

# **MOTION PICTURE MOODS**

*For*

*Geeks and Non-Musicians*

**An Automated, Intelligent Music-Generation System**

*Proposed By*

## **SCOTT DOWNIE**

*B.S.E.E., Kansas State University, 1977*

*B.M., University of Kansas, 1983*

*M.A., University of Oklahoma, 1991*

**Submitted to the School of Fine Arts and the  
Faculty of the Graduate School of the  
University of Kansas**

**In partial fulfillment of the requirements  
For the degree of Master of Arts in Special Studies**

-----  
**Faculty Name, Chair**

-----  
**Faculty Name**

-----  
**Faculty Name**

**Date Defended:** \_\_\_\_\_

**Copyright 2006 Scott Russell Downie**

# Acceptance Certification

The Thesis Committee for Scott Downie certifies  
That this is the approved version of the following thesis:

*Motion Picture Moods for Geeks and Non-Musicians: An Automated,  
Intelligent Music-Generation System*

Committee:

Chairperson

---

---

---

---

Date Approved: \_\_\_\_\_

# Table of Contents

ABSTRACT .....	5
INTRODUCTION: VIDEO MOODS AND SELF-ORGANIZING SOUNDTRACKS ...	6
PURPOSES OF THIS PROJECT .....	7
AUDIOVISUAL "LANGUAGE" .....	7
SATISFACTION .....	7
CONTROL .....	8
FILM-MUSIC PSYCHOLOGY.....	8
CUSTOMER: "STEVE".....	11
PERSONA.....	13
SCENARIO.....	13
APPLICATION NICKNAME: "AL" .....	14
REAL-WORLD CUES AND CLUES .....	15
PROJECT BAR-B-Q'S PRINCIPLES OF GOOD INSTRUMENT DESIGN.....	15
GARAGEBAND 3.....	19
MUSICBED DV .....	20
BAND-IN-A-BOX.....	21
MUSIC MOUSE .....	23
IBM'S QSKETCHER .....	25
SUMMARY OF THE STATE OF THE ART .....	28
MUSICAL STYLES .....	29
AL VERSION 0.1: THE FIRST ITERATION .....	30
SCOTT'S DIY SOUNDTRACK INTERVIEW SCRIPT - 3 FEBRUARY 2006.....	33
<i>GENERAL AUDIO/VISUAL BACKGROUND</i> .....	33
<i>VIDEO SAMPLE: GENERAL IMPRESSIONS</i> .....	35
<i>VIDEO SAMPLE: MUSICAL SUGGESTIONS</i> .....	35
<i>AUDIO/VISUAL CONTROLLER</i> .....	36
<i>AUDIO/VISUAL CONTROL FEEDBACK</i> .....	37
AL 0.1 SURVEY RESPONSE .....	38
<i>GENERAL AUDIO/VISUAL BACKGROUND</i> .....	38
<i>VIDEO SAMPLE RESPONSE/GENERAL IMPRESSIONS</i> .....	38
<i>VIDEO SAMPLE RESPONSE/MUSICAL SUGGESTIONS</i> .....	39
<i>AUDIO/VIDEO CONTROL PREFERENCES</i> .....	40
CONCLUSION: THE FIRST ITERATION .....	40
WEB SURVEY RESPONSES: THE SECOND ITERATION .....	41
THE SURVEY SITE .....	41
THE VIDEO CLIPS AND THEIR LABELS.....	43

SURVEY RESULTS.....	44
SURVEY RESULTS DISCUSSION.....	45
<b>CURRENT SOUNDTRACK APPLICATIONS: THE THIRD ITERATION .....</b>	<b>48</b>
PRE-TEST TERMINOLOGY SURVEY .....	49
GARAGEBAND 3 AND MUSICBED DV .....	49
<i>GARAGEBAND IN USE</i> .....	50
<i>MUSICBED DV IN USE</i> .....	52
<i>SUS SURVEY RESULTS FOR GARAGEBAND AND MUSICBED DV</i> .....	53
M WITH A JOYSTICK .....	53
MUSIC MOUSE .....	55
<b>CONCLUSION: MORE ITERATIONS, PLEASE! .....</b>	<b>56</b>
MEANING AND STRUCTURE.....	56
ONE POSSIBLE AL CONFIGURATION .....	56
USER CALIBRATION .....	58
EXAMPLE STORYBOARD: STEVE MEETS AL.....	59
<i>FEMALE LION</i> .....	61
<i>MALE LION</i> .....	62
<i>LION CUBS</i> .....	63
FUTURE DIRECTIONS .....	63
REFLECTIONS ON MY PROCESS.....	64
MUSIC, EMOTION, AND INTERACTION DESIGN .....	66
CODA .....	69
<b>APPENDIX A: VIDEO CLIPS.....</b>	<b>70</b>
ANGER (BBC MOTION GALLERY #5853-72_65) .....	70
CALM (BBC MOTION GALLERY #819-31-65) .....	70
CUTE (BBC MOTION GALLERY #5075-194_65).....	71
SLAPSTICK (BBC MOTION GALLERY #6046-116_65).....	71
SPOOKY (BBC MOTION GALLERY #450-49_65) .....	72
WALKING (BBC MOTION GALLERY #830-25_65).....	72
WORK (BBC MOTION GALLERY #3873-15_65) .....	73
<b>APPENDIX B: WEB SURVEY RESPONSES .....</b>	<b>74</b>
<b>APPENDIX C: ADJECTIVE-PAIRS SURVEY.....</b>	<b>79</b>
<b>APPENDIX D: MUSICBED DV GENRES AND MOODS.....</b>	<b>80</b>
<b>APPENDIX E: QSKETCHER QUOTATION .....</b>	<b>81</b>
<b>APPENDIX F: COBI VAN TONDER'S <i>SKATESONIC</i> INTERVIEW.....</b>	<b>83</b>
<b>APPENDIX G: GLOSSARY OF TERMS.....</b>	<b>87</b>
<b>REFERENCES .....</b>	<b>90</b>

## **Abstract**

A soundtrack-generation program (nicknamed "Al") is to be designed for an idealized customer named "Steve." Given Steve's interests and skills, this paper looks at a number of the questions—practical, personal, psychological, musical, technological, and cultural—that should be addressed before the soundtrack-generation software can be written. Three methodologies are employed to determine how users who have received little or no musical training might best respond to the opportunity to create their own soundtracks for their home videos.

The answers to these questions may contain implications that are applicable to a broad spectrum of interaction design problems.

## Introduction: Video Moods and Self-Organizing Soundtracks

The title of this thesis is derived from a sheet-music recipe book called *Motion Picture Moods For Pianists and Organists: A Rapid-Reference Collection of Selected Pieces, Adapted to Fifty-Two Moods and Situations* (Davis 1999, 19). It was arranged by Erno Rapée and published by G. Schirmer in 1924. Rapée's tome was one of a number of "fake books" that provided musical raw materials from which silent-movie pianists and organists could assemble accompaniments for the picture-only projections of the day.

Fifty years later, electronic musician Brian Eno seemed to be willing to let the fake books play themselves.

Since I have always preferred making plans to executing them, I have gravitated towards situations and systems that, once set into operation, could create music with little or no intervention on my part.

That is to say, I tend towards the roles of the planner and programmer, and then become an audience to the results. (Eno 1975)

The short-term goal of this project is to take the musical wisdom contained in contemporary "fake books" and inject that wisdom into a digital system that is able to intelligently assemble those musical pieces into soundtracks. The question is this: What sort of music-generation system would encourage a user with no formal training in music to express his or her musical thoughts in the context of composing a soundtrack for a video?

## **Purposes of this Project**

### **Audiovisual "Language"**

Gather clues as to how people who have not been trained to speak the "language" of music (e.g. key signatures, time signatures, reading printed musical notation, etc.) may manipulate music (or "sound design") in structured ways. For example, what mental images form in conjunction with a "classical" sounding piece of music as opposed to a "jazzy", "rock", or "modern" tune? In other words, what sort of codes do listeners attach to various sonic stimuli?

This coding operates on a number of levels. For example, how does the emotional response to a "bassoon" sound differ from the response to a "glockenspiel" sound? Does a rising pitch usually imply movement up and/or to the right (e.g. a piano keyboard)? If a soft flute generates that rising pitch, what is the difference between the reaction it receives and the reaction to a loud, distorted electric guitar rising through the same set of pitches? Broadly speaking, how can we apply "soundtrack psychology" to a practical problem?

### **Satisfaction**

Determine the thresholds of musical "quality" at which untrained musicians express satisfaction. In short, when is the music "good enough"? This project is not concerned with turning untrained musicians into virtuoso soloists or composers of the first rank.

## Control

Seek out the kinesthetic operations and display-control compatibilities that best facilitate translating musical preferences to tangible audio output. Composer, instrument, and musical output constitute an expressive ecosystem. Tinker with one element and the other two will feel the impact. What are the connections?

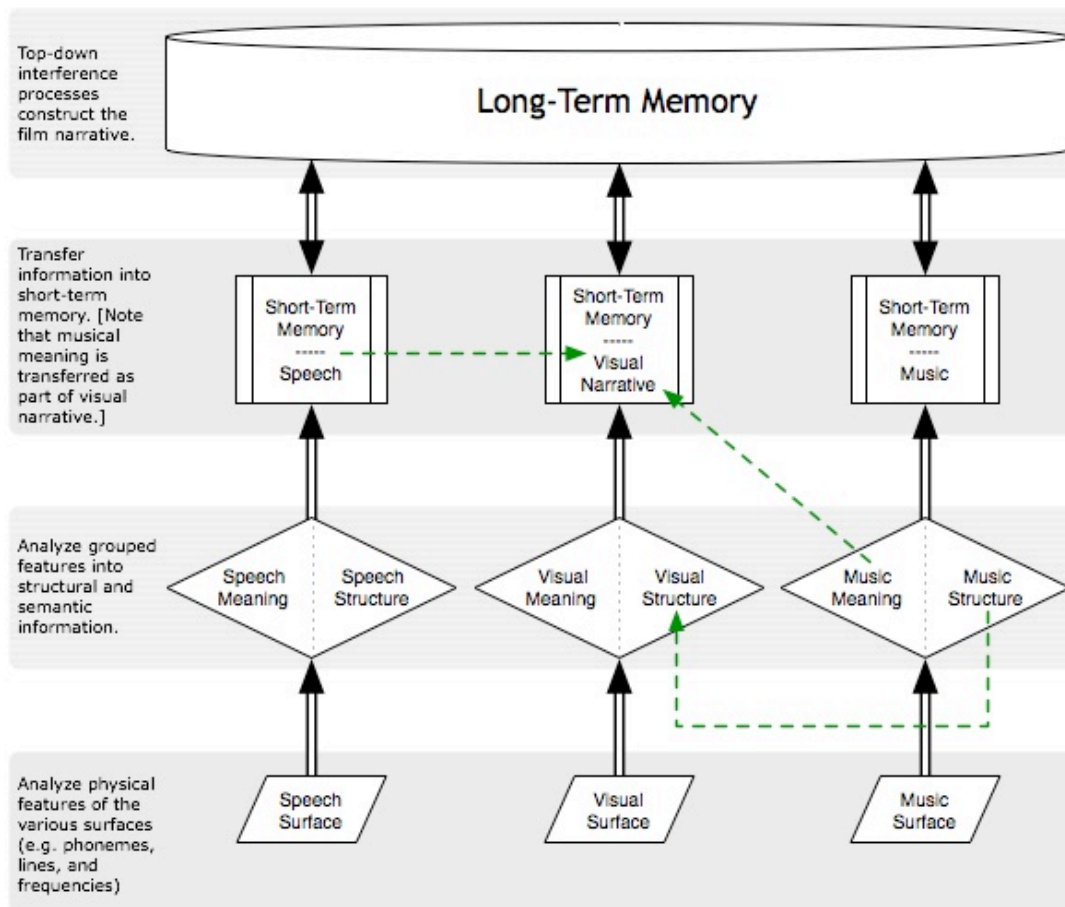
Before digging into the details of finding potential customers and figuring out what those customers want, it is important to research the psychology of simultaneously processing music and video.

## Film-Music Psychology

[F]ilm music scholarship ... and the art of film-score composition ... reveal that the application of music in film is not arbitrary; certain combinations of film and music are more effective than others. From the experimental psychological perspective, the appropriate application of music in film reflects psychological rules. Critics of film music, composers of film scores, and audience members tacitly understand these rules. It is the job of the experimental psychologist to make such implicit rules explicit. (Cohen 2005, 16)

Dr. Annabel J. Cohen has published a number of papers that discuss the psychology of film music. She distinguishes between two aspects of film-music psychology: structure and meaning. "Structure refers to systematic relations among sounds that characterize the style or grammar of music originating within any time period or culture. ... Meaning includes the emotional aspect as well as the associations that the music brings to mind ..." (Cohen 2005, 17). She applies the notions of structure and meaning to a well-regarded theory of how human memory works, and proposes her Congruence–Associationist framework, presented below in slightly simplified form. (In her latest model, there are five channels: Text, Speech,

Visual, Music, and Sound Effects. I trimmed the Text and Sound Effects channels from the diagram in order to focus on the elements at the heart of this thesis.)



*Congruence-associationist framework for understanding film-music communication (Cohen 2001, 259)*

Dr. Cohen's framework proposes parallel information-processing channels. For the purposes of this paper, the most important channels are speech, visual, and music. These channels span four processing levels. The "bottom-up" physical-feature analysis level breaks the input down into components (e.g. phonemes, frequencies, and lines). The "bottom-up" structural analysis

level extracts structural characteristics (e.g. accent patterns, contours, and motions) and meaning characteristics (e.g. associations brought to mind). At this level, information can "leak" from one channel to another. In the diagram above, information about the musical structure may intertwine with information about the visual structure. The short-term memory level puts working memory to use; processing preattended and attended memory while also participating in a dialog with long-term memory. The "top-down" long-term memory level is the vast repository containing our lifetime's worth of experience. It provides the context into which the received information fits.

The mind compares bottom-up and top-down elements, takes the best matches between the two, and produces a working narrative, which is dominated by the visual domain. However, musical meaning and speech may, for example, "leak" into the visual narrative either directly (as indicated by the arrow that runs from Music Meaning to Visual Narrative) or by providing associations that establish inferences in long-term memory (Cohen 2005, 31). This means that musical structure may be used to direct the visual attentional focus in situations where the visual focus is ambiguous (Cohen 2005, 24). This directing of focus can also affect the way an incident is remembered. Dr. Marilyn Boltz demonstrated this effect by accompanying visually ambiguous sections of Alfred Hitchcock's *Vertigo* with "positive" and "negative" soundtracks.

[I]n the positive-music condition of *Vertigo*, participants were more apt to remember the bright, sunny day and bouquet of flowers than the dark sedan and deserted alley, whereas in the presence of negative music, the opposite pattern of results occurred. In addition to effects of selective memory enhancement, music also led to the confabulation of new schema-consistent information in memory. As in the case of recognition accuracy, these false alarms displayed an affect similar to that of

the accompanying soundtrack. In the negative-music condition of *Vertigo*, for example, participants misremembered the presence of an ice pick and open grave whereas in the positive-music condition, an engagement ring, old letters, and a worn photo were more likely to be misremembered. Although response bias effects failed to emerge in these mood congruent conditions, participants were more inclined to identify objects as 'new' when their associated mood conflicted with that of the accompanying music. Given that the affect of these items was inconsistent with one's schematic framework, participants were biased toward believing they had never been presented (2001, 445).

The work of Dr. Cohen and others reminds us that the process of combining music with visuals places quite a load on a composer's cognitive system. This means that a piece of soundtrack-composition software should make every effort to minimize demands on the user's cognition.

A complete psychological theory of film music would necessarily represent the subdisciplines of perception, cognition, social psychology, and emotion. In addition, such topics as lifespan development (how psychological principles and functions change with age), cross-cultural psychology, gender and individual differences would necessarily be included. One can predict a new subdiscipline of multimedia psychology in which the psychology of film music would contribute an essential part. (2005, 19)

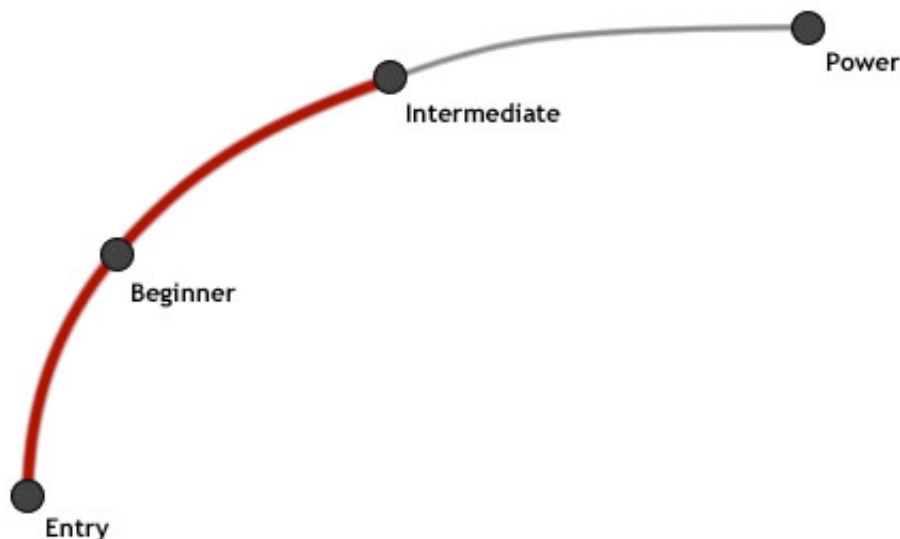
With this basic outline of film-music psychology in hand, we can create a fictional person—a persona—who will be this application's ideal customer.

### **Customer: "Steve"**

When discussing which line of thesis inquiry would be most profitable, Dr. Deron McGee, Associate Professor of Music Theory at the University of Kansas, suggested that the most intriguing findings could be culled from the widest and potentially most interested set of customers: minimally musical amateur soundtrack composers. With the computational horsepower now available in home computers, more and more individuals are using software

applications such as GarageBand to generate soundtracks for their digital home movies. Likewise, directors of films intended for broad release (whether in theaters or on DVDs) may be looking for quick, royalty-free, easy-to-use ways to create either "temp tracks" or completed scores. Could an intelligent music manipulation system help musical novices create more musically satisfying output?

The musical instrument proposed in this thesis—nicknamed "Al"—will allow a neophyte to explore soundtrack possibilities. Depending upon how much musical style knowledge its database contains, Al should remain useful to users up to the intermediate stage. After that, I assume users will have accumulated enough musical expertise to move on to applications that are optimized for detailed musical manipulation, such as Apple's Logic Pro or Digidesign's Pro Tools. Plotting the musical and computer sophistication of potential users on a technical-information user curve (Coe 1996, 47), the thick section identifies the overall musical and technical ability expected of an Al user.



*The intended range of user technical and musical ability for Al*

## Persona

"Steve" is a 41-year-old software engineer who likes to play with his camcorder and video-editing software. He works with this software, but he would not consider himself a power user of that genre of applications. Even though deep down he is a Macintosh geek, he is a realist about living in a Windows-dominated world. His tastes in music are fairly diverse—he has collected many gigabytes of MP3 files—and he has taken music lessons on and off over the years, but he has never had the time to polish his DJ-ing skills or dig into music theory for the purposes of musical composition. Over the years, he has accumulated quite a bit of video footage from family get-togethers, road trips, and sporting events, not to mention the sundry clips of his only child, Scott.



When it comes to adding audio to a freshly edited home video or slide show presentation, he typically lines up a playlist of MP3s that seem pertinent to the subject matter, fading the music at the end of the video. He has become bored with the pick-my-favorite-rock-opus approach because a) after a while, all such soundtracks sound the same, and b) he has used all of his rock favorites more than once already and wants to break out of that musical rut.

## Scenario

Steve may be a geek, but when it comes down to expecting a tool to look good and work well, he is no different from any other software customer.

He is willing to learn some of the basics of music theory to start putting his own stamp on the music he wants to attach to his videos. However, if he did not have to learn anything about key signatures, time signatures, or musical notation, he would be inclined to spend more time playing with the software and, possibly, generating results. After all, video production is just a hobby for him. The musical aspect of creating a soundtrack should be fun!

The soundtrack application should not require special hardware (e.g. a piano keyboard) or exceptional manual dexterity. Although in some respects it should remind the user of a video game, it should not demand every ounce of Steve's concentration to keep track of the audiovisual environment. The application should encourage play, but in a relaxing, non-competitive way. The application should also generate output with a minimum of tinkering and fuss. Steve might be interested in music, but he does not want to practice for hours and hours to master the performance or compositional aspects of the program. It was the threat of practice that drove him from the piano bench years ago!

### **Application Nickname: "Al"**

To facilitate discussion of this project, I have bestowed this mythical soundtrack-generation system with the name "Al." If you interpret the "l" in "Al" as a lower-case "l," you understand that the name is in honor of Alfred Newman, long-time head of the 20th-Century Fox music division and winner of nine Oscars®. If you interpret the "l" in "Al" as an upper-case "i," you understand that the name also alludes to the much-maligned notion of "artificial intelligence."

## Real-World Cues and Clues

I began this project using guidelines and applications from the commercial world as my starting points. First, I will cite the pragmatic wisdom of Project Bar-B-Q. Project Bar-B-Q is a yearly conference that brings designers, engineers, musicians, and business folks to a big ol' Texas brainstorm session, where they puzzle out the future of multimedia. Then I will provide brief overviews of two shipping applications—GarageBand and Musicbed DV—that are intended to generate soundtracks, as well as one application—Band-in-a-Box—that automatically generates music in recognizable styles. I will also mention a long-discontinued-but-still-beloved program—Music Mouse—whose appeal has never faded. Finally, I will discuss an intriguing piece of vaporware: IBM's QSketcher.

### Project Bar-B-Q's Principles of Good Instrument Design

At Project Bar-B-Q's 2004 Texas hoedown, one of their working groups enunciated nine principles of good instrument design. The following extended excerpt describes Project Bar-B-Q's nine principles (Project Bar-B-Q 2006).

#### *Principle 1 - Flow*

The core principle of good digital instrument design is that it should be designed to allow the user to achieve a state of “flow” as defined by psychologist and philosopher Mihaly Csikszentmihalyi in his groundbreaking theories on creativity and quality work. ...

These are the characteristic dimensions of the flow experience:

1. Clear goals: an objective is distinctly defined; immediate feedback: one knows instantly how well one is doing.
2. The opportunities for acting decisively are relatively high, and they are matched by one's perceived ability to act. In other words, personal skills are well suited to given challenges.
3. Action and awareness merge; one-pointedness of mind.
4. Concentration on the task at hand; irrelevant stimuli disappear from consciousness; worries and concerns are temporarily suspended.
5. A sense of potential control.
6. Loss of self-consciousness, transcendence of ego boundaries, a sense of growth and of being part of some greater entity.
7. Altered sense of time, which usually seems to pass faster.
8. Experience becomes autotelic: If several of the previous conditions are present, what one does becomes autotelic, or worth doing for its own sake.

#### *Principle 2 - Progressive Sophistication*

... According to [the principle of progressive sophistication or progressive complexity], the design should invite the novice, but reward the sophisticated user by progressively revealing more and more advanced functionality. The initial experience of the instrument should not be intimidating, but it should allow the user to grow into deeper and more complex uses and encourage the user to move to ever-higher levels of skill. ...

#### *Principle 3 - Activity-Centered Design*

A close correlation to the principle of progressive sophistication is the principle of activity-centered design. By presenting a simpler interface initially, a problem is created that only certain functionality is available. This problem, however, can be overcome by activity-centered design, which presents only those features required to achieve a particular activity. The interface should be based around the completion of activities as opposed to accessing specific functions. ...

#### *Principle 4 - Responsiveness*

Like traditional instruments, digital instruments should provide the user immediate, and possibly even physical feedback. Such feedback should be consistent and non-arbitrary so that cause and effect is always maintained. ...

#### *Principle 5 - Noodle-worthiness*

The instrument should be capable of just being played without the intrusion of any technology. It should invite exploration and "noodling."

#### *Principle 6 - Ergonomicalicious*

Traditional instruments have developed their ergonomics over centuries through trial and error. Digital instruments should apply this knowledge as well as the modern knowledge of human-factors engineering to develop ergonomically appropriate digital instruments. ...

#### *Principle 7 - Capture Everything*

Digital instruments should attempt to capture and record as wide an array of human performance parameters and inputs as possible. With physical instrument interfaces, finger pressure, body tapping, torsion, licking, slapping, scraping, banging and blowing should all be captured digitally for later editing and prosecution. ...

#### *Principle 8 - Positive Emotional Response*

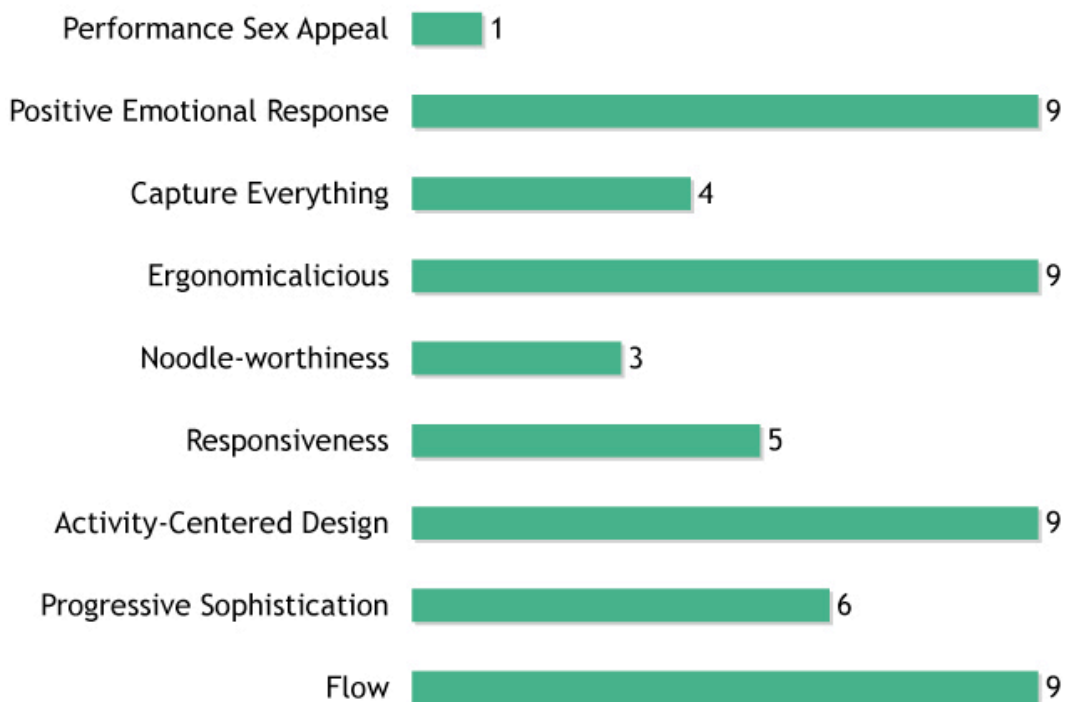
The instrument should evoke an immediate positive emotional response and continue to gratify throughout its useful life. Physical instruments should take industrial design seriously and software instruments should be designed with both usability and beauty in mind. ...

#### *Principle 9 - Performance Sex Appeal*

Digital instruments need to look cool and be capable of usage on stage where the audience can see and appreciate the connection between the performer's actions and the musical result. By way

of comparison to traditional instruments, the guitar is the obvious example of performance sex appeal. It provides physical freedom of movement and presents the actual musical performance (fingers on a fretboard) directly in front of the audience. ...

To refine my initial focus on Steve's expectations, I created a matrix—using Project Bar-B-Q's categories—to indicate degrees of interest in the facets of my target customer's relative interest in AI.



*Scott's Project Bar-B-Q Breakdown, as applied to AI*

*(The scale—from low value (1) to high value (9)—is my own arbitrary creation.)*

As the chart indicates, I began with the assumption that the most important principles to associate with AI should be flow, activity-centered design, ergonomicaliciousness, and positive emotional response. Since AI will be a relatively blunt tool, interactively speaking, it should be responsive. It does not have to be the most sensitive or high-strung of systems. AI is probably not something a user might be inclined to "noodle" with (although Steve

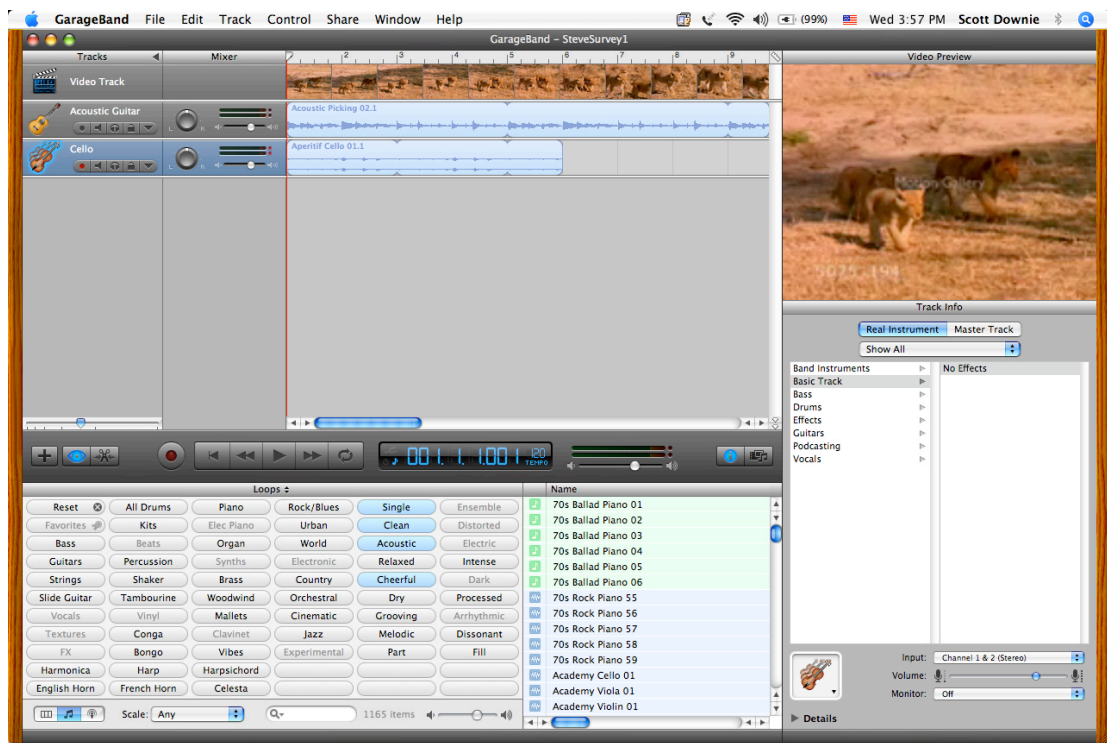
should be able to do so) nor is an AI performance intended to be a form of visual entertainment in its own right. Of course, these initial assumptions are all subject to change if users clamor for a different set of priorities!

### **GarageBand 3**

Here is an excerpt from Apple's promotional copy for the "Movie Score" functionality that made its debut with GarageBand 3. (Apple 2006)

The new video track in GarageBand makes it easy to create an original score for your movie. And don't worry about your musical talent—or lack thereof. Just use GarageBand's included loops, or try a combination of loops, software instruments, or any previous audio recordings you created. You can even use GarageBand to add cinematic "foley" sound effects such as footsteps and creaking doors.

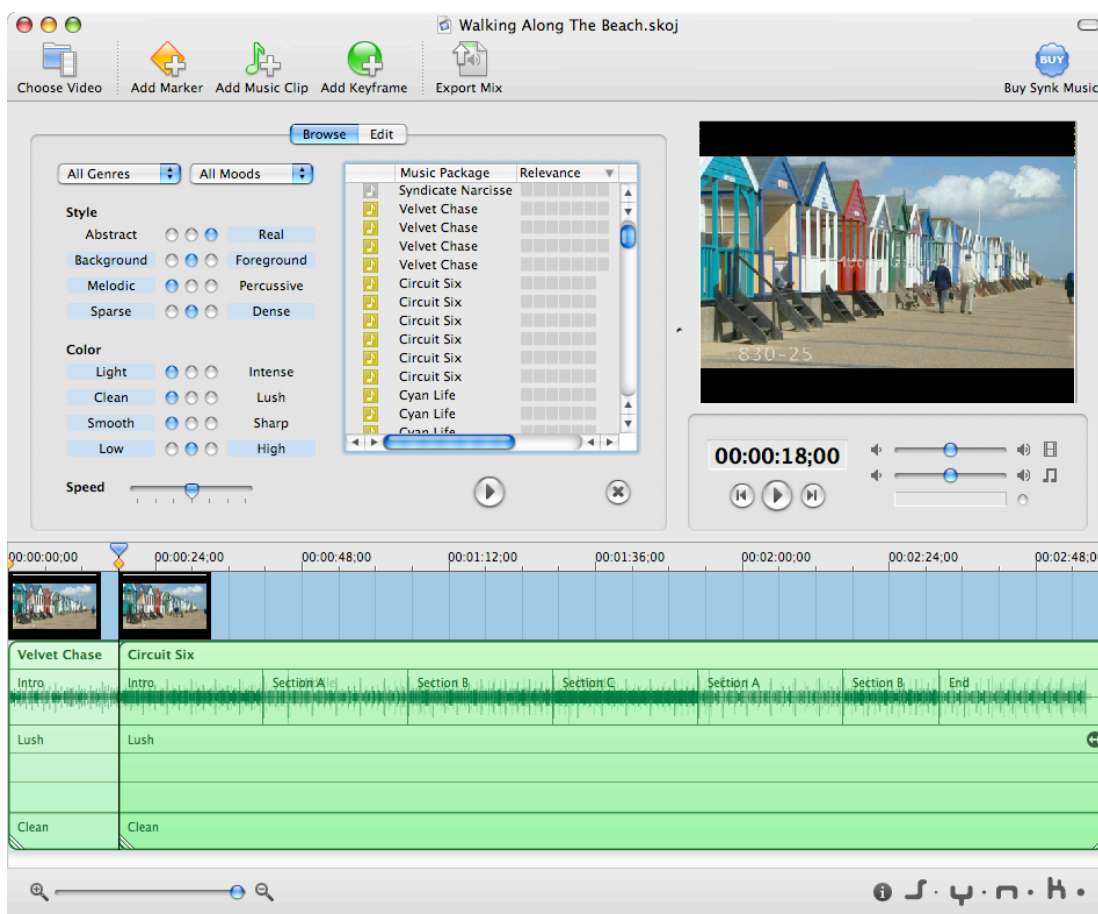
GarageBand 3 allows non-musicians to assemble a soundtrack by dragging and dropping prefabricated musical chunks with names like "70's Ballad Strings 02" and "Edgy Rock Bass 03." The problem is, nobody has any feeling for what "70's Ballad Strings 02" may bring to the musical endeavor, or how it may fit with "Edgy Rock Bass 03." This seems to be true of virtually all of the soundtrack-generation systems currently available: The user must play drag-and-drop roulette with a pile of cryptically named audio snippets.



*GarageBand 3 in "Movie Score" mode*

## Musicbed DV

Musicbed DV is billed as "an easy-to-use music-customization application designed for visual media professionals." It uses buttons and sliders to assign levels of musical relevance to its own proprietary library of film music building blocks. Unfortunately, the user is asked to associate musical meanings to labels like "Velvet Chase," "Circuit Six," and "Cyan Life."



*Musicbed DV 1.2*

## Band-in-a-Box

Since I sprinkle references to Band-in-a-Box throughout this paper, I will briefly describe it. Here is the official Band-in-a-Box blurb.

Just type in the chords for any song using standard chord symbols (like C, Fm7 or C13b9), choose the style you'd like, and Band-in-a-Box does the rest. Band-in-a-Box automatically generates a complete professional quality arrangement of piano, bass, drums, guitar and strings in a wide variety of popular styles. (Jazz, Pop, Country, Classical and more.) (PG Music 2006)

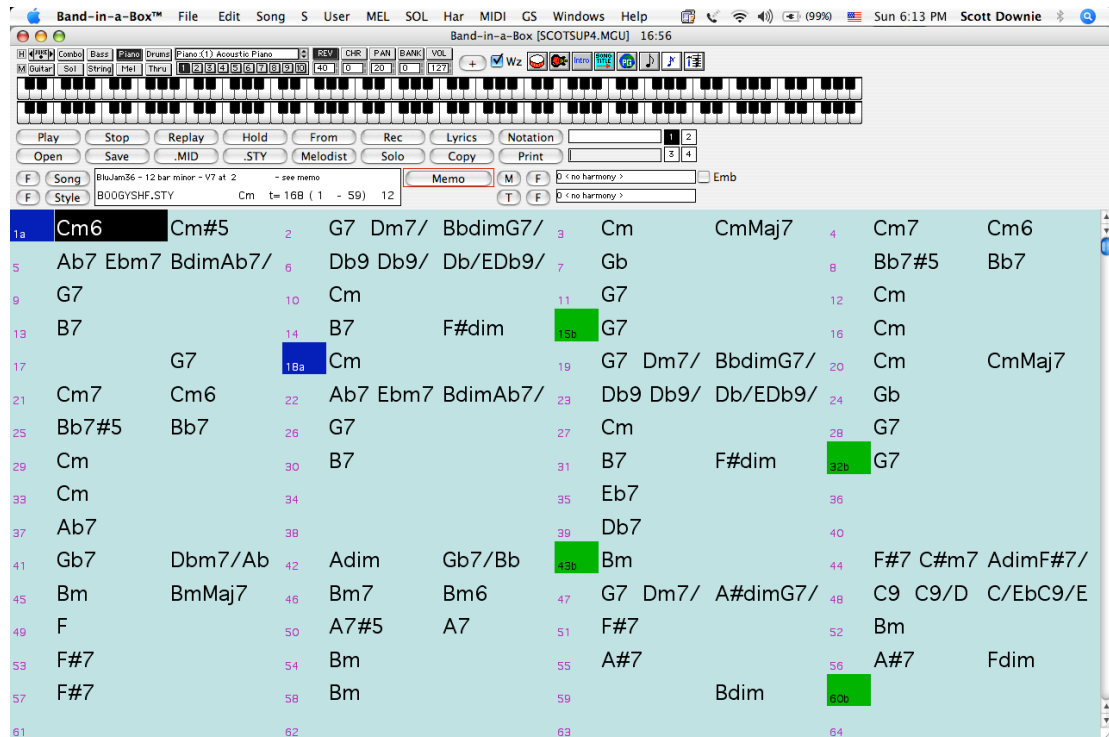
Band-in-a-Box uses Style, Melodist, and Soloist "sets" that contain not only basic musical "sequences" but also meta information that allows Band-in-a-Box to string sequences together into longer, stylistically varied musical phrases. The Melodist and Soloist sets contain not only the harmonic and rhythmic information of a Style set but also "melodic" data. Here are the announcements of the latest Melodist and Soloist releases. Notice that, coincidentally, these new releases focus on soundtrack music!

Soloist Set #18 - Soundtrack Soloists - Soloist Set 18: Soundtrack Soloists adds 3 great new background Soloists - "Peaceful" (slow), "City" (medium) and "Exciting" (fast) - to any of your Band-in-a-Box soundtracks or compositions. The Peaceful New Age Soundtrack Soloist will create peaceful, tranquil solos for any composition or soundtrack project. City Soundtrack Soloist will generate a medium tempo corporate type of solo. The Exciting Soundtrack Soloist generates busier, high-energy solos. ... Soundtrack Soloists are ideal for adding musical content to your soundtracks or music compositions. With the press of a button, you can easily add a great solo of any length to your tune or soundtrack. These Soloists create solos for any chord progression that you type

Melodist Set #6 - MOR Pop/Rock & Soundtracks - This exciting new set adds 3 brand new Melodist categories to your Band-in-a-Box. The MOR (Middle of the Road) Melodist can instantly create tunes in an up-tempo even 8ths Pop-Rock style. Start with your own chord progressions or have Band-in-a-Box generate a whole tune for you at the press of a button. The Soundtrack Melodists includes 2 very different Melodist styles. The New Age Slow Soundtrack Melodist will create tranquil, harmonious background melodies for your movie or production, while the High Energy Soundtrack Melodist will generate exciting, electrifying compositions. These 2 Soundtrack Melodists are perfect if you are looking for copyright free music for use in your home movies, corporate presentations, voice-over backgrounds or other projects. (PG Music 2006)

While Band-in-a-Box may be related to AI as far as algorithmic-composition philosophy goes, the application itself is geared toward users who know

something about chord progressions, time signatures, General MIDI sound sets, and other musical details that AI should handle invisibly. A busy screen displaying many musical control options is not really what my persona Steve is looking for. Therefore, I did not evaluate it alongside GarageBand and Musicbed DV.



A screen shot of the main Band-in-a-Box 12 for Macintosh window

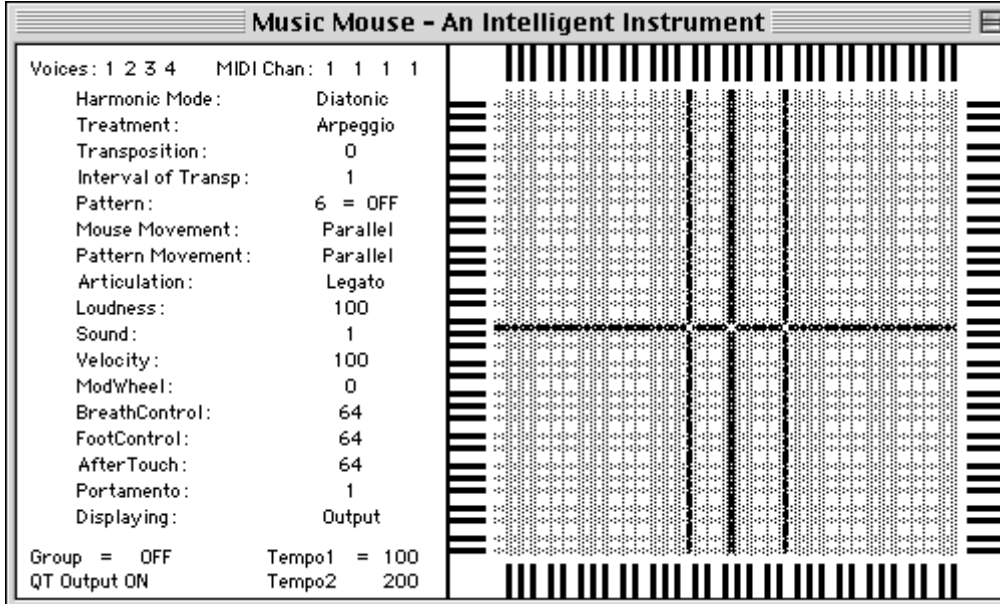
## Music Mouse

Music Mouse first appeared in 1986 and was updated most recently in 1998. At some point in the distant past (probably the late 1980s), Keyboard Magazine wrote the following.

The product of desire to move sounds intuitively ... Music Mouse is not a sequencer, but rather a mouse-driven music generator ... equally useful to trained musicians and rank amateurs. This

program will help you explore and develop modal compositions, and then let you carry those ideas far beyond the capabilities of your ten fingers. Has performance controls that allow you to change almost every aspect of the music you produce... You can really get some great progressions ... crank up the tempo so fast that individual notes become indiscernible ... produce really nice simulated filter sweeps and other timbral changes. This is more than a toy. Whether you're using it as your only musical tool for experimental things or creating background washes for more traditional music, this product is definitely worth \$\_\_\_\_\_ of your Earth dollars. (Spiegel 2006)

With Music Mouse, you simply start the program and begin moving the mouse (or running your finger around a track pad): The music begins immediately! You can pick various harmonies (Chromatic, Diatonic, Pentatonic, Middle Eastern, Octatonic, and Quartal) and various performance styles (Chord, Arpeggio, Line, and Improvisational) to alter the basic musical character of your mouse manipulation.



*Music Mouse, circa 1998*

## IBM's QSketcher

In March 2006, I ran across an IBM project that also attempted to use software to facilitate soundtrack creation. Alas, the IBM project ended in 2001, apparently without producing usable software. Nonetheless, given the similarity between the QSketcher project and my own, sections of the QSketcher introduction are pertinent.

The music composition process has evolved over millennia as a balance of opposites: inspiration versus perspiration, broad formal approaches versus minute detailed work, macro-level (or structural) conceptualization versus micro-level (e.g., note-level) editing, and so on. When composing in the traditional ways, one moves frequently from mode to mode: an idea pops up, one captures it, and then thinks critically about and develops it. Another idea pops up, and the process iterates. One can easily shift focus to thinking about the entire composition and how a small motif relates to the whole, for example.

...

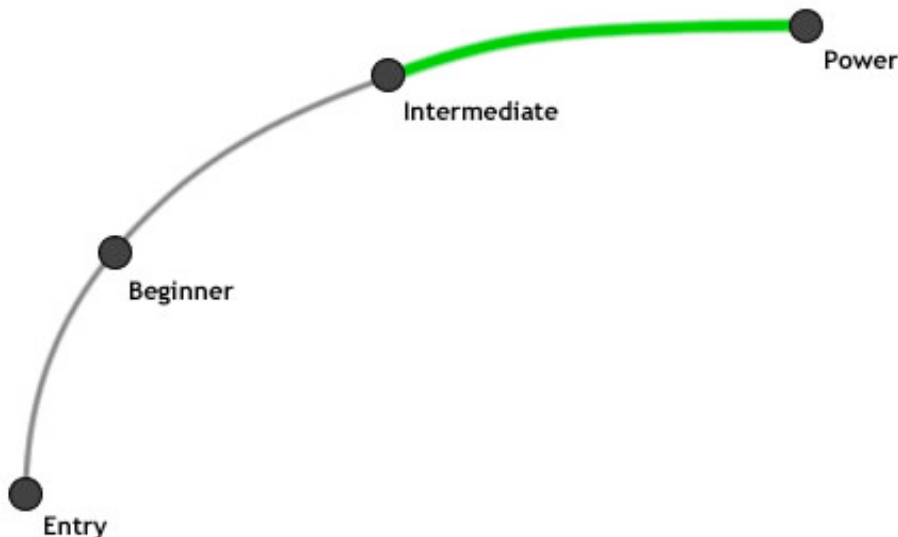
[Th]e environments for composing music that have evolved over five decades in academic and commercial settings still fall short in their ability to manipulate music directly in terms of musical concepts. Computer languages for composing music, for example, typically model limited structural aspects of music through a score, e.g. instrument, part and score, and provide a suite of algorithms to manipulate that score. The score is often expressed in physical terms, such as frequency and amplitude, and the tools provided for its manipulation are often borrowed from computer science with few changes. This detachment from the intuitive musical concepts in which composers think, and from the musical experience itself, places an unnecessary cognitive burden on the composer. Computers are supposed to make work easier, not harder!

...

Interestingly, these dichotomous modes [inspiration/perspiration, capture/manipulate, and macro/micro editing levels] are also

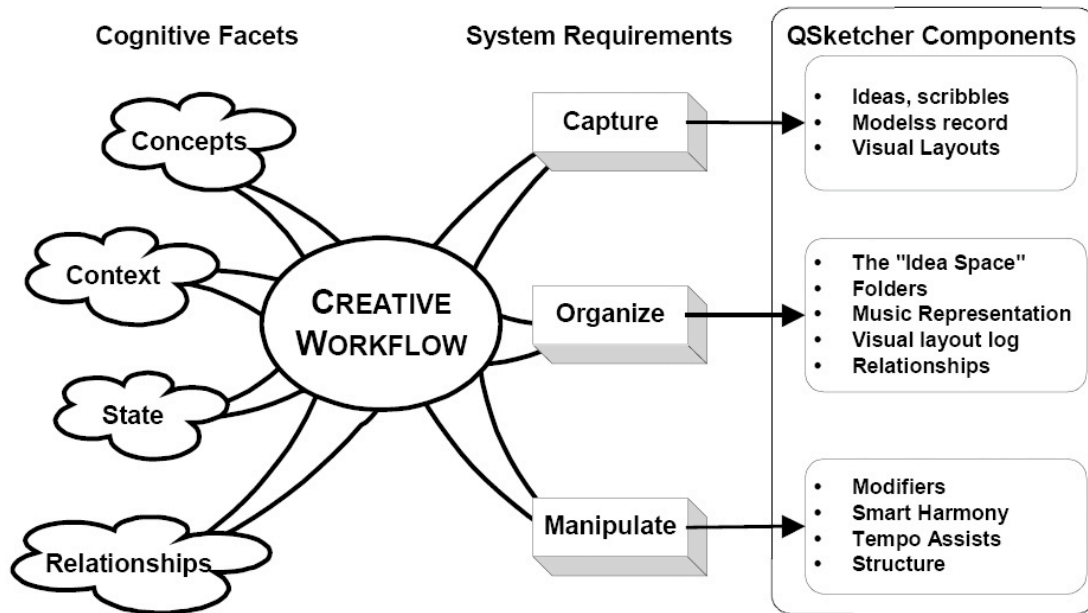
present in a wide variety of creative tasks—writing, drawing, preparing presentations, as well as more technically-oriented activities such as software design and development, architecture, and even the act of research itself. The concepts uncovered during the QSketcher project are important in these domains, and many of the solutions implemented would prove quite useful there. Several of the authors are currently adapting these techniques to other areas. In fact, we believe that the “tip of the iceberg” of an important research area in HCI is emerging: developing information technology tools to support creative work. (Abrams 2001, 1-2. The complete extended excerpt is printed in Appendix E beginning on page 81.)

QSketcher was designed to be used by the film music professional, not the multimedia beginner. For example, QSketcher expects that the user will be entering his or her own musical ideas and manipulating those ideas in reasonably original ways. AI, on the other hand, offers the user a library of prefabricated ideas—no keyboard skills necessary—and then combines those ideas in relatively standard—even “clichéd”—ways. I indicate in green where I believe QSketcher would land on the technical-information user curve below.



*The intended range of user technical and musical ability for QSketcher*

The QSketcher team diagrammed the basic elements in their system as follows.



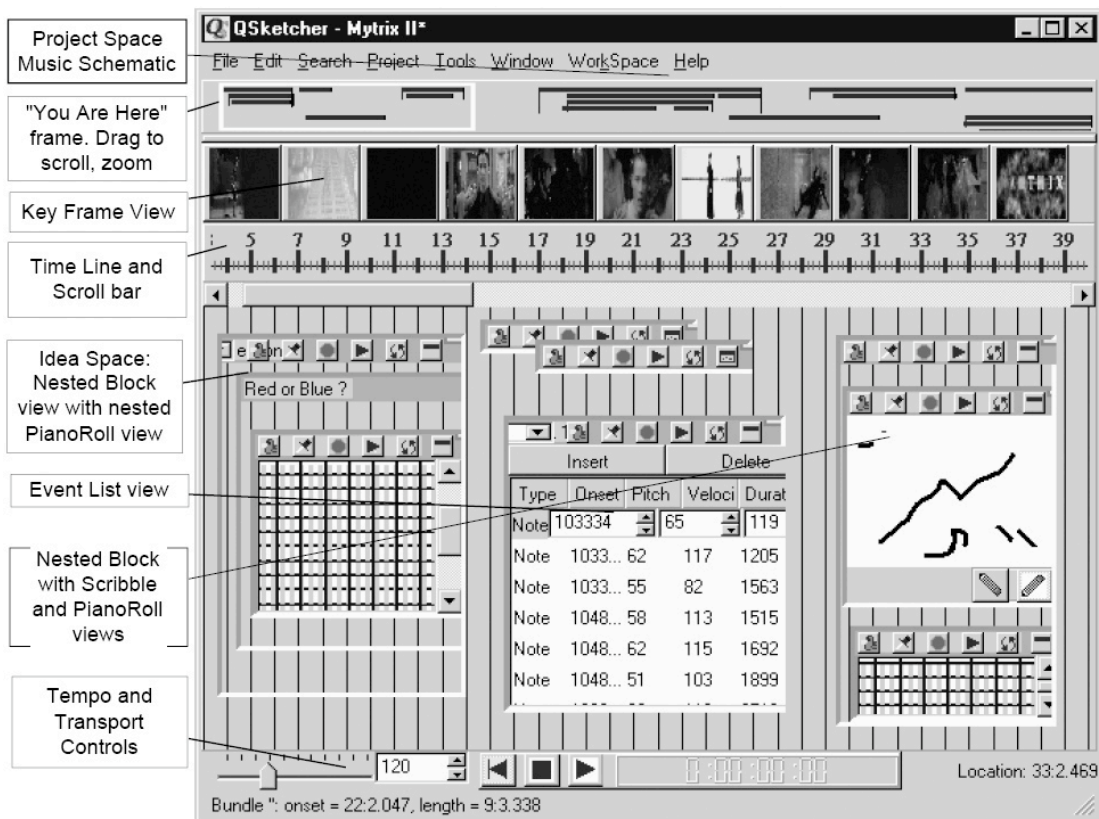
*Overall cognitive aspects, system requirements, and the related QSketcher components (Abrams 2001, 3)*

Comparing QSketcher to AI, the Capture phase—in which the application records the (sophisticated) user's musical input—is not necessary in AI, since AI draws upon libraries of extant material (e.g. Apple Loops, Apple Jam Packs, Band-In-A-Box Style files, etc.). [Ideally, AI would take advantage of the network, allowing the application to download soundtrack style information from a server in the same way iTunes currently facilitates downloading music and video files.]

Where QSketcher makes it very easy for the user to drill down to the musical details in the Organize and Manipulate stages, AI would allow the user to organize and manipulate material, but from a "higher altitude." If a user wanted to dive into the lowest-level details of an AI-generated score,

the user would probably export a project created with AI to a more powerful music editor, such as Logic Pro. In fact, Apple's GarageBand and Logic Pro, with their ability to share musical projects, already work together that way.

One possible arrangement of QSketcher's Idea and Project spaces was mocked up for the IBM report.



*QSketcher's Idea Space and Project Space (Abrams 2001, 4)*

## Summary of the State of the Art

My inspection of updated and new software packages (such as GarageBand 3, Musicbed DV, and Apple's Soundtrack) as well as my discovery of soundtrack-generation projects of the past—most notably IBM's "QSketcher"

project—reinforced my sense that I was on the right track. The up-front complexity of these software packages suggested that, comparatively speaking, I was more interested in working "backward" from the non-musical user to the computer instead of employing the standard design-and-engineer-outward-starting-with-the-computer approach. While all of these applications have much to recommend them, they place heavy demands on users who have little or no musical experience.

## **Musical Styles**

The musical scope that AI is expected to cover is relatively narrow. The initial thrust of software development aims to produce 18th- through 21st-Century Western tonal music. Persona Steve is most familiar with this type of music. There is no technical reason why AI could not synthesize styles from other cultures and historical periods. However, one might as well use the most commercially viable musical vocabulary to initiate the development process. If users demand them, AI 2.0 can include the Palestrina, Raga, and Gamelan upgrade modules.

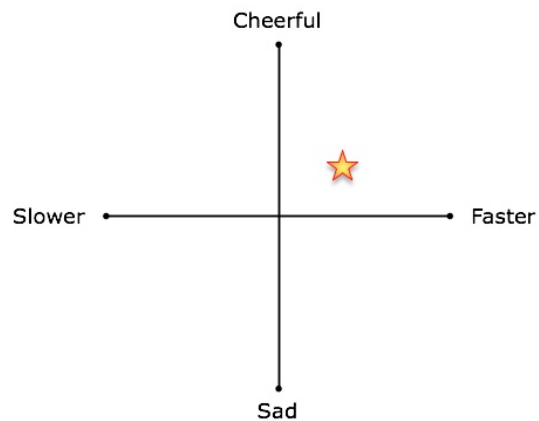
The ideal solution would be to extract the musical intelligence contained in Band-in-a-Box style definitions, GarageBand Jam Packs, and online music-analysis services, using that intelligence to weld together and modify the musical phrases on the fly. Assuming ownership and copyright issues were properly addressed, this would enable the user to immediately call upon a significant body of musical styles. For users more musically sophisticated than persona Steve, an advanced version of AI (marketed as "AI Pro") could also enable style creators to use existing tools to create new style files.

## AI Version 0.1: The First Iteration

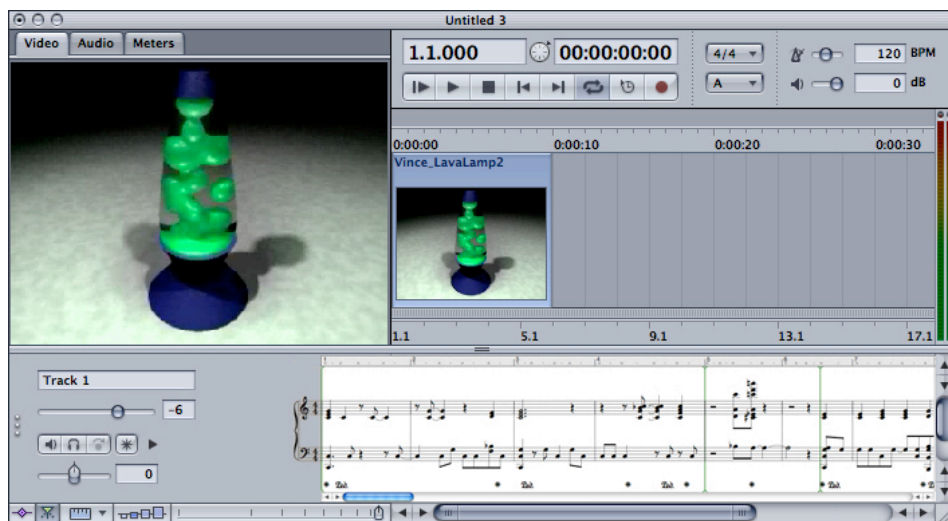
Even though the most important aspect of this proposal has very little to do with particular hardware/software combinations, the interaction design process has to start somewhere. Therefore, I assembled the following cluster of images in an attempt to give people a better idea of my proposal.



*Game Controller*



*Controller Input Map*



*Video/Audio Synchronization Software*

Al 0.1 consisted of two main elements: a standard game controller and software running on a Macintosh computer. To generate a soundtrack, persona Steve would watch the video stream and use the game controller to dynamically register musical choices. Manipulating the music would be like playing a game. In the simplest scenario, a movie would play on the screen. As the user moved the joystick across the input map illustrated above, Al would intelligently assemble and play music—with varying degrees of cheerfulness and sadness at varying tempi—in real time.

On a real-world computer, the rendering of the audio might be at a lower resolution during soundtrack creation than it would be once the soundtrack is complete and ready for the final output. Even relatively picky users might not find this medium-quality real-time audio track to be problematic: Dr Shin-ichiro Iwamiya discovered that users are not as fussy about audio quality if a significant fraction of their attentional bandwidth is occupied with video.

When the sound quality is reduced by band-limitation, the addition of visual information decreases auditory sensitivity to the reduced sound quality. Therefore, the visual information appears to compensate for the poor sound quality of the audiovisual stimuli. The sound quality of the audiovisual stimuli does not appear to affect visual processing. (Iwamiya 1994, 152)

For an input device, I took the well-known PlayStation game-pad layout as a starting point. The Logitech "Dual Action" USB game controller sports one joystick, one four-way arrangement of buttons, and three single-action pushbuttons per hand. With this game controller, left-hand joystick controls could, for example, be assigned to the "emotional" aspects of the music: an upward motion makes the music "happier"; downward, "sadder"; rightward, faster; leftward, slower.

Additional programming effort could enable the application to make finer psychological and musical distinctions, the right hand could control the shape of the melody—which, as it is generated, would of course fit the automatically generated harmony—as well as the texture and timbre of the work-in-progress. Perhaps some of a game controller's other "shooting" controls would be assigned to tasks like generating markers or synchronization points, allowing a well-coordinated user to "Mickey Mouse" the action.

Allow me to reiterate the provisional nature of the hardware scenario presented above. Although it is hard (and perhaps counterproductive) to completely avoid presenting some traditional music-making (e.g. piano-keyboard, frets-and-strings, pipe-with-holes) and game-playing (e.g. joystick, computer-keyboard, game pad) controller options, I want to initially frame the choice of music production as broadly as possible. For example, one person I have asked about music making said she would like to make music by swinging her legs. What kind of music that would generate, I have no idea, although Cobi van Tonder's *Skatesonic* project may offer one answer. (For information on *Skatesonic*, see Appendix F on page 83.) By simply letting users' imaginations run wild and iterating throughout the process, we run a better chance of coming up with a truly novel solution to a difficult problem.

As far as musical content is concerned, it would probably be best (in the interest of time) to narrow the focus to only two typical soundtrack styles: Romantic and Modern, for example. Ultimately a real product would include many other options—jazz, rock, hip hop, ambient, etc.—but the addition of other styles can wait until the first wave of satisfactory preferences have been determined.

The two suggested categories—"Romantic" and "Modern"—would contain an inventory of musical "licks" (a.k.a. "clichés") associated with a classification scheme ("Cheerful," "Sad," "Excited," "Phlegmatic," etc.). This approach is similar to that used by silent-movie-era musicians in compiling books of musical cues. In our case, the musical content for the testing phase could be drawn from all sorts of sources: soundtrack CDs, MIDI files, Band-in-a-Box styles, GarageBand Jam Packs, Apple Loops, online music information databases, etc. These snippets will be associated with the aforementioned classification scheme in a matrix of styles, speeds, key signatures, time signatures, and whatever else seems to be appropriate.

### **Scott's DIY Soundtrack Interview Script - 3 February 2006**

To test the assumptions contained in my AI 0.1 prospectus, I created the following interview script. My goal was to be able to complete an interview using this script in about 30 minutes, 45 minutes tops.

To supply the video I need to perform the testing, I discovered a wonderful online library of short, stock video clips from the BBC's Motion Gallery. Appendix A, beginning on page 70, describes the seven clips I used throughout this survey.

### *General Audio/Visual Background*

1. Do you have any hearing impairments (e.g. deafness, tinnitus, etc.)?
2. Do you play an instrument or sing?
3. Have you had any formal training in music theory or composition?
4. Do you have a general idea of what a "time signature" or a "key" might be?
5. Have you had any formal training in film studies or video editing?

6. How much do you want to know (in depth) about "music"? If you want to learn more, how much time would you be willing to spend per week?
7. How many hours per week do you spend listening to music?
8. What sorts of music do you listen to (e.g. jazz, pop, rock, classical)?
9. When you watch movies, have you ever paid attention to the soundtrack? If so, under what circumstances?
10. How many soundtrack composers can you name?
11. Have you ever wanted to set your own home movie or slide show presentation to music? If so, what sort of movie/slide show was it and what sort of music or musical instruments would you have used?
12. If a computer could assist you in creating a soundtrack for your videos and slide shows, how good would the results have to be to maintain your interest in a soundtrack-generation product?
  - Perfect - indistinguishable from what you would get with a Hollywood blockbuster
  - Very good - an experienced movie goer would have to listen closely to tell the difference between the computer-generated music and the "real thing"
  - Good - An average Joe could tell that the computer-generated "orchestra" wasn't real, but the music still got the job done
  - Acceptable - An average Joe could tell that the music was pretty cheesy, but the music is still better than what an amateur could come up with using other means (e.g. MP3s, CD rips, GarageBand, Band In A Box, or other musical sources)
13. What sort of movie-making hardware are you familiar with (e.g. camcorder, digital camera, film camera, etc.)?

14. What sort of computer controlling hardware are you familiar with (e.g. typewriter keyboard, MIDI piano keyboard, game controllers, etc.)?
15. What sort of computer software are you familiar with (e.g. iMovie, Photoshop, multimedia-authoring software, etc.)?
16. Do you play video games on your computer?
17. How much would you pay for soundtrack-generation software?

*Video Sample: General Impressions*

Show the user a very short video clip with either just dialog or else no audio at all. The video clip should probably not suggest a nuanced interpretation.

1. How did this clip make you feel (e.g. happy, sad, bored, amused, etc.)?
2. Do you think it would be enhanced with a soundtrack?
3. If so, what are the main qualities you would want to hear in that soundtrack (e.g. happiness, ironic counterpoint, humor, etc.)? In other words, would you want the soundtrack to reinforce the feeling you got from watching the video or would you want the soundtrack to emphasize a feeling you did not get from the video?
4. Do any specific pieces of music come to mind as being potential soundtrack material?
5. Does the sound of any particular instrument (or just a sound of some type) come to mind as being potential soundtrack material?

*Video Sample: Musical Suggestions*

Play very short audio clips—no longer than ten seconds—and ask which clip best fits the video clip we have just seen. The video and audio will not require tight "cartoonish" synchronization. In the first round of testing, the

audio sample comparisons will focus on the timbre of the sound. Subsequent iterations will introduce more "musical" content (e.g. melody, harmony, rhythm). The idea, which may well be invalid, is to try to distinguish the effects of musical timbre (which are immediately apparent) from the effects of musical content (which takes time to develop in the listener's mind).

1. [Play the video clip with a very simple audio sample #1 (some orchestral stringed instrument) and then play the same video clip with audio sample #2 (some orchestral brass instrument).] Which audio sample do you think best fits the video? Or is neither sample satisfactory? Do you have any comments?
2. [Take the preferred sample from Step 1 and use the same procedure to compare that sample to a third sample (some orchestral woodwind instrument).] Which audio sample do you think best fits the video? Or is neither sample satisfactory? Do you have any comments?
3. [Take the preferred sample from Step 2 and use the same procedure to compare it to a fourth sample (some synthesizer sound).] Which audio sample do you think best fits the video? Or is neither sample satisfactory? Do you have any comments?
4. Do you want to reconsider any of the samples or is your preference clear?
5. This was a very simple example. What do you think the next step would be?

#### *Audio/Visual Controller*

1. How do you want to provide input to a soundtrack-generation system? With a standard computer keyboard? With a mouse? With a piano

keyboard? With a game controller? With a guitar- or clarinet- or trumpet-shaped object? With a Theremin?

2. Would you prefer broad controls that direct general musical tendencies or would you want tight, fine-grained control over all the musical elements?

### *Audio/Visual Control Feedback*

The idea was to provide an audio "mockup" by crudely modeling various musical responses while the user manipulated a game controller. This step was not intended to prompt an overall evaluation of a potential soundtrack generation system. This step was intended to give the user a better idea of the project's ultimate goal.

1. [Using a video or two chosen from the Video Sample Response sections above, have the user "play" a soundtrack to the video.] Does this controller feel "right" for the job? Is there another input technique you could suggest?
2. Given the layout of the controller you have in your hands, which buttons would you associate with what sort of musical responses?
3. At what level of detail do you want to "catch" the action on the screen?
4. How well does the music have to be rendered in real time? For example, due to the amount of computational horsepower required, it may be necessary to provide you with only a "piano" sound in real time. After you have made a complete pass through a video, the music would then be rendered as an "orchestra." Could you live with that?

## AI 0.1 Survey Response

I recruited a potential customer: a successful software engineer (a.k.a. "geek") who is a gadget addict and who is very interested in video but who hasn't really had time to edit his accumulated videos into presentable form. He just started taking lessons on violin, so he is perhaps a bit more knowledgeable about music than my persona. Nonetheless, he matches my "Steve" persona very well.

### *General Audio/Visual Background*

None of his answers marked him as an oddball. He had had a little video production experience in high school, knew the names of three soundtrack composers (John Barry, John Williams, and Randy Newman), and had wished he had had a soundtrack for a "dramatic" slide show presentation in the not-too-distant past. In response to question 12, he would expect a good-to-very-good result from a piece of soundtrack software for which he would want to spend no more than \$100.



### *Video Sample*

### *Response/General*

### *Impressions*

I showed him the "Cute" (lion cub) video sample sans sound. He called this the "wonder of Nature." He also expressed a desire to see a preset range of emotional choices (e.g. a Likert scale) rather than answering the question in

such an open-ended fashion. He thought that the clip would benefit from a musical soundtrack, but he also could not imagine a clip that would not benefit from a musical soundtrack. Of course, if these clips had ambient sound of their own, the response may well be different.

When it came to estimating what sort of soundtrack might fit, he noted that the minimal amount of context—the clip is only ten seconds long—made it somewhat difficult to contextualize the video. However, he suggested simple, "uplifting" music, with, maybe, a beat that matched the footsteps of the cubs but was not too fast. No instrumentation or specific piece of music came to mind.

#### *Video Sample Response / Musical Suggestions*

With the "Cute" video still visible, I played some audio samples, all using the same pattern (a C-major arpeggio up, a scale-wise descent in c minor, and a C-major chord (C, E, G, C). Since these sounds were played through the tiny, built-in speakers on a late-2003-vintage 15" PowerBook, it quickly became apparent that he had a hard time making distinctions between the traditional set of orchestral instruments to which I had initially intended to limit my investigation. Therefore, I decided to widen the sonic field by going with my "Plan B" set of samples. The glockenspiel ("whimsical") made the best impression, followed by the xylophone and the Yamaha piano. The bassoon and the solo violin were marginally acceptable. The alto sax and church choir drew the strongest negative responses.

Beyond the issue of the timbres themselves, he took issue with the sequence that presented them: It was too long. He suggested a single-note sample to minimize confusion between the quality of the sound and the appropriateness of the sequence of notes followed by the sound. He also

noted that, while it was difficult to pick a clear winner, a "wrong" sound stood out immediately. Finally, instead of me controlling the sound while the listener sits passively, trying to recall and compare sonic memories, he suggested I set up a "Variations" matrix, similar to the visual matrix Photoshop uses to display the before-and-after effects of a particular filter.

### *Audio/Video Control Preferences*

I presented him with a few hardware controller options. Since these controllers were not actually musically controlling the computer, these questions were purely speculative. He thought that, for the purposes of following the music, he would prefer a standard joystick to a game controller. After I heard his response, I explained how a more complex soundtrack-generation system might be able to take advantage of the game controller's buttons, and he was a bit more receptive to the idea.



He wondered if I had looked into the current crop of digital DJ software. He also wondered about how "edit mode" would work—how would a user go back and tweak a soundtrack that had already been recorded? A finished soundtrack product would ideally offer a setup phase, a recording phase, and a post-production phase.

### **Conclusion: The First Iteration**

I had expected this survey to take approximately 30 minutes. Instead, it took about 100 minutes ... and I still did not make it through the whole set of questions. The open-ended nature of some of my questions was consuming much more time than anticipated! However, it was time well

spent. The participant was very articulate and willing to express opinions at great length. I realized that even if I were to interview only a few more people, the time constraints would not allow me to stick with my original set of questions.

## **Web Survey Responses: The Second Iteration**

I recognized that I had to perform major surgery on my initial set of questions and provide users with a system they could control themselves. Since such a revision would take time (see "Current Soundtrack Package Responses: The Third Iteration" on page 48), I decided that I needed to collect information in a much more simple way over the web. Users could take the surveys at their convenience, with the software emailing results to me as they came in. There were seven video clips—"cute," "anger," "calm," "slapstick," "spooky," "walking," and "work"—as drawn and labeled from the BBC Motion Gallery. Other than the video clips and page titles, all of the survey pages were exactly alike. You can see two frames from each of the video clips I selected in Appendix A on page 70.

### **The Survey Site**

The site consisted of the following:

- A home page that described the goals of the project and contained links to the survey instances
- Seven survey forms, each of which contained the following elements.
  - A QuickTime 7 movie of six to 15 seconds in length
  - A Flash 8 audio matrix containing the 16 test samples (each with its own start and stop button)
  - Three text fields into which users could type a clip characterization, a musical suggestion, and comments

- A matrix of 16 check boxes that allows users to register their sonic selections. This audio matrix included: Alto Sax, Hammond B3 Organ, Bassoon, Blues Harp [Harmonica], Church Choir, Distorted Guitar, Fat Synth, Glockenspiel, Goblins, Horn Ensemble, Lead Guitar, Shakuhachi, Timpani Tremolo, Violins, Xylophone, and Yamaha Grand Piano. Each sound button in the audio matrix played a simple single-note sample of that instrument on an appropriate C in that instrument's range.
- A Thank You page, from which you could either select another survey form or return to the home page

## Video Meets Audio #2



In a few words, how would you characterize this video clip (e.g. "violent," "crazy," "soothing," etc.)?

Do any specific pieces of music come to mind as being potential soundtrack material? If so, what are they?

Comments? What other sounds should be added? What other video clips would be helpful? What are some circumstances unique to you that would affect your soundtrack-generating thought process?

Alto Sax <input type="checkbox"/>	B3 <input type="checkbox"/>	Bassoon <input type="checkbox"/>	Blues Harp <input type="checkbox"/>
Church Choir <input type="checkbox"/>	Dist. Guit <input type="checkbox"/>	Fat Synth <input type="checkbox"/>	Glockenspiel <input type="checkbox"/>
Goblins <input type="checkbox"/>	Horn Ens <input type="checkbox"/>	Lead Guitar <input type="checkbox"/>	Shakuhachi <input type="checkbox"/>
Timp. Trem <input type="checkbox"/>	Violins <input type="checkbox"/>	Xylophone <input type="checkbox"/>	Yamaha Pno <input type="checkbox"/>

Click a sound button to audition a sound. Click a Stop button to stop the sound.

**Which sounds best fit the video?**

Alto Sax <input type="checkbox"/>	B3 <input type="checkbox"/>	Bassoon <input type="checkbox"/>	Blues Harp <input type="checkbox"/>
Church Choir <input type="checkbox"/>	Dist. Guitar <input type="checkbox"/>	Fat Synth <input type="checkbox"/>	Glockenspiel <input type="checkbox"/>
Goblins <input type="checkbox"/>	Horn Ens <input type="checkbox"/>	Lead Guitar <input type="checkbox"/>	Shakuhachi <input type="checkbox"/>
Timp. Trem <input type="checkbox"/>	Violins <input type="checkbox"/>	Xylophone <input type="checkbox"/>	Yamaha Pno <input type="checkbox"/>

Send my preferences to Scott!

Clear all of the fields on this page.

*The "anger" page from my online "timbre" survey*

An advantage to performing a web-based survey—in addition to the convenience of not having to arrange a meeting with a participant at a specific location—is that users can experience the video and the audio on their own computer systems instead of my (non-representative?) PowerBook. Users could bounce back and forth between clips at their leisure, answering as many questions as they see fit. If a user wants to run through a survey twice, that is fine.

The main disadvantage of a web-based survey is that I, as an observer, do not have access to the rich data provided by watching facial expressions and hearing off-the cuff comments. While there is value in having the user perform a survey in his or her home, I do not have the ability to inspect that home environment for potential enhancements or distractions.

### The Video Clips and Their Labels

As mentioned earlier, the BBC's Motion Gallery provided not only a rich source of perfectly sized video clips (in QuickTime format) but also a vocabulary that can be used as a baseline for "emotional" labels. Each BBC clip is cross-referenced to a number of terms.



For example, the most popular clip in the survey depicted two lion cubs wandering across the bush. The BBC label I applied to that clip was "Cute." The BBC cross-referenced "Cute" with "Botswana (Africa)," "Small Animals,"

"Bushland," "Cute Animals," "Camouflaged Young Animals," "Lion Cubs," "Young Animals," "Walking," and "Two (Number)."

I chose the BBC's "Walking" as the label for the sixth clip. The BBC provides these other keywords for the same clip: "Beach Huts," "Suffolk (English County)," "Pensioners," "Seaside Resorts," and "Promenades (Esplanades)." We can drift from lion cubs



walking across the bush to pensioners walking past the beach huts with only one degree of separation!

### Survey Results

As of 12 April 2006, ten unique visitors provided 54 responses: cute (10), anger (8), calm (8), slapstick (8), walking (7), work (7), and spooky (6). The table below summarizes the choices participants made in the audio matrix.

Page Name	Alto Sax	B3	Bassoon	Blues Harp	Church Choir	Distorted Guitar	Fat Synth	Glockenspiel	Goblins	Horn Ensemble	Lead Guitar	Shakuhachi	Timpani Tremolo	Violins	Xylophone	Yamaha Piano	Total
Anger	0	0	1	1	1	2	1	0	1	1	5	0	3	1	0	1	18
Calm	1	1	3	4	2	0	0	0	3	2	0	5	0	3	0	1	25
Cute	1	1	6	1	2	0	0	0	0	2	0	6	2	3	1	0	25
Slapstick	2	0	0	0	0	0	0	0	0	1	2	1	2	1	4	3	16
Spooky	0	2	1	1	3	0	3	0	4	2	1	1	1	4	0	0	23
Walking	3	1	1	2	0	0	1	2	0	1	0	0	0	1	1	3	16
Work	4	1	2	1	1	2	3	3	0	3	1	1	0	1	1	2	26
<b>Total</b>	11	6	14	10	9	4	8	5	8	12	9	14	8	14	7	10	

### *Timbre Survey Matrix Results*

Since the users saw clips labeled only with numbers, most of the respondents used the video-characteristics text box to provide their takes on the character of the video clip. A few participants supplied examples of specific pieces of music and general comments that came to mind when they watched the clips. Appendix B on page 74 lists the information the users provided.

### **Survey Results Discussion**

The most frequently selected instruments were shakuhachi, bassoon, and violins. The least frequently selected were distorted guitar, glockenspiel, and (Hammond) B3. Speaking generally, the popularity of certain sounds comes as no surprise. Orchestrators have known for centuries that the bassoon's timbre is pleasing yet not obtrusive. [Disclaimer: I own a bassoon.] Shakuhachi sounds have been some of the most (over)used synthesizer presets for almost 20 years. The violin sound is not only the definitive "classical" sound but it may also be the sound most immediately associated with film music.

Reviewing the participants' matrix choices on a sample-by-sample basis,

- Anger: Lead Guitar was the top vote getter with five. Five instruments received no votes.
- Calm: Shakuhachi snagged five votes with Blues Harp close behind at four. Six instruments received no votes.
- Cute: Bassoon and Shakuhachi were the big favorites with six votes each. Six instruments received no votes.
- Slapstick: Xylophone was the favorite with four votes. This is the clip demonstrating the greatest timbre consensus: Eight instruments received no votes.
- Spooky: Goblins and Violins each received four votes. Five instruments received no votes.
- Walking: Alto Sax and Yamaha Piano each collected three votes. Six instruments received no votes.
- Work: Alto Sax won with four votes. This is the clip demonstrating the least timbre consensus: Only two instruments received no votes.

The survey participants' remarks, reproduced in Appendix B beginning on page 74, provided some clues as to which simple characterizations may be more useful than others. With the "Cute" clip, the BBC's label lined up most closely with the participants' take. "Walking" and "Work" appeared to be labels that would not have been chosen by consensus.

- Anger: Based on the comments received, the ambiguity of Anger appeared to be about average. "Tension," "confrontational," "intense," "conflict," and "protest" were some of the sentiments that matched a number of the BBC's keywords. Some of the participants focused on the temporal aspect—the 60s-vintage look—rather than the angry or

confrontational facet. Even then, the time was eventually associated with conflict.

- Calm: Ambiguity expressed in the comments was lower than average for this clip while the ambiguity expressed in the timbre selection was about average. "Peaceful" and "tranquil" were the terms most frequently used. Interestingly, this clip seemed to inspire more talk of sound effects—"ocean sounds," gulls, wave, etc.—than of music. One remark—"ominous"—reminds us of how easily the mood could be shifted with just a touch of music (e.g. John Williams' *Jaws* motif).
- Cute: This clip also had lower-than-average ambiguity. The BBC and the survey participants agree: This was a "cute" clip! This was another instance where there was a significant discussion of sound effects ("animal noises," "jungle sounds," and National Geographic) in addition to music.
- Slapstick: This was arguably the least ambiguous clip of the bunch. While the sound selected most frequently from the matrix was xylophone, comments suggested "horn" sounds and "parlor" pianos. Although a number of visitors suggested a comedic atmosphere, the "time" stereotypes seemed to have the strongest hold on the participants' imaginations when it came to imagining timbres.
- Spooky: I thought "Goblins" would be the runaway consensus favorite for this clip, but users were not of the same mindset: The ambiguity was average. There were two mentions of Stanley Kubrick's *2001: A Space Odyssey* and, by extension (I assume) Richard Strauss' *Also sprach Zarathustra*. A few sound effects—hooting owls, blowing wind, and crickets—were mentioned. One user wrote, "It would depend on the music: either soothing or scary." Indeed!
- Walking: This clip prompted interesting results in that even though the timbral consensus seemed relatively high, the comments on the

clip suggested a wide variety of opinions on what captured attention. Leisure? The colors of the huts? The humor? The quaintness or nostalgia? This suggests that "walking" is probably not the most appropriate label for this clip. This looks like the reverse situation from "calm" above. With "calm," the ambiguity expressed in the comments was relatively low while the ambiguity of the timbre selection was about average.

- Work: This clip appeared to provide the most ambiguity on both the timbral and verbal fronts. "Busy" was mentioned most often in the remarks, but two people mentioned "process," one person went the cultural route ("Japanese"), and another wrote "instructional, exotic." The pieces of music suggested were Flight of the Bumblebee (at two-thirds speed) and "almost anything by Philip Glass," apparently alluding to a *Koyaanisqatsi*-esque vibe.

Some of the BBC's label choices are interesting. Why, for example, is the "calm seas" clip associated with colors ("yellow" and "orange") while the "spooky" clip isn't also labeled "black"? Would an American taxonomist use "USA Nationals" as a synonym for "anger"? I would be interested to find out what process the BBC uses to classify its clips. However, no matter how one might feel about the BBC's video taxonomy, at least it gives us something to work with. You have to start somewhere!

## **Current Soundtrack Applications: The Third Iteration**

In the Real-World Cues and Clues section beginning on page 15, I discussed the feature sets of GarageBand 3 and Musicbed DV. I decided it would be a good idea to see how my target user would respond to those applications. In addition, I also rigged up a funky joystick-control test to see if a joystick was capable of providing as much kinesthetic joy as the participant

suggested it might. Finally, I created a fallback plan in case my jury-rigged joystick test crashed: I made sure that Music Mouse was healthy enough to run on my 1999-vintage PowerBook G3.

### **Pre-Test Terminology Survey**

Since one of the biggest challenges is to determine what terminology works best when attempting to label musical and emotional qualities, I wanted to find out how much sense the adjective pairs used in GarageBand and Musicbed DV made in the absence of any experience with the application. Appendix C on page 79 contains the test form.

Although I originally intended to ask the students in a Design Management class to take this survey, I decided that it was important to ask this question of someone who had some direct experience with what I was trying to accomplish. Therefore, only one person has completed this survey to date.

Overall, the participant thought that Apple's GarageBand descriptors were more helpful than those in Musicbed DV. The pairs that resonated best were "Relaxed - Intense" (GB) and "Melodic - Percussive" (MDV). Other pairs that scored well were "Light - Intense" (MDV), "Single - Ensemble" (GB), "Acoustic - Electric" (GB), and "Melodic - Dissonant" (GB). The pairs that made the least sense were "Part - Fill" (GB), "Abstract - Real" (MDV), and "Low - High" (MDV).

### **GarageBand 3 and Musicbed DV**

The test for both GarageBand 3 and Musicbed DV was as follows: Select a video clip and work for 15 minutes creating a soundtrack for that clip. The composition choices were completely up to the tester: I provided no criteria indicating what may or may not constitute an acceptable soundtrack.

Musicbed DV allows the user to choose from its proprietary music library that consists of "tens of thousands of music theme variations." Oddly enough, many of the snippets listed separately in Musicbed DV share the same names. This makes it difficult to differentiate a sample you may have already auditioned (and rejected) from a sample you have not yet heard.

In my GarageBand installation, I have over 8000 loops from which to choose. Even if a user engages GarageBand's style filters, he or she may still face a daunting number of relevant loops. For example, there are at least 78 loops called "Orchestra Strings." There are 31 loops labeled "Funky Electric Piano." The novice user who does not have hours of time on his or her hands can be overwhelmed by these choices.

### *GarageBand in Use*

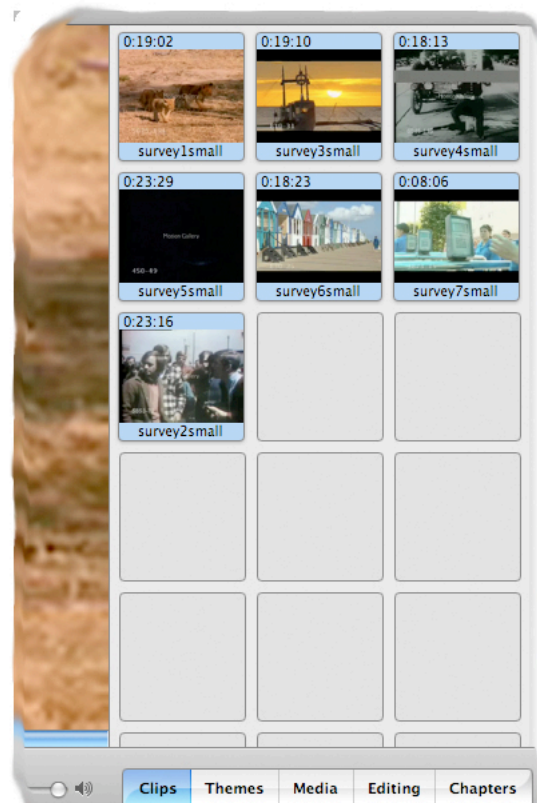
The participant chose to score the "Cute" video clip and started in by seeking out a rhythm and/or bass line to set the overall tone. Using GarageBand's "mood" filtering buttons, he tried to minimize the number of loops presented by choosing "Cheerful," "Relaxed," "Acoustic," and "Melodic." Even after applying these filters, a long list of loops remained visible. He previewed two loops, selected the second one ("Acoustic Picking 02"), and dragged it to the tracking area. He then looped this one sample to repeat for the duration of the track.

At this point, he wondered how to go about selecting a second track that would fit with the one he had just chosen. He was looking for help in assembling loops in ways that made musical sense, based upon the one track he had chosen. Since he is a software engineer with years of experience with a variety of programs, I asked him if a cyber-agent like Microsoft Word's infamous Clippy—the animated paper clip that pops up to

solve word-processing problems—might be useful. He did not think help in that form would be good. Nevertheless, some sort of suggestion would be helpful.

After more trial and error, he found a second loop ("Aperitif Cello 01") that fit with the first, dragged it to the tracking area, and looped it.

At this point, he had spent 30 minutes selecting two basic tracks for a 19-second-long video clip. Of course, the test environment slowed the process down somewhat. I asked if he thought that amount of time and effort was excessive, given the length and subject matter of the clip. He was not sure. On the positive side, he figured that if he were adding music to a home movie, he would probably score only one small section per setting. I asked him how much time it should take to score a 10-minute video clip. He thought that two hours would be acceptable. Even then, he would only spend that amount of time and effort on video that he **really** wanted to score. He would not work that hard on a regular basis.



iMovie's Clips Thumbnail Area

As far as being able to score small sections in separate sittings, he thought it would be nice to have an iMovie-like clips thumbnail area where the

various sections, scored and unscored, could be stored. In his opinion, the easy part of the scoring process would be to score the sections. The hard part would be scoring the transitions from one section to another.

### *Musicbed DV in Use*

To test Musicbed DV, the participant chose "Cute" once again! While he liked the idea behind the "Style" and "Color" radio buttons (see the introduction to Musicbed DV on page 20), some of the category labels (e.g. "Low - High") were not helpful. The "Genres" and "Moods" drop-down lists went unnoticed for about ten minutes. Once he explored the genres and moods, he thought the options were too limited. (Musicbed DV's genres and moods are listed in Appendix D on page 80.)

Musicbed's concept of "sections" was confusing: He had no idea what they meant or how to manipulate them. The "Relevance" bars—little "signal-strength" boxes that indicated how closely a particular sample matches the currently selected set of mood preferences—were good. The names of the musical clips, which come from Synk Audio's proprietary music database, were repetitive and not descriptive. When auditioning samples, playback did not stop immediately when you clicked the Stop button. Samples deemed relevant did not match his musical expectations. Generally, the selection of musical samples seemed "woefully inadequate."

Adding music to the timeline was not as easy as it looked. Neither dragging and dropping nor right-clicking samples produced the desired effect: He had to click the "Add Music Clip" button to insert a sample on the timeline. He chose one of the many "Cyan Life" music "packages." After auditioning more samples, he added a second package—"Magenta Passage"—to the first. Once

the second layer was added, there was no obvious way to return to the "Cyan Life" layer.

After about thirty minutes of exploration, he decided that he did not think he would buy this product at **any** price. It was not amenable to exploration. Even though music could be layered, he thought it invited using only one clip at a time. Some controls he regarded as fundamental were not intelligible. In short, Musicbed DV did not earn his vote of confidence!

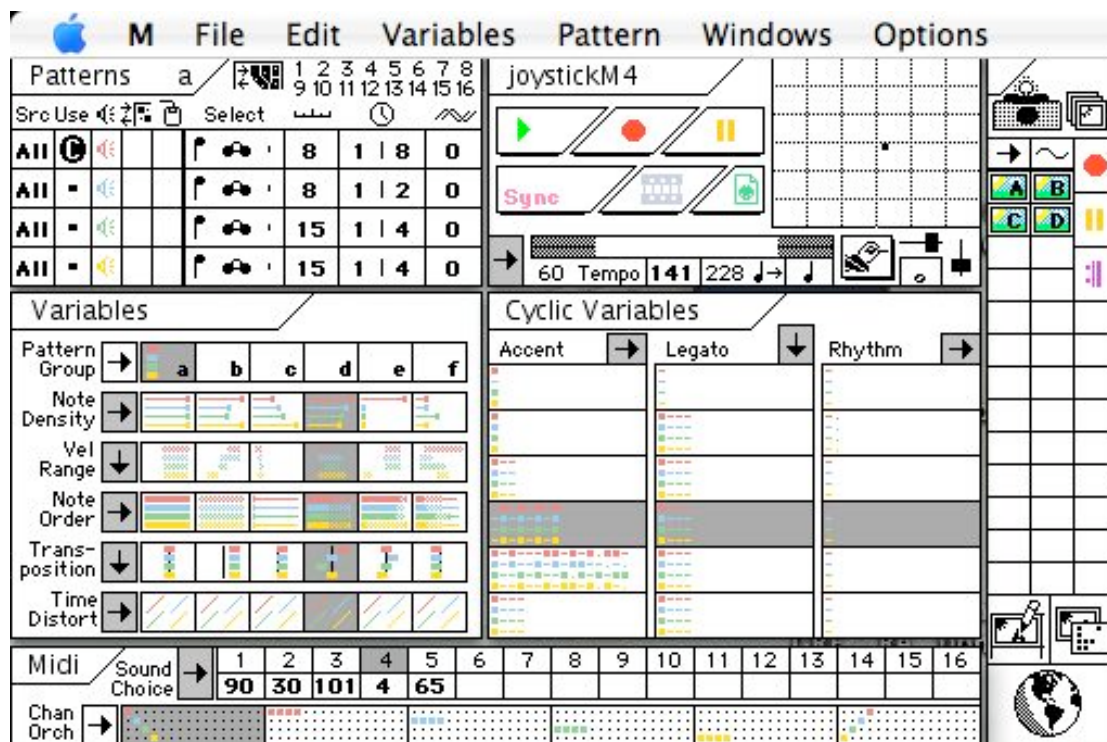
### *SUS Survey Results for GarageBand and Musicbed DV*

Even though he had some experience with GarageBand before this test began, the participant was not smitten with soundtrack-generation prowess of either of these applications. Using the Digital Equipment Corporation's System Usability Scale (SUS) (Brooke n.d.), he gave GarageBand a score of 40. Musicbed DV received a score of 25. When considering results from an SUS review, most usability experts consider a score below 50 to be unacceptable.

### **M with a Joystick**

I asked the participant to try my Rube-Goldbergian concoction of QuickTime, SimpleSynth (a program that allows the internal QuickTime voices to be controlled like any other MIDI instrument), M (an algorithmic-composition program from Cycling '74), JunXion (a program that converts information from controllers into MIDI information), and a Logitech Attack 3 joystick. In an earlier test, this same participant had suggested that joystick control would be more immediately gratifying than game pad control. This was an experiment to check that hypothesis. By moving the joystick, the user controlled the algorithmic-composition parameters of M in real time. In the screen shot below, you can see a grid in the upper-right corner of the

application. The various musical parameters—note density, velocity range, note order, transposition, etc.—are assigned to track movement across the grid in either the X or Y axis. In the example illustrated below, as the joystick moves to the right, tempo increases and note density, note order, time distortion, the accent pattern, the rhythm, and the timbre (sound choice) change. As the joystick moves up, the velocity range, transposition, and degree of legato changes.



*Cycling '74's M algorithmic composition program*

Two scales—one in C major and the other in c minor—provided the raw musical material M used in this trial as the basis for its statistical manipulations. The note order parameter was set to favor the minor scale as the joystick moved to the left and favor the major scale as the joystick moved to the right. (Perhaps adapting motives from Terry Riley's *In C* would have produced slightly more interesting results, but time was short.)

Alas, the software on my PowerBook proved to be quite unstable on test day. The participant was able to associate joystick movement to "musical" output—the "music" was not pretty—for only a few minutes before M crashed. The system would not remain alive long enough to produce meaningful results.

The participant's opinion was that it might be an appealing way to control a soundtrack system, but there was no sure way to judge given the circumstances. However, the joystick experience did prompt him to say that more than more than two axes of musical control may be necessary after all. (In an earlier test, he was the one who suggested my game controller idea was overkill and that a joystick might strike the best balance between musical control and ease of use.)

### **Music Mouse**

Anticipating the aforementioned joystick-related software problems, I had prepared a 1999-vintage PowerBook (which runs the older Macintosh OS 9.2) with Music Mouse and the standard QuickTime set of internal sounds. The standard PowerBook track pad controlled Music Mouse.

Music Mouse was a smash success! The tester really liked the fact that the music began immediately. He described the interface as "safe" and "direct." The application "invites you to play." He gave Music Mouse a SUS score of 92.5 (compared to 25 for Musicbed DV and 40 for GarageBand)!

This also prompted him to rethink the track pad/joystick/game controller issue once again. Music Mouse reminded him that a simple track pad has considerable potential, if used cleverly. This led into a discussion of how a more sophisticated music system could use multiple track pads or a multi-touch sensitive controller (e.g. JazzMutant's Lemur touch-screen music

control surface) or a graphics tablet (e.g. a Wacom Graphire). A graphics tablet could use the pressure sensitivity of the stylus (or maybe a control "thimble") to vary another musical parameter in real time.

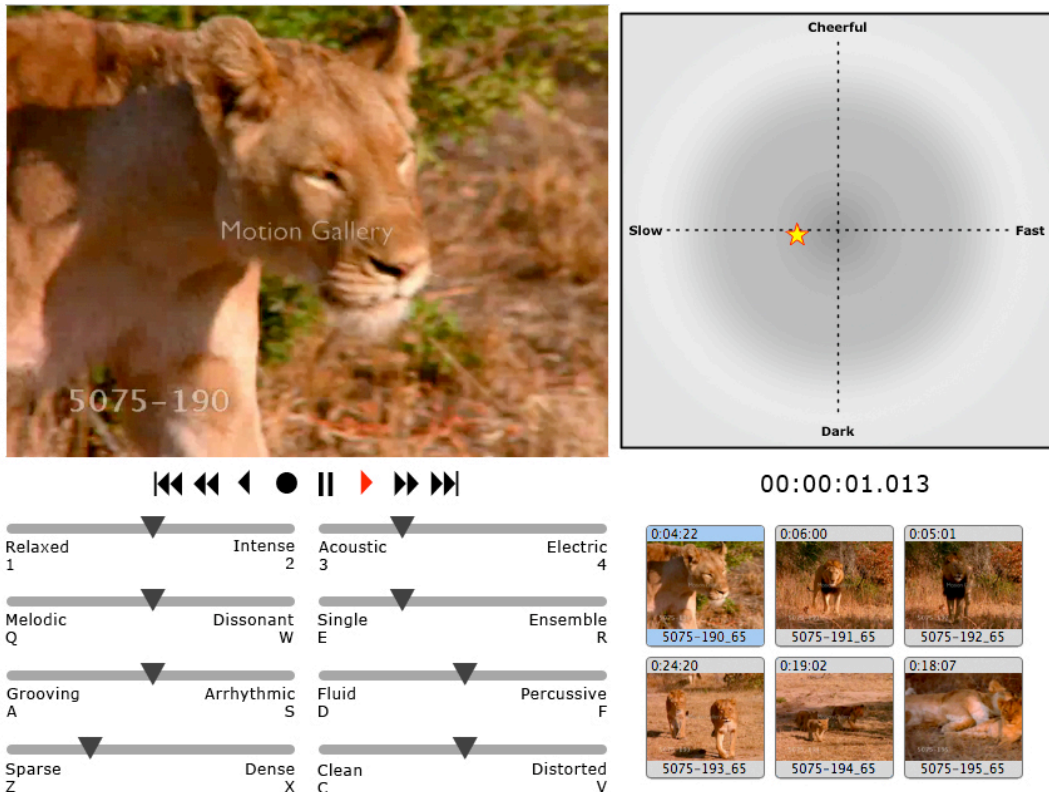
## **Conclusion: More Iterations, Please!**

### **Meaning and Structure**

Recalling Dr. Cohen's model of how the mind interprets audiovisual input (page 8), the goal is to design an application that allows a user to identify the *meanings* of the musical and visual elements they want to combine and then impose *structure* on those elements. Even though the research referred to in this paper hints at combinations that may be more effective than others, there is still a very broad spectrum of configurations from which to choose.

### **One Possible AI Configuration**

This section presents one way to arrange the major components. In the best Frank Gehry tradition of remaining "liquid" (Boland 2004, 20), this illustration has been kept simple by design (so to speak). The intention behind presenting relatively skeletal mockups is to encourage further speculation on how the elements discussed in this thesis may be added, deleted, reformatted, and reshuffled: video clip presentation (with shuttle controls), X-Y axis map, music "quality" sliders, and an area for clip thumbnails.



The numbers and letters beneath the various musical-characteristic-shaping sliders represent keyboard shortcuts. For example, to generate more "relaxed" music, the user would tap on the "1" key to nudge the slider to the left. For more percussive music, the user would tap the "F" key to move the relevant slider to the right. Ideally, the user could assign any of the musical continua to either axis of the X-Y display or any of the sliders.

The practical problems of assigning keywords to video clips and adjective pairs to points along the musical spectra became more intriguing as the project evolved. As I move forward with this project, the next step will be to do a considerable amount of user testing to get a general sense of which terms have garnered the greatest consensual powers in their various cultures. Perhaps it will turn out that, after 82 years, Erno Rapée will be

vindicated concerning the fifty-two piano-playing moods and situations he chose!

As far as other interface elements go, there seems to be little reason to display a staff containing music notation, as I did in the initial AI 0.1 proposal (page 30). Persona Steve will not be able to read sheet music so why clutter up the screen with gratuitous widgetry? More importantly, the furrowed brows of the participant prompted by the relative visual complexity of GarageBand and Musicbed DV stood in stark contrast to the joyful expressions inspired by Music Mouse's simple interface.

### **User Calibration**

In keeping with psychological findings on the impact of culture and age, it would be nice if there were a "calibration" function in AI. A user would calibrate AI in the same way most voice-recognition programs are calibrated: The user tells AI what sorts of music he or she prefers and provides some basic preferences on how that music should be applied to the soundtrack-generating tasks. AI's calibration material could come from a variety of sources. It could be a simple user survey form, a list of favorite songs, an analysis of an iTunes library, a feed from a music "analysis" site like Pandora (<http://www.pandora.com>), or maybe information provided by some collaborative online environment. Based on the stated preferences, cultural traits, age, and other relevant factors, AI would adjust the way it assembled the information contained in its musical database. As Bullerjahn and Gldenring (1994) put it,

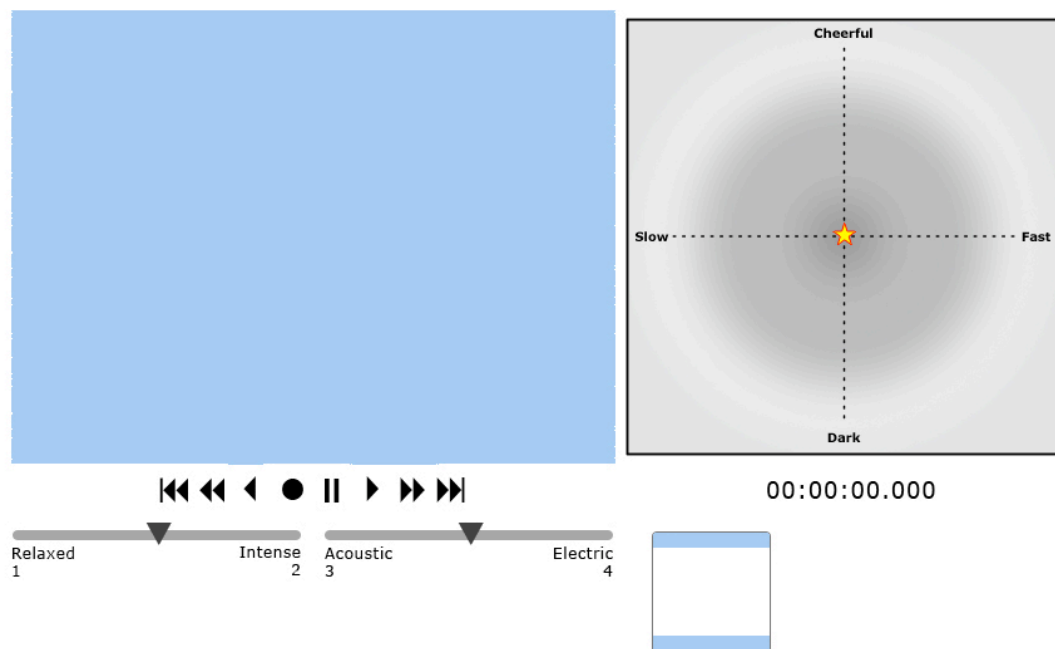
[F]or subjects without media competence, music enhanced evaluations of the screenplay and actors. Subjects with musical experience hear music more precisely, and structural elements of the music shape their interpretation of film. Subjects without musical experience, though, are easily influenced by well-known

clichés. Finally, older adults are less aware of background music than younger adults and prefer scores by earlier composers (102).

Customizing the output of a design tool such as AI to the personalities of its users will not be easy. As the QSketcher team wrote, "Creators have many different work-styles: no single approach or process is sufficient" (Abrams 2001, 3).

### Example Storyboard: Steve Meets AI

The first time Steve runs AI, Steve sees a default interface configuration that looks something like the following illustration.



*AI's default configuration*

At this point, AI has not yet been calibrated to Steve's soundtrack expectations. Therefore, AI exposes an "absolute novice" set of controls, exposing only four sets of "emotional" sliders to the user. An absolute novice would see "Slow–Fast" and "Cheerful–Dark" assigned to the X-Y coordinate

controls while "Relaxed—Intense" and "Acoustic—Electric" would be assigned to sliders.

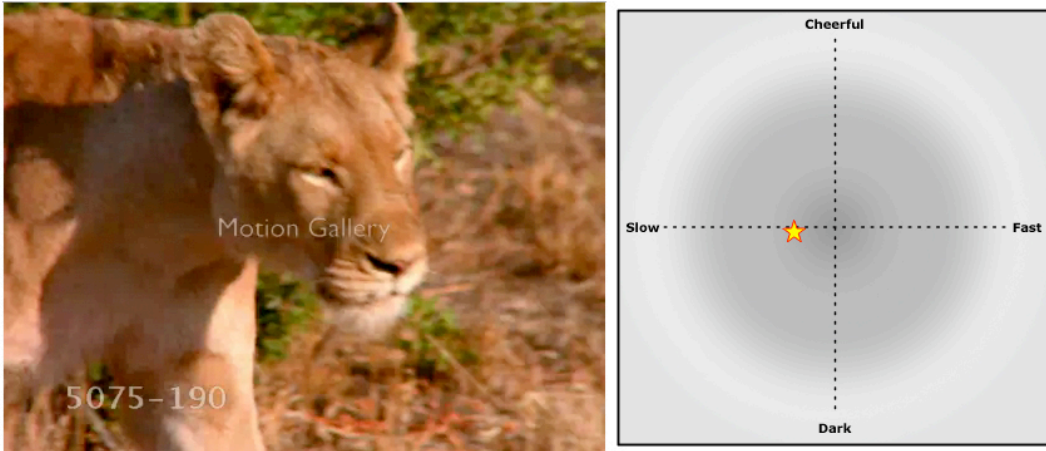
After Steve answers Al's calibration questions, Al rates Steve's sophistication at a level above "absolute novice," realizing that Steve is comfortable juggling a number of parameters simultaneously. To match Steve's expected soundtrack "composing" style, Al adds six more sliders—"Melodic—Dissonant," "Single—Ensemble," "Grooving—Arrhythmic," "Fluid—Percussive," "Sparse—Dense," and "Clean—Distorted"—to the interface.

Steve has video footage of his recent trip to Africa that he wants to edit and present to a gathering of his friends at a party next week. He has followed Richard Davis' advice:

Have a strong idea in your mind of what you want to say with the music. Should it be funny or should it heighten tension? Should it be somewhat neutral, or melodramatic? The mood or emotion you want to express will suggest a tempo. (Davis 1999, 163)

Overall, the video storyline moves from a shot of a female lion to a male lion (who may or may not be on the prowl) to a bouncy den of lion cubs. Steve "spotted" six short sections in his video. This storyboard will examine Steve's musical impressions for three of those six spots. Here is a sequence of three screen mockups that trace the evolution of Steve's soundtrack.

## Female Lion



5075-190

Motion Gallery

Cheerful

Slow Fast

Dark

00:00:01.013

Relaxed 1 Intense 2 Acoustic 3 Electric 4

Melodic Q Dissonant W Single E Ensemble R

Grooving A Arrhythmic S Fluid D Percussive F

Sparse Z Dense X Clean C Distorted V

0:04:22 0:06:00 0:05:01

5075-190.65 5075-191.65 5075-192.65

0:24:20 0:19:02 0:18:07

5075-193.65 5075-194.65 5075-195.65

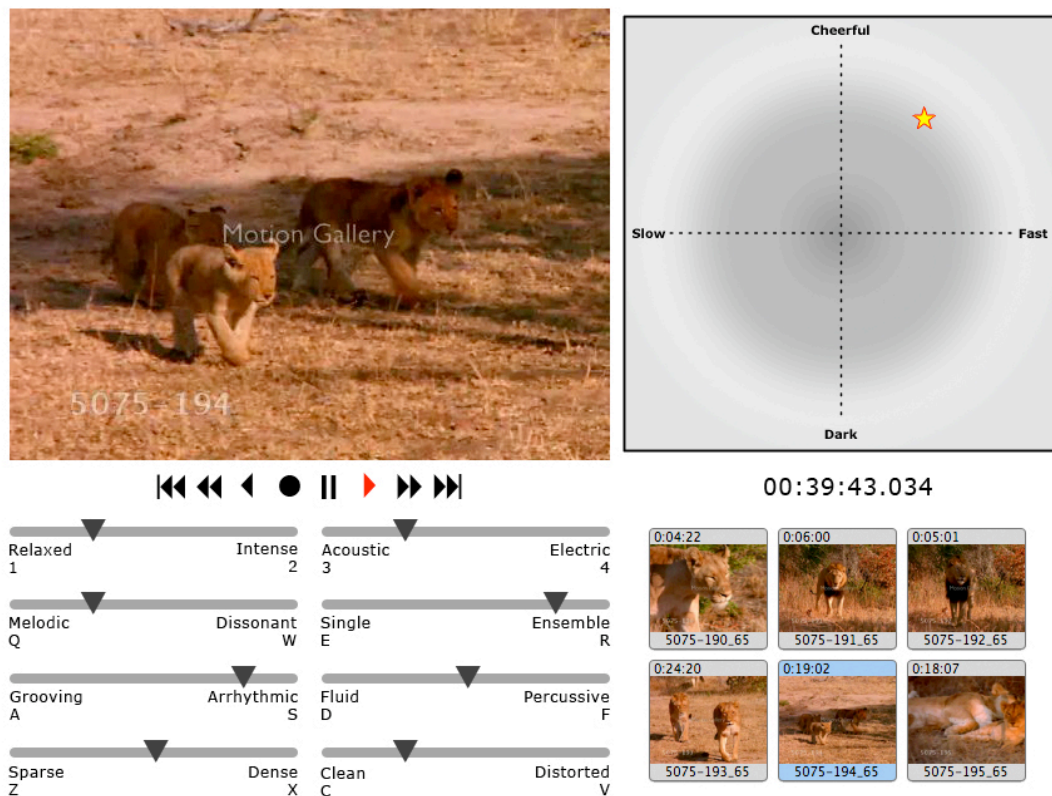
Since the female lion makes her initial appearance solo, Steve's music starts out on middle ground, with the texture on the sparse side, "performed" by only one or two "acoustic" virtual instruments. Steve chooses a tempo that reflects her deliberate pace.

## Male Lion

The image displays a male lion in a savanna environment. To the right is a circular mood chart with axes for Cheerful (top), Dark (bottom), Slow (left), and Fast (right). A yellow star is positioned in the lower-left quadrant, indicating a mood that is slightly slower and darker. Below the lion image is a video control interface with playback buttons and a progress bar. The video title is "5075-191". To the right of the controls is a timestamp "00:04:22.103". Below the controls are eight sliders for adjusting various audio parameters: Relaxed (1) to Intense (2), Acoustic (3) to Electric (4), Melodic (Q) to Dissonant (W), Single (E) to Ensemble (R), Grooving (A) to Arrhythmic (S), Fluid (D) to Percussive (F), Sparse (Z) to Dense (X), and Clean (C) to Distorted (V). On the bottom right is a grid of video thumbnails with timestamps: 0:04:22, 0:06:00, 0:05:01, 0:24:20, 0:19:02, and 0:18:07.

When the male lion makes his appearance, Steve increases the tempo a bit and suggests a slightly darker mood because we do not really know what this male's intentions are. Steve taps the "2" key a few times to increase the intensity. He also taps the "F" key to make the music a bit more percussive. The musical changes suggest this male may be an unwanted outsider, although not necessarily a looming threat.

## Lion Cubs



*Lighthearted music accompanies footage of the cubs. Steve chooses more relaxed, melodic, fluid, and clean options to indicate youthful innocence. Steve increases the arrhythmic aspect to emphasize the cubs' clumsy nature. Finally, since there are a number of cubs, Steve thickens the texture by increasing the bias in favor of dense and "ensemble-sounding" musical output.*

## Future Directions

I am glad Dr. McGee suggested investigating musical timbres first. As far as I can tell, there has been very little research into how timbre affects the overall perception of an audiovisual presentation. Given the frequent reference to sound effects in my web survey, it would be interesting to get a better handle on where one draws the line, for instance, between a

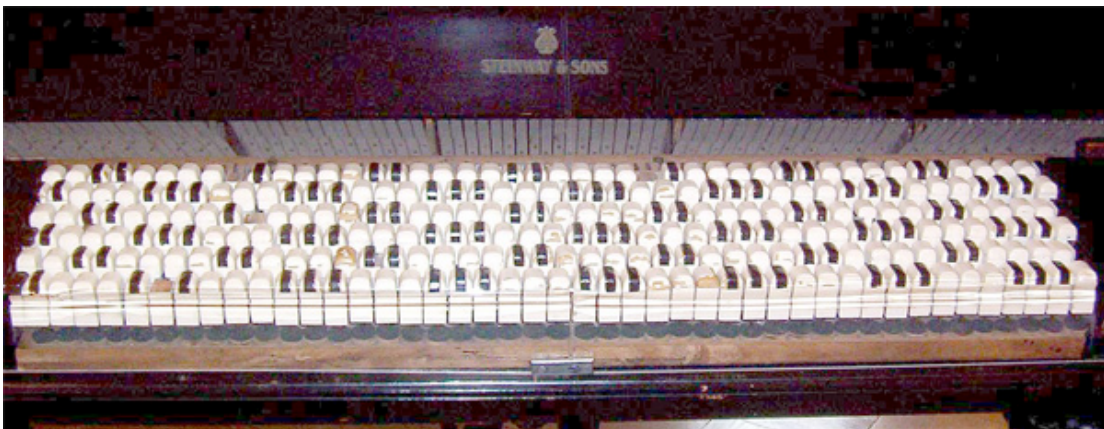
"musical timbre" and a "sound effect." Who knew better: Rimsky-Korsakov or Cage?

More research into applicable business models is called for as well, particularly regarding the mechanics of accumulating a style library or collaboratively creating soundtracks with friends and family, either locally or over the Internet. How many economic ideas can AI borrow from the online gaming industry, for example?

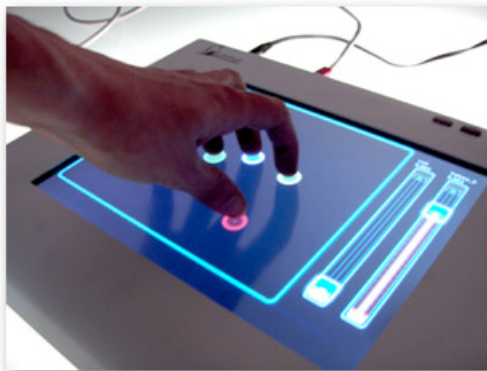
### **Reflections on My Process**

The evolution of this project has been so tortured that even Darwin may not be able to fully comprehend how my work ended up promoting cyberharmony between "Steve" and "Al."

I began laying the groundwork for research into alternative music hardware controllers (e.g. the Janko keyboard layout, the JazzMutant Lemur, the Theremin, etc.) in the fall of 2004. My search for a new musical hardware controller was selfishly motivated by a lifetime of crippling piano performance anxiety.



*A Janko keyboard installed in a Steinway grand piano, circa 1910*



*The JazzMutant Lemur control surface*



*Moog's MIDI Theremin*

My initial hardware-only quest morphed into an imagined hardware + software system that augmented a musical controller's physical virtuosity with AI (artificial intelligence)-enhanced stylistic information. The intention was to employ a language like PROLOG to extract stylistic info from Band-in-a-Box libraries and Apple Loops, putting that info to work in a real-time performance environment.

The pragmatist in me soon realized that a new, "revolutionary" musical hardware controller stood about as much chance of widespread adoption as the Dvorak—or the Janko—keyboard. Focusing my attention on the software side, I expanded the scope of the software that was originally intended to provide "virtuoso" performances. I divided the musical capacities of the software into three user categories:

- People like me who have struggled with performance anxiety and have been on the lookout for an instrument that would somehow put nerves at ease.

- People who are using Soundtrack (or GarageBand) to generate soundtracks to their digital home movies.
- Orchestrators and composers who know quite a bit about music and who are under the gun to produce lots of fully orchestrated music under tight deadlines but who suck at playing traditional keyboards.

Since the odds of working out the beginnings of a solution for even one set of customers listed above were slim, I took Dr. McGee's suggestion and went with the option that not only poses the most interesting psychological questions but also has the greatest potential commercial upside.

Even after I had narrowed my project to designing an "AI" for "Steve," my goals remained very ambitious. I not only intended to collect and analyze user data, I was also going to have a rudimentary system running—built with Max, Reaktor, Cocoa, PROLOG, or a combination of all of them.

On the design-thinking front, I initially intended to link my research—particularly the decision-making aspect of choosing one timbre or piece of music over another—to recent developments in "bounded rationality." The notion of "bounded rationality" is commonly ascribed to Herbert Simon (Simon 1996), has been interpreted by investigators like Gigerenzer (Gigerenzer 1999, 2000, 2002) and Gilovich (2002), and was popularized in *Blink: The Power of Thinking Without Thinking* (Gladwell 2005). It would be interesting to see how musical decision making fits into a bounded-rationality theory.

## **Music, Emotion, and Interaction Design**

Emotion is one of the hottest topics in interaction design now. Since soundtracks have been manipulating emotions for many years, interaction

designers may be able to formulate theories of emotional "design" based on soundtrack studies.

Because emotion and meaning characterize the film experience, film music provides one route for exploring the relations between music and meaning and between music and emotion. (Cohen 1994, 2)

Alvin Collis was until recently Muzak's former senior vice-president of strategy and brand. In the eighties, he worked as a freelance sound engineer. Here is what author David Owen had to say about Collis and Muzak.

... Collis was doing an engineering job for Muzak. He told me, 'I walked into a store and understood: this is just like a movie. The company has built a set, and they've hired actors and given them costumes and taught them their lines, and every day they open their doors and say, 'Let's put on a show.' It was retail theatre. And I realized then that Muzak's business wasn't really about selling music. It was about selling emotion—about finding the soundtrack that would make this store or that restaurant feel like something, rather than being just an intellectual proposition.' (2006)

Recent studies of how emotions are related to "classical" music promise to contribute to our understanding of this complex issue. For example, researchers are working with the Boston Symphony to analyze the emotional responses to music of audience members and musicians.

Instead of his usual suit jacket or tuxedo, [Boston Symphony conductor Keith] Lockhart will wear an armband and a black Lycra top that looks more like a biking jersey. They will both be wired with a series of sensors that will measure his heart rate, movements, muscle tension, and other physiological evidence of emotion. Five musicians will be similarly wired to measure how they react to his conducting and to playing the music.

In the audience, some children and adults will wear these same sensors on their arms and fingers, allowing their bodies to tell the scientists what kinds of emotional intensity they are feeling. Others will monitor themselves, sliding a switch on a hand-held box to indicate the emotional intensity they perceive in the music.

"We are trying to get a fingerprint of the emotional landscape of the music as it is transmitted from the conductor's baton to the musicians' fingers to the hearts and minds of the audience," said Daniel Levitin, a professor of psychology and music at McGill University in Montreal ....

...

"Ideally, we will see something we recognize from the conductor's reading in the musicians' readings about two seconds later, and then in the audience's readings," Levitin said yesterday.

...

The scientists are also going to videotape Saturday's performance and show it to a similarly wired audience in Montreal to measure whether audiences react differently to live or taped performances. ... (Elton 2006)

Broader applications of "emotion" were on the QSketcher team's mind as well. "Several of the authors are currently adapting these techniques to other areas. In fact, we believe that the "tip of the iceberg" of an important research area in HCI is emerging: developing information technology tools to support creative work." (Abrams 2001, 2)

"More generally, we have found the music domain to be an excellent vehicle for developing new metaphors and mechanisms for supporting creativity, and believe that these ideas apply to many other domains. The key aspects of tools needed to support creative workflow (i.e. capture, organize, and manipulate), and the ways in which they relate to the cognitive aspects of creativity (i.e. context, relationships, state, and domain-concepts), have the potential to grow into a general model for

framing research and development in tools that support creativity across many domains." (Abrams 2001, 8)

## **Coda**

The process of researching and writing this thesis has generated more questions than answers. The aggregate question, however, provides a filter that will enable us to refine the next generation of questions and methodologies. By balancing an accumulation of psychological information with the practical goal of creating an automated, intelligent music-generation system, we stand to learn something new about interaction design as well.

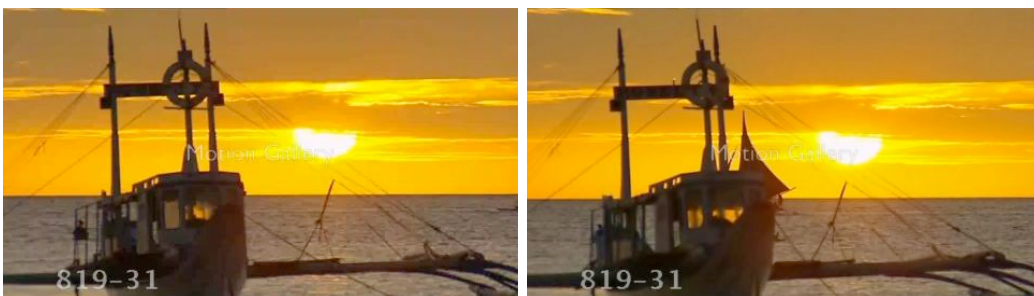
## Appendix A: Video Clips

### Anger (BBC Motion Gallery #5853-72\_65)



BBC Keywords: Opposition, US Foreign Policy, Suits, Civil Disobedience, Heated Arguments, Vietnam War (1964), Men, Body Language, Views (Opinions), Microphones, Universities, Cleveland (USA), Students, Facing (Opposite), USA Nationals, Standing, Peace Movements, TV Reporters, Sit Ins (Occupation)

### Calm (BBC Motion Gallery #819-31-65)



BBC Keywords: Cirrus, Raised Ship Sails, Yellow, Sailing (Action), Ropes, Ship Masts, Calm Seas, Orange, Philippines, Sailing Dinghies, Ship Sails, Oceans, Bobbing (Action), Wooden Ships, Sun Set (Celestial Motion)

**Cute (BBC Motion Gallery #5075-194\_65)**



BBC Keywords: Cute, Botswana (Africa), Small Animals, Bushland, Cute Animals, Camouflaged Young Animals, Lion Cubs, Young Animals, Walking, Two (Number).

**Slapstick (BBC Motion Gallery #6046-116\_65)**



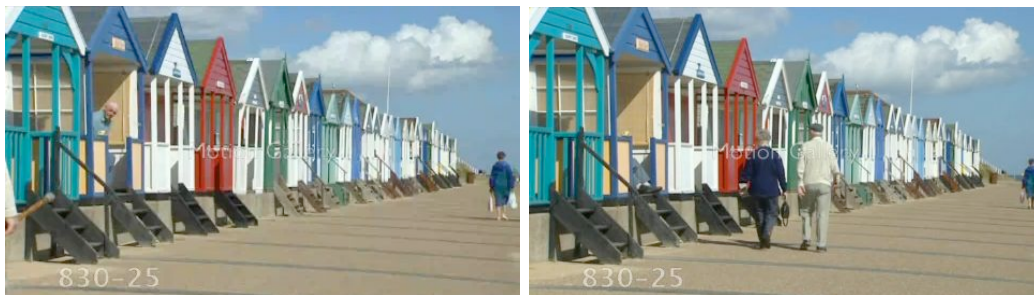
BBC Keywords: Motor Cars, Slapstick Comedy, Central Reservations, Car Chases, Slapstick (Funny), Roads, Police Vans, Robbers, Slapstick Comedy Films, The Keystone Cops (Fictional Characters), Policemen

**Spooky (BBC Motion Gallery #450-49\_65)**



BBC Keywords: Spooky, The Moon, Moon Rise, Cloud, Shadows

**Walking (BBC Motion Gallery #830-25\_65)**



BBC Keywords: Walking, Beach Huts, Suffolk (English County), Pensioners, Seaside Resorts, Promenades (Esplanades).

**Work (BBC Motion Gallery #3873-15\_65)**



BBC Keywords: Identity Cards, Employees, Blue, Lines (In Line), Clocking In, Factories, Work (Employment), Uniforms, Shenyang (China)

## Appendix B: Web Survey Responses

The respondents supplied the following terms in the video characterization text box. A **red** entry indicates a match with the BBC.

- Anger
  - pending trouble, tension
  - confrontational, newsy, historical
  - confusion, hubbub
  - fragile, escalation, **angry**
  - Earnest
  - intense, dangerous
  - Conflict
- Calm
  - peaceful
  - soothing, peaceful, travelogue
  - ocean sounds, water waves, awesomeness of nature, enz
  - peaceful, finalizing, relaxing
  - tranquil, lazy, ominous,
  - Tranquility
  - **calm**
  - Serene and slightly seasick
- Cute
  - maternal
  - wondering
  - bugga
  - **cute**, natural, scientific, exotic
  - darling
  - calming, **cute**, reassuring
  - Lazy, transitional

- **cute**, funny, relaxing
- **cute**
- **Cute**
- Slapstick
  - funny
  - calamity... silly chaotic chase
  - funny, crazy, historical
  - wacky, anarchy
  - **slapstick**
  - exciting, humorous, zany
  - Hitler
  - Funny hectic
- Spooky
  - distant
  - imagining... exploration, what are the possibilities?
  - It would depend on the music: either soothing or scary.
  - humble, quiet
  - creepy, scary, revealing
  - Eerie, gloomy, ominous
- Walking
  - leisure... time is not an issue
  - colorful, intriguing
  - funny, quaint
  - peaceful, warm
  - nostalgic, happy, gentle
  - Nostalgic
- Work
  - busy
  - process

- instructional, exotic
- puzzling, intriguing
- repetitive, busy
- busy, chaotic, massive
- processional busy

The following sets of bullet points report the combined specific-pieces-of-music suggestions and general comments received from the web-survey participants.

- Anger
  - Snare drums ... the call to military action or climatic drama
  - Building of pitch and speed towards what seems an inevitable confrontation
  - Early Bob Dylan
  - It depends on my audience; if I was scoring this as part of, say, a political commentary clip I meant for You Tube or something, I'd want a voiceover and/or music; if this was a home movie of people I know, I'd want to hear what they were saying.
  - "Southern Man," "What's Goin' On"
  - If the Timpani Tremolo were used it would sound past tense, like a documentary sound track
  - A 60's or 70's war protest song like "Fortunate Son" or something.
  - The time period was what influenced me the most on the decision. Timpani just sounded like a nice crescendo before something violent happened.
  - Not really. Maybe some Beatles song revolution or something like that
  - With all the conversation going on this would have to be low in the mix

- Calm
  - Something by John Williams
  - Sea gulls
  - Wind, waves. I like sailing and the ocean.
  - Ambient sounds typical of the sea with man made objects, like ships bells creaking rigging
- Cute
  - Something playful (theme to Addams family?)
  - African folk music, The Circle of Life
  - Other animal noises; some sample clips, with background noises intact, from "Animal Planet;" I've never been to Africa, or seen lions in the wild, so all of my opinions about what would sound "right" come from TV or the movies.
  - Ambient jungle sounds
  - Reminds me of a national geographic show, where they don't really have music most of the time.
  - The only thing that comes to mind immediately is from Peter and the Wolf, but not the wolf music.
  - Playful songs, safari songs
  - Something with an African Percussion, I don't know of anything by name, generate a beat to go with the walking of the cubs
- Slapstick
  - A kazoo
  - Just about any ragtime piece
  - Honking horns, shouts.
  - Silent film organ music
  - Old brassy horn sounds

- An old parlor tune on a piano, something upbeat with a fast tempo to match the fast film speed.
- A parlor piano would be perfect for this
- Spooky
  - Also sprach Zarathustra from *2001*. Sorry, can't help the association.
  - Werewolves of London
  - A hooting owl
  - *2001* space odyssey, ET
  - Reminds me of an old horror movie.
  - Wind blowing, crickets
- Walking
  - Had trouble finding anything that connoted leisure
  - Might help if the xylophone was a rolled note.
  - Some Rudy Vallee
  - Gulls, wind, voice over conversation by the couple
  - Muted voices, seagulls
- Work
  - Flight of the bumblebee but at  $2/3$  speed
  - Mostly influenced by the Japanese nationality
  - Ding sounds, crowd noise
  - Turnstiles, foreign-language chattering, an industrial hum
  - Almost anything Philip Glass composed, everything he does sounds dynamic

## Appendix C: Adjective-Pairs Survey

This was a survey designed to determine if the adjective pairs used to describe musical qualities in GarageBand and Musicbed DV make sense to a user coming unprepared to these applications.

### Attach Concrete Terms to Abstract Notions!



Two "soundtrack" applications use the following sets of terms to define their musical categories. If you were whipping up music for your video, would these classifications make sense to you? How would you rank the importance of these categories. If a particular term makes little sense, what term would you substitute? If you have more ideas, write them on the back of this form. Thanks!

#### Musicbed DV

Your Ranking	Category Names and Subcategory Pairs		Your Substitute Term	Comments
	Style			
	Abstract	Real		
	Background	Foreground		
	Melodic	Percussive		
	Sparse	Dense		
	Color			
	Light	Intense		
	Clean	Lush		
	Smooth	Sharp		
	Low	High		

#### GarageBand

Your Ranking	Category Pairs		Your Substitute Term	Comments
	Single	Ensemble		
	Clean	Distorted		
	Acoustic	Electric		
	Relaxed	Intense		
	Cheerful	Dark		
	Dry	Processed		
	Grooving	Arrhythmic		
	Melodic	Dissonant		
	Part	Fill		

## Appendix D: Musicbed DV Genres and Moods

In addition to the categories listed in the Adjective-Pairs survey, Musicbed DV offered two more controls intended to help direct the user to relevant musical content. These controls were two drop-down lists labeled "Genres" and "Moods."

- Genres
  - Action
  - Biography
  - Crime
  - Drama
  - History
  - Mystery
  - Sports
- Moods
  - Aftermath
  - Change of Plans
  - Climax
  - Final Push
  - Higher Stakes
  - Major Setback
  - New Situation
  - Opportunity
  - Point of No Return
  - Progress
  - Setup

## Appendix E: QSketcher Quotation

This is an excerpt from the QSketcher document that is quite relevant to the soundtrack project in particular and interaction design in general.

The music composition process has evolved over millennia as a balance of opposites: inspiration versus perspiration, broad formal approaches versus minute detailed work, macro-level (or structural) conceptualization versus micro-level (e.g., note-level) editing, and so on. When composing in the traditional ways, one moves frequently from mode to mode: an idea pops up, one captures it, and then thinks critically about and develops it. Another idea pops up, and the process iterates. One can easily shift focus to thinking about the entire composition and how a small motif relates to the whole, for example.

Unfortunately, the technological tools available for composing music do not readily support this kind of creative workflow - they are geared towards realization of preconceived ideas. Thus, the user interface found in many commercial applications models recording studio machinery, namely the multi-track recorder and mixer. This model has little to do with the traditional process of creating music. Therefore, although the multi-track recorder metaphor is appropriate for the mixing and audio post-production stages, it provides little advantage in capturing and developing musical ideas, which lie at the heart of the early stages of the creative process.

In addition, the environments for composing music that have evolved over five decades in academic and commercial settings still fall short in their ability to manipulate music directly in terms of musical concepts. Computer languages for composing music, for example, typically model limited structural aspects of music through a score, e.g. instrument, part and score, and provide a suite of algorithms to manipulate that score. The score is often expressed in physical terms, such as frequency and amplitude, and the tools provided for its manipulation are often borrowed from computer science with few changes. This detachment from the intuitive musical concepts in which composers think, and from the musical experience itself, places an unnecessary cognitive burden on the composer. Computers are supposed to make work easier, not harder!

Our goal was to develop a tool that overcomes these limitations. To focus our efforts, we addressed a specific musical task: that of scoring film. We felt that this would not overly restrict the musical style, that general issues relating to supporting computer-assisted composition would still need to be addressed, and that the film environment in particular provides structure that resonates with many of our ideas. We formed a cross-disciplinary research team consisting of computer music researchers, system developers, and composers (note that all team members have strong musical backgrounds; most have strong technical backgrounds). Together, we examined the creative workflow, critiqued existing solutions, and focused on areas that we felt needed attention. We also employed a highly iterative development strategy with our composers involved in most stages. Our focus on the creative workflow and musical concepts had a profound impact on the system design, from the underlying data structures used for music representation to the overall user environment.

The result is QSketcher, an environment designed to allow composers to move fluidly between dichotomous modes (inspiration/perspiration, capture/manipulate, and macro/micro editing levels), while directly supporting a variety of common compositional concepts, so that composers can work using the terms in which they think.

Interestingly, these dichotomous modes are also present in a wide variety of creative tasks—writing, drawing, preparing presentations, as well as more technically-oriented activities such as software design and development, architecture, and even the act of research itself. The concepts uncovered during the QSketcher project are important in these domains, and many of the solutions implemented would prove quite useful there. Several of the authors are currently adapting these techniques to other areas. In fact, we believe that the “tip of the iceberg” of an important research area in HCI is emerging: developing information technology tools to support creative work.

## Appendix F: Cobi van Tonder's *Skatesonic* Interview

Sylvie Parent interviewed Cobi van Tonder about the IDEO-sponsored *Skatesonic* project for the ZeroOne San Jose Festival 2006 web site.

Cobi van Tonder: *Skatesonic* taps into skateboard culture, using the motions and sounds of skateboards to create music. In a way, each move translates to musical parameters and the rider ends up skating through a landscape of music (which s/he influences over time).

...

Sylvie Parent: Skateboarding is related to urban space and although it is an individual activity it often involves peers. Teens and young adults organize gatherings and share their experience. What attracted you to this subculture? How is *Skatesonic* going to be an individual and/or group experience?

CvT: It is wild/fearless/extreme/punkd/crazy. Also, I love the sea, but end up spending so much time in cities. Skateboarding brought the waves into the city, or it turns the city not just upside down, but into waves, and waves are music! *Skatesonic* will work in both solo and group situation. There are going to be 4 boards, each board mapping to a unique sound and structural parameters, so if there are 4 riders they will be able to jam like a band.

SP: Standing on a skateboard and riding it supposes certain ways of moving your body in space. It involves a play with balance, the loss/recovery of stability. Skateboarding can also be a mode of transportation. It implies mobility. Is this specific body language important to the development of the project?

CvT: Yes, all of this implies rhythm and mood. Music is an emotional kind of landscape that enables travel. Connecting the gestures of a physical reality with musical form and texture implies superimposing these two realities on top of each other so that they fold back on themselves, thus creating a strange new space.

S.P. : What kind of sound or music can we expect from riding/playing the skateboard?

CvT: Crazy music. It should be a surprise.

SP: As a participant experiments with *Skatesonic*, s/he will learn how to play it, how to produce sounds and effects. The skateboard is also recording new sounds. Can you tell us a few words about the process? How does the *Skatesonic* work?

CvT: There are four basic elements: input from sensors on the board, sounds, a musical score and a system to interconnect these elements. Input from the sensors includes acceleration, 3D-spatial positioning, optical input and a microphone inside the skateboard trucks. Sounds include a library of prepared samples and virtual instruments as well as the raw microphone input. The musical score consists of a vocabulary of events corresponding to each possible move, as well as a musical timeline. The core Max/MSP patch is a set of conditions that links up input situations to score and sounds.

All kinds of things will happen. You can, for example, buffer through a sound file in Max, meaning that as the rider rolls over a certain distance it is as if s/he has a record needle under the board, and every inch of movement progresses the sound. The live microphone input also reveals rich information about the texture of surface under the board and intensity of movement. This can be mapped to all kinds of sound parameters and has a very connected, live effect since it is so direct.

...

SP: Yes, exactly. Another interesting aspect of the work is that it involves the whole body. Can you comment on that?

CvT: Yes, the body is the form of us. The abysmal librarian/collector of experience and perception. It is so expressive. I like to look at people - such infinite stories people tell with their bodies. During my residency a local orthopedist showed me some of the technology they use with patients, which I found very interesting. The next step would be to work directly with pulse, EEG, skin response, eye-movement, xray, sonar, etc. It is fascinating and also very real - people dealing with injury or

illness. A hospital is also a space that puts my imagination in overdrive. I think there is much potential for theatre in that. Beauty is often found in things that are "wrong," you know? Show me your weakness.

Speaking of which, I am @#\$%\* scared of skateboarding. I didn't do these kinds of things as a kid. Too risky. I was a pianist and violinist. So I am amazed at what these skaters can do - they seem fearless. A kind of freedom.

...

SP: How will *Skatesonic* be shown during ZeroOne San Jose/ISEA2006? Will it be installed for the general public in a venue? And will it be used in a special performance with experienced skateboarders (or performers/dancers, acrobats!)?

CvT: I hope that you will be able to hear *Skatesonic* from afar. It will be in front of the new City Hall, starting with a performance and flowing into a free improv/play space at designated hours.

SP: What is the relationship of *Skatesonic* with space? How does it relate to its environment?

CvT: *Skatesonic* "listens" to space through movement, which it maps out and translates into music. I like to think of the skateboard as a creature that I call the LicKr. Imagining what the world sounds and feels like living and moving at great speed that close to the ground. The LicKr is a brainless body, or actually a flattened skull, with long tongue on wheels that licks the asphalt, tasting every bit of it. It's a sensor creature dependant on its riders; mostly a moody and easily aggravated creature and super hero of hers own transmission fiction, spitting out syllables of noise, texture, beats and spatial patterns.

Physically, the skateboard is absorbed into the body (similar to the way a music instrument becomes part of a musician), and allows a skateboarder to use the architecture of a city. This is such a creative way of dealing with the city! Every nook and cranny becomes a surface that is encountered in a direct way. This reminds me of human touch on a music instrument (and the touch of the instrument on the musician). Sound reveals space, be it physical, social or imaginary.

Of course the real sounds (unamplified) of skateboard riding are also directly connected to the environment in which they are created. One could say it is a layer of the environment. *Skatesonic* partly records data and sounds, allowing for repetition, difference, layering, erasure, memory. A pre-composed musical architecture or form to allow boundaries within which to control the musical outcome is created. (Parent 2006)

## Appendix G: Glossary of Terms

All definitions are courtesy of Wikipedia.Org, as of April 22, 2006, unless otherwise noted.

Alfred Newman (March 17, 1900 - February 17, 1970) was a major composer of music for films. He received 45 Academy Award nominations (a record in the music categories, now shared with John Williams), winning nine times. He also composed the familiar fanfare which accompanies the studio logo for 20th-Century Fox, where he headed the music department.

Artificial Intelligence (AI): AI is concerned with producing machines to automate tasks requiring intelligent behavior. Examples include control, planning and scheduling, the ability to answer diagnostic and consumer questions, handwriting, speech, and facial recognition. As such, it has become a scientific discipline, focused on providing solutions to real-life problems. AI systems are now in routine use in economics, medicine, engineering and the military, as well as in many common home computer software applications, traditional strategy games like computer chess and other video games.

Autotelic (from Project Bar-B-Q's text): An activity worth doing for its own sake.

Bounded Rationality (from the Oxford University Press web site): An influential concept developed by the Chicago School of the 1950s and 1960s. It challenged rationalist perspectives and suggested that the rationality of actual economic and firm behaviour was always partial, or 'bounded' by human limitations.

**Fake Book:** A fake book is a collection of simplified sheet music. A song in a fake book contains the melody line, basic chords, and lyrics, the minimal information for musicians adept in improvising to make an impromptu arrangement of a song, or "faking it". Fake books are not intended for novices: The reader must follow and interpret the scant notation. For someone playing music as a hobby, fake books can be an avenue to playing songs very quickly.

**Interaction Design:** Interaction design is the branch of user experience design that illuminates the relationship between people and the machines they use. While interaction design has a firm foundation in the theory, practice, and methodology of traditional user interface design, its focus is on defining the complex dialogues that occur between people and interactive devices of many types—from computers to mobile communication devices to appliances.

**Kinesthetic:** The sense of the position of parts of the body, relative to other neighbouring parts of the body.

**Pandora.Com** (from the Pandora web site): Pandora is a music discovery service designed to help you find and enjoy music that you'll love. It's powered by the Music Genome Project, the most comprehensive analysis of music ever undertaken. Just tell us one of your favorite songs or artists and we'll launch a streaming station to explore that part of the music universe.

**Spotting** (from Davis 1999, 83): [To 'spot' video is to determine] where the music will begin and end for each cue, what it should sound like, and what role it will play in relation to the drama.

Temp Track (from Davis 1999, 96): A temporary track of music laid into the work-print of the film in order to give studio executives and test audiences an idea of what the film will be like once the final score is completed.

## References

Abrams, S., Bellofatto, R., Fuhrer, F., Oppenheim, D., Wright, J. (IBM Research); Boulanger, R., Leonard, N., Mash, D., Rendish, M., Smith, J. (Berklee College of Music). (2001). "QSketcher: An Environment for Composing Music for Film", *Proceedings of the International Computer Music Conference*, Havana. Retrieved March 1, 2006 from, <http://www.openmuse.org/noncpl/QSketcherICMC2001Paper.pdf>.

*Apple GarageBand 3 Movie Score promotional web page*. Retrieved April 10, 2006, from <http://www.apple.com/ilife/garageband/features/scoring.html>.

Bailey, R. W. (1996). *Human Performance Engineering: Designing High Quality Professional User Interfaces for Computer Products, Applications and Systems*. Upper Saddle River, NJ: Prentice Hall PTR.

*BBC Motion Gallery*. Retrieved February 1, 2006, from <http://www.bbcmotiongallery.com>.

Bell, David. (1994). *Getting the Best Score for Your Film: A Filmmakers' Guide to Music Scoring*. Los Angeles: Silman-James Press.

- Boland, R. J. and Collopy, F.(Eds.). (2004). *Managing as Designing*.  
Stanford: Stanford Business Books.
- Boltz, Marilyn. (2001). "Musical Soundtracks as a Schematic Influence on the  
Cognitive Processing of Filmed Events," *Music Perception* 18: 427–  
455.
- Boltz, Marilyn. (2004). "The Cognitive Processing of Film and Musical  
Soundtracks," *Memory and Cognition* 32: 1194–1205.
- Borchers, Jan. (2001). *A Pattern Approach to Interaction Design*.  
Chichester: John Wiley & Sons.
- Borja de Mozota, Brigitte. (2003). *Design Management: Using Design to  
Build Brand Value and Corporate Innovation*. New York: Allworth  
Press.
- Brooke, John (n.d.) *SUS - A quick and dirty usability scale*. Retrieved  
February 1, 2006 from  
<http://usabilitynet.org/trump/documents/Suschapt.doc>.
- Brougher, K. et al. (2005). *Visual Music: Synaesthesia in Art and Music  
Since 1900*. New York: Thames & Hudson.

- Bullerjahn, C. and Güldenring, M. (1994). "An Empirical Investigation of Effects of Film Music Using Qualitative Content Analysis," *Psychomusicology* 13, 99–118.
- Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience*. New York: HarperCollins.
- Coe, M. (1996). *Human Factors for Technical Communicators*. New York: Wiley.
- Cohen, Annabel J. (Ed.). (1994). *Psychomusicology: A Journal of Research in Music Cognition*, Vol. 13. Tallahassee, FL: Florida State University.
- Cohen, Annabel J. (2001). "Music as a Source of Emotion in Film," in *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.
- Cohen, Annabel J. (2005). "How Music Influences the Interpretation of Film and Video: Approaches from Experimental Psychology," in *Perspectives in Systematic Musicology*, Kendall, R. A. and Savage, R. W. H. (eds.). Los Angeles: Department of Ethnomusicology, UCLA.
- Cooper, A. and Reimann, R. (2003). *About Face 2.0: The Essentials of Interaction Design*. Indianapolis, IN: Wiley.

- Cope, D. (2001). *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, MA: MIT Press.
- Davis, Richard. (1999). *Complete Guide to Film Scoring: The Art and Business of Writing Music for Movies and TV*. Boston: Berklee Press.
- Dean, R. (2003). *Hyperimprovisation: Computer-Interactive Sound Improvisation*. Middleton, WI: A-R Editions.
- Deutsch, Diana. (1999). *The Psychology of Music, Second Edition*. Academic Press: San Diego.
- Elton, Catherine. (2006, April 5). "Measuring emotion at the symphony," *The Boston Globe*. Retrieved April 22, 2006 from [http://www.boston.com/ae/music/articles/2006/04/05/measuring\\_emotion\\_at\\_the\\_symphony/](http://www.boston.com/ae/music/articles/2006/04/05/measuring_emotion_at_the_symphony/).
- Eno, Brian. (1975). Liner notes from *Discreet Music*. Retrieved April 21, 2006 from [http://music.hyperreal.org/artists/brian\\_eno/discreet-txt.html](http://music.hyperreal.org/artists/brian_eno/discreet-txt.html).
- Gigerenzer, G. and Todd, P. M. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World*.

New York: Oxford University Press.

Gigerenzer, G. and Selten, R. (Eds.). (2002). *Bounded Rationality: The*

*Adaptive Toolbox*. Cambridge, MA: MIT Press.

Gilovich, T. D. Griffin, and Kahneman, D. (2002). *Heuristics and Biases:*

*The Psychology of Intuitive Judgment*. Cambridge: Cambridge

University Press.

Gladwell, Malcolm. (2005). *Blink: The Power of Thinking Without*

*Thinking*. Little, Brown and Co.: New York.

*IBM Computer Music Center 1993 - 2001*. Retrieved March 1, 2006 from

<http://www.openmuse.org/noncpl/aboutcmc.html>.

*IBM Research: Services Sciences, Management, and Engineering*.

Retrieved November 6, 2005, from

<http://www.research.ibm.com/ssme/index.shtml>.

Iwamiya, Shin-ichiro. (1994). "Interactions Between Auditory and Visual

Processing When Listening to Music in an Audio Visual Context: 1.

Matching 2. Audio Quality," *Psychomusicology* 13, 133–154.

Juslin, Patrik N. and Sloboda, John A. (2001). *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.

Laurel, B. (1993). *Computers as Theatre*. Reading, MA: Addison-Wesley.

Laurel, B. (2003). *Design Research, Methods and Perspectives*.  
Cambridge, MA: MIT Press.

Lipscomb, S. and Kendall, R. (1994). "Perceptual Judgement of the Relationship Between Musical and Visual Components in Film," *Psychomusicology* 13, 60–98.

Löwgren, Jonas and Stolterman, E. (2004). *Thoughtful Interaction Design: A Design Perspective on Information Technology*. Cambridge, MA: MIT Press.

Parent, Sylvie. (2006). "Skatesonic," *ZeroOne San Jose web site*. Retrieved April 22, 2006 from <http://01sj.org/content/view/461/49/>.

Simon, H. A. (1996). *The Sciences of the Artificial, Third Edition*. Cambridge, MA: MIT Press.

Spiegel, Laurie. (2006). *Music Mouse* [Computer software]. Retrieved April 1, 2006 from <http://reitary.org/ls/programs.html>.

- Nemeth, C. P. (2004). *Human Factors Methods for Design: Making Systems Human Centered*. Boca Raton, FL: CRC Press.
- Nielsen, J. (1993). *Usability Engineering*. Boston, MA: AP Professional.
- Norman, D. (1988). *The Design of Everyday Things*. New York: Doubleday.
- Owen, David. (2006, April 10). "The Soundtrack of Your Life: Muzak in the Realm of Retail Theatre," *The New Yorker*. Retrieved April 11, 2006 from  
[http://www.newyorker.com/fact/content/articles/060410fa\\_fact](http://www.newyorker.com/fact/content/articles/060410fa_fact).
- PG Music Band-in-a-Box web site promotional blurb*. Retrieved April 10, 2006 from, <http://www.pgmusic.com/band.htm>.
- Preece, J., Y. Rogers, and Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction*. New York: Wiley.
- Pressing, J. (1992). *Synthesizer Performance and Real-Time Techniques*. Madison, WI: A-R Editions.
- Project Bar-B-Q's Ninth Interactive Music Conference Group Report: The Stroke-a-phone: A New Digital Instrument for Troubled Times*. Retrieved November 6, 2005 from  
<http://www.projectbarbq.com/bbq04/bbq04r6.htm>.

- Schaffer, J. and McGee, D. (1997). *Knowledge-Based Programming for Music Research*. Madison, WI: A-R Editions.
- Schön, D. (1987). *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass.
- Simon, H. A. (1996). *The Sciences of the Artificial, Third Edition*. Cambridge, MA: MIT Press.
- Sonnenschein, David. (2001). *Sound Design: The Expressive Power of Music, Voice, and Sound Effects in Cinema*. Studio City: Michael Wiese Productions.
- Synk Audio Studios. (2006). *Musiced DV* [Computer software]. Retrieved March 1, 2006 from, <http://web.synkaudiostudios.com>.
- Todd, P. M. and Werner, G. M. (1998), "Frankensteinian Methods for Evolutionary Music Composition." In *Musical Networks: Parallel Distributed Perception and Performance*, edited by N. Griffith and P. M. Todd. Cambridge, MA: MIT Press.
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire, CN: Graphics Press.

Walker, Michael. (2006, April 3). "Software That Can Turn You Into a Songwriter," **The New York Times**. Retrieved April 3, 2006, from <http://www.nytimes.com/2006/04/02/arts/music/02walk.html>.

Windsor, P. (1992). *Automated Music Composition*. Denton, TX: University of North Texas Press.

Winkler, T. (1998). *Composing Interactive Music: Techniques and Ideas Using Max*. Cambridge, MA: MIT Press.