Ritter, Don (2013) "A Flexible Approach For Synchronizing Video With Live Music."
in Live Visuals, Leonardo Electronic Almanac/MIT Press, Vol 19, No 3. p. 30-43.

A Flexible Approach for Synchronizing Video with Live Music

by
Don Ritter
School of Creative Media
City University of Hong Kong
aesthetic-machinery.com

INTRODUCTION

Performances of synchronous video and music have become increasingly popular over the past 15 years, supported by a proliferation of appropriate software. [1] Live music events, nightclubs, operas, and world fairs are now popular venues for a performance format that has existed for nearly 300 years under various monikers, including ocular music, visual music, color-music, music for the eyes, interactive music-video and, now, VJing. [2] [3] [4] [5] Louis-Bertrand Castel (1688–1757), a Jesuit priest, mathematician and philosopher, is acknowledged for creating the first device that synchronized imagery with music in 1734.[6] Castel's instrument, known as an "ocular clavichord," exposed small pieces of colored glass when its keys were pressed, each key being associated with a different color. [7] Karl von Eckartshausen (1752–1803) created an instrument similar to Castel's in 1791, which he describes in his book *Disclosures of Magic from Tested Experiences of Occult Philosophic Sciences and Veiled Secrets of Nature:*

> *I have long tried to determine the harmony of all sense impressions, to make it manifest and perceptible. To this end I improved the ocular music invented by Pere Castel. I constructed this machine in all its perfection, so that whole chords can be produced, just like tonal chords. Here is a description of the instrument. I had cylindrical glasses, about half an inch in diameter, made of equal size, and filled them with diluted chemical colors. I arranged these glasses like the keys of a clavichord, placing the shades of color like the notes. Behind these glasses I placed little lobes of brass, which covered the glasses so that no color could be seen. These lobes were connected by wires with the keyboard of the clavichord, so that the lobe was lifted when a key was struck, rendering the color visible. Just as a note dies away when the finger is removed from a key, so the color disappears, since the metal lobe drops quickly because of its weight, covering the color. The clavichord is illuminated from behind by wax candles. The beauty of the colors is indescribable, surpassing the most splendid of jewels. Nor can one express the visual impression awakened by the various color chords.* [8]

ISSN 1071-4391 ISBN 978-1-906897-22-2

ABSTRACT

Performances of synchronous imagery and music have existed for almost 300 years, but they have become popular only in the past 15 years. Synchronous imagery and music refers to their simultaneous presentation, but synchronization refers to an interactive correlation between the imagery and music. The article will discuss pioneering works that use synchronization by examining their strategies for correspondence. An absolute approach to synchronization uses fixed correlations between musical and visual characteristics, but a flexible approach permits the selection of music, imagery, and forms of correspondence. The paper discusses the characteristics of a computer-based system that uses a flexible approach for synchronizing video with live music. This flexible system imposes no specific rules for synchronizing video with music, and it provides a programming environment that enables the creation of a composition using various forms of correspondence. The paper is based on the author's work since 1988 on the creation of software and performances of video synchronized with live music.

I consider synchronous imagery and music to be the simultaneous presentation of these two media, such as a piano accompanying a silent film or a VJ performing with recorded music. The synchronization of imagery and music also involves the concurrent presentation of these two media, but the imagery and music are intricately correlated. Using this distinction, performances of synchronous imagery and music can be differentiated according to the presence or absence of synchronization. The ocular clavichord created by

ISSN 1071-4391 ISBN 978-1-906897-22-2

Castel enabled synchronized performances because pressing its keys produced a musical pitch and a specific color of light corresponding with the pitch.

The synchronization of imagery and music can be created by having music correlate with imagery, imagery correlate with music, or though an independent action that directs the correspondence in unison. An important discriminator for this synchronization is whether an *absolute* or a *flexible* approach is used for the correspondence. The instruments created by Castel and von Eckartshausen used absolute approaches because their devices would always provide the same color of light in response to a particular note being pressed, regardless of the fingering style of the performer or the desires of the composer. In contrast, a flexible approach imposes no specific rules for synchronizing imagery with music, and it provides versatility by permitting a composer to change the imagery, the music, or the rules of correspondence between imagery and music.

This aim of this paper is to discuss the differences between an absolute and flexible approach for synchronizing video with live music during a performance. These ideas were developed between 1988 and 1992 when I developed a software that provided interactive control of video through live music. This software was born near the end of my graduate studies at MIT, and I continued developing it for four more years. The software, later christened *Orpheus,* enabled interactive control of full-screen, digital video through any musical instrument. Orpheus could be controlled by electronic music instruments that provided MIDI output [9] or by acoustic instruments in conjunction with a hardware device that converted audio into MIDI.

I collaborated with various musicians during those years, and we presented performances of interactive video controlled by trombone, synthesizer, saxophone, trumpet, vocals, electronic drums, electric guitar, and sampling keyboard. These events were performed at numerous venues, including the MIT Media Lab, New Music America in New York and Montreal, The Art Institute of Chicago, STEIM in Amsterdam, and the Verona Jazz Festival. Most of these performances were collaborations with trombonist and interactive music pioneer George Lewis. Our performances featured electronic music and large video projections that were both controlled by Lewis's improvisation on trombone. Orpheus provided the interactive video for these events, while Lewis's interactive music software – called *Voyager* – produced the electronic music. [10]

In addition to creating Orpheus and presenting many performances, the other outcome from that work was the development of a general strategy for synchronizing video with live music. The primary problem I had during the creation of Orpheus was not the writing of the software, it was deciding how music and video should correspond. What eventually evolved was a strategy that I called *a Flexible approach for synchronizing video with live music.* The goal of this method is to provide flexibility when synchronizing video with music by enabling the use of any type of video imagery or musical instrument, and by providing a high level of versatility when composing the correspondence between music and imagery.

The following pages present a detailed description of this method as it was finalized in 1992. I will occasionally discuss how this approach was used within Orpheus, but the core ideas are independent of specific computer hardware, software, musical styles, musical instruments, or media formats. Many of the references cited in this paper were obtained between 1988 and 1992, but I believe they are still relevant today, including Adrien Bernard Klein's observation from his 1927 book, *Color-Music: The Art of Light:* "...it is an odd fact that almost everyone who develops a color-organ is under the misapprehension that he, or she, is the first mortal to attempt to do so." [11]

ISSN 1071-4391 ISBN 978-1-906897-22-2

SYNCHRONOUS IMAGERY AND MUSIC

Prometheus: The Poem of Fire is the symphonic composition written by Alexander Scriabin in 1910 that included synchronous imagery. [12] The part of the score labeled as "Luce" contained directions for playing a keyboard that controlled 12 colored lights. This keyboard was played as accompaniment to the live music, though it produced no music. Scriabin had initially used an absolute correspondence between the music and the lights – assigning a different colored light to each of the 12 pitches in the chromatic scale – but he later rejected that approach and began organizing color into counterpoint relationships [13] with the music. The first performance of *Prometheus* was presented in 1915 at Carnegie Hall in New York City, and it was presented more recently at Yale University in 2010 using modern lighting and MIDI technologies. [14]

Film director and theorist Sergei Eisenstein (1898–1948) was a pioneer of synchronous music within film, which he discusses in his books *The Film Sense* (1942) [15] and *The Film Form* (1949). [16] Eisenstein proposed that sound and music should be combined with imagery to create a specific mood, rather than attempting to create a realistic correspondence between imagery and sound. He writes, "The search for correspondence must proceed from the intention of matching both picture and music to the general, complex 'imagery' produced by the whole." [17] Experiencing synchronous imagery and music within a film is obviously different from attending a performance of imagery synchronized with live music. An audience that attends a film does not experience the creation of imagery and sound, they experience a recording of disconnected elements that were created at various production locations, a process that is similar to the creation of recorded music. Miller Puckette, a pioneer of interactive music software, wrote in 1991 that "An essential part of 'real' music is the live element, the indefinable but undeniable interaction between players and audience which makes music exciting. It is hard to prove this interaction is there...but any musician will swear to its reality, and to its importance to the music-making process." [18]

SYNCHRONIZED IMAGERY AND MUSIC

The instruments that Castel and von Eckartshausen created in the eighteenth century for synchronizing imagery with music would be categorized today as being interactive media devices. The definition of interactive media that I will adopt in this article is the following: a media environment that enables a user to direct significant perceptual changes in media content.

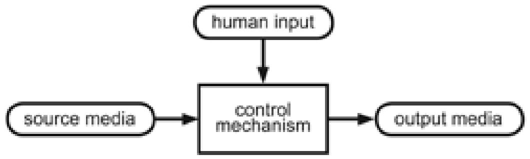

Fig 1. *Primary components of an interactive media system.* © Don Ritter, 2013. Used with permission.

Figure 1 diagrams the primary components of an interactive media system based on this definition. The *human input* component refers to the actions required by a user for directing the interactive media, such as a finger gesture, body motion, or voice. The *source media* component refers to media content that can potentially be output by the system, such as a specific collection of text, graphics, moving imagery, or sound. The *control mechanism* activates the correspondence rules that specify which source media will be output in accordance with the human input.

The *degree of interactivity* within an interactive media system is dependent on the number of opportunities provided to a user for directing the media output. For example, an interactive system has a low degree of interactivity if it provides a user with 2 options every

ISSN 1071-4391 ISBN 978-1-906897-22-2

30 minutes, but it has a high degree of interactivity if 100 options are available every 2 seconds. The degree of interactivity within an interactive system can be enhanced by increasing the number of human gestures that are recognized by the human input component or by increasing the amount of media available to the control mechanism, but only if these increases are incorporated into the control mechanism.

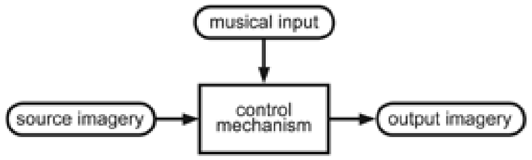

Fig 2. *Primary components of a system for synchronizing imagery with music*. © Don Ritter, 2013. Used with permission.

Figure 2 diagrams the primary components within an interactive media system that provide synchronization of imagery with live music. In this system, the control mechanism directs the correspondence between the music being performed and the imagery being output by the system. The quantity of source imagery and the number of musical input characteristics that are incorporated into the control mechanism determine the system's degree of interactivity. The control mechanism has the following functions: to recognize the format and quantity of source imagery; to recognize characteristics of the musical input; to maintain a collection of correspondence rules between the musical input, source imagery, and output imagery; and to direct the output imagery according to activation of these rules. An example of a correspondence rule is the pairing of a specific image with a particular musical pitch. Castel used this approach in his ocular clavichord, which used the following 12 correspondence rules for pairing color with musical pitch. [19]

C : blue
C-sharp : blue-green
D : green
E-flat : olive green
E : yellow
F : yellow-orange
F-sharp : orange
G : red
A-flat : crimson
A : violet
B-flat : agate
B : indigo

Scriabin had initially used a similar method of correspondence in his *Prometheus* symphony; though his approach had paired different colors with the pitches. [20]

C : intense red
C-sharp : violet or purple
D : yellow
D-sharp : flesh color
E : sky blue
F : deep red
F-sharp : bright blue or violet
G : orange
G-sharp: violet or lilac
A : green
A-sharp : rose or steel
B : blue or pearly blue

Three areas of composition are involved in the creation of synchronized imagery and music: the imagery, the music, and the correspondence between them. The instruments created by Castel and Scriabin used 'absolute approaches' for synchronization; their control mechanisms, source imagery, and the musical characteristics associated with the correspondence are fixed: a certain color is displayed when a specific key is pressed, regardless of the composition being played.

ISSN 1071-4391 ISBN 978-1-906897-22-2

ARTICLE

Castel's approach was based on Isaac Newton's proposal that the relative wavelengths of different hues could be paired with the relative pitches of notes, which he justified mathematically. [21] Two centuries later, Sergei Eisenstein criticized an absolute approach to image and sound correspondence, "The question must, nevertheless, be undertaken, for the problem of achieving such absolute correspondence is still disturbing many minds, even those of American film producers. Only a few years ago I came across in an American magazine quite serious speculations as to the absolute correspondence of the piccolo's tone to – yellow!" [22]

Eisenstein's approach differed from Castel's by proposing that a 'desired mood' could be the determining element for the correspondence between music and imagery, rather than a fixed relationship based on the physical properties of light and sound. He describes several methods for combining sound and imagery within the *Film Sense*, including correspondence according to the meter, phrase, rhythm, or timbre of music:

> *It is important to keep in mind that our conception of synchronization does not presume consonance. In this conception full possibilities exist for the play of both corresponding and non-corresponding "movements," but in either circumstance the relationship must be compositionally controlled. It is apparent that any one of these synchronization approaches may serve as the "leading" determining factor in the structure, dependent on the need. Some scenes would require rhythm as a determining factor, others would be controlled by tone, and so on.* [23]

A FLEXIBLE SYSTEM FOR SYNCHRONIZING VIDEO WITH LIVE MUSIC

Eisenstein's writings were very influential in the creation of a flexible system for synchronizing video with live music because his ideas emphasized the creation of multi-sensory composition rather than a synchronization mechanism. An instrument that uses an absolute approach for synchronizing imagery with music provides a composer with unchangeable rules of correspondence, thereby limiting the potential of a composition. The alternative to an absolute approach for synchronizing imagery with music is a 'flexible' method that permits a composer to decide the correspondence between music and imagery. A flexible approach has a practical advantage over absolute methods because it is encompassing. If a composer wants a particular color to be presented consistently in response to a particular pitch, a flexible system permits this relationship, along with many other types of correspondence. A system that uses absolute correspondence can only be used in accordance with the rules inherent in itself.

My knowledge of music theory and composition was quite limited when I started to create Orpheus 24 years ago. Consequently, I adopted certain terminology and a perspective of musical structure that reflected the mechanical operation of this software. The purpose of the terminology was to create a conceptual model that assisted in the development of a system that used a flexible approach for synchronization. Figure 3 depicts this model, followed by explanations of its components and terminology.

The *music-listener* component of this system provides a continuous stream of features and categorizations that continually describe changes in the structure of the incoming live music. This analysis is used for activating video actions in conjunction with correspondence rules that are contained within the 'control mechanism.' A *video action* is comprised of single or multiple video sources – such as video clips, live video,

ISSN 1071-4391 ISBN 978-1-906897-22-2

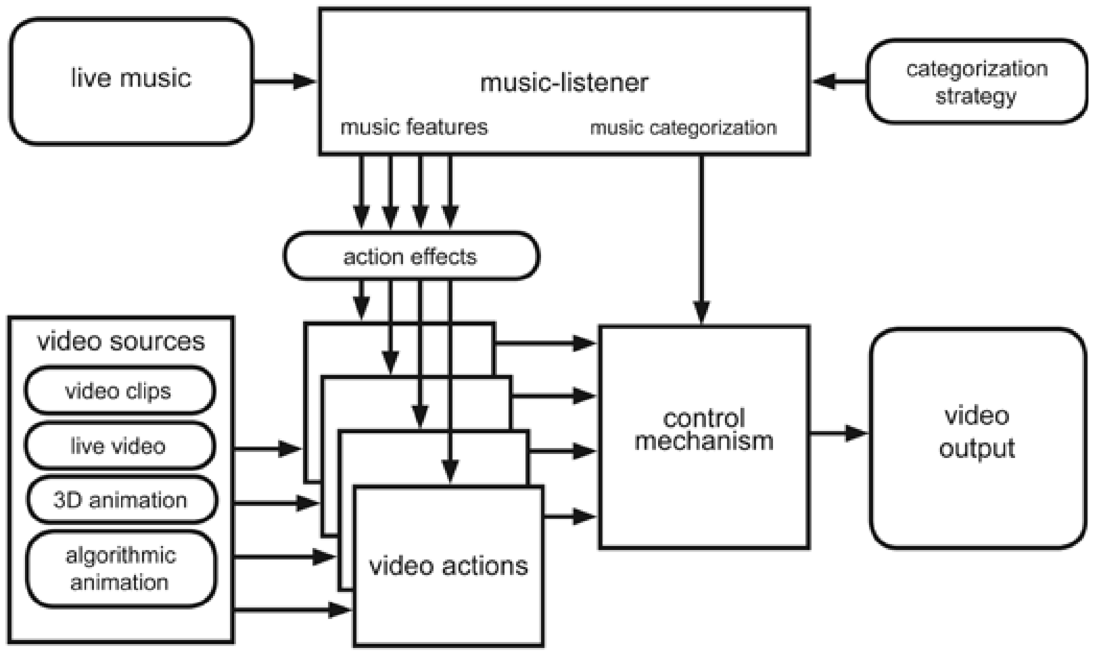

Fig 3. *A flexible system for synchronizing video with live music*. The music-listener extracts features from live music for directing video output using the correspondence rules contained in the control mechanism. © Don Ritter, 2013. Used with permission.

3-D animation or algorithmic animation – in combination with a method for playback, called an *action effect.*

The specific musical characteristics that control the video actions are determined by a strategy for categorizing the music. This configuration provides three forms of flexibility for composing a performance of video synchronized with live music: the construction of video actions, selection of a strategy for analyzing music, and selection of correspondence rules between video and music. The remainder of this article describes the components of this system in more detail.

VIDEO ACTIONS

An initial limitation of Orpheus was the duration of video it could control. The software used full-screen, digital video clips that were activated to loop forward or backward from a computer's memory. The maximum amount of memory that the computer could accommodate – which was 18 MB in 1988 – determined the total number and duration of clips that could be used within a performance. That year, the average price of memory was $333 per MB, or 8000 times more than today's price. [24] The limitation imposed by my financial empire was overcome by 1991, and the computer's limitation was solved through the notion of a video action.

Within Orpheus, a video action is comprised of a specific collection of video frames, an action effect, and various control parameters. An example of an action effect is *skip-to-note-oscillate.* If this action effect is used within a video action, the relative pitch of the incoming music determines the playback speed of the video; lower pitches produce slower moving imagery and higher pitches produce faster imagery. An adjustable parameter for this action effect is the reference pitch for determining the playback speed. When a

ISSN 1071-4391 ISBN 978-1-906897-22-2

video action uses the skip-to-note-oscillate effect, the frame order oscillates forward and backward, while the speed of the clip is determined by the pitch difference between the incoming music and the reference pitch. The skip-to-note-oscillate effect can be adjusted to play for certain duration of time, or it can play continuously until stopped by a specific musical event.

The use of video actions was created to overcome the memory limitation imposed by the computer, but this notion is independent of the format of the source video. A video action can be comprised of a digital video clip, live video from a camera, pre-rendered animation, algorithmic animation generated in real time, or 3D animation generated in real time. Orpheus has 70 different action effects that can be used to create video actions, each adjustable through various parameters. By using a specific collection of frames within different video actions, 15 minutes of video footage can be created from 15 seconds of source video stored in the computer's memory. At the early stage of the software's development, different video actions were activated sequentially and according to the number of musical notes received by the software. For example, video action 1 would respond to the first 10 music notes, video action 2 would respond to the next 15 notes, video action 3 would respond to the next 12 notes, and so on. Eventually the software could analyze music to a greater extent, and various musical characteristics became associated with video actions in a non-sequential manner.

The next level of flexibility to be incorporated was the arrangement of video actions into lists, with each list associated with a particular category of music analysis. By that stage of development, a specific video action was not associated with a particular musical feature; it became part of a list that was activated according to combinations of music features and categorizations. Stately simply, the software became more interactive by recognizing additional features within the music and using these numbers to control the video more intricately.

MUSIC-LISTENER

In order for an interactive system to synchronize video precisely with musical input, it must recognize the continuously changing characteristics of the music. Using music as an input to a human interface is significantly different from using finger gestures. Most interactive media use a human interface that is activated through fingers on a touch screen, computer keyboard, or mouse. With these devices, the number of options available to a user at a decision point is typically determined by a two dimensional arrangement of buttons. This spatial approach, however, cannot be used effectively with music because sound is a temporal phenomenon.

Research on the automated analysis of music – also known as *machine listening* – is a complicated endeavor that is undertaken for various reasons, including comprehension of music, automatic transcription of music into scores, music therapy, teaching of music, identifying music, identifying instruments, and within interactive performance systems. [25] [26] [27] [28] The characteristics of live music that can currently be identified through computerized techniques include the pitch, duration and timbre of individual notes, and the key, rhythm and melody of a series of notes. [29] This discussion will provide a brief overview of the process of computerized music analysis and its use within a performance system for synchronizing video with live music. Specifically, I will discuss the use of machine listening within the Orpheus software and its use of a flexible music-listener.

ISSN 1071-4391 ISBN 978-1-906897-22-2

From a practical perspective, automated music analysis can evaluate the actual sound produced by musical instruments, or it can analyze the MIDI data corresponding with the music. The use of MIDI within music analysis is convenient because the pitch and loudness of musical notes are conveyed as discrete units of digital data, though MIDI messages cannot convey the actual timbre or volume envelope of a note. The MIDI note number 60, for example, indicates that the pitch of a note is C4, but there is no information specifying its actual harmonics. More advanced techniques of music analysis use fast Fourier transform analysis (FFT), a computerized algorithm that can identify the harmonics that make up a note's timbre. [30] Machine listening using any computerized technique is not yet a perfected technology; Robert Rowe states in his book *Machine Musicianship,* "We must labor mightily to make a computer program perform the analysis required of a freshman music student." [31]

A detailed discussion on the current state of automated music analysis is beyond the scope of this article and my ability, therefore I will primarily describe how I have used machine listening when synchronizing video with live music. The music analysis used within my interactive performances and installations until 1998 was accomplished solely through MIDI. When working with musical instruments that did not provide MIDI output – such as a trombone or saxophone – an instrument's sound was converted into MIDI messages using a hardware pitch follower. [32] Since 1998 I have also used the analysis of audio signals within various projects that employed machine listening. My primary interest in machine listening is to develop practical methods for identifying changes in musical structures that are perceivable and relevant for controlling video.

The music-listener diagrammed in figure 3 is the component within the system that executes a specific strategy for analyzing music. It provides an ongoing stream of features pertaining to the incoming music that are used by the control mechanism for directing the video output. The music-listener provides two categories of features, instantaneous and temporal. *Instantaneous features* are characteristics of the music at a specific moment in time, while *temporal features* are those that span a longer duration of time or calculated from combinations of features and parameters. Creating a music-listener that can recognize a large variety of musical features is an important strategy for enhancing the system's degree of interactivity.

The smallest musical event recognized by the music-listener is the *note*. The features within a note include its pitch, loudness, and duration. The pitch of a note is an example of an instantaneous feature, such as an A note in the fourth octave (A4). The term *pitch* in the music-listener refers to the fundamental frequency of the sound produced by the note, such as A4 being 440Hz. A temporal feature of a note is its *pitch interval,* the difference in pitch between this note and the one previously played. The calculation of a pitch interval within the music-listener is determined according to the chromatic scale that is comprised of 12 pitches per octave: C, C#, D, D#, E, F, F#, G, G#, A, A#, and B. The pitch interval between adjacent notes on this scale is 1, also known as a semitone. For example, if the notes played sequentially were C4, C4, C#4, D#4, E4 and C4, the series of pitch intervals would be as in the following example.

C4 to C4 pitch interval = 0
C4 to C#4 pitch interval = 1
C#4 to D#4 pitch interval = 2
D#4 to E4 pitch interval = 1
E4 to C4 pitch interval = -4

The pitch of a specific note can be determined independently of other notes or time, but a pitch interval can only be determined from the occurrence of two sequential notes. This type of pitch interval is also called a horizontal or melodic interval. [33] In contrast, a vertical pitch interval is the pitch difference between two or more notes playing simultaneously – known as a chord. A vertical pitch interval is classified as an *instantaneous polyphonic feature* because it describes the relationship between two or more notes playing at a particular instant. An *instantaneous monophonic feature* refers to the characteristic of single note playing at particular moment. Eventually Orpheus was able to recognize these four categories of music features: instantaneous monophonic features, temporal monophonic features, instantaneous polyphonic features, and temporal polyphonic features. Listed below are some of the features that were recognized by the music-listener used by Orpheus beginning in 1990. All of these features were extracted from music that was represented as MIDI messages.

Instantaneous monophonic features used for a series of single notes:

- Active note pitch: note currently playing
- Active note loudness: loudness of note currently playing
- Active note octave: octave of note currently playing
- Active rest: no note playing

Temporal monophonic features used for a series of single notes:

- Designation of active note for use in higher level analysis
- Active note start time relative to beginning of music
- Active note interval, relative to pitch of previous active note
- Active note loudness interval, relative to previous active note
- Active note duration relative to its start time
- End time of previously active note relative to beginning of music
- Number of occurrences of individual pitches
- Tempo: rate of occurrence of active notes
- Rest start time, relative to beginning of music
- Rest duration, relative to its start time

Instantaneous polyphonic features used for chords:

- Total number of notes playing at one time
- Pitch of each note currently playing
- Loudness of each note currently playing
- Octave of each note currently playing

Temporal Polyphonic Features used for chords:

- Start times for all notes playing relative to beginning of music
- Durations of all notes currently playing
- End times of notes previously played relative to beginning of music
- Designation of active note: most recently played note or with lowest pitch
- Tempo: rate of occurrence of active notes
- Pitch interval of active note relative to previous active note
- Loudness interval of active note relative to previous active note
- Designation of active note to be used in higher level analysis
- Number of occurrences of all pitches played

Figure 4 provides an example of the music-listener within Orpheus using a reduced collection of features. The top section of the example indicates a sequence of notes represented as black rectangles, with pitch indicated vertically, duration horizontally, and a note's loudness as the white number within a rectangle. MIDI note numbers are indicated to the right of the note names on the music keyboard. A list of instantaneous and temporal features is listed beneath the notes, corresponding with time units 1 to 19. Each time unit represents 0.425 second, but in actuality the sampling time used by Orpheus was 0.0333 second. The center

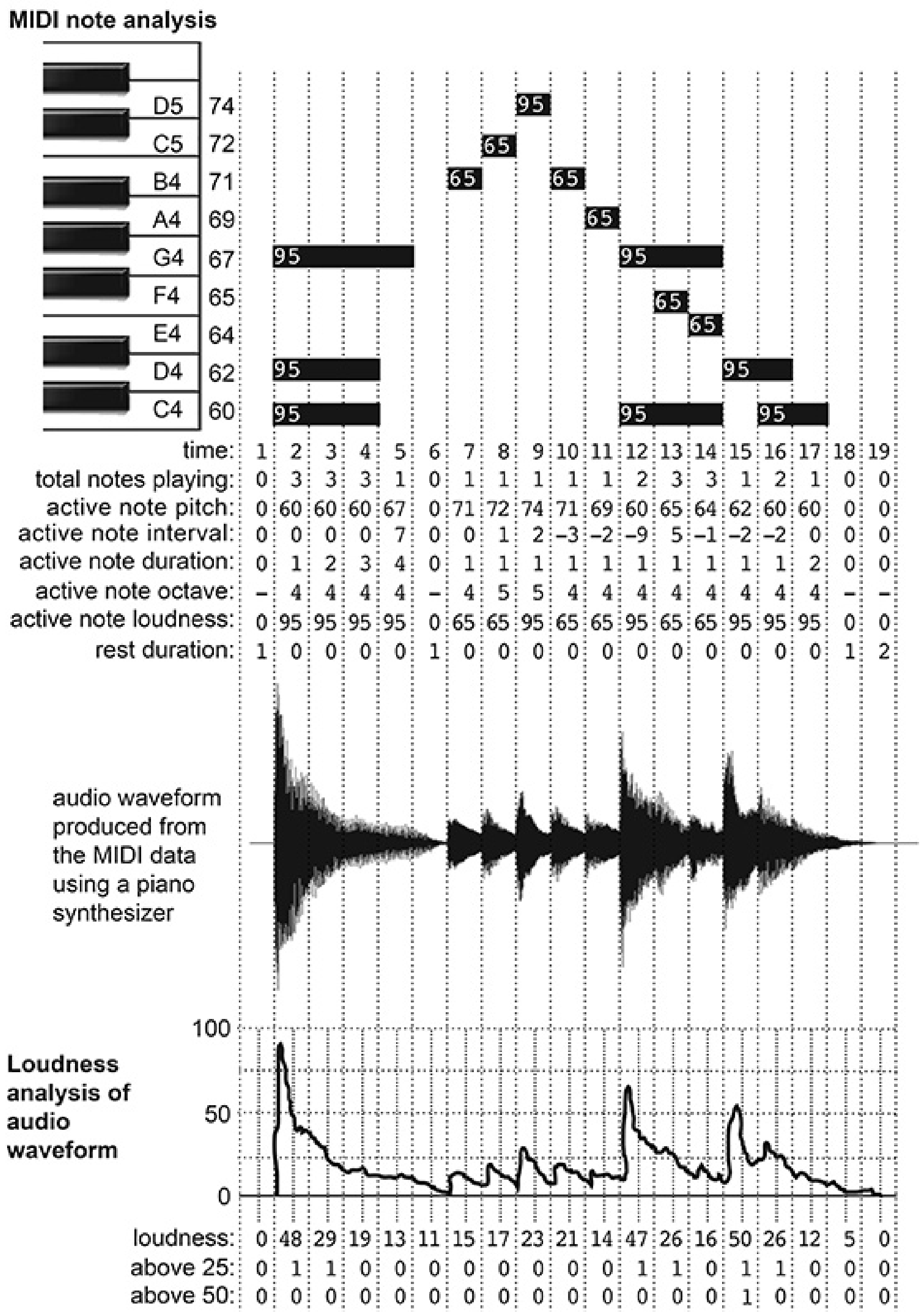

Fig 4. *Music-listener analysis using MIDI and audio wave.* An example of analyzing the MIDI data output by a piano synthesizer that provides a continuous stream of instantaneous and temporal features. An analysis of the instrument's audio wave provides additional details of the loudness envelope. © Don Ritter, 2013. Used with permission.

of the diagram is a representation of an audio wave that was created by sending the MIDI messages into a piano synthesizer. The lower section of the example depicts an analysis of the loudness envelope associated with this audio wave.

At time *1* within the example, no music is playing and the music-listener indicates a musical rest of duration 1. At time *2,* the *active note* is MIDI note 60 (C4). When multiple notes are played simultaneously, the note with the lowest pitch is designated as the active note. If a new note begins while a previous note is still playing, the new note is designated as the active note. For example, two notes are playing at time *12,* but MIDI note 60 (C4) is the active note because it has the lowest pitch. At time *13,* MIDI note 65 (F4) is the active note because it occurred more recently than MIDI note 60 (C4). The note designated as being active is used within the next level of analysis, when the music is categorized into different *music-types.*

The bottom of the example depicts a simple audio analysis of the loudness envelope that was created by sending the MIDI messages into a piano synthesizer. The result of the loudness analysis is a stream of numbers – ranging from 0 to 50 – that correspond with the loudness of the music at the 19 sample times. Orpheus never used this method because the software's only form of input was MIDI, but I have used this form of analysis since 1998 within installations and performances that did not use Orpheus.

RELATION SHIPS BETWEEN MUSIC FEATURES, MUSIC CATEGORIZATION, AND VIDEO ACTIONS

The music-listener contains 3 levels of analysis:

1. acquisition of instantaneous features, such as the pitch of the active note
2. calculation of temporal features, such as the duration of the active note
3. categorization of the music into a music-type

The categorization of music into a *music-type* is made according to different groupings of instantaneous and temporal features. For example, incoming music could be categorized according to pitch ranges, loudness ranges, or a combination of pitch and loudness ranges. Below are two simple examples of categorization strategies that can be used within the music-listener. In these examples, pitches are designated using MIDI note numbers. In actual practice, categorization strategies were more complex than what is listed below, and up to 999 different categorization rules could be created within a strategy.

Example of a music categorization rule according to loudness:

» the loudness of an active note corresponds with the music-type; for example, the music-type is 96 when the loudness of the active note is 96.

Example of music categorization rules according to pitch ranges:

» active note below 36 is music-type 1,
» active note between 37 and 42 is music-type 2,
» active note between 43 and 46 is music-type 3,
» active note between 47 and 52 is music-type 4,
» active note between 59 and 127 is music-type 5.

Example of music categorization rules according to note duration:

- » active note duration under .25 second is music-type 1,
- » active note duration between .25 and .5 second is music-type 2,
- » active note duration between .5 and 1.0 second is music-type 3,
- » active note duration between above 1.0 second is the music-type 4.

When music is analyzed by the music-listener, it provides a continuous stream of instantaneous features, temporal features, and music-types. The specific manner for categorizing music into music-types is completely adjustable: any combination of instantaneous features and temporal features can be used within a strategy.

Figure 5 provides an example that depicts the relationships between music features, music categorization, and video actions. In this example, a music-listener strategy has been selected for categorizing music into three types: music-type O when no music is playing; music-type 1 when the pitch of the incoming note is less than MIDI note 60 (C4); and music-type 2 when the incoming note is greater than MIDI note 60. I am using an overly simple strategy for categorizing the music to make the flexible structure of the system more apparent. Thousands of different strategies for categorizing music can potentially be created using various combinations of musical features, and a single strategy can create a large number of music-types. Because a video action is not video media – it is only a manner for playing video media – the amount of memory occupied by a video action is small and thousands of different video actions can be used without reducing the system's performance.

Returning to the example in figure 5, the total number of music categories created by this music-listener is 3. The number of video action lists that can be controlled by a music-listener corresponds with the number of music categories it can identify. Therefore, this listener can control 3 different action lists. Each of these lists contains 3 video actions, though a video action list can potentially hold thousands of video

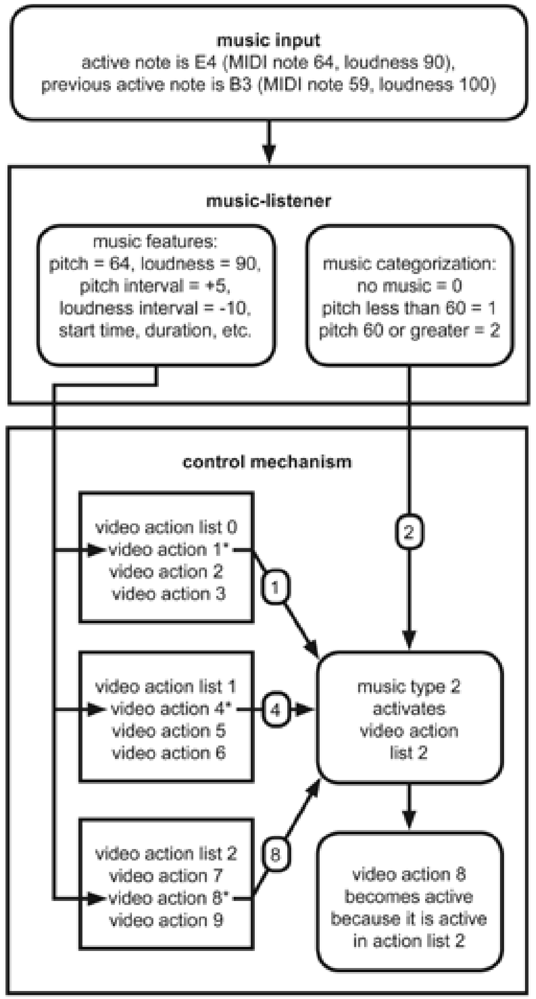

Fig 5. *Example of a video action activated by a music-type.* A specific video action is activated according to musical features and music categorization. Don Ritter, 2013. Used with permission.

actions. A video action list becomes active when the active music-type corresponds with that list, such as video action list 2 being activated when the incoming music is categorized as type 2. A specific video action within a list is available for activation according to the number of times the list has been activated. The video actions marked with an asterisk are available for activation. In the example, video action 8 is designated as being active within action list 2, so the video output of the system is video action 8 when the music-type is 2. The categorization of music into different music-types is simply a method of using a single number to represent the overall musical activity at a certain moment in time, similar to a stock market index representing the overall activity of a stock market at a certain instant in time. The more intricate synchronization of the system is provided through the various instantaneous and temporal features that control the various parameters within the video actions.

CONCLUSION

The interactive system described in this document evolved over thousands of hours of software development, testing, and live performances. I typically collaborate with an improvising musician for the creation of a performance of synchronized video and music, and I am fortunate to have worked with some outstanding musicians. The process of composing a performance typically begins by giving priority to either the music or the imagery, or by jointly developing a unified concept with the musician. If the video imagery is given priority, the musician selects a specific musical instrument, compositional structure, or a collection of sounds that relate to the predetermined video material. If the music is given priority, I create new video material that relates to the predetermined music. Creating a work that uses new video material and new music according to a mutually selected concept is usually the most fulfilling approach, but very laborious. The next step is to create numerous correlations between the specific musical and visual elements and then incorporate them into the system's control mechanism.

The primary advantage of a flexible system for synchronizing video with live music is the compositional freedom it provides. After working with various musicians, I found that different types of correspondence were more effective according to the instrument and musical style employed by a musician. For example, synchronization according to subtle changes in loudness and pitch were most effective with saxophone, but this approach was less interesting when used with percussive instruments. A flexible approach for synchronizing video with live music enables the creation of a composition that uses counterpoint relationships between imagery and music rather than a strict trigger and response approach. ■

ACKNOWLEDGEMENTS

I would like to thank the musical collaborators who were involved in performances using Orpheus since 1988, including George Lewis, Thomas Dimuzio, Trevor Tureski, Amy Denio, Nick Didkovsky, Ben Neill, Geneviève Letarte and Tom Walsh, and also Robert Rowe who has answered my questions on machine listening for many years. Portions of the work described in this paper received support from the Canada Council during 1989-1991.

REFERENCES AND NOTES

1. David Eagen, "VJ Software Listings," Audio Visualizers, http://www.audiovisualizers.com/toolshak/vjprgpix/softmain.htm (accessed July 18, 2012).
2. A. B. Klein, *Color-Music: The Art of Light* (London: Crosby Lockwood and Son, 1927).
3. Jack Burnham, *Beyond Modern Sculpture* (New York: George Braziller, 1968), 286–288.
4. Kenneth Peacock, "Instruments to Perform Color-Music: Two Centuries of Technological Experimentation," in *Leonardo* 21, no. 4 (1988): 397–406.
5. D-Fuse, *VJ: Audio-Visual Art and VJ Culture (*London: Laurence King Publishing, 2006).
6. Kenneth Peacock, 398–400.
7. Louis-Bertrand Castel, *L'Optique des couleurs* (Paris, 1740). Online version of this illustration at http://en.wikipedia.org/wiki/Louis_Bertrand_Castel (accessed November 25, 2012)
8. Karl von Eckartshausen, "Disclosures of Magic from Tested Experiences of Occult Philosophic Sciences and Veiled Secrets of Nature," (1791) quoted in Sergei Eisenstein, *The Film Sense,* ed. and trans. Jay Leyda (New York: Harcourt, Brace & World, 1942), 88.
9. MIDI is an acronym for Music Instrument Digital Interface, a digital language for describing music in real time. MIDI messages specify musical characteristics – including the pitch, loudness and timing of musical notes – but contain no actual audio.
10. George Lewis, "Too Many Notes: Computers, Complexity and Culture in Voyager," in *Leonardo Music Journal* 10 (2000): 33–39.
11. A. B. Klein, *Color-Music: The Art of Light* (London: Crosby Lockwood and Son, 1926), 21.
12. B. M. Galeyev, "The Fire of Prometheus Music-Kinetic Art Experiments in the USSR," in *Leonardo* 21, no. 4 (1988): 383–396.
13. Counterpoint in music composition refers to independent rhythms being combined to create a unified structure.
14. Youtube, "Scriabin's Prometheus: Poem of Fire," Yale University, February, 2010, http://www.youtube.com/watch?v=V3B7uQ5KoIU (accessed October 15, 2012).
15. Sergei Eisenstein, *The Film Sense,* ed. and trans. Jay Leyda (New York: Harcourt, Brace & World, 1942)
16. Sergei Eisenstein, *The Film Form,* ed. and trans. Jay Leyda (New York: Harcourt, Brace & World, 1949).
17. Ibid., 78.
18. Milller Puckette, "Something Digital," in *Computer Music Journal* 15, no. 4 (1991): 65.
19. Kenneth Peacock, 400.
20. B. M. Galeyev and I. L. Vanechkina, "Was Scriabin a Synesthete?," in *Leonardo* 34, no. 4 (2001): 357–362.
21. Kenneth Peacock, 398–400.
22. Sergei Eisenstein, *The Film Sense*, 109.
23. Sergei Eisenstein, *The Film Sense*, 85.
24. Memory Prices (1957–2012), http://www.jcmit.com/memoryprice.htm (accessed October 15, 2012).
25. Robert Rowe, Skype interview by author, October 19, 2012.
26. Robert Rowe, *Interactive Music Systems: Machine Listening and Composing* (Cambridge, MA: The MIT Press, 1993).
27. Nick Collins, "Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems," (PhD thesis, University of Cambridge, 2006).
28. Brian Whitman, Gary Flake, and Steve Lawrence, "Artist Detection in Music with Minnowmatch," *Proceedings of the 2001 IEEE Workshop on Neural Networks for Signal Processing*, Falmouth, Massachusetts, (September 2001): 559-568.
29. David Gerhard, "Computer Music Analysis," (Technical Report CMPT TR 9713, Simon Fraser University, 2002).
30. Wikipedia, "Fast Fourier transform," http://en.wikipedia.org/wiki/Fast_Fourier_transform (accessed October 22, 2012).
31. Robert Rowe, *Machine Musicianship* (Cambridge, MA: The MIT Press, 2001), 8.
32. Synthony Music, "IVL," Synth and Midi Museum, http://www.synthony.com/vintage/pitchrider.html (accessed November 20, 2012).
33. Wikipedia, "Interval (music)," http://en.wikipedia.org/wiki/Interval_music (accessed October 28, 2012).

 ISSN 1071-4391 ISBN 978-1-906897-22-2