3, as an example of parameterizing expressive content. Expressive performance parameters play a key function in many of the systems described in Chapter 3. Expressive parameters are the essence of Media Scores and a consideration of their properties and semantics will be treated in greater depth in Section 4.3.2.

Composer, Author, Artist, and Creator

Media Scores finds applicability in the hands of creative practitioners, whether they are generating a new work of art or working, often collaboratively, to realize a new or existing work of art. As such, for the purposes of the discussion of Media Scores, particularly from Chapter 4 on, I will often use the terms *composer*, *author*, *artist*, and *creator* interchangeably, as these are all titles for the role of one who generates a work of art. As opposed to sculptor or photographer and so on, the names "composer" and "author" have the connotation of creation through notation, be it the script for a play or film, a novel itself, or a musical score.

Thus, it is appropriate to use them for the creators of Media Scores. Of course, with respect to Media Scores, I do not intend to convey the particular media with which these terms are usually associated. The term artist is more sufficiently generic and medium agnostic.

Conductor, Director, and Designer

Similarly, *conductors* and *directors* are the individuals responsible for interpreting a notated work of art through performers or other performance apparatus. *Designers* are also responsible for interpreting the notated work of art, but generally not through performers. Individuals in these roles may work with a Media Score that has been created by the artist. Whether or not in conjunction with the creator, this set of individuals would likely form the core of a creative team that would use Media Scores as a collaborative document.

Performance

I define *performance* to be the realization of any shaped work of art for an audience member in a spectator or participatory role such that the realization is unique to that particular viewing of the work. Oftentimes, in the case of theater, dance, or music performance, the variability of each instance of the work comes from the imprecision or adaptability of human performers. Elements of physical chance, internal emotional states, and audience reactions can all contribute to some degree of change in a performer's intended behavior from one performance to the next. This gives theater its ephemeral quality: its *liveness*. However, by my definition, performance is not limited to variability introduced by human performers. Aleatory and pseudo-random processes, the incorporation of external time-varying data sources, and other forms of audience interaction can also yield a performance in media that is otherwise precisely reproducible. Thus, an entirely computer-generated audio or visual artwork may constitute a performance, if it is uniquely situated in time and responds to the conditions of the time in which it exists. On the other hand, a film in the common traditional sense may involve human performances and itself be an act of Gesamtkunstwerk, though each screening of the finished film does not constitute a performance.

Gesamtkunstwerk

Ideally, an instance of *Gesamtkunstwerk* is a unification of all creative and expressive disciplines. For Wagner, a Gesamtkunstwerk was the union of the three humanist and three plastic arts. However, one could extend the constituent art forms to include many other endeavors that result from the human capacity for creativity. I will, however, use the term in a less restrictive form to refer to artworks that employ in concert a significant number of complementary modalities. This definition is inclusive of musical theatre, film, video games, and other aggregate media that rely on complex relationships among modes of representation.

1.6 Thesis Structure

This thesis comprises six chapters, including this introduction.

Chapter 2 brings together a diverse set of disciplines ranging from perception and information theory to abstract art and visual music in order to lay the theoretical foundations upon which Media Scores can stand. In this chapter, I will also engage in a discussion of the projection of time as a fundamental concept for the representation and notation of artistic experience. I will also survey conventional notation practices employed in the creation of scripts and scores for performance works and how they have been extended by practitioners in the 20^{th} and 21^{st} centuries, in order to achieve additional flexibility, deliberacy, and greater expressive power.

In Chapter 3, I describe several diverse projects to which I have contributed during my time at the Media Lab. In doing so, I focus on the unique systems for control and content creation that I created for each in order to meet the demands of technologically-mediated multimodal representation and authoring in technically complex productions. This leads to a set of design criteria for technological performance systems. The needs of these projects motivate the creation of Media Scores.

Chapter 4 brings the theoretical and historical perspective together with the lessons learned in creating performance systems. Coupled with the motivation for this work of creating composition tools for Gesamtkunstwerk, an implementation of a Media Scores authoring and playback environment is presented. Initial experiments and ideas that led to the genesis of Media Scores are presented, along with a set of design goals that address the needs of creative practitioners for multi-modal composition tools.

Example projects that used Media Scores both as a guiding concept and the current implementation are presented in Chapter 5. These projects will explore the scope in which Media Scores may be applied and evaluate their role in the ideation and generation of artworks.

This thesis concludes in Chapter 6 with a reflection on the work presented and a discussion of how Media Scores could impact traditional practice, evolving roles of personnel and technological systems, and the future of performance and the arts.

2 FOUNDATIONS AND PRECEDENTS

What's best in music is not to be found in the notes.
—Gustav Mahler

As a method to convey expressive detail about the intended shape of a work of art, a Media Score must represent that intention in a form other than the final work of art itself and it must also reflect exactly what constitutes expressive shape. Thus, to embark on the creation of Media Scores as a means to notate artistic intention, we must begin by exploring answers to the question: How is expression encoded in a work of art?

In this chapter, we will take a brief look at a number of different subjects in order to construct an argument for the types of abstract representation that make Media Scores possible. This exploration will also inform how Media Scores can function as a representational tool, positioning it as a novel approach to the composition of Gesamtkunstwerk.

2.1 Perception, Representation, and Mapping

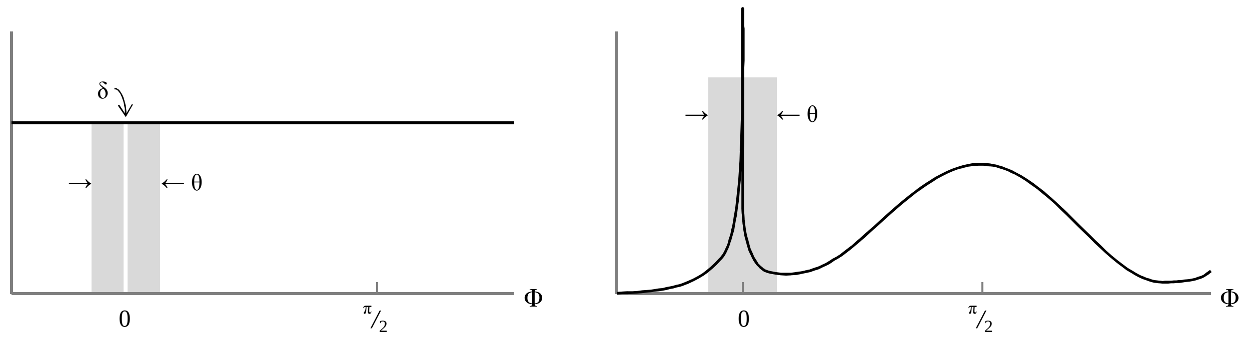
We will begin by considering the process of understanding sensory stimuli. This understanding is, in effect, a mapping from sensation to perception. When we look at a work of art, hear music or the spoken work, or interpret a diagram, our senses are transducing signals from the external world into representations within our mind about which we can reason. The method of transduction is not of interest for our present purposes, but rather the way in which we come to understand the meaning of the frequencies, scents, patches of light and dark that we sense. We're not concerned with the details of shape and form in our search for meaning, but an understanding of what the impression suggests about the state of the world, why it exists, and why it is being presented to us at a particular moment.

Relevant research into understanding perception has been done in the field of computer vision. Witkin and Tennenbaum posit that perception is the process of explanation [114]. When presented with a set of stimuli, the mind attempts to identify the process that generated those stimuli based on cues specific to the modality perceived. The percept is the most likely assumption, at an unconscious level, of the reason the stimuli were produced. Multistable perceptual illusions arise when the stimuli cannot be reduced to a single generating phenomenon [87]. The reason that our minds can apply heuristics in parsing a sensory scene is that the external world is both consistent and "non-accidental", to borrow Binford and Lowe's terminology. If any possible stimulus had an equal likelihood of occurring, i.e. if it were accidental, no generalizations could be made about the cause of that stimulus nor could any predictions be made about the meaning of a stimulus or the likelihood of any stimulus occurring at some future point. A random world—or specifically, a random set of sensible features—contains no meaning, as in the random collection of lines in Figure 2 top. Fortunately, the world is not random. Perceptually salient features often occur in modal distributions, which our mind relies on to arrive at a meaningful percept (Figure 3) [88]. In Figure 2 middle, we can immediately tell that the lines are arranged in a non-accidental manner. In fact, the percept is not that of lines, so much as it is of a representation of objects arranged in physical space, and even more specifically of buildings with windows and a railing. Our perception has done the work of not merely sensing the existence of lines, but generating a plausible explanation of what objects in the world would generate those lines. What we perceive are the buildings. When processing the original stimuli (represented photographically in Figure 2 bottom), the perceptual system also is responsible for feature extraction of the lines that generate the same perceptual explanation. Other formal parameters of color and texture further contribute to our understanding of the scene: a stone building, a brick building, leaves, grass, and so on.

Figure 2: Random and ordered worlds
A random assortment of lines with uniform distribution (top)
An assortment of lines with modal distribution representing the physical world (middle)
The scene represented by the ordered lines (bottom)

The principles of Gestalt psychology also play a role in our perception and understanding of stimuli by suggesting higher-order structure among perceived elements. The Gestalt theory codifies cognitive mechanisms or heuristics, such as common fate and constancy, that the mind employs when trying to make sense of a set of stimuli [51]. These phenomena can be leveraged in the creation of a meaningful representation and are often perceptually accessible in different modalities.

Representation is fundamental to communication. A work of art seeks to convey some story or feeling held within the artist's mind to an audience in order to evoke the intended feeling or understanding. It is this communicative process of transferring an idea from one mind—that of the artist—into the mind of others—the audience or spectators—that poses the fundamental challenge in the creation of an expressive work of art. How



UNIFORM DISTRIBUTION

MODAL DISTRIBUTION

Figure 3: Uniform and modal distributions
Considering the orientations of lines and edges in representations of the built environment (Φ), modes of the distribution occur for parallel ($\delta = 0$) and near perpendicular lines ($\delta = \pi/2$).
(Redrawn from [88])

can this intent, this complex idea, be copied from one mind to another? The work of the artist is to encode the idea into some intermediary form that exists as a vessel outside of the mind and can be absorbed by and reconstituted in the mind of the audience. This vessel is the work of art. However, there is no sufficient medium to represent explicitly the structures and concepts that can be held in the mind. Thus, the work of art is a reduced and incomplete form that, if done well, captures enough of the intent that it can produce a similar effect in the observer of the work. It cannot exactly copy an idea from one mind to another, but it can often produce a reasonable approximation. The ability to extract and convey the salient features of a represented concept, foregoing the noise of formal artifacts that do not clearly reinforce the intention and can detract or add overhead to reasoning about the concept, makes for an appropriate means of representation [79]. The choice of medium, semiotics, and formal encoding of an idea in the physical world is a problem of representation.

The mapping process to and from the intermediate medium must also be considered. In traditional artistic practice, this is primarily left to the intuition of the artist for encoding the intent and the intuition of the observer to decode it. An artist's skill is in knowing how to encode the message such that the common faculties of the consumers of the work will interpret it sufficiently. G. Spencer Brown, as quoted in [70], summarizes:

> The composer does not even attempt to describe the set of feelings occasioned through them, but writes down a set of commands which, if they are obeyed by the reader, can result in a reproduction, to the reader, of the composer's original experience.

Although this process is intuitive for both parties, it is studied at length by artists and scholars, distilled into rules and guidelines and stylistic conventions which can be learned to inform the work of other artists, adapting practice for their own intuition and unique message [10].

In the technological realm, the mapping process is pervasive at all levels of representation and translation of intent. It is effectively a mathematical function that must take as input a representation—for our purposes, some gesture of the artist's intent—and, by manipulating the data vector, produce some analog or digital representation that is then passed through another function to create the output that audiences will perceive. The function itself may be linear or non-linear. The dimensions of the data may be mapped in a dependent or independent way. However, this conception has the tendency to yield arbitrary transformations. The formal parameters of one modality are mapped to the formal parameters of another modality without a means of ascribing semantics to the intentional concepts. A mapping designer may apply intuition to the process, but this view does not provide a criterion for what constitutes a good mapping or how to deliberately create mappings that are meaningful and not arbitrary.

In my master's work, I introduced the concept of *reified inference* as an alternative to linear mapping as methodology for deriving an output representation in the same or in a different modality than some original input representation [107]. Reified inference is inspired by the notion of perception as the explanation of a stimulus by surmising the generating process that created the stimulus. It attempts to derive an intermediate model from the input that represents the generating process. Formal parameters in the input medium are first mapped into the metric space of the model and then, from that model, mapped into the formal parameters of the output medium. The model is an approximate description of the generating process and, therefore, represents the semantically meaningful facets of expressive intent. A sufficiently accurate model is then capable of producing novel output representations of the generating process, regardless of medium, as well as recreating the input data. This provides a useful schema for creating mappings and a method for validation.

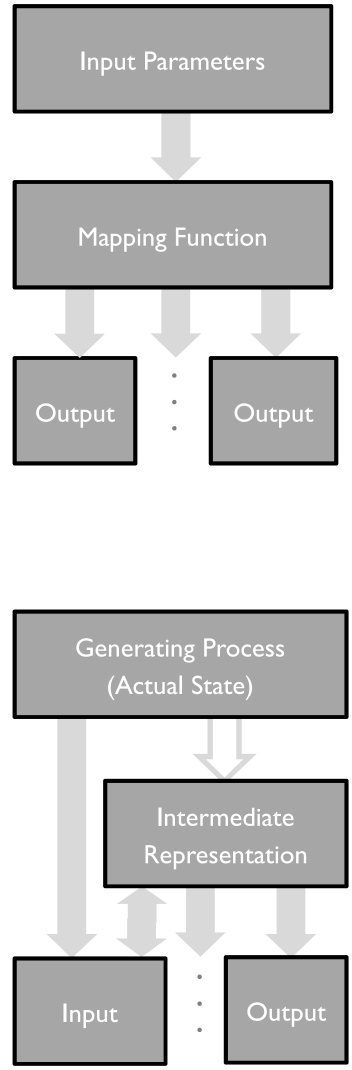


Figure 4: Mapping and reified inference
Traditional mapping procedure (top) and mapping through reified inference (bottom)

Most systems or datasets are too high-dimensional to be transformed through a constructed mapping. The number of dimensions required to completely model stimuli both by their formal properties and their semantic explanations is unwieldy. As an analogy, let us consider the process of modeling a digital color image. The image is represented in the computer as a two-dimensional grid of three-tuples representing the red, green, and blue components of each pixel.

$$I(x,y) = (r,g,b)$$

This representation describes the image, but not the world, i.e. the generating process, that it represents. Viewing this image on a display or printout, the human mind and perceptual system can analyze and make inferences about what is represented. We can describe not only the image, but the light field using the plenoptic function, therefore representing the set of light rays impinging upon a focal point.

$$L(x, y, z, \theta, \varphi, t) = (r, g, b)$$

The plenoptic function goes further, describing the light in the physical world that an image captures. However, the dimensionality of the domain increases. This function will describe all possible images at a given point in space, orientation, and time, but it doesn't encode any further semantic *explanation* of the image. There's no notion of objects or contours or what is being depicted in the image and why. This is the role of perception. If the model were expanded to include the type and state of all objects within the scene, their locations, and orientations, the description of the image would more closely represent the scene which was mapped into it. Again, the dimensionality of the representation would grow astronomically. As a representation increases in absolute fidelity of that which it represents, then it becomes identical to reality and is no longer a representation. However, if we can make a parameter space that is descriptive of the *what*—the intention or causal circumstance—not a simulacrum of the formal properties intersecting with a sampled instance, then we can capture the story or situation that generated the image.

Thus, any act of representation, and therefore mapping, must be one of dimensionality reduction. This incurs a loss of information, which ideally, with an appropriate model, is not information relevant to the encoded intention. The model underconstrains the set of possible output mappings. When expanded to the dimensionality of the output medium, additional information may need to be generated to compensate for the initial loss and any mismatch of dimensionality between the input and output representations. This leads to a multiplicity of possible output variations for a given model state, though the influence of the model semantics is preserved.

Media Scores rely on the mapping of expressive content across media and modalities. The score itself is a representation that must both store the expressive intent and allow it to be understood and manipulated. The process of composing is one of mapping the intent into the score form. The score then must be mapped into the final artwork that will be consumed. The mapping processes convert intangible ideas into one set of parameters in a fixed number of dimensions, the score, and then into another set of formal parameters, those of the medium in which the work is realized. The applicability of the reified inference model and the process of modeling appropriate parameters is discussed in Chapter 4. We'll also consider later in this chapter and in Chapter 6 how the underconstrained nature of modeling artistic expression is an important feature of generating artworks.

2.2 Expectation and Information

We've seen how certain perceptual features exhibit an intentional or non-accidental salience. This aids in the determination of which features of a particular modality should be examined in a meaningful representation. However, this is only the beginning of the process. In order to preserve expressive content through the mapping of one modality into another, often when they have disjoint feature sets, we need to look at the continuous changes in the values associated with these features. This situates the mapping process in a parametric space that is independent of the input and output modalities. It is in this space that we can begin to model expressive function and, in turn, intention rather than creating a model that reproduces reality.

As a prime example of a medium that is both expressive and inherently abstract in nature, I look to music theory and analysis to shed some light on how expressive intent is encoded in a set of features. First, let us take a moment to consider how music is represented and reasoned about in terms of parameters that, while still formal in nature, describe the medium in an intentional manner versus a reproducible image. Music theory operates on the level of musical parameters of notes and rhythm and harmony, the concepts that we will see later are captures in conventional musical scores, rather than sound waves in time, as would be the case in the image representation example above. An image of music is an audio recording and must be experienced in time in order for the musical features to be perceived from it. Reasoning about music at the level of frequencies and wavelengths. which are the substance of the stimulus as we sense it, is unproductive. Rather, we favor a model that more closely represents the features we experience in time and from which we can infer expressive intent.

In music, more so than other non-figurative art forms, there has been considerably more research into the constituent perceptual parameters and their potential expressive interpretation. The parameters do not exist in isolation, but interact in contradictory or reinforcing ways and have varying weights of contribution to the overall assessment of the musical piece. For example, expression in music can be modeled as degrees of tension and resolution in an ensemble of formal parameters over the course of attentional windows at different scales [34]. The moment-to-moment response is the result of a continuous integration over this memory window to produce a net effect over both time and the ensemble of parameters. At a slightly higher level, Daniel Levitin argues that the patterns of tension and release can be thought of as establishing expectations and intentionally violating them for expressive effect [56]. Levitin proposes that the emotional responses from such violations result from a dopamine response caused by the mesolimbic system, which is involved in reward and pleasure responses.

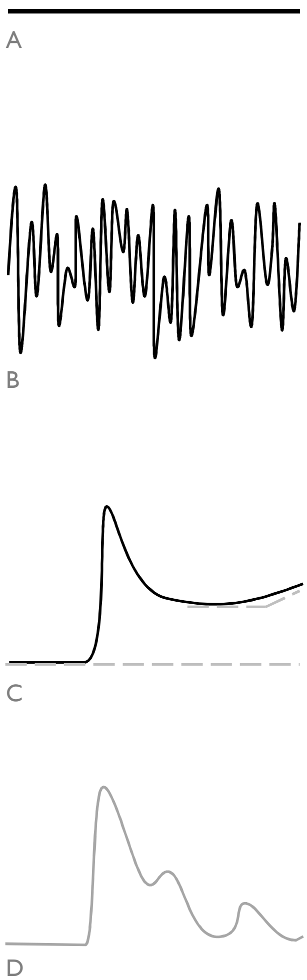


Figure 5: Expectation curves
A. A constant parameter contains no information.
B. A random parameter contains no information.
C. An intentional change in value produces information, as evidenced by having a similarly meaningful derivative (D).

David Huron provides a more nuanced description of the role of expectation in music by proposing that the processes of assessing and evaluating the success or failure of a prediction about what will occur gives rise to the sense of emotion. His ITPRA theory divides the evaluation process into five phases [45]. The *imagination response* and *tension response* occur in advance of the event in question. The former is a slow-timescale prediction of the outcome of an event. The latter *tension response* is an anticipation of an event. Immediately following the event, the *prediction response* evaluates whether or not the expected outcome occurred, positively reinforcing an accurate prediction and negatively reinforcing an incorrect one. The *reaction response* occurs simultaneously and is a reflex that assumes a worst-case, failed prediction regardless of the actual event. The appraisal response is a longer-term evaluation of the success or failure of the expectation. While Levitin's suggestion most cleanly maps to the prediction response, Huron demonstrates that the reward and negative reinforcement mechanisms are likely considerably more complex.

The investigation of the emotional response to music, particularly when coupled with neurophysiological explanations, begs the question of evolutionary theories of music. That music functions emotionally could be argued to be the exaptation of processes that have evolved to reinforce an individual's ability to accurately predict the outcome of survival-related events. Marvin Minsky suggests that music may have been developed as a tool to hone predictive abilities by segmenting and differentiating states in time [69]. There is considerable debate over the evolutionary role of music in human society and the role in expectation is only one possible function. Although much of this research is focused on musical experience, I believe that these ideas, regardless of their evolutionary function, are generalizable to other media, given an appropriate selection of formal and perceptual parameter semantics.

What these models of perception and expectation provide is a scaffolding for identifying and quantifying expressive content in abstract, non-figurative representations. Information theory provides a lens through which to consider and understand the significance of change in these parameters. I relate the information theoretical notion of *entropy*, as defined in Shannon's treatise *The Mathematical Theory of Communication* [95], to analysis of expression. Specifically, an instantaneous or windowed entropy metric would seem to reveal the encoding of information content as expressive gestures in arbitrary values. In this light, it is clear that information, and thus expression, is non-existent in the cases where the formal parameters are either unchanging or completely random (Figure 5). An incessantly repeating value, a constant value, is predictable *ad infinitum* and, thus, its repetition or persistence is meaningless. Similarly, as we saw in the previous section, a value that varies randomly with a uniform distribution, and therefore contains no modes or non-accidental features, is similarly predictable in the statistical sense. In between these two extremes, changes are non-uniform and unpredictable. Shannon indicates that such uncertainty is essential for a representation to encode information. A meaningful representation must be intentional and, therefore, represent the significant qualities we wish to preserve during mapping.

To see how the dual requirements of non-accidental and unpredictable variation are essential for meaning-making, let us consider a narrative plot. In a narrative story, consistency arises from causality and character motivation. The causal linkage between events makes them non-random, though we cannot reasonably predict the events of the story or else we would not want to read or watch it. An understanding of character motivations, although often implicitly evaluated with respect to our own human nature, provides consistency for the character's actions and choices and how events are set in motion. The interaction of distinct motivations of multiple characters or the reaction of the story world adds the necessary element of uncertainty.

The net effect of the information encoded within the piece is, as noted earlier, a form of tension and release. We feel anticipation of an outcome and the resolution of a correct prediction or a heightened state from an incorrect one. In time-based scenarios, this ebb and flow of information-expression is familiar to us. If we think about hearing an unexpected sound, say the strike of a chime, we have an immediate response that rises proportionally with the amplitude of the sound. Once heard, we attend to the sound while the bell rings, but as the amplitude decays, we return to or less aware state. In audio synthesis, the amplitude of the sound is modeled by an attack-decay-sustain-release curve that very closely parallels our low-level emotional reaction to the sound itself (Figure 6). This example is an oversimplification of expectation and affective response to a stimulus, but serves to show that we are familiar with reasoning about these types of

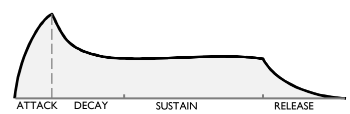
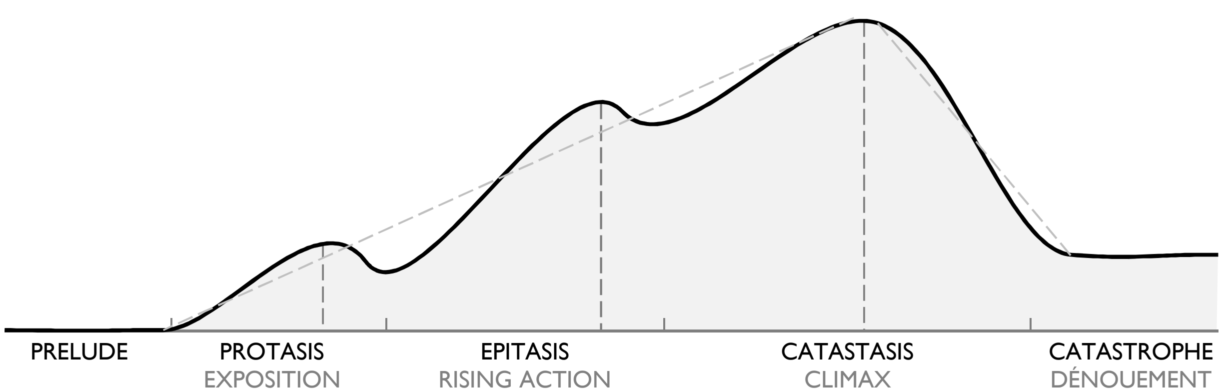


Figure 6: ADSR envelope

Figure 7: Narrative arc



events and reactions. Not only are we familiar with their form, but also their visual representations. Consider the triangular shape of a story that we learned in grade school (Figure 7). The story has a beginning, middle, and end. After the beginning, there's rising action and conflict that build toward a climax at the apex of the triangle, and then descends as the story resolves. The three- or five-act narrative structure that this triangular curve represents comes from Horace [44] and later Freytag [36]. We recognize it as the shape of any good story, due both to having been taught it as such, but also because it does reflect our *feeling* of a story. What really does it indicate? What is the *shape* of a story? This triangular shape is actually a plot of the tension over time during the course of the telling of the narrative. Furthermore, this is the overall, net shape a story takes. It is generally a function of multiple such character arcs that are inflected at plot points. Other stylistic and expressive arcs can intertwine to create the impression these curves represent. We'll see this idea reappear in Sections 2.9 and 4.1.2 as foundational ideas leading toward Media Scores.

The notion that information is encoded in the establishment of expectation and unpredictability of a medium is consistent with the role of formalism in artistic practice. Artistic disciplines or stylistic schools often define a set of conventions for the form and composition of a work. Students of artistic media learn these conventions as the foundation for their artistic practice. However, if every work of art followed the rules with exacting precision, there'd be little room for originality and individual, auteur-like ways of encoding intent. These conventions operate on a variety of scales ranging from the medium as a whole, to artistic and cultural movements, to artists, to individual works, and from moment-to-moment within a work.

With time, some deviations from conventions may be widely adopted and artistic practice and understanding evolve. Conventions make the work accessible to a public and define a set of expectations that can then be diverged from to create profound meaning.

Manfred Clynes suggests that individual composers have a unique pulse, a particular variation in timing of events that is not notated, but suits their compositions and can be mathematically modeled as a microstructure of the work [23]. The composer's pulse typically identifies the composer and spans their compositions. Clearly, we can often identify a new work as originating from a particular composer, even if we have not been exposed to that particular work before. One can recognize Bach as the author of a Bach piece that one has not previously heard, regardless of the interpretation of the conductor or performer. Even if I missed the opening credits, I can tell if a film's score is by John Williams or James Horner. Each has a unique style and voice that transcends their work, the quality of the work, and even the orchestra performing it. Philip Glass's music is characteristic, and despite being in the minimalist style, not easily confused with that of Steve Reich, just as we can tell a Monet from a Renoir.

Variations in practice distinguish artists from each other and the effect of particular works. Within an individual work, following the rules doesn't make something inherently meaningful. Breaking the rules with the incorrect intention can yield an incoherent mess. In some sense, the creation of a work of art is conforming to the rules where it supports meaning and then violating them or bending them where it creates new meaning. This gives rise to the expressive power of metaphor, counterpoint, and montage in verbal, musical, and cinematic media, respectively.

Analysis or reasoning about the expectation and the information content, and resulting expressive effect, of a medium requires that the information be quantified or otherwise represented in some temporally- or spatially-varying manner. In the artistic context of interest, these may be formal or expressive parameters in one or more modalities and can interact with each other and across modalities. It is here that we can formulate a general typology of relations. Much like the characterization of the movement of melodic lines in counterpoint, expressive parameters can relate to each other in ways that reinforce, diminish, or contrast their individual effect. This is related to notions of expectation, affective intensity, and expressive information content. For some spatial or temporal window of analysis (the domain and timescale depending upon the modality being explored), a single parameter may behave in one of seven ways as shown in Figure 8. Continuous changes may or may not be linear, but are monotonic for the region in question. Instantaneous changes are discontinuous.

Layering multiple parameters as they vary independently provides richness to the analysis of the emotive content of the medium. For a given region, the way in which these parameters interact can then be explored. In the simple case of two equally salient parameters, the parameters with respect to each other:

- Move in the same direction
- Move in contrasting directions
- One may move relative to a constant state in the other

If two parameters move in the same direction, it is presumed the effect would be additive, where one parameter resonates with the other, enhancing the combined percept. If the parameters move in opposing directions, the effect is diminished or—if proportional and the parameters have equal weight—nullified. When one parameter moves with respect to a constant parameter (non-existent, constant, or random), the effect is one of contrast. The constant parameter serves as a reference frame for the change

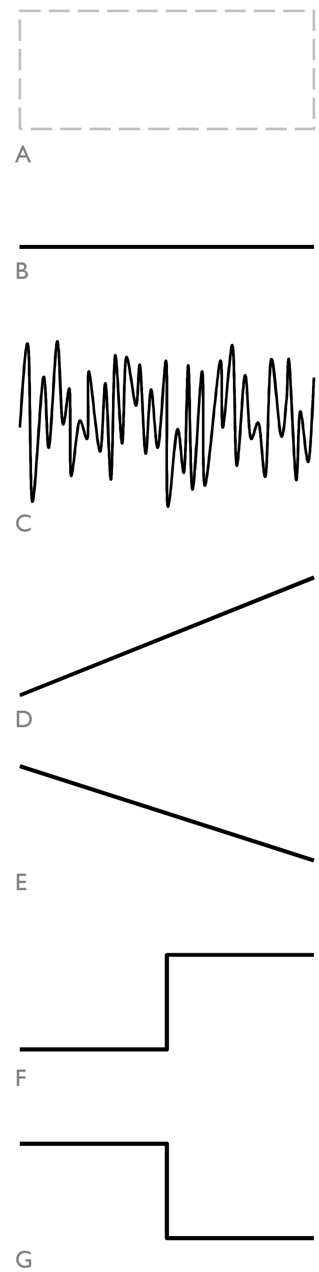


Figure 8: Types of parameter movement
A. non-existent (not salient, absent)
B. remain constant
C. behave randomly
D. increase continuously
E. decrease continuously
F. increase instantaneously
G. decrease instantaneously

in the other. As additional parameters are considered, the net effect is some linear combination of the relative movements of the parameters.

It is important to note, however, that the story is a bit more complex than this. Simply adding or attenuating feature parameters is ineffective at accurately describing the percept. Work in modeling the percept of *tension* in Western tonal music has shown that the result is neither additive nor noncongruent (emerging from the difference or contrast of parameters changing in contrary motion). Instead, feature parameters must be weighted for relative salience before modeling their interactions (the weights of which, in the case of musicians versus non-musicians, are also influenced by the perceiving population's experience).

Juxtaposition and contrast intuitively suggest the emphasis of features that are perceptually salient and are, in essence, related to the violation of expectation. The sharper, more distinct the contrast, the greater the salience of the form. The meaning of that contrast is contextually dependent and, in most cases, must also be judged relative to prior schemata. What is important about the juxtaposition? There are two possibilities. The first is simply that a difference exists, regardless of the content of the stimuli being compared. That something is not the same as something else—that a change has occurred in time—is significant. The second possibility is the difference between the figure and the ground or one parameter with respect to another.

This is a form of contrast and relates to the dialectical approach that Eisenstein theorizes about in film montage. He asserts that this "collision" approach is applicable to the creation of meaning in all arts and, indeed, the constituent modalities of film in which these collisions occur are analogous to other media [30]. There is conflict within the frame (visual composition), between frames (montage), and in conjunction with other stimuli (music). Meaning is also constructed through the juxtaposition or superimposition of the filmic work, in his view, and the schemata of the surrounding context. He asserts that there is an inherent conflict between art and its context through illuminating the disparity between the viewer's mental conception and the socially normative and also through its intentionality over the chance ordering of nature.

In a sequence, he argues that each element, each frame or shot, is inherently abstract and meaningless. Although this follows from his claim that meaning is the result of the dynamism of the expressive utterance, this does disregard any inherent denotation of the subject of an image or sound and any personal or social context that may impart meaning to the still image.

Eisenstein, favoring the power of a conflict in stimuli giving rise to meaning is particularly dismissive of the previous work of Lev Kuleshov. Kuleshov established that the creation of meaning arises through editing of film from the juxtaposition of shots [15]. This idea laid the groundwork for Eisentein's theory of montage. However, Kuleshov's work generally created meaning though a more constructive or additive cinematic language than that which Eisenstein espouses. The renowned experimental work Kuleshov presented assembled shots of subjects that were not created for a specific intent or story in such a way that a story emerged. Viewers imposed a causal relationship on the sequence of images and that was the source of meaning. An expressionless face of a man followed by a shot of a baby crying would suggest to viewers that the man was reacting with displeasure at the baby. The same dispassionate shot of the man when followed by a plate of food would be read as hunger. And so on. The meaning—and ultimately a narrative, if intended—is constructed from the relationship of one shot to the next. The causality is inferred from the presumed intentionality of the editing and the plausible scenarios or motivations that can be ascribed to the subjects of each image. The effect is not unlike the pairing of subject nouns with verbs and verbs with objects. Meaning is assembled from potentially inert constituents.

The consideration of expectation in creating meaning and expressive content is at the very core of the process of composing a Media Score. The Media Score is an object that must encode the particular path that an artist chooses between the predictably consistent and the erratically arbitrary. Composing a Media Score is one of adding moments of interest, contour that elicits the particular effect, and constraining the range of variability where appropriate in order to sculpt the overall experience. The Media Score is a tool for representing these trajectories of experience and, like any form of artistic creation, is open to the higher-levels of significance offered by expectations about story and composer. Indeed, the uniquely identifiable style and taste and method of creating experiences employed by a particular Media Score composer will likely be preserved by the process. Multiple Media Scores by a single seasoned composer should be identifiable as such.

2.3 Emotion

In the previous section, we focused on the capacity for emotional response to emerge from structured changes in time or space. I touched on a few examples of this in practice in particular media. Artistic expression often encodes meaning through a medium's particular formal parameters by presenting stimuli intended for an audience to find emotionally resonant. Studies of the various media examine how their formal parameters can contribute to intentional expression. This is important for understanding and analyzing representations, as input to a system or the score as an artistic representation itself, in addition to generating representations from an expressive intent. The formal parameters model this representation of emotion, but we also need to look toward models of the emotion itself from and to which we can translate.

Many theories of emotion discretely classify affective states. Most notable, Paul Ekman proposed a set of six basic emotions based on findings related to recognized facial expressions [32]. Ekman claimed that the six discrete emotional states were universally identifiable both in experience and the manner of facial expression, thereby suggesting a consistency of experience and representation. The work of Manfred Clynes identifies seven such states and demonstrates how each can be manifest consistently among individuals in a simple motion [24]. Clynes's sentograph was a device that measured the horizontal and depth displacement of a finger press in response to emotional prompts. The resulting sentograms collected reveal characteristic contours associated with each emotional prompt, suggesting an intuitive connection between affect and gesture (Figure 9). Perhaps they bear some resemblance to the expressive curves representing tension and information in the previous section, although Clynes suggests that these curves may be more figurative. He goes on to demonstrate how the form of these contours may be read in appropriate emotional contexts in works of visual art, sculpture, and music. This *sentic equivalence* , to use Clynes's terminology, demonstrates one way in which expressive intent may be encoded in works of art, particularly in a cross-modal way.

Intuitively, these discrete models of relatively few states fail to capture the breadth and nuance of our emotional experience. Variations on some relatively small sets of states can be thought of as co-occurring primitive states or as interpolation among these states [83]. The latter view leads to the suggestion of metric spaces for modeling affect. A number of such spaces have been proposed, though the most widely accepted is James A. Russell's two-dimensional circumplex model [90]. In this model, two orthogonal dimensions have come to be termed: *arousal* and *valence*. Arousal indicates the intensity or investment of the individual while valence reflects whether the response is favorable and positive or unfavorable and negative. Russell used experimental data to place commonly identified emotional states along the unit circle of these axes. The metric space is useful because it can represent the magnitude of an affective response, as well as account for trajectories from one emotional state to another, a parameterization that is important for representing emotional expression over time. In the previous section, we saw how features of the stimulus medium can be integrated by our perception to arrive at a continuously-varying affective and expressive representation of the source. Thus, even when the source medium is discrete, our experience of it is one of continuous modulation. With an appropriate parameterization, responses to expectation fulfillment and information can be mapped into a trajectory of affective response.

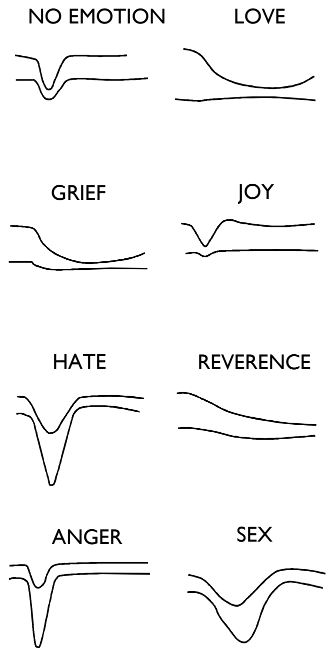
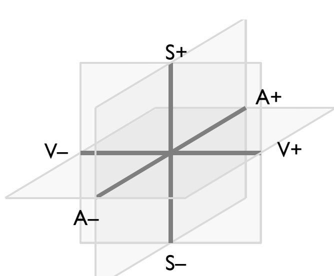


Figure 9: Essentic forms
Sentograms for seven emotional prompts and a neutral state. The time line in each response indicates the vertical displacement of the sentograph key while the bottom is the horizontal skew. Each trace is about two seconds in duration. (Redrawn from [24])

Figure 10: Affect space
A three-dimensional metric space for modeling affect with orthogonal axes of arousal, valence, and stance



Metric spaces like these suggest models for representing expression within a Media Score. This approach is used to infer the instantaneous affective state of a character as part of the mapping of real-time performance across modalities in Disembodied Performance (see Section 3.1.1). However, affective state is not the only means for conveyance of expression. Variations in expectation can occur with respect to other parameters that represent abstractions of cross-modal congruencies that do not necessarily have clear affective definitions. We now turn our attention to synaesthetic experience as a means for exploring cross-modal interpretations.

2.4 Synaesthesia and Multisensory Integration

Media Scores are non-figurative representations of story and intent. They are symbolic and expressive indications of the essence of a work of art. Referential meaning, as well as context—situational, cultural, and personal—can complicate the interpretation of a work of art. For the purposes of this research, I am concerned primarily with abstract representations. These representations have the capacity to convey emotional content without the confounding factors of referents and context relative to the spectator. Music is an excellent example of an entirely abstract medium. There are no inherent referents in music, though music can evoke particular imagery or concepts when framed by a context and musical symbols can be established within a work. Visual arts are capable of explicit reference by depicting likenesses or using well-known symbols. The abstract visual arts create visual representations that avoid referents, favoring the pure communicative quality of the medium.

Synaesthesia can provide a lens through which we can attempt to understand how stimuli can be assembled into meaningful representations. As a neurological phenomenon, synaesthesia occurs when a stimulus in one modality activates a sensory pathway for a different modality [26]. True synaesthesia is relatively rare and generally considered an abnormality, possibly resulting from a lack of pruning or differentiation of sensory pathways during brain development. However, to some extent, we all have synaesthetic experiences. It is that synaesthetic awareness that makes multimodal artworks resonate with us, which can be seen as the basis for Gesamtkunstwerk. It is this emotional connection to stimuli and the ability to respond similarly to different stimuli that makes representation possible. We can use the same words to describe a percept or an experience, regardless of the modality in which it arises. Words themselves carry out this interplay through synaesthesia's cousin, metaphor. The link between the two was first suggested by Charles Osgood and Changizi and others later suggested the neurological basis for the link between synaesthesia and metaphor [22]. Metaphor applies meaning from one domain abstractly to another. We can observe "rhythm" in the spatial domain architecture as a periodicity that, in music, would occur in the time domain. Describing an instrument's timbre as "bright" does not suggest that it is emitting photons, but evokes a higher-level concept that the visual and auditory modalities have in common. The power of language through metaphor is important in defining the formal and expressive parameter spaces that enable the mapping of evocative content from one medium into another, which we will explore in more detail in Section 4.3.2.

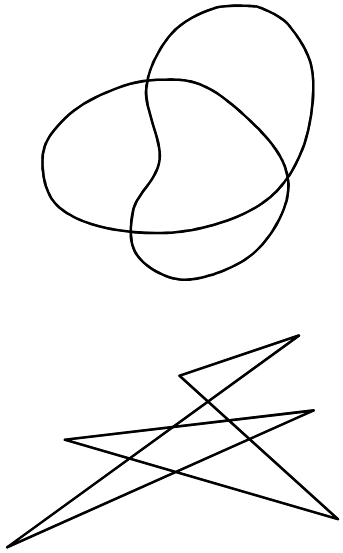


Figure 11: Maluma and takete
(Redrawn from [51])

While true synaesthesia is the result of neurological function, as is the derivation of emotional responses to sensory stimuli, synaesthetic experiences can occur by means of cognitive processes, as well. A classic example of cross-modal correspondence is presented in *Gestalt Psychology*. Köhler conducts an experiment in which subjects were presented with two images (Figure 11) and two names—*maluma* and *takete*—and were asked to assign each name to one of the images [51]. With a high degree of certainty, subjects associated the maluma name with the rounded figure and takete with the angular figure. One possible explanation for this is that the morphology of the mouth resembling the visual shape when producing the phonemes of the names. The sounds themselves have qualities and spectral formations that are similar to the features of the shapes, with the staccato consonants in takete producing "sharp" features and frequency spikes, whereas the "smooth" elongated vowels of maluma resemble the curved shape. Regardless of the mechanism by which these correspondences occur, they clearly illustrate the capacity for a concept to be perceived across modalities.

The Gestalt view has also led to the study of multisensory integration at both the neurological and cognitive levels. Multisensory integration is particularly useful in making perceptual inferences about our environment. The co-presentation of stimuli in one or more modalities can reinforce the causality of sources that produce effects in multiple modalities [98]. Stimuli across modalities may reinforce each other in space and time, indicating that the Gestalt laws of grouping can apply across different senses. Cross-modal congruency also can enhance the perception of weak stimuli [100]. This suggests that multimodal representations can reinforce content, expression for our present purposes, through matched stimuli, as we discussed with respect to parameter ensembles and the creation and manipulation of expectation.

Multisensory integration is important for understanding our perceived environment, but the implication for the creation and interpretation of synaesthetic experiences is also apparent. The congruence of concepts across modalities and the perception of correlation, whether at the perceptual or cognitive level, defines the relationships that can be constructed for the encoding of expressive intent. Synaesthetic responses, like metaphor, propose a connection of concepts across modalities. I believe the presumption of a connection reveals a commonality in aesthetic percept independent of modalities.

2.5 Semiotics and Abstract Representation

Media Scores is an abstract medium in that it represents intention without explicitly defining the form of the final representation. As we will see, assets and notations can assign specific signs and referents to parts of the piece, but the core of representing expression is devoid of explicit referents and figurative form.

In representational art forms, the creation of meaning comes not only from the synaesthetic assembly of form, but also from figurative and symbolic reference to concepts and objects. If we use regularities learned about the world in order to perceive and make meaning of it, as we saw in Section 2.1, how then does an abstract representation function as a communication medium? The visual arts can certainly communicate in a figurative form. Text is a medium entirely of signifiers and the spoken word can be borrowed in theater and song to convey denotative meaning. Music and dance, however, function without explicit denotation.

Music is an excellent reference for considering the communicative power of abstract representations. It is a medium that cannot have explicit referents represented into it, unless that reference is to another work of music, as in musical quotations of "God Save the Tsar!" and "La Marseillaise" in Tchaikovsky's *The Year 1812*. Music can also contain other types of signs to create meaning. A trill of a piccolo might be iconic of birdsong. The practice of leitmotif establishes an indexed sign in the form of a musical motif to represent a character and its meaning must generally be established within the piece and persists only within the piece [43].

For the most part, however, music is devoid of explicit reference and is one of the more abstract media for expression. Despite this fact, it has the power to communicate very directly and clearly the emotional intention of the composer. For this reason, the study of music and emotion has received a considerable amount of attention since antiquity. Music also has the advantage of generally being a discrete and well-defined language and readily quantifiable, making mathematical and computational analysis considerably easier and less ambiguous than with other media.

2.6 Abstract Art

Abstract art, particularly the art that arose from Futurism in the early 20th century, is replete with examples of works in a visual medium that evoke powerful responses. Many of the abstract artists of the time theorized or felt a strong synaesthetic connection with the visual and other modalities. Such examples serve as a strong influence on the design of Media Score representations and as a validation of the potential for multimodal non-figurative meaning. Artists such as Wassily Kandinsky and Paul Klee employed a strong synaesthetic sense in their work [99]. Klee referred to some of his paintings as operatic and often included text or allusions to music notation to suggest a synaesthetic interpretation. Kandinsky, who is believed to have been a true synaesthete, similarly considered his paintings to have a musical quality and theorized about the interplay of color, form, and emotion. Kandinsky also created synaesthetic theatrical experiences that sought to unify color and sound in performance, a notable example being his work *The Yellow Sound*.

Figure 12: *Composition V* (W. Kandinsky, 1913)

The abstract expressionist style of Kandinsky's *Compositions* series of paintings is of relevance to the development of Media Scores representations. These paintings generally have stories associated with them, that is to say they are in some way referential, but are non-figurative and any sense of time or progression is revealed in an implicit projection. Kandinsky sought to invoke emotional and spiritual responses in viewers in reference to depictions of particular events. The paintings themselves, however, contain little if any figurative material, relying on the relationship of form and color to convey their message. He focused particularly on dialectics of color as a means for prompting an "inner resonance", as he referred to it, in response to viewing the artwork [49]. Like Goethe and others, Kandinsky theorized on the function of color and dialectical color pairings, ascribing an intrinsic meaning and value to them. He also identified the affective character of visual forms. Although intuitive and clearly communicative, as evidenced by his works, these relationships are difficult to quantify or derive as a computational grammar. Nevertheless, the ideas of contrast and dynamism of form in the visual field are consistent with our discussion of expectation and change as the source of expressive meaning in a representation.

Figure 13: *Cosmic Synchromy* (M. Russel, 1913–14)

In Futurist works, we find examples of pairing abstract visuals with poetic text as a multimodal form of experience. One such work is *La Prose du Transsibérien et de la Petite Jehanne de France* (Figure 15), a collaborative piece that features a concrete poem by Blaise Cendrars situated alongside an abstract painted accompaniment that echoes the emotional qualities of the text by Sonia Delaunay [82]. This example is of interest because it is a work in two distinct modalities that are displayed in alignment, with color invading and infusing the text. In essence, it is an explicit-time representation, with the imagery following the essential progression of the text which is read linearly, line-by-line. Sonia and her husband Robert Delaunay championed the creation of meaning through juxtaposition and, more commonly, congruence, what they termed *simultaneity* (*simultanéisme*). This concept was inspired by Michel Eugène Chevreul's theory of simultaneous contrast with regard to color perception. Simultaneous artworks, like *La Prose du Transibérien*, pair colors with each other or with other forms and modalities.

Of particular interest is the Synchromism movement that arose out of Futurism, including painted works by artists such as Morgan Russel and Stanton MacDonald-Wright [17]. Synchromist works, often called synchromies, were intended to be visual analogs of an expressive musical experience. They rely on color scales, rhythmic repetition of form, and abstract structures to create visual compositions. The Synchromists looked up to music as the ideal art form of pure expression and sought to emulate its ineffable qualities in the medium of paint [109].

Joseph Schillinger's 1948 treatise *The Mathematical Basis for the Arts* proposes a formalization of artistic expression that relies on mathematical models to describe expressive content [94]. As a composer himself, the Schillinger System gained particular prominence in the music world. Schillinger used a mathematical formulation to define the regularities of aesthetic expression across the formal art forms. This theory is notable for it describes a methodology for composition—in any medium—that is independent of stylistic constraints. It focuses on the mathematical effects of formal properties and their combinations.

As part of his theoretical formulation, Schillinger defined a comprehensive classification of eighteen art forms defined by their formal properties (Figure 14). He categorizes the art forms first by *general components*, which consist of *space* and *time*—thereby accounting for kinetic and static artworks—and *special components*, which are the formal parameters of the specific modalities in question. While this reduction of artistic expression into a common set of quantifiable parameters is in the same spirit of Media Scores, the use of formal parameters obfuscates the nature of expression that can be encoded in the art forms. Media Scores defines content independently of these formal properties, so that a Media Score can be realized in any of these eighteen art forms by interpreting expressive parameters into the formal parameters of these modalities.

2.7 Visual Music

The conflation of color, abstract imagery, and music, since antiquity, also has been employed in performance and novel mechanisms for creating these visual forms evolving in time. The synaesthetic art form of visual music has taken the form of colored light or imagery produced in conjunction with music, often live. It is the extension of the abstract forms of visual art explored in the previous section into the dimension of time. These color

Figure 15: *La Prose du Transsibérien et de la Petite Jehanne de France* (opposite)
(B. Cendrars and S. Delaunay, 1913)

Figure 14: Schillinger's complete table of individual art forms
(Reprinted from [94])

instruments include light organs, such as Louis-Bertrand Castel's *clavecin pour les yeux* of 1734 [81] to Oskar Fischinger's 1940s Lumigraph [77].

Castel's color organ took the form of a keyboard instrument in which the keys, instead of producing sound through mechanical means, opened or closed curtains over colored glass filters, thereby producing multicolored illumination in response to playing the manual. Castel envisioned a version that would also produce pitches in correspondence with the color. A variety of color instruments have been created since. Some produce color and sound or sound alone. With the proliferation of inventors and artists building new forms of the instrument, just as many theories arose about the congruence between color and pitch or the nature of color music performance.

Sensation		General Component	System of Special Components	Title
Kinetic	Hearing	Time [1]	Sound	1. The Art of Audible Sound
	Touch	Time [1]	Mass	2. The Art of Touchable Mass
	Smell	Time [1]	Odor	3. The Art of Smellable Odor
	Taste	Time [1]	Flavor	4. The Art of Tastable Flavor
Static	Sight	Space (X_1, X_2) [2]	Light	5. The Art of Visible Light
	Sight	Space (X_1, X_2) [2]	Pigment	6. The Art of Visible Pigment
	Sight	Space (X_1, X_2) [2]	Surface	7. The Art of Texture of Visible Surface
Kinetic	Sight	Time [1], Space (X_1, X_2) [2]	Light	8. Kinetic Art of Visible Light projected on a Plane Surface
			Pigment	9. Kinetic Art of Visible Pigment transforming on a moving surface
			Surface	10. Kinetic Art of Visible Texture transforming on a moving surface
Static	Sight	Space (X_1, X_2, X_3) [3]	Light	11. Static Art of Visible Light placed inside of a 3-dimensional spatial form
			Pigment	12. Static Art of Visible Pigment covering the surface of a 3-dimensional form
			Surface	13. Static Art of Visible Texture of 3-dimensional forms
			Mass	14. The Art of Static 3-dimensional visible mass
Kinetic	Sight	Time [1], Space (X_1, X_2, X_3) [3]	Light	15. The Art of Visible Kinetic Light projected on a 3-dimensional or a 2-dimensional screen in motion
			Pigment	16. The Art of Visible Kinetic Pigment covering 2- or 3-dimensional surfaces in motion
			Surface	17. The Art of Visible Kinetic Texture of surface or volume
			Mass	18. The Art of Visible Kinetic Mass

In 1893, Alexander Wallace Rimington introduced the *clavier à lumières* along with the terms *color music* and *color organ*. Rimington's instrument gained success, particularly among composers, including Wagner [118]. The instrument was often performed with music, as part of an orchestral arrangement. Color instruments evolved from Castel and Rimington's light-producing organs to systems that produced abstract imagery from colored light. Thomas Wilfred coined the term *lumia* for these images and his 1922 instrument the Clavilux [81]. Wilfred's instrument did not produce sound and he believed that there was no reasonable correspondence between image and tone. Thus, lumia should stand as dynamic works on their own. Wilfred's work has since given way to more recent forms of performance projection, live abstract imagery in light and color, both paired with music and otherwise, such as the oil light shows of Glenn McKay in the 1960s and the like [116].

Figure 16: *Untitled*, Op. 167
Example of lumia
(T. Wilfred, 1965)

The tools for performing color and light over time in connection with music and text furthers the idea of the Gesamtkunstwerk. Following Kandinsky's imagination of *The Yellow Sound*, in 1915, Alexander Scriabin premiered his *Prometheus: Poem of Fire* scored for orchestra and including a color organ, the Chromola by Preston S. Millar, based on Rimington's *tastiéra per luce* [118]. Stage designer Adolphe Appia also believed in the intimate connection of the visual, spatial, and musical in storytelling. His designs, for Wagner's operas in particular, relied on dynamic control of light in color as an integral part of the scenography [12]. Today, the integration of light and dynamic moving image projection in performance and theater are an essential part of the scenographer's palette.

Visual music is not only composed for live performance and color instruments, but has found a home in recorded form through film and now video and computer technologies. The abstract films of Mary Ellen Bute and Ted Nameth echo the spirit of simultaneity and synaesthetic performance of color with music, relying on both color and form to accompany and intertwine with music [76]. Bute's work pairs the photographic and the musical, exploring the congruence across modalities, in collaboration with Joseph Schillinger, in her earlier work.

Norman McLaren followed a similar pursuit, creating abstract films through animation and celluloid drawing that were often less photographic in form than Bute's [86]. McLaren also explored the connection between visual and sound by animating not only imagery for display, but also drawing in the optical sound track of the film to visually compose sonic frequencies and timbres.

Fischinger is also well known not only for his performance of color music using the Lumigraph, but also for his animations meticulously designed to accompany recorded music [17]. They are at once simultaneous and

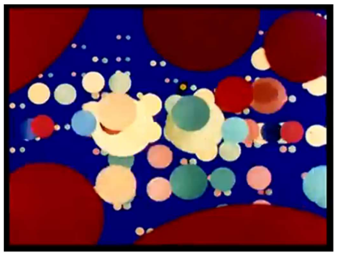


Figure 17: Frame from "An Optical Poem"
(J. Fischinger, 1937)

synaesthetic works of art as much as they are visualizations of the music. While they may not literally reflect the entirety of the notes and parts in the musical score, their forms bear a very close correspondence to the formal properties of beats, note onsets; size correlated with salience and amplitude; color and shape reveal qualities of timbre and instrumentation. A notable example of this work is Fischinger's famed "An Optical Poem" in which his animated object visuals accompany Franz Liszt's Hungarian Rhapsody No. 2. The title in the prologue to the film sums up the synaesthetic goal of visual music:

> To most of us, music suggests definite mental images of form and color. The picture you are about to see is a novel scientific experiment—its object is to convey these mental images in visual form. [4]

In the subsequent years, many such pairings of abstract moving imagery and music have been made.

Today, visual music has its most common descendant in the algorithmic visualizations that are generated by computer software, often in media players, in real-time correspondence with live or recorded music. This has given rise to a host of computer tools for video DJing (VJing) to combine the visual and musical media in close congruence of abstract form. However, many of these tools merely perform trivial mappings of the formal parameters of an audio waveform into the formal parameters of pixels and primitive shapes [2]. They do nothing to try to understand the expressive intention of the music and create a phenomenological link at a higher semantic level of abstraction. It is up to the performers of these systems to create or invoke any meaningful relationships in real-time.

Visual music has seen resurgence with the rise of creative coding. However, in many cases the music and visuals are synthesized together or the visuals are synthesized in response to audio or note event data. Thus, the process of abstraction of quality and perceptual effects of the events is lessened as the visual elements bear a stronger, literal correspondence to low-level details of the musical source material or input. While highly responsive, the visuals have a musical quality solely from their connection to the music. They are illustrative, rather than a complimentary medium. As was the case with "An Optical Poem", visual music exists on a continuum from performance to visualization, particularly in the pairing of the visual modality with music. As performance, the representation is dynamic and expressive. The visual component can accompany music or another modality or stand on its own. As accompaniment, it should feel related and connected to the other modalities, but it is not essential that every event or every formal parameter of the companion forms be shown. As visualization, it is diagrammatic and demonstrates something of a one-to-one correspondence with the medium it accompanies. Visualizations may take the form of rendering audio data, as mentioned with media player visualizers, or musical events, as in the quintessential example of Stephen Malinowski's Music Animation Machine [66]. The latter essentially produces a piano roll notation from musical data. Music visualizations are, as opposed to audio visualizations, in essence, a recreation of a score for the music. In many cases, the music could be recreated almost exactly from such a low level visualization. More complex visualizations reveal the piece's structure rather than form, but may lose expressive power in doing so.

It is also interesting to note that the creation of color music predates recording technology and cinema. Thus, color music paralleled auditory music in that it was created by instruments and could only be viewed live. Painting (implicit representations) and dance (dynamic representations) also existed. The former could exist independent of time. The latter also needed to be performed. It is not uncommon to associate music visualization and motion graphics today with recording media and, through the mediatization of live performance practices, it is quite common for truly composed abstract visual representations to be pre-recorded and not performed live. Of course, this is not exclusive. There are numerous interactive and responsive artworks that do generate visuals in real-time, though perhaps not in as composed a manner, as music and theater in general are performed in real-time. In the work presented in Chapter 3, some examples from my own recent work illustrate the process of creating shaped visual and multi-modal works that are performed live and highlights the need for tools for composing and annotating this process.

2.8 Projections of Time

An abstract visual depiction of the shape of expressive parameters over the course of a time-based work poses the peculiar question of how best to represent *time*. Is time a critical element of a story or expressive utterance? How can non-time-based art forms be capable of conveying an emotional arc? Certainly, we can experience a compelling journey within the static frame of a painting, for example; so too in the winding melodies and counterpoint of an orchestral symphony. In some media, such as film, the timing of the experience is fixed and reproducible. Onstage or in an installation setting, time is fluid. The rate of information presented is constantly varying and how an audience may experience it—including both linear and non-linear or interactive experiences—will be different from one enactment to the next. Media Scores must consider this temporal variability and the embodiment of time, in order to be capable of representing both time-based and static, as well as time-varying and temporally scalable, realizations of works.

I define three modes by which time can be represented. Media Scores should be capable of modeling media that rely on any of these modes of representation and the different views afford useful perspectives on the same information. This classification will come into play in the discussion of the Media Scores application in Section 4.3.3.

The *explicit* projection of time is perhaps one of the most familiar representations of time in composing tools. This is essentially a timeline, where events or values are plotted against time, which occupies one axis, typically the horizontal axis. The explicit projection is particularly illustrative and diagrammatic, since the precise correspondences of value and time can be seen at once, and reflects how we reason about temporal relationships. This representation is common in software environments for editing video, audio, and animation. Indeed, it is also the representation of time used in standard music notation scores and many graphical scores, which we will look at in the next section.

An *implicit* representation of time does not reveal a single axis along which events are chronologically organized. Like the explicit representation, it is a static image, but retains a perceptual sense of dynamism despite not demarcating time. Implicit representations tend to function in roles that are more artistic. They favor experiential and synaesthetic quality over diagrammatic modes of presentation. The Synchromist paintings are a good example of this sort of representation. An expressive journey can be apprehended in a single glance.

In contrast to the implicit projection, a *dynamic* representation of time, such as in an animation, a performed musical or theatrical work, or a film, must be viewed over time to perceive and understand the changes of parameters or onset of events over the course of the work. In a dynamic representation, at any moment only a single instant of the piece's arc is presented.

The projections of time reflect our different understandings of how events unfold. The discussion of expectation above looked at two notions of temporal experience. The more intuitive version compares experience in time from one instant to another. The reality of the experience, though, has lasting temporal effects, accounting for the integration of stimuli, establishment of expectation, desensitization, and comparison of the current instant relative to a past impression. In both cases, we are confronted with the notion of the *instant*, a durationless moment in time. Perception is a phenomenon that occurs only in the present [53]. The past is but memory. With respect to static representations of the entire work, such as scores with explicit or implicit projections of time, the present and future are shown to us at once in notated form. The instant is any point in that score. In static works of art in general, we know that the work exists concretely at any point in time, but we do not perceive it instantaneously. We must move around the sculpture to see its various sides and use perceptual cues to interpret the interplay of light and shadow. The painting may fill the visual field at once, but our attention shifts as our gaze and focus wanders over its surface. Our perception of these static works is always dynamic, there is an implicit temporality that exists in art that, objectively, exists in the instant. It is from this that simultaneity and contrast can occur in static representations and through which expectation, meaning, and therefore expression can be constructed.

2.9 Scripts and Scores

Scripts and scores are the backbone of many forms of performative arts, particularly those that are associated with Gesamtkunstwerk. The source text penned by an author or the notated music of a composer, through interpretation by the artist or a creative team and performers, are turned into what is presented to an audience as a performance. The text and stage directions of a script can vary considerably to the extent in which they explicitly denote the author's expressive intention. Similarly, a musical score indicates which notes should be played and at which time by a certain instrument, as well as the quality or expression with which each note should be performed. However, neither of these representations truly captures the communicative nuances intended by their creators. Yet, somewhere amidst the symbols of notation and words, somewhere between the lines, the composer and author have encoded a reasonable semblance of their intent. Directors and conductors then have the charge to extract this latent message and work to realize it in its final form. Scripts and musical scores are effectively lossy compression codecs through which ideas are indirectly communicated from artist to audience.

The *piano roll* is one of the most basic forms of music notation: a record of pitch over time. A piano roll is an expressionless, literal score; a deterministic computer program. The more traditional *standard music notation*, as performed by humans, is subject to interpretation and an elaboration of the musical and expressive content not explicitly denoted in the score. At the opposite end of the spectrum from the piano roll is simply a musical instrument. A performer of any skill level can produce any sound that the instrument is capable of producing in an arbitrary sequence. There is no guide to the level of expression.

Standard music notation occupies a symbolic middle ground. Notes on staves define musical parts. Each note indication denotes a pitch and duration. Marks around the notes and staves give cues to the dynamics, articulation, and expression the composer intends. There is a large degree of imprecision in these indications, but the system of notation and the practice of its interpretation have evolved such that the composer's intention is preserved. The visual representation of standard music notation is iconic. Even if one doesn't know how to read music notation, it is still readily identifiable and has certain cultural connotations. This visual legacy of music notation—staff lines and square or dot-like notes—dates back at least to the 11th century when Guido d'Arezzo set out to notate Gregorian chants [80].

The intuitive, but inconsistently-defined connection between the visual and the sonic or the musical has extended from synaesthetic pairings in artworks to the composition process itself. Composers, particularly in the 20th century, have used graphical musical scores as a means of expressing sonic ideas that standard music notation is not well-suited to conveying.

Even given the freedom artists have to develop any form of representation, the connotative affordances of standard musical notation are difficult to escape, regardless of the actual interpretation of lines and notes. This is evidenced by the multitude of unconventional music notations and graphical scores that explicitly invoke these iconic visual elements. The meanings may be entirely different, nonsensical, or even non-essential. Yet, composers feel the compulsion to invoke the aesthetic of staff lines and note-shapes in their otherwise radical departures from convention, as if to assert that the artifact is indeed a musical score and should be performed and interpreted. For example, Cornelius Cardew's score for *Treatise* borrows the visual elements of standard music notation, but twists and combines them in ways that do not have any explicit interpretational semantics [21]. This practice shifts the role of the score from the symbolic to expressive. In general, however, lines, curves, abstract forms, and color may all play a role in graphic scores [17]. John Cage's 1969 compendium, *Notations*, contains numerous examples of the innovations of his contemporaries to standard music notation, in order to extend the symbolic system's expressive range [18]. Following Cage's example, Theresa Sauer recently compiled *Notations 21*, an updated sampling of the breadth of non-standard notation and graphic scores [92].

Graphic notation is subject to a greater degree of variability than standard music notation. In some cases, graphic notation can be more literal: expressing exact values, typically for the direct control of automated, electronic systems or algorithms. Although, standard music notation has a variability and subjectivity that we embrace: the influence and interpretation of the conductor and performers, as well as that of the listener. Non-standard techniques can emphasize the co-creation by participants in the performance-listening process beyond the constraints defined by composers. The aleatory and indeterminacy techniques, of which the New York School composers were proponents, are examples. While John Cage focused on chance, Earle Brown emphasized 'choice', affording the performers with a set of possibilities that could be chosen at performance time [91]. He wanted to bridge the divide between composition and improvisation. His graphic scores varied from those that allowed the conductor to transition sections in varying order to scaffolding time, but not specifying the durations or order of notes, leaving those decisions to the players. This 'open form' technique heightened the variability of a piece ensuring that no two performances would even have the same form, let alone the same expression. However, as Schillinger points out, they retain consistency by resulting from the same initial score [94]. Brown, in extreme cases, left the interpretation of the notation up to the performers, as in his "December 1952". When no two performances are alike, does the work even exist? Brown believed so. He didn't want to *own* the work in the auteurial sense, but he still felt that the score was his. Members of Fluxus, originally students of the New York School, brought this same sort of variability to their 'event scores' for happenings [28].

Graphical scores are particularly common in electronic music. While standard music notation reflects our perception of discrete pitches—which also is embodied by many types of instruments, such as the woodwinds—and often a regular meter, analog and digital systems make continuous variation in pitch and time possible [52]. Unlike standard music notation which must be parsed and interpreted into a sequence of pitches and other parameters in order to perform music, parametric scores can be used by music systems to directly generate sound. Graphical scores are then necessary to notate such continuous changes and often reflect the actual parameters used by electronic music systems. This type of graphical score is diagrammatic in nature, as opposed to expressive. Like symbolic forms, the meaning of visual gestures may be explicitly defined. Karlheinz Stockhausen used this approach for several of his electronic pieces, such as *Studie II* [103]. Iannis Xenakis also relied on graphical scores to provide electronic instructions for generation and playback of the work. Xenakis also developed a computer system on which to compose such scores, the UPIC, which used drawings of strokes in pitch over time to modify and play back recorded sound [57].

More recently, Hyperscore, like UPIC, is an example of an environment for composing graphical scores for music. However, Hyperscore manipulates and shapes musical material, motivic fragments composed of discrete notes, rather than sampled audio. It uses a visual language of color, shape, and texture to represent high-level musical features. Fragments of motivic material are referenced by strokes, which continuously transpose the notes to match the drawn intention. A harmony line represents a sort of musical tension that allows the system's automated harmonization to be sculpted by the composer. Early versions of Hyperscore have more in common with Media Scores, in that the graphical representations underwent a more

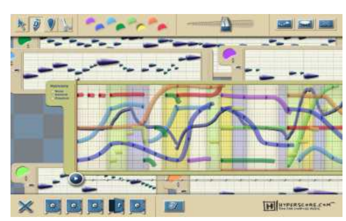


Figure 18: Hyperscore

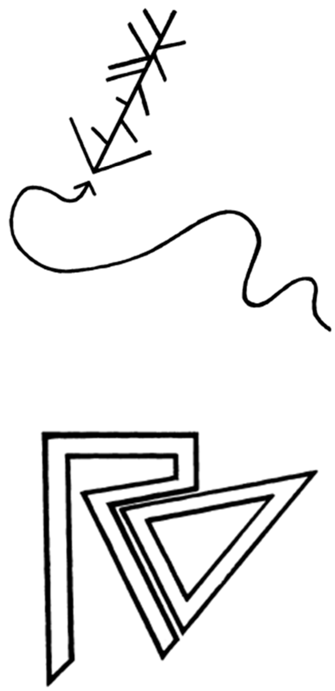


Figure 19: Visual scripts
Example visual scripts for exercises
(Reprinted from [25])

complex and more expressive mapping to musical output [35]. While Futurist and Synchromist attempts at painting in a musical fashion did not yield scores, environments like UPIC and Hyperscore make it possible to draw music. Additional work has been done to automate the musical interpretation and performance of more abstract and painterly methods of composing, as well [97].

The advent of graphical scores inspired a similar process in drama. Experimentations with visual scripts were used to convey qualities of action and branching scenarios in stage acting [25]. This systems-view of dramatic performance extends the purely textual indications of dialog and stage direction that often convey only the what, not the how, of narrative action. The extension relies on a synaesthetic interpretation, the use of visual metaphor, to extend the influence of the scripts author into the performance of the piece. Visual scripts are non-representational indications that can be used to provide source material for inspirations or stylistic indications for prescribed actions.

Graphical scores and visual scripts augment or supplant textual and symbolic representations of their artistic media with notations in the visual modality, in order to convey additional expressive parameters. While these techniques are still considered avant-garde, related techniques to interpret conventional sources (scripts and scores) have become established practice in theater, cinema, and video game production. Such tools include mood boards, storyboards, color scripts, animatics, and pre-viz. Mood boards assemble visual reference for tone, color, texture and abstract feel as part of early stages of the creative process. Storyboards begin to propose the mise-en-scène over time through sketches illustrating action given in the script. The illustrations, sequenced in an explicit projection of time, give an impression of the image's composition and editing, as well as begin to lay out the logistical needs of identifying shots and assets, defining the scope of the setting, used in generating a production schedule, and so on. This vision is refined through the addition of color palettes, presentation in real time, and animation through the remainder of the techniques mentioned. These visual tools primarily serve the function of collaborative communication among members of a creative team and are essential for asserting a creative vision that personnel can apprehend and follow.

Interesting amalga of many of these techniques, from standard music notation, graphical scores, and storyboards are rare, but do exist. An apropos instance is Sergey Eisenstein's diagram of his film *Alexander Nevsky* (Figure 20). Eisenstein did not use this diagram as a part of the creative process, as a tool for collaboration or generating the film, but rather as an analysis technique [31]. With it, he highlights the synaesthetic reinforcement of narrative content through the multiple modalities of film: musical underscoring, visual form, gaze, and so on. It can easily be seen

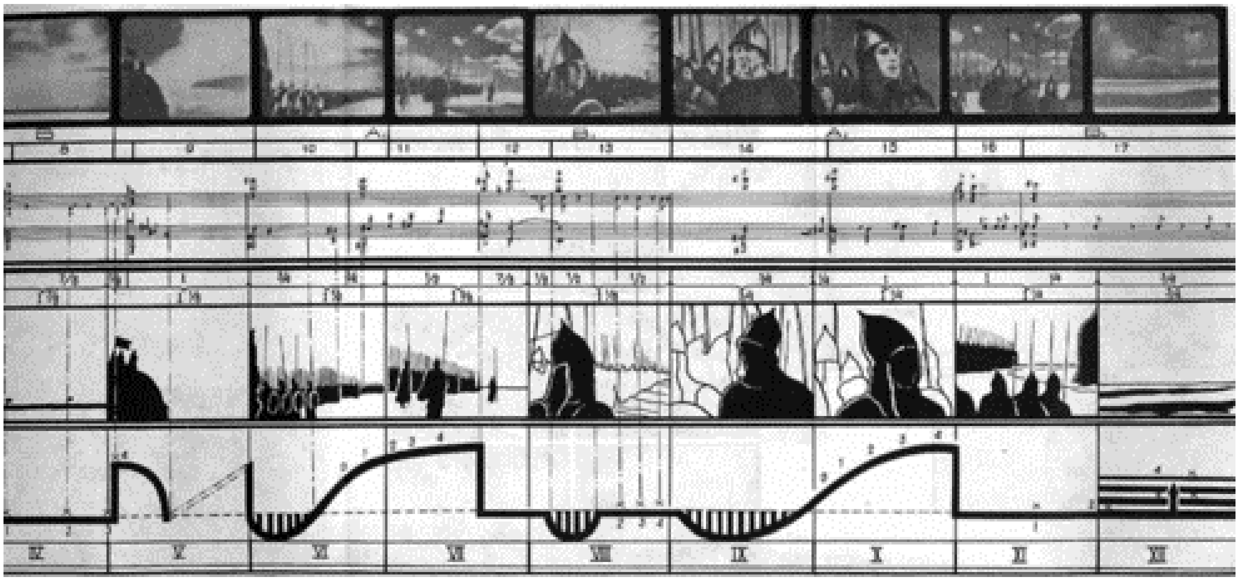


Figure 20: *Alexander Nevsky* diagram
This diagram relates frames from the final film with the musical notation for the films score, an illustration of each frame's composition, and a plot of the viewer's gaze. (Reprinted from [31])

how such a representation would have utility in the design and planning phases of production.

The goal of a collaborative reference, such as a script or a storyboard, is in many ways the same as the objective of an artwork itself: to take an idea from the creator's mind and place it in the mind of others. For previsualization tools, the vision for a work is conveyed from the creator (author, director, designer) into the mind of other designers and production personnel so that the vision can be faithfully executed. Previsualization tools also provide a way of indexing or referring to parts of the vision and a tangible means by which alternate ideas may be proposed. With the right tools, the flow of vision and ideas from one mind to another can be not just faithful, but bi-directional.

The script or score is not the whole of the experience, perceived or intended. For works of live performance, the script or score is what exists of the work when it is not being performed. It is both incomplete and whole at once. It encodes the intention and the possible worlds in which it could exist. Yet, it is a dormant object; a shadow of the work itself. Scripts and scores are, in and of themselves, seldom intended to *be* the work. They are intended to *become* the work.

A sculpture or an etching are works of art that exist timelessly. They can be observed and appreciated at any point. They are complete works in a way that scripts and scores are not. With recording technologies, cannot an instance of a work persist beyond the time of its performance? Indeed, this is essential to the medium of cinema. A Hollywood blockbuster is borne of a script, but goes on to become the film that is consumed. That final form becomes the work that is known; the singular canonical instance of the work the script encoded. A magical evening at the theater may be recorded by a multi-camera video crew for later television broadcast or DVD distribution. It features different vantage points from the singular one I'd have sitting in the proscenium house; close-up shots of performers in exaggerated makeup that would look natural from my seat 140 feet away. The magic of being in a shared space with 1199 other audience members, the ephemeral quality of a moment transpiring onstage that will never be precisely duplicated again is absent from the video recording. A compact disc release of Beethoven's Symphony No. 7 is a realization of the score Beethoven crafted. It is frozen in pits and lands on the optical disc, we can hear it again and again, repeated each time impeccably identical to the instance of the piece that the recorded orchestra played. But does that recording become the work or does Symphony No. 7 exist both apart from the score and apart from the sum total of all recordings and live performances?

Media Scores favor a well-defined interpretation of the expressive intention of the score, even if intuitive and open-ended. This preserves intention, even though the vocabulary to describe how the score should be interpreted may not be made clear. It is subject to interpretation, but no more so than a symbolic notation, perhaps even less so, since it prescribes the expression of the piece, something for which symbolic notations are weak.

2.10 Production Practice

Most production workflows for complex artworks can be broken down into five phases. The pre-production process precedes the actualization of the work and includes the composition, design, and rehearsal processes. Over the course of these three phases, control of the work moves from the composer or author into the hands of the director or conductor and performers. In the production phase, the work is realized. In live media, this takes place before a performance. In recorded media, this is the performance that is captured and later distributed to audiences. The distribution phase can also incorporate additional types of engagement with the work. During the latter phases, control of the interpretation of the work migrates yet again into the hands of performers and audiences.

In theater, cinema, television, and computer games, the personnel involved in the production of the work often takes on a hierarchical structure [38]. Authors and composers create a work and, if contemporaneous, sometimes collaborate with the creative team on a production. Generally, thought, the interpretation of the source text becomes the responsibility of a director. During the pre-production process, the director works with a team of designers to develop a production concept from the script and a vision for the realization of the work that reflects the director's interpretation. Each designer, in turn, is responsible for teams within their department who

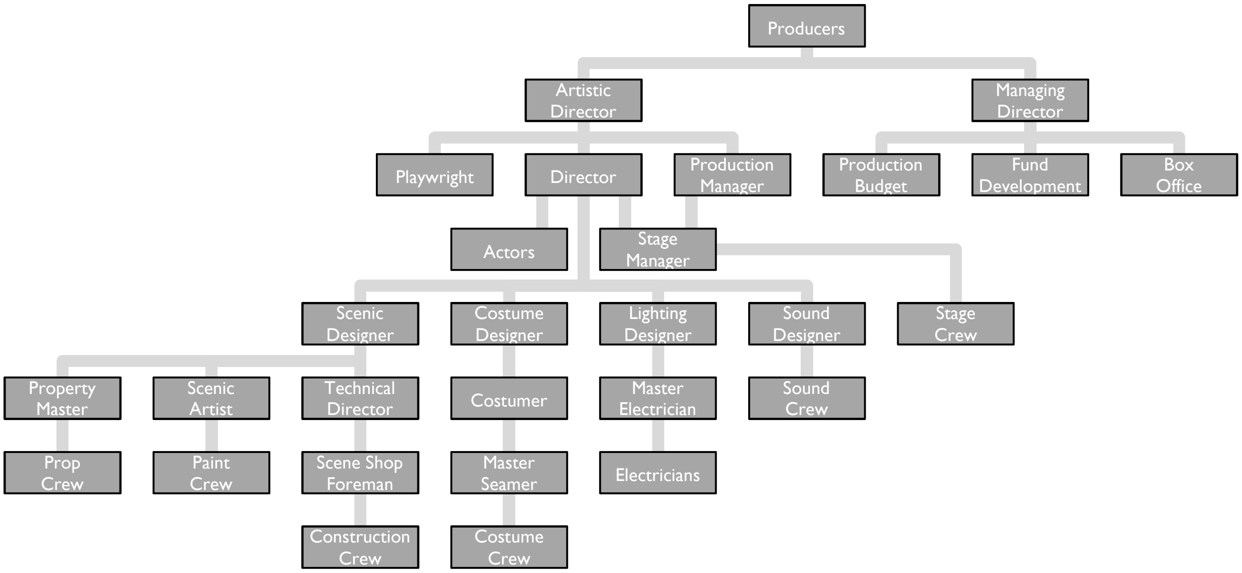


Figure 21: Organization chart for theater production
A typical organization for a theatrical production company centralizes creative control within a small team of designers around the director and sometimes playwright. (Reprinted from [38])

construct and realize the production concept. The departments are then fairly insular and creative collaboration and communication often happens only at the highest levels of the hierarchy. The work of the designers and their departments continues into the rehearsal and production phases as elements are brought together and final refinements are made. During the rehearsal process, performers are introduced and the director works with them to shape their portrayals. Once in performance and distribution, the creative work of the director and design team is done. Other creative teams may form for marketing and ancillary events or products related to the production, but these efforts draw from the production as a source rather than the original script or score.

The insular nature of creative departments functions well from an organizational perspective when there isn't much need for interaction across departments. The design processes and tools employed by each department depend on the modality, but generally follow a common form. After the script is analyzed and a production concept exists, reference materials are collected, sketches of the final form of each department's contribution are made, and ultimately construction drawings are rendered, in order to be turned into the final products of sets, costumes, lighting installation, and so on. A great deal of the design process is the development of a visual language that echoes the concept and the necessary planning required to realize that language during the production process.

The production process must also take into account not only the design language, but the role of time as an expressive modality. In musical composition, particularly in opera and program music, the composer has the ability to move the story along, to distort time—to dwell in a moment, a thought, an action or to elide and omit—to change both time and place, to isolate one or more characters and to show how they connect and interact. In these ways, the process of composition is very much analogous to editing in film. However, while editing is done once the entire story and production design have been realized and committed to celluloid or digital tape, the score for a musical work exists well before the singers or scenography take the stage.

Unlike film, live theatrical performance is necessarily improvisatory and malleable. There is no final version, as in the final cut of a film that is then distributed to cinemas. Many projects have explored the idea of performative or interactive film and video, essentially allowing the viewer to edit the playback of media while viewing.

The tools used for design and planning in these disciplines are of several forms. Scripts and scores are the symbolic guides that must be interpreted. The artistic vision is reconstituted from what the director and designers read between the lines of these source texts. Expressive tools include concept sketches and mood boards, which assemble imagery that feels related to the production concept in tone, texture, color, and other qualities. Similarly, for the creation of musical content, inspirational sounds and musical pieces can be assembled. This exploratory process can involve dramaturgical and historical research to create a design language that is consistent with period events or social and cultural influences that must be referenced to tell the story. The scripts and scores are often translated into storyboards, a more concrete representation of the production [50]. The use of storyboards is most common in television and film, but have been used in theater as well, such as by Bertolt Brecht. Storyboards begin to visualize the final form that the production will take, illustrating what will be seen and when, in reference to the source text, and how the window into the world represented will be composed. Storyboards allow the director and designers to determine aspects of blocking and, in cinema, camera placement. From this, the requirements of lighting and setting can be inferred. Scene and shot breakdowns, lighting and sound cues, and other lists of events that will need to occur in time can be assembled for each department. All of these design references inform the rehearsal and production phases.

Most of these design documents are created by analog means. Tools are available to assist with script writing and formatting, possibly augmented with storyboard illustrations or visual annotations. Software exists to collect repositories of reference imagery and display them. Previsualization software assists in constructing animatics and 3D mockups of sets and camera moves. Asset management systems can track versions of scripts and scores or digital images and 3D models from previsualization that may end up being fabricated or used in visual effects. However, these are essentially just databases of assets and metadata and do not provide a meaningful way of integrating the various forms these documentation and planning

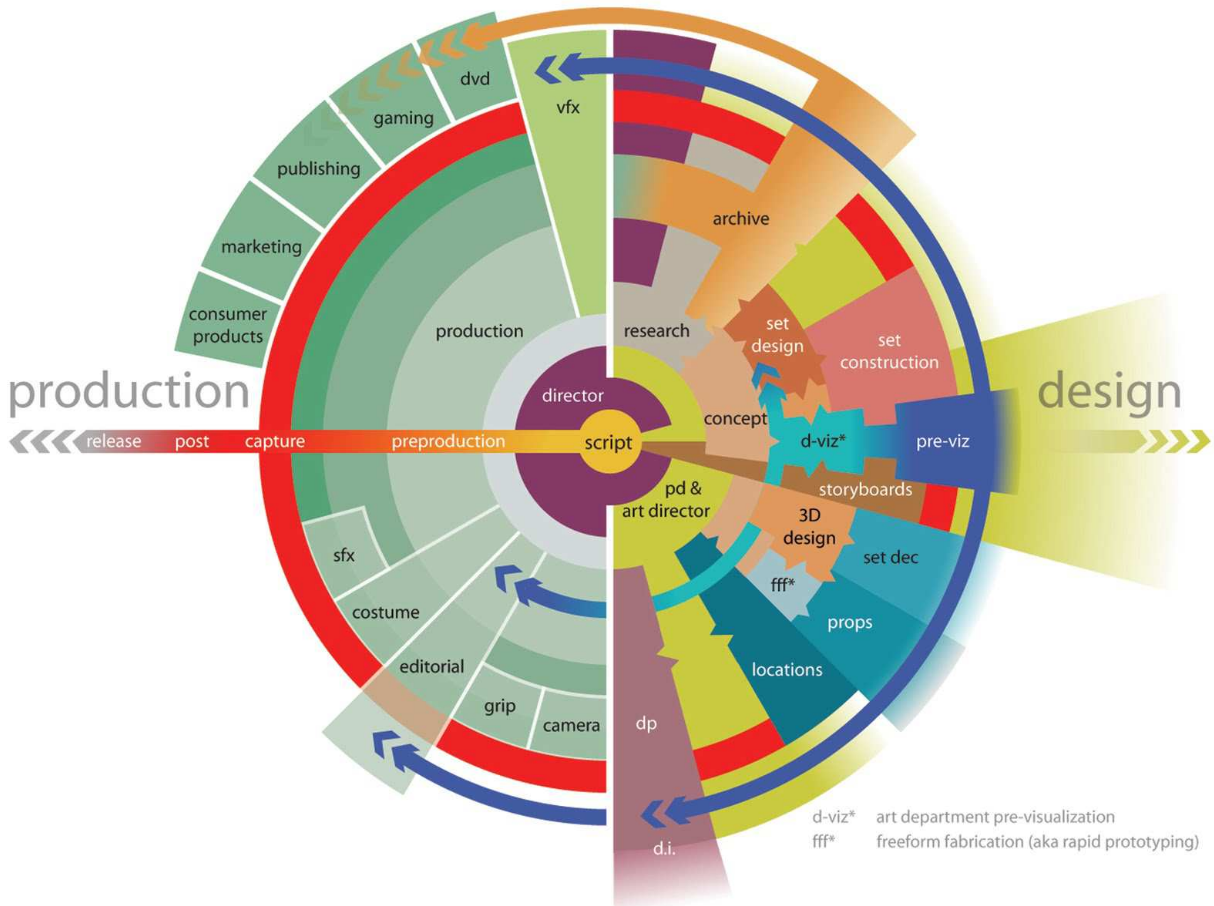


Figure 22: Immersive design mandala
This diagram reflects the interrelation of departments and tools during the production of contemporary cinematic work using digitally-enabled workflows.
(Image courtesy of Alex McDowell)

methods take. Furthermore, these resources, with the possible exception of 3D models, generally don't have any direct function in the production phase. They exist as references only.

One of the outcomes of incorporating new digital technologies into the filmmaking process was the advent and importance of the previsualization (pre-viz) animatic. This is an extension of storyboarding, in that it allows the director and creative staff to create a visual representation of the film in time before it is shot. The camera angles and editing are essentially determined in advance of production on the film, before an actor steps before the lens. Still, this pre-viz often takes into account elements of the production design (sets, color, et cetera), so it represents a more literal interpretation of the work than does the operatic score.

As we'll see in the next chapter, the introduction and interconnection of modalities of expression necessarily breaks down the divisions of creative responsibility. Film production practice is undergoing a similar process with digital pre-production, production, and post-production processes being integrated. Production designer Alex McDowell, through a process he calls *immersive design*, explores how digital production practices more closely

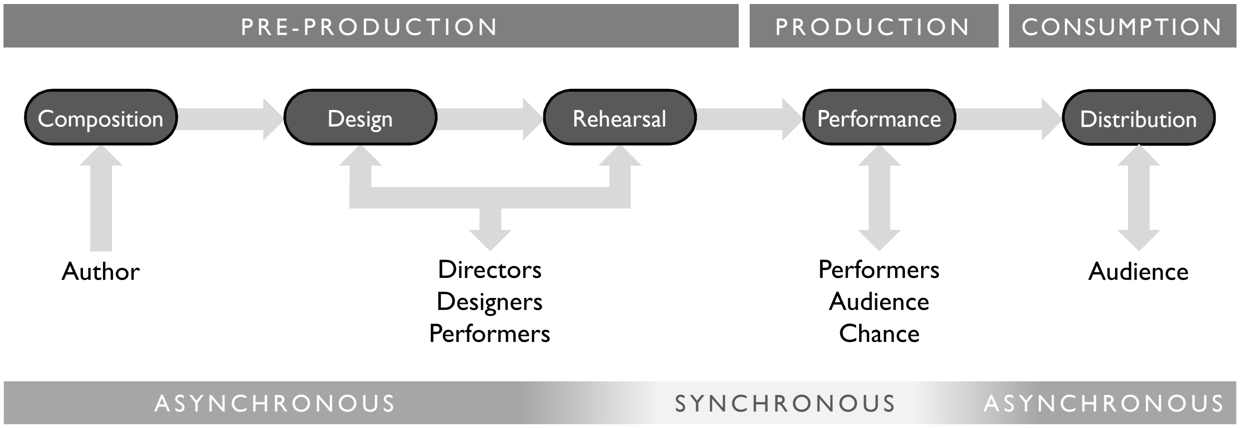


Figure 23: Phases of production
The production process of many Gesamtkunstwerk-like forms generally falls into a five categories—composition, design, rehearsal, production, distribution—each demonstrating different levels of synchronicity of interaction and sources of creative control.

integrate departments in production workflows. He illustrates how the increasing multitude of production departments is not as strictly hierarchical or as isolated as conventional organizational models of production would be [67]. McDowell's mandala visualization (Figure 22) also indicates the flow of information and resources among the departments. Nevertheless, aside from database repositories, there is no unifying software or representation that maintains references to assets and is able to document aspects of the story and production design. The need for a common tool for assisting in the collaborative creation of Gesamtkunstwerk and multidisciplinary art forms inspires the development of Media Scores.

2.11 Performativity, Liveness, and Immersion

The technological complexity that can be introduced into a modern-day Gesamtkunstwerk can pose a risk of ossifying the valuable and essential qualities of *liveness* that define theatre, real-time performance, and shared experiences. Theater creates for us a story in the moment. Video games, as a form of Gesamtkunstwerk, take a different approach to liveness by engaging the audience as participants, as performers, in the storytelling. Art technologies, however, particularly the technologies of digital projection and control that are finding their way into theatrical productions, borrow from film a deterministic, fixed sense of time, thereby having a tendency to counter the live, ephemeral quality theater enjoys. Our mediatized culture has desensitized us to a passive, yet unavoidable, screen on which glow imagery and illusions composed of pixels, as well as a didactic mode of discourse. How then can liveness be preserved in a new work such that it enhances the telling of a story or the communication of an expressive intent? How can new works of art satisfy our natural craving to assume, from time to time, an active role in our experience of art?

There is nothing inherent in technology, particularly given present-day computational capabilities, that disfavors live interactions. Indeed, the popularity of technologically-mediated installation art has grown in response to tools and software environments that facilitate forms of live processing. However, on stage and in concert halls, the most readily available technologies serve only as playback mechanisms of pre-recorded content. It is the very notion of recorded media that creates the distinction between that which is "live" and that which is not [8]. There was no such distinction prior to the advent of recording technologies. Prior to the phonography, music had to be performed, either by professionals in concert or by individuals at home, in order to be heard. Before the printing press, the mass distribution of stories was through telling and enacting them, not through the static, written page. Today, the vast majority of the media we consume is recorded: television, cinema, downloaded music on our MP3 players. This, however, has not diminished our desire for live experiences and perhaps has only heightened the need.

Intermediary forms of live experiences include shared experiences of pre-recorded media, such as going to the cinema with an audience of at least a few dozen strangers versus watching a DVD at home alone. The on-demand nature of recorded media and the ability to watch it (increasingly so) when and wherever we please has not discouraged our need to partake of the shared experience. The shared experience serves several social functions. There is social value and comfort in attending an experience with friends and acquaintances and, from the planning or dinner preceding an evening at the theater to coffee afterwards, the anticipation or preparation for such an event is part of the experience. Once the performance or screening ends, we often discuss it with those that attended it with us, sharing perspectives and reactions. Such experiences also make an individual feel as if they are a part of a community of like-minded people who have similar preferences or desires for the type of performance. During the performance, audience reactions—laughter, frisson, suspense, applause, et cetera—validate our own emotional reactions. Technologically mediated forms of entertainment can adopt aspects of the shared experience of live performance. The concept of "appointment television" underscores the ephemeral sense of the live experience, even when we are watching in solitude or separated from our friends and co-workers with whom we will later discuss the events of the program. Computer and video games in the form of massively multiplayer online role-playing formats and other networked multiplayer scenarios encourage engagement with the media experience and others despite distance.

Live performances offer additional aspects to the experience beyond the immediate social functions of a shared viewing. The ephemerality of a stage show or the virtuosity demonstrated in a live concert brings us in contact not only with other audience members, but with performers. We may not interact directly with those onstage, but we share presence with them, as well. We are aware of them as present and existing in the moment and sense that they are also aware of us in the audience. The live performance creates this connection between audiences and performers in a way that cinema and television cannot and, consequently, the demands on our attention are also greater out of the sense of being in the moment and mutual respect. Live performance also introduces notions of jeopardy and privilege into our experience. As contrasted to film, which can be reshot and edited until the assembled performance matches the artists vision as precisely as possible, live performance is subject to error, variation, and imperfection. This makes a flawless virtuosic performance that much more impressive. It keeps us on the edge of our seats for an improbable acrobatic stunt, a ballerina's solo leap, an actor's conjured rage and sorrow, or a feat of legerdemain. There is a sense of potential failure or the unexpected and we are rewarded in either case by the impressive or the singularly unique. Liveness privileges the audience—that particular audience at that particular performance—with a one-of-a-kind experience that can be related and compared to others' recollections of different performances of the same production or simply treasured in one's memory.

The mediatization of live forms of performance dulls and detracts from the sense of liveness by introducing static elements in the form of playback and projection that extends beyond merely defining the mise-en-scène. Static forms of media introduced within the live context break the immediacy and ephemeral connection audiences experience. Yet, the scale of the projected image or its emissive, luminous quality is distracting, diverting our attention away from the dwarfed human performers toward the bright flicker of the recorded image [75]. The language of the visual vocabulary of mediatized forms follows that of film and television, bringing passive screen-based experiences and dialects into the theatrical realm. The live performance medium becomes one of cinema augmented by relatively perceptually insignificant live action rather than theater augmented by additional expressive modalities. While the balance of the stage picture must account for the scale and physical quality of the mediated elements, it is the lack of liveness that presents the true experiential problem. As soon as we see a technologically reproduced image, we lose the sense of risk, knowing that the presented image will perform as expected. Our sense of privilege shifts from the live moment to one of the camera's gaze. However, this problem is not inherent with the modalities, just the way in which they are used. It is difficult to create live images of sufficient quality in a meaningful and compelling way with existing practices. As we'll see in the next chapter, my work in the Opera of the Future group has taken strides to create technologies and methodologies that enable the incursions of liveness into the technologically-laden form. This simple fact is often one that defies expectation. In describing the methods of Disembodied Performance (Section 3.1.1) and other techniques discussed, I often receive the surprised and awed reaction, "so what you see is actually different each night?" The answer is, of course, “Yes.” Technology can be used in performative, ephemeral, and essentially live ways.

Audiences and experience creators are also craving more *immersive* experiences. Immersive experience can be interpreted in several ways, of late. Immersion can be used in a more literal sense to describe promenade theater performances or other non-proscenium forms of presentation where the action happens in or around the audience. Immersion also may include additional modalities in the presentation that are not typically associated with the medium. For example, a theatrical performance or installation that included an olfactory component might be billed as immersive. The immersive qualifier can refer to higher-fidelity reproductions, as well. A greater sense of realism, higher resolution displays, surround sound systems, three-dimensional projection techniques may all be considered to create a sense of immersion.

Immersion also may entail agency and presence. Is it simply enough to passively be surrounded by the world in which the story we are viewing unfolds, or should we become a functional entity within that world? The additional, enveloping modalities make the constructed or fictional experience feel more real and, given that illusion of present reality, it would follow that we, as audiences, would be able to take on a performative role.

From this line of reasoning, I would posit that both liveness and immersion are addressing the same need: the desire to have an experience that feels like a plausible reality. What we seek is not simply the novelty of stimuli in additional modalities, but a sense of being there in the moment. Regardless of the level of interactivity afforded by the experience, once immersed, we are no longer passive observers of the world and the narratives it contains. As audience members, participants, players in an immersive experience, we do not passively watch an established character with whom we identify in that other world, but we readily suspend our disbelief and step wholly into another world and become another character within it. We take on a role—perhaps an alter ego or a character that has admirable traits or that is less inhibited than our real-world persona—and, though performative, within a sufficiently plausible world, we don’t merely play that role, but we assume that role. For a time, we live also as that individual, immersed within that world.

If we accept evolutionary theories of music that claim that music is an exercise in time segmentation, change perception, and prediction all with the unconscious intent of honing survival skills, then let us consider the need for liveness and immersion in entertainment experience similarly. Plausible realities serve as hypothetical circumstances, simulations, simulacra of life itself. They are safe, but believable scenarios for testing one’s ability to navigate, defend, make decisions, interpret causality or motivation, and predict outcomes. Immersive experiences can test and refine fitness and adaptability to function in fantasy world scenarios where the rules of common, every-day existence need not apply. Through them, we can cope with violations of expectation that do not have grave life consequences and incorporate these lessons into our models for interaction.

These sorts of immersive experiences can take numerous forms from computer games to live action role-playing (LARP). The theatrical work of the British theater group Punchdrunk provides a good example of a middle ground between open-ended role-playing and a conventional theatrical context. Punchdrunk performances—such as their successful production of *Sleep No More*, which will be discussed in Section 3.3 as part of a technological intervention into the location-based form on which I worked—are typically situated in an impeccably detailed multi-room space that defines the physical world of the show. Dancers and actors move about the space, acting out the performance's story. Audiences in the Punchdrunk experience have the freedom to move through the space and discover both the action and the world as they please. Often given masks, audience members are both anonymized and transformed into elements of the diegesis [58]. They are empowered to act and assume a performative role through the agency afforded them by the medium, but, unlike LARPing, the audience members are not essential to the unfolding of the narrative. The story progresses and is acted, regardless of the presence of any audience. This balance between immersion and theatricality ensures that a particular story is presented in a particular way.

It is not merely enough to exist in an open-ended world, just as words on a page are insufficient to create an engaging narrative. Like theatrical experiences and books, music must also have a structure with a beginning, middle, and end. We crave a story. We understand our world and our goals through story. Our lives are not episodic or structured, but we tell *stories* from the haphazard events of our life, discarding unessential moments, confabulating causal relationships where perhaps there were none, and giving ourselves goals. We communicate, reason about, and make sense of events through story. Thus, we expect musical works, prose, theatrical experiences, and live, immersive experiences must also be shaped.

Because of the narrative independence of the audience, Punchdrunk experiences don't impose goals. They are not games, but autotelic goals are brought to the experience by veteran audiences, a process that is happening at both diegetic and non-diegetic levels [29]. The behavior of the individual audience member is governed by their own desire and free will, curating their encounter with the work. The abstracted nature of the performances, relying more on movement in space and dance-like expressive vocabularies for storytelling over text, leaves considerable room for interpretation of the story. Furthermore, the scale and temporal parallelism of the performance ensures that no one audience member can experience the entirety of the narrative content along a multiplicity of threads within a single viewing. The production is intentionally open-ended. Thus, it is up to the audience to arrive at a meaning and interpretation of the experience by interpolating from what glimpses of the world they captured and the order in which they happened to come across them. Scenes may be viewed out of plot chronology with numerous elisions, yet, as when threading together the fragments of a dream upon waking, each audience member must piece together meaning from an assembly of scenes, gestures, and a pervasive musical soundtrack. Although the interpretation of the story is open-ended and subject to an individual's unique experience, the entirety of the experience is indeed shaped and finely crafted.

For all of these reasons, as designed, expressive, artistic experiences become increasingly live, individualized, and immersive methods and tools for creating, documenting, and storing experiences that span multiple modalities and with both linear and varied ways of unfolding as a result of the audience's agency are required. As we've seen, conventional scores can create a shape in time and structure the creation of expressive content in real-time. They effectively negotiate the fluidity of time and the symbolic communication of intention. However, the conventional score does not fully capture the breadth of flexible, interactive, and open-ended storytelling experiences in multiple modalities. The current techniques of creating technological or technologically augmented works define the rules and the vocabularies of expression, but lack the communicative power to create story and shape seamlessly across a range of modalities. Media Scores take a step towards this compositional paradigm.

3 SYSTEMS IN PERFORMANCE

We're not a film in a can on a piece of film or digital tape. We can be live.

—Diane Paulus on *Death and the Powers*

This chapter will highlight a few of the significant projects to which I've contributed that involved the creation of complex performance technologies in order to realize an artistic vision. The capabilities of each of the resulting systems vary with the needs of each project, which are somewhat diverse in scope. However, several commonalities can be found among the system architectures. These can be generalized into a set of design goals for technological performance systems that facilitate intricate multi-modal artworks.

All of the systems discussed here were developed for live performance works. However, that being a more demanding condition than static artworks (a painting, a novel, a film; those that are not themselves altered at the time of consumption), the principles derived from my experiences creating and using these systems are also applicable to, or at least do not interfere with, systems for non-performance art. In many cases, the responsive feedback necessary for the rapid creation of content and realization of live performance makes possible considerable experimentation during the composition process, regardless of the intended presentation format of the work.

The arts are communicative forms of expression that leverage our experiences, our awareness, and our sensitivity to the synaesthetic and the intangible to communicate an idea from the mind of the artist into our own. In particular, the magic of live theater and live performance emerges from several factors, extending the power of the experience. In several ways, it is about presence. Audience members occupy the same space as performers, as well as other audience members. Together they co-experience the performance providing subtle feedback about the presentation, validation of each other's reactions, and fulfilling a social and communal function [19]. In addition to a sense of presence, liveness is a function of time. There is risk and uncertainty inherent in the experience. Whether it is technological or human, regardless of the amount of planning and rehearsal, the performance is executed in the moment and we implicitly feel the variability, the potential for anything to happen. Oftentimes, the more memorable moments are when something unexpected does occur—something that would not have made it into the final form of a static work, that would have been cut from the film, or discouraged in rehearsal had it happened before. This risk, this uncertainty, this uniqueness makes attending a live event a personal experience. The audience is privileged with seeing and participating in something that will never happen in the same way again. It is ephemeral. Even more so in interactive works, the experience is something that might only have happened because they were there and apart of the co-creation of that instance of the work. Common to artworks beyond live performance is the suspension of disbelief. However, when this happens in a live context—when the impossible appears to happen before one's eyes or one loses the sense of one's self and is transported into a different reality—we are moved.

In mediatized culture, we have come to expect the impossible. From George Méliès early cinematic special effects to the latest in Hollywood's visual effects, the illusion of reality on the screen has grown to encompass worlds we know to be impossible [15]. Méliès, however, introduced the techniques of illusionists who convincingly defy reality in live performance to the recorded medium where it has grown beyond what could be created live. Nowadays, we take for granted that the world projected onto the flat screen in front of us can take any form and have the acceptance that it is not live, static, and manipulated. The layer of indirection that recording has added to our sense of experience has all but removed a sense of presence from mediatized forms. We cannot reach into those worlds. They are constructed and trapped forever behind a plane of glass, a moving picture, but not a plausible, truly immersive reality. Their ubiquity, their emissive light, however, can't help but draw our attention in. Conversely, when introduced into the theatrical or live context, they can't help but to create a sense of distance [9]. Pre-recorded and technologically-presented media are not live, not dimensional, not fluid in time, present, or subject to variation in that ephemeral way the live context is. This need not be the case.

To create liveness and the ephemeral sense of immersion in a live context, the technological tools of creation and reproduction of mediated content must be flexible enough to meet the needs of the live performance scenario. They need to handle time in a fluid and malleable manner during a performance. This also impacts the rehearsal and content-creation phases in developing a performative work. In these contexts, narrative-time or experience-time is continually being stopped and restarted. Variations in content, material, and story can be suggested, tried, and modified in a rapidly iterative process. Human performers can extemporize and improvise efficiently within these situations, but technological performance systems are generally less flexible. For performance technologies to be fully integrated into the creative process and be used as part of the ephemeral moment in theater, the technologies need to support and facilitate the essential qualities and broaden the palette of what's possible; extending presence, connection, and the expressive power of the Gesamtkunstwerk.

What is needed are technologies that function not unlike instruments and actors, adapting quickly and producing variable and expressive output synchronously with the remainder of the production process. Acoustic musical instruments are tools on which a performance may be rehearsed, but there is nothing intrinsic in the instrument that necessitates preparation. Any content can be performed on the instrument, within the constraints of the physical nature of the instrument, at any time.

Live television production has evolved tools that enable a story to be told from a number of different live or existing sources. The video switcher is a tool for real-time editing with the storytelling eye of a skilled television director. The cameras can quickly train in on important action and the technical director can quickly configure the switcher to take the appropriate source and execute the desired transition.

The traditional technological infrastructure within the theater can be quite responsive to real-time creation, but generally has an implicit assumption in the systems design that a planning and rehearsal phase is part of the production practice. Most theatrical systems are cue-oriented. Cues are created in advance of the production and executed at the appropriate time during the live event. In part, this is due to the complexity and large number of variables being controlled by the system. Present-day lighting consoles, for example, are capable of controlling hundreds or thousands of lights independently. It is unrealistic to expect that an operator or designer be able to manipulate each individually at performance-time, so cues are essential. As a result, cues are used to save an ordered set of states designed for the show. These are discrete states in order for them to be executed flexibly in time. A stage manager would call lighting cues during a performance at the appropriate time accounting for any variable in the timing of the performance. This also has the effect of rendering the lighting reproducible from one performance of a production to the next, despite the complexity or variations in timing. However, even if not fixed in time, this is an ordered set of predefined states.

Changes in time are well-accounted for, but not changes in order or spatial variations. This is why it is crucial for actors to hit their marks onstage, because the technologies are not robust to these other sorts of variations. Relatively recent features being added to lighting consoles do move toward being able to manage the complexity of the medium while taking into account more types of variation. These consoles allow instruments and properties of the lighting to be grouped and recalled instantaneously. An operator can independently recall colors, lighting functions, and intensities in a manner very similar to how video switchers are used. The creative process and tolerance for change and potential for technological improvisation is arriving in the theater.

Though these capabilities are starting to appear and be refined in the context of performance technologies, many technologies and approaches to their use in a theatrical context do not yet embrace these essential qualities of variability and ephemerality. Particularly, forms that derive from other media, such as video and film, bring with them the static nature of their source medium. Magic lanterns and moving image projection in theater requires that the image already exist on glass or celluloid [37]. The avant-garde practice of video Foley in theater, counters this with live image creation for stage projection and has its roots in the performance color organs (see Section 2.7) and more advanced technological analogues in VJing and live cinema [2].

For use in live performance, technological performance systems should be able to function with aspects of all these paradigms of interaction and responsiveness. They need to be tolerant of variations in time, as well as space, as necessary. More advanced systems to support improvisation or unexpected events could also be imagined to be tolerant to variations in content or intention. The performance can be planned, scored, cues written in advance, while still affording flexibility. The systems should serve to simplify the control over complex sets of parameters through grouping related functions or abstracting the details of the medium into readily manipulated and easily reasoned about high-level concepts. As we'll see in the following examples and discussion, to shape live content in the rehearsal and design process in this way, performance systems need to provide immediate feedback and forms of direct manipulation of the piece, not solely simulation or representation of the final form, but the final form itself. These principles will ultimately be broadened from the rehearsal and performance process to all phases of production through Media Scores.

3.1 Death and the Powers

Death and the Powers, finalist for the 2012 Pulitzer Prize in music, is an opera by Tod Machover with a libretto by former US Poet Laureate Robert Pinsky that tells the story of a wealthy and eccentric entrepreneur and inventor, Simon Powers [62]. Powers is weary of the world, but does not wish to relinquish his powerful business influence or become disconnected from his family. His ailing health leads him—with cooperation from his adopted research assistant, Nicholas—to develop The System, a technological infrastructure embedded throughout his home, into which he can upload his essence and consciousness upon his death. While The System represents a form of immortality, Simon particularly embraces the transcendence of the body and the material world. Theatrically, we see Simon Powers onstage at the start of the opera and then watch as he is transformed into the scenic environment and the setting comes alive with his character, physicality, and larger-than-life presence for much of the production. The remainder of the opera explores the implications of this potentially eternal and non-corporeal form and how Simon's family—namely his third wife, Evvy, and daughter, Miranda—and the world at large cope with his new form of existence.

Death and the Powers was given its world premiere performances in Monte Carlo, Monaco in September of 2010 under the haut patronage of His Serene Highness Albert II, Sovereign Prince of Monaco with contributions from the Futurum Association. In 2011, *Powers* played at the Cutler Majestic Theater in Boston and the Harris Theater in Chicago. Additional performances worldwide are expected.

At the core of the opera's design was a desire by Machover to innovate on the operatic and theatrical form, looking at ways to physicalize the experience of the music and storytelling onstage beyond conventional approaches. This is a reaction to the mediatization of the theatrical form. Machover notes, as do Auslander and others, that the proliferation of screens and projection in theater reduces the dimensionality and the sense of humanity and ephemerality that make live performance uniquely rewarding [59]. Machover believes that commonplace use of projection and image magnification of performers have a distancing and dehumanizing effect. When video is used to increase the scale of a performance, the "lure of the screen" supersedes the direct view and palpable connection of audiences and performers sharing the same space. Machover's guiding vision was to escape the flat confines of the screen and not simply reproduce the image of a performance, but to extend the human qualities of that performance into the entirety of the physical environment. He envisioned a "choreography of objects" and an immersive sound design associated with these objects that would draw audiences closer to the character, to feel the performance around them. Early in the project's development, several stories were explored, ultimately settling on what would become *Death and the Powers*. The story necessitated that the theatrical environment become animated as a believable, present character.

This story serves as an inner narrative that is framed by a mysterious prologue and epilogue set in the distant future featuring a cadre of robots that have been programmed for reasons incomprehensible to them to ritualistically perform this pageant-play about the Powers family. In a cinematic transition from the prologue to the Powers household, four of the robots transform into the four principle human characters that they portray for the pageant. Within the inner narrative, the robots, *operabots* as the production refers to them, function in diegetic roles, as setting and illumination, and as a Greek chorus reacting to the action.

Figure 24: *Death and the Powers* onstage
Simon Powers (James Maddalena) is surrounded by his family in Scene I. (Photo by Paula Aguilera)

In addition to the chorus of operabots, the set features three large periaktoi that are also robots capable of moving independently about the stage and a chandelier-like structure suspended above the stage. The animatronic nature of these set elements enables Simon in The System to omnipotently inhabit the objects from time to time, emoting and communicating through them. Each of the periaktoi has three translucent faces which can be illuminated from within the structure, as well as a low-resolution LED display embedded in the walls resembling a bookshelf. The chandelier is an intricate, stringed object with internal LED illumination. In one scene, Simon occupies the chandelier and it descends and encloses Evvy as they reminisce. She then begins to play the strings of the chandelier by plucking and strumming them, producing characteristic sounds and modifying Simon's voice.

Audio production plays an important role in *Death and the Powers*. Two formats of sound spatialization are used in order to both connect sound to the dynamic physicality of the set, as well as to extend the sense of presence and omnipotence of Simon in The System from the stage and into the house. Third-order ambisonics is a surround sound technique that relies on a ring or spherical section of speakers placed around the audience to create the impression of moving sound sources at the periphery of the space. The ambisonic format was particularly advantageous, as it does not rely on consistent speaker placement, enabling the setup to tour easily [48]. Content could also be authored in a studio setting and faithfully presented in the theater without additional modification. Additionally, smaller satellite systems can be added for improved spatialized sound reproduction in loge or box seating areas. The second format of spatialized sound is wave field synthesis (WFS) which uses a long array of small drivers across the stage apron and is capable of synthesizing sound waves that appear to originate from any point in space. The location of objects, robots, and singers on stage is tracked in three dimensions using an Ubisense ultra-wideband active RIFD system. The position data can then be used as the computed origin for a sound source produced by the WFS array, providing natural sound reinforcement for singers, as well as clearly voicing physical objects onstage with a fidelity that could not be achieved using conventional sound reproduction techniques. The application of WFS in *Powers* is one of the first examples of using this technique on live sound sources in performance.

My involvement with *Death and the Powers* began in 2007 as the production moved into a critical phase of translating the creative team's vision for the opera into theatrical reality. I worked closely with composer Tod Machover, production designer Alex McDowell, and director Diane Paulus, to explore and refine design options for both the theatrical set elements and the visual representation of Simon in The System, consulting on the theatrical and technical feasibility of various options and the aesthetic implications of design decisions and how they would serve the story. As discussed in the remainder of this section, the needs of *Powers* necessitated the creation of novel software systems and show control architectures. I wrote initial versions of control software and authoring tools for the robotic elements, operabots and walls, as a unified show control system intended to control sonic and visual aspects of the production. As the needs of *Powers* evolved, control of sound and visuals branched into independent development. I ultimately focused my software contributions on the creation of a mapping system and a system for authoring and rendering the visual representations of Simon in The System. I also developed the visual language for graphical elements on the walls and created dynamic content using these systems. For further descriptions of the overall design process and implementation of *Death and the Powers*, I refer the reader to Chapter 3 of [107] as well as to [106] and [48]. It should be noted that the first resource, which details the design process, was written while the opera was in development and does not entirely reflect the final implementation of the walls, robots, and visuals systems. In the remainder of this section, I'll describe three of the Powers' systems in additional detail.

3.1.1 Disembodied Performance

An important creative challenge in creating *Death and the Powers* was to create a believable and compelling impression that Simon in The System is omnipotent and omnipresent through the various set elements and spatialized audio. All of these representations need to feel connected and alive as a unified, larger-than-life presence. To this end, I developed the technique of Disembodied Performance, as detailed in [107]. Using this approach, the theatrical environment of the opera is animated by the offstage performance of opera singer James Maddalena. We see Maddalena onstage as Simon Powers in the first scene after which he begins the process of entering The System and his presence is transformed from his human body to the theatrical set. The actor himself then enters an isolation booth in the orchestra pit. There among the musicians, in view of the conductor, and still very much a part of the performance, he continues to sing and act his role for the remainder of the opera until he momentarily re-emerges as an apparition in the final scene.

While offstage, Maddalena dons gesture and physiological sensors designed by my colleague Elena Jessop. Accelerometers on his arms and hands

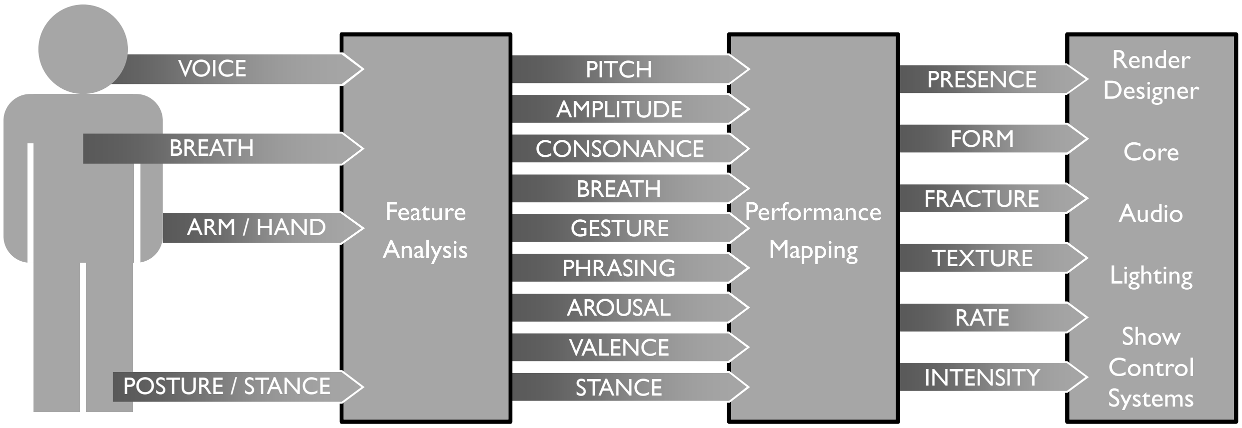


Figure 25: Disembodied Performance mapping

measure his expressive movement and were positioned to sense his normal mode of performance which focuses on emoting through subtle hand movements. A breath sensor around his torso captures the expansion and contraction of his chest cavity in response to his singing. Breath is a particularly important feature to sense as it has a characteristic signature of "being alive" and closely follows the natural phrasing of the music as Maddalena sings it. These sensors are wireless, so as not to impede his performance, and the data from them are sent along with audio data of his voice are sent to the mapping system described in section 3.1.3 for analysis and translation into various modes of representation onstage.

Disembodied Performance abstracts Maddalena's performance away from his human body and re-embodies it in a variety of non-anthropomorphic forms. It reduces the sensed data into a set of meaningful expressive parameters that then can be used to create and shape various aspects of the performance from visual elements, animatronic movements, to transformations of the singer's own voice. In this approach, we focus on the quality of the physical and vocal gestures made by the performer, thereby reflecting the emotional journey of the character rather than the experience of the performer. The performance parameters used include both aspects that the performer is aware of and those that are unintentional, but all are part of the actor and singer's craft: in what he is trained and skilled. We're looking to capture the same essence of the performance that would normally communicate the character's emotional journey to an audience, were the performer onstage. From this we are able to recreate equivalent stimuli in the mise-en-scène with different form. It matters not so much what the performer is doing—what gesture he makes with his hands at any given moment or what note he sings, the latter being prescribed by the score—but rather *how* he is gesturing. This emphasis on quality recalls the tradition of Hyperinstruments in our research group by extending the range of a virtuosic performer and leveraging his natural talents and abilities. However, the actor is not playing the instrument. He is simply performing. There is no new gestural language or instrument that needs to be learned.

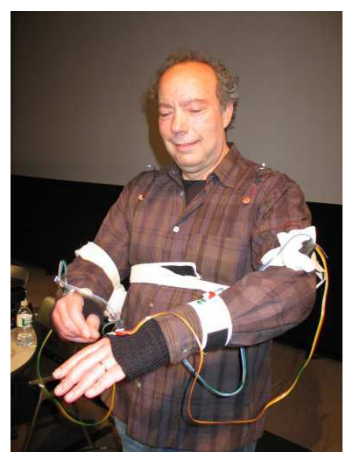


Figure 26: Disembodied Performance sensors
James Maddalena wears wireless gesture and physiological sensors during testing and calibration.

Disembodied Performance was an essential technique for telling the story of *Death and the Powers* and the artistic conception of an environment that comes alive as the main character. The translation of a single live performance into the multitude of required modalities necessitated a unique set of systems. We will now look at three systems that were created to address the needs of Disembodied Performance in *Powers*.

3.1.2 Robot Choreography

Figure 27: Operabots
Operabots onstage during *Death and the Powers* (Photo by Paula Aguilera)

Disembodied Performance provides the model or essence of Simon in The System, but it relies on that information being manifest by different modalities in the theatrical environment in order to represent the character's performance. Simon's presence in the system takes on physical, visual, and sonic forms. Once technical development on the opera began in 2008, I set out to create a unified control system that could handle the demands of interpreting the performance through the use of robotics and light. The purpose of a single system would be to design gestures and effects in a single place that could extend across multiple modalities. This control system, implemented in Java 6, became known as Core during its development. I envisioned Core as a type of non-linear animation system that could easily execute pre-programmed animations, respond to live input, and procedural effects. These types of control would then be blended and transitioned as part of the design of the show to produce complex, hybrid behaviors. Core's architecture was originally inspired by the Architecture for Control Networks (ACN) entertainment control protocol suite [33], though due to lack of robust library implementations at the time, Core itself did not implement the protocol itself. The architecture is device-agnostic, relying on modular *device descriptions* to specify the capabilities and requirements of connected devices, such as the operabots. As the overall development of the different modalities representing Simon in The System evolved, the control needs for the visual components changed correspondingly. At that point, led by my colleague Michael Miller, development on Core then focused primarily on creating a system capable of designing and controlling the production's animatronic systems for the operabots and periaktoi [68]. Local encoder feedback combined with the 3D position data captured by the Ubisense system gives onboard robot systems closed-loop feedback over their location and attitude onstage. This information is then relayed to Core in order to ensure that each robot hits its mark.

Core makes a distinction between devices and roles. For example, a particular operabot may be assigned to perform a particular role. If the physical operabot experiences a failure, an operator can quickly assign its role to a different hardware unit. Each device in each role is composed of one or more axes that represent a controllable parameter. All parameters have normalized signed or unsigned floating point values in the ranges [−1.0, 1.0] or [0.0, 1.0], respectively. Sequences in Core are sets of devices

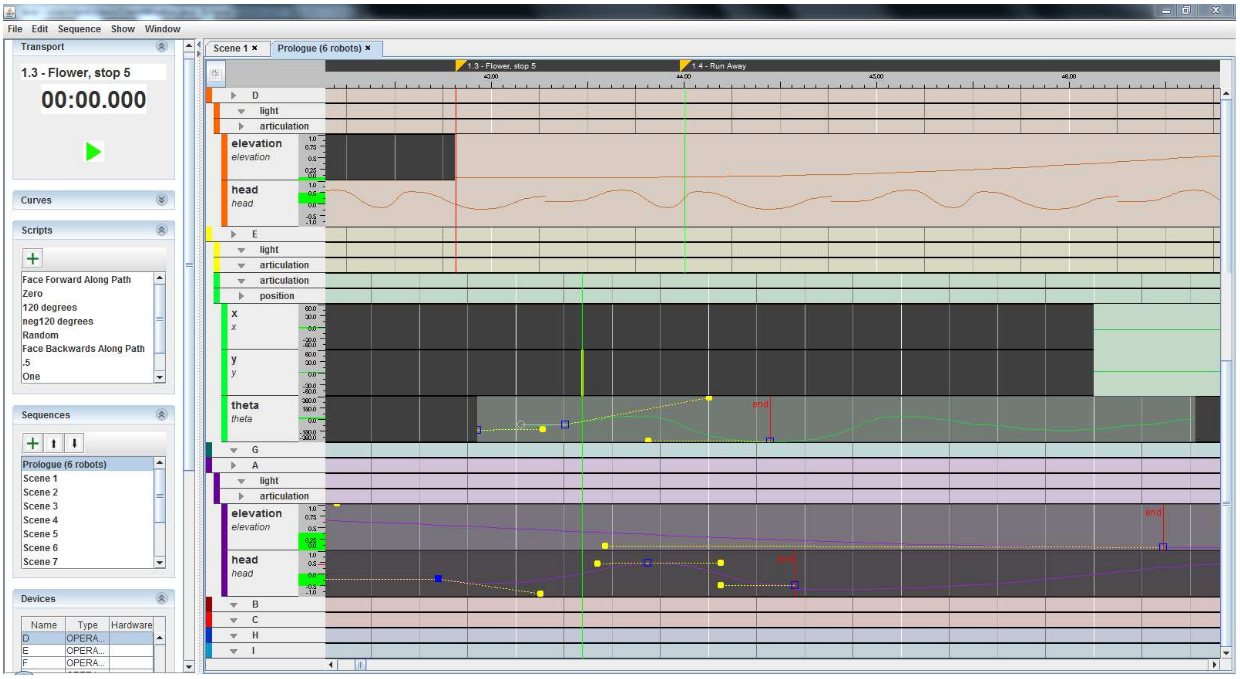


Figure 28: Core robot control system

Screenshot of the robot control system for operabots and periaktoi. Each track represents a parameter of each robot for a given role.

and their axes and are presented to the user in a track view. The user can specify manual, procedural, or keyframe control over the value of a particular axis at a given time. During playback of a sequence, the output values of each of a device's axes are sent to that device over wired and wireless networks using the Open Sound Control (OSC) protocol over multicast UDP [117].

In *Death and the Powers*, all of the operabot functions—omnidrive translation and rotation, elevation and head articulation, and lighting—could be programmed for control by live input from the Disembodied Performance or human operabot operators, as well as pre-programmed sequences. For purposes of safety in critical configurations and moments of difficult robot choreography, much of the operabot translation is manually controlled. Drive systems for the periaktoi, however, are pre-programmed in order to reproduce precise transitions in setting from scene to scene. Although as many as nine operabots are onstage at any point in time, only four operators and one person to oversee Core's operation are used in the production. Each operator has a networked computer with a display showing the current state of the robots overlaid on live infrared video of the stage from above. Additional status information includes the current sequence, cue, and time of the choreography being executed by Core. Each terminal is equipped with a video game controller that operators may use to drive or puppeteer functions of the operabot. To account for the disparity in number between operators and operabots, each operator can assume control over an operabot by selecting an unselected robot from the interactive display. Core then connects the operator's controls and overrides

Figure 29: Core operator console and simulator
During performance, the operators can monitor playback of the core sequence window (top) along with a 3D simulation (center) and robot status overlays (bottom left).

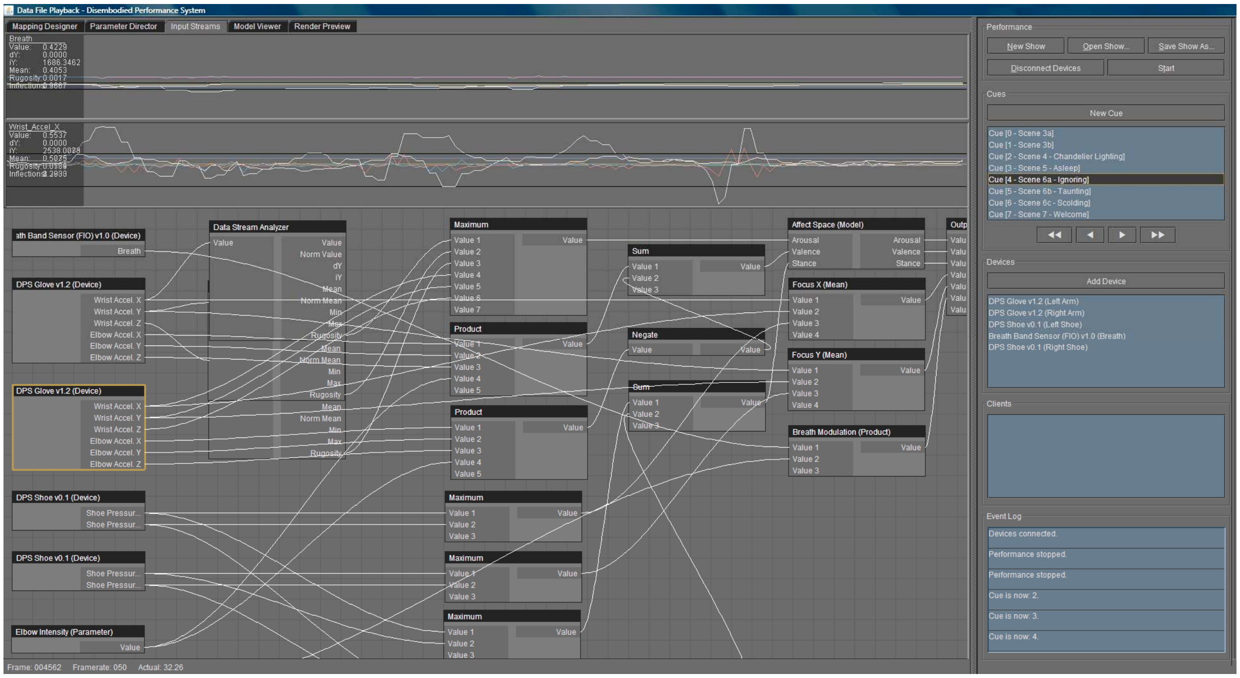
to the appropriate device as defined at the time for the operabot's current role.

In addition to its track views and show management functions, Core displays a 3D simulation of the operabots and walls onstage. This gives real-time feedback about the current state of all of the show's robotics to the Core operator during performance. The simulation view is also useful for offline editing. Sequences can be choreographed in Core and visualized in the simulator, playing them back on a model of the stage in time. Much of the robot choreography for *Powers* was accomplished offline by the Core operator and directorial staff before rehearsals began and the systems could be assembled in the theater. Once in the theater, the graphical timeline and keyframe interface of Core allowed for robot choreography to be quickly refined and adapted during the rehearsal progress.

3.1.3 Mapping System

Conceptually, Disembodied Performance relies on the process of reified inference discussed in Section 2.1. This is realized through the analysis of real-time performance data through an intermediate parametric model and mapping the resulting state to parameters of the various systems that realize the Disembodied Performance onstage. I wrote a mapping software system tailored specifically to this application that uses node-based flows to transform data from input sources to output destinations.

Within the application, the user can specify a set of input devices and named output parameters that are common to all flows associated with a particular show document. These appear as input nodes and output nodes in the mapping interface. The show may contain one or more cues that define a particular mapping flow. The user constructs a mapping for each cue by adding processing nodes to the interface, which typically represent arithmetic or statistical operations, and connecting the output ports of nodes with the input ports of other nodes. This ultimately creates a tree structure from the output node which can be evaluated to produce the



output parameters. The values of data within a mapping use a common floating point type, unlike other node-based systems. Regardless of the input devices, all data entering the mapping flow is nominally in the normalized signed or unsigned unit ranges of [0.0, 1.0] or [−1.0, 1.0], just as the values of axes in the Core control system. All operations preserve these ranges through to the output.

Figure 30: Mapping System
Screenshot of the mapping system showing live performance data analysis (top), the currently selected flow-based mapping (center), and a cue list of other mappings (right).

The application interface also provides real-time feedback on input data and its analysis over a defined window of time through dynamic plots for each data stream. These plots are useful for verifying that the Mapping System is receiving data from the input device and for investigating the ranges and behavior of the data. Parameter nodes can be added to the flow to specify user-adjustable values. Another view displays a slider for each parameter node in a layout similar to faders on an audio console or submasters on a lighting console. Using these sliders the behavior of the mapping can be mixed by the user to refine the mapping in response to input.

The application initially starts in an offline mode. No data is processed, but cues and flows can be created and edited. Some input devices are capable of streaming data directly to the Mapping System when they are online, such as data sources that generate OSC over UDP. Other input devices, such as serial devices or networked systems that require the use of the TCP protocol, must be connected to the Mapping System. The system can be brought online and initiate connections with such devices if they are available. When online, data streams are analyzed and the active mapping is computed and output. Cues, and therefore mappings, can be switched from the user interface or through remote commands and triggers as part of a larger show control network. Cues may have a transition time and executing a cue can interpolate output values between two mappings. The values of parameter sliders and the structure of the mapping can be modified at any time for the active flow displayed in the user interface. Thus, when connected to input devices and output systems, modifying the mapping has an immediate effect, enabling rapid experimentation during the creation and refinement of mappings and tuning adjustments during performance.

In *Death and the Powers*, many dimensions of vocal, gestural, and physiological performance data are analyzed and reduced to a relatively small-dimensional space to represent the affective quality of the performance within the Mapping System. It is the continuous value of these few parameters that encode the perceptually and intuitively meaningful aspects of the performance independent of the physical form through which they were expressed, thereby enabling the reconstitution of the performance through visuals, robotic movement, and transformation of sound. The small-dimensional parameter space is not some arbitrary reduction. Rather, the semantic definition of each parameter was carefully chosen to both span the set of expressive behaviors that were observed from initial performance data experiments and to span the range of physical, visual, and sonic effects that were envisioned for the production. As part of this mapping process, a three-dimensional circumplex model of affect is used in the intermediate stages of the Disembodied Performance mapping [16]. The affective parameters of arousal, valence, and stance were then mapped into a six-dimensional parameter space that defined expressive parameters suitable for interpretation by the visual, sonic, and animatronic output systems. The six performance output parameters are:

PRESENCE	how attentive or involved in the scene is Simon Powers, scale
FORM	how nebulous or defined is Simon's presence
FRACTURE	how focused or atomic is Simon's presence, versus disjoint and omnipresent
TEXTURE	how complex or varied is Simon's behavior
RATE	how animated or excited is Simon
and INTENSITY	the degree of spectacle or assertion Simon is imposing

The parameters are defined such that they have meaning in the various modalities used in *Powers*. Several were inspired by Machover's descriptions of his sonic vision for Simon's processed voice in The System. The show systems that receive these parameters as input, such as the visuals system discussed in the next section, each interpret the parameters appropriately for their output modality.

3.1.4 Render Designer

In order to generate the dynamic visual imagery that is displayed on the LED faces of the periaktoi, I developed a system called Render Designer that could be used to create a variety of abstract, non-anthropomorphic representations of Simon in The System using the real-time data from the Disembodied Performance analysis. After unsuccessfully searching for software that would function in the manner required, I elected to develop my own. Like the other systems mentioned so far, Render Designer was implemented as a Java 6 application with hardware accelerated rendering through the JOGL OpenGL API.

Render Designer is effectively a combination between a lighting console and video compositing software. A Render Designer project consists of a raster configuration that describes the scale and proportions of the graphical output it produces for display. In *Death and the Powers*, this raster definition configures the graphics system for the four display faces of the periaktoi. The center display region can be mapped differently on the center-most two display faces depending upon the configuration of the walls. Additionally, each show document contains one or more cues that define modes of rendering in the same manner as cues define flows in the Mapping System. Again, cues can be executed in sequence or remotely triggered and can have transition durations resulting in the interpolation of cue parameters and a crossfade of the rendered image. Cues may also retrigger other systems. In *Powers*, this capability enabled mappings in the Mapping System to be reused with different graphic configurations. While

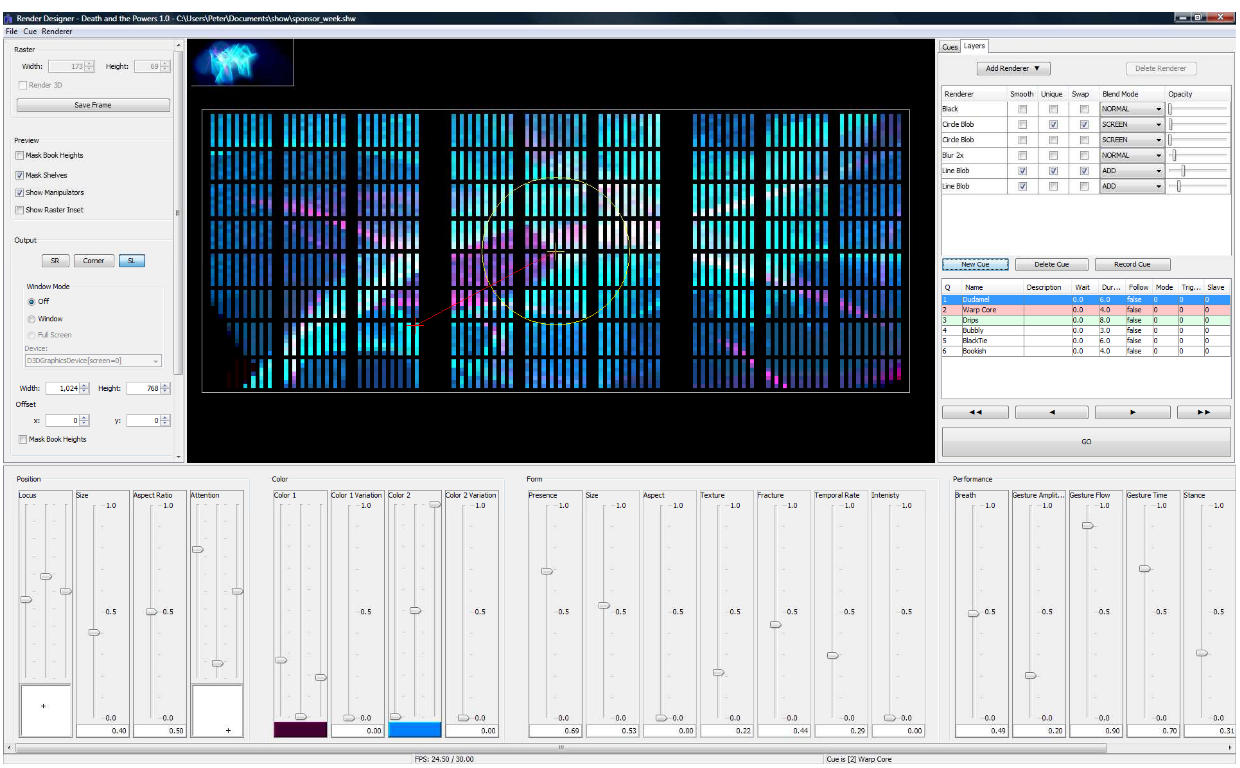


Figure 31: Render Designer
Screenshot of the rendering system design for *Death and the Powers*, illustrating the generated image (center), renderer stack (top right), cue list (middle right), and continuous parameter values (bottom).

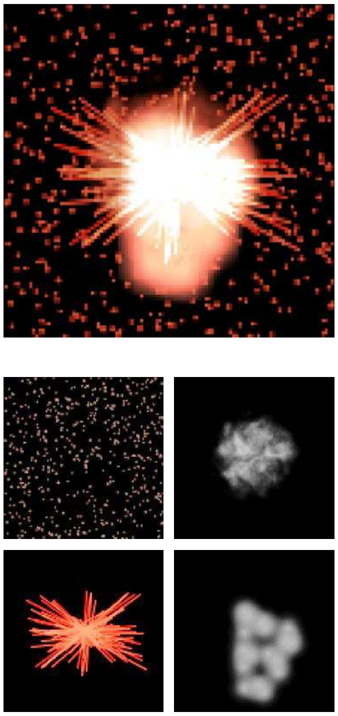


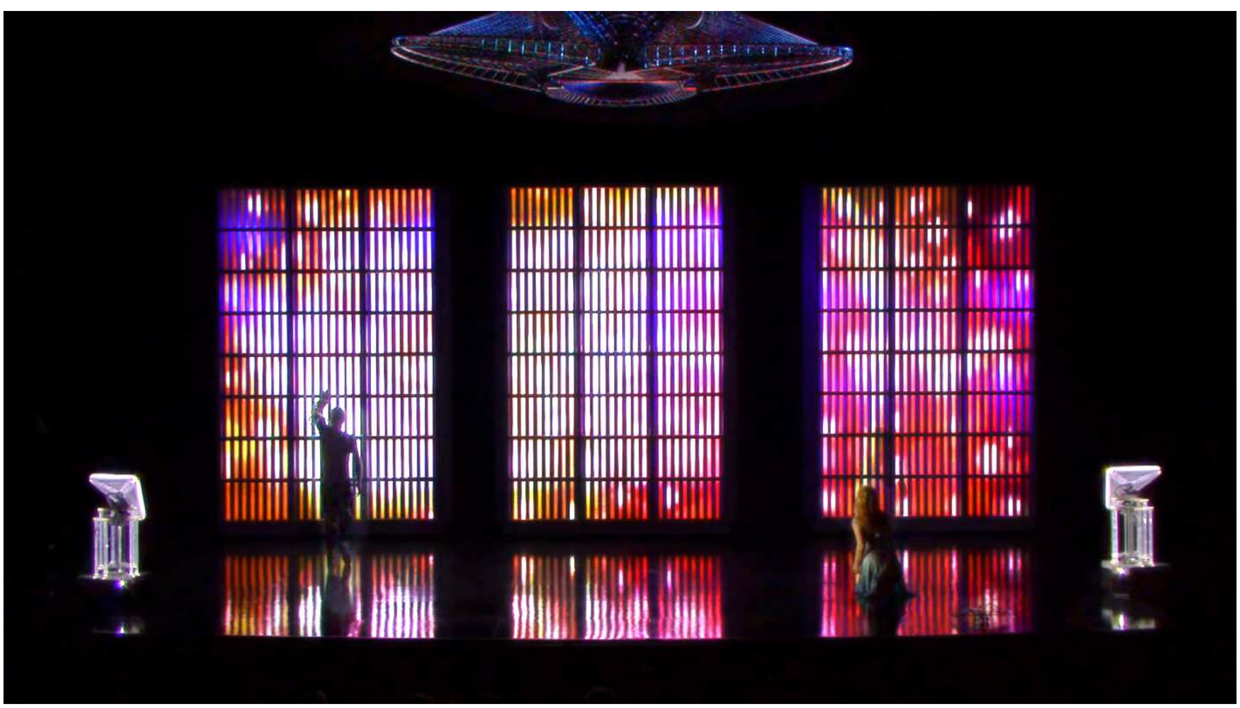
Figure 32: Example Render Designer composite and individual layers

cues could be duplicated and triggered for both systems, associating the mapping with visual cues was more convenient, since there were considerably more visual cues in the show and they were typically executed in sequence. This slaved the Mapping System to the cue list in Render Designer. In *Death and the Powers*, the part for Keyboard 2 notated in the score contains triggers for both playback of electronic sound and cues to the visual system. Render Designer receives the MIDI note events from the keyboard in the orchestra pit and interprets these as GO messages for executing cues, which in turn place the Mapping System in a particular mode.

Each cue stores two types of states for the rendering mode. The first is a set of parameters that define the baselines for the live Disembodied Performance parameters, enumerated in the previous section. The second classification includes formal parameters for the cue's look, including scale, color palette, and notions of *location* and *attention*. Location is roughly Simon's position in the abstract representation and attention represents something akin to gaze.

Additionally, each cue defines a renderer stack that is used to produce imagery from the stored and live performance parameters. A renderer stack behaves like layers in Adobe Photoshop or After Effects. Renderers are added to the stack and composited using opacity and blending modes in order to build up the graphical behavior for each cue. An extensible set of renderers are available in the application. Each procedurally generates dynamic imagery given the formal and performance parameters. The renderers are fairly simple graphical primitives. Complexity and visual

Figure 33: Visuals on display walls

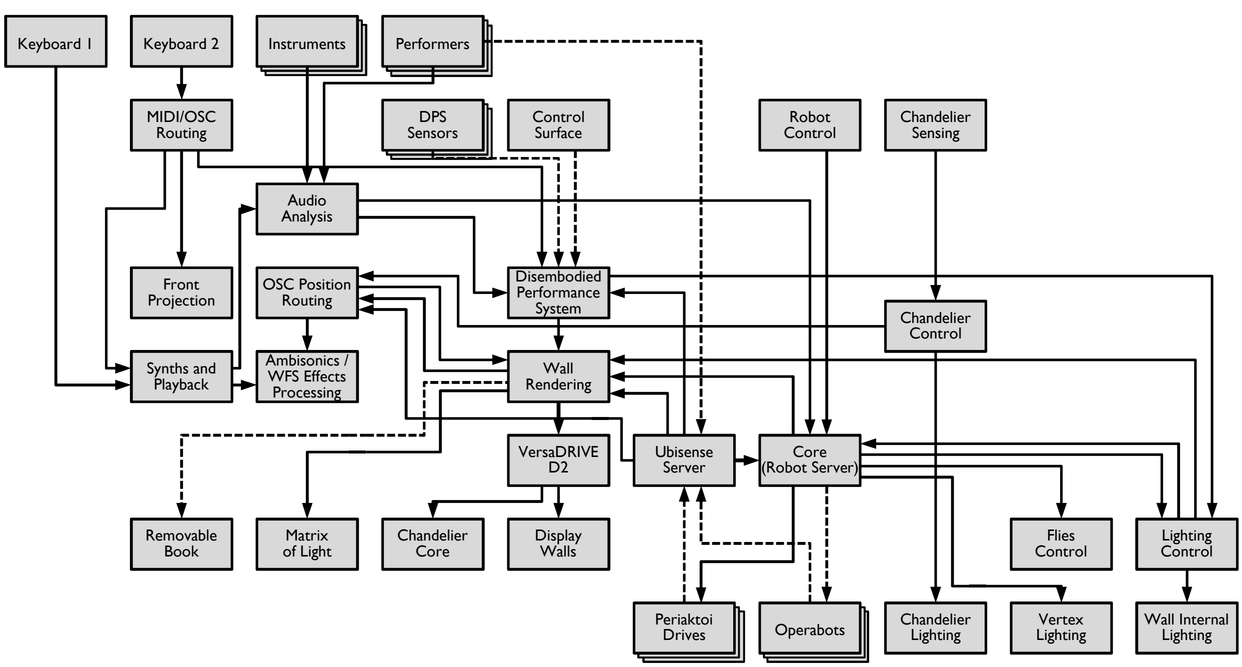


interest is created through compositing multiple renderers. This approach makes it easy to define a consistent visual language, by reusing particular renderers throughout, while still maintaining a large variety of possible looks. Furthermore, since the renderers are relatively simple, the set of performance parameters is sufficient for modifying the procedural graphic elements in appropriate ways for the semantics of each parameter. Some renderers expose additional configuration operations through a dialog window.

Like the Mapping System, Render Designer does not present any intrinsic notion of time. There is no way to schedule an even to occur at an absolute time or for each renderer to assess a relative time. All graphical renderers compute their output explicitly on each frame. A global frame counter is maintained and can be used by renderers to drive an oscillator or evaluate a parametric noise function. At the cue level, transitions between renderer stacks and parameter values may be set to interpolate linearly over a given duration or occur instantaneously and the start of a transition may follow the completion of the previous cue with an optional delay. These follow-on cues can be used to make simple animations, but are primarily for the purpose of orchestrating complex changes to the overall visual look in response to a single trigger. The absence of a well-defined sense of time means that significant visual changes must be the result of change in the live input of performance parameters.

At the center of the application's user interface is a preview window that displays the real-time output of the current renderer stack with the current cue's parameters and is influenced by incoming performance data. Like the

Figure 34: *Death and the Powers* system flow diagram
This diagram indicates how information and control data is shared among the production's systems.



simulation view in Core, this display is essential for offline content creation and for monitoring the result during live performance. Controls for direct manipulation of formal parameters are also available in this view. The cue list and renderer stack controls occupy the right side of the application's user interface. On the right-hand side are general controls for configuring Render Designer, including the display of rendered imagery to a second graphics interface that can be used for output on display technologies. Instances of Render Designer can be run on multiple networked computers to drive high-resolution displays. The display walls for *Powers* were of a sufficiently low resolution that only one system was required. Across the bottom of the window are slider displays of the formal and baseline parameters associated with the current cue, as well as the live performance parameters. Manipulation of the renderer stack or the parameter values produces an immediate effect on the rendered preview and output of the system. Changes are stored by recording them into a new or existing cue. All or some of the groups of cue properties may also be copied from one cue to another to quickly enable duplication and variation of composed looks.

The three systems described, and other show systems for audio reproduction and position sensing, exchanged information over IP-based networks during the production. Interconnected networks minimized unnecessary traffic along busy links, but allowed critical information to be routed to any of the show's computers. A common *lingua franca* of OSC with appropriate conventions for names and values meant that any piece of information could be shared with and understood by any program that required it. The ability to exchange information across systems meant that elements could be designed that responded to each other in complex ways. Sound could be reproduced to appear to originate from a singer or robot. Operabots and singers can pass a wall and the visuals displayed would react to the proximity of the character. An emphatic gesture by James Maddalena would result in an equivalent gesture on the walls and the timbral and spatial transformation of his voice. This approach was essential for creating a believable performance through all of the modalities employed in *Death and the Powers* that felt evocative and present in the moment. Indeed, the walls, operabots, chandelier, electronic sounds, voices, and orchestration do come together as a single performance of an eccentric, powerful character that has transcended his human form.

3.2 Spheres and Splinters

Following the 2010 premiere of *Death and the Powers*, several of the opera's audio systems and the Mapping System were used for a new performance work: *Spheres and Splinters*, also by Tod Machover, commissioned for Aldeburgh Music's Faster than Sound event [1]. In *Spheres and Splinters*, a solo cellist performs the piece using a sensor-augmented bow. The cellist is surrounded by visual display elements in the form of addressable LED strips and sound reinforcement in the performance venue is spatialized. The bow data and cello sound are analyzed and used to control the sound processing, layers of additional sound, sound spatialization, and generative visuals. For the premiere performances of the piece in the UK in 2010, United Visual Artists created the visual component based on data from the Disembodied Performance analysis and Mapping System. For the US premiere in 2011, the visuals were generated by the Render Designer system described above for *Powers*. While *Death and the Powers* used visual and sonic components to replace and translate the presence of the performer through Disembodied Performance, *Spheres and Splinters* reintroduced the human performer into the mise-en-scène. The surrounding environment serves as an extension of the performance. The correlation between the cellist's behavior, the music, and the additional sensory layers is made clear. Due to the modular and connected design of the *Powers* systems and the convenient parameterization of performance expression, the systems were readily adapted to this new performance situation.

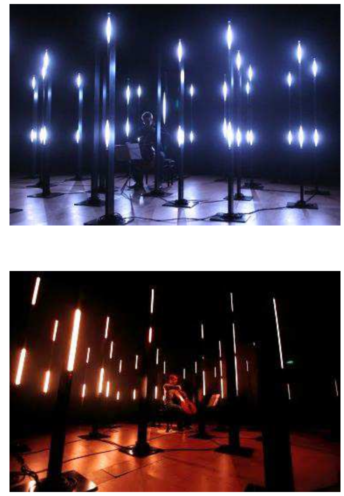


Figure 35: *Spheres and Splinters* Cellist Peter Gregson is surrounded by LED strips that reflect the electronic, spatialized accompaniment generated by his performance. World premiere performance. (Photos courtesy of Jana Chiellino)

Spheres and Splinters was composed by Machover in five stylistically-distinct sections. The design of the sound transformations and generative visuals emerged from the qualities that Machover defined for each section. Beyond these sections, the musical performance provided a finer arc and shape to the piece. Although no planning tools were used to orchestrate the entirety of the experience, this piece would have been well-suited for Media Scores as a means to compose the various facets and perform them by defining sections, modes, and specific events that can be triggered within them. Additionally, modulation of the live performance data by the shape of composed parameters would add subtlety and contour to the experience.

3.3 Extending Sleep No More

Sleep No More is a successful immersive theatrical production created by the innovative British theater group Punchdrunk. Currently running in New York City, the production tells the story of Shakespeare's *Macbeth* intertwined with Alfred Hitchcock's 1940 film adaptation of *Rebecca*, based on the Daphne du Maurier novel of the same name. The New York production is set in the fictionally named McKittrick Hotel, a converted warehouse with a set consisting of over 100 cinematically detailed rooms distributed over the seven story space. Audience members are masked upon entering the experience and instructed to explore but not to speak. The audience proceeds to wander through the space as they please and soon discovers actors who, mostly through dance, enact the story over the course of the three-hour performance.

Punchdrunk approached the Media Lab in the fall of 2011 with the challenge of extending their existing production into an online experience. They provided the caveat that, in their experience, video and imagery of the physical experience falls short of capturing the immersive nature of being there and limits the sense of agency afforded to audience members. This began a research collaboration spanning roughly six months to develop a new model for distributed live performance that was termed Remote Theatrical Immersion. The collaboration was funded by a grant from the NESTA Digital R&D Fund for the Arts with gracious contributions from Cisco Systems Inc., Intel Corporation, Time Warner Inc., and Motorola Inc. and production support by NYC Resistor. As part of the Media Lab team designing and building the augmented *Sleep No More* experience, I was responsible for the control system that mediated the experience by linking the virtual world to online and onsite participants, software interfaces, video mixing and distribution, and the overall visual design of the online experience.

The augmented experience that the team at the Media Lab created connected individual audience members online with partners from the audience in the physical space. Each pair ended up following a narrative thread from the show that was further developed for this experience. The story content presented was constructed such that each pair of participants needed to rely on each other in order to discover more of the narrative and piece together the meaning behind content to which they were exposed. This dependency would motivate the online and onsite participants to both explore and periodically seek out each other to exchange information. The experience was carefully constructed to avoid over gamifying the interaction. In our design of the experience, we avoided goal-oriented interactions, in favor of a more exploratory and narrative-driven framing. Our experiences with interactivity in a fictional context, particularly in computer- and screen-mediated scenarios, is typically one of *gaming*. We tried to minimize such affordances common to games beyond basic MUD (multi-user domain) capabilities. As there is no way to "win" when attending a normal performance of *Sleep No More*, so too for the augmented experience.

One particularly challenging aspect of the project was to develop an experience for the online participant that would feel as immersive and compelling as being at the physical show. The online audience would need to have a similar sense of freedom and agency without feeling confined or tied to a particular point of view. Extensive video or a 3D virtual recreation of the McKittrick Hotel, or the broader setting of Gallow Green, would fall short of the feeling of standing amidst the rich multisensory settings of the physical show. The level of detail required would necessitate extensive content and modeling to reproduce. It would still fail to capture the sense that the virtual world is completely plausible, infinitely detailed, and extends indefinitely in the way that the over 100 rooms in the physical environment does. The bounds of the virtual space would quickly become apparent. Darkness is key to Punchdrunk's scenic design. When one first enters the performance venue, one is plunged into complete darkness before arriving in the show world. From that point on, there is a constant play of light and shadow, seen and unseen, that hides the finitude of the experience.

The solution we arrived at foregoes the explicit visual world in favor of text, in the manner of traditional text adventure games such as Zork [47]. The online participant interacts with the experience through a web interface that displays slightly stylized text and evokes the spaces of the physical show. Upon arriving in the virtual show, the web browser's display quickly goes black leaving only a blinking cursor on the screen. The participant must discover that typing is the method of interaction. After typing a line or two, they begin to discover evocative descriptions of the virtual world in which they are now a part. The text interface is accompanied by an immersive soundscape that is streamed to the user's browser. The text feedback they receive subtly introduces them to the methods of navigation and exploration within the system and provides narrative context as they begin on their way through the virtual Gallow Green. As the user moves about, the binaural soundscape that they hear and subtle changes in text style, formatting, and color provide additional cues to location and are also influenced by the events of the story and their onsite partner's experience. Occasionally, the text interface is supplemented with moments of live video of the scenes or their partner in the physical show, pre-recorded video that provides flashback glimpses of narrative importance, and still imagery that references the show's actual designed spaces in iconic rather than expository ways. Again, the presentation of these visual media is affected by events in the production as a whole, as well as the unique experience of the connected pair. We aimed to preserve a feeling of liveness, with a sense of the physical show and its real-time nature bleeding into the virtual world and the impression of presence of the participants' counterparts.

Figure 36: Screenshots of the online *Sleep No More* interface Online participants interact with the extended experience through a text-based console interface implemented within the web browser. The style and formatting of the text changes to reflect the participant's location in the virtual environment, in addition to the sensed experience of their onsite counterpart. Initially black, the activities of both members of a connected pair trigger occasional glimpses of imagery, pre-recorded, and live video.

Each onsite audience member participating in the experience was fitted with a specially enhanced version of the *Sleep No More* mask that had physiological and environmental sensors with bone conduction transducers to reproduce audio inside the participant's head without obstructing their ears (Figure 37 top) [108]. The sensor data provided a level of communication about the individual's experience to their online counterpart—to influence both narrative-textual elements and the visual quality of the web interface—while the transducers allowed for some communication—triggered based on the circumstance or mediated from the online participant through an actress to satisfy the conceit of the storyline—back to the onsite participant. The computing hardware incorporated into the mask, in the form of a Google Android mobile

Figure 37: *Sleep No More* mask and portal objects
For the extension of *Sleep No More*, masks augmented with computation, sensors, and acoustic feedback in the form of bone conduction transducers were developed to connect onsite audience members with their online counterparts. Portal objects are augmented props (typewriter and mirror shown) within the physical show environment that enabled mediated simplex or duplex communication between online and onsite participants.

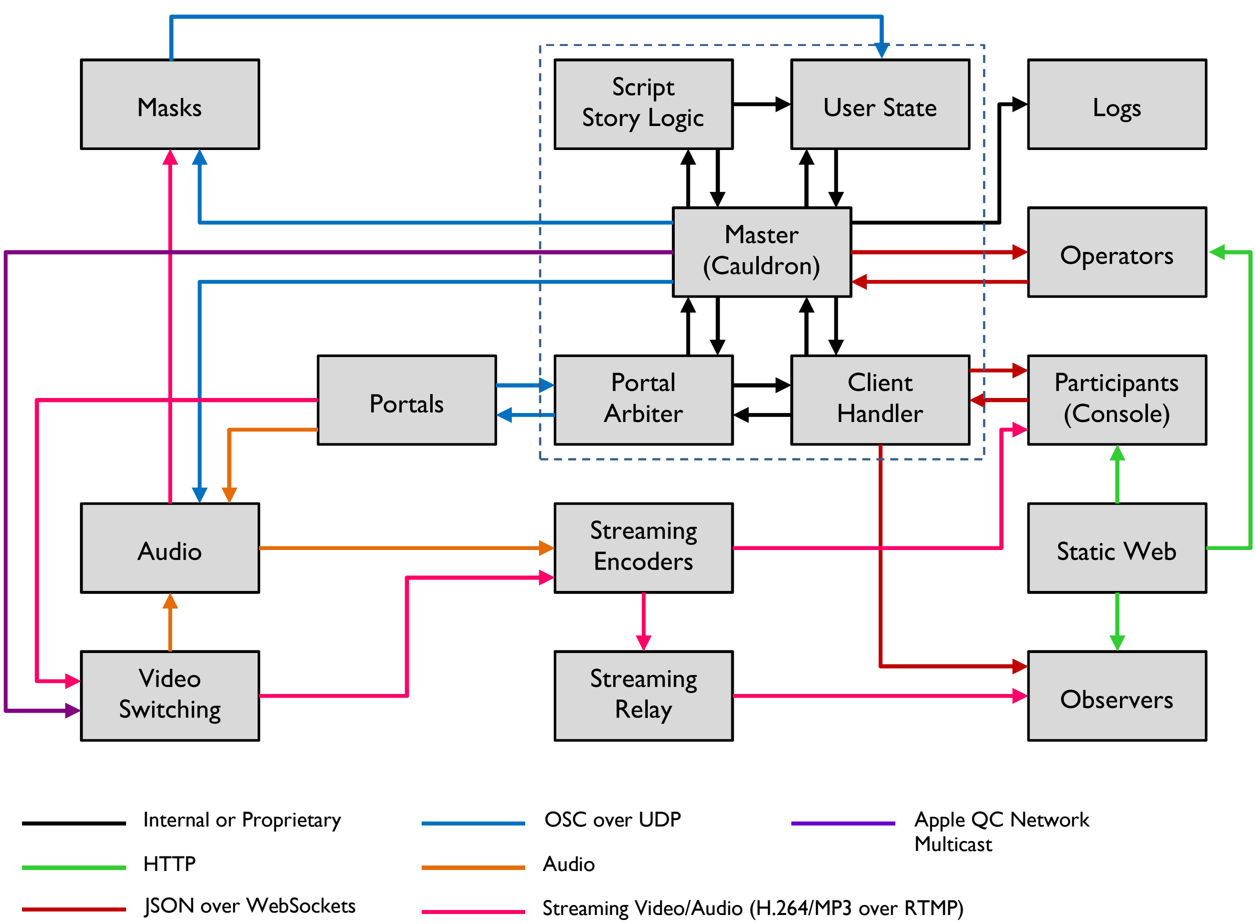
device, relayed sensor data to the central control system and played back pre-recorded and streaming audio to the transducers. An extensive Wi-Fi network was deployed throughout the performance space to maintain wireless communications with the masks. Location tracking throughout the six-story performance space was also handled by the Android device by tracking the signal strength of nearby Bluetooth beacons distributed discretely throughout the space.

Physical *portal objects* in the form of actuated props were installed discreetly throughout the physical set to provide moments of greater connection between the online and onsite participants. When the online audience member occupied the virtual portal space while their onsite partner was in the corresponding physical space, the participants could communicate through the portal object. In addition to Bluetooth positioning, RFID was used at some portal locations to determine the onsite participant's proximity and orientation. Portal objects allowed for higher-bandwidth communication between the online and onsite participants in each pair. Some portals were simply triggered by the presence of both participants in a corresponding space, such as ejecting a book containing important information off of a shelf. Most portals relied on the text-based interaction of the web interface. When such a portal is active, the text that the online participant types is redirected to the physical portal. Examples include a typewriter that would type what the remote user types, an Ouija board that spells out the submitted message, or a mirror that appeared to write the message in dust by the finger of an invisible hand. Additional portals relied on telephony. Triggering a portal could call the telephone number of the online participant that had been provided during registration and simultaneously ring a telephone located in the set in the physical show. When the onsite participant answered the phone, they could speak and be heard by the person online. However, whatever was said by the online person was restated by the aforementioned actress, in order to maintain the established characterization of the pair with respect to each other and to ensure that what was being communicated was consistent with the world of the show, before being heard by the onsite participant on the set telephone.

I developed the central control system that coordinated the interactions among masks, portals, audiovisual playback, time, and story. It was the link among the physical, personal, and the remote. This system, called Cauldron, is at its core a variation on an interactive fiction engine. It uses an XML script format—named JEML for my colleagues Jason Haas and Elena Jessop, who authored much of the story content—to define the virtual world, characters and objects in it, and the range of possible actions that can take place in the virtual and physical spaces. Using this script, Cauldron is essentially a show control system for the coupled online-onsite performance, capable of triggering events in both the real and virtual worlds. Although a Cauldron performance uses a single JEML script, it can manage the independent story state of multiple pairs of participants. Furthermore, encoded in the script, are properties and timings of the physical performance. Pairs could navigate the real and virtual spaces in their own time, but events that were happening in the physical show and potentially experienced by the onsite participant, also had effects in the virtual space. Cauldron was also used to discourage activity near portal objects when an important scene in the physical space was occurring and to avoid different pairs converging on a single portal at the same time.

Cauldron maintained connections with the remote web interface using JavaScript Object Notation (JSON) messages over the full-duplex connection-oriented WebSocket protocol through the Internet and operator interfaces over the private show network. Static portions of the web interface and content were hosted on offsite servers and connected to the Cauldron server once loaded within the browser. The OSC protocol was used over the local network for exchanging information with portals, masks, and media systems. Video playback, switching, and compositing was achieved using several computers running Apple QuartzComposer, each generating video for up to three pairs of participants. The video output was segmented and encoded using VideoLAN VLC and streamed by the Wowza Media Systems streaming server as H.264 via RTMP to an Adobe Flash client

Figure 38: Extended *Sleep No More* system flow diagram



embedded in the client web interface. Cameras distributed throughout the performance space streamed their video over the wired show network and that video could be recalled in QuartzComposer by Cauldron, as well. The digital audio workstation Reaper was automated to interface with telephony systems and provide binaural playback of a variety of different audio, as controlled from the script actions through Cauldron. Audio output of Reaper was streamed to mask devices via Icecast and to web clients through Icecast and Wowza.

In addition to the web interface for remote audience members, browser-based interfaces for control room operators monitored and influenced the interactions online and onsite participants have with Cauldron and the story. The interface allows the *operator*—a narrative orchestrator—to monitor online and onsite users' progress and activity, along with the availability of portals and status of mask connections. In normal operation, Cauldron responds to text commands and input from remote participants automatically with information encoded in the JEML script. Unexpected input is handed off to the operators, who can override automatic system responses and step in for the system such that the online participant never would feel stifled by computer-like responses. Drawing a parallel between the extended experience and *Sleep No More*, the role of the operator is something like that of a stage manager, steward, and an actor in the way that the operator can gently guide both the online and onsite participants, either directly or by issuing commands to Cauldron.

Figure 39: Control room and operator interface
Operators and the rest of the team monitored the extended *Sleep No More* experiences from a control room onsite (top). Operators used a web interface to track the status of online and onsite participants and to override the Cauldron system as needed. Monitors displayed live video feeds from throughout the physical space and the switched streaming output. (bottom)

When we sit down in a proscenium theater to watch a staged performance, the entire audience receives essentially the same show. There may be slight variations in point of view, but they all see the production from the same direction. One set of actors perform a single script lit by lighting controlled from a single console and the audience hears the same sound. In many ways, the one-to-many nature of performance to audience in this context is identical to cinema or television. The normal *Sleep No More* experience is nearly identical, with the exception that audiences can choose their vantage point. Nevertheless, the lighting, sound, and action are all follow a conventional, singular performance paradigm. Traditional theatrical control systems enable such a singular performance to be created in real time. Similarly, online experiences in conjunction with live events follow the same broadcast model, where online users all receive the same content. However, due to the open-ended nature of the virtual world, the control architecture for the augmented *Sleep No More* experience required that each pair of online-onsite participants receive what is essentially an individualized performance. Audio, visual, and story content had to be generated and served to each pair individually—the equivalent of creating a virtual theater for each pair—while also responding to events that occurred in time with the live performance.

A pilot of the resulting experience was held over a one-week period with as many as five pairs of participants in each of five performances of *Sleep No More* during May 2012. Both online and onsite participants were selected to represent a cross-section of the typical *Sleep No More* audience with a broad range of familiarity with the show in particular and Punchdrunk productions in general. Individuals selected for the trial experience were informed that they were participating in a research project exploring connecting audience members in a live performance context and were instructed not to discuss their experiences until the pilot period was concluded.

Online and onsite participants were given a date and performance time to appear for the show. Online participants received a URL by e-mail thirty minutes prior to their staggered entry point into the virtual experience and given the exact time they were to log in to the web interface. Onsite participants were taken aside from the show's bar waiting area at their start time and masked out of view of other audience members before being sent into the performance space.

Much was learned from the development of this project and the pilot performances. The necessity of complex control infrastructure for generating independent audio and video streams was an unexpected challenge, as no existing software or media distribution models reflected the needs of the project. While managing the state of the story logic for multiple pairs is a relatively inexpensive process, the processing requirements of creating independent media content is a barrier to scale in the present system. Another barrier to scale is the use of human operators in the co-creation of the live performance, particularly for the online portion of the experience. While this textual teleperformance was deemed one of the more successful aspects of this experimental project [29], contributions of advanced natural language processing and artificial intelligence techniques within the story engine could make the automation of the experience within Cauldron more robust to user input, minimizing the load on operators and reducing the barrier to scale. Many of the constraints of discretely integrating the extended experience into an existing performance work led to decisions that both strengthened and hindered the overall experience that could be created. In future iterations of Remote Theatrical Immersion, it would be worthwhile to design the entirety of the live experience and the remote experience together so that technologies and interactions wouldn't need to be so obscured or limited within the physical space. This would likely allow a greater depth of communication between online and onsite audiences and strengthen the sense of connection, awareness, and presence over distance.

As with *Spheres and Splinters* and *Death and the Powers*, the resources created and referred to during the design of the experience were disjoint from the implementation of the experience. A Media Score for the augmented performance would have provided a useful reference on such a collaborative project and could be integrated into the control architecture of Cauldron. The JEML script, as implemented, had to be modified in by hand; there was no authoring tool for Cauldron that could provide immediate feedback and rapid modification. The augmented *Sleep No More* experience presents particular challenges for Media Scores that are not immediately addressed in the proposed research described in this document. Namely, the score metaphor would need to be adapted to represent properly the nonlinearity of events necessary to model the virtual component of the experience, as well as the multiplicity of spaces in the physical experience. The incorporation of branching structures and constraints into the scoring metaphor would be needed to represent these types of experiences and define an even more flexible relationship with time. Scoring for such open-ended events will likely be critical, as performance contexts like these become more common. The exploration of nonlinearity in Media Scores is important future work.

3.4 A Toronto Symphony

Composer Tod Machover was named curator for the Toronto Symphony Orchestra's 2013 New Creations Festival, which presents a series of concerts highlighting new and innovated works for orchestra. The first two programs in the festival featured one of Machover's signature works: *Jeux Deux* and *Sparkler*. For the final concert, Machover was commissioned to create a new work for which he elected to undertake the ambitious goal of composing a symphonic piece about Toronto in collaboration with the citizens of Toronto. This is *A Toronto Symphony: Concerto for Composer and City*, which premiered at Roy Thomson Hall with the Toronto Symphony Orchestra under the baton of Maestro Peter Oundjian in March of 2013 [60]. The process of assembling contributions of Torontonians for the piece will be described in detail in Section 5.1. For now, we'll focus our attention on the actual premiere performance of *A Toronto Symphony* and the systems designed to realize it.

The symphony was originally intended to feature only the acoustic orchestra with no technology involved in the actual performance of the piece. Machover envisioned that contributions of sounds of the city, recordings of Torontonians telling stories about their relationship to the city, and the like would be fully translated into the orchestral music. As the variety of sound samples accumulated in the development of the piece, he decided that they should, in fact, be featured in the performance. At many times throughout the piece, the electronic samples and the orchestra complement each other and, in several places, the orchestra quite accurately imitates the recorded sounds, blending in and out from acoustic to electronic and back. With technology now in the mix, Machover also added several of his characteristic electronic textures to augment and richen the acoustic timbres, as well as musical keyboard passages for synthesizers. As a result, the final score features a keyboard to perform the electronic elements of the piece.

One of the concepts Machover had for *A Toronto Symphony* would be that the final orchestral work document its own creation. The structure of the piece roughly follows the chronology of activities that were conducted to engage Torontonians in the collaborative development of the score. It was important that individuals who participated felt like their efforts were represented in the symphony; that they would feel a sense of pride and ownership and community with respect to the work. While this should be true musically, when working with children and other groups to compose music as part of projects in Opera of the Future, we generally also make a point of acknowledging the individual, inviting them onstage for bows and recognition after a piece to which they have contributed is performed. Due to the large number of contributors to *A Toronto Symphony*, Machover and the TSO wanted to acknowledge their contributions during the performance in time with the material they created or modified, so it was decided to accompany the orchestral performance with projected visuals that would display content to this effect.

The visuals, which I created, featured a broad range of content from titles briefly introducing the sections of the piece to photographic imagery taken at various locations around Toronto from which sounds of the city had been recorded. Video of events held as part of the development of the piece were incorporated, along with video and images of individuals who had contributed stories. Titles were displayed that attributed individuals who submitted specific fragments or ideas as they occurred. All of the visual elements were pre-rendered with animations and video treatments in Adobe After Effects. The visuals were constructed as a sequence of approximately 240 short videos ranging from a few seconds to nearly 30 seconds in length. The clips were triggered at specific moments over the course of the 30-minute-long symphony. I annotated a copy of the score where I felt a new clip onset was required and Machover incorporated these trigger events into the electronic keyboard part. (See Section 3.5.1 below for more on how keyboard parts are used to trigger events.) Each clip was longer than it needed to be to account for variations in time, but designed to transition smoothly from one cue to the next.

The imagery also included the graphical sketch of the Media Score for the piece, in addition to live performance versions of two of the web applications developed for the project, discussed in more detail in Chapter 5. The visual component of the performance versions of the applications were implemented in Apple's Quartz Composer and responded

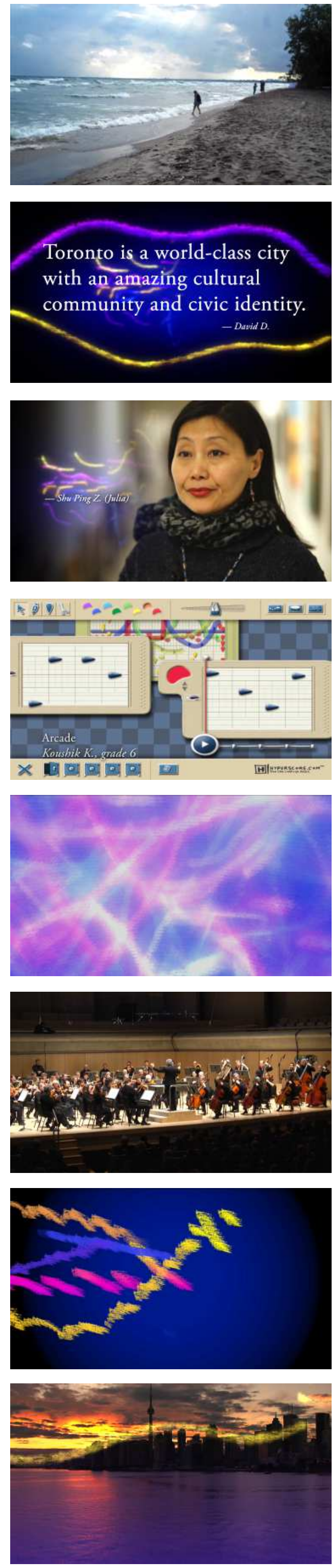


Figure 40: Projected visuals frames from *A Toronto Symphony*

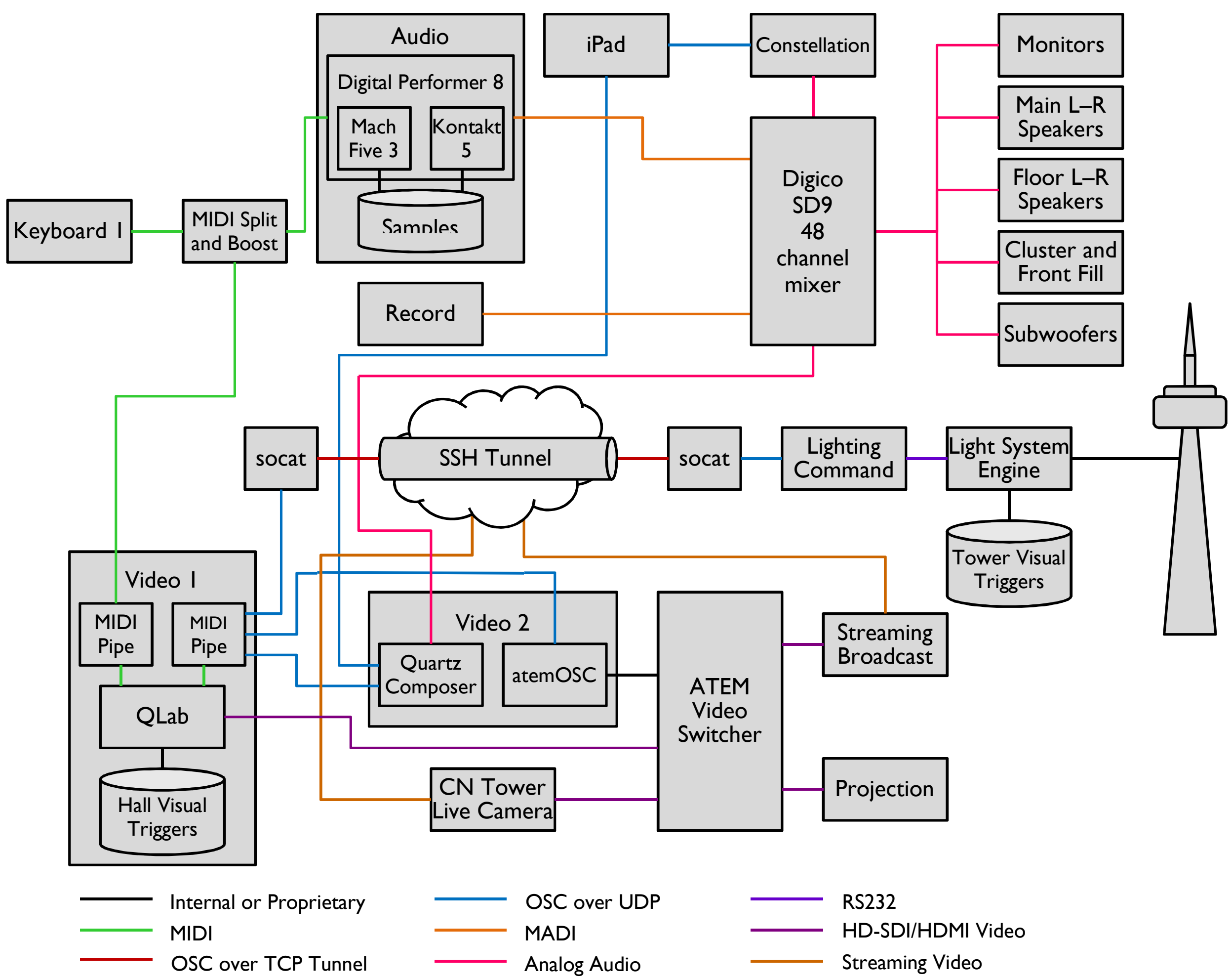
Figure 41: CN Tower illuminated for *A Toronto Symphony*
(Image courtesy of the Toronto Symphony Orchestra)

to an interactive interface running on an Apple iPad played by Machover during the concert via OSC messages over a show network. The corresponding audio was generated separately using the SuperCollider real-time synthesis engine and triggered by the same OSC messages. Each performance application had several different modes that would influence the visual quality of the imagery rendered. The modes were trigged by the keyboardist from the notated part in the score. Live audio analysis of the orchestra was also used in conjunction with data from the iPad interface to create visuals that responded and felt connected with the live music.

In addition to the visuals projected in the hall, we partnered with the city's iconic CN Tower to create a synchronous spectacle for all of Toronto to see. For a symphony celebrating the city and created in part by its residents, it made sense to share the experience with the city as broadly as possible, beyond only those in attendance at the concert in Roy Thomson Hall. I developed a separate visual program that abstractly reflected the music using the Tower's extensive Color Kinetics LED architectural lighting system. Unlike the dynamic visuals for the walls in *Death and the Powers*, a system like Render Designer could not be immediately integrated with the existing lighting infrastructure with the resources allotted. Instead, I used an identical approach to the projected hall visuals, creating a sequence of over 130 separate animations in After Effects that accompany the music. These animations were loaded into the Color Kinetics Light System Engine that controls the Tower's lighting for playback. A live camera feed of the CN Tower was fed back into the concert hall, where it could be incorporated into the projected visuals so that audience members could see glimpses of the Tower. The TSO, in conjunction with the Canadian Broadcasting Company, streamed the concert live online including audio of the performance, the projected visuals, live shot of the CN Tower, and live video of the orchestra. Thus, Torontonians around the city could listen to the live web stream and watch the CN Tower light show. To maintain synchrony between the Tower lighting and the live performance, the visual animations were triggered from the same keyboard in the orchestra. The trigger message was routed to the CN Tower as OSC over an SSH Internet tunnel where the message was translated by custom software into a RS232 serial command to the Light System Engine to recall and playback a particular animation sequence. Due to encoding and buffering latency of the web stream, the Tower lighting cues were delayed by an estimate of the average latency. However, during the moments when video of the CN Tower was shown live in the concert hall, the delay was temporarily removed so that the Tower lighting would be in synchrony with the live performance rather than the stream of the performance.

Since *A Toronto Symphony* originally did not have any technological needs for performance, the technological requirements grew incrementally: first audio samples, then electronic textures, video projection, and finally integration with the CN Tower. My colleague Ben Bloomberg and I set out to design the concert's technological infrastructure using existing hardware and software solutions, expecting that it would simplify the development process. Compared to *Sleep No More* and *Death and the Powers*, *A Toronto Symphony* is technically a very simple production. The core show control system used for the visual components was Figure 53 QLab. Using AppleScript automation in the application, a cue list was generated for all of the visual triggers and each was associated with the appropriate MIDI value for keyboard triggers from the orchestra. However, QLab's MIDI triggering of cues and MIDI output did not sufficiently handle the needs of the show architecture in various ways. SubtleSoft MidiPipe was used to modify the keyboard MIDI data coming into QLab, in order to achieve the desired behavior. QLab was responsible for video playback of the pre-rendered animations for projection in the hall. AppleScript cues in QLab were used to run a shell application to generate OSC messages for control of other systems from QLab. These messages were relayed to Quartz Composer running on a separate computer, in order to set the appropriate modes of the performance versions of the web applications. The OSC triggers were also routed through `socat` and SSH tunnel to the CN Tower

Figure 42: *A Toronto Symphony* system flow diagram



system for cuing lighting animations. A Blackmagic ATEM video switcher was used to switch projection sources from QLab playback, Quartz Composer, and the live feed of the CN Tower, as well as to overlay titles on these sources using a downstream key. The video switcher could be controlled through a customized C++ application via OSC out of QLab as well, so all of the transitions in video source could be added to the cue list and thus performed from the symphony's score.

In the end, the decision to use existing software systems may not have been the most efficient for this project. A lot of time was spent developing workarounds for limitations or assumptions in the existing applications that made it difficult for programs to communicate with each other. AppleScript was used extensively to extend the functionality of QLab and MidiPipe in ways that functioned more closely to the principles described in the next section. Custom Python, C++, JavaScript, and shell scripts were written to translate messages from one program into a format that would be understood by another. In contrast, the control systems that I described for the other projects above use protocols and message presentation conventions that make it easier to interface connected software systems.

3.5 Practice

These example projects we've looked at so far demonstrate the complexity of integrating new technologies into live performance contexts. The systems created for *Death and the Powers* did not aid in the design creation during the pre-production phase, but did facilitate implementation, rehearsal, and performance. Each was designed such that the content could be edited in real-time, with no distinction between editing and playback modes, and with a local visualization so that content could be created offline or live during rehearsal conditions. This made the technological components as agile as possible so changes could be made very easily during the rehearsal process.

Since the systems were designed for content creation, editing, and performance during the production and performance phases, the overall design of *Powers* took shape in pre-production somewhat independently of the software environments. Certain needs or design motifs that were envisioned in advance of the rehearsal and content creation periods did inform the software architecture and capabilities, but the systems themselves were intended to be general enough to be able to flexibly create and adapt to changes in design at all phases of the pre-production process. The need to unify the design process with the implementation of the performance systems became apparent, motivating what would ultimately become Media Scores. Although this need could not be fully addressed with the time and resources available for the completion of development on

Powers, some design approaches developed that would inform the work at hand. These are addressed in Section 4.1.3.

In general, the systems created seek to provide as much flexibility as possible during both the creation and performance of complex works incorporating performance-driven technologies. They function in a continuum of interactivity and variability that is essential to both live performance itself and the process of creating it. As such, the systems serve as a common set of tools for both authoring and performing a work.

3.5.1 Cues, Modes, and Triggers

Based on practices developed for the creation of Hyperinstruments, in the Opera of the Future group, we generally consider interaction with performance systems in terms of two types of cues: *modes* and *triggers* [63]. This organizing principle establishes a hierarchy of temporal information in the composition of artworks that rely on technological mediation during performance. A piece that exists in time—a performance or installation—can be organized into a series of states—the modes—and events within or between states—the triggers. All the while, continuous properties of performance and expression add the essential continuity and detail to the experience.

Modes are discrete changes of continuous controls or changes in configuration of the system. They create a higher-level function of structure by defining the rules of the performance system at a particular time. Cues in the Mapping System and most of the cues in Render Designer for *Death and the Powers* are modes. In the Mapping System, a cue defines a mapping which is the way in which the input performance data is interpreted while that cue is active. Cues in Render Designer set the procedure by which graphical material is procedurally generated in response to input. Modes may also take other forms, such as determining how continuous performance parameters are interpreted. Modes may also define how triggers are interpreted—the meaning of a particular trigger is related to the mode during which it occurs, thereby multiplexing trigger values or providing multiple senses for a trigger—or be used to allow triggers or performance data to have any effect.

Triggers are discrete events that initiate an action. Transitions among triggers, like crossfades of cues in a lighting board or sources on a video switcher, may have a continuous change associated with their onset. The trigger itself is momentary, unlike a mode which persists until the mode is changed. However, the trigger can, by initiating a mode, have effects that last longer than the instant of the trigger. This action may be a change of mode, the playback of media, or the deliberate setting into motion of a sequence (of varying degree of flexibility). The triggers may be more finely crafted in pre-production than what could be created or manipulated at performance-time, such as the creation of pre-recorded media for playback or elaborate choreographies of robots. Audio triggers in Machover's compositions generally take two forms. The first are one-shot sounds that are not in themselves single notes as would be played by an instrument, but brief collages, generally with a well-defined attack. Given their short duration and punctuating role, these types of triggers generally play in their entirety. In Render Designer, the playback of pre-rendered video or animation can be triggered in the same manner. Other, longer pre-recorded material is used as what Machover terms *textures*. In his work, these form an electronic backing around which acoustic orchestration can weave and are generally held for as long as needed and can be stopped at any time by releasing the trigger.

An analogy for the relationship of modes, triggers, and continuous performance data can be found in instrumental performance. The modes for a piece of music might include the orchestration, tuning, and key in which the piece or sections of the piece are written. Triggers would be the individual note events that actually produce sound. The continuous performance data includes the articulations of each note event—attack and decay, vibrato—that give it its character.

Both types of cues afford a robustness to time that is essential in live performance. An entire piece does not play back to a click track or timecode, fixed in time. As in traditional practice, cues are executed at the appropriate time taking into account any variations in performance without needlessly restricting the ephemeral quality of liveness and burdening human performance with keeping up to an arbitrary, fixed sense of time. In music-driven performance contexts, such as concerts and operas, the conductor paces the orchestra to conform the musical timing with other stage business and for expressive effect. Thus, it makes sense that the cuing of events within the show be intimately connected with the music and its timing. In Machover's projects, such as *Death and the Powers* and *A Toronto Symphony*, the triggers and modes are written into the orchestra score and given to a special part for MIDI keyboard. *Powers* features two electronic keyboards, one playing a musical part and one playing the audio samples, textures, and visual cues. For *A Toronto Symphony*, only one keyboard was used and mode changes determined whether the keyboard's notes were in musical passages or trigger events. The MIDI messages are parsed for modes and triggers and sent to the appropriate control systems directly or, in some cases, first converted to OSC messages. It is also possible that continuous performance data when analyzed can also invoke changes in modes or fire triggers. This approach was used in *Spheres and Splinters* based on audio and bowing analysis. Additional higher-level structural mode changes were triggered by an operator and not the performer.

Notating triggers and modes within the orchestral score has several advantages. Instrumentalists performing the trigger parts allow events to be precisely timed with the music—with as much precision as any instrumentalist in the orchestra. Additionally, the score becomes a means for documenting not just the musical aspects of the production but any number of technological events. In *Death and the Powers*, the visuals representing The System are part of the orchestration, not just a layer that is added by the production. I worked with Machover to define where these visual cues occur as he was notating his own triggers for electronic sounds. This practice expands the modalities associated with particular source texts and provides the composer with additional means to sculpt the form of the final production. We'll see in the next chapter how Media Scores takes this idea further.

By notating triggers as a keyboard part, each is given a distinct identity and physical means of firing it. While cue lists in lighting consoles and other theatrical automation systems are generally numbered and can be accessed by number, this capability is not used during performance. Instead, the cues are arranged in sequence and executed in sequence, one after the other, by pressing a single GO button or firing an external trigger control signal. In all the systems described above, cues are arranged in sequenced lists and the interfaces and remote control protocols provide a mechanism for executing cues in that sequence. The notated part in the score often assigns sequential triggers to chromatically increasing notes. However, any of those notes *can* be pressed in any order. This technique is capable of variation in timing and in order. If a cue is accidentally missed by the performer, the technological systems will not reflect the missed cue, but will not fall behind in sequence, either. Thus, in these systems, cues can be randomly accessed. As mentioned previously, it is in this way that cues in the Mapping System can be reused by cues in Render Designer. This versatility is helpful during the rehearsal process. Without needing to reset systems, the conductor can just cue the orchestra to return to a particular measure or may jump around as needed, and the instrumentalist follows, playing the appropriate notes and triggering the appropriate state and behavior of show systems. This random access behavior, in other contexts, means that non-linear structures for the piece can be explored. Variations in order and improvisation are just as easy as performing the technological components of the system in sequence.

3.5.2 Communications Protocols

We've seen so far that complex performance works necessitate systems that exchange information with each other. At times, this information is simply a cue message to synchronize output in different media with music, time, or performance. At other times, it's sharing real-time performance parameters to allow performers and audiences to expressively impact any of a show's modalities. Whether it is connecting the 43 computers that it takes to run *Death and the Powers* or linking sensors in a *Sleep No More* mask to a web browser halfway around the world, the way in which systems communicate is of critical concern. As these technologies begin to combine the theatrical disciplines, it is not only the communication among the humans designing and operating a production that matters, but the communication among the systems that run the performance.

Well-established theatrical control protocols are an important reference point to consider how systems should communicate. In theater, lighting is most often controlled using a RS485-based serial protocol, DMX512 to connect lighting consoles with dimmers and intelligent fixtures. The protocol is relatively simple in that it specifies up to 512 8-bit values per frame on a single link. Devices usually pass through the control signal allowing multiple devices to be chained. Splitters may also be used to create a star topology with DMX. Although it has been co-opted for additional purposes, such as the Remote Device Management (RDM) extension for bi-directional communication of control information, DMX at its core is a unidirectional broadcast of values for continuous parameters or channels. Another common theatrical control protocol is MIDI Show Control (MSC). Curiously, MSC is built on top of MIDI, as RDM is built on DMX, adding discrete triggering operations on top of a protocol that otherwise emphasizes continuous parameters. Also a serial protocol, MIDI links are often daisy chained or split, as well. The topology of MIDI and DMX networks can become cumbersome in large installations. Separate DMX universes are required for every 512 channels used and many complex devices can easily use on the order of 40 channels each, thereby necessitating a fair amount of cable and distribution equipment. Many theatrical devices rely on MSC for triggering and DMX for continuous control, but require specialized hardware in order to interface with software-only systems running on typical computers.

Internet Protocol-based communications—generally over a wired Ethernet physical layer or wirelessly with 802.11 networks—offer greater flexibility. Networking hardware for IP networking can be much less expensive than specialized devices and can be received and produced by computers without additional hardware interfaces. These networks are also capable of handling large numbers of devices over longer distances and multiple types of messages at a time, from control messages to high-fidelity multimedia. Local show networks can also be readily connected to the Internet, enabling distributed performances and remote interactions. Although dedicated show networks offer more ideal performance and security, for some applications it may be suitable to use existing infrastructure within a performance venue.

To take advantage of IP-based networking, several transport container formats have been developed for traditional theatrical control protocols and other applications. For example, MIDI-over-IP capabilities are not uncommon and RTP-MIDI has been integrated into CoreMIDI functionality in Apple OS X and iOS [5]. The Art-Net protocol developed by Artistic License Engineering provides similar functionality by encapsulating DMX512 data in User Datagram Protocol (UDP) packets for IP network transport [7]. In 2006, the Entertainment Services and Technology Association first released the Architecture for Control Networks (ACN) protocol as a forward-thinking replacement for DMX512 [33]. For use predominantly with IP-based networks, ACN builds on the functionality of DMX and RDM to manage complex networks of a variety of theatrical devices. Device functionality in ACN can be described using the Device Definition Language to controllers and devices can report their functionality and provide other mechanisms of reflection in addition to control. This makes for flexible, reconfigurable, and fault-tolerant control networks. Adoption of ACN has been slow, though most theatrical control systems now support some parts of the specification.

As noted earlier, the ACN protocol was a source of inspiration for the control architecture used in *Death and the Powers*. However, due to the complexity of the protocol specification and lack of robust implementations at the time, I ultimately chose to implement device communications using the Open Sound Control (OSC) protocol, which is designed for IP-based networks and typically relies on the lower latency of UDP transport. Protocols such as OSC provide a common presentation format for exchanging both continuous and discrete control information among typical computing hardware. It is straightforward to incorporate such IP-based control into new software systems with ample libraries available for a variety of programming environments. Additionally, hardware support for OSC is increasingly common in devices such as audio mixers and commercially available software-based audiovisual applications. The flexibility of IP networks, routing, and OSC allows new devices or components to be added to the network at any point as needed. As an example, mobile devices and tablet computers can be used as wireless display or remote control interfaces in show networks. During the development of *Powers*, with the OSC control scheme already in place for our custom systems, we found that we could very easily use Apple iPads running the Hexler TouchOSC application as remote interfaces for the production's systems because they shared a common protocol and manner of information presentation.

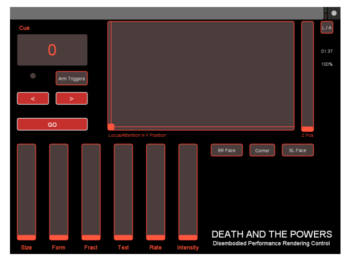


Figure 43: Remote interface for Disembodied Performance

TouchOSC interface used during performances of *Death and the Powers* to monitor and control the Disembodied Performance of Simon Powers and the Render Designer application

Although flexible and increasingly ubiquitous, IP networks are not without their challenges in performance contexts. Perhaps the most critical issue for such control networks is to minimize latency as much as possible. Dedicated, private show networks with only essential hosts connected are important for minimizing network latency. In *Death and the Powers*, five networks were used to localize traffic only on links that required it. The robot systems, and positioning systems generated large amounts of packets, so they were given their own networks. Networks are bridged at multihomed hosts or routers, so that the subset of traffic that requires it can be routed to other show systems. The topology of dedicated wired show networks would not likely contain cycles, so concerns about packet delay variation and order are negligible for UDP connections are, and therefore generally reliable. For higher-level protocols that require connected transport or for connections to hosts through the Internet that are not tolerant to packet loss, TCP is used. However, these are concerns for wireless networks and when routing show data over the Internet. Only networks that absolutely require Internet access, if any, are routable to an Internet gateway and physically connected to an Internet link. Wireless links are also subject to radio interference and complications due to overcrowding and security. *Death and the Powers* relies on several wireless networks. With audiences entering the venue carrying devices trying to pair with available networks, we had concerns about the additional traffic. Wireless show networks do not broadcast their SSID and a dummy public, open network was created, so that devices would associate with it and not continually search for available networks.

A common show control specification for OSC emerged during the development of *Death and the Powers*: OSC Show Control (OSCSC). I have since used OSCSC as the common language for the exchange of show control information and performance data wherever possible in projects such as *Sleep No More* and *A Toronto Symphony*. Conforming to this data-presentation convention allowed existing systems to be connected with little effort in a variety of configurations as dictated by each project's needs. Inspired by ACN and MSC, it can easily be transformed into other protocols for interoperability with non-OSC systems. In brief, OSCSC specifies naming conventions for OSC addresses and the types of arguments associated with each address. It handles the needs of cuing, triggering, transport control, and specifies messages for reflection in the system and device feedback for monitoring and control. For the exchange continuous performance and control parameters, a hierarchy is specified. A device may have zero or more addressable subdevices identified by name. Values are associated with named channels or axes of the device or subdevice. In the following chapter, we'll see how the implementation of Media Scores relies on OSCSC conventions to integrate with performance systems.

3.6 Design Principles for Performance Systems

The systems described in this chapter augment conventional production for live theater with novel technologies for show control as well as authoring and generating dynamic content in live performance scenarios. The needs of each system grew out of the particular challenges associated with each project. However, there are commonalities among the system architectures and, all together, these systems demonstrate essential features for the fluid integration of robust and responsive technologies into live performance practice. I conclude this chapter with a summary of design principles extracted from the systems.

Continuous Parameters

In addition to setting discrete states of operation by executing modes and triggers, continuous parameters let control systems produce more responsive and expressive control and feedback. Additionally, constraining the values of these parameters to a known range with a started interpretation means that values can be routed freely among the systems. Any system can respond to any value as desired. An application such as the Mapping System takes advantage of this by allowing any output to be connected to any input with any set of operations in between. The output could then be routed, as needed to robot systems, audio effects, and Render Designer without need for conversion and scaling of operational ranges.

Interconnected Systems

In traditional theatrical practice, the control architecture is centralized with a stage manager who typically audibly calls the cues for the show, usually via closed circuit intercom. Operators then respond by executing the cue on their respective system: lighting board or automation system operators press a GO button, audio engineers press a button for sound playback or bring up a fader for a microphone, rail crew manually run a lineset to fly a batten in or out, actors step onto the stage. These cues are discrete triggers and skilled operators and the design of the systems ensure that the cues are executed with as little latency as possible.

In complex scenarios where multiple systems generate output for the performance, it is advantageous to have the systems communicate with each other. Interconnected systems are capable of achieving the same effects. Technologies such as MSC and other serial protocols have made this sort of triggering possible for some time. As the systems become more complex and the particulars of the routing of triggers, as well as the introduction of continuous parameter data is introduced to the functioning of the show, more flexible types of interconnection are required. Gestures and events can easily be coordinated across all of the show's systems. Conductor control models and star topologies are not absolutely required for all functions. Hybrid systems allow individual components to communicate information among each other in a heterarchical manner with each system routing data to a subset of systems or only listening to information that is relevant to it.

All of the productions described in this chapter utilized IP-based networks to link systems, as described in the previous section. Having a common protocol and format for information presentation using that protocol enables complex show systems to be assembled with ease from existing software in a modular fashion. Systems are built up from distributed components as needed and easily connected to a common link medium where possible. Messages from the common network protocol may be translated in software to other hardware protocols as necessary for legacy systems. Using IP-networking also allows systems to communicate beyond the show networks, sending information over the Internet to enable remote control and distributed performance.

Modeless Operation

Performance systems should not have distinct modes for editing and live output. The edit-audition cycle should become instantaneous. Render Designer and the Mapping System are both examples of modeless performance systems. As changes are made to parameters, visual qualities, and data mappings, the output is continuously updated and can be previewed in local views within the application or, when connected to the show infrastructure, viewed live onstage. Since edits are reflected instantaneously, modifications can be made quickly without the need for separate processes such as editing code or rendering content. Operators and members of the creative team can experiment with ideas quickly with immediate feedback and the system state can take direction from production personnel. If a mapping or look needs to change, this change may only take a few seconds or minutes and the scene can then be rerun, much the way an actor would receive direction and notes and be able to incorporate the changes.

The playback of time-based or physical media may still require time to reset and preview. For example, a modification can be made to the parameters of a cue in Render Designer while the cue is live, but if the cue contains a video asset, it would need to be re-fired to preview the changes in conjunction with the video. For the short video triggers used in *Death and the Powers*, this did not effectively reduce the efficiency of modeless editing. Audio textures and triggers behaved similarly. A slightly slower edit-audition process was required in Core due to the physical nature of the robots. To re-run a sequence or cue in a sequence that involved operabot or wall translation, the robots needed to be reset to their position at the top of the cue before running the cue again, just as actors would need to reset before running a scene again and hitting their marks. Otherwise, the robot's offset in space would produce movements that would not reflect the final form starting from the initial positions. Robots might also enter into dangerous situations if moving from an unexpected location, such as a movement that would extend off the edge of the stage or collide with another object or person. Like all physical objects, a certain minimum amount of time would be required to reset the operabots and walls and the reset trajectory would need to avoid collisions with other elements being reset for the cue.

Local Feedback

Systems should visualize their state to designers and operators in real-time within the application. Software applications for lighting and show control typically provide support for transmitting data to a separate visualizer program for previewing content. In the systems described above, this preview is integrated into the application and is a central part of the user's workflow within the application.

Core and Render Designer provide views or simulations of the robot behaviors and content within the application interface. The Mapping System also features dynamic plots of input and output data. Though not a visualization, this diagrammatic representation serves the same purposes when input data is available. Thus, while looking at the interface, changes in behavior are reflected on the screen without needing to view the stage. This preview is also useful during rehearsals and performances to verify that the system is producing the expected output, particularly if something appears anomalous onstage. Local feedback also allows these systems to be useable offline, when not connected to the necessary infrastructure. Robot choreographies could be edited and previewed with just the software application, not requiring physical robots. This allows content to be modified or developed in advance of rehearsals or while other activities are happening onstage. Coupled with a "blind" mode of operation, local feedback could also support editing of content or other cues without running them live onstage. Updates can be made in blind without affecting the final output of the system to display or automation infrastructure.

Cue-based

As described in Section 3.5.1, cue-based systems enable the performance to be shaped by providing the ability for composed modes to respond to how systems interact and to triggers for particular discrete events. As implemented in the systems above, cues can be readily accessed in any order and enable technological systems to be robust to variations in time and operator error. Cues also form a scaffolding around which behaviors across multiple modalities can be synchronized by a single trigger. Without cues or modes, systems would only be able to perform actions in one way. Particularly when incorporating live data from performer or audiences, the cuing of modes allows the interaction to be modulated in meaningful ways and avoid the monotony of a single method of interaction.

4 CREATING MEDIA SCORES

> *I am not an abstract painter. I am not interested in the relationship between form and color. The only thing I care about is the expression of man's basic emotions: tragedy, ecstasy, destiny.*
>
> —Mark Rothko

Through the various projects described, a need was uncovered for tools that are well-suited to the creation of multidisciplinary works. How can an individual artist or a creative team reason about the complex interrelationships of story, form, and expression? How can these abstract and essential concepts be represented and communicated? Once represented, how do they become a final work of art? These are the questions that Media Scores and the process of working with them strive to address.

The discussion of live performance and example productions in Chapter 3 suggest a need for an adaptation of practice that is better suited to the integration of new technologies in live performance. A critical concern is one of communication among practitioners during the development and realization of a complex work of art. Additionally, creative decisions in such contexts—either made by members of a team, the team as a whole, or an individual creator—have repercussions for other aspects of a work's design. With the aid of technologies such as interactive and interconnected systems, creative choices become increasingly interrelated. However, in service of a common expressive intent, these changes should be able to be managed or at least represented at a common point. Applying the design goals outlined in the previous chapter to a system for composing works of art opens the door for a number of unique capabilities that redefine the role of composition in the creation of Gesamtkunstwerk.

The role of the Media Score can be defined with respect to traditional scripts and scores by considering the production process of

Gesamtkunstwerk within the frame of reified inference that was introduced in Section 2.1. When applied to the mapping problem, reified inference posits that the input representation is as much a result of applying a particular model of a generating process as the form into which the input is mapped. Thus, artifacts are both the input and output of a process that can be modeled.

In the production of an existing artistic work, we can consider the source text—the script or score—as the input representation and the final form of the production as the output of a mapping process that is the realization of the work through interpretation, design, and production. If we apply the process of reified inference to this workflow, then we can see how a typical production process resembles the flow of information in reified inference. The artist encodes his intent in the source score. The score is then interpreted by a production team. From the reduced form that is the score, they reconstitute a vision of the final work that remains faithful to the artist's intent, regardless of how explicitly it is encoded in the source score. The result is an instance of the artistic work; one possible output of the score or, more precisely, one possible output of the artistic intent. In the reified inference approach, the score is an output like any other that captures the artist's meaning. The artist is the generating process. The production team models this meaning in order to produce another possible output in the form of the realized work. Many possible interpretations of the model may be made. The modeling process here is one of resolving or inferring the implicit information in the source score as a model of artistic intent.

The Media Score situates itself in this process as an encoding of the parameters to realize the instance of a production from the model of the artist's intent. As with the final production or range of possible productions, the Media Score is an instance of the artwork itself. However, in the creation of an original work, the Media Score can be created directly by the artist. The artist, in effect, is creating directly the model or the system of his intentions from which any output can be produced. Doing so eliminates the need to recreate the model through inference from a symbolically notated score.

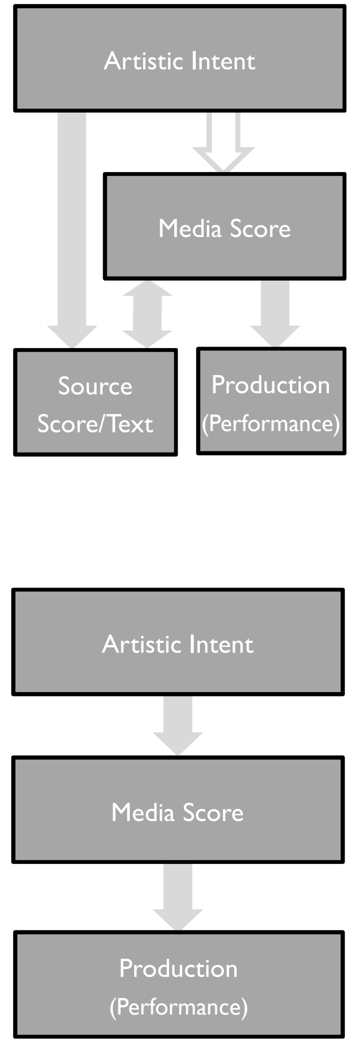


Figure 44: Production process as reified inference
The Media Score created by a production team from a source text or score models the artist's intent. (top)
The artist can compose the Media Score directly. (bottom)

4.1 The Beginnings of an Idea

My formulation of Media Scores has its roots in several early projects and concepts that focused on the representation of expression and narrative structure over time. In these experiments, I began exploring the application of the metaphor of musical score to the creation of composed works in multiple media. The score representation seems appropriate for the creation of time-based works both for the functional role of organizing multiple parameters or events and as an entry point for conceptualizing the creation process. Personal Opera, the first early project I will mention below, sought to create tools for novice users to facilitate this process. As such, a familiar metaphor would ease the barrier of entry to a complex process, even in the case when the users are not skilled in the medium from which the metaphor is borrowed. As we saw in the discussion of scripts and scores in Section 2.9, there has been considerable innovation and creative liberty taken by expert practitioners. Similarly, the metaphor of score serves here only to loosely define the function of the organizing artifact and not the particulars of notation or theoretical underpinnings.

4.1.1 Personal Opera

In 2011 and 2012 Tod Machover and I collaborated with the Royal Opera House, Covent Garden in London to offer Personal Opera workshops to senior citizens in London. During these workshops, participants would develop short, two- to three-minute long operas about a personal story or impactful moment in their life. Each personal opera began with participants creating a text that they then set to music. They also created visuals to accompany their performance of their own pieces. For these workshops, material was created mostly by analogue means—though Hyperscore was used in the 2012 workshop to aid in music-creation—with the assistance of professionals in each modality.

The idea of Personal Opera dates back several years prior. The workshops with the Royal Opera House were pilots for a number of imagined activities. Personal Opera touches on many of the same themes of individuality, preservation of legacy, and perpetuation of self as were explored in the story of *Death and the Powers*. In general, Personal Opera aims to give senior citizens the ability to express themselves through a music-driven medium in a way that can be shared or created through intergenerational collaboration. It is a way of communicating the important aspects of one's life through an accessible medium and therefore expressively documenting and preserving one's legacy. As such, for Personal Opera, we envisioned a set of easy-to-use software tools that can allow users to assemble the text, music, and visual components in expressive and meaningful ways. Ideas included building personal operas from users' own music playlists and photos on their mobile devices, to desktop applications, to interactive installations where visitors could quickly create their own piece and view those of others. Connections among the collection of operas by theme or temporal and historical events were considered. The final form could be distributable online, exhibited in the installation space, or embedded into an heirloom object that could be given as a gift or displayed in the home. It was important that personal operas felt richer and more individual than simply a slideshow of images set to familiar music. They needed to convey the essence of the person telling their story as much as the events and images being recounted.

As part of this brainstorming, I began thinking about what such a software interface might look like. In doing so, I created several conceptual mockups of a Personal Opera system, through which themes emerged. One was to base the interface on a metaphor of a musical score, a form of notation for arranging elements in time to create the operatic finished product. The score-like interface, inspired in part by Mark Podlaseck's Glass Engine as a lifetime compendium of the works of Philip Glass [74], would allow users to find connections among expressive, thematic, and temporal events and sequence storytelling materials accordingly. For the installation version, I considered a model where users would assemble their media and materials that are then sequenced and blended by conducting their opera through expressive gesture and, at the same time, essentially performing a visually rich story. One of the conceptual drawings considered a visual notation that was not merely symbolic, but also expressive of the story it represents. Although Media Scores targets a more professional user, the concepts that grew out of work on Personal Opera remain key aspects of the system.

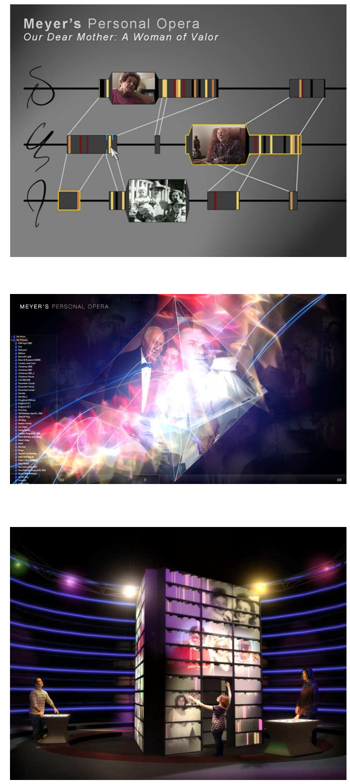


Figure 45: Personal Opera score and installation concepts
A musical-score-inspired notation for composing stories and media, revealing connections (top)
A vision of composing personal operas where the score becomes a dynamic, expressive representation of the story (middle)
Installation version where gesture is used to compose a piece from assembled assets (bottom)

4.1.2 Story Plot

The concept drawings for Personal Opera imagined possible representations of events in time, in a small set of audiovisual media, and the connections across those events. In 2009, while surveying computational storytelling, including the work of Propp [84] and Siskind [96], I wrote a small software application called Story Plot that looked at more continuous changes that occurred in narrative events, including emotional and qualitative events, particularly with respect to characters in the story.

The Story Plot application displays the overall structure of a narrative plot segmented into named sections and time proceeding along the horizontal axis. Superimposed over these section boundaries are continuously-varying curves, one color for each character in the story. The curves are continuous for the duration or section that the character is present in the scene, with discontinuities and when the character is absent. The curves for each character may be overlaid with each other within an absolute range or separated vertically into tracks for each character. Different properties or functions could be presented, one at a time. Switching from one parameter to another would animate the relative values of the curve for each character. While multiple properties for each character could not be observed at once, the animation made the relationship between one parameter and another clear when transitioning between them.

Story Plot is simply a visualization tool for character-specific parameters. No automated analysis was used to generate the data, though that would likely be possible using computational storytelling techniques. Rather, for my own experiments with the application, I annotated the data manually based on personal judgment or situational fact. The types of parameters I looked at typically related to the character's cognitive or affective state, as

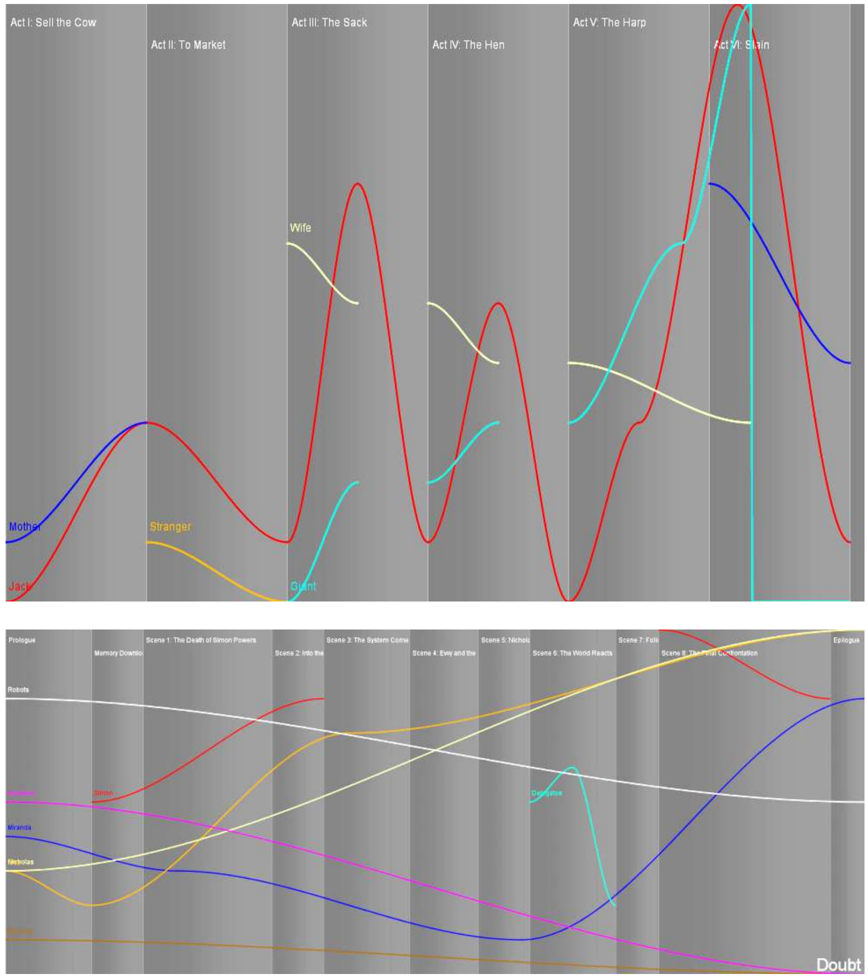


Figure 46: Story Plot
Jack and the Beanstalk illustrating characters' *fear* as a parameter (top)
Powers characters' *doubt* as a parameter (bottom)

well as the circumstance of the character. For example, in a plot I created based on the fairy tale of "Jack and the Beanstalk," one expressive parameter depicted is *fear* while a situational parameter is *vertical height*. At this time, I was also frequently in discussions about the story for *Death and the Powers* based on the action and text of the libretto. I used Story Plot as a tool to analyze and notate more complex cognitive states such as "doubt that Simon is in The System" and "excitement/agitation" while considering aspects of staging and production design in these discussions.

4.1.3 Visual Score for Death and the Powers

As production development on *Death and the Powers* progressed, a number of details about the design began to take shape. With an overall structure in place, a palette of set elements had been decided upon: the three periaktoi with display walls, operabots, and a chandelier. While the constituent elements were there, the moment-to-moment look and feel of the stage experience needed to be crafted. The opera was given a workshop presentation in 2009 for which the entire score was completed for keyboard, electronics, and singers. This provided a timeline of text and events to which the time-varying aspects of the production could be designed. I began considering the visual language to represent Simon in The System, including color palettes, textures, and other visual qualities that were informed by the music and the emotional arc of the character. At this point, the role of Disembodied Performance in the shaping of The System had also been established and several experiments had been conducted with baritone James Maddalena to determine the types of performance parameters that would be used.

Early in the production process, production designer Alex McDowell had assembled a set of reference images to guide the scenic design. Working with McDowell to shape the dynamic visual language, I had added additional reference imagery of my own and started to envision the stage picture in time. In order to communicate my vision to McDowell and the rest of the creative team, I first drafted a document describing the visual aspects of the production by scene or scene part along with a selection of reference imagery to support my descriptions for each moment. This lengthy document was primarily textual and, while it communicated the ideas in great detail, it had to be read, carefully parsed, and considered in relationship to other aspects of the production, such as the music, to fully envision the entirety of the work. Essential content was there, but the method of presentation was not well-suited for easy communication among a variety of people with differing concerns.

McDowell and I then imagined that there could be a more graphical representation of the information that would encapsulate a variety of different types of information in an easy-to-parse document that could be shared among the creative team individually and in meetings. The document would not be static, but would need to be updated to reflect new ideas and changes when they arose.

Previously, I had merged several versions of the *Powers* libretto, stage directions, and staging ideas into a single annotated and versioned document using Extensible Markup Language (XML). I amended this document to include some of the visual references and descriptions from the lengthy design document. Blocking diagrams and storyboard renderings were also added, each associated with the line in the libretto as set at the appropriate time, thus organizing the design information at a more granular time scale. I further extended the information contained in the document with a suggestion of the color palette, indicated by two colors at any point in time and varying with transitions between fixed color annotations, as well as continuous curves representing the anticipated influence of certain types of Disembodied Performance data.

Having all of this information arranged in time within a single, editable XML document solved part of the challenge. However, it needed to be presented in an easy-to-read format. For this purpose, I created an

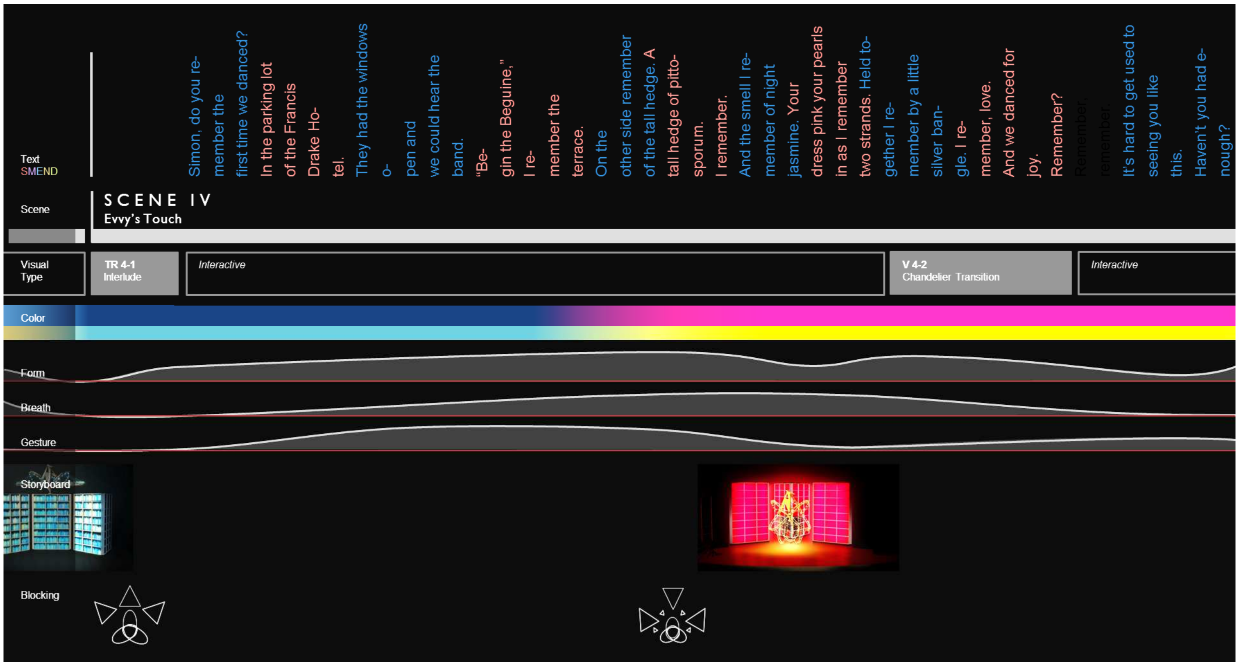


Figure 47:Visual score for *Death and the Powers*
Visual score illustrating libretto text in line with sections, events, color palette, parameter envelopes, reference images, and blocking diagrams

Extensible Stylesheet Language Transformation (XSLT) from the original XML markup to the XML-based Scalabale Vector Graphics (SVG) markup language. Thus, the source XML document could be opened in a web browser using the XSLT style sheet to render the information in a graphical format resembling Eisenstein's *Alexander Nevsky* diagram shown in Figure 20. The resulting *visual score* for *Death and the Powers* featured annotations for scenes and sections of the opera above which the libretto text was arranged in time, as it was presented in the music and annotated using color for each singing role. Below were several tracks showing the reference images, blocking diagrams, plots of the Disembodied Performance data types, and color palettes. Later, I added indications of interactive modes for the Disembodied Performance mapping and Render Designer based on notated triggers in the musical score.

The ability to distribute this information in a single document and to easily open it with the web browser was useful for sharing the state of the opera's design with the team. However, due to the length and complexity of the information, it was not possible to view the entire opera at once, but the representation needed to be scrolled within the browser window. For one meeting of the creative and design teams, the entire length of the visual score was printed on letter-size paper and taped to a wall, so that notes and references to the document could easily be made. The printed document was nearly 20′ long. In addition to the challenge of viewing the entire document, it was not possible to directly edit the rendered document. Edits needed to be made to the XML file, which is a tedious process and a deterrent for collaborative editing by the production team.

4.2 Design Objectives

The implementation of a system for composing Media Scores should account for the use-cases described in the preceding section. Additionally, a tool for designing and shaping and creating the modern-day Gesamtkunstwerk needs to flexibly encode the artist's intention regardless of the modalities employed by the final form of the work. It can also facilitate collaboration and realization of the work of art. Software systems act as translators, turning stored information from one form to images or text on the screen and interpreting user's key presses and mouse clicks and manipulations of the represented data. Software can also be used to navigate events and modifications of data in time and communicate state with other software applications. Thus, a software based score exists in a format that has considerably more flexibility than a paper notation which must be interpreted by a human in order for it to have any function.

This spirit is related to the design principles for performance systems outlined in Section 3.6. By creating a software tool that adheres to those principles, the functionality of the system grows considerably from merely facilitating the editing and display of a score-like notation. In fact, as we will see below and in the discussion in Chapter 5.3, this extends not only the capability of the score, but the function of the score and the composer.

Before detailing the implementation of a Media Scores composition tool, I will describe the design objectives for a Media Score itself. These objectives suggest the types of information that a Media Score should encode, as well as the roles the Media Score can take in the process of producing a work of art.

Medium Agnosticism

The goal of Media Scores is to capture the artistic intent for a work of art by its creator. Following the discussion of expressive representations in Chapter 2, I consider the expressive intention independent of the form the work of art takes. An idea, a story, the evocation of an emotional journey can be realized in any of one or more sufficiently expressive modalities. In the case of multimodal works and Gesamtkunstwerk, a number of modalities with independent sets of formal properties represent the intention in concert with each other. Multiple modalities are used to tell the same story or communicate the same expressive ideas together and through their interaction, juxtaposition, and reinforcement.

As such, Media Scores should be medium and modality agnostic. The score itself is a distillation of the author's intention and thus may be interpreted in any medium comprised of any set of modalities. This is done by describing the score in terms of an ensemble of parameters with a semantics that can be applied to or interpreted in a variety of modalities. This parameterization will be looked at in more detail in Section 4.3.2. The result is a single Media Score that can be performed or realized in any number of ways, from an onstage production of an opera to an interactive installation to a web-based experience. This truly transmedia approach to content creation has implications not only for the artist, making available a broad palette of modalities and media for presentation of a work, but also opens up new opportunities for content distribution. Effectively, the *same* work—or variations and extensions of a work—can be presented through multiple channels of distribution, regardless of the particulars of the media.

Temporal Scalability

As a corollary to medium agnosticism, the expression of an artistic idea or story should also be temporally independent. A story can be told through an instantaneous representation or a lengthy symphonic work. One may be figurative, the other abstract, but the time it takes to experience the artwork can also vary. For this purpose, Media Scores should be able to scale the performance of their intent in time. This includes not only being able to produce works with both dynamic and implicit projections of time, as defined in Section 2.8, but also for dynamic representations to be compressed or extended in time. This variability in temporal representation may be used to tell the story within the conventions of particular media, or can reflect an audience member's or viewer's preference.

In many ways, temporal scalability is not all that different from scaling a representation in physical space. In the spatial domain, these sorts of scale transformations are common: viewing a film on a 50′-wide screen in a cinema or the 5″-wide screen of a mobile device. This broad range of scales is a consideration in content creation. as digital works can be reproduced in such a wide variety of forms and viewing contexts. There is increasingly greater variation in how artistic works are being consumed that are as much about transmedia forms, technological innovations, social factors, and economic factors of distribution as they are about the work itself. Artists can no longer have a clear expectation of how their audiences will experience and interact with their work. The size of a painting or a work of architecture is itself communicative and can influence the perception of the piece and often constitute creative choices. So, it should be kept in mind that while intention may be preserved over expansion or compression of the work in space and time, the experience of the work may not be identical.

Parametric Model

By using a parametric model for Media Scores, complex meanings and intentions may be encoded numerically. This quantitative approach enables many possibilities for the application and transmission of the score in multiple forms. It extends the ideas set forth in the development of Disembodied Performance (see Section 3.1.1) beyond live input into a form that can be constructed and shaped both offline and online. A parametric model facilitates many of the design goals for Media Scores, particularly medium agnosticism and temporal scalability, as mentioned above. An abstract, expressive parametric model enables the score to be readily interpreted by computational systems into a variety of modalities. It also allows time to be expanded or compressed while preserving information content. Temporal scalability in its simplest form could be implemented by linearly scaling the times of events and notated values in the score relative to a fixed point in time (likely the start of the score). However, more advanced approaches could use an information-theoretical analysis or perceptual analysis of the score (such as the method for obtaining a perceptual model of the tension of the score described below in Section 4.3.3.1) to scale events in time non-linearly, in order to preserve perceptual relationships and the establishment and violation of expectation that may require different time scales to achieve. In this way, temporal scaling preserving salient features of the time-varying parametric model is analogous to the technique of seam carving used for expanding and compressing images in the spatial domain. Seam carving is an application of dynamic programming that preserves higher-energy content and removes or adds narrow bands of lower-energy content as needed [11]. The quantitative, parametric nature of the underlying representation need not be presented directly to users of Media Scores. However, various representations may allow users to interact with the model.

Unification of Production Phases

Media Scores are intended to play a role at all stages of the production process. By representing the expressive intention of the score's creator in a parametric and digital form, the score can span what are ordinarily distinct phases of the production process. In some sense, it coalesces the continuum from composition to performance and distribution, from creation to consumption. The Media Score is the document that is used in the artist's creative development process, like a sketchbook or collection of inspiration. As the pieces of the experience fall into place, it becomes the score for the piece, encoding the necessary information to convey the composer's intent to systems and creative personnel for realization of the artwork. In the case of generative computation systems and show control scenarios appropriate to performance media, the score's data representation can then become the language that ties the production's systems together. A single artifact can then evolve with the production and serve it in all of these roles. Not only may the phase of production remain distinct, but having a single tool with which to interact with the document blurs the distinction of each discrete stage and means that activities normally reserved for one production phase are available in all phases in which Media Scores is used. This does not mean that Media Scores are the only tool or technique necessary, but that it can function alongside existing methods and integrate with typical practices, all the while providing a new level of consistency and communication among people and systems. This capacity of Media Scores will be analyzed in Section 6.3.3, after we've seen Media Scores put into practice.

Collaborative Document

In the production of a complex, multidisciplinary Gesamtkunstwerk, many factors about the production must be considered at once by personnel in all departments. For *Death and the Powers*, these design and logistical considerations pointed to a need to enhance communication and organize ideas among a diverse team of people. This led to the development of the visual score described in Section 4.1.3. Media Scores should be able to function both as a notation of a sole creator's artistic intent and also as a collaborative document in which to share ideas, document the evolving production, and communicate across departments, distance, and time. Designing Media Scores with collaboration in mind facilitates a broad variety of workflows.

Using Media Scores, creators of works that require collaborations for realization and performance can take on a role greater than that of composer and author. Typical scripts and scores do encode the artist's intent, but still leave the expressive realization to the interpretation of conductors, directors, and performers. By composing a Media Score, the creator can take on a more auteur-like role in the development of a work by shaping the feel of the experience and contributing to the design process, not just the source text. The Media Score composer is able to author not just content, but expressive form in a manner similar to a painter or sculptor who works independently to move from concept to realization.

Whether a creative team of directors and designers are mounting an existing work for which they create the Media Score from a source text or are using the Media Score composed by the artist, the score becomes the reference for realizing a production or multiple productions of the piece. It is a common language for all departments to reference and communicate the current state of the production. Multiple people from different disciplines can work together on an evolving, living document. As changes are made, versioning the score can also allow users to compare the current state with past iterations.

This collaborative capability isn't just limited to a production team realizing a work, but can be utilized during the original composition phase, offering the possibility of collaborative composition rather than single-author composition. Additionally, the score need not be handed off to the production team when the artist has completed the original composition. With applicability in later phases of production, the authorial role can also be extended through performance-time.

Documentation is always an important part of the creation of a complex work of art. Performance works that are meant to be reproduced—a play being performed over a run of many days, a touring production that must be reassembled from one venue to the next, changes of cast or musicians—need to be thoroughly documented, so that they can be recreated. Of course, there is essential and valuable variability in a work from one performance to the next and the Media Score does not of itself describe the technical implementation of a production and all of its systems. Nevertheless, the Media Score can encompass enough information about a complex Gesamtkunstwerk to enable it to be performed, even with variations in the technological infrastructure. In *It Worked Yesterday*, Sebastian Berweck is critical of technological performance, specifically that of electroacoustic music, for not being easily reproducible or performable without the assistance of the composer [14]. For new works that rely on custom systems, the composer is often the system creator and knows how the system should be configured to function and what the output of the system should be like. Berweck argues the need for sufficient modes of documentation in order to allow performers to be able to create the piece at any point. Media Scores allow for the preservation of artistic intent and documentation of many types of information about the work or a production of the work. The Media Scores application is neither the entirety of the infrastructure required to perform the work nor does the score document the infrastructure requirements as Berweck suggests. Infrastructure and connected systems still need to be created or assembled for a Media Scores-run production. In fact, from the perspective of the score, no specific infrastructure is required. Thus, to create a production, any compatible and suitable systems can be used with Media Scores to faithfully reproduce the composer's intent. The range of possibilities can, however, be assessed from the score.

Composing for Live Input

Conventional scripts and scores are subject to interpretation by directors, conductors, designers, and performers. Furthermore, every artistic experience is subject to the interpretation of audience members. However, scripts and scores explicitly say little about the role the audience has in co-creating the work. They are static objections that prescribe the cues that communicate artistic intent. Stage directions or non-standard notations in contemporary scores may invite performers to take additional liberties with the performance of all or part of a work or to provide their own material or suggest chance procedures for creating or organizing what is performed. These approaches are inadequate for describing how the continuous qualities of performance should interact with the expressive essence of a work. The problem is compounded when the interpretative and performative role is not just given to trained performers, but audiences. With increasing number of works seeking to engage audiences, creating augmented or immersive experiences, or inviting direct audience interaction with the development of the piece, the results seem endlessly unpredictable. How can artists encourage the influence of real-time performance technologies and audience involvement while maintaining the integrity of their artistic intent in fluid and meaningful ways?

To achieve this, a goal of Media Scores is to present a model for allowing the composer to define when and how performers, data from the environment, aleatory processes, co-located audience members, and remote audiences can interact with the performance. This is distinct from performance-time composition activities that can be achieved by having the composer directly manipulate the score during performance, which is an open-ended pursuit guided by the artists intuition just as it would be during a separate composition phase. Instead, in these circumstances, Media Scores adopts the notion of composing for interactive and live input as authoring constraints on that input. Regions of time can be defined in which live input is considered by the performance systems, enabling it only where the composer feels it will benefit his storytelling. This contrasts with many installation-type systems, in which interactivity is always incorporated into the systems' functions. The resulting experiences consequently tend to feel unshaped, overly open-ended, and not as communicative or powerful as structured linear experiences like cinema or a novel. Within regions of interactivity, the composer can then constrain how the live data influences the generation of content.

Using the parameters from a Media Score during performance in this way allows composers to easily incorporate live input during the performance. A work scored for live input uses modes defined in the score to enable the input and then parameters can modulate the performance data, giving the composer deliberate control over how the live information is utilized. Live data may originate from sensors on instruments, as in *Spheres and Splinters*, performers, as in *Death and the Powers*, or from the audience, as in *Sleep No More*.

Aesthetic Artifact

The Media Score itself is a work of art that conveys the expressive intent of its creator. Unlike traditional symbolic forms of notation, which must be interpreted or even performed to be understood, an impression of the emotional journey intended can be apprehended by viewing the Media Score. The score itself is an artwork or a view of *the* artwork for which it encodes. The underlying parametric model enables the rendering of this aesthetic representation to be modified in real-time. The representation of the score is expressive and not figurative or diagrammatic. In practice, all types of views of the score are available, as explained in Section 4.3.3. Figurative design information, representing the actual form of a possible production of the score, can be included as assets associated with the score. Diagrammatic representations offer alternative ways of thinking about and manipulating the score information. However, the core representations of the score are abstract, expressive views that produce a phenomenological appreciation of the artwork from the score.

Control System Architecture

As noted in the discussions of *Death and the Powers* and *Sleep No More*, above, and in traditional theatrical practice, the control information used in running the technological components of a performance or experience is distinct from the script or score on which the performance is built. Only the actors or musicians reference the source texts directly during performance. Media Scores merge both of these types of information into a single document. The Media Score itself is a data format that can be exchanged and shared with other software and systems to realize a performance.

In order to function fluidly in live performance settings, when applicable, and to serve as part of a greater ecosystem of performance technologies, Media Scores is designed to adhere to principles of performance systems defined in 3.6. For example, interconnected systems allow for the exchange of information in real-time during a show. Therefore, the parametric expressive representation and its associated metadata of events contained in a Media Score should be used in performance settings as a form of show control. Playing a Media Score runs the performance. Connected output systems for visuals, sound, et cetera can utilize the expressive parameters, assets, and events emitted by the Media Score to generate output by interpreting the semantic parameters and mapping them into formal parameters.

In designing systems for *Death and the Powers*, I advocated for interfaces that were modeless and offered local feedback. Modeless interfaces, in this context, means that there is no distinction between an editing mode and a playback mode. Content may be created and altered while the system is processing live data in real-time and displaying output. Local feedback allows the output to be monitored and viewed as it is being generated during the performance, as well as a preview when offline and not connected to the entire performance infrastructure. Incorporating these behaviors into Media Scores gives immediate feedback about changes during composition and opens the door for composition-like or improvisational activities at performance-time.

4.3 An Implementation

With these particular design goals in mind, we now turn our attention to an implementation of Media Scores that addresses each goal. To summarize the structure of a Media Scores composition and playback environment, let us consider what elements are a core part of the implementation and how information about these elements are encoded in and used by the score itself.

The process of creating a Media Score may often begin with a source text, an existing work of literature or music around which a particular multimodal production will be designed. The relevant elements from the source text should become a part of the Media Score. As a tool in the design process, additional information, including notes, stage directions, and design reference—such as inspirational images—should also be incorporated into the score. These assets function primarily as annotations on a set of expressive parameters that form the core of the score. The principle act of the Media Score composer is to represent the expressive arc of the experience of a work in the score. Together, all of these elements are input and combined into the Media Score document.

The score itself is represented by one or more views that render the parameters and reference assets in a variety of different manners providing multiple perspectives to the composer and production team, enabling each to understand and manipulate the score.

During performance, when used in show control applications, the score then transmits its data for use by other systems that serve as content generators for the desired output modalities and styles for a work of art. In cases where there exists interactivity or real-time performance data influencing the generated content, the performance or audience data would be analyzed and reduced to the appropriate expressive parameter set. At that point, the live data can then be modulated by the Media Score before being relayed to content generators, in order to preserve the composer's intention, but still preserve a sense of immediacy and apparent influence.

The Media Scores implementation described here slightly favors time-based ways of reasoning about the score and its performance, but such a bias is considered as a superset of possible temporal representations. That is to say that, should the implementation be capable of creating, managing, and performing the time-based works, it is also capable of serving time-less media.

The schematic diagram in Figure 48 summarizes the architecture of the Media Scores implementation and its usage. At the core of the implementation is the Media Score data model, a persistent and mutable representation of the data encoded by the score. This data model is described in detail in Section 4.3.1. Looking toward the left-hand side of the diagram, we can see that the Media Score incorporates user-supplied assets that describe the artistic intent by example and potentially includes a source text such as an existing script or score, if the Media Score is used to realize

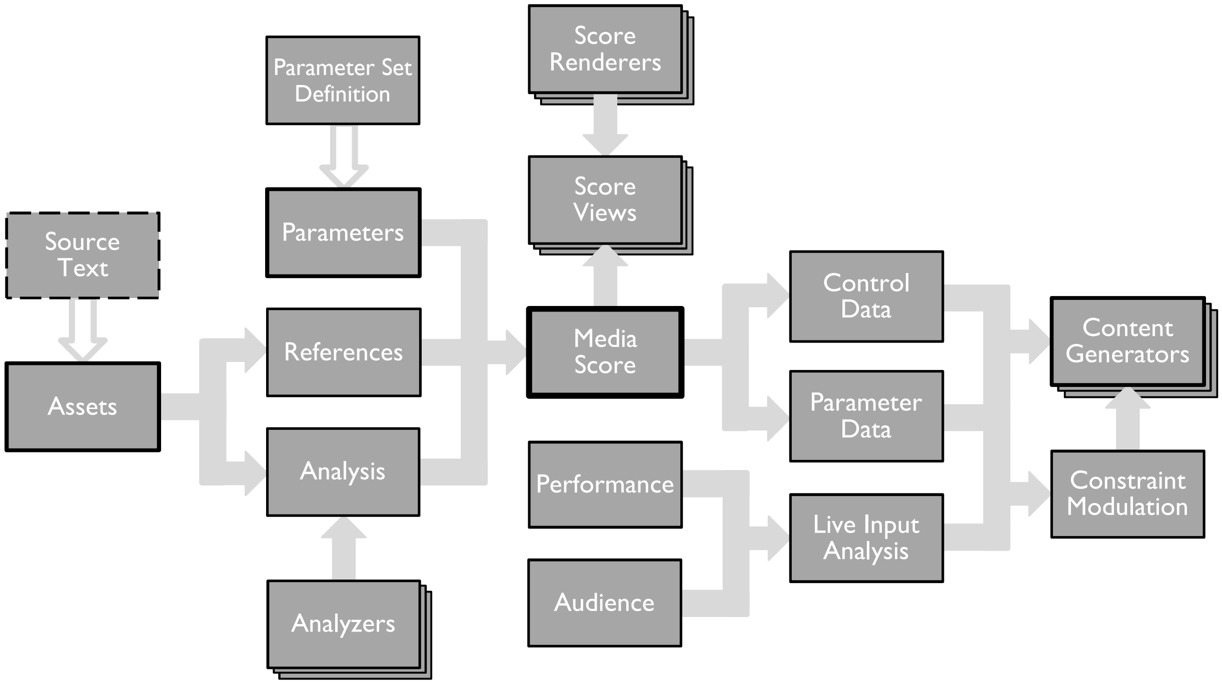


Figure 48: Media Scores conceptual flow diagram
User-specified assets and parameter values define the Media Score. The score data is visualized by renderers during interaction within the Media Scores application. During performance of a score, the system emits control and parameter data to content generating systems to produce output representations, potentially incorporating real-time data from performers and audiences.

an existing work. The other critical type of information that a Media Score contains are a set of expression parameters, as explained in Section 4.3.2. These parameters encode the expressive intention of the score and should conform to an established semantic convention—a parameter set definition—in order to be faithfully interpreted by the Media Scores environment and connected systems. Within the software implementation, a set of views present information from the data model to the user and allow the user to interact and modify the score. An assortment of score renderers does the work of representing score information for each view. At the right-hand side of the diagram, we see that the Media Score can emit show control information, as well as the expression parameters in order to drive content generators that realize the work of art. This capability is covered in Section 4.3.5 Content generators include audio and visual output systems, animatronics, and other theatrical technologies or processes that can present the artwork encoded for by the score. Media Scores allow composers to incorporate interactivity into the performance of a work with constraints defined by the composer in the score. Live performance data from performers or audiences is analyzed and then modulated and combined with data from the score to be used by content generators.

This architecture has been designed to be extensible in several ways. The diagram indicates several points where the implementation of Media Scores is modular and may be extended: analyzers, views, renderers, and content generators. Additional renderers for views of the score may be created and used by the software application to provide new methods of representing the score. Assets may be analyzed to generate metadata and information that is meaningful given the semantics of the parameter set definition. New analyzers can be written to handle additional asset types or use different analysis techniques. For playback in show control scenarios, Media Scores implements control messages that can readily be translated to a variety of connected systems and content generators allowing a broad range of equipment and software to be incorporated into a performance environment alongside the Media Scores application.

The implementation described in the remainder of this chapter has been implemented as a web-based application. While still relatively uncommon for creative tools and performance technologies, recent advances in browser support for web technologies including HTML5 and related application programming interfaces (APIs) mean that browser-based applications are sufficiently capable of the necessary interaction, graphic, audio, storage, and networking capabilities required. Due to current browser support of these features at the time of this writing, this implementation targets Google Chrome 21.0 and later. As standards continue to be adopted, other browsers will likely be capable of using Media Scores, as well. Furthermore, the implementation of Media Scores as a web-based application means that it is readily accessible from any computer with a support browser without the need to deploy and install the application. For a tool that is intended to be used in collaboration, this makes accessing and sharing Media Scores particularly easy among large teams and individuals with various comfort levels with creative software applications. A relatively simple markup page forms the structure of the user interface and the visual presentation of the interface is managed using Cascading Style Sheets. The remainder of the application is written in JavaScript (ECMAScript 5) in a pseudo-classical object-oriented style. Code dependencies and dynamic loading are managed through the RequireJS asynchronous module definition framework. Renderers in the current implementation take advantage of the WebGL API for hardware-accelerated rendering of 3D and painterly representations using the Three.js library.

The Media Scores user interface features a central workspace comprised of one or more user-configurable views of the score. Available views and modes of interaction with them are listed in Section 4.3.3. Above the workspace, the title of the score project is shown with the composer. These values can be edited directly and update the score metadata. To the left of the score title is a dropdown list of sequences within the score. All scores have at least one sequence and only one sequence may be active at a time. The active sequence is selected and displayed here. Controls for creating new sequences and removing existing sequences from the list are provided. Additional sequence configuration options are accessible via keyboard shortcuts.

Below and to the left of the score and sequence metadata is a toolbar consisting of buttons for three types of operations. All toolbar commands

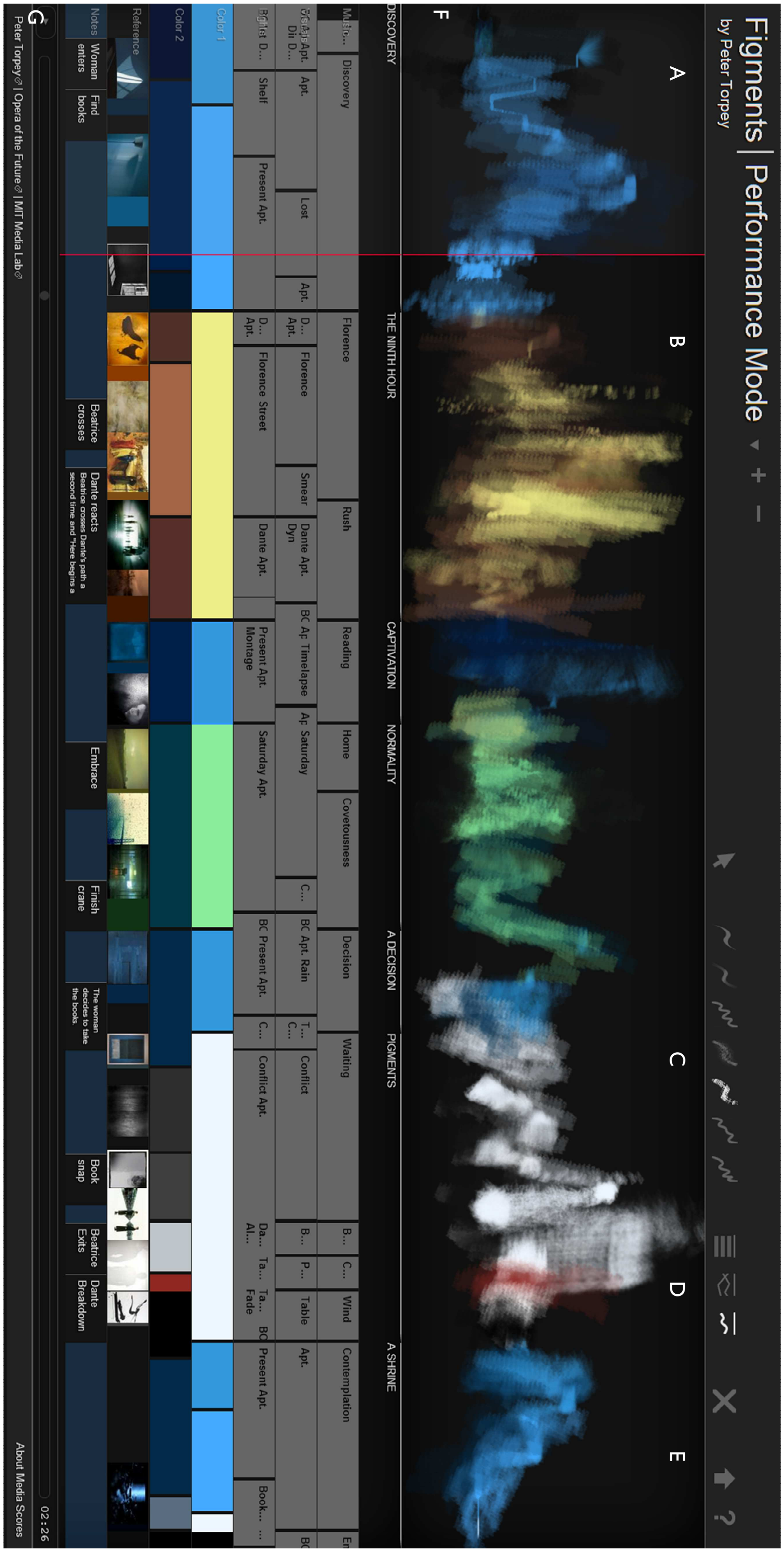


Figure 49: Media Scores application

The user interface of the web-based Media Scores application features (A) editable metadata, (B) sequence management and selection, (C) parameter "brush" selection tools, (D) view selection tools, (E) score management functions, (F) workspace with score views, (G) transport controls

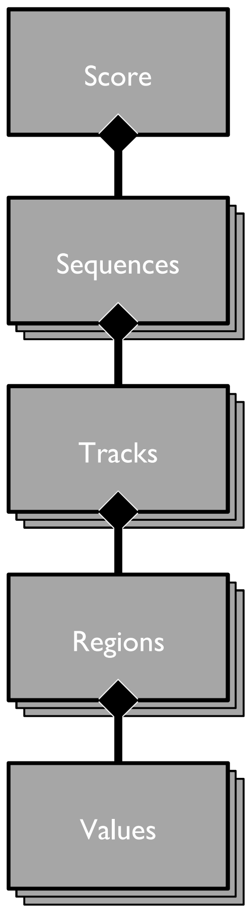


Figure 50: Media Score data model scheme

have keyboard shortcuts associated with them, as well. The first command group selects expression parameters for editing, a process explained in the following sections. The second command group provides options for opening new views and renderers. Within the workspace, the focus tracks interaction by view pane. Selecting a new view or renderer will replace the focused view with the specified new view. If no view is active, a view is added to the workspace. The third command group presents score-level operations, such as enabling show control, creating a new score, and committing the current score to the server-side repository. Below the workspace are situated the transport controls: a play/pause button, time track with an adjustable current time indicator, and a display of the current sequence time.

In addition to the Media Scores web application itself, which allows for editing and playing back scores, other web pages were created to link the user experience with a backend database. This database enables additional sharing and collaboration features, as well as a repository for scores and assets. User can log into the website and retrieve their scores, loading them into the editor, from the website.

4.3.1 Data Model

At the heart of the Media Scores implementation is a data structure that represents the expressive parameters and other information that is useful to represent and communicate as part of the composition of a work of art. All of the information contained is essential to this vision of composition through an expressive and collaborative score. However, as is the theme of Media Scores itself, the data and its presentation are kept distinct. To the composer, the organization of information within the score should appear logical and intuitive, but the details—what is stored and how—should be opaque. It is the task of the composition interface and the systems that rely on this data to represent it in a meaningful and malleable way.

Nevertheless, I will describe the schema of the data structure and then subsequently show how this information is presented to the user. For the current web-based implementation, the data model of a Media Score exists in memory as a hierarchical JavaScript collection of objects and lists of objects of well-defined types. A classical object-oriented approach is used to define the types and their capabilities. The data structure can then be serialized to JavaScript Object Notation (JSON) for persistence and relevant parts of the structure in JSON format can be transmitted for collaboration and show control applications.

Each Media Score document contains a set of metadata applicable to the entire score or project it represents. This information includes the title of the score, the composer, and score format version. Although not currently used, should the specification of the data model change, the format version can be used to properly interpret the document. Additional metadata about the score document, including versioning and access control is available when the score is stored and retrieved from the Media Scores server backend. This data is external to the score file, residing in a database, as explained in Section 4.3.4.

In addition to global metadata, a Media Score document contains one or more *sequences*. A sequence contains the actual descriptive data for a score. Sequences may be named and each specifies a duration for explicit or dynamic time representations of the sequence. For renderers within the interface that represent time implicitly and for output media that are implicit projections of time, the actual durations of sequences and timing of values have no sense with respect to a temporal measure. Instead, times and durations are effectively abstract quantities that can describe a relative locus in the viewer's experience of the work (see Section 2.8 for a discussion of Projections of Time). Sequences can represent different parts of a large work—such as symphonic movements or acts in a drama—or variations on a work for different purposes, such as interactive versus non-interactive modes of presentation of a work in cases where the telling of the story would not be adequately achieved by simply not having live input into the system. Sequences can also function as logical divisions of the work and be nested into other sequences. This enables the reordering of large portions of score data by composers and collaborators, as well as independent concurrent editing of separate sequences. Like many show control systems that feature multiple cue lists, sequences may also be loaded and executed during a performance in response to triggers.

Sequences are composed of a set of *tracks* that store the expressive content and annotations encoded by a score. Each track, in turn, contains one or more *regions*. A region associates the track's value with a particular time or range of time (specified by a start and end time or a start time and duration) in the sequence and may, itself, be named and described for ease of use. Regions are the principal point of interaction with the data model when composing and manipulating a Media Score. The composing environment principally provides multiple methods for adding, removing, and moving regions, in addition to modifying their value. Regions may have one or more values, depending upon the type of track. Multi-valued regions have values that are associated in time relative to the start time of the region and are only valid for the duration of that region. Thus, as regions are moved in time, their values move correspondingly. Regions in tracks may not overlap; a track may have only one value at a given time. The value of a track at a given time is the value of the region at that time or the linear interpolation of the nearest values if a multi-valued region is present at that time. If the instant in time is not contained by a region, its value is computed based on a configurable interpolation function applied between regions for tracks that have a data type that is interpolatable.

Tracks have a default value appropriate to their data type, which is used when the value is otherwise not defined by a region and the track's interpolation setting.

Multiple track types are available, each having a particular purpose defined by the data type they contain and their function. The track types can be organized into three classes: *events*, *parameters*, and *assets*. Each track can be named to describe its function and has a Boolean property that determines whether the values of that track are output as show control data when performing a Media Score online.

Event-type tracks contain discrete information that function as user annotations or show control data. Event tracks can also serve as markers for advancing playback of a Media Score in response to external commands and for moving around the score in a random-access fashion. These event track types include Sections, to define named regions of a sequence, Notes, which provide textual annotations, and Triggers, that enable Media Scores to send show control messages to other systems. A fourth event track type, Time, provides some control over how Media Scores are played back in time-based works, including defining regions of time that may repeat for a fixed number of iterations or vamp until triggered. Events in the Time track can also internally trigger non-linear jumps in playback of the score. While a sequence may contain any number of tracks of any other type, it only may contain at most one Time track. If no Time track exists or if it contains no data, time-based playback will proceed linearly.

The second class of track types is parameters. Parameter-type tracks represent continuously-varying data with floating-point values in the range of [0.0, 1.0]. Control tracks represent keyframed parameter data that may be used to drive other control systems in a typical fashion. This behavior is akin to the track data in the *Death and the Powers* robot control system mentioned in Section 3.1.2. These tracks generally have a literal semantics in that their values directly influence a physical property or a formal property of connected systems and output renderers. In contrast, the semantics of Expression tracks specifically suggests the use of continuous data to represent expressive parameters. Expression tracks represent the fundamental principle of Media Scores: the encoding of expressive intent in a modality-agnostic ensemble of high-level, abstract parameters. Expression tracks may contain not only a continuous function, but an envelope around that function allowing a composer to modulate the influence of live performance data. Expressive parameters will be discussed further in the following subsection. A third kind of parameter track is Parts. While Expression tracks are used to shape the overall expressive intent of the work of art, Parts specify the expressive state of an element of the composition—a character, a theme, an instrumental part—relative to the overall shape given by the Expression tracks.

The third class of track types is assets. Assets are annotations or resources that do not directly represent the expressive or functional properties of a score. However, they can serve as important information during the creative and development process and can be used to shape the expressive parameters of the score or as source material for output renderers. The Media Scores implementation at the time of this writing features six asset types: text, color, image, audio, video, and music. The Text asset type is the simplest, with a value that represents a line of text that represents some element of content, in contrast to the Note track type which is used for production annotations or stage directions. These values may be from the source text for a Media Score, when creating a score for a work of literature, poetry, or prose with a previously existing script. Text tracks could also be used to contain lines of dialog or lines from the libretto of an opera as with the text in the visual score for *Death and the Powers* (Figure 47). Color tracks define a set of colors that vary during a sequence and can be used by visual renderers and output to other systems. Many of the renderers described in 4.3.3 expect at least two color tracks in each sequence in order to define a color palette for the piece. When a color is not available, renderers default to black and white.

The remaining asset types of image, audio, video, and music are presented in some score views, but primarily serve as references. The actual data for these media are stored on the Media Scores server and indexed by the database for later access and sharing across projects. Consequently, the actual value stored in the score for each of these track types is the uniform resource identifier (URI) required to retrieve the asset from the server. Assets are not embedded into the score document itself. Asset tracks can have *analyzers* associated with their type. These analyzers are modular algorithms that derive metadata or expressive parameter values from the asset. For analyzed assets, the values stored with each region in the score include both

	Track Type	Data Type	Default Interpolation	Default Output
Events	Triggers	String, Integer	None	■
	Sections	String	None	
	Notes	String	None	
	Time	Commands		
Parameters	Expression	Float	Cosine	■
	Parts	Float	None	
	Controls	Float	Linear	■
Assets	Text	String, Analysis	None	
	Color	Color	Linear	■
	Image	URI, Analysis	None	
	Audio	URI, Analysis	Linear	
	Video	URI, Analysis	Linear	
	Music	URI, Analysis	None	

Figure 51: Track types
This table shows the currently implemented track types along with the data type of their values, default interpolation function, and whether or not the value of the track is output by default during playback for show control. The interpolation function and whether output is enabled may be configured by the user.

the URI of the resource and the results of the analysis. The analyzed data can be viewed by the composer as another tool in the creative process or may be applied directly to other aspects of the score. Thus, as inspirational assets are assembled during the initial phases of creating a score, they can be readily used to begin shaping the expressive parameters of the score. This process is described in more detail below.

4.3.2 Expression Parameters

While many types of information may be incorporated into a Media Score document, as explained above, the unique aspect of Media Scores is the emphasis on notating the expressive quality of the work. This sets Media Scores apart from other general-purpose software for scripting and production planning, as well as existing show control and timeline management systems. As such, the most important parts of a Media Score are its expression parameters that represent continuous data about the intended feel of the artwork.

Expression parameters represent continuously-varying values of a particular set of semantically meaningful qualities. This use of parametrical models of artistic and human expression builds on the approach laid forth in my formulation of Disembodied Performance [107] and is grounded in the assemblage of theory offered in Chapter 2 of this thesis. The live performance data used in Disembodied Performance and mapped into a parametric space is an example of modeling the affective and expressive intention of the performer in real-time. In Media Scores, the artist is given the ability to score out these parameters offline; to sculpt the performance or artwork in advance of its realization. This type of expression is what is inferred from in between the lines of a traditional script or is evoked by the notes of a musical score. In those instances of symbolic representation, the expressive intent isn't explicitly notated. However, as we have seen, it is this dynamic that lends works of art their emotional resonance. The perception of expression is inferred from the formal parameters of the performance by an audience. In the creation phase, it is notated in terms of formal parameters by the composer. The process by which the composer maps his intention into a set of formal parameters is largely intuitive and is the result of experience. It is a synaesthetic process. Similarly, the process of reconstituting that intention in the minds of audiences is an intuitive and learned process for the audience members when perceiving stimuli produced by the artwork or the performance of the artwork, with additional translations from form to form by production personnel and performers along the way. At all stages, this interpretation introduces variability, but seeks to preserve intent. In Disembodied Performance, the expressive performance parameters were an intermediate model—in the reified inference sense—of instantaneous expression. In Media Scores, the expression parameters are the intermediate model for the entire work of art. Media Scores minimize the indirection between intention and encoding or encoding and realization in the various phases of the production process. This does not remove the possibility of variation and interpretation, but provides a more direct pathway for composition and performance to take place.

Modeling expression, particularly for application to Gesamtkunstwerk, places certain requirements on the semantic definitions of the parametric space. The meaning of abstract expressive parameters may be difficult to articulate or to have associated with words. What is needed is a parameter space that spans the range of expressive intents that are otherwise intuitively translated into formal properties. The dimensions of such a model space need to be reasonably orthogonal in order for each to be independent and concisely descriptive. The expression parameters need to be defined at a higher level of abstraction than the formal parameters of any particular modality. They must be capable of being applied to or represented using one or more formal parameters of any modality such that they can be represented in both score form and through the actual production of the work of art in multiple media. They represent, at an abstract level, commonalities across modalities. As we saw earlier, the intuitive pairing of a name with a shape in Köhler's "takete" and "baluba" example illustrates a synaesthetic gestalt that transcends modality [51]. The shape and word representations share a common concept. Metaphors are a good place to start when thinking about commonalities of formal parameters across modalities: the notion of rhythm in music and architecture, brightness in timbre and color, sharpness in touch and taste. We can understand these concepts because they reveal some commonality. We perceive the same qualities in multiple modalities and they are generally consistently-interpreted. The metaphor communicates from one person to another the concept of the quality being discussed. Beyond a multimodal abstraction of formal parameters, expression parameters must describe a perceptual effect, as well as an emotional one, describing what concept of *feeling* is invoked by a quality of stimulus.

As with the metaphorical descriptions of formal parameters that are mostly well understood, there is perhaps a parametric model space that could be defined that would suitably capture the shades of expression and be universally applicable. I believe that such a space likely exists. It is how art, in a sense, works; that an artist can transmit an idea from his mind into that of another through multiple layers and modalities of indirection. The challenge is to arrive at such a space. I believe this to be a useful, but lofty goal that likely will not be achieved for some time to come. I think that one of the significant challenges in attempting to derive such a space is one of language. At the required level of abstraction, the words we have to talk about and classify the concepts of expression that are so much a part of our experience may be an impediment. To decide on the axes of expressive parameters, and to establish consensus on them, generally requires naming them. However, the commonalities we feel and want to capture may elude verbal designations, which is why we consider them *felt* and why we use non-verbal or poetic media to represent and transmit them.

Since the definition of a truly universal parameter set is deserving of considerable research in its own right, neither Media Scores nor this thesis specifically defines a particular, universal set of parameter semantics. The data model described in the previous section makes no mention of the specific set of expression parameters used. Most of the Media Scores implementation does not need to treat parameters with any particular semantics. Ideally, short of arriving at a universal set of expression parameters, users should be able to define their own desired set. However, non-diagrammatic renderers, discussed in the next section on Views and Interaction, and content generators do need to interpret expression parameters within a particular semantic framework in order to represent them in a manner consistent with the modality at hand. Thus, renderers and content generators expect a specific set of parameters and identify them within the score's data model by name. Scores should therefore have a common set of expressive parameters across all contained sequences and their expression parameter tracks designated by the expected parameter names. The renderers implemented will simply fail to incorporate unexpected parameters or not reflect the appropriate variation if an expected parameter is missing. If a user chooses to specify an alternate set of expression parameters, the renderers will need to be modified or new renderers created to interpret the new set properly.

During the development of Disembodied Performance, the set of performance parameters used to represent the character across modalities evolved considerably. It began simply with a three-dimensional circumplex model of affect: arousal, valence, and stance. While the sensed performance could be mapped into this space, it did not seem to span the range of expressive gesture used by the opera singer nor was it of sufficient dimensionality to affect the visual and sonic content generated by the performance. The circumplex model was used as part of the analysis of the performance and augmented with the six parameters listed in Section 3.1.1. My definitions and designations for these six parameters continued to evolve in subsequent projects. For the current implementation of Media Scores, I have settled on a seven-dimensional space using the following parameters:

WEIGHT	magnitude, degree to which attention is attracted, relative impact or proportion
INTENSITY	strength or amount of presence
DENSITY	number of events per instance of time or point in space, amount of activity, layeredness
COMPLEXITY	degree of variation at different scales
TEXTURE	quality of detail, smooth or continuous to varied, timbre
REGULARITY	degree of metrical order, temporal or spatial consistency
and RATE	frequency of events per unit of time or space

Media Scores treats color—a property composed of formal parameters of hue, saturation, and value, or other bases typically in three dimensions—as an asset rather than something that is derived from expression parameters. This may seem counterintuitive when color is so commonly viscerally associated with emotion and the semiotic denotations of colors are deemed to be intimately tied with established feelings, situations, and evocative symbols. Color is treated distinctly from expression, as it was in the development of Render Designer for *Death and the Powers*, for two reasons. The first reason is one of semiotics. Ingrained in us are clear correlations between colors and emotional states. Some of these arise from physiological responses to color. Brighter, warmer colors such as yellows and reds activate alertness and arousal responses. Cooler colors tend to have an inverse effect. However, particular significations—red with passion or life, green with jealousy or prosperity—are not universal. Rather, they are often culturally specific, socially-constructed meanings and have evolved and changed with time. That there is a connection between color and emotion seems to be commonplace, though the associated denotation for particular hues is not something that can be reliably considered.

The second reason for not deriving color from expression parameters is a practical one. In such an artistic context, color is left to be a parameter of design. The artist or creative working with Media Scores is free to define a suitable color palette for the work's aesthetic and intent they infer from the score and any other sources. Media Scores does not restrict color to a particular mapping, but favors the expertise and creative talents of humans to use color in unique and compelling ways. When production personnel from a variety of departments such as scenic design, costume design, and lighting come together, their color needs or creative approaches are influenced by the color selections of their peers. A single palette of colors may not suit all modalities equally. Color may have other functions in the course of a production: for example as leitmotif or to segment space and time with locations or distinct scenes having unique color palettes. Complex artistic, psychological, and even physical processes come into play. Designers are free to annotate color as they see fit within the Media Score.

This distinction of color from expression parameters is simply a convention. The currently implemented renderers look to color asset tracks for their coloration. Expressive parameters, however, are often used to control the balance or variation among this set of user-specified colors, but do not decide the colors from the set of all possibilities themselves. Despite the emphasis on expressive representation, this serves an intuitive diagrammatic function even in the most abstract and painterly of renderers: to visually indicate specific portions of time or regions of values based on their color. If a scene in an opera is intended to have blue-ish hues and this is indicated in color asset tracks, this scene can be identified in the various renderings of the score by its blue-ish color. Since the set of available renderers is extensible, renderers could be designed that generate color automatically from expressive parameters, foregoing color asset annotations. Similarly, analyzers can be created to derive values for the set of expression parameters given a color value. The analyzed values can then, if desired by the user, be applied to the expression parameter tracks at the appropriate location in the sequence.

Expression parameter values can be composed in a variety of ways using the Media Scores application. As shown in the next section, this is often through painterly interaction with different representations of the score, or by setting the values of regions explicitly through independent views of the track. Expression parameters should reflect the overall impression of the artwork; the summation of what is happening in the music over time or the overall stage picture. However, stories and artworks are often more complicated than presenting a single expressive trajectory at a time. There is often musical or visual counterpoint, multiple characters, elements juxtaposed in time and space. These different parts of the composition weave in and out of consistency in parallel or contrasting motion through the expressive parameter space.

Instead of describing each element of the composition as its own ensemble of expression parameters and then attempting to determine the net effect, Media Scores provides the part track type to assist the composer in defining these relationships. For each element or part in the piece, a corresponding track is added to the sequence's score. Part tracks represent a one-dimensional parameter value that gives their deviation from the overall effect of the expression parameters. Part tracks, by having a default interpolation of "none", are similar to the per-character expression curves shown in Story Plot (see Section 4.1.2). This way they can indicate the presence of a part and how it relates to the whole of the experience. The construction of parts as relative to the overall experience encourages the shaping of the piece as a whole, instead of one composed from the integration of potentially unrelated elements. This top-down approach to storytelling ensures continuity and cohesiveness, as we are aptly reminded in Aristotle's *Poetics*:

> We maintain, therefore, that the first essential, the life and soul, so to speak, of Tragedy is the Plot; and that the Characters come second—compare the parallel in painting, where the most beautiful colours laid on without order will not give one the same pleasure as a simple black-and-white sketch of a portrait. [6]

The structure must first be present and, even in the abstract, the embellishments and constituent parts must be guided by that form. It is from this variation of the form that they acquire their meaning, their information in counterpoint to the expected norm at the core of the work. Therefore, the actual values of a part track are derived by applying weighted constraints to the expression parameters and deviating from them by the value of the part at a given time in the sequence. This constraint is solved such that the amount of deviation is satisfied across all of the expression parameters.

Expression parameters are also critical to the process of composing for interactivity in a Media Score work. Regions in expression parameter tracks can define an envelope or range as upper and lower bounds relative to the region's value. This range constrains the excursion of that parameter from its specified value in response to live input at performance time. Live input may come from a variety of sources including the sensing of performers, information about environmental conditions, aleatory processes, sensing of or input from co-located audience members, or input from remote audiences. Audience inputs may be collected through sensing or interaction with elements of the performance, specialized devices and controllers, or software or computers or mobile devices. As with any live performance data, the raw data must first be analyzed and translated into the same parametric space of expression that the Media Score uses.

Incorporating live input into the performance of a Media Score differs from performance-time composition activities that the Media Score application permits. The former is used in a manner specified by the composer in the score with the modes of interaction defined in advance within the composition. During the latter, the composer can manipulate the score using the application interface in real-time as it is being performed. Performance-time composition is not unlike the activities of DJs or VJs in live performance, as it involves improvisatory shaping of content to tell a story. The difference here is that what is being manipulated is the affective influence rather than the content itself and its formal properties. The operation is carried out on a dynamic notation of the work that is subsequently interpreted, instead of sequencing and manipulating the performance materials directly. Like other dynamic score systems, this type of performance-time composition can influence not only technological productions in various modalities, but human performers, as well.

4.3.3 Views and Interaction

Within the Media Scores software environment, the score is presented to users through a variety of views. The views allow for the display of parts of the data model with different representations highlighting different aspects of the score or ways of reasoning about the score. The views are also responsible for how the user interacts with and manipulates the score data model. For the most part, views are visual, onscreen representations of the score in the web application. However, views also may serve other functions, as we will see shortly.

Views are implemented as subclasses of a generic `ScoreView`, which represents a user interface component that can be added to the web application's layout for display. There are six subclasses of `ScoreView`, one for each projection of time (see Section 2.8)—explicit, implicit, and dynamic—in each of two flavors for two- and three-dimensional representations. Views manage the mapping to and from how time is represented in the Media Score data model and how it is represented onscreen. These mapping functions are exposed to rendering delegates in order to draw the score representation and are used in interaction to translate the user's mouse interactions into manipulations of properties of regions in the data model, converting 2D or 3D space to time and value. They also are used as part of the application's focus and selection management for actions that are invoked at the application level, such as keyboard events or changes of views. Views are notified of changes to the data model and can update themselves accordingly in response to such events. Thus, all views remain in sync with the current state of the model affording a real-time interaction with the score and output values.

Each view is associated with a subclass of `ScoreRenderer` when it is constructed. The `ScoreRenderer` instance is delegated the responsibility of actually drawing the score in a particular style or otherwise computing an output from the current state of the data model. Renderers vary from producing expressive to diagrammatic representations or functional output. A special subclass of `ScoreRenderer`, `CompoundRenderer`, can be instantiated to composite the rendering of multiple renderers into a single view. The modular nature and weak dependency allows many `ScoreRenderer` types to be created or extended, so new representations can easily be added to the application. As noted in the previous section, in the current implementation, representations that generate output using a particular semantic interpretation of the information contained within a track expect to identify that information through a naming convention.

4.3.3.1 Explicit Views

Explicit views of the score are generally familiar from other time-based editing software, such as digital audio workstations (DAW), non-linear video

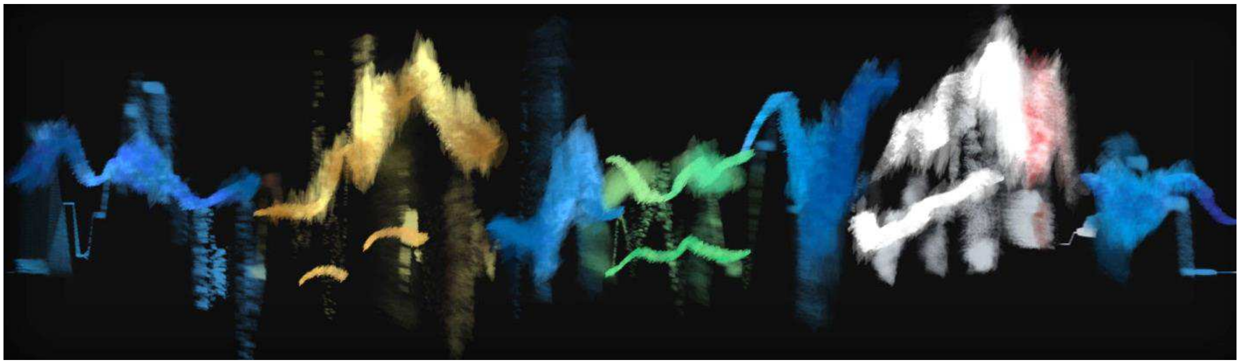


Figure 52: Sketch renderer
Expression parameters of the selected sequence are rendered as a line in a painterly style that visually communicates the magnitude of each parameter with an explicit projection of time. Defined parts are shown as lines for regions that deviate from the overall shape.

editors (NLE), and computer animation packages. Time, for the duration of the sequence, is presented from left to right in the horizontal dimension of the view pane.

The cornerstone representation in the Media Scores application is the Sketch. This explicit-time view uses a painterly renderer to represent the expressive parameters for a sequence. The Sketch renderer draws a single line through its view, varying the quality of the line to reflect the values of the expressive parameters. These visual qualities include the vertical displacement of the line, its thickness, softness, diffusion, texture, and complexity. An expressive parameter, "texture" in the current implementation, transitions from the value of the first color track in the sequence to the second. Thus, the transition of color is an expressive parameter, but not the color itself. It is through representations such as the Sketch, that the Media Score can function as an aesthetic artifact. The explicit-time projection of the view means that the entire arc of expressive shape for the work can be seen at once, impressionistically conveyed through the ever-changing quality of the line. The color of the line still provides a cue to the designed aspects of the work and how it might feel when performed. Part tracks may optionally be displayed and are rendered as independent lines that deviate from or conform to the central sketch line.

Users interact with the Sketch representation through a painterly interaction appropriate to its visual style. For each of the tools that represent the current set of expressive parameters, a paintbrush in the Sketch view is activated. The user can easily switch from painting with one expressive parameter to another using buttons in the application interface's toolbar or using keyboard shortcuts. Painting over the Sketch by clicking and dragging the mouse will increase the value of the active parameter at the time in the score continuously computed from the horizontal position of the mouse cursor. Depressing the [`Shift`] key while painting has the effect of decreasing the value, as if "erasing" the parameter during the paint action. Internally, the painting action updates the value of regions in the data model, creating regions if they do not yet exist at the appropriate time.

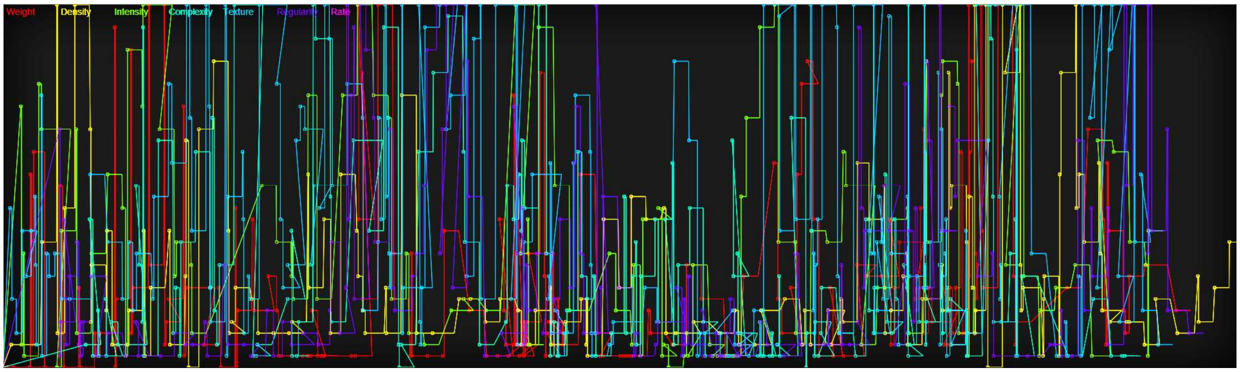


Figure 53: Parameter renderer
Expression parameters of the selected sequence are rendered as individual curves in an explicit-time view showing the interpolation of values between defined regions.

Since the data model is updated continuously during a paint stroke, the change is reflected immediately in all other views of the score.

Another explicit-time representation using the `ParameterRenderer` provides a different, more diagrammatic view of the same information as the `SketchRenderer`. In this representation, the ensemble of expressive parameters is drawn as independent curves overlaid with each other, sharing the same time and value domain. Each parameter's curve is given a unique color to visually distinguish it and a legend is presented. Unlike the `SketchRenderer`, this representation indicates the relationship of regions and their values. Single-valued regions appear as horizontal lines of constant value for the duration of the region with values between regions drawn according to the interpolation function set for the track. The start times of all regions are indicated with a small open square at a vertical position according to the region's value.

The interactions that are capable in this view are slightly different than those of the Sketch. As before, the user may activate the tool for the expressive parameter they wish to edit. Since this view can be rather cluttered and it would be difficult to mouse to a particular curve or value on a curve, this modal selection of the curve to be manipulated is useful. The user can then click and drag to move a region in time, reordering regions as necessary, or to change the value of a region (and relative values for multi-valued regions) by moving the region along the vertical axis. While selected through a drag operation, the region may also be removed by pressing the [`Delete`] or [`Backspace`] keys. Clicking on an interpolated part of a parameter's curve between regions, or at the beginning or end of the curve before or after any regions, creates a new region with the value of the interpolated curve at its start time. As in the Sketch interactions, the duration of the region created depends upon its proximity to other regions.

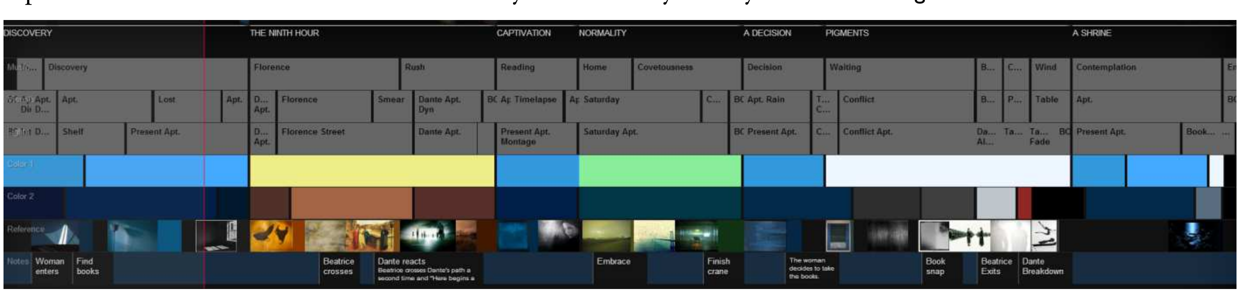
An additional subclass of `ScoreView2DExplicit`, `ScoreViewTracks` provides a special explicit-time view of score information that is the primary method for interacting with the properties of the Media Score other than expressive parameters, though expressive parameters may optionally appear in this view, as well. The track view is a composite of multiple explicit-time subviews that span the width of the viewport, and are stacked vertically. This part of the application most resembles traditional timeline-centric DAW and NLE user interfaces. Each track in the view uses a subclass of the `TrackRenderer`, itself a subclass of `ScoreRenderer`, specific for its track type to provide appropriate visual representations. However, `TrackRenderer` provides the common interface to track views that allows for the manipulation of regions in a track. Regions can be created by dragging to define a duration in a track. Double-clicking on an empty portion of a track will create a region that will occupy the space between the previous and the next regions. Once created, the left and right edges can be dragged to adjust the start and end times of the region. Dragging the region itself moves it in time along the length of the track. In all of these operations in the track view, depressing the [`Shift`] key while dragging will snap region boundaries to 10-second interval of sequence time or to the nearest region boundary if it falls within a 10-second interval of the mouse cursor position during the operation. Double-clicking on a region will raise a dialog window that allows the user to edit properties of the region including its name, timing, and value. Double-clicking on a region with the [`Control`|`Command`] key depressed will remove the region.

Figure 54: Edit region dialog

The track view of the score provides an easy way to add event regions, such as triggers for external control, notations, such as section titles or stage directions, and assets to a score. Parameter tracks can also be edited as curves rendered within each region of a track, though the interpolated values between regions are not shown as they are in the `ParameterRenderer`. Since each track is managed by a subtype of `ScoreRenderer`, in addition to being displayed in the multi-track `ScoreViewTracks`, each track can be rendered into its own view.

The track view also enables several drag-and-drop type interactions with the score, particularly for the handling of assets. Image, audio, video, and text assets may be dragged from the Media Scores web gallery interface, other websites, or the operating system's local file browser onto a track of the appropriate type in the track view. This action will produce a region with the asset as its value. If the asset did not originate from the Media Scores server repository, the asset is uploaded to the server for indexing. The asset is also processed by the analyzer attached to the track type, if one exists. When the asynchronous analysis operation completes, the region value is replaced with the asset and the results of analysis. The analysis may be

Figure 55: Track view



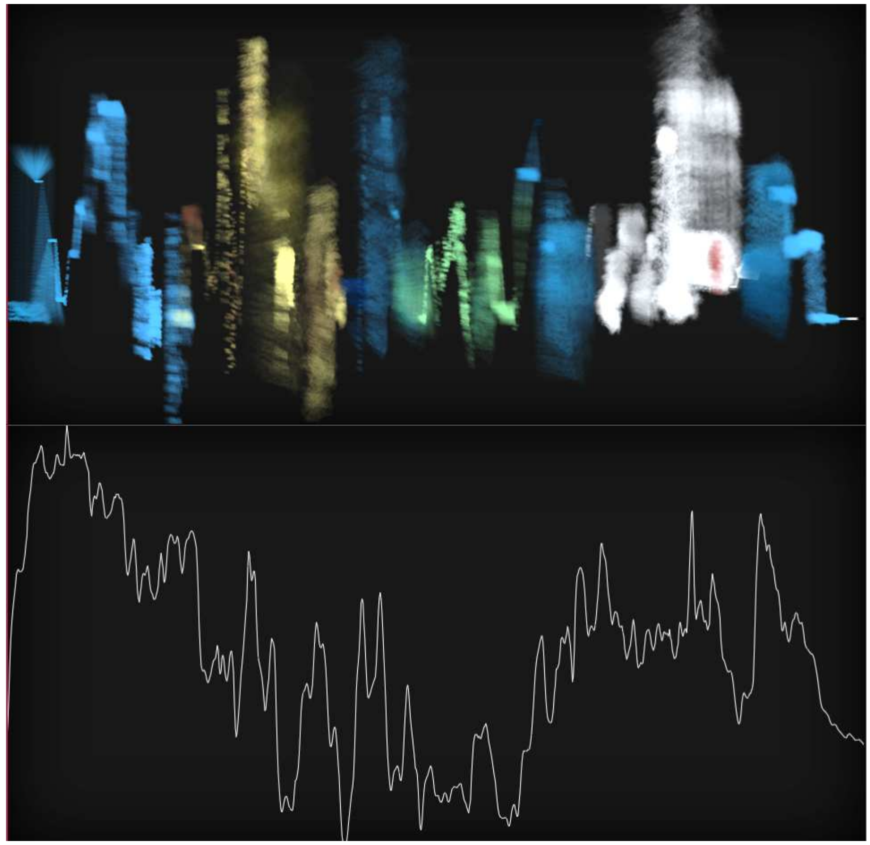


Figure 56: Tension plot
This non-interactive renderer (shown below the sketch it reflects, for comparison) depicts an analysis of the perceived tension produced by the formal parameters over the course of the experience of the work.

viewed for reference in the region's edit dialog or it can be used to set the value of other regions. For example, when adding an image asset, it is analyzed according to the ensemble of expressive parameters and by other metrics, including producing a summary of color information in the image. If the image is intended to be used as an inspiration for color, the composer or designer can then [`Control`]-drag the image region onto a region in a color track or an empty part of the color track to set the value or create a region with the image's most dominant color. If the image functions as a reference for mood or tone, then its region can be [`Control`]-dragged onto individual expression parameter tracks or other views to set the value of regions or create regions at the appropriate time in the sequence with the given expressive qualities. Through the inclusion of reference assets into Media Scores, this implementation not only provides tools for practices such as mood boards and image banks, but it also allows those resources to have a direct impact on the creation of the work. The process is always under the control of the composer or designer, though, meaning that a Media Score isn't simply a naïve translation and visualization of input.

While most score representations allow the user to interact with the score, some renderers may only make sense for visualization. One such explicit-time renderer depicts an analysis of the expressive parameters of the current sequence. This render produces a plot of the perceived tension over the course of the experience represented in the sequence. This continuous tension value is computed in the manner described in [34] and relies on sliding attentional and memory windows to account for consistency and variations in expectation over time. The original technique was developed to model perceived musical tension from a weighted ensemble of formal parameters, such as onset frequency, harmony, pitch height, tempo, rhythmic regularity, and dynamics. In this modified form, the ensemble of parameters used consists of the seven expression parameters of the Media Score. The tension plot provides Media Score composers with a key to the overall experience of the work, in addition to the renderings of the expressive parameters themselves. As noted in Section 2.2, variations in expectation are related to information content and, in turn, to perceived tension. Patterns of increasing and resolving tension contribute to the emotional impact of a work of art. Since the data for each parameter is aggregated in an accumulative and non-linear manner, it is nontrivial to produce a reasonable inverse of this curve. The values of the input parameters would be underconstrained by the tension curve and, thus, there are many possible solutions that would generate the same perceptual result. Therefore, the tension view only provides feedback about the expressive parameters in the score and cannot be edited to influence the score's parameters.

4.3.3.2 Implicit Views

Arguably, explicit projections of time are more easily reasoned about when considering multiple parameters of a work of art, rather than the overall impression or experience of the work. This is especially true for works that are assumed to be performed in time. However, one of the objectives of Media Scores is to not make assumptions about the time-dependency of a work of art. Another is to provide multiple representations of the work so that it can be considered by the composer and designer in different ways. To this end, views that rely on implicit and dynamic projections of time are also provided by the Media Scores composition environment.

Implicit views depict sequence information without a clear sense of time as a dimension along a single axis. As with explicit views, the entirety of the score is visible and can be seen in an instant. This view type is inspired by the expressive power of abstract and Synchromist styles of painting, though the quality of rendering need not be similar, as long as the expression parameters are conveyed. It is not particularly useful to represent event regions in implicit time views, though renderers could incorporate asset values into their representation, for example, creating a collage of reference imagery based on the analysis data for each image.

Looking at a Synchromist painting or a painting by Wassily Kandinsky, it is clear that these pieces have a structure to them. They convey a dynamism and orientation that is important in understanding the work. There are

Figure 57: Kandinsky renderer
An example of a painterly renderer in an implicit-time view

areas of common color and form that could function like scenes or sections of a time-based work. While time is not explicitly represented, there is an experiential path through the work that gives it structure. Time is *implicit, not absent*. Renderers of implicit views function by creating a mapping of the sequence's time to the spatial domain in order to achieve this kind of narrative consistency. This path may be curvilinear and circuitous throughout the visual field and its shape may take into account section boundaries notated in event tracks and some of the expression parameters. A ground for the visual field is rendered, upon which events are emitted along this spatiotemporal path. So, while there isn't an explicit, linear, orthogonal temporal dimension, a trajectory guiding the experience of the piece is created. Even if visual objects are presented sparsely along the path, Gestalt grouping principles suggest the temporal structure of the piece. The color of the emitted graphical or painterly elements can derive from the color asset tracks of the score. Other visual properties—formal parameters such as intensity, scale, texture and so on—are then free to reflect the sequence's expressive parameters.

Users may interact with implicit views in ways similar to how they interact with explicit views, such as the sketch view. The expression parameter brushes can be used to paint parameter values into the viewport. The view instance managing the viewport can then translate the spatial position of the mouse cursor into the time-value space used by the data model, projecting the temporal-narrative trajectory back into linear time. With that coordinate transformation complete, the values of regions for the active parameter can then be set and the viewport re-rendered to take into account the change.

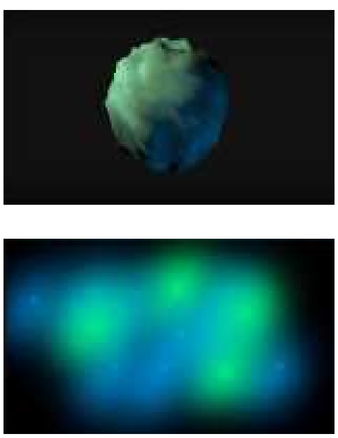


Figure 58: Dynamic views
The sculpture renderer generates an 3D object that deforms during playback (top)
The lumigraph renderer produces a dynamic image that resembles Fischinger's visual music apparatus (bottom)

4.3.3.3 Dynamic Views

Dynamic views do not show the entirety of a sequence at once. Instead, the rendering evolves in time as the sequence is being played. Dynamic views are, in effect, animations generated from the score. They provide an impression of the state of the score at a specific time and visualize the continuity in time of the changes of expressive parameters. Whereas implicit views can render scores in the form of abstract paintings, for example, dynamic views produce representations more like visual music.

For dynamic views with renderers that depict the expression parameters for the sequence, the parameter brush tools are used to manipulate the score, as before. However, the values being set occur at the current time of the score's playback. They may be painted as the sequence is playing or the current time indicator may be moved to a specific point while the score is not being played back. Painting would set the values at that time. Just because the value of a parameter is being set at a particular time, it does not mean that the value that parameter is unmodified at other times. Painting at a particular time sets the value of the current region or creates a region at the time and regions would likely have a duration greater than a single frame. Also, changing the value of a region would influence the interpolated values between adjacent regions for most interpolation functions that can be applied to a track. Thus, the painted value may affect the state of the score at times before or after the current time, as well.

Renderers for dynamic views are not intended to be the generated output for use during performance, but are just another way to explore and interact with the Media Score. However, equivalent visual content generators could potentially be implemented as renderers within the Media Scores application, in which case they could provide a preview of the output. For example, a subset of the capabilities of Render Designer developed originally for *Death and the Powers* could be rewritten as a Media Scores renderer. Additional visualizations of an output representation could potentially be implemented as renderers, such as a three-dimensional simulation of theatrical lighting, as an example. In general, however, connecting the Media Scores application to separate content generation applications is a more robust and feature-rich approach.

Dynamic views can serve useful functions with renderers that primarily rely on assets, as well. For example, the `ImageRenderer` creates a simple dynamic view directly associated with an image track. It displays the images for each region in time for the duration of each region. A smooth interpolation between image values results in a crossfade from the current image to the next. This is essentially a slide show of the image track, which can be useful for reference imagery or in previewing a sketch of a score with accompanying audio or music reference. When applied to an image track containing storyboard frames, the output of `ImageRenderer` is effectively an animatic accompanying the playback of the score and can be used to evaluate timing and pacing of the sequence. The related `VideoRenderer` behaves in a similar manner, being associated with a video track acting as a form of playlist list for video playback.

Views need not be added to the user interface when they are created. While renderers for explicit and implicit views only make sense to use for producing visual output to the screen, some dynamic-time renderers produce no visual output and, thus, their viewports do not need to be displayed. Like the `VideoRenderer`, an `AudioRenderer` treats the assets arranged in an audio track as a playlist for audible preview implemented using the browser's Web Audio API. Although not user-selectable as a view, the output of show control data and expression parameters from the visual score, explained in Section 4.3.5, is implemented as a non-visual dynamic-time renderer, since it relies on updates to the score's data model and the state of sequence playback.

4.3.4 Collaboration

Media Scores are intended to be collaborative documents in the sense that an entire team—single or multiple composers, directors, designers, performers and others—can reference a single vision of a Gesamtkunstwerk through representations that are intuitive and visually concise. The creative team can work individually on different aspects of the score using the views that are most appropriate for their needs. They can make annotations to share with each other remotely, and, in group settings such as meetings and rehearsals, have a common image and frame of reference to look at and point to.

Several features have been developed for the implementation of the Media Scores application to make it even more useful as a collaborative tool. Scores are generally stored in a custom server-side SQL database to facilitate collaboration and access to projects, regardless of location. Users may register and log in to a site on the Media Scores web server that hosts the application interface in order view their scores. A user's scores can be loaded from this database through a gallery-like interface into the composition interface for editing. The gallery displays a listing of scores created by the user or to which the user has been invited to collaborate, along with basic metadata about the score and a thumbnail image of the score. The database maintains additional metadata about each score, beyond what is contained within the JSON-encoded file itself, and provides simple version control of the score. Each saved commit to the server is recorded under the user that made the commit, which may be different than the composer of the score in collaborative use case. A timestamp, a unique identifier for that version, the ID of the project shared by all versions, and a thumbnail of the sketch view of the score, so that it can be visually distinguished in the web interface, are also recorded. The database also provides a basic form of access control, allowing score creators to grant read or read and write privileges for a score project to other users. This web-based sharing and collaboration model closely parallels the functionality of network storage services, such as Google Drive Documents [39], that are commonly used among teams to organize information.

Although real-time concurrent modification is not supported in the current implementation, in a manner similar to multiple users editing a single Google Drive document, a future update to the Media Scores application could make such operations possible. Since modifications of the data model already take the form of atomic transaction generated by interaction with a view, it is a relatively straightforward task to synchronize client data models for each modification by relaying the transaction to other client applications viewing the score through the web server. Updating the data models in response to a remote modification would function similarly to a local user's modification and views would automatically and continuously reflect changes as they occur.

In certain use cases, it may make sense for users to assume sole responsibility for some part of the Media Score. Sharing a score from the database provides user-level access control over whether the score may be edited or only viewed with a set of other users. More granular control over sequences, tracks, and regions is also available. A user with write access to the score may lock one or more of these structures. Each has a single lock associated with the user that requested it. The locked state of the object and the user owning the lock is stored within the serialized Media Score data model. Once locked, only that user is able to edit the object. This mechanism is handled at the user interface level and is not intended to be a secure method of enforcing access restrictions, but primarily an indication to users that some part of the score should not be modified and for preventing accidental changes while another user is working to achieve a particular effect.

4.3.5 Show Control and Playback

So far, we've seen how the Media Scores application functions during the creative process from design through content creation. By being able to communicate and interface with other systems and technological components in performance, Media Scores can be used as part of the production phase of a project, as well. Typically, pre-production processes are rather distinct from the production phase, particularly in the context of live performance. In visual effects production for film, assets—3D models for animatics, location plates, material textures—from the design process can be transferred to production systems for further refinement and incorporation into the final film. In other domains, the resources developed during ideation and design do not themselves translate into the final product. From an orchestral score, parts must be made and musicians must be assembled to rehearse and perform them. A scenic designer's watercolor sketches of the setting for a play must then be drafted, constructed, and painted onstage. A painter's sketch must eventually be recreated at scale on a canvas. Most importantly, the script or score guiding the production of such works of art must be interpreted through the machinery of the production itself.

Media Scores, outside of the trivial case where the score and the variety of representations within the application are deemed sufficient, does not replace these needs. From a Media Score for an opera, annotated by the director and design team, a set must still be constructed; costumes still need to be fitted and sewn. However, Media Scores can directly serve as the central conductor for performance technologies, whether they are discretely cued as a video switcher or lighting board might be, are animatronic machinery responding to continuous control values, or generative visual or audio systems. Complex systems can react to events in the score, triggering cues or modes, adapt in real-time to expressive parameters and live performance data, or draw on assets such as color and image defined in the score.

This is achieved by transmitting data from the Media Score to other systems during performance and supports the principle explained in Section 3.5. I will refer to the variety of other systems that rely on this control data as *content generators*. Content generators may simply be triggered by events and produce an output that has been programmed or may be controlled by the scores continuous parameters.

As the most straightforward form of show control, the values of Trigger tracks are typically output in order to cue states of connected systems. These triggers can set the state of real-time performance analysis systems, such as the Disembodied Performance mapping system (discussed in Section 3.1.3), allowing for variations in the way live performance or audience data is captured and processed, before it is modulated by the score itself. The triggers can also be used to cue output changes, such as taking a cue on a lighting board or video switcher or placing a generative content system, such as Render Designer (discussed in Section 3.1.4), into a particular rendering mode. An example of this control flow using Media Scores is given in Section 5.2.

Different systems have different conventions for triggering events. Common theatrical control systems, such as lighting boards, typically use a fixed-point number to identify cues, some may rely on alphanumeric names to identify states or modes, and many simply require a "GO" command to execute the next transition in a sequence of cued states. Media Scores can support these types of triggers by setting the appropriate value for regions in a trigger track. During performance, when the start of a trigger region is encountered as time progresses during playback, the trigger command with the appropriate value is emitted to the content generator.

Since the focus of a Media Score is on its expressive parameters, these are generally output to content generators for interpretation. Parameter values are output in the range of [0.0, 1.0] to be compatible with existing systems, such as those created for *Death and the Powers*, as described in Section 3.1. Values in this normalized range can then be easily scaled as needed by content generators. A bundle of values for output-enabled parameter tracks is transmitted at each time step during playback or when the current time within the sequence is otherwise updated. The values of output-enabled tracks are also sent when the track contents are modified. This provides real-time control and feedback with content generators online during editing, as well as performance.

Since the expression parameters in a Media Score do not represent formal parameters of output, the process of generating content from the expression parameters is underconstrained. If the output of a content generator was governed solely by the expression parameters, the result would be at best a diagrammatic representation of the parameters' values, not an aesthetically communicative production. A grammar is required to create meaningful utterances given a particular style, otherwise a realization may fit the parameters and intended expressive contours, but lack thematic or stylistic continuity. Humans do this in the design process realizing a production from a Media Score. As with view renderers, content generators must also encode this grammar in their implementation. In the case of renderers, the style and grammar is hard-coded into the implementation of the renderer. Content generators, such as Render Designer, expose tools and options that allow users to build up the visual language that can then be influenced by the expression parameters.

During playback, the Media Score application will also by default transmit information about the current time along with transport control operations, namely when playback is started or resumed and when it is suspended. Time is represented in two formats for the convenience of connected systems: the absolute sequence time in floating point seconds and the fraction of the sequence elapsed as a floating point number from 0.0 at the start of the sequence to 1.0 at the end. The linear representation of playback time from the score need not be interpreted literally, as it would for playing back audio or video of fixed duration. The time track can influence how time flows in a discontinuous fashion. The rate of playback may be adjusted using the interface or through remote commands. This playback rate affects how Media Scores interprets the progression through time, but does not influence the frequency at which events are emitted, as the Rate expression parameter does. The playback rate can be 0.0, effectively holding time frozen, but the system will still be playing back, producing output values and able to affect and respond to live performance data.

Changes of the active sequence are reported so connected systems can configure themselves for a different mode of operation, if necessary. The Media Scores application can also respond to external messages to change sequences, start and pause playback, advance the current time to the next sectioning region, and set the playback time to a proportional or absolute value, either directly or in response to a named trigger. This remote control over Media Scores allows it to slave to other show control applications, if desired, or remote control interfaces. For example, a TouchOSC interface could be used on a mobile device to wirelessly control Media Scores playback [42].

Output to remote systems can be enabled or disabled in Media Scores using a toolbar command or keyboard shortcut. When disabled, Media Scores functions in an offline mode and will not receive control input, as well. In online mode, sequence and playback state is sent along with values for all output-enabled tracks. Since Media Scores is a browser-based web application, all incoming and outgoing communications are implemented using JSON over WebSockets. To communicate with content generators and other systems, the JSON messages must be sent and received using the WebSocket protocol and translated to and from Open Sound Control (OSC) or other protocols. A companion software application written in Java handles this translation and can be configured to route messages to OSC and MIDI Show Control (MSC). On private local networks, this translation step introduces a negligible latency of less than 12 milliseconds, maintaining real-time performance. A new version of the Disembodied Performance mapping system also accepts the WebSocket messages and can be used to format messages using OSC, MIDI, MSC, and Art-Net protocols, not just OSC as the original version for *Death and the Powers* supported. The mapping system is ideal for analyzing incoming live data during a performance from performers or audiences and modulating the output by the expression parameters and any specified envelope constraints on the live input. Media Scores follows the show control message and formatting conventions of the OSCSC protocol developed during *Death and the Powers* (see Section 3.5.2), so the reformatting of JSON messages as OSC is a trivial implementation. Future versions of the Media Scores application may be modified to run as a privileged application—Google Chrome Packaged Apps or Mozilla Open Web Apps—and consequently have access to TCP/UDP sockets for direct OSC output and hardware serial devices without the need for a WebSocket bridge application.

Some applications may not require connected systems for Media Scores playback. Independent devices capable of interpreting the score and generating content need not be networked or have a star topology control architecture. In such applications, the entire JSON score file itself could be loaded onto the device. Imagine, for example, a self-contained electronic device, say a luminous orb, capable of producing abstract colored lighting: lumia. The score could be loaded onto the device and the device itself could then perform it independently of any other systems. Sensors in the device could also respond, in accordance with the score, to user interaction. Appliance versions of Media Scores interpreters could also be implemented in this fashion, when a rich user interface or web front end is not required.

5 APPLICATIONS OF MEDIA SCORES

The stage is not so much a sequence as a tangle of diverse intentions. A threading, mirroring, echoing, space. A dramaturgy of knots, collision, tangles.

—Tim Etchells

Having defined what a Media Score is and how the Media Scores application functions, we will now turn our attention to projects that have leveraged Media Scores in their creation. These projects represent the beginning of what is possible with the Media Scores application and serve as concrete examples of some possible use cases. The flexibility of Media Scores suggests a broad range of other possible use cases and we will briefly look at some.

5.1 A Toronto Symphony

In Section 3.4 I introduced *A Toronto Symphony: Concerto for Composer and City* by Tod Machover and the citizens of Toronto. We looked at the systems that were used to produce the premiere concert. Although Media Scores was not used as part of the control architecture for the live performance, it did play an important role in the year-long development of the work. Media Scores was key to the massive collaboration required to create what Maestro Peter Oundjian described as, "the most collaborative piece of music that has ever been written" [13].

Machover wanted to redefine the public's role in the creation of such a piece beyond recent attempts to connect the public with composition and orchestras, leveraging the ubiquity of newer web-based technologies. Existing models for engaging the public in musical co-creation, particularly in distributed technologically-mediated ways, follow two different paradigms. The *remixing model* encourages music consumers to make derivative works from the musician's source material. This type of activity tends to occur even when not invited through the ready access to musical recordings and tools for audio manipulation. Online forums for posting media are replete with amateur performances of popular music and remixed works. This has been the subject of some consternation among artists. Others, however, embrace this by releasing materials or tools for the public to interact and create derivative works. For example, in 2008 the popular music group Nine Inch Nails released an album entitled *Ghosts I–IV* under a Creative Commons license to encourage music consumers to actively engage in manipulating and repurposing the music for non-commercial uses [54]. An edition of the album was distributed with source audio stems for each track, thereby providing the raw materials for remixes or alternate mixes of the album to be created. In 2011, Icelandic artist Björk released her album *Biophilia* as a series of interactive Apple iPad applications, as well as traditional compact disc release, installations, and live performance versions [93]. The iPad applications allowed users to play back the album's music and interact with it in limited ways by creating sequences with timbres from a track or through game-like behaviors that play back the track's music. Sections are reordered in response to user interaction. Each activity features different visualizations of the music and each track's theme. The applications function more as specialized musical instruments tailored to the performance of the album's pieces. In both of these examples, the public does not have any creative control in the genesis of the musical content. Rather, the artists maintain creative control over the range of possibilities, leaving the consumer with a limited, but still creative, opportunity to reorganize and potentially reconceptualize the existing material.

Another type of interaction that leverages the capabilities of the World Wide Web is the *crowdsourcing model*. As contrasted to the remixing model where consumers and audiences manipulate existing works, the crowdsourcing model encourages participants to engage in the creation of the work, often by contributing sounds or ideas. A notable example of this is cinema composer Hans Zimmer's score for the 2012 film *The Dark Knight Rises* for which he requested audio recordings of fans' voices to be submitted online for inclusion into the score as a chant theme [85]. Zimmer's goal here was to create an impressive mass of voices, though typical studio production techniques could have accomplished this, as well. This project also served as a form of viral marketing and outreach by engaging Batman fans with the film prior to its release and by giving them a sense of participation, having contributed and, in some ways, being present in the film. However, with thousands of submissions and the collaging of voices, it's unclear that audience members would be able to discern their individual contributions. Other examples of crowdsourcing music make individual contributions more evident. An example of this is the single "Know Your Exit" by Robert Morris and Vivian Darkbloom [89]. For this project, a request was published for online submissions of short snippets of recordings of the singing voice and hand clapping. Again, like Zimmer's score, these audio fragments were mixed together by the musicians into an existing song. However, presented through an online interface, the playback of the piece attributes the contributors and their locations on a virtual map of the world. The crowdsourcing paradigm, as used in these examples, affords even less creativity to participants, though their contributions do contribute directly to the final work.

Previous projects in the Opera of the Future group have given the public a more active creative role. The *Brain Opera*, which premiered in 1996, incorporated musical material from audiences in the performance space and the influence from online audiences into live performance. In 2002 and 2003, the *Toy Symphony* project toured the globe working with symphony orchestras to perform music composed for children and in performance with children onstage playing specially designed instruments. Children in each city were also able to compose pieces for the orchestra using the software program Hyperscore [64]. For *A Toronto Symphony*, we sought to extend the way in which crowdsourcing can be used beyond merely the contribution of necessary material, but to be essential to the creative development of a musical work.

Machover envisioned using web-based technologies such as social media and innovative web applications to not simply crowdsource material that would be buried within an independent musical experience, but to actually engage the public in key aspects of the piece's development. He defined three types of interaction that would be used in the project and would generally be associated with different sections of the piece: Yours, Mine, (H)ours. In the Yours sections, Machover would invite the public to submit materials in responses to different requests. The content for these sections would be created primarily from the contributions and interwoven with each other during the composition process. Within the rough constraints of the overall shape of the piece, each of the Yours sections would take on a form based on the Torontonians interactions and ideas. The Mine category would consist of musical material composed mostly by Machover, influenced by contributions, but not necessarily referencing them directly. While these two categories were not intended to draw a complete distinction between composer and city in isolation of each other, a third category, (H)ours, would represent a true compositional dialogue between Machover and participants. In hour-long sessions either in Toronto or held live online with the public or select community and music groups, material for the final symphony would be composed by interacting and actively collaborating with Torontonians. This structure allowed Machover to be responsible for the overall shape and the final form of the piece while engaging in a massive collaboration to create it. Thousands could contribute asynchronously in the Yours method while the (H)ours sections would rely on tens or hundreds working together.

5.1.1 A Graphical Media Score

At the outset of the project, Machover defined the named sections of the piece that gave structure to the story of the work. He envisioned the piece as a documentation of the process of its creation in the form of a story about an outsider (himself a non-Torontonian) visiting the city, becoming acquainted with its sounds, stepping back to reflect on his impressions of the city and its inherent music, and finally bringing all of what had come before together in a cohesive, celebratory fashion, constructing a musical image of the city from perspectives both within and without. Machover also imagined the quality of each section and the overall shape of the piece.

With this information, I created a Media Score sketch of *A Toronto Symphony* that would visually indicate all of these aspects of the piece (Figure 59). This explicit-time score sketch would be a starting point for Machover's collaboration with Torontonians, giving them an impression of what they were working together to create long before any musical material existed. It served as a scaffold for the assembly of the multitude of different types of inputs and could be populated with contributions as they were added. As Torontonians submitted material and Machover wove it together, the sketch would evolve into a visualization of the final form of the symphony.

Figure 59: Original Media Score sketch of *A Toronto Symphony*
This sketch illustrated Tod Machover's conception of the overall work in advance of the collaboration and composition phases. It defines the thematic sections of the piece with relative timing and how user material would be contributed in a variety of ways. (opposite top)

Figure 60: Final Media Score sketch of *A Toronto Symphony*
This sketch reflects the final form of the piece. Differences from the original sketch include section order, section timing, and overall contour. Still, it shows how the original form influenced the structure of the final work. (opposite bottom)

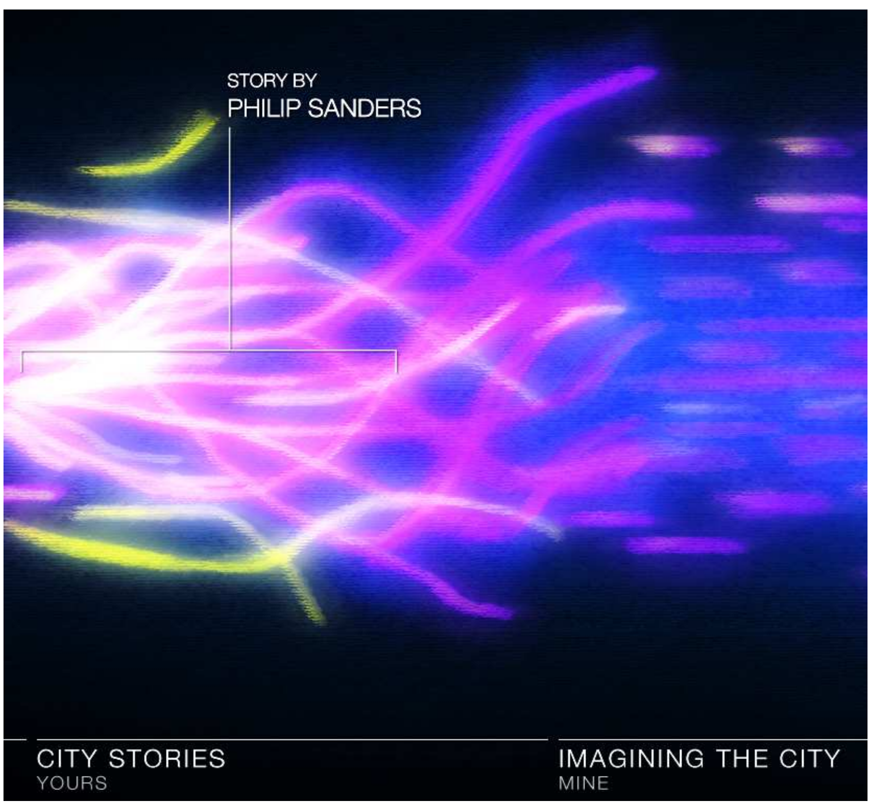
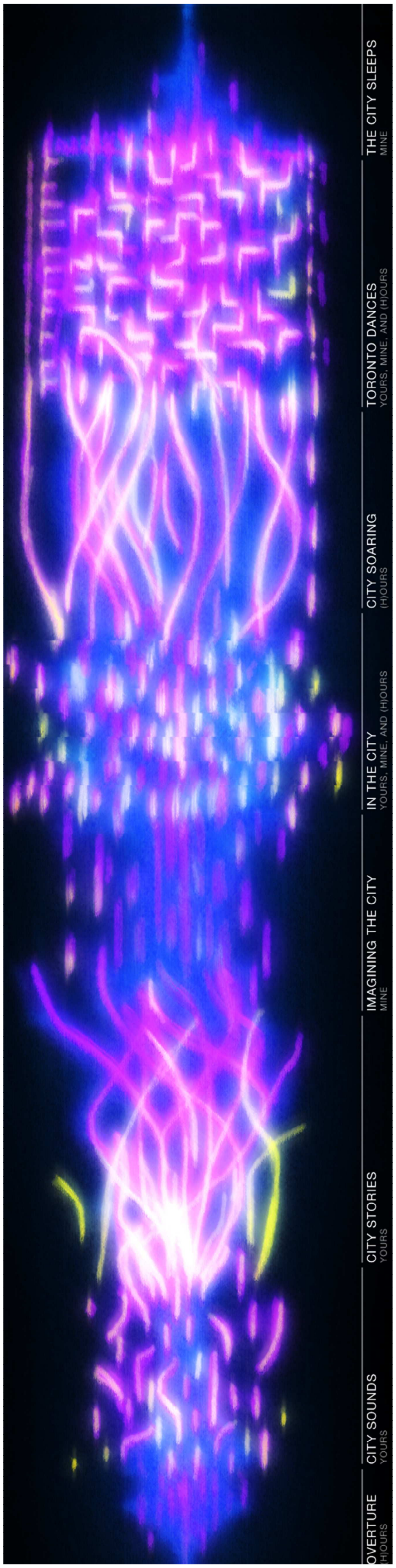
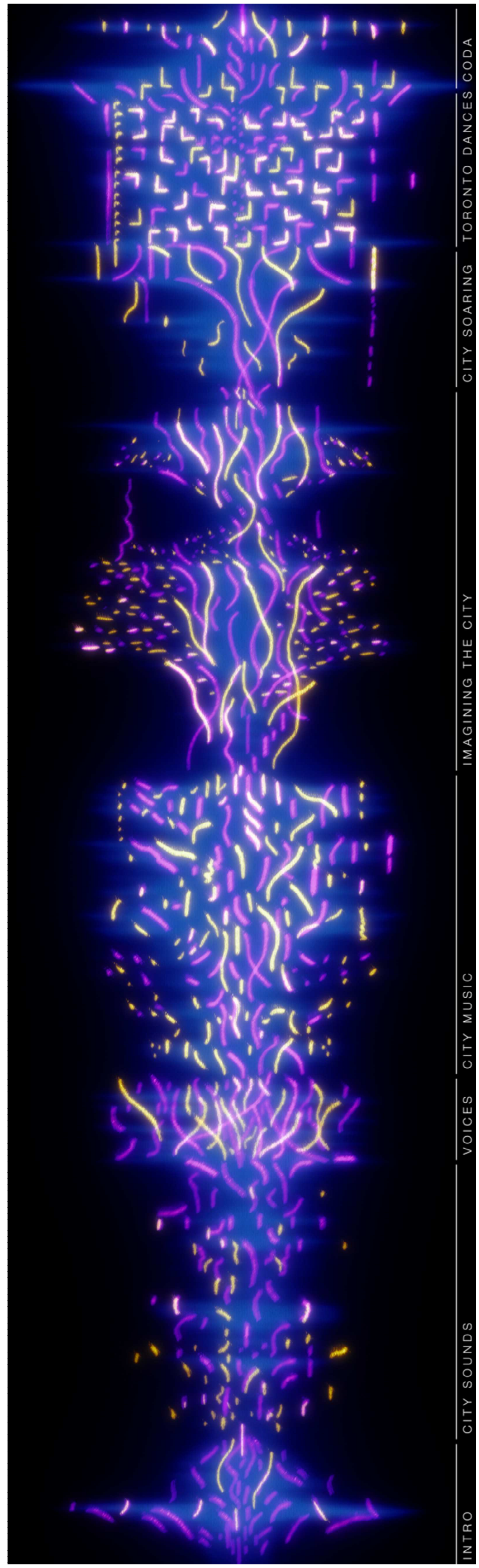


Figure 61: Detail of *A Toronto Symphony* Media Score
Section name and mode of collaboration are shown along with attributions of contributions.


OVERTURE
(H)OURS
CITY SOUNDS
YOURS
CITY STORIES
YOURS
IMAGINING THE CITY
MINE
IN THE CITY
YOURS, MINE, AND (H)OURS
CITY SOARING
(H)OURS
TORONTO DANCES
YOURS, MINE, AND (H)OURS
THE CITY SLEEPS
MINE


INTRO
CITY SOUNDS
VOICES
CITY MUSIC
IMAGINING THE CITY
CITY SOARING
TORONTO DANCES
CODA

The visual sketch depicts each planned section for the piece along with its name. It also indicates the primary mode of collaboration—of the Yours, Mine, and (H)ours varieties—for each section. The shape of the sketch was rendered to reflect the intensity contours within the piece. A blue bed frames the expressive qualities of the score at any given point. Atop this structure are purple and yellow strokes that intertwine and further communicate the expressive quality of the piece. These are parts (see Section 4.3.2), but not necessarily in the musical sense of material being assigned to a particular instrument group. Their deviation from the central axis running through the sketch suggests the continuously-varying salience of each part. Purple strokes represent musical elements composed or created by Machover: Mine material. Yellow strokes represent contributions from Torontonians: Yours material.

The original Media Score was speculative. Strokes represented expected contributions or representative qualities. This sketch evolved as Torontonians participated and as the piece was assembled through activities and the development of the orchestral composition. The final form of the Media Score bears an expected resemblance to the original, though it is not identical (Figure 60). Machover reordered and renamed some sections as the piece came together. The relative timings of the sections were adjusted to reflect the actual length of material in each. The overall dynamics of the piece were also refined. Comparing the final sketch with the original, the symphony has greater shape and variation to it. However, the quality that Machover originally envisioned for each section was generally preserved.

Additional Media Scores were created at various points during the project. These were generally incorporated into special presentations and (H)ours events in an illustrative capacity. They reflected a portion of the work: a section or an experimental example. These smaller scores were either generated from input content or annotated to illustrate the structure and organization of the example.

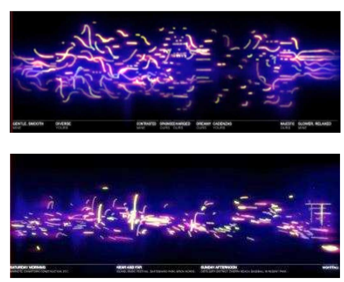


Figure 62: Additional Toronto Media Scores
Launch Music (top)
Sound samples and cello collage for a Glenn Gould event (bottom)

The visual presentation projected during the premiere of *A Toronto Symphony* in Roy Thomson Hall made extensive use of the original and final versions of the Media Score. Intended to document the creation of the piece, the transition from one to the other over the course of the performance was appropriate. The explicit-time representation also gave feedback to the audience about the structure of the piece and context. Between sections, the score sketch was often shown close up, with the current position in time with the performance indicated on the scrolling sketch to visually recap the concluding section and introduce the next. Additionally, the sketch aesthetic and color palette were used as visual motifs throughout the imagery, not only when the Media Score itself was displayed.

Figure 63: CN Tower lighting
Lighting on the CN Tower was triggered from the live performance of *A Toronto Symphony* and represented a visual impression of the music.

As mentioned in Chapter 3, the premiere performance of *A Toronto Symphony* was accompanied by a synchronized light show on the CN Tower. The visuals content for display on the Tower had to take on a very different form from that of the visuals projected in the hall and streamed online. It's unique form factor and extremely low horizontal resolution precluded the use of most figurative imagery. Instead, color and dynamics were the primary forms of expression. Like the Media Score sketch, the Tower visuals had to feel like the music itself; a quintessential example of visual music not unlike the light shows of Louis-Bertrand Castel's Clavecin Oculaire or the abstract animations of Oskar Fischinger (see Section 2.7). The visual accompaniment for the CN Tower is essentially a dynamic-time rendering of the Media Score. It is a means of representing the music, its structure, and expressivity in a different medium that could both standalone and accompany the live web streaming broadcast of the concert. I animated it in Adobe After Effects as a series of video triggers, due to the technical constraints of the installed lighting control system available at the time. Such a visual representation could have easily been derived from the Media Score and its real-time playback, implemented as a procedural content generator. With this approach, accents and variation from the live orchestra could have been incorporated in the same way that the live web applications projected in the hall.

5.1.2 Outreach and Events

At the start of the project, we envisioned that the majority of the collaboration between Machover and the city would take place through Internet channels. A blog and website were launched to offer to the public avenues of engagement, including a series of "musical journeys" that would invite the initial contributions to the symphony. Requests were posted for the public to submit material such as personal or family stories and reactions to life in Toronto, favorite or readily identifiable sounds of the city, music that feels connected to the city, and so on. Submissions for these Yours-type activities could be made using text or media files through the website or by linking to media posted on other sites, such as YouTube and SoundCloud. The use of social media was encouraged early on as a form of viral promotion for the project. The hope was to reach large numbers of people and diverse communities by having contributors' submissions posted to their Facebook pages, Twitter accounts, and the like, so that their social networks would become aware of *A Toronto Symphony* and the activities available to the public. Machover would respond to some submissions and react to contributions on the blog. The submitted assets were catalogued for later use in the piece or as seed material for future activities. On the website, Machover would also contribute musical ideas he was developing and offer the public the opportunity to comment on them, as well as to experiment with them and post their variations and elaborations.

One of the first (H)ours sessions was a live improvisation demonstrating how participants could influence aspects of the composition. Pianist Tae Kim performed at the MIT Media Lab while being streamed live to the project website. On the website, users could continuously rate the qualities and styles of the music they would like to hear. The data from all users was aggregated and graphically presented to Kim to guide his improvisation.

To lay out a shape for the symphony, Machover composed a series of chords that would inform the harmonic structure of the piece. The chord progression was released on the website and the public was invited to play them, experiment with them, and compose melodic material using them. The chords were also sent to musicians from the Toronto Symphony Orchestra, who were encouraged to compose material for their own instruments around the chords. Machover assembled and combined the responses from the musicians into a short piece that was then performed by the eight members of the TSO at the ideacity conference to formally launch the project in June of 2012 [46]. The resulting Launch Music featured improvisatory sections for the musicians and provided another example of the collaborative model for the larger piece. The Launch Music also provided seed material for the Constellation web application mentioned below and was transformed into portions of the final work.

While a significant portion of the collaboration for *A Toronto Symphony* was completed through online activities, the project ended up relying on more sessions and events in Toronto than originally anticipated. A variety of community outreach events and (H)ours sessions were held during the year-long development of the piece, ranging from interviews with Torontonians and vocal workshops with young children to involvement with music festivals around the city and a variety of public presentations, many of which resulted in additional content and directions for inclusion in the symphony.

We collaborated with the Toronto District School Board (TDSB) to engage schoolchildren in grades 3 through 6 in the greater Toronto area to contribute to *A Toronto Symphony* as part of their studies. This aspect of the project began by introducing teachers in the TDSB to the Hyperscore composition software (see description in Section 2.9). The teachers then developed a curriculum around the use of graphical scores and musical storytelling for use in classes throughout the TDSB. At the conclusion of the program, we held a small concert with Tod Machover and an ensemble from the Toronto Symphony Youth Orchestra, with over 500 students from schools in the TDSB who had submitted Hyperscore pieces about Toronto. A selection of pieces was performed at the event and those pieces were later incorporated into the City Music section of the symphony.

5.1.3 Web Applications

Three web applications were created to facilitate a compositional dialog between Tod Machover and participants from Toronto. Although not synchronous sessions, by allowing participants to interact with Machover's material, they function as part of the (H)ours category. Each provides material that users can explore, shape, and listen to the effects of their changes. When a user is satisfied with their work, they can submit it to an online gallery. Visitors to the website can browse the gallery, listen to others' contributions, comment and share them as links or through social media websites, and open them to modify or remix their peers' projects. Each submission is stored with their user name and a reference to the version that it remixes, if not an original interpretation. Given the visual nature of each of the web applications, a thumbnail image of each is also presented in the web-based gallery view so that users can quickly get a sense of what it will sound like based on how it looks.

The two web applications I created for *A Toronto Symphony* focused on allowing users to orchestrate accompaniment and embellish material composed by Machover for the final two sections of the symphony: City Soaring and Toronto Dances. These applications were implemented as subsets of the Media Scores application described in Section 4.3 and will be covered below.

The third app, Constellation, was created by my colleague Akito van Troyer and served more of a sequencing role, allowing users to compose with sampled sounds of the city and sampled musical fragments [71]. Constellation was not based directly on the Media Scores implementation, so I will not discuss it in depth. However, it provides a useful model for potential extensions to Media Scores interaction by adding an intricate sequencing component. In Constellation, the user is presented with a field of dots each representing a grain of sound. The user can drag over the dots to hear a sound or create a sequence in time by drawing a line through the dot field. Playback of the Constellation follows the line, producing the same trajectory through sound that originally created the line. Currently, sequencing within the Media Scores application involves moving regions in explicit-time views, such as the track view. Future renderers could take advantage of this drawing model to organize sequences or, by allowing branches to be drawn, creating more complex, non-linear sequences.

Toronto Dances

The web application for Toronto Dances, the finale of *A Toronto Symphony*, featured two types of interaction with pre-composed musical material [72]. A core melody composed by Machover is always present and is visually represented by a white line in the interface. The line essentially plots source time on the vertical axis against playback time on the horizontal access. The line may be bent and distorted by the user, thereby altering the playback of

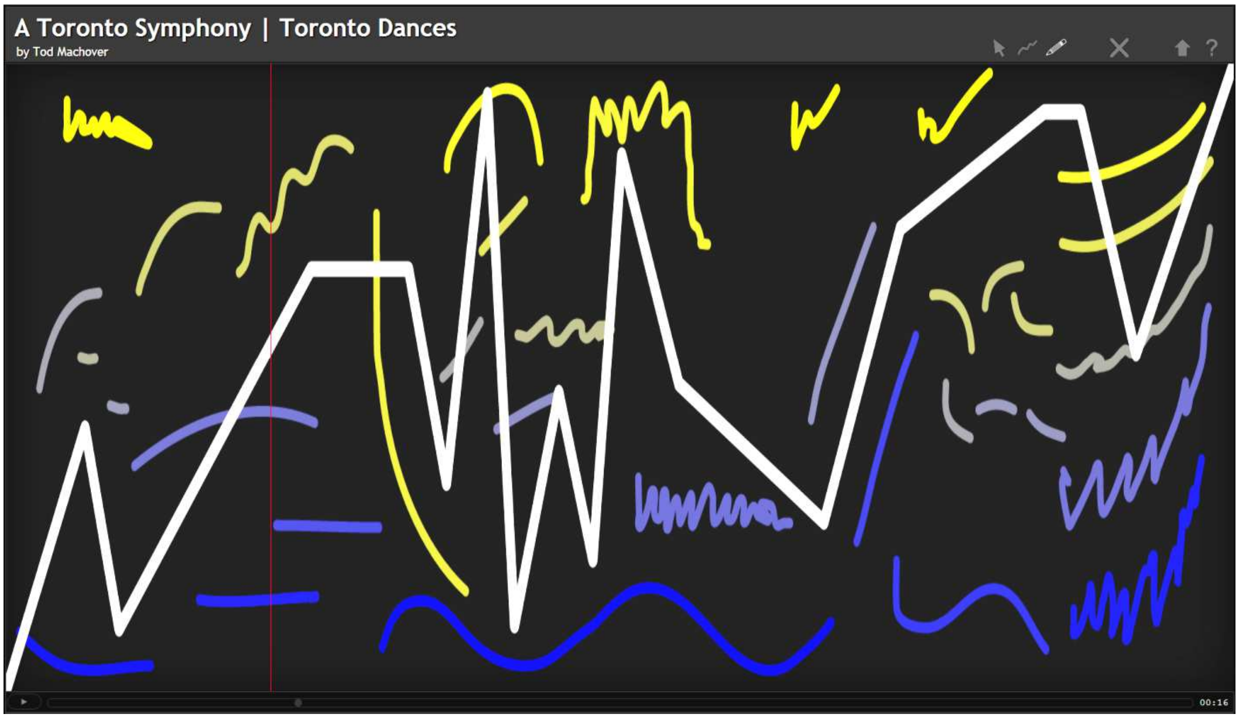


Figure 64: Toronto Dances web application

the melody. A segment of the line increasing at a 45° angle (a 1:1 slope) will play Machover's melody as written. A steeper slope progresses through time more quickly. A horizontal segment will repeat a portion of the melody. A decreasing slope will move toward earlier material. Using this line, the melody can be re-sequenced and extended in time with considerable variation. Discontinuities in the line permit reordering of source material. Machover intended this section of the symphony to be dance-like with a well-defined rhythm, having an irregular progression, thus this remapping of time allowed contributors and Machover himself to explore possibilities.

The manipulation of time is not continuous, as this would produce distortions in pitch and timbre. It is different from changes in playback rate or tempo, as can often be found in audio editing software. Instead, Machover segmented his original melody into chunks of varying lengths, each ranging from one to several beats in length, that reflected the phrasing of the music. Changes in time happen at these chunk boundaries, so that a chunk of time is always played in its entirety at tempo and multiple chunks do not overlap.

In addition to shaping how the melody of Toronto Dances unfolds, users can add accompaniment to the melody in a variety of instruments by drawing additional lines onto the sketch. The position and shape of the lines affects the color of the line and the quality and timbre of the musical material added. The vertical position selects the pitch register of the accompaniment and the curvature of the drawn line influences the complexity of the embellishment. As these additional parts were composed to accompany the melodic line, they are also segmented into chunks, so that they remain in synchronicity with the user's perturbations of time.

City Soaring

The penultimate section of *A Toronto Symphony*, City Soaring, is intended by Machover to assemble the ideas presented in the earlier portions of the work into the symphony's main theme [73]. Users were invited to contribute to the final form of the section by orchestrating Machover's theme using another web application variant of Media Scores. Like Toronto Dances, the user is at first presented with a line representing Machover's unadorned melody. In this case, the vertical displacement of the line is fixed to reflect the pitch contour of the melodic material. The user can then paint over this line with four paintbrushes representing different expressive qualities of orchestration and accompaniment. The four parameters that can be painted are:

WEIGHT	instrumental density
COMPLEXITY	accompanying melodic variation, timbral variation
TEXTURE	timbral selection
INTENSITY	overall amplitude and attack

Each parameter alters the appearance of the line in a visually analogous manner to the changes in the resulting sound. The painterly representation using the sketch renderer provides a concise and intuitive view of how the abstract expressive parameters influence the orchestration. An additional explicit-time view of the function curves using the parameter renderer is available for more precise editing, but is also more cumbersome to

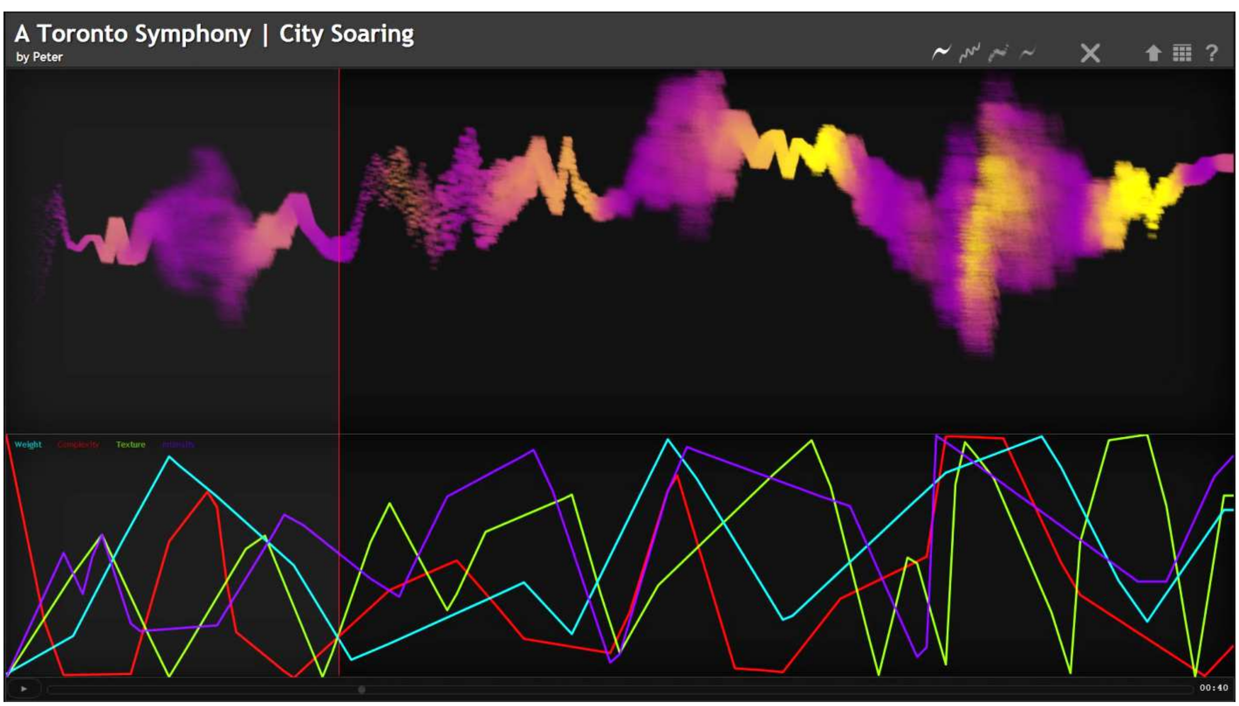


Figure 65: City Soaring web application

manipulate. See Section 4.3.3.1 for descriptions of these renderers.

Both the Toronto Dances and City Soaring applications were implemented as a subset of the full Media Scores web application. The applications allowed users to play back immediately the Media Scores to hear the results of their manipulations. In both cases, the audio production was implemented as a dynamic-time renderer within the Media Scores infrastructure, relying on the Web Audio API for scheduling, mixing, and playback. Upon loading, each application asynchronously downloaded a set of audio files necessary to produce the variety of effects. No audio is synthesized by the application itself. An additional table of "chunk" times is loaded, as well. This was used to schedule the playback of distinct regions of time from the audio file so that certain types of changes occurred at musically logical times. Audio files interpreted as chunks had silent tails inserted after each chunk to prevent audible discontinuities and truncation of note decays when switching chunks discontinuously.

5.2 Figments

Figments is a short theatrical work commissioned for the MIT Media Lab's inaugural The Other Festival of art and design in 2013. The piece is inspired by a number of source texts, most substantially portions of Dante Alighieri's 1295 prosimetrum *La Vita Nuova* [3]. Other texts that serve as inspiration for *Figments* include writing by Peter Carey, Karl Iagnemma, William Shakespeare, Maira Kalman, François-Marie Arouet, George Gordon Byron, Alexis Felix Arvers, Graham Davies, and personal writings. The production evokes a form of promenade theater in a multimedia work that can function both as an installation through which audiences may move and that can be activated for performance at a specific time. In both the acted and installation versions, the representation is intentionally left abstract and open to interpretation.

In creating and directing *Figments*, I set out to use Media Scores to coordinate the design of multiple modalities and use it as show control for various outputs during performance. I also wanted to use the Media Score as part of the direction of the human performers in the piece. The performance mode features one male performer and two female performers and no spoken dialog. In the installation mode, the abstracted story of *Figments* is told solely by the lighting, visuals, and sound elements. When performers enter the installation space, the Media Score reconfigures the output to incorporate the performers into the telling of the story, providing necessary illumination and visual context for their choreography, which is itself dictated in advance by the Media Score for the piece. Within the Media Score itself, the two presentation modes are stored as different sequences (see Section 4.3.1).

Figure 66: Photographs from *Figments*
(Photos by Andy Ryan)

The play area for this production of *Figments* is a circular space 20′ in diameter situated in a black box-type theatrical venue. Situated in the center is a small, square table with a white surface. To one side of the table is an understated chair. The circumference of this area is a porous ring of 12 Barco MiSTRIP LED display units rising vertically from custom-built stands. Between two of the LED strips stands a bookcase supporting numerous handwritten notebooks. (In reality, the contents of the handwritten notebooks are fragments of the aforementioned source texts.) The LED strips face inward, toward the center of the space, and provide both a horizontally sparse display surface for visuals generated in response to the Media Score playback. At times, the LED strips act as a form of lighting for the performers in the interior. This functions a bit like image-based lighting techniques, in which naturalistic lighting can be recreated in the space [27]. Additional conventional and intelligent theatrical lighting overhead is required to properly illuminate the scene and is also under the control of the Media Score playback. Discretely placed within the space, surrounding the audience are several speakers that produce the musical accompaniment for the experience.

The action of the final performance version of the piece begins with a woman entering an apartment. She is possibly a real estate agent or social worker arriving for business purposes. Finding the apartment uninhabited, she soon discovers a shelf laden with handwritten notebooks. She takes a few moments to begin reading one or two before realizing that her responsibilities require her to be elsewhere. The scene transitions back in time, possibly to 13th century Florence, where a Dante-like character enters, sits, and begins writing poetry or journal entries in his notebook. Momentarily, he rises, exits his apartment, and soon encounters a Beatrice-like character that so affects him that he promptly returns to furiously document his thoughts in words. The scenario continues alternating between the present, as the woman returns to the apartment to continue reading the notebooks, and two additional vignettes with the Dante- and Beatrice-like roles, though possibly portraying other pairs hundreds of years apart. As the present-day woman becomes more engrossed in the writings, she decides that they need to be shared with the world and prepares to bring them to a publisher. However, eventually she realizes that these texts were never meant to be read, and returns them to their shelves in the abandoned apartment as she takes her leave of the space and its literary diversions one final time.

The process of composing the Media Score for *Figments* followed aspects of my typical practice for designing lighting or visuals for a performance, though streamlined by use of the Media Score application. For *Figments*, I was also developing story points, the action of the piece, and the musical accompaniment as part of this process. The piece began with a rough concept and ideas for the source texts. I entered specific lines from source

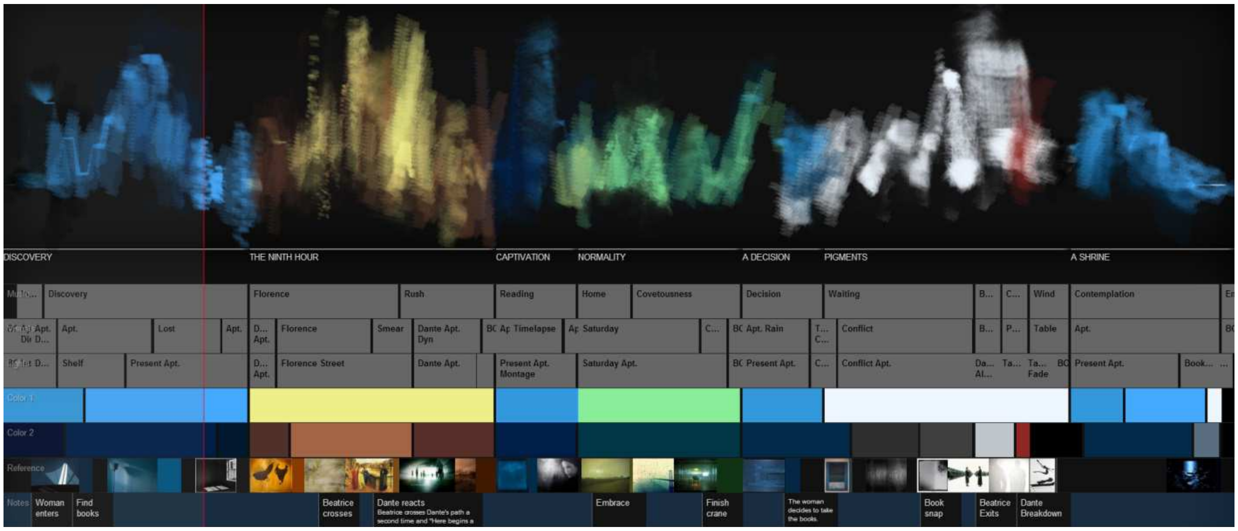


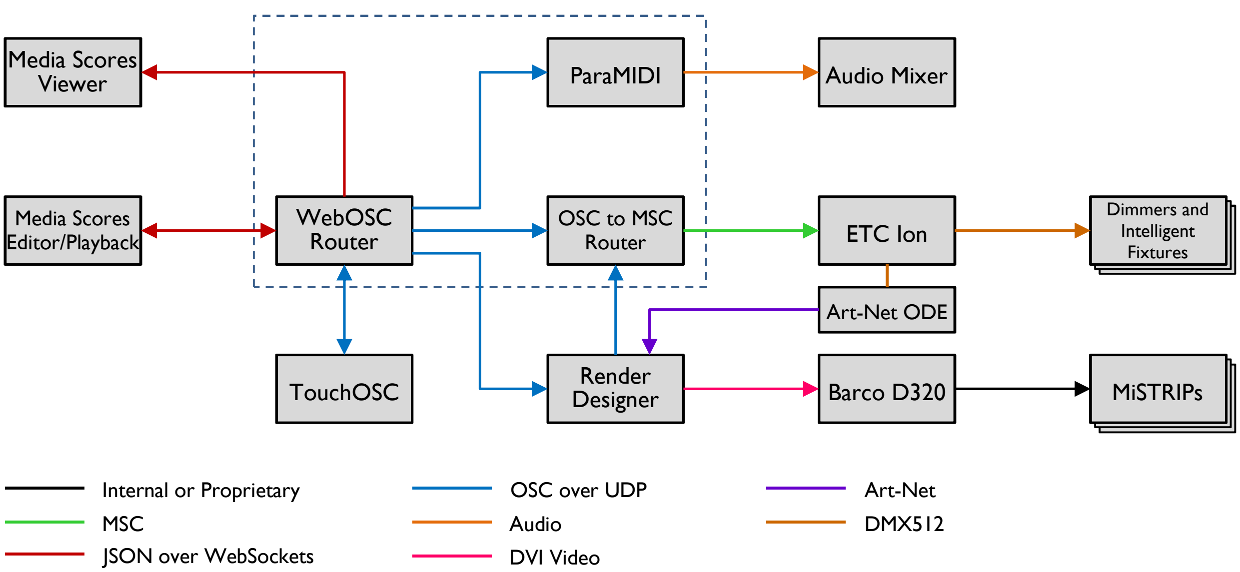
Figure 67: *Figments* Media Score

texts or moments of action, even before I had fully developed the context for the actions, into the note track of a new Media Score sequence. The duration of the sequence was set to the approximate time I wanted the performance piece to last. To this, I began an image bank for the production, adding visual references to an image track in the sequence and arranging them in time for a sense of visual and emotional continuity around the existing annotations. In parallel, more of the story's structure began to develop, particularly the idea of the two parallel narrative spaces taking place in seven scenes. This provided the necessary structure to add section regions defining the scenes and their durations. With the sections and an image bank in place, the images were used to compose two color tracks, defining palettes for each scene, distinguishing the present-day narrative scenes from the embedded vignettes, and moments of significance requiring changes in color. At this point, I began shaping the expression parameters by painting in the sketch renderer. This provided the additional continuity and transitions among the key moments already defined and led to the development of additional narrative-driven stage direction for the actors. Event tracks were then used to create trigger regions for cuing the lighting, visuals, and music.

The Media Score was now ready to be used in the rehearsal process. Initial rehearsals began before the set for Figments had been assembled in the performance venue. The rehearsal process was largely improvisatory. Actors would respond to the score and my prompts and we'd run through scenes repeatedly, refining the action and performances, sorting out what actually happens to tell the story and fit the expressive contours of the performance. During these rehearsals, the Media Score served as both a visual stimulus in the direction of actors illustrating the emotional quality of their performance at a particular point in time. It provided a clear means of introducing performers to the structure of the piece and its shape. I also referred to the score myself when directing and made additional annotations as the timing of scenes and actions changed and modifying the sketch of the expression parameters to reflect the development of the performances and new ideas that emerged from the improvisation process. By the time rehearsals were completed, the score was in a form that both documented the piece and could serve as show control for the technological aspects of the production. The documentation aspect, as well as the ability to read both stage directions and expressive and emotional contour proved particularly useful when the scheduled date for the festival during which *Figments* was to be performed changed due to unfortunate circumstances. Two of the three cast members with whom the piece had been rehearsed and developed were unavailable for the new performance date, so the parts were recast and re-rehearsed using the Media Score to bring the replacement cast quickly up to speed.

The Media Score for *Figments* served as the primary control system for the performance. Events and parameter data influence lighting, visuals, and musical accompaniment at performance-time. The Media Score application was connected via OSC to Render Designer (described in Section 3.1.4) to trigger and modify visuals displayed on the LED strips surrounding the set. Visual looks were composed for each trigger, generally defining the look for each scene along with color information from the asset tracks within the score. Some of the looks could be dynamically modified by expression data to reflect changes in the emotional timbre within a scene. Musical accompaniment was also shaped by a set of triggers from the score and the expression parameters using a system described below. Finally, the Media Score application output cues to an Electronic Theater Controls Ion lighting console using the MSC protocol. Like the visuals, changes in lighting were also cued from the Media Score and used to create looks for scenes and important moments. The cues were pre-recorded in the Ion,

Figure 68: *Figments* system flow diagram



including movements of intelligent fixtures to follow blocking. Data from the color asset tracks were also sent to the Ion via MSC to move submasters to which the color channels of key fixtures had been associated. During the performance, the Media Score could be played back, automatically triggering changes in all of the connected modalities. The time track was used to specify moments where the score time would pause (essentially vamping the visuals and music by remaining in a specific mode) until cued by an operator based on onstage action.

5.2.1 ParaMIDI

Music accompanying the performance was produced by a system created for the production to receive data from the Media Score. I wanted the music to feel as though it had been composed and to have stylistic variations appropriate to each scene, so I opted not to use a purely generative music system. Instead, I composed source material using standard music notation software. The pre-composed music is orchestrated for two flutes, shakuhachi, oboe, bassoon, timpani, vibraphone, tubular bells, harp, piano, percussion, and string orchestra with occasional viola solos. It consists of sixteen sections (approximately one to two sections per scene). Each section consists of one or more "chunks" that are one to several measures in length with an optional introductory chunk for each section. The chunks and orchestral parts were both composed to be modular, so that their order and instrumentation is variable. The score was exported in the MIDI file format with each chunk named by section using marker meta-events. The MIDI file is read by the music system, called ParaMIDI, written in Java. ParaMIDI emits the modified music in real-time via MIDI to a software sampler with a General MIDI patch for audio production. When triggered, the system begins playback at the start of the first section. For each section, it will play the introductory chunk first and exactly once, if such a chunk exists. Then, until triggered to proceed to a different section, ParaMIDI will repeatedly play the subsequent chunks within the section in a pseudo-random order, always playing chunks to completion. Triggers are sent from the Media Score to cue ParaMIDI section changes, specifying the name of the next section to play. Introductory chunks provide transitional musical material from one section to the next in order to provide a coherent sense of movement, rather than abrupt changes. This indeterminate and procedural process is not unlike many examples from 20th century music, such as Stockhausen's *Klavierstrück XI* [102], or Tod Machover's *Chomsky Suite* [61].

ParaMIDI also receives the seven expressive parameters from the Figments Media Score. Several parameters continuously influence properties of the playback, such as tempo and attack velocity. Other parameters influence structural changes within a chunk and take effect only at the start of a chunk's playback. This approach is similar to how expressive parameters in the web applications for *A Toronto Symphony* altered the playback of pre-rendered audio tracks in response to user interaction (see Section 5.1.3). The likelihood of transition from one chunk to the next within a section of the music is governed by the Regularity parameter. The Density parameter influences the number of instrumental parts being performed at a time. Since the importance of instrumental parts varies based on the musical texture of each section, a lookup table is embedded into the MIDI file that specified the relative importance of a part for each chunk and thus the probability that a chunk will be played in a given section. A minimum of one part is always played with all parts with non-zero probability played for a Density value of 1.0. Within that range the number of parts playing is guaranteed to increase linearly with the value of Density. While Density influences the vertical likelihood of note events, the Complexity parameter affects the horizontal likelihood by varying the number of note events for each part within a chunk for notes below the duration of a beat. This has the effect of removing ornamental figures written in the source music and replacing these short notes within a beat with a single note of beat duration, musically grounding the piece in its harmonic context. The Texture parameter provides harmonic variation by transposing chunks. The degree of transposition is computed by mapping the linear Texture value to a range of intervals sorted by a tension score for each where tension is correlated with psychoacoustic consonance [41].

Since the source music was composed specifically with these seven degrees of freedom in mind, this hybrid music generation/playback system effectively produced a musical accompaniment to the stage action that worked in concert with the lighting and visual design to set the tone for each scene. The resulting music reflected the subtle emotional variations encoded in the Media Score and could fluidly expand to fill time required to play out each scene without simply repeating material. The source music played straight through would last for three minutes and thirty seconds. However, through ParaMIDI, it naturally conformed to the 18-minute duration of a typical performance of *Figments*. The actors had the freedom to take their time in their portrayals, without needing to follow the music or match musical cues. Instead, the music followed the natural and variable pacing of the performance. Typically, in stage performances accompanied by live musicians, the conductor, or musicians themselves in small ensembles, can influence the tempo, so that the accompaniment coordinates with dialogue or stage business. Vamps and safeties are often included in scores for musical theatre, should action take longer than expected, allowing the orchestral music to resynchronize with the performance. Otherwise, though the conductor has control over tempo and the cuing of instrumentalists and stage performers, there isn't a very large range of variability. Using ParaMIDI, the entire piece is capable of vamping at any point to accommodate the variability in the duration of performance and the realization of the Media Score. In future work, it would be interesting to explore using ParaMIDI in conjunction with a real-time notation system for performance by live instrumentalists rather than computer playback. As with most real-time notation systems, the lack of context and frequent changes can be difficult for musicians. However, like the aleatory pieces mentioned above, instrumentalists would still be able to rehearse the core of the music from the source score.

5.3 Feast

Another component of the Media Lab's The Other Festival was a culinary event composed by my colleague Janice Wang referred to as *Feast* [113]. Wang's research explores the use of multisensory stimuli on taste and food experiences. She develops meals that resemble theatrical experiences, incorporating a variety of modalities, in order to produce unique and, at times, unexpected shifts in perception. The medium-agnostic nature of Media Scores provides an ideal environment for designing and coordinating such experiences.

Feast was composed as a dining experience in five acts serving over 100 people. Each act had a theme associated with it that was related to the prescribed activity or aesthetic quality of the food and ambiance. In addition to eating, diners were encouraged to complete certain tasks, such as constructing their own utensils for the remainder of the meal from

Figure 69: Media Score for *Feast* (Score by Janice Wang)

supplied materials that were served in illuminated packages. The remainder of the acts featured appropriate menus in keeping with the acts theme. Music or soundscapes, typically one per act, added an additional modality. At the start of each act, a volunteer subset of the diners was electronically signaled—using discrete infrared-controlled LED pins worn by each—by Wang to begin serving the next course. Two of the acts also featured performative spectacles to ceremoniously or ritualistically introduce the act.

5.4 Workshops

In May of 2013 my colleague Elena Jessop and I presented two workshops focusing on the use of parametric abstractions and scores in performance. The first workshop was held at the Research@ML conference at the MIT Media Lab. Inspired by the curriculum developed by the Toronto District School Board teachers for *A Toronto Symphony*, the workshop featured an exercise in which participants were instructed to create a parametric graphical score for a multi-modal performance work using yarn. The participants had to work in small groups to develop the language of notation, defining the expressive parameters needed for their performances. Next, participants performed their work by interpreting the score through movement and percussive or vocal sounds.

The second workshop Jessop and I conducted was at the ArtEZ Institute of the Arts in Enschede, Netherlands as part of the symposium entitled A New Dawn: Search for a Spring, which was commemorating 100 years since Stravinsky's *The Rite of Spring* by exploring the future of performance for the next 100 years. Participants from the workshop included students and instructors from ArtEZ and other nearby institutions with backgrounds in music performance and composition, music therapy, informatics, and storytelling. We began the workshop with the same activity: composing and

Figure 70: Yarn parametric scores
A group of participants at the ArtEZ workshop composes a Media Score out of yarn before performing it.

performing yarn scores to demonstrate the concepts of expressive parameters and graphical scores. For this workshop, participants used musical instruments to perform their scores. In both workshops, the graphical scores are, in reality, a subset of Media Scores. Without explicit instruction, all groups of participants ended up using the yarn to create explicit-time representations and the parameters they elected to depict included notions of intensity, variation, complexity, and density, with definitions not unlike some of the seven parameters outlined in Section 4.3.2 for the Media Scores application. During the performance of these scores, all of the groups concluded that they would require one person to function in the role of a "conductor," someone who would keep time and indicate the position in the score at which performers would interpret the parameter values.

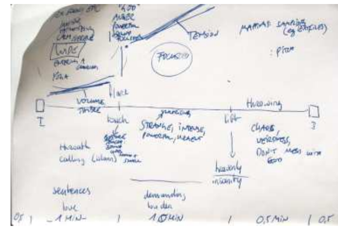


Figure 71: Hand-drawn score for ArtEZ performance
This hand-drawn Media Score guided the development for one of the interactive group performance pieces composed at the ArtEZ workshop. (Drawing by Rck Steggerda)

Subsequent activities in the ArtEZ workshop included having participants interpret through movement and vocal sounds a collection of abstract imagery—paintings and photographs—as implicit-time scores. Participants were also introduced to the Media Scores application and were assigned the task of using it to create Media Scores for fairy tales of their choosing, which they then presented to the group. The weeklong workshop concluded with performances of several music and sound-related group projects, each incorporating performance technologies and audience interaction. Some groups used the Media Scores application or hand-drawn notations to develop their pieces, though the show control functionality of the application was not utilized.

5.5 Extending Death and the Powers

In February 2014, *Death and the Powers* will receive its third showing in the United States with The Dallas Opera in Dallas, Texas. For the final performance of that run, the opera will be broadcast live to remote venues around the world. The broadcast will feature a multi-camera video switch and surround sound audio of the performance. In the spirit of Remote Theatrical Immersion and the experimental extension of *Sleep No More* discussed in Section 3.3, work is underway to augment the remote experience, giving these audiences the ability to vary and curate their experience of the opera, escaping the passive screen-centric mediatization of typical live theatrical broadcasts. Current plans include providing additional audio content to audience's mobile devices, generating interactive glimpses into The System and Simon Powers's experience there, and offering a backchannel for audiences to influence The System, including having a presence as light within the Moody Foundation Chandelier in the Winsepear Opera House [115]. In a sense, remote audiences will actually be a part of The System, dwelling in that transcendent realm alongside Simon. From there, they will have an omniscient view of the action of the opera, as presented through the multi-camera video stream, surround sound. Additional audio, video, and multimedia content privileges remote audiences with perspectives that can only be obtained by having entered into the story's "world of light."

This project will require the creation of several layers of additional parallel content that must complement the existing production and, thus, must be designed in collaboration with the opera's creative team and new partners, utilizing existing and new assets. Media Scores is the ideal tool for the planning and coordination of such a project. The visual score for *Death and the Powers* can now graduate to a Media Score, the tool that was only beginning to be conceived of at the time of its creation. The score would be updated to reflect the current state of the opera itself. Media Scores will not be used to influence the existing production of the opera, as the extensions being created are designed to be as independent from existing show systems as possible, in order to avoid adversely impacting the stage production due to insufficient time to shake down the changes. However, Media Scores can function as the central pre-production tool and show control for additional systems. It can receive triggers from the Keyboard 2 part in the musical score for the opera in the same manner as the existing Powers systems and, in turn, trigger new content generators designed to produce the augmented audio and video material, control for remote audience members' mobile devices, and responsive visual content for display on the Moody Foundation Chandelier. Media Scores can also be used to modulate feedback from remote audiences and integrate that as necessary with the existing arc of the opera, as well as the live Disembodied Performance data, to create a completely integrated experience.

6 CONCLUSION AND DISCUSSION

A work such as this is never really finished; one must simply declare it finished when one has, within the limits of time and circumstances, done what is possible.

—Johann Wolfgang von Goethe

In this thesis, I have introduced a practice for composing the modern-day Gesamtkunstwerk. Media Scores, and the current implementation described, lay the groundwork for a streamlined approach to artistic expression across media and modalities. I have demonstrated how Media Scores can be used through example projects and presented the procedural mechanics of composing with the Media Scores application. This method bears with it important implications for the role of the artist, audiences, and the role of technology in the creation of new multidisciplinary works.

6.1 Quod erat faciendum

The formulation of Media Scores, the Media Scores application, and the accompanying discussion presented in this document offer the following contributions:

- A conceptual framework and design guidelines for the role of technology in the creation of Gesamtkunstwerk including its impact on the traditional model of production phases, collaboration, authorship, interactivity, and liveness

- A method for representing evocative story and expression parameters using the metaphor of a score for the creation of a Gesamtkunstwerk based on research in cognition, perception, and emotion

- The implementation of a software tool for the creation of these parametric score-like representations

- A systematic description of the multiple ways in which the story parameters, including time, can be considered semantically and intuitively visualized

- A catalogue of essential properties for flexible show control systems used in designing and performing live, multimedia experiences, building on the philosophy of control system architecture developed for *Death and the Powers*

- A formalization of the types of events that can occur in control architectures for live performance based on *Powers* systems and methods that have arisen in Hyperinstruments and Opera of the Future work over the years

- An example work that uses Media Scores to compose and to generate control information for use in the production, i.e. the performance of *Figments*

Media Scores constitutes a new framework for approaching the composition of a general notion of art. It is a revised, holistic practice that draws from existing approaches and my personal experience creating works that approach the Wagnerian ideal of Gesamtkunstwerk. It fosters the notion that expression is the key to the artistic experience and that this essence can be reinforced through the juxtaposition and conglomeration of expressive modalities. I have shown how multiple representations can be used to interact with this abstracted form of an artwork that can exist in a continuum of interactivity. I am hesitant to describe it, particularly in its current implementations as described in this thesis, as a universal language for the arts. More appropriate might be to refer to it as a "language of universals". There may be many such languages, implementations, or styles of representation, but they have at their core a set of intuitive abstractions that transcend media.

The Media Score is at once expressive and practical. Scores become the language of communication among artists and systems, unifying them in time and intent. It is a tool for the development and realization of works, a nexus for the integration of technologies and human collaborations. The Media Scores application demonstrates a user interface that presents relatively simple interactions for manipulating complex and abstract parameters and annotations. A low barrier to entry or simple learning curve makes it functionally accessible to creative personnel and artists for composing without technical expertise or much training.

In *A Toronto Symphony*, we saw how Media Scores can be used as a guide to shape and manage a large-scale collaboration and provide tools for collaborators of different skill levels and abilities to interact with an evolving work and compose aspects of it together. In *Death and the Powers*, the visual score served many of the functions that would eventually become a part of Media Scores, namely that of a collaborative document and repository for different types of information. The proposed extensions for *Powers* suggest how the Media Score can be integrated into an existing control architecture to manage and shape interactive components of a production, from remote broadcasts to the incorporation of performance and audience data into an expressive presence through different modalities. The use of Media Scores in the production of *Figments* demonstrates the role that such a tool can have at all phases of production, from ideation and authoring to show control.

6.2 Next Steps

The concept of Media Scores, as I have presented it in this thesis, is merely the first step in the aim to develop a uniform method and notation framework for the composition of expressive works of art. The ideas and implementation are but in their early stages and a variety of avenues of exploration remain. Using Media Scores to create additional performances and works of art will provide invaluable feedback on the current implementation, elucidate limitations and shortcomings of the concept as presently formulated, and suggest new functionality or refinements. With each future production and the application of Media Scores in projects with different contexts and considerations than those for which it has already been used and that inspired its creation, new needs and possibilities will be identified. I intend to continue the development of Media Scores in the following ways and look forward to seeing how others contribute to the evolution of Media Scores through their use and extension of its present functionality.

6.2.1 Towards a Universal Parameter Set

The core component of a Media Score is its expression parameters. These represent the essence of the artwork and the journey on which the composer takes the audience or participant. I tentatively propose that expressive gestures in different modalities may in fact have a common set of descriptive properties. Such a parameter set would exhibit a perceptual congruence across modalities, which can then facilitate their comparison. Domain-general properties, such as cross-modal expression parameters, do not preclude additional domain-specific formal properties afforded by various media or modalities.

As it is, the Media Scores framework makes no implicit assumptions about the number and definition of the expression parameters used within a score document. The semantics of the parameters is left both to the artist's use of the expression parameters and how they are interpreted by renderers and content generators in a manner that faithfully reproduces the artist's intent. In Section 4.3.2, I defined seven expression parameters that I am currently using with Media Scores. We saw different parameter sets for use with Disembodied Performance and with *A Toronto Symphony*. What makes the mapping of expression parameters work is that they represent common high-level concepts that each of the modalities are capable of conveying. The parameter set definitions I will use with Media Scores in my work will likely continue to evolve with each subsequent project.

Nevertheless, intuition supported by the background domains described in Chapter 2 would seem to suggest that there may be a universal set of abstract parameters with well-defined semantics that could model expression across any modality and that would be unambiguously understood by artists and interpreted by readers of a Media Score. The difficulty in identifying such a set is twofold. First, additional research into perception and cognition is needed to more precisely formalize how expression is encoded in each medium. Narrative analysis is well-developed for the representational medium of story-driven text. Considering abstract expressive representation, the field of music cognition has made great strides toward this in its own medium, in part due to the readily parameterizable and quantifiable nature of formal parameters in music. The research into understanding and parametrically modeling other modalities is not as developed. The dimensionality of the other modalities is greater, making reasoning and parameterizing them an additional challenge. In visual media, analysis and formalizations of practice tend to be either qualitative and subjective or take the form of generative grammars, such as the shape grammar proposed by Stiny and Gips [101]. Their formulation borrows from Chomskian linguistics and therefore bears a resemblance to Lerdahl and Jackendoff's famed *A Generative Theory of Tonal Music* [55]. The shape grammar, and many other procedural methods for generating imagery, operate on discrete visual units. This, in turn does, not make the formalization descriptive of the broad range of visual gestures. In the case of music, the fundamental elements of Western music, particularly for the purposes of analysis, are discrete in both pitch and time. The visual field, especially in abstract works, is not solely limited to discrete objects. Comprehensive theories and formal theories of abstract visualization are needed, in addition to similar hypotheses for other abstract sensory modalities.

Another confounding factor in arriving at a universal parameter set is simply that of language. The concepts that expression parameters attempt to describe are perceived and felt above the modality-specific level.

Therefore, they are at a higher level of abstraction than the text itself. However, we use text to name each parameter and define it. Thus, there may necessarily be a mismatch between the definition of the parameter and what it is intended to represent. Identifying an expression parameter space that is orthogonal in its dimensions entails finding the right words—the metaphor that best approximates the intention—to describe the parameters, which may simply be a process of trial and error towards group consensus. It is in this way that the use of multiple renderers and views in the Media Scores application frees the definition of the parameters from linguistic shortcomings. The behavior and representation of a parameter in multiple modalities, styles, and perspectives give a more refined intuition about what expressive information that parameter should encode.

6.2.2 Analyzers, Renderers, and Content Generators

The extensible nature of the Media Scores architecture and implementation make it straightforward to add new analyzers and renderers to the application. Analysis of different asset types can be refined over the current version to better reflect the expressive nature of the source material. A library of analyzers would be appropriate for inferring information from a variety of different kinds of assets. The current image asset analyzer treats an image as a field of color and texture. Taking into account the style of the image, if photographic or an artistic work, would refine the derived expression parameter values. If the images are storyboards, could expressive inferences be made from an analysis of the composition of the frame? Could this also be achieved if the images were more diagrammatic, such as blocking diagrams of set pieces? Music analyzers could similarly be refined by a variety of analysis techniques that are available. Audio analyzers would also benefit from knowledge of the context for the audio they are operating on. Text analyzers could also leverage story archetypes, advancements in natural language processing, and common-sense data regarding the affective content that may be implicit, explicit, or contextually encoded in the rich medium of language. The possibilities for enhancement and refinement are endless across the variety of asset types. Analyzers play an important role in linking conventional practices and existing resources, source texts, and media with the Media Score.

Renderers are another prime opportunity for further exploration. Artists should be able to interact with the score, viewing it and shaping it in a manner that's most intuitive for them. Common notational forms can be added, as well as abstract renderers that emphasize different aspects and provide new perspectives on the expressive contour of a work of art.

Content generators are intended to be production-specific. Earlier, we saw content generators that produce abstract visual imagery, cue playback of pre-rendered imagery, control the generation and playback of sound and music, and influence lighting. Many more modalities can be controlled by

Media Scores during performance. Digital fabrication techniques could also be connected to the Media Scores application or process a score file directly to create realizations of the artwork in physical modalities such as sculpture. Media Scores could also be used with a content generator that accepts media asset track data, as well as expressive data in order to edit and assemble media. An example of this is a video editing application that could use the analysis of video, image, music, and audio assets to switch or transition among them based on the expressive parameters of a score. The resulting edit could then be performed from the score or rendered to an independent video file. This shaping or sculpting of media would satisfy the usage envisioned for Personal Opera in Section 4.1.1.

Additionally, content generators could result in new media or contexts for experience. For example, objects or appliances could be designed that interpret and play Media Score documents at any time. Imagine an object—a desk ornament, for example—that could continually perform a Media Score in your home. Like a digital picture frame, you could download Media Scores of your choosing, or perhaps a playlist of scores, onto the object. They could include the Media Score for your favorite Bach concerto or one that your niece made to commemorate your most recent visit. The performance of the scores could perhaps take the form of swirling colored light, as in Wilfred's lumia, or a soft, pleasant musical sound. The performance could exist subtly in your periphery as a dynamic artwork, reminding you of your loved one or a particularly memorable time. However, when you picked it up, it could sense the quality of your interaction and the performance of the score could adapt to reflect your mood or reveal additional aspects of the abstract story.

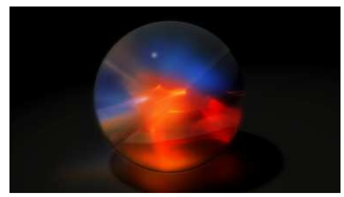


Figure 72: Orb player
A conceptual rendering of an object that could ambiently perform Media Scores and through which one could interact with the score

6.2.3 Enhanced Real-Time Capabilities

As a scoring tool, the Media Scores application is well-suited to creation in pre-production for realization at a later performance time. However, as I demonstrated in Chapter 4, by adhering to the design principles for performance systems that I presented, Media Scores can function interactively in all phases of production. The artist can engage in compositional activities in a directed or improvisatory manner at performance-time, since the output will respond immediately to the user's actions. In performance contexts, live performance and audience data can be incorporated into the score's output. Assets can also be used by content generators during playback, so that Media Scores can be used to flexibly define content for use by other systems.

There are still opportunities for Media Scores to become even more integrated into performance-time behaviors, including facilitating more types of compositional and participatory interactions. The application could be extended to allow not only the manipulation of assets in sequence time and the shaping of expression parameters and performance data, but also to incorporate the inclusion of assets in real time. Consider a Media Score composer to allow audiences to contribute imagery at performance-time for display in an installation or to be reflected through an expressive translation into a different medium. How could the composer orchestrate constraints on the live input of assets from audiences such that they can be meaningfully incorporated? The application would need to be extended to define appropriate constraints for assets that would be used directly. This modulation of the real-time supplied assets by the score is conceptually equivalent to the enveloping of performance and audience parameters. The assets could then be relayed to content generators as they arrive at the score. Asset analyses could also be used in real-time, though there is generally a slight processing delay that varies due to the complexity of analyzers' algorithms, so that assets could function as performance data and be blended with the scored expression parameters.

Additional real-time functions would also be advantageous for the composition of a Media Score, not just its performance. If data from performance sensors were analyzed in real-time and sent to the score, not modulated downstream by the score, then live sound and gesture could be used as modes of composition through recording. This would augment the existing mouse-based user interactions of painting and direct manipulation, effectively creating a way to record a performance as a score; capturing a dynamic-time representation. This model is quite familiar from digital audio workstations, into which audio tracks can be recorded in real-time and later manipulated. This approach to score creation would be useful for asynchronous improvisation, mapping performance modalities into other modalities via the score, and editing performed data.

The idea of composition as the process of defining constraints on the form that the work of art will take when realized is critical to the creation of an artwork in a medium-agnostic fashion. Constraints presently are explicitly defined by the composer as values for tracks, as well as optional envelopes around those values for shaping performance-time variations. The Media Score application could be enhanced by incorporating knowledge of shape and an additional understanding of parameter semantics, so that it could assist composers in the composition process. These constraint-based guides could help automate the composition and performance coordination of modalities. Several examples of this approach can be found in other types of applications. The use of a compressor in audio processing ensures that audio levels remain within a specified range while avoiding discontinuities in inputs that exceed that range. The music composition application Hyperscore automatically applies constraints on the harmonic relationships among notes, freeing the user to focus on expressive contour, while the software keeps track of the details required to produce musically consonant pitches that have the intended transitions in harmonic tension. Similar rule- and constraint-based approaches have been employed experimentally in theatrical control systems. Incorporating constraints and heuristics into the response of the Media Score application to the composer's interaction would broaden the efficacy of compositional activities at performance-time, thereby attempting to ensure expressively meaningful operations. Thus, the score becomes more of an instrument that can be performed by the composer.

6.2.4 World-Building and Open-Ended Experiences

At present, Media Scores is an approach to composing the framework of a story or an experience of an artistic medium. What it does not fully address is the process of world-building that is an essential part of production design for creating narrative-driven experiences. When writing a story, the author envisions the world in which the plot unfolds. The world has conventions, laws, rules, and life beyond the window through which we see it in a book, on stage, or on the cinema's screen. In production of a television series, for example, it is common practice for the creators of a program and the initial writers to document this world in what's known as a *writer's bible*. This document serves as a reference for future writers of the series, detailing the world of the series, its rules, significance, character backstories, and notable events. It is a document that evolves and grows as the series progresses, so that future episodes can reference the past and remain consistent with a plausible world. The production design process similarly is one of taking the vision of the world in which a story or types of stories can transpire and physicalizing it through setting with a look that evokes the tone of the designed world.

World-building is of particular importance in open-ended media, such as computer games and immersive experiences. Not only must the rules of the world be represented to an audience, the audience must be able to enter into that world and interact with it. This was entailed in the challenge of developing the remote augmented *Sleep No More* discussed in Section 3.3. In that instance, the JEML language, as interpreted by the Cauldron control system defined the space of the world, the events that could occur within it or that occur at particular times, how participants could interact with the world, and the visual style of the world. It also accounted for changes based on location within the world and the unfolding of the narrative. As I noted before, there was no authoring tool for creating this script. A Media Score could provide an overall shape for the experience, incorporating expressive variation based on the activities of the online and onsite participants, but it could not easily handle sculpting various expressive trajectories that are specific to the events and paths encountered by the participants.

Media Scores is an experience-composing tool, but could potentially evolve into a world-building tool, creating worlds through which arbitrary trajectories of experience can be shaped. Currently, the world-building process for a single trajectory can be aided by the Media Scores application through the extensions of conventional practices that it affords. Annotations and assets can be added to the score to build up notes, storyboards, and animatics. Renderers or content generators could be written or connected to the application that would respond as previsualization of the world created. Additional asset types could be created to facilitate this, as well.

The next step in the evolution of Media Scores would likely be a means for representing reconfigurable stories, taking into account non-linear and random access experiences, such as those in the extended *Sleep No More*. While the Media Scores application can be triggered and controlled via the time track to move about in time or across sequences, this does not represent the degree of variability in trajectory that Cauldron and JEML afforded. Non-linear experiences and the art of *database cinema* construct meaning from disjoint alternatives in a similar manner. Database cinema arises from the quest for a medium that exploits the symbology of meaning prevalent in the computer age, though related examples extend back to films such as Dziga Vertov's 1927 Man with a Movie Camera. In this paradigm, the world is a collection of objects or fragments of scenarios that get assembled into a narrative as played, performed, or edited. The library of possibilities does not itself contain a narrative thread, but exists to enable many paths from many beginnings to many ends. It is a difficult challenge to conceive of a visual notation for representing the multiplicity of branching and converging paths that a participant in such an open-ended narrative experience—particularly when there are aspects that are also constrained to the linear or monotonic elapse of time in a live performance—as in *Sleep No More*. This particular area of research is of great interest for making Media Scores and the Media Scores application a truly flexible storytelling tool. Breaking the paradigm of linear-time or linear-space that is implicit in the current architecture and implementation would require new ways to visualize the complex array of possibilities in a clear and evocative manner. It would be an even more malleable approach to reasoning about time and progression through a sequence. The score would no longer be an aesthetic artifact representing an experience, but one concisely and readily representing the set of all possible experiences within a world. A composer could create open-ended structures, with constraints on interaction, movement, and expressivity capable of intentionally sculpting multiple trajectories and perspectives on a single Gesamtkunstwerk. Then, Media Scores would be capable of interacting with systems like Cauldron or become an interface for world-composing, an authoring tool beyond what would be required for building worlds within the object- and rule-based model of JEML.

6.3 What is a work of art?

Throughout this thesis, I have referred to a work of art as separable from its material realization. This abstraction is at the very core of the concept of the Media Score. Is a work of art the final product? Or is it the artist's idea in its purest, potentially uncommunicable form? Is it each audience member's individual experience of an artistic effort presented to them? The script or score that is used to create the artist's vision? The video or audio recording documenting a particular production? The catalogue photo of a painting or sculpture or the painting and sculpture themselves?

There are no clear-cut answers to these questions. Through the present, we often consider a work of art to be its material form: the painting by Rembrandt, the sculpture by Giacometti, the building by Saarinen. Recording and reproduction of artworks makes the question more complicated. We still have a sense that a halftone print of an artwork in a book is not the artwork itself and, while we can appreciate the work from the print, factors of scale, texture, presence, and contact are missing. The image is not a substitute for the original. Cinema, by its very nature, on the other hand, necessitates a multiplicity of duplicates. Films are distributed as copies and there's no sense of viewing the duplicate as being inferior to the master. Remakes of films separate the reproduction in time, not only the multiplicity in space. Few would argue that the 2004 film of *The Manchurian Candidate* is the same work of art as the 1962 release. What is more, both films are adapted from the 1959 novel of the same name by Richard Condon. Are all three manifestations of the same work of art composed by Condon? Yet, the 1962 recording of Beethoven's Symphony No. 7 under the baton of Maestro Karajan would be considered to be of the same work as the symphony's 1813 Vienna premiere. Variations in the conducting and interpretation, tuning, performers, location, audience don't alter the material to the point of no longer being the same work. For all of the variation, something authentic, uniquely identifiable, and essential remains and is preserved among the notes. What do these performances have in common? The score.

It's clear that what constitutes a work of art is dependent upon the mechanics of the medium and the degree of influence of creative input and interpretation during the realization of a work. Different designs and different screenplays create different works, even if they share the same story. The role of a conductor and musicians does not sufficiently alter the intention notated in a musical score from what was implicitly encoded by the composer. By blending disciplines and abstracting artistic intention into a score form that could be used to create an artwork in any medium and by allowing various forms of interactivity to be composed to influence the work, Media Scores touches on issues of authorship and production organization that will continue to be explored as the arts evolve.

6.3.1 Authorship

Typically, the roles of author, director, or composer are regarded as originators of a work. The intention of Media Scores is to preserve this. However, in the creation of interactive and multidisciplinary media, the role of the author comes into question. In Chapter 5, the projects discussed highlighted two of the several possible authorship models that Media Scores can accommodate. *Figments* and *Feast* both relied on a single author to compose a score for a work and, in these cases, the composer also oversaw the realization of the project from score to performance. Although capable of supporting such an auteur-like control over a piece, Media Scores is also inherently a collaborative tool. In one use case, leveraging sharing and versioning technologies, a score can be composed by multiple authors. Pieces can easily be the result of a collaboration of equals in assembling and shaping the work's intent. This was the approach adopted for *A Toronto Symphony*.

As François Truffaut sought to reestablish the role of the artistic creator in the complex, collaborative, and multidisciplinary world of film [15], one usage of Media Scores is to further the influence of the auteur of the Gesamtkunstwerk. A single artist can assume ownership of all avenues of the creative direction of his work by sculpting the intention across the constituent modalities in which the piece is realized. This recalls Wagner's detailed scores for his operas and his involvement in their production. In opera, the composer is regarded as the creator, but is generally responsible for the music and the setting of the libretto, while the design and performance of the production often falls to others. Wagner did more than compose the music, but also wrote his own libretti. Wagner's scores supply more than the composer's mandate, but extend to scenic and stage direction, as well as musical performance. In his essay "On Conducting," he wrote in the voice of a true auteur, "Composers cannot afford to be indifferent to the manner in which their works are presented to the public" [111]. He suggests that the conductor can become too much of an interpreter, diminishing or transforming the composer's intent. The construction of the Bayreuth Festspielhaus, brought the composer's influence to the very architecture of the theatrical venue in which his works would be performed, specifically to satisfy the musical and scenographic demands of the *Ring* cycle. Due to the abstract nature of the expressive representations in Media Scores and the underconstrained parameterization of expression in the abstracted work, the auteur model for the Gesamtkunstwerk does not eliminate the necessity or minimize the significance of the creative team that it takes to realize the work from its score.

With respect to audiences, a truly interactive work, where all aspects can be shaped by the interactions of the audience or users, can be said to belong to each user. Each interaction leads to a wholly unique experience of the work.

What then is the role of the author? To realize such a high level of interactivity in an experience requires that the user create the system for the experience, which is typically not the case for interactive art works. Thus, the author's role is, at the very least, to create the system and to define the tools used, the methods of interaction, and their consequences.

In his article "A Framework for Digital Expression", John Maeda points out that, in technologically mediated works, the result is not solely a function of the artist, but also the system and, therefore, the designer of the system, as well [65]. A work is co-constructed by the system and its user. If the artist is also the developer of the system and the work is static, then the model collapses back to that of the auteur. However, if the artist uses an existing system, then he is, in some sense, collaborating with the creators of his tools and the history of these tools, as is the case in traditional media, as well. Regardless of whether the technical implementation of the system originates with the artist or elsewhere, the artist's objective is to communicate to the audience through direct manipulation of the medium or through performing agents. The contributions of singers, musicians, dancers, and actors in any performance work is arguably also that of co-creation with the author or composer or director. In interactive performance and installation works, the audience assumes an active role in how the experience is played out, fragmenting the authority of the artist.

The author is then responsible for, at the very least, defining the constraints of the experience. The constraints are the rules by which the system responds and the aesthetic of the representation of that response. Through this process, by selectively constraining response over time or other aspects of the interaction, the author can then tell a story. In musical composition, for example, the assembly of an orchestra has the potential to play an infinite combination of notes in a myriad of possible sequences. It is the composer's score that constrains the system to a particular sequence of notes. In interactive systems, the author and user interacting with the work can be thought of as co-creators of each unique experience of the work. However, through the constraints set out by the author during the creation of the work, there is still a directorial role that concerns itself with the communication of a particular message or quality of experience.

The lossiness or potential for variation and interpretation in reconstituting a script or score is celebrated by our desire to critique different performances or interpretations and to create different styles of retelling. While Media Scores give composers the opportunity to be more specific when conveying their intent, they are not intended to dictate the entirety of a production. They allow the composer some freedom to choose when to be specific and when to be ambiguous and leave aspects of the work open for performative interpretation. Media Scores enable this co-creation to varying degrees. By giving the composer the tools to design more than simply a source text or musical score or storyline for an experience, Media Scores allow the creator to be a true auteur of the experience, specifying in detail a unique style and program for the enactment of an experience. By providing the capability for live input into the system from performers, aleatory processes, and an audience, Media Scores give the auteur a means of assigning, modulating, and attenuating the possible influences in advance by encoding this into the score for the piece. However, the extent to which this happens is in the purview of the composer. If the composer desires fine-grained control over the interaction, it is possible. The composer can create modes over the course of all or part of a piece that defer much of the realization of the piece to performance-time input.

6.3.2 Redefining Traditional Roles

Technology and technological performance systems are but additional tools in the artist's arsenal. Placing technology onstage should not privilege it beyond any other modality of performance. Visual projection is inherently no more significant than costume design or acting. It is up to the design of the production to emphasize any one over another for particular effect.

Technology does have two important implications in the realization of a Gesamtkunstwerk, though. For one, technology can be used to produce content in multiple modalities, whereas the conventional disciplines are generally divided by modality. Lighting, costume, and set designers need to consult on color palettes in order that their designs faithfully conform to the production concept, but also so that their respective designs don't negatively influence each other. A saturated blue fabric in a costume when lit by strong red light would appear black and is likely not the intention of either the costume or lighting designers. Communication is essential among the creative team during the creation of such multidisciplinary works. New performance technologies tend not to fit cleanly within the existing organizational structures of production departments. Technologies that produce output in multiple modalities and are managed independently of conventional departments necessitate clear channels of communication and understanding of the implications that changes in one modality would have on others, for all involved.

All of the productions in which I've personally been involved and that are described in this thesis required additional communication among departments in ways that were at times unfamiliar or unexpected from the vantage of traditional practice. Introducing projection imposes certain constraints on stage lighting for the projection to be effective. A creative decision to change a color in visuals as related from the director to the visuals designer may also necessitate a change in color palette for the lighting designer, as well. If changes are communicated only to part of the team, the designers that are not aware of the decision will not be able to assess whether or not they need to compensate accordingly and to do so.

This leads us to the second implication of technological performance systems: technology can serve as the glue that unifies the modalities and other elements of performance and stagecraft. Through sharing control data and parameters, technological and human aspects of performance can interact. This is an essential component of allowing technologies to respond to the fluidity of time common in live performance contexts and preserve a sense of liveness and ephemerality. However, this is another way in which technology in performance necessitates heightened communication among personnel. Decisions made in one department are more likely to affect others when everything is so interconnected. When systems are connected, executing an action with one system could cause another system to respond, so care must be taken when systems are being used online. If the reactive visuals in *Spheres and Splinters* are being refined and are still receiving data from the cello bow, the visual designer may not be able to properly evaluate or manipulate the system if the cello happens to be played. Cuing based on triggers notated in the score, as in *A Toronto Symphony*, requires that notes in the trigger part be unique unless a trigger is to be repeated. If there is an audio event and a visual event associated with a particular trigger, then both will always occur when that trigger happens. If the sound designer decides that the audio for that trigger should be repeated later in the piece and writes in the same trigger note to the score without consulting with the visual designer, then the visuals associated with that trigger will recur unexpectedly at the later time.

This communication is not only necessary among designers, but across all of the production staff. An anecdote from rehearsals for *Death and the Powers* illustrates this. A technical rehearsal call was held for a particular scene and the singers onstage for that scene were called. The director wished to develop the singer's performances, including reactions to Simon Powers's manifestation on the walls. However, the singer playing Simon was not called to rehearsal, since he does not appear onstage during the scene. Though the second keyboardist dutifully triggered the visual modes during runs of the scene, without a singer connected to the Disembodied Performance System, very little appeared on the walls.

Although the introduction of new technological practices necessitates greater communication, it can also provide a solution to the problem. This is where Media Scores steps in by providing a central reference point for the design and cuing of performance systems. The Media Score begins by serving a dramaturgical function and extends into the realms of design, coordination, and collaboration. Serving as a point of collaboration, the Media Score document remains updated with changes made to the work, so that the production team is aware and the systems can respond directly, as well. This use case demonstrates the true power of technology as the connective tissue in complex performance infrastructures.

Technologies such as Media Scores and the complex performance systems we've looked at will be increasingly commonplace in performance practice. The Gesamtkunstwerk will accumulate more constituent art forms and the role of audiences will increase. These factors will require organizational shifts in the production process. New roles in production hierarchies will need to be established to manage the added modalities and the interaction among them. Producing organizations will likely acquire master networkers to manage the communications among systems, just as they have master electricians to handle connecting and powering lights and technologies. Many performance venues rely on labor union crews, most commonly the International Alliance of Theatrical Stage Employees. The divisions of labor typically fall along the conventional disciplines of stagecraft. With complex show architectures, which union or which member of a union is responsible for a part of the production becomes the subject of question. When the keyboardist in the orchestra of *A Toronto Symphony* triggers video to playback and be projected, this entails, at the least, the considerations of the orchestra's union and the projectionist's union.

Another concern that often comes with the introduction of technology into any practice is that it will supplant humans or replace skill and specialized training. In the case of Media Scores, it is important to point out that it does not eliminate the role of any production personnel or designers. Media Scores does not make dramaturgical or design decisions for production staff. It broadens the palette of what creators and designers can do and provides a common language for their interaction. Schillinger states it well when describing his proposed methodology:

> It does not circumscribe the freedom of an individual artist, but merely releases him from vagueness by helping him to analyze and to realize his own creative tendencies. It gives him a universal knowledge of his material: the principles and the techniques of this system permit an infinite number of solutions, which satisfy any requirements set forth by art problems. [94]

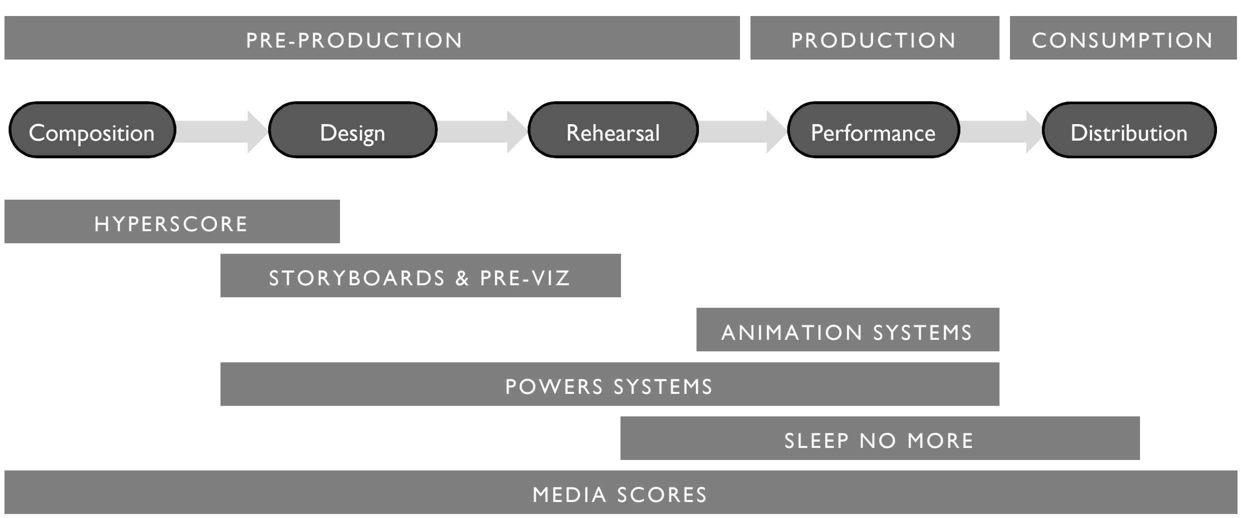
Media Scores are intentionally open-ended. What the implementation visualizes is the score. Renderers in the application are not meant to become the fully-realized form of the work, but a representation of its essence. There is no one absolute view of the work. The act of creating the score is a creative one that remains in the hands of the artist. The Media Score facilitates the creation process. Designers can then bring their skills to the interpretation of the score in the final set of modalities. This does not preclude the possibility of automating the realization of the work from the score, but the practice of developing that process would then be a creative one, as well. Furthermore, Media Scores offers a continuum of control for the creator. An artist may compose more aspects of a piece than they ordinarily would—such as the composer of a musical assigning reference images alongside musical notation—but the realization process can take many forms. The framework itself offers options. It does not replace or eliminate them.

6.3.3 Phases of Production

In Section 2.10, we talked briefly about the five phases that the production of a performative work of art goes through and how control of the work gets handed off at each phase. As a tool for composition, a collaborative design reference, and a show control system, Media Scores can have utility during pre-production and production phases. In technologically mediated live performance, both situated in physical space and virtual space, the show control function of Media Scores can influence the distribution of the work, as well. Technologies that can influence a recorded or existing work, essentially making what would otherwise be a static artwork at the time of distribution into a responsive live experience, can also rely on Media Scores to shape the interaction. Thus, the Media Scores application can span the range of production processes and afford a level of control or shaping at all phases.

Figure 73 shows how Media Scores relates to other tools that have been discussed. Conventional musical composition tools—from pencil and paper to notation editing software and Hyperscore—function only in the pre-production phases. They enable the creation of the core of the work and some opportunity for notating aspects of the design, which might include instrumentation in music. Scripts for plays and films, as well as libretti, are also situated at these early stages. Unlike the software tools, the paper scripts and scores have an influence that often extends into the rehearsal process. Actors can read their scripts during some rehearsals. Musicians often retain their scores through performance. Storyboards and previsualization techniques originate during the design process, as the script and production concept are translated into more concrete forms. As visual references, they guide the rehearsal process. In film, previsualization may

Figure 73: Utility of applications in phases of production
Various applications and practices discussed are shown in correspondence with the phases of production which they serve.



also guide the capture of the production and post-production, since these tasks happen asynchronously. Animation and non-linear editing software can assemble existing content, often created as part of the production phase and guided by pre-production planning, into a form that can be distributed. These systems themselves are not used in the pre-production or distribution phases. The performance systems created for *Death and the Powers* that were described in Section 3.1 allow for content to be designed and assembled within the software applications. They can be used for offline prototyping and visualization, edited flexibly during rehearsals, and perform the content during the production. Cauldron and other systems created for the extended *Sleep No More* experience did not facilitate content creation. However, they were critical to running the production environment during performance and, given the nature of remote theatrical immersion, the distribution of the content.

As we've seen, Media Scores can function at all of these phases. In this way, Media Scores can evolve with a production, providing a uniform notation and collection of show assets that can become the language that runs the final performance. By being able to serve in all of these capacities, it can be used for authoring interactive works and works in multiple modalities. The score can be an object of distribution itself, or the application performing the score can control the distribution. Interaction can happen once the work is distributed, not solely in production. Composition activities can also occur at any of these phases. Therefore, control, centralized through the Media Score, need not necessarily be handed-off as the production progresses. All of the phases of production become less discrete operations in a continuum of creation, performance, and consumption. In the previous section, we saw that the nature of Media Scores dissolves the boundaries between departments in the development of a work, compressing the hierarchy of control and influence of production personnel to accommodate the conceptual and technological interdependence of disciplines in a Gesamtkunstwerk. We can now see that this property extends into time, spanning and coalescing the previously discrete phases of production.

6.4 The Future of the Art-Work

What kinds of story can only be told with these new technologies? What previously unimagined experiences will emerge from a new approach to creating art?

Media Scores is yet another stepping stone in the evolution of the arts. From the time that the spoken word and music were combined, the human form donned a mask and costume and stepped into the limelight, the architecture of space and the movement of perspective through time all combined to create image, theatre, cinema and beyond—the artistic forms of expression have evolved and merged. The sculpting of space and light in time that is cinema or the interactive narrative of a computer game are not the culminations of artistic practice, but waypoints on the asymptotic approach toward the true Gesamtkunstwerk. In time, all of these media will coalesce into simply the artistic medium. Art will become whole. At the same time, no medium shall become obsolete. The live proscenium stage production will not give way to the holodeck [78], as cinema has not crumbled under the ubiquitous presence of television. We will still prefer at times a painting or curling up with the written words in a good book—regardless of the physical nature of the book's form—or attending a concert in the world's great music halls, passively sitting and watching virtuosi awe us with talent and instruments hundreds of years old. Not every experience will require all modalities and the full range of interactivity and abundance of technological reproduction, though such possibilities will be at the artist's command. The Media Score opens the door for creating the truly grand unified art form that can encompass and subsume all that has come before. Artists can then choose the modalities and conventions of expression that they please. The single score can be realized in any one medium, as much as it can come to life through all of them.

In practice, Media Scores can be the genesis of an ecosystem of hardware devices and software applications that perform the scores: players that can parse the score directly and are themselves the modestly-named content generators in the system architecture (recall Figure 48). Content generators will aid in the production of Media Scores works in forms that rely on conventional techniques, but can also be the home of purely technological performances of the work in their own right.

What is more, the audience will also be afforded the choice of how they wish to consume a work of art. At present, we are seeing this occur at a crude level with the availability of television programming over network broadcasts, as well as on the Internet and mobile devices. However, the model presented through Media Scores gives creators of art the option to give their consumers the potential to choose the modalities and scale in which to experience art works. One audience member may choose to watch the latest episode of *Law and Order* as a 10-minute-long video on their Android tablet, attend the operatic production, or listen to it on their commute home. The transmedia culture of today is one of disparate media. Film and television properties expand into ancillary media, such as DVDs with bonus features, action figures, fast-food meal memorabilia, and spinoff series. These multiple incarnations of story elements are driven more by economic and commercial factors than by storytelling need. This is beginning to change as stories bleed from one medium into another. The transmedia of tomorrow will emerge not solely from bottom lines, promotions, and ratings, but from creative desire. With the entire palette of human expressive capabilities, creators will invent works that suit the variety of ways audiences wish to consume them while maintaining their vision and artistic integrity.

Tools such as Media Scores will also contribute to personal storytelling. A low barrier to entry and intuitive representations will allow individuals of all ages to share their stories by creating art in a multitude of forms, extending the vision of initiatives like Personal Opera (see Section 4.1.1). Anyone can compose in any medium.

Performance media will become malleable and responsive. The infrastructure of theatrical environments will become richly connected, providing a cohesive support for quickly or spontaneously generating meaningful works. Stories and artworks will preserve the unique hallmarks of their creators, and also invite audiences to enter into the world, to participate as much or as little as they please, and to experience a truly personal version of every work not just by the interpretation they bring to it, but by being a part of it: a personal performance of artwork from a shared score. From the individual to the communal, the technological mediation of the coalesced medium will make remote theatrical immersion commonplace. The liveness of a performance is the result of the audience participating in the performance of the work and the sum of audiences and performers distributed in space and possibly time. The distinction between audience, composer, and performer will not dissolve, but will continue to blur.

By proposing a new method for composing Gesamtkunstwerk, it is my hope that Media Scores will open up new opportunities for creating expressive, interactive performances and storytelling experiences. It is a prototype of a language for a union of all of the arts, not a new art form, but an amalgam of the entire history and future of the human capacity for expression—for communicating that which can only exist in the mind and only be felt in the soul.

BIBLIOGRAPHY

[1] Aldeburgh Music. Faster than Sound. [Online]. http://www.fasterthansound.com/spheres-and-splinters/, November 2010.

[2] A. Alexander and N. Collins. "Live audiovisuals" in *The Cambridge Companion to Electronic Music*, eds. N. Collins and J. d'Escriván. Cabridge: Cambridge University, 2007, pp. 126–139.

[3] D. Alighieri. *La Vita Nuova*, trans. A.S. Kline., 1295 (2000-2002). [Online]. http://www.poetryintranslation.com/PITBR/Italian/TheNewLife.htm

[4] *An Optical Poem*. dir. O. Fischinger. Metro-Goldwyn-Mayer. 1938.

[5] Apple Inc. iOS Developer Library. [Online]. http://developer.apple.com/library/ios/DOCUMENTATION/MusicAudio/Conceptual/CoreAudioOverview/WhatsinCoreAudio/WhatsinCoreAudio.html

[6] Aristotle. *The Rhetoric and the Poetics of Aristotle*, trans. R. Roberts and I. Bywater.: McGraw-Hill, 1984.

[7] Artistic Licence Holdings. "Specification for the Art-Net 3 Ethernet Communication Standard." , 1998-2012.

[8] P. Auslander. "Digital Liveness: A Historico-Philosophical Perspective." *PAJ: A Journal of Performance and Art 102*, vol. 34, no. 3, pp. 3–11, September 2012.

[9] P. Auslander. *Liveness: Performance in a Mediatized Culture*, 2nd ed. London: Taylor & Francis, 2008.

[10] L. Austin and T. Clark. *Learning to Compose: Modes, Materials and Models of Musical Invention*. Dubuque: Wm. C. Brown, 1989.

[11] S. Avidan and A. Shamir. "Seam carving for content-aware image resizing." *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH 2007*, vol. 26, 2007.

[12] C. Baugh. *Theatre, Performance and Technology: The Development of Scenography in the Twentieth Century*. New York: Palgrave Macmillan, 2005.

[13] BBC News. BBC News Magazine: Living Online video. [Online]. http://www.bbc.co.uk/news/magazine-21719783, March 2013.

[14] S. Berweck. "It Worked Yesterday: On (Re-)Performing Electroacoustic Music." University of Hudderfield, PhD Thesis, 2012.

[15] D. Bordwell and K. Thompson. *Film Art: An Introduction*, 5th ed. New York: McGraw-Hill, 1997.

[16] C. Breazeal. "Emotion and Sociable Humanoid Robots." *International Journal of Human-Computer Studies*, no. 59, pp. 119–155, 2003.

[17] eds. K. Brougher et al. *Visual Music: Synaesthesia in Art and Music since 1900*. London: Thames and Hudson, 2005.

[18] J. Cage. *Notations*.: Something Else Press, 1969.

[19] K.M. Cameron, P.P. Gillespie, J. Hunter, and J.A. Patterson. *The Enjoyment of Theatre*, 5th ed.: Allyn & Bacon, 1999.

[20] R. Canudo. "The Birth of a Sixth Art" in *French Film Theory and Criticism*, ed. R. Abel.: Princeton University Press, 1988, pp. 58–65.

[21] C. CARDEW. *Treatise*. Buffalo: The Gallery Upstairs, 1967.

[22] M.A. Changizi, Z. Qiong, and S. Shinsuke. "Bare skin, blood and the evolution of primate colour vision." *Biology Letters*, vol. 2, no. 2, pp. 217–221, 2006.

[23] M. Clynes. "Microstructural Musical Linguistics: Composers' pulses are liked best by the best musicians." *Cognition*, vol. 55, pp. 269–310, 1995.

[24] M. Clynes. *Sentics: The Touch of the Emotions*. Garden City, New York: Anchor Press/Doubleday, 1978.

[25] D. Cole. "The Visual Script: Theory and Techniques." *The Drama Review*, vol. 20, no. 4, pp. 27–50, 1976.

[26] R.E. Cytowic. *Synaesthesia: A Union of the Senses*, 2nd ed. Cambridge, Massachusetts: MIT Press, 2002.

[27] P. Debevec et al. "Acquiring the Reflectance Field of a Human Face." *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 145–156.

[28] S. Dixon. *Digital Performance: A History of New Media in Theater, Dance, Performance Art, and Installation*. Cambridge: MIT Press, 2007.

[29] D. Dixon, J. Rogers, and P. Eggleston. "Between Worlds: Report for NESTA on MIT/Punchdrunk Theatre Sleep No More Digital R&D Project." University of Dundee, University of West England Bristol, 2012.

[30] S. Eisenstein. "A Dialectic Approach to Film Form" in *Film Form: Essays in Film Theory*, ed. J. Leyda. New York: Harcourt, 1949 (1929), pp. 46–63.

[31] S. Eisenstein. *The Film Sense*. New York: Harcourt, Brace & Jovanovitch, 1942.

[32] P. Ekman. "An argument for basic emotions." *Cognition and Emotion*, vol. 6, no. 3/4, pp. 169–200, 1992.

[33] "Entertainment Technology - Architecture for Control Networks." American National Standards Institute, ANSI E1.17, 2006.

[34] M.M. Farbood. "A parametric, temporal model of musical tension." *Music Perception*, vol. 29, no. 4, pp. 387–428, 2012.

[35] M.M. Farbood, H. Kaufman, and K. Jennings. "Composing with Hyperscore: An Intuitive Interface for Visualizing Musical Structure." *Proceedings of the 2007 International Computer Music Conference*, San Francisco, 2007, pp. 471–474.

[36] G. Freytag. *Technique of the Drama: An Exposition of Dramatic Composition and Art*, 3rd ed., trans. E.J. MacEwan. Chicago: Scott, Foresman, 1900.

[37] G. Giesekam. *Staging the Screen: The Use of Film and Video in Theatre*. New York: Palgrave Macmillan, 2007.

[38] J.M. Gillette. *Theatrical Design and Production*, 2nd ed. Mountain View: Mayfield, 1992.

[39] Google. Google Drive Support. [Online]. https://support.google.com/drive/answer/49008

[40] G.W.F. Hegel. *Hegel's Aesthetics: Lectures on Fine Art*, trans. T.M. Knox. Oxford: Oxford University Press, 1975, vol. 1.

[41] H.v. Helmholtz. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 2nd ed., ed. A.J. Ellis. New York: Dover, 1885.

[42] hexler.net. [Online]. http://hexler.net/software/touchosc

[43] C.R. Hoffer. *The Understanding of Music*. Belmont, California: Wadsworth, 1967.

[44] Horace. Ars Poetica. 2005.

[45] D. Huron. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge: MIT Press, 2007.

[46] ideacity. ideacity Online. [Online]. http://www.ideacityonline.com/talks/tod-machover-the-toronto-symphony-orchestra-ideacity-2012-crowdsourced-symphony/, 2012.

[47] Infocom, Inc. *Instruction Manual for Zork I: The Great Underground Empire*. Cambridge, 1984. [Online]. http://infodoc.plover.net/manuals/zork1.pdf

[48] E.N. Jessop, P.A. Torpey, and B. Bloomberg. "Music and Technology in Death and the Powers." *Proceedings of the 2011 Conference on New Interfaces for Musical Expression*, Oslo, 2011.

[49] W. Kandinsky. *Concerning the Spiritual in Art*, trans. M.T.H. Sadler., 1911.

[50] S.D. Katz. *Film Directing Shot by Shot: Visualizing from Concept to Screen*. Studio City: Michael Wiese Productions, 1991.

[51] W. Köhler. *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. New York: Liveright, 1947.

[52] T. Kvifte. *Instruments and the Electronic Age: Toward a Terminology for a Unified Description of Playing Technique*.: Solum Forlag, 1989.

[53] R. Le Poidevin. *The Images of Time: An Essay on Temporal Representation*. New York: Oxford, 2007.

[54] J. Leeds. "Nine Inch Nails Fashions Innovative Web Pricing Plan." *The New York Times*, March 2008. [Online]. http://www.nytimes.com/2008/03/04/arts/music/04nine.html

[55] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. Cambridge: MIT Press, 1983.

[56] D.J. Levitin. *This Is Your Brain on Music: The Science of a Human Obsession*. New York: Plume, 2006.

[57] H. Lohner. "The UPIC System: A User's Report." *Computer Music Journal*, vol. 10, no. 4, pp. 42–49, 1986.

[58] J. Machon. *(Syn)aesthetics: Redefining Visceral Performance*. New York: Palgrave Macmillan, 2009.

[59] T. Machover. J. Kim interview "The idea behind *Death and the Powers*" on 18 September 2010. [Online]. http://operaofthefuture.com/2010/09/18/interview-with-tod-machover-1-the-idea-behind-death-and-the-powers/

[60] T. Machover. *A Toronto Symphony*. New York: Boosey & Hawkes, 2013.

[61] T. Machover. *Chomsky Suite*. New York: Boosey & Hawkes, 2011.

[62] T. Machover. *Death and the Powers*. New York: Boosey & Hawkes, 2010.

[63] T. Machover. "Hyperinstruments: A Progress Report." MIT Media Laboratory, Cambridge, 1992.

[64] T. Machover. "Shaping Minds Musically." *BT Technology Journal*, vol. 22, no. 4, pp. 171–179, October 2004.

[65] J. Maeda. "Digital Expression: A Framework for Artistic Creation of Forms On and Using the Computer." University of Tsukuba, Japan, Ph.D. Thesis, 1995.

[66] S. Malinowski. Music Animation Machine. [Online]. http://www.musanim.com/

[67] A. McDowell. Art Directors Guild Art Direction Wiki. [Online]. http://www.gestaltung.hs-mannheim.de/designwiki/files/2523/mcdowell_design%20mandala.pdf

[68] M.R. Miller. "Show Design and Control System for Live Theater." Massachusetts Institute of Technology, MEng Thesis, 2010.

[69] M. Minsky. "Music, Mind, and Meaning." *Computer Music Journal*, vol. 5, November 1981.

[70] M. Minsky. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*.: Simon & Schuster, 2007.

[71] MIT Media Lab. A Toronto Symphony. [Online]. http://toronto.media.mit.edu/?p=508, December 2012.

[72] MIT Media Lab. A Toronto Symphony. [Online]. http://toronto.media.mit.edu/?p=458, December 2012.

[73] MIT Media Lab. A Toronto Symphony. [Online]. http://toronto.media.mit.edu/?p=556, January 2013.

[74] B. Moggridge. *Designing Interactions*. Cambridge: MIT Press, 2007.

[75] N. Moran. "Resisting the lure of the screen." *International Journal of Performance Arts and Digital Media*, vol. 6, no. 1, pp. 77–88, 2010.

[76] W. Moritz. Animation World Network. [Online]. http://www.awn.com/mag/issue1.2/articles1.2/moritz1.2.html, 1996.

[77] W. Moritz. "The Dream of Color Music, And Machines That Made it Possible." *Animation World Magazine*, vol. 2, no. 1, April 1997.

[78] J.H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge, MA: MIT Press, 1997.

[79] D.A. Norman. *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*. New York: Basic Books, 1994.

[80] M. Ondaatje. *The Conversations: Walter Murch and the Art of Editing Film*.: Knopf, 2002.

[81] K. Peacock. "Instruments to Perform Color-Music: Two Centuries of Technological Experimentation." *Leonardo*, vol. 21, no. 4, pp. 397–406, 1988.

[82] M. Perloff. *The Futurist Moment: Avant-Garde, Avant Guerre, and the Language of Rupture*, 2nd ed. Chicago: University of Chicago, 2003.

[83] R.W. Picard. *Affective Computing*. Cambridge, Massachusetts: MIT Press, 1997.

[84] V. Propp. *Morphology of the Folktale*, 2nd ed., trans. L. Scott. Austin: University of Texas, 1968.

[85] C. Radish. Collider.com. [Online]. http://collider.com/hans-zimmer-sherlock-holmes-2-dark-knight-rises-interview/, December 2012.

[86] V.T. Richard. *Norman McLaren, Manipulator of Movement: The National Film Board Years, 1947-1967*. Newark: University of Delaware Press, 1982.

[87] W. Richards. *Anigrafs: Experiments in Collective Consciousness*. Cambridge, Massachusetts, 2007.

[88] W. Richards, J. Feldman, and A. Jepson. "From Features to Perceptual Categories." *British Machine Vision Conference*, 1992, pp. 99–108.

[89] Rob Morris Music. [Online]. http://www.robmorrismusic.com/knowyourexit/

[90] J.A. Russell. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology*, vol. 50, no. 6, pp. 1161–1178, 1980.

[91] D. Ryan. "Between Construction and Spontaneity: Earle Brown and the Visual Arts" at Beyond Notation: An Earle Brown Symposium on 18 January 2013.

[92] T. Sauer. *Notations 21*. New York: Mark Batty, 2009.

[93] S. Schiesel. "Playing the New Bjork Album, And Playing Along, With Apps." *The New York Times*, p. A1, October 2011.

[94] J. Schillinger. *The Mathematical Basis of the Arts*. New York: Philosophical Library, 1948.

[95] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication.*: University of Illinois Press, 1998.

[96] J.M. Sisking. "Grounding Language in Perception." University of Pennsylvania, Philadelphia, 1993.

[97] H. Solís García. "Improvisatory Music and Painting Interface." Massachusetts Institute of Technology, Cambridge, S.M. Thesis, 2004.

[98] C. Spence. "Crossmodal correspondences: A tutorial review." *Attention, Perception, and Psychophysics*, vol. 73, no. 4, pp. 971–995, May 2011.

[99] A. Steggell. "Mind, the Gap: Synaesthesia and Contemporary Live Art Practice." Østfold University College, Norwegian Theatre Academy, Fredrikstad, 2006.

[100] B.E. Stein and M.A. Meredith. *The Merging of the Senses*. Cambridge: MIT Press, 1993.

[101] G. Stiny and J. Gips. "Shape Grammars and the Generative Specification of Painting and Scultpure." *Proceedings of IFIP Congress 1971*, Amsterdam, 1972.

[102] K. Stockhausen. *Nr. 7 Klavierstück XI*. London: Universal Edition, 1963.

[103] K. Stockhausen. *Studie II, Nr. 3 von Elektronische Studien*. London: Universal Edition, 1956.

[104] Style Boston. Plugged In episode 229: Death and the Powers. [Online]. http://vimeo.com/21721641, 2010.

[105] The Metropolitan Opera. Der Ring des Nibelungen. [Online]. http://ringcycle.metoperafamily.org/

[106] P.A. Torpey. "Digital systems for live multimodal performance in Death and the Powers." *International Journal of Performance Arts and Digital Media*, vol. 8, no. 1, pp. 109–123, 2012.

[107] P.A. Torpey. "Disembodied Performance: Abstraction of Representation in Live Theater." Massachusetts Institute of Technology, Cambridge, S.M. Thesis, 2009.

[108] A. van Troyer. "Hyperaudience: Designing Performance Systems for Audience Inclusion." Massachusetts Institute of Technology, SM Thesis, 2012.

[109] P. Vergo. *The Music of Painting*. London: Phaidon, 2010.

[110] R. Wagner. *Götterdämmerung: Third day of Der Ring des Niebelungen*. Huntington Station: Kalmus.

[111] R. Wagner. On Conducting: A Treatise on Style in the Execution of Classical Music. 1869.

[112] R. Wagner. "The Art-Work of the Future" in *Richard Wagner's Prose Works.*, 1895, vol. 1, pp. 69–213.

[113] J. Wang. "Music, Mind, and Mouth: Exploring the Interaction Between Music and Flavor Perception." Massachusetts Institute of Technology, Cambridge, SM Thesis, 2013.

[114] A.P. Witkin and J.M. Tenenbaum. "What is Perceptual Organization For?", 1985.

[115] M. Wood. "The Retractable LED Chandelier." *Lighting & Sound America*, pp. 66–70, February 2010.

[116] M. Woolman. *Sonic Graphics: Seeing Sound*. London: Thames & Hudson, 2000.

[117] M. Wright. "The Open Sound Control 1.0 Specification." , 2002. [Online]. http://opensoundcontrol.org/spec-1_0

[118] J. Zilczer. "Color Music: Synaesthesia and Nineteenth-Century Sources for Abstract Art." *Artibus et Historiae*, vol. 8, no. 16, pp. 101–126, 1987.