2014

# Audio-Based Visualization of Expressive Body Movements in Music Performance: An Evaluation of Methodology in Three Electroacoustic Compositions

Yeamin Oh

*Louisiana State University and Agricultural and Mechanical College*

AUDIO-BASED VISUALIZATION OF
EXPRESSIVE BODY MOVEMENTS
IN MUSIC PERFORMANCE:
AN EVALUATION OF METHODOLOGY
IN THREE ELECTROACOUSTIC COMPOSITIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Music

by
Yeamin Oh
B.M., Kyung Hee University, 2002
G.P.D., University of Hartford, 2006
M.M., Georgia Southern University, 2010
August 2014

*To my dear parents*

*Chaing-il Oh and Young-nam Han*

*And*

*To my dearest wife*

*Yujin Im*

# Acknowledgments

I would like to thank the School of Music and Center for Computation and Technology at Louisiana State University. I would also like to thank Dr. Jesse Allison for guiding me and sharing his wealth of experience and knowledge. Special thanks go to Dr. Stephen Beck for his encouragement and compositional guidance. Very special thanks go out to Dr. Edgar Berdahl for his excellent feedback, and to Dr. Brett Boutwell for sharing his expertise on the history of music. I would like to acknowledge Dr. James Byo’s guidance of the survey. Thanks also to fellow student and good friend Benjamin Taylor for his help editing.

## Table of Contents

List of Tables

Lists of Figures

# Abstract

An increase in collaboration amongst visual artists, performance artists, musicians, and programmers has given rise to the exploration of multimedia performance arts. A methodology for audio-based visualization has been created that integrates the information of sound with the visualization of physical expressions, with the goal of magnifying the expressiveness of the performance. The emphasis is placed on exalting the music by using the audio to affect and enhance the video processing, while the video does not affect the audio at all. In this sense the music is considered to be autonomous of the video. The audio-based visualization can provide the audience with a deeper appreciation of the music. Unique implementations of the methodology have been created for three compositions. A qualitative analysis of each implementation is employed to evaluate both the technological and aesthetic merits for each composition.

# Chapter 1: Rationale for the Research

## 1.1 Origin of Multimedia Art and its Definition

Artists and musicians have the passion and desire to create innovative works of art. This can be seen throughout music history. Musicians constantly evolved the meaning of art and cultivated new forms of art. Not only did they try to harmonize different categories of art, but also attempted to combine technology and art into a new form in the 20th century. This attempt can be found in the work of many contemporary artists and composers. One might wonder where the origin of multimedia art can be found in history. In this chapter, the people who influenced the others are briefly discussed.

### 1.1.1 Origin of Multimedia Art

The initial idea to merge different art forms into one total art form can be found in the work of the 19th century. The composer Richard Wagner discussed the term “Gesamtkunstwerk”, which translates to “total work of art" or "ideal work of art,” to suggest a direction for a new artistic endeavor which he realized through his work. He was confident that new poetic heights could be accomplished by the separate branches of art when put to the service of drama, which he regarded as the ideal medium for achieving his vision.[1] While Wagner’s hyper-romantic rhetoric may seem out of date, his acumen and foresight of the synthesis of artistic practices can be appreciated for illuminating contemporary notions of multimedia.

After Wagner’s idea, artists, composers and scholars handed down the idea and expanded it even further. At the beginning of 20th century, Filippo T. Marinetti and his colleagues, called

---

[1] Richard Wagner, “ ‘Outlines of the Artwork of the Future,’ The Artwork of the Future.” *Multimedia: From Wagner to Virtual Reality,* Edited by Randal Packer and Ken Jordan, (New York, NY: W.W. Norton & Company Ltd., 2001), 3-9.

Futurists, foresaw the need to integrate different aesthetics in art, as well as the potential of new art forms such as cinema. They believed that the moving picture was a rising technology that could actualize their artistic endeavors. Marinetti indicated that artists must liberate film as an expressive medium in order to make it the ideal instrument of a new artistic practice, immensely vaster and nimbler than all the existing arts.[2] However, the Futurists' prediction about the moving picture as a new art form was not realized until sufficient technologies were available.

After the Futurist movement in 1910s, László Moholy-Nagy, painter and photographer, introduced his idea of the "Theater of Totality," which is a different interpretation of Wagner's concept of Gesumtkunstwerk. While Moholy-Nagy's concept would reduce the importance of the written word and the presence of the actor, his idea of the "Mechanized Eccentric" introduced the qualities of machinery into every aspect of stage performance. This idea resulted in a technologically innovative theater that emphasized the physical rather than the literary, and that reflected the speed, dynamism and precision of state-of-the-art technology.[3]

Marcel Duchamp in visual art and John Cage in music influenced many other artists and composers with their conceptual approaches to create new artistic practices. They emphasized using a variety of media that had never been employed in artistic practice. For example, John Cage's *Imaginary Landscape*, a series of compositions, was composed for a variety of mediums, including turntables, audio frequency oscillators, and radios. In Duchamp's famous sculptures,

---

[2] Fillippo T. Marinetti et al., "The Futurist Cinema." *An Futurism: Anthology*, Edited by Lawrence Rainey, Christine Poggi and Wittman Laura, (New Haven: Yale University Press, 2009), 230.

[3] László Moholy-Nagy, " 'Theater, Circus, Variety,' Theater of the Bauhaus" *Multimedia: Wagner to the Virtual Reality*, Edited by Randall Packer and Ken Jordan, (New York, NY: W.W. Norton & Company Ltd., 2001), 16-26.

“Fountain” and “Bicycle Wheel”, one can see the possibility of a daily-life medium, called *ready-made*, as an art form. This influence of new media hardened the ground for building the field of modern interdisciplinary art.

1.1.2 Definition of Multimedia Art

In the time after World War II when many technologies were introduced to the public, the interest of artists and composers to create interdisciplinary works between technology and art increased rapidly, and were actualized into real artistic practices. In music, composers and engineers tried to incorporate technology and music, giving rise to new forms of musical art such as *musique concrète*, acousmatic music, e*lektronische musik*, and tape music. After the first commercial video camera, the Sony Portapak in 1965, video art gradually became a new form of visual art also.[4]

In 1965, the term multimedia was first coined and introduced by artist and singer Bobb Goldsteinn for describing his show, “LightWorks at L`Oursin,” which integrated music, lights and film among other media.[5] After his introduction of the term, many other scholars from several interdisciplinary areas defined the term “multimedia” in their own ways. For example, two years after Bobb Goldsteinn, artist Dick Higgins introduced the term “intermedia” to describe a myriad of emerging genres that spilled across the boundaries of traditional media.[6] In

---

[4] Holly Rogers, *Sounding the Gallery: Video and the Rise of Art-Music.* (New York, NY: Oxford University Press., 2013), Chapter 1, Kindle.

[5] Richard Albarino, “Goldstin’s LightWorks at Southhampton,” *Variety*, 213, no.12 (August, 1966).

[6] Dick Higgins, “Statement on Intermedia” (August 3, 1966), reprinted in *In the Spirit of Fluxus,* ed. Joan Rothfuss and Elizabeth Armstrong, (Minneapolis: Walker Art Center, 1993), 172– 73.

the 1990s, scholar of journalism and educator Steven Maras discussed ten different definitions of the term, including an artwork using multiple media, a hybrid definition, and a megamedium.[7]

In psychology, Richard E. Meyer proposed a three-part view to define the attributes of multimedia: the artwork needs two or more delivery devices, the artwork needs to have two or more modes of presentation; and it needs to require two or more sensory modalities.[8] These categories of the term multimedia by Meyer seem reasonable in terms of a scientific approach. However, in Meyer's three-part view, the third definition on sensory modalities is not clear, and can be debated.

Among the many other definitions, graphic designer and educator Brian Slawson described multimedia as "the use of a personal computer to compose, display and manipulate a variety of electronic media simultaneously, combining elements of text and speech, music and sound, still-images, motion video and animated graphics."[9] His interpretation of the term is appropriate not only for this paper but also for the general meaning of the current era. Additionally, since most artists and musicians use a computer to realize their multimedia art currently, the use of the word "computer" in his definition is appropriate to this paper. If the meanings of multimedia in art and computer science are combined, a similar definition can be

---

[7] Steven Maras, "Convergence/divergence: Ten Definition of Multimedia." *Metro Magazine: Media& Education Magazine* no.108 (1996): 9-13.

[8] Richard E. Meyer, *Multimedia Learning*. 2nd ed. (New York, NY: Cambridge University Press. 2009), 8-10.

[9] Brian Slawson, "Interactive Multimedia: The Gestalt of a Gigabyte." *Art Education*, 46, no. 6 (Nov. 1993): 15-22.

concluded; in art, multimedia means “using more than one medium, expression or communication,” and in computer science, “the incorporation of audio and video.”

Several 20th-century artists and composers foresaw the potential of integrated arts and technology as a new art form, and they were convinced that the new form of art might be the future of artwork. With our available technology, including computers and multimedia, we can achieve their vision. It might be our generation’s mission to manifest this future of arts as our contribution to art history. Several people differently defined multimedia for their own purposes, and by integrating their definitions, we can develop one suitable for this paper: multimedia is the use of computation to compose, display and manipulate more than one electronic medium simultaneously, while requiring two or more sensory modalities such as auditory and visual.

### 1.2 Congruency between Visuals and Music

Interest and attention to multimedia, especially for relationships between visual and sonic stimuli, can be found in many psychological studies. For researchers outside the field of psychology, particularly those who are in the fields of music and visual arts, a connection with the scientific approach can be a challenge from both conceptual and practical standpoints.[10] However, since researchers in psychology gather data from humans or animals under controlled conditions for the purpose of addressing specific questions with an empirical experiment, artists and composers, especially those who deal with multimedia, can acquire the psychological rationale behind their arts and music.

[10] Annabel Cohen, Scott Lipscomb, Siu-lan Tan, and Roger Kendall, “Introduction: The Psychology of Music in Multimedia,” *The Psychology of Music in Multimedia.* ed. SIu-Lan Tan, Annabell Cohen, Scott Lipscomb, and Roger Kendall (Oxford, UK: Oxford University Press, 2013), Chapter 1, Kindle.

Through this empirical approach, psychologists investigated the relationship between visual and music, and proved their assumption about human's perception on multimedia. Before multimedia, they already had researched music and human auditory cognition for a century, which is almost the same length of time as multimedia's history. They have researched not only psychological phenomenon on listening to music, but also human's perception of music based on music theory.

As an extension of this research on music, since the 1980s several psychologists have specifically focused on the Congruence-Association Model (CAM) of multimedia. The history of psychology on CAM emitted not only the research and theoretical outcomes but also an example of the process of discovery, which is mostly based on an empirical approach.[11] Since the goal of this study is related to psychological views for the reasoning and purpose of composing with multimedia, the basis can be found in the empirical approaches and experiments of psychology on cross-modal perception in multimedia.

Bolivar, Cohen, Fentress, and Iwamiya investigated the congruency between music and visuals with many different experiments, and they categorized the perceived congruency in two aspects.[12] One is formal congruency: the matching of auditory and visual temporal structures, which produces a unified perceptual form to auditory and visual information. The other is semantic congruency: the similarity between auditory and visual affective impressions. This

---

[11] Annabel J. Cohen, "Congruence-Association Model of Music and Multimedia: Origin and evolution," *The Psychology of Music in Multimedia,* ed. SIu-Lan Tan, Annabell Cohen, Scott Lipscomb, Roger Kendall (Oxford, UK: Oxford University Press, 2013), Chapter 2, Kindle.

[12] Shin-ichiro Iwamiya, "Perceived Congruence Between Auditory and Visual Elements in Multimedia," *The Psychology of Music in Multimedia,* ed. SIu-Lan Tan, Annabell Cohen, Scott Lipscomb, Roger Kendall (Oxford, UK: Oxford University Press, 2013), Chapter 7, Kindle.

similarity allows the listener to tie the sound to the visual event, but allows more freedom for artistic parings of visual and sonic event. Both types of congruency have important roles in the cross-modal interaction between auditory and visual domains. While formal congruency produces a united perceptual form between auditory and visual information, semantic congruency helps to convey the meaning of audiovisual content to the perceiver.

### 1.2.1 Formal Congruency

Formal congruency is the matching of auditory and visual temporal structures, a phenomenon which has been well documented. Iwamiya, Sugano, and Kouda experimented on perceived congruence between sound and moving pictures to discover the effect of audiovisual synchronization. The experiment used a visual stimulus of a ball on a grid surface, and a change in visual perspective of the ball coincided with auditory events. Auditory events and stimuli were based on a simple rhythmic melody played on a bass and drums.[13] As a result of the experiment, when visual events synchronized with auditory events, formal congruency was perceived, and when the study used higher synchronization between events, the perceived congruence was higher. This result showed the effectiveness of formal congruency in raising perceived congruence, because the participants' rating of their impression of synchronization was also higher when the perceived congruence was higher.[14]

---

[13] Shin-ichiro Iwamiya, Y. Sugano,and K. Kouda, "The Effects of Synchronization of Temporal Structures of Sound and Motion Picture on the Impression of Audio-visual Contents." *In Proceedings of the 2000 IEEE International Conference on Systems, Man, and Cybernetics* (2000): 1222-225.

[14] Iwamiya, "Perceived Congruence,“ Chapter 7.

Another example can be seen in a second experiment by Sugano and Iwamiya where they examined the effect of synchronization between auditory and visual accents.[15] In this study, in addition to synchronized and asynchronized conditions, a temporal phase-shift condition between video and music was incorporated. For example, a constant time difference between auditory and visual accents occurred during the time interval between events. Their result demonstrated that the relatively higher perceptual congruence of the phase-shift condition indicated the significance of correlated periods between periodic auditory and visual events. Sugano and Iwamiya did a third similar experiment on formal congruency. In this experiment, by adjusting the speed of a visual object's movement, participants were asked to match the music presented at a constant tempo. As a result, perceived congruence was created when the periods of auditory and visual events corresponded at integer ratios, such as 1 to 1, 2 to 1, 3 to 1, or 4 to 1.

As a consequence, it was determined that higher synchronization between auditory and visual events follows a direct causation of subjects' rating an increase in formal congruency between music and video. In addition to one to one audio-visual synchronization, integer ratios of periodic patterns also cause direct congruence.

### 1.2.2 Semantic Congruency

While formal congruency is the perceived correlation of auditory and visual events, semantic congruency is the similarity between auditory and visual affects and impressions on humans. Iwamiya, Jogetsu, Sugano, and Takada examined semantic congruency using computer-generated moving objects and short pieces of music. In this experiment, participants rated the semantic differences of scenes on a computer monitor depending on congruent affect due to

[15] Iwamiya, "Perceived Congruence," Chapter 7.

moving objects and music. Visual stimulus was performed under two conditions: the speed of the objects and the density of the objects. Auditory stimulus was also performed under two conditions: tempo of the music and mode of the music (major or minor). Results of this experiment revealed two types of high semantic congruency. First, high densities and fast speeds of images corresponded with major-key and fast-tempo music. Second, low densities and slow speeds of images correlated with slow-tempo and minor-key music.[16]

Iwamiya and Hayashi's experiment using piano music and computer graphics of a piano performance under various colors of lighting also reveals the effect of semantic congruency. They used eight short musical passages from popular piano tunes combined with lights colored red, yellow, yellow-green, green, cyan, blue, violet, and purple. In this experiment, they examined that blue, violet, purple, and red were highly matched with minor-key and fast tempo music, while green, yellow-green, yellow and cyan were matched with bright, major-key and fast tempo music. In addition, the results of this experiment on affective impression showed that yellow, green, and cyan colors made the music seem brighter.[17]

Several empirical studies on formal and semantic congruency between sound and moving pictures have been discussed, providing various suggestions for producers, sound designers, and composers of visual media productions to assist in creating effective audiovisual works. However, in real applications of these suggestions to multimedia art, there are a lot more issues to consider. Furthermore, because the aforementioned psychologists' results are based on simple tasks and unified situations for drawing scientific conclusions, it could be argued that those

[16] Iwamiya, "Perceived Congruence," Chapter 7.

[17] Ibid*.,* Chapter 7.

studies cannot be directly applied to the actual creation of multimedia arts, especially that which is avant-garde. Nevertheless, if the results of the congruencies are intertwined, these studies can be qualitatively interpreted within the context of creating music and art correlating and non-correlating aspects. Details of this direct connection will be discussed more in Chapter 2.

### 1.3 Definition of Expressive Body Movement

In most cases, musical performers use their bodies to interact with their musical instruments when they perform. The role of body movement in musical performance conveys the musical intention of the composer not only through the sonic properties of the music performance, but also through the visual properties of the performance. In other words, performative body movement has two functions: first, it inevitably creates inflections in the sonic properties of resulting tones, and second, its visual properties assist to elevate the audience's appreciation of expressiveness of the music. Several scholars define and categorize performative movements from different perspectives. Among the definitions and categorizations, some were adapted to this study, and re-classified according to two distinct aspects of performance movement: sonic properties and visual properties.

#### 1.3.1 Sonic Property of Body Movement

The sonic properties of performance movement are described using several terminologies. Delalande's categorization of body movements designates effective gesture as those which affect the sonic property, which means the movement actually produces the sound.[18] Cadoz uses another term—instrumental gestures—to define such body movements. Instrumental

[18] F. Delalande, "Human Movement and the Interpretation of Music." (presented at *the Second International Colloquium on the Psychology of Music*, Ravello, Italy, 1990).

gestures are regarded to be a complementary mode of communication to empty-handed gestures,[19] and are characterized as a gesture that is applied to a concrete object with a physical interaction and its phenomena during the interaction. These two terms have different perspectives, but a similar meaning in terms of the sonic significance of the gesture.

### 1.3.2 Visual Property of Body Movement

On the other hand, as a visual property, musicians also move their bodies in a way that is not directly related to producing musical tones. Body movements that assist to increase the expressiveness of music are often called expressive movements, ancillary gestures, expressive body movement or accompanist gestures. Each term has a slightly different meaning. Ancillary gestures in electronic music include the actions that are not used to produce sound, such as walking, posture adjustments, facial expressions, dance movements, muscular tension, and even motionless signals such as brainwaves.[20] In Delalande's definition, accompanist gestures are a body movement such as shoulder and head movements, which support expressiveness but can also be related to producing sound. Among the several definitions, the musicologist, Jane Davidson’s terminology is suitable to this study in terms of its visual property. Davidson has conducted significant research projects on movement that does not produce actual sound, and has used several different terms to designate the movements: expressive movements, expressive body movements, and expressive bodily movements. Among the terms, expressive body movement is used throughout this paper.

---

[19] Claude Cadoz, “Instrumental Gesture and Music Composition.” In *proceeding of the 1988 International Computer Music Conference.* (San Francisco, CA: ICMA, 1988): 1-12.

[20] Eduardo Miranda, and M. Wanderley, *New Digital Musical Instrument: Control and Interaction beyond the keyboard.* (Middleton, WI: A-R Editions, Inc., 2006), 10.

1.4 Prior Studies on Expressive Body Movement in Music Performance

Visual aspects of performance can distract from the appreciation of music, and sometimes closing one's eyes can increase the enjoyment of music. However, there are many people prefer to go to a live-concert to appreciate music instead of listening to a high-quality of recording of the same performer. They want not only to hear the sound of music but also to see the expressiveness of the music performance. The former might have a similar philosophic idea of acousmatic music in electroacoustic music. In Scruton's perspective, visual aspects of the performance of traditional music can be a hindrance to the reception of its sonic expression.[21] In what he calls "acousmatic experience," music ideally exists by itself to an audience in terms of how its sounds are made, just as Pythagoras' students listened to their teacher speaking from behind a veil.

In Scruton's opinion, the expressiveness of the performer's movements is an obstacle to appreciate the music, and the audience can receive the expression of the music without seeing the expressive body movements. Deniz Peters discusses the importance of a performer's expression in electroacoustic music. Peters starts by mentioning the way a listener can hear the musician's body in his or her music. Peters describes three aspects: the presence of the performer's body to the listener in the act of listening, the body's appearance of composer in music, and performative listening. He argues there is a residue of bodily presence in the sounds people hear, both in the activity of sound and on the side of the listening experience. Peters concluded his thesis:

> Despite the often noted 'disruption' of bodily sound-making in music using electronic media, and despite listening attitude and theories that construe listening as disembodied,

[21] Roger Scruton, *Aesthetic of Music,* (New York: Oxford University Press, 1997), 2-3.

> there is much more bodily presence in what people hear when people hear music than hitherto acknowledged, and that, in terms of aesthetic perception, it may even be omnipresent.[22]

In Peters' writing, he aims to convince the reader that expressiveness in music is significant enough to merit appreciation, and that an audience can hear the performer in the music itself.

Furthermore, there is much other research to support the importance of the expressiveness of music performance. In musicology, Davidson suggests that vision can be more informative than sound in the perceiver's understanding of a performer's expressive intentions. She proves this through several related experiments in live music performance. In one of her experiments, she separated three modes of observation: vision alone, sound alone, and sound and vision together. The result revealed not only that vision is a useful source of information about expressive action, but also that a combination of vision and sound more clearly communicates the expressive action.[23]

Rodger, Crig, and O'Modhrain examined the significance of visible performance when people rated the level of expertise of performers. In their experiments, musicians and non-musicians rated performances by novice, intermediate and expert clarinetists from a sound recording, a point-light animation of their movement, or both. When they switched the sound recording and visual movement, ratings of novice musicians' music was significantly higher

---

[22] Deniz Peters, "Touch: Real, Apparent, and Absent: On Bodily Expression in Electronic Music." In *Bodily Expression in Electronic Music,* ed. Deniz Peters, Gerhard Eckel, and Andreas Dorscel (New York, NY: Routledge Taylor&Francis Group, 2012), Chapter 1.

[23] Jane W. Davidson, "Visual Perception of Performance Manner in the Movements of Solo Musicians." *Psychology of music* 21 (1993), 112.

when paired with an expert's movements, although the opposite was not found for an expert's sound presented with a novice's movements.[24] This result suggests that musicians and non-musicians perceive the expertise of a performer from a musician's body movements in addition to their sound.

In an article by Friedrich Platz and Reinhard Kopiez, they describe a meta-analysis of several existing experiments that investigate expressiveness in music performance.[25] By analyzing the data from fifteen studies, their article examined the relative contribution of sound, music, and visual signals for effective multimedia performance. Similarly, Bergeron and Lopes examined and compared several studies on expressive body movement. Their discussion of expressive properties in music performance mentions not only sonic properties but also visual properties.[26] Their article ends with the composer Robert Schumann's words, which are taken from his book. The same quote can be found in articles by Davidson. [27] [28]

---

[24] Matthew W.M. Rodger, Cathy M. Craig, and Sile O'Modhrain, "Expertise is Perceived from Both Sound and Body Movement in Musical Performance." *Human Movement Science* 31 (2012): 1137-50.

[25] Friedrich Platz, and Reinhard Kopiez, "When the Eye Listens: A Meta-analysis of How Audio-visual Presentation Enhances the Appreciation of Music Performance." *Music Perception: An Interdisciplinary Journal,* vol. 30, no. 1 (Sep. 2012): 71-83.

[26] Vincent Bergeron, and Dominic M. Lopes, "Hearing and Seeing Musical Expression." *Philosophy and Phenomenological Research* 78, No. 1 (January 2009): 2.

[27] Davidson, "Visual Perception," 103.

[28] Jane W. Davidson, "What Does the Visual Information Contained in Music Performances Offer the Observer? Some Preliminary Thoughts" In *Music and Mind Machine: The Psychophysiology and Psychopathology of the Sense of Music,* ed. Reinhard Steinberg, (Berlin, Germany: Springer-Verlag, 1995), 105-6.

> Within a few seconds tenderness, boldness, exquisiteness, wildness succeed one another; the instrument glows and flashes under the master's hands . . . He must be heard and seen; for if Liszt played behind a screen a great deal of poetry would be lost.[29]

Schumann's words imply not only the importance of the sonic properties of musical expression, but also the significance of the visual properties of expressive body movement on the stage. Schumann is reminding us that there is a great deal of expressiveness in music performance, which might be lost when people listen to audio recordings instead of seeing live performances.

[29] Robert Schumann, *On Music and Musicians.* eds. Konard Wolff, trans. Paul Rosenfeld, (Berkeley, US: University of California Press. 1946), 156.

# Chapter 2 : Objectives of Audio-Based Visualization

## 2.1 Congruency between Sound and Visuals in Electroacoustic Music

Audio-based visualization is a visualization driven by audio-derived data to interpret the audio in visual sense. (see Figure 1.)

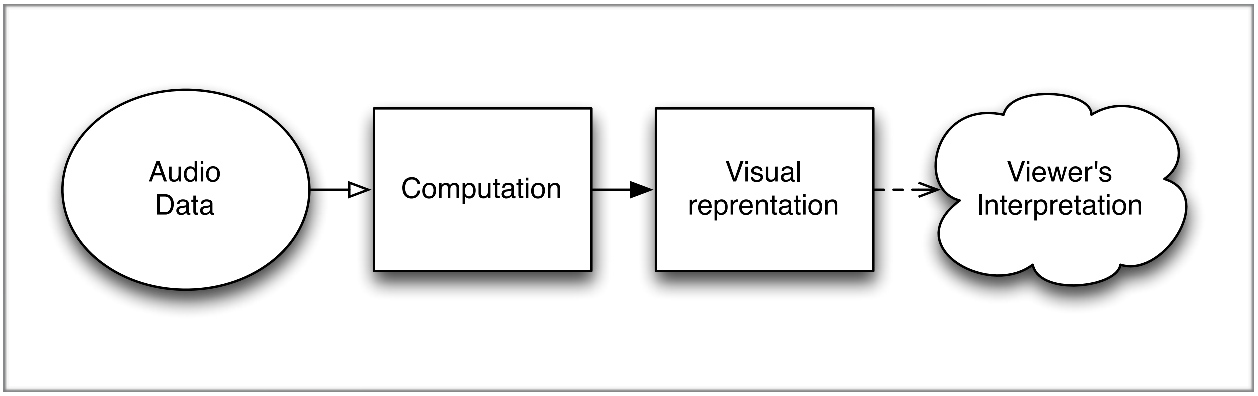



Figure 1. Prior work of audio-based visualization

It is generally used for scientific work to study the audio information in visual form. The audio-based visualization of this study is a performance-oriented visualization driven by audio-based data, with the goal of helping the audience appreciate expressive music performance. (see Figure 2.)

Instead of view's interpretation of the visual representation, the goal of this paper is to create the better environment of music appreciation of live music performance. For actualizing the audio-based visualization which can be universally used in electroacoustic music, the categorization of characteristics and the defining of fundamental attributes of electroacoustic music are necessary.

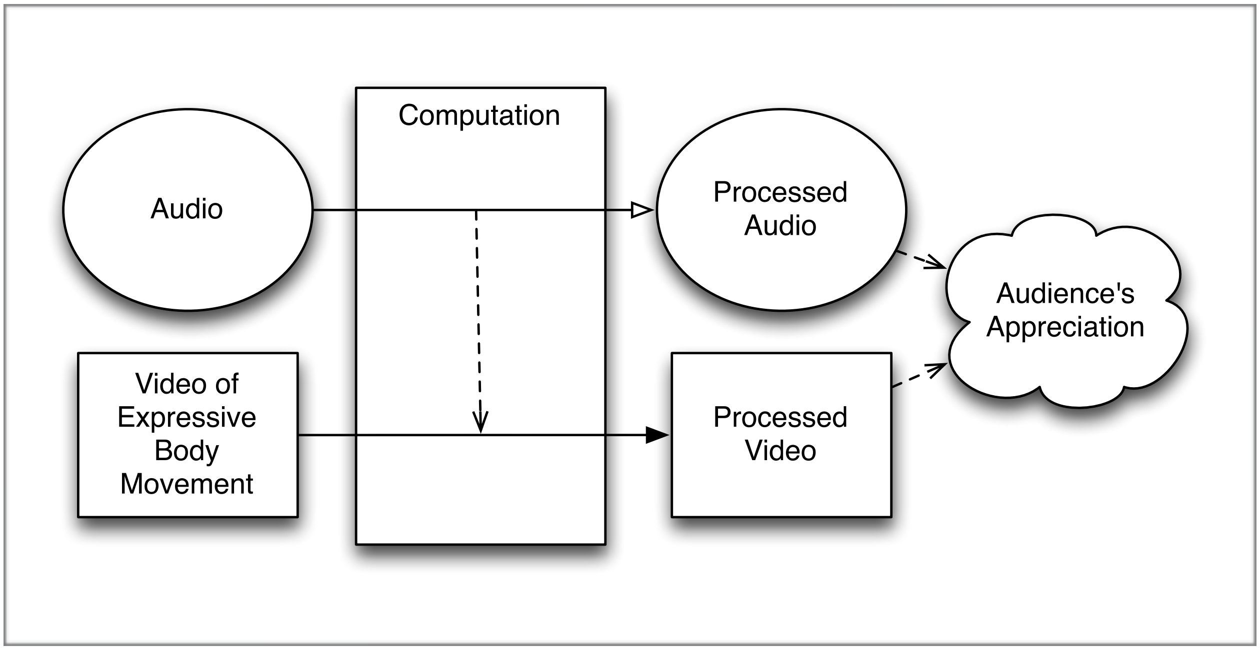


Figure 2. Scheme of the audio-based visualization in this paper

2.1.1 Fundamental Attributes of Electroacoustic Music

Since avant-garde artists and composers push the limits of their fields, generally they are working outside of prior styles. Avant-garde artists and composers are not simply creating art to please the people's ear; they are attempting to create a style of art in a way that no one else has created before.[30] At the same time, their new style of art does not replace the old one, instead they coexist.[31] Because of these reasons, the investigation of standardized characteristics of all avant-garde music would seem to be an impossible task. Even if one found a general musical rule in contemporary music, being avant-garde implies that the art form is constantly evolving and that rule may already be broken by new works. If one has only read through the table of

---

[30] Diana Crane, *The Transformation of the Avant-Garde: The New York Art World, 1940-1985.* (Chicago:University of Chcago, 1987), 20-5.

[31] Joaquim M. Benitez, "Avant-Garde or Experimental? Classifying Contemporary Music" *International Review of the Aesthetics and Sociology of Music,* Vol. 9, No.1 (June, 1978): 53-77.

contents of books on contemporary music, one notices that authors have conceded to the impossibility of attempting a unified approach.[32]

However, the definition of some general attributes of electroacoustic music deriving from the avant-garde is possible for the scope of this research. Instead of trying to unify the diverse set of musical contexts of the avant-garde, the generalized fundamental attributes can be obtained in scientific research based on physics and acoustics, research which forms the basis for much electroacoustic music. Since electroacoustic music creates musical objects from sounds that people hear in daily life, the musical sound must be defined in terms of physical phenomena which include not only pitched tones but also noisy sounds.

### 2.1.2 Amplitude and Timbre as Fundamental Attributes

In physics, sound is described as the passage of pressure fluctuations through an elastic medium as the result of a vibrational impetus imparted to that medium.[33] Because sound is a phenomenon from the nature of waves, it may be comprised of only one frequency, as in the case of a pure steady-state sine wave, or it may be comprised of many frequency components, as in the case of noise generated by construction machinery or a rocket engine.[34]

Sound may or may not be audible to the human ear, depending on its frequency content and intensity; oscillation between 20 Hz and 20kHz, with intensity (amplitude) above a sufficient decibel level, depending on the environment, will be audible. More specifically, for the sensation

---

[32] Benitez, "Avant-garde or Experimental?," 53-77.

[33] Daniel R. Raichel, *The Science and Applications of Acoustics, 2nd eds.* (New York, NY: Springer Science+Business Media, Inc., 2006), 13.

[34] Ibid., 13.

of sound to be audible to a human, it needs to be comprised of a certain number of aperiodic or periodic sinusoidal waves, each above a certain amplitude, which depends on the environment.[35]

A pure tone is known as a sine wave and can be described with the equation:

$y(t) = A \sin(2\pi v t + \varphi)$ or $y(t) = A \sin(\omega t + \varphi)$ [36]

where "$A$" is amplitude, "$v$" is frequency in Hertz, "$\varphi$" is the phase in radians, and the quantity $\omega = 2\pi v$.

"A" is always positive by convention, and "A" can be zero. However, since "A" multiplies the sine function, if "A" is zero, there is no sound. As a result, amplitude is a primary component to produce a sound in both physical and mathematical models.

If the amplitude component is excluded from the equation, it remains a weighted-sum of a series of sinusoidal waves. Because the information of a set of sinusoidal waves determines the timbre of the sound and the pitch of the sound, that timbre and pitch are the second and the third components of sound can be inferred while amplitude is the first component.

These three essential components of sound are corroborated by the definition of sound in psychoacoustics. Poynting and Thomson categorize sound in three characteristics for human sensation: loudness (amplitude), pitch (frequency), and quality (timbre).[37] However, the

---

[35] Hugo Fastl, and Eberhard Zwicker, *Psychoacoustic: Facts and Models.* (Berlin, Germany: Springer-Verlag, 2007), 62-63.

[36] Dave Benson, *Music: A Mathematical Offering.* (Scotland, UK: University of Aberdeen, 2008), 17, accessed April 25, 2014, http://www.maths.abdn.ac.uk/~ bensondj/html/maths-music.html.

[37] John Henry Poynting, and Jeseph John Thomson, *Text-Book of Physics: Sound. Vol. II, 3rd eds.* (London: Charles Griffin & Co Ltd., 1904), 7-8.

definition of pitch in acoustics is differently used in music. While pitch in music is the position of a single sound in a particular tonal standard, in acoustics pitch is the relative degree of highness or lowness of a tone. In avant-garde electroacoustic music, many composers are concerned with pitch in its acoustic meaning, and in fact many composers place more importance on timbre than on musical pitch. In other words, their attention is still on the acoustic meaning of pitch, but the attention they place on musical pitch is decreased. This is a good reason to use timbre instead of musical pitch as the essential components for the universal application of the audio-based visualization.

In fact, some electroacoustic compositions are composed only with timbral aspects of sound. One may recall the concept of *musique concrète*, and how Pierre Schaeffer compared the aesthetics of serialism and electroacoustic music:

> Why twelve notes when electronic music has introduced so may more? Why a series of notes when a series of sonic objects is so much more interesting? Why the anachronistic use of an orchestra whose instruments are handled with such obvious anti-naturalness by Webern and his imitators? And above all, why limit the horizon of our research to the means, usages and concepts of a music after all linked to a geography and a history; certainly an admirable music but still no more than the Occidental music of the last few centuries.[38]

Schaeffer pointed out the potential of sound objects as musical notes and he indicated that there might be a limitation of serial music which is still based on pitches of music. Although the evidence of the application of serial techniques can be found in Schaeffer's tape collages, one can recognize that he treated sound objects as musical notes without the aspect of pitch.[39]

---

[38] Carlos Palombini, "Pierre Schaeffer, 1953: Towards an Experimental Music." *Music & Letter* vol. 73, no. 4 (Nov. 1993): 542-57.

[39] Pierre Schaeffer, "Introduction à la musique concrète" *Polyphonie*,vi (1950): 48-50.

In this work, and in an effort to make the accompanying software immediately compatible with many electroacoustic compositions, pitch is excluded as a general attribute of electroacoustic music that is avant-garde. Instead of pitch, timbre is chosen for the second attribute of electroacoustic music. In a manner similar to how purely pitch-based notions of dissonance have informed the syntax of tension and release in tonal music, timbral consonance and dissonance of sound objects as measured by spectral flatness can describe the tension and release in electroacoustic or timbre-based music.

In electroacoustic music, loudness is the primary component of musical sound because it has a relationship to the actual energy that makes sound, and timbre is the secondary component because it is the most elementary and convenient component that differentiates one sound object from the next. With those two components of sound, a congruency between electroacoustic music and visual processing can be formed.

### 2.2 Mapping Audio Data on Visuals

Previously, two components of sound—loudness and timbre—were defined and chosen as fundamental attributes of electroacoustic music. In the visualization of these components, the method of mapping is a critical matter. What visual aspect can be connected to those attributes? What degree of mapping can best achieve congruency? For the mapping between music and visuals, a discussion of which attributes of visuals would be naturally connected to the attributes of music is demanded, and a rudimentary component of visuals should be used for the mapping.

#### 2.2.1 Amplitude to Visual Brightness

Amplitude might be the easier attribute to map, as the amplitude can be applied to visual brightness in intuitive ways, so that the presence of sound corresponds to the presence of light.

As the amplitude of a sound is the primary energy used to generate the sound, visual brightness can also be considered a fundamental aspect to see the visuals. Conversely, no amplitude would result in no sound, causing a lack of visuals.

Several psychological studies already reveal correlation between amplitude and visual brightness. Lawrence Marks researched the relationship enthusiastically. He wrote several articles related to the perceptual congruence between pitch, loudness, and brightness of light. Throughout his articles, he consistently argues that there is a positive correlation between them.[40] In one of his experiments, participants in two groups, four year-old children and adults, adjusted the degree of similarity between pairs of visual stimuli and auditory stimuli. Among mixed pairs of sounds and visuals, both groups perceived the congruence between brightness and loudness and between pitch and brightness.[41]

### 2.2.2 Noisiness to Visual Distortion

A timbre of sound was discussed as the second fundamental attribute of electroacoustic music, and the mapping of timbre to visual effect can be much more complicated. Because of an almost infinite number of possibilities of sound in real life, finding standardized mapping methods for timbre cannot easily be done. Even if the timbral information is obtained with a Fast Fourier Transform, audio visualization requires many complex mappings such as one-dimensional data to multidimensional data or multidimensional data to one-dimensional data.

---

[40] Lawrence E. Marks, "Cross-Modal Interactions in Speeded Classification." In *Handbook of Multisensory Processes*, ed. Gemma Calvert, Charles Spence, and Barry E. Stein, (US: MIT Press), 85-102.

[41] Lawrence E. Marks, "On Cross-Modal Similarity: Perceptional Structure in Pitch, Loudness, and Brightness." *Journal of Experimental Psychology: Human Perception and Performance*, 15, (1989): 586-602.

Even though many other applications and studies have done audio visualization for scientific or graphical representation with spectral data, since our visualization in this study is derived from live video of a performer's movements, we don't require all the spectral information used in scientific studies. Instead, our research requires an intuitive method that can be easily correlated with visual effects for the audience's perception of congruency. In this study, Spectral Flatness Measure (SFM) of sound was employed as an evaluation of the quality of timbre. SFM is defined as the ratio of the geometric mean of the power spectrum to the arithmetic mean of the power spectrum.[42] By using SFM the value representing a sound's place on a spectrum of tonality to noisiness can be found, which can be used for measuring the timbral quality of music. In this work, for lack of a better term, the word "noisiness" is employed to describe how spectrally flat a signal is. A signal with the highest measure of noisiness (1) is white noise, and a signal with the lowest measure of noisiness (0) is a pure tone.

There are many ways that the visual image can be altered according to the timbre of sound for audio-based visualization. For simplicity, in this study, it has been decided to use the mass spectral flatness to blur or distort the visuals. Since the value of SFM is representing the degree to which a sound is tone-like as opposed to noise-like, it can be correlated with the noisiness of visuals. In other words, the quantity of the noisiness in the sound can be associated with the level of distortion of the visualization. Subsequently, the value of noisiness by SFM is connected to the value of the distortion of the visualization. Depending on the type of visual

[42] James D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria" *IEEE Journal on Selected Areas in Communications.* vol.6, no.2 (Feb. 1988): 318.

effects, the final result of the visualization can be varied. Detail of the visual effects by noisiness will be discussed in Chapter 3.

## 2.3 Visualization of Expressive Body Movements

Because the number of video cameras determines a variety of visuals, the proper number of cameras should be contemplated, and because the placement of the camera is directly related to the visuals' perspective of the performer, each camera must be positioned with careful consideration for the final result of the visualization. In addition to camera arrangement, lighting is an essential element in the visual presentation. By projecting appropriate lighting, clear contrast of the light and shade can be created, and by clear contrast, several levels of silhouette can be produced. These three components—the number of cameras, the placement of cameras, and the manner of lighting—must be taken into consideration.

### 2.3.1 Multiple Video Camera and Placement

Multiple Perspective Interactive Video (MPI-video) is a strategic system that allows a viewer to view an event from multiple perspectives interactively, even based on the contents of the events.[43] The idea of MPI-video has been useful for various research such as immersive video for telepresence systems, traffic monitoring and control, and the analysis of physical performances, sports and dance.[44] Using MPI-video, a set of sequences is obtained from an object or event, and used to build a multi-dimensional model of the important events in a scene.

---

[43] Arun Katkere et. al., "Multiple Prospective Interactive Video" In *Proceeding of the International Conference on Multimedia Computing and Systems* (1995): 202.

[44] Ibid., 202-11.

In the case of audio-based visualization, the information of audio decides the best camera angle at every time instant, while other systems consider user request for camera selection.[45]

One of the purposes of this research is to project the performer's expressive body movement in multiple perspectives which cannot be easily viewed from the audience. Depending on the placement and angle of the camera, the expressiveness can be magnified or reduced. If the placement and angle is not in an appropriate position, the cameras might not contribute to the visualization of the performance. Thus, the composer should carefully place and angle the cameras, according to the final visualization on the screen.

There are two aspects to be maintained when the cameras are positioned for the performance. The first aspect is finding suitable locations to place the cameras. A location is acceptable if it meets the following three requirements: the camera is not blocking the view of the audience, the camera view is not blocked by the equipment or instrument, and the camera is not disrupting the performer's movement. If the camera is placed in a suitable distance from the performer, and if the space between performer and audience is not disrupted on the stage, a feasible location might be found.

The second aspect is the angle of the cameras. The angle of the cameras should be varied depending on the final aim of the visual effects. If multiple camera views are to be projected in a distinct place on the screen, the orientation of the subject in the views should be separated by physically adjusting the angle of the cameras. For example, this would be necessary if the performer's hand is positioned on the left side of the screen, and the performer's face is placed on the right side of the screen. However, the positioning of each view can be controlled

---

[45] Arun Katkere et. al., "Multiple Prospective Interactive Video", 202-11.

computationally, which is not considered in this study, but even in this case physical positioning on the proper area of the screen is needed.

Since the brightness of each camera view is controlled by the amplitude of audio processing, each subject of the view can be separated depending on angle of cameras and the loudness of each audio effect. However, because there is a possibility that multiple views can be overlapped simultaneously, depending on the composer's intention the separation of views of subjects might be necessary. (See Figure 3.)

Figure 3. Comparison of the overlapped view and juxtaposed view.
Left: Two views are overlapped in the center, Right: Two views are juxtaposed.

In addition, the final decision of equipment setting was based on two criteria; whether the final visuals displayed each gesture and its connection to sound, and whether each video was adequately magnified to show the expressiveness of gesture. For example, when each view of the gesture overlapped and it obscured its connection to sound, the positions of each view were juxtaposed. When expressive body movement is too small to be recognized, it is magnified and overlapped in the center.

### 2.3.2 Three Point Lighting and its Usage

In video production, quality lighting is one of the most important elements. When everything is brightly lit with a fine use of lighting and shadow, the best images can be made.[46] In the actual application of lighting for images, several factors were considered: ambient light, white balance, softness or hardness of light, light bouncing, shadow, and depth of subjects. Among them a couple of factors can be modified physically by the placement of lighting. More specifically, "three-point-lighting," the standard ideal lighting arrangement for video production was adapted. In this system three lights are required: a key light that is the hardest and brightest light on the subject, a fill light that is a soft light to fill the shadow cast by the key light, and a back light that visually differentiates the subject from the background. By using three separate positions, the subject can be illuminated properly, and the desired fine images or vivid videos can be achieved.

In the performances of the compositions, the usage of a fill light was not implemented. Because the fill light faces toward the audience, it might distract the view of the audience. Another reason is that the fill light might eliminate the shadow which creates "depth" and "emotion" in the images.[47] Only the key light and the back light were used.

While the key light can be placed in front of the performer, the back light can be created by using the house lights of the concert hall. Most concert halls have their own house lighting system to point at the performers, one can take advantage of the house system. The combination

---

[46] Philips McIntosh, *Introduction to Lighting*. (Seattle, WA: Amazon Digital Service Inc., Feb. 2012), 3.

[47] Ibid., 12.

of the house light and point light can help to illuminate expressiveness of the performer's motions. Depending on the direction, house lighting can be used as key light or back light. During the preparation of the performance, different lightings using the adjustable portable light were attempted and compared, and appropriate lighting arrangements were chosen by the composer. More details on the lighting for each composition is discussed in next chapter.

### 2.3.3 Formal and Semantic Congruency

In the first chapter, two kinds of congruency in the psychological research were reviewed: formal congruency and semantic congruency, and in the previous section of this chapter the amplitude and noisiness applying to visual brightness and distortion were discussed. By connecting the data of the amplitude to visual brightness, the formal congruency between the visuals of performer's instrumental gesture and instrumental sound can be created. By connecting the noisiness to visual distortion, the audience can feel another Semantic congruency between timbral dissonance and visual distortion.

However, some portion of expressive body movement is missing in the visualization. Since in the audio-based visualization the audio data is controlling the brightness of visuals, audio always precedes visuals. It means that some portion of expressive body movement cannot be seen through the visualization when there is no background sound or no sustained instrumental sounds previously. In other words, the audience cannot see the expressiveness of the performer in the visualization if there is no sound. Nonetheless, this missing portion of expressive body movement is not considered as important by the author, because in the actual performance it was too subtle to affect the congruency of the particular performances for this work, and it can be seen only in the beginning of the performance.

### 2.4 Procedure and Hypotheses of the Survey

Surveys were conducted in order to verify the validity and value of the audio-based visualization.

#### 2.4.1 Procedure

Surveys were processed during the doctoral recital of the author, and the recital was open to the public. Three electroacoustic compositions were performed without an explanation of the objectives of this research, and then a lecture followed the performance. The lecture was given following the performances in order to help avoid biasing the perception of the participants according to the opinion of the author. After each composition, the audience filled out a questionnaires about the composition.

Forty-two people attended the concert, and everyone participated in the survey. Among them, twenty-five people were male and seventeen were female. The average age was twenty-nine years with a standard deviation of 13 years and 6 months. Participants were sorted into two groups by categorization of their musical background: years of musical training, the number of avant-garde music concerts they had previously attended, and the number of electronic music concerts attended. The participants who have been trained more than four years were twenty five, while seventeen people had musical training of less than four years or none. Twenty participants had attended an experimental music concert once or never, and the participants who had attended more than one numbered twenty two. Half of participants had no experience, or had attended electronic music concerts once, and the other half had attended two or more. (see Appendix D)

### 2.4.2 Hypotheses

There are four hypotheses in the survey. The first hypothesis is that the audience can perceive congruency between amplitude and visual brightness, and between noisiness and visual distortion in the audio-based visualization in the three compositions. Second, the audio-based visualization is helpful to appreciate the music performance and expressive body movement. If the first two hypotheses are true, these two hypotheses become the assumption of the third and fourth hypotheses. The third hypothesis is that the degree of perception for the congruencies of loudness and noisiness is different depending on musical background, and the fourth is that the degree of appreciating the performance and expressive body movement is different depending on musical background.

Based on these four hypotheses, a questionnaire was made with the following factors: (a) musical background of the participants, (b) perception of the congruency between visuals and music, and (c) whether the visualization is helpful for music appreciation and expressive body movement. By these three categorizations of the musical background of the participants, the degree of perceived congruency and appreciation of the performance can be compared within the groups. (see Appendix C)

# Chapter 3 : Implementation

## 3.1 Music and Audio processing

Designing the audiovisual system requires a specified methodology to implement the objectives. Even though a concrete theory is well-defined, without the example of a real application many details and issues might not be discovered. Cognitive scientists Minsky and Papert indicated the importance of real cases in their book:

> Good theories rarely develop outside the context of a background of well-understood real problems and special cases. Without such a foundation, one gets either the vacuous generality of a theory with more definitions than theorems—or a mathematically elegant theory with no application to reality. Accordingly, our best course would seem to be to strive for a *very thorough* understanding of well-chosen particular situations in which these concepts are involved.[48]

They indicated the importance of the real application and actual implementation. The purpose of this chapter is the specific implementation of audio-based visualization and its performance. From the general scheme to the actual usage, the specification and methodology of audio-based visualization is discussed, and, from software to hardware, the issues and problems of musical concepts and audiovisual processing are examined.

Three electroacoustic compositions, *Piece for Solo Snare Drum and Electronics, Memoriam,* and *Synesthetic Moment,* are examined as examples of audio-based visualization. The audio processing which controls video processing is discussed, and the concept of the

---

[48] Marvin Minsky, and Seymour Papert, *Perceptrons: Expanded Edition.* (Cambridge, London:The MIT Press. 1988), 3.

[49] David Zicarelli, "An Extensible Real-Time Signal Processing Environment for Max" In *Proceedings of the International Computer Music Conference 1998,* (San Francisco: ICMA, 1998): 466.

programming in Max 6[49] (a graphical audio-programming language) is examined. Figure 4 shows the overall scheme of the programming of audio-based visualization.

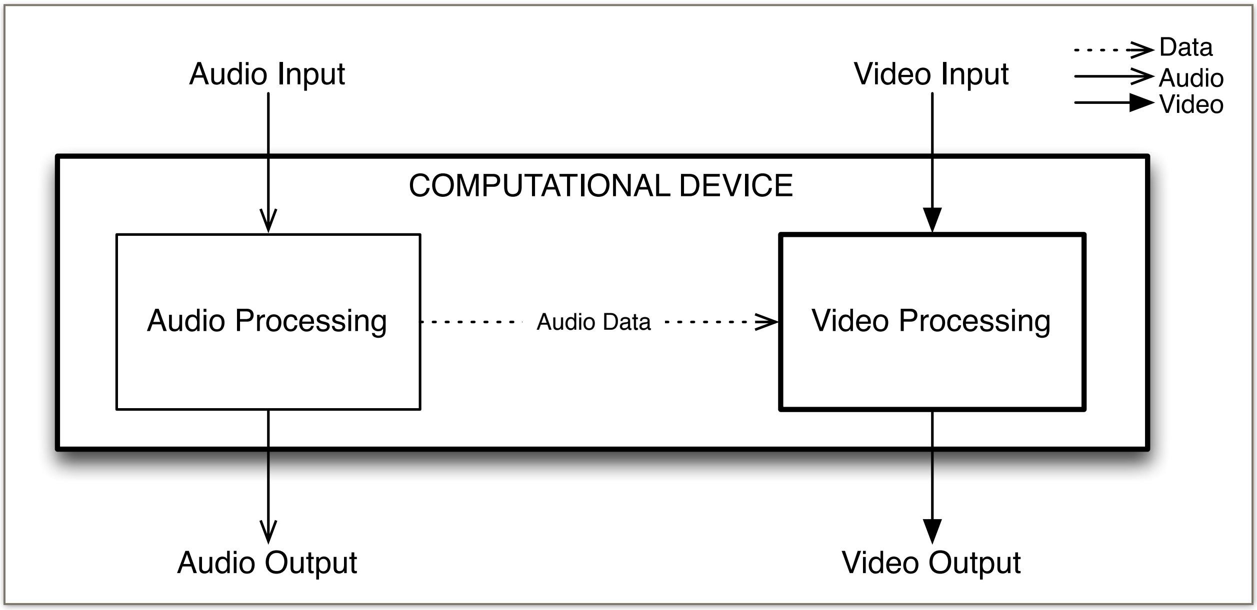


Figure 4. Overall Scheme of audio-based visualization for Live Performance.

3.1.1 Musical Concept of the Works

Each composition uses different approaches to convey its own artistic design and aesthetic intention. All three compositions are musically notated in scores, with some portion of the scores directing improvisation by the performer. Each composition has different musical ideas, and each idea is conveyed in a slightly different way.

The main musical idea of the first piece, *Piece for Solo Snare Drum and Electronics,* is to explore the tension and relaxation created by increasing and decreasing dynamics and rhythmic accents. Originally the piece was composed in 2006 for three snare drums, and it was revised for

solo snare drum and electronics in 2011. In the original version, the poly-rhythmic patterns such as hemiola were repeated and exchanged between three performers, but in this revised version, the computer repeats and alternates the patterns between virtual performers. The repeated sections are replaced with delayed sounds which have irregular durations determined by the computer within a certain range of durations. The overall form of this piece is ternary (A-B-A'). The A section starts with irregular rhythmic patterns, which is the prepositional section, and after the motivic pattern appears the performer begins to explore the tension and relaxation of rhythms in earnest. At the end of the B section, there is a Cadenza-like section in which the performer improvises the motives and materials freely, culminating in a return of the original motive of the A section.

The second piece, *Memoriam*, has two ideas to achieve. One is to express a person's complicated feelings at a funeral of a close friend or family member. In the situation, the person does not want to recall the memories of sadness, but he is forced to remember it by the others, whether they intend to remind him or not. As a metaphor for the memories of sadness, video and audio are recorded and played. More specifically, at the climax of the piece, the computer plays back the recorded video and audio which were captured in each section of the piece. The second idea is based on an exploration of the percussive sounds of a guitar. Conventionally a guitarist plays on the strings and neck, but in this piece other extended techniques were attempted. The score was written with special consideration of the possibility of percussive sounds and resonant sounds of the guitar body. This piece is a solo guitar piece, but is played by a percussionist since most of the piece is based on percussive techniques.

The last piece, *Synesthetic Moment* is for piano solo and electronics. The main intention is to explore some of the timbres of the piano, from tapping sounds on the keyboard to sweeping sounds on the strings. The piece has three different modes of playing: below-keyboard, on-keyboard, and above-keyboard. The performer starts the piece with tapping sounds on the side of the keyboard, and then plays notes with normal piano technique. When the piece reaches its climax, the performer sweeps across the strings. In the piece, the audience can appreciate increasing and decreasing tension not only through the timbral contour of music, but also the elevating body movements of the performer.

### 3.1.2 Audio Processing Units

In the digital signal processing of the three compositions, the Max 6 audio programming environment was used. Inside of the programming there are two distinct units: the processing unit and the analyzing unit. The processing unit consists of several sub-units including recording, playback, delay, pitch-shift, granular synthesis and spatialization. The analyzing unit is comprised of two sub-units, a noisiness unit and an amplitude unit. Figure 5 illustrates the overall scheme of the audio processing.

In the processing units, there are a couple of Max bpatchers (imported patch encapsulation) designed by the composer. The first bpatcher, “10tapsPitchShift” created for a customized delay effect and pitch-shift of which parameters can be manipulated differently. This bpatcher can produce up to ten distinct delay sounds, and all the parameters, such as delay-time, feedback-rate, and interval of pitch-shift, are randomized within a certain range of numbers. When the amplitude of the input sound passes a certain level, the analyzing unit sets the delay-time of the sound and the interval of the pitch-shifting and triggers to play it. The delay units of

the bpatcher do not loop constantly, but play back only certain times in short delays (0-500ms) and irregular feedback rates. Sounds are then sent through the “gizmo~” object, which performs frequency domain pitch shifting on the sound. Even though a standard delay effect technique is the main signal processing of this bpatcher, the actual output does not sound like a delay effect because all the parameters are manipulated differently by randomized numbers. This bpatcher was commonly used throughout three compositions.

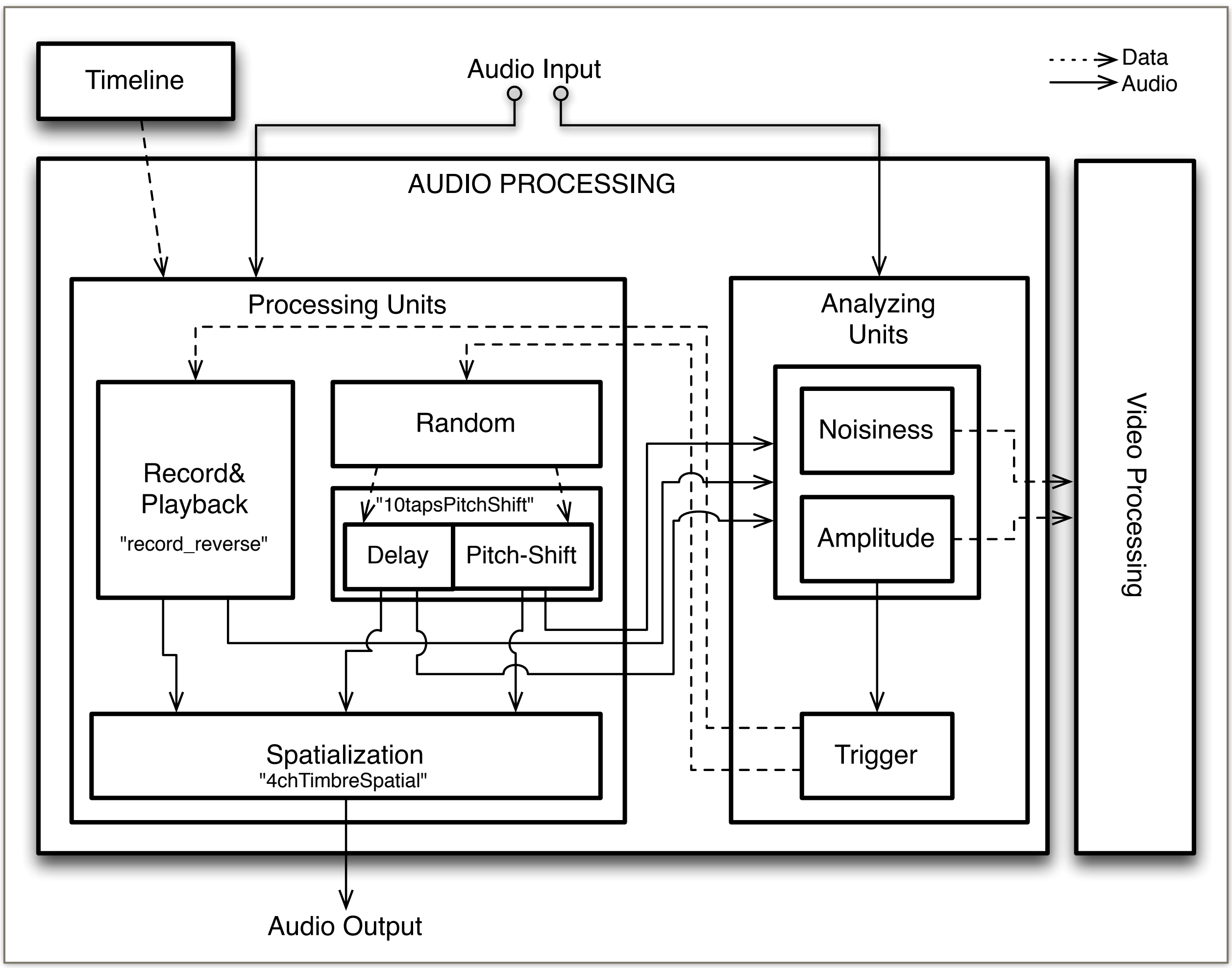



Figure 5. Overall Scheme of Audio Processing: Signals and Data streaming of Processing Unit and Analyzing Unit.

The second bpatcher, “record_reverse”, was made for the function of recording a sound into a buffer and playing it back at different rates with or without pitch-shifting. The recording is

activated by the amplitude of live sound input, and playback is triggered either at the end of the recording or by the events of the compositions. This bpatcher consists of “record~”, “buffer~”, “groove~” and “gizmo~” objects. The “record~” object records the input sound into “buffer~” object, and the “groove~” objects play the recorded sound by playback rate. “gizmo~” shifts the pitch by FFT data. This patch is only used in the third piece, *Synesthetic Moment*.

A third bpatcher, “4chTimbreSpatial”, is also only used in the third piece. The main function of this bpatcher is to diffuse sound through a multi-channel speaker array, and the current version of the patch includes a quadraphonic speaker arrangement. In the bpatcher, a couple of objects were used for spatialization, for example the “cross~” object separates high and low frequency sounds at the crossover frequency or cut-off frequency, and the two separated sounds are spatialized in different positions. Two separated sounds were controlled by randomized position with spherical coordinates: azimuth, elevation, and distance. Two objects made at the Institute of Computer Music and Sound Technology, ambiencoder~ and ambidecoder~, were used for the spatialization. The function of these objects is to allow encoding and decoding using real-time spatial data of virtual sound sources and speaker configuration in three dimensions of up to third order B-format ambisonics.[50]

Besides those bpatchers, several audio processing patches were employed, including a delay effect with pitch-shifting, spectral synthesis, reverberation, and granular synthesis with the idea of a sound particle cloud. For delay effect with pitch-shifting, “+pitchdelay~” made by Tom

---

[50] Jan C. Schacher, and Philippe Kocher, “Ambisonics Spatialization Tools for Max/MSP.” In *Proceeding of International Computer Music Conference 2006,* (San Francisco, CA: ICMA, 2006).

Erbe was used for all the compositions, and the sound of this object can be heard in several parts of the music. The function of the object is to delay a sound with continuous pitch-shifting. More specifically, each delayed sound has a different pitch as time moves forward, either consistently increasing or decreasing. For spectral synthesis, “tap.spectra~” object from the tap.tools object collection which is created by Electrotap was employed. The object remaps the spectral data from FFT bins to different bins of an IFFT. For instance, if 100hz - 500hz as source range and 500hz - 900hz as destination range, frequencies from 100hz to 500hz of the sound are shifted up to the destination range of 500hz to 900hz. As an output, depending on the parameter of frequency range, pitch and timbre of an input sound can be modified to create different sounds.

Granular synthesis of a sound particle cloud was used in the first composition. Using the randomized particle within the range, the stochastic delayed sound is produced,[51] and the delayed sound represents the virtual performer in the piece For granular synthesis, “gran.cloud.live~” object created by Nathan Wolek was employed. One of the purposes of the composition is to produce a virtual performer’s sound, and usage of this object in the piece is to sample the actual snare drum sound and reproduce the sound as granular synthesis. Because a percussive sound has a fast attack, an exponential decay envelope was used for windowing each grain.

Reverberation with different parameters was used in three compositions. Two reverberation objects were used: “newverb~” made by CNMAT of University of California at Berkeley, and “yafr2~” designed by Randy Jones from Madrona Labs. Besides the objects, a reverberation with over 30 seconds was demanded for the musical concept of the second piece.

---

[51] Curtis Roads, *Microsound*, Cambridge, MA: MIT Press. (2001): Chapter 1.

The “yafr2~” object was customized to achieve a longer reverberation time by changing its internal feedback parameters.

### 3.1.3 Audio Analyzing Units

There are two components derived in the audio-analyzing unit—noisiness and amplitude—which provide two functions for influencing the processing units. The first is to send data to the video-processing unit for manipulating the visual effects dynamically, and the second is to trigger audio effect events in the audio-processing unit. The triggering can be decided by the amplitude of sound or the velocity of sound impact. For analyzing the noisiness of a sound, the “noisiness~” object created by Tristan Jehan was used, and a built-in object “peakamp~” was used for amplitude data.

There are a couple of issues in analyzing audio data. Because several “noisiness~” and “peakamp~” objects were implemented along with various audio effects, the computer slowed down, or had high latency in the processing. This issue can be solved by changing the polling interval of the objects. After trying many different intervals, a number above 50 milliseconds was quite stable for the visualization. There is another reason to choose 50 ms as the frame rate, which will be discussed in the next section of this chapter.

For better performance of visual processing and audio processing, it is necessary to minimize computational delay or latency. In the “noisiness~” object, one of the arguments is the order of onset, and by providing different arguments each object can be separately executed in succession, which reduces the possibility of slowing-down the CPU. Because there are several ways to analyze amplitude, several objects were tested to find out which is most efficient and least computational. Max objects “levelmeter~”, “meter~”, “live.gain~”, “avg~”, “average~”,

and “peakamp~” were examined, and “peakamp~” was less computational than the others by up to 20% when 1000 copies of the objects were executed.

## 3.2 Magnification and Augmentation of Live Video

The visualization requires specific circumstances and particular programming which are related to audiovisual mapping, visual processing, and installation of visual equipment. Through experimentation, various methods were examined. According to the varying successfulness of those methods, the most efficient method was determined.

### 3.2.1 Mapping Audio Parameters to Visualization

The visual-processing unit receives data on the amplitude and noisiness of sound from the audio-analyzing unit. After receiving the data, the visual-processing unit converts amplitude and noisiness values to visual brightness and distortion values. Since each sensation has different perceivable measurements, the transformation of the data to the correct scale is essential; logarithmic decibels of sound should be changed to linear numbers of visual brightness, and the spectral flatness value should be transformed to the percentage of visual distortion, which can be represented by the number from zero to one.

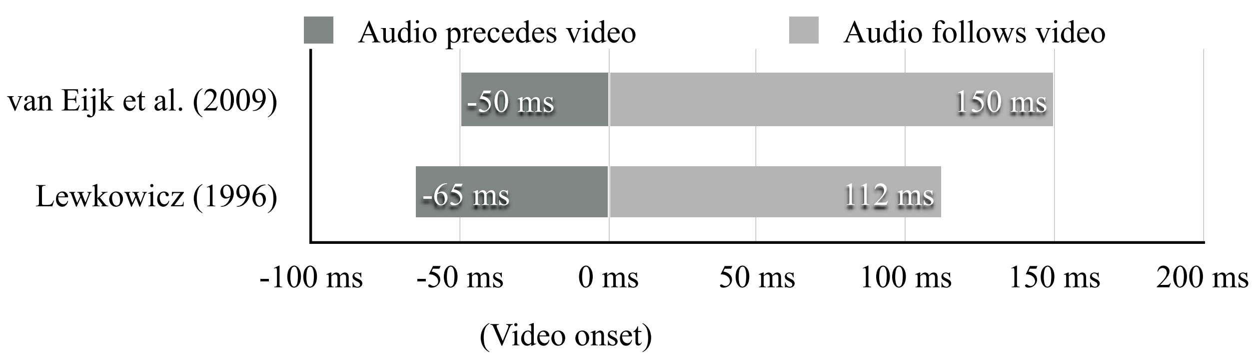
For the mapping of amplitude to brightness, two steps of unit changing and one of scaling were applied. The first step is to map the exponential amplitude data of “peakamp~” to the logarithmic scale of human hearing, which more closely approximates the audiences’ perception. The second step is mapping the logarithmic decibel data to the linear value of brightness. The third step is the scaling of the minimum value of amplitude mapping, which is optional depending on the overall situation of brightness. For instance, when a motivic sound is played with loud background sounds, the visualization of the motivic sound is hardly recognizable,

because the loud background sounds raise the brightness of all the visuals. This issue results in no congruency of motivic  sounds. In this case, additional scaling of the loud background sound is required. The function of the scaling is to align the brightness and perceived amplitude by changing the minimum value of the units.

3.2.2 Audiovisual Synchronicity

According to Van Eijk, Kohlrausch, Juola, and Van De Par, cross-modal events will be perceived as simultaneous, when the audio component precedes the video by fewer than 50 ms or the audio follows the video by less than 150 ms. The range is called *point of subjective simultaneity* (PSS) by the authors. More research on the perception of synchrony is Lewkowicz's article.[52] In this experiment, he compared adults and two to eight-month old infants, and discovered the threshold for perceiving asynchrony. When the sound preceded the visual, adult participants perceived asynchrony starting at 65 ms, and when the visual preceded the sound the threshold was 112 ms, while the infants required less than 350 ms and 450 ms of difference, respectively. Table 1 show the audio range of perceived synchrony from the research.

Table 1.  Range of Perception of Synchrony by Audio onset or PSS




---

[52] van Eijk, R. L., A. Kohlrausch, J. F. Juola, and S. van de Par, "Temporal Interval Discrimination Thresholds Depend on Perceived Synchrony for Audio-visual Stimulus Pairs." *Journal of Experimental Psychology: Human Perception and Performance*, 35(4) (2009): 1254-63.

Because part of this research is to examine the way to manipulate visuals with sound, audio processing is executed first, and the visualization is operated by the data from audio-analyzing unit afterward. Thus, the numbers after the visual onset are not related to this study. (see Table 1.) 50 ms was selected to be the interval to capture audio information, because it is the maximum acceptable number and meets the conditions of the perceived synchrony of the two researchers when there is no computational delay.

3.2.2 Visual Effects and Processing

For computational effectiveness, the visualization is mostly programmed with OpenGL (Open Graphics Library). OpenGL, developed and optimized by Silicon Graphics, Inc., is a software interface for graphics hardware.[53] More specifically it is an interface that your application can use to access and control the graphics subsystem of the device upon which it runs.[54] A modern day computer assigns complex visual processing to its Graphics Processing Unit (GPU) to produce images for the display with its own graphics memory that is called Video RAM or VRAM. Current GPUs consist of a large number of small programmable processors called shader cores which run mini-programs called *shaders*, and from a few tens to a few thousand of these cores contained in the GPU greatly increases the processing of graphics by parallel computing.[55] Because OpenGL is mostly executed by the GPU, by using GL objects in

---

[53] Richard S. Wright Jr., and Michael Sweet, *OpenGL Superbible: The Complete Guide to Opengl Programming for Window NT and Window 95.* (United States: Waite Group Press. 1996): 15.

[54] Graham Seller, Richard S. Wright, Jr., and Nicholas Heamel, *OpenGl SupterBible: Comprehensiv Tutorial and Reference.* 6th eds. (Upper Saddle River, NJ: Addison-Wesley, 2014): 3.

[55] Ibid*.,* 5.

Max 6, video-processing of the compositions can be carried out mostly on the GPU. As a result, the visualization is executed effectively, and does not interrupt the audio-processing.

The usage of GL objects in the compositions were programmed in 2-dimensional textures with OpenGL. The textures are rendered on planes facing the screen with z coordinates set to zero. In OpenGL graphics, there are two stages that are programmable: vertex processing and fragment processing. Since this project is dealing with the video from the camera as 2-dimensional textures, only the texture was modified by loudness, and noisiness in fragment processing.

In Max 6, OpenGL graphics are accessed by GL objects, which are named with the prefix "jit.gl." A streamlined interface of the objects performs the general purpose of GPU-based grid evaluation, with each shader object processing the 2D texture mapped onto a 3D object. In this way, a 2D composite image can be created highly efficiently. By using the "jit.gl.slab" object, the customized texture processing in 2-dimensional view can be executed, and by referencing GL Shader Language(GLSL) files as an argument, the customized visual effect based on GLSL programming can be displayed. In the three compositions, three built-in shaders and one customized shader were implemented. Figure 6 shows the arrangement of the shader and data flows.

For the connection between amplitude and visual brightness, a function to control the luminance of each video source is required. To accomplish this, a customized shader was employed, which was programmed by the composer. When using the shader, the value of the amplitude data is multiplied with the value of luminance, and consequently the brightness of the video is manipulated by the amplitude. The "jit.gl.slab" object referencing the customized shader

file can handle up to eight inputs of videos, and in this project, three video inputs were implemented.

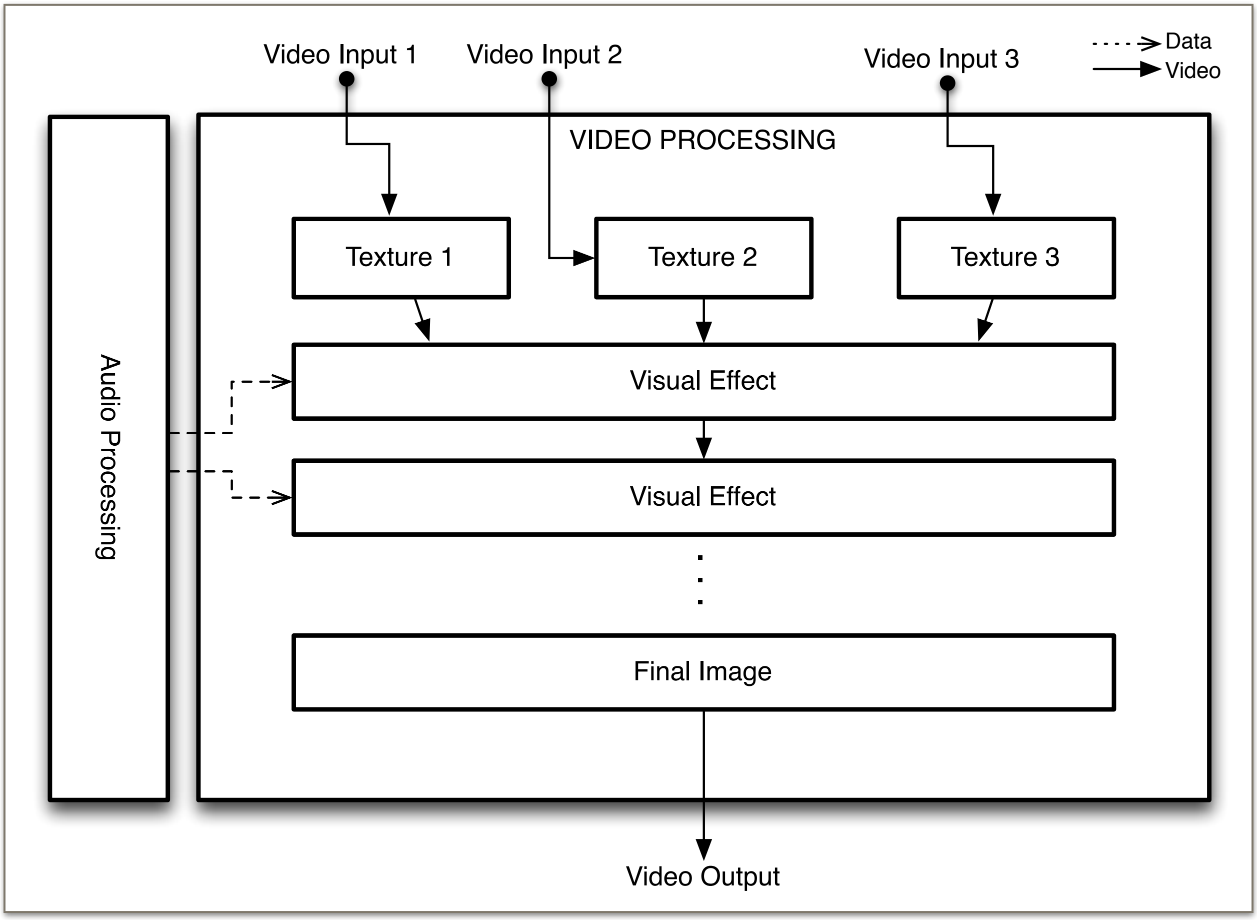


Figure 6. Overall Scheme of Video Processing

The second goal is to connect the noisiness to visual distortion in an artistic way. The artistic goal with this type of processing is to find a visual boundary between clarity and distortion of the video. When the noisiness is high, visual distortion is maximized, and when the noisiness is low, a clear view of the video is displayed. However, if the visual is excessively distorted the expressive body movement cannot be seen, but without enough distortion no congruency is perceived. Thus, the type of visual effect and the degree of distortion are critical. Among many other built-in shaders, three shader effects were implemented.

The first shader is “td.lumadisplace.jxs”, which provides a visual effect that spatially displaces the texture by the luminance. There are two parameters to control the effect: offset, and amplitude. Each parameter needs two float numbers (-1. to 1.) to determine the X and Y coordinate value for the direction and the width of the displacement from the center of picture. Figure 7. shows the difference between a processed image and unprocessed image.

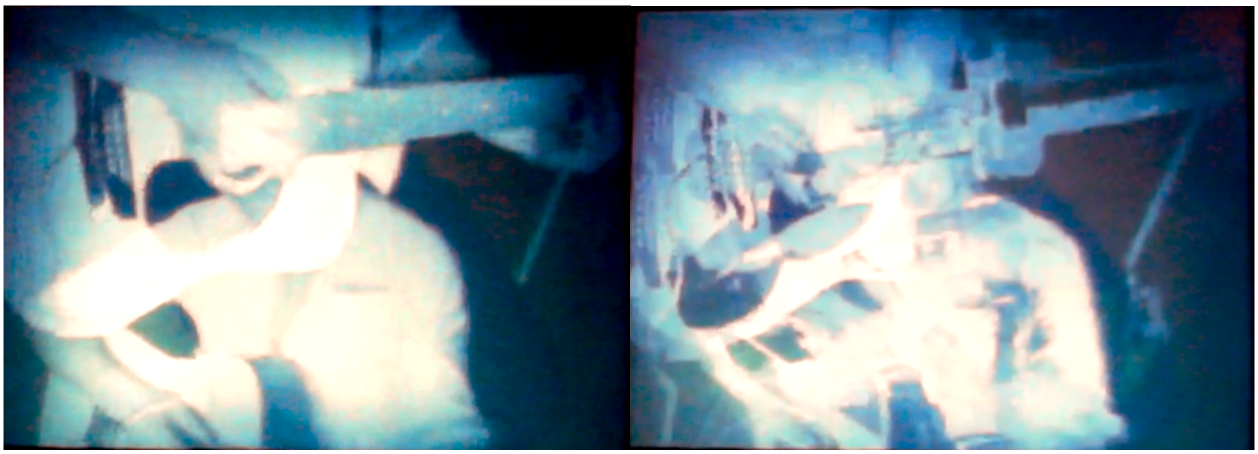


Figure 7. Views with or without luma displacement processing.

Left picture: image without luma displacement processing, Right picture: image with luma displacement processing controlled by noisiness value.

The second shader is “tr.edgeblend.jxs”, which generates gradient alpha for edge-blending. This object requires a parameter that consists of four float numbers. Each float number represents the width of edge-blending of each edge in a 2-dimensional view. The function of this object in the visualization is not only to focus the audience’s vision on the main subject in the video, but also to eliminate the opaque rectangular edge of the video, which might look less artistic.

The third shader is “td.sinefold.jxs”, which results in a ripple-like spatial displacement. By controlling several parameters such as center position, frequency, amplitude, and phase of the ripple shape, a wavy-surface texture of the image can be created. This object was used for

visualizing the spatialization of sound in *Synesthetic Moment,* the third composition. The coordinate of the sound source from the spatialization bpatcher was sent to control the center position of the wave effect. The values of frequency, amplitude and phase of the ripple effect were fixed. As a result, the randomized movement of sound from the bpatcher is presented on the screen by the movements of a wavy texture. The difference can be seen in Figure 8.

Figure 8. Views with or without sinefold processing.
Left picture: image without sinefold processing, Right picture: image with sinefold processing controlled by the spatialization of sound; position of the sound source is middle-left.

3.2.3 Installation of Video Cameras and Lighting

Depending on the lighting situation of the concert hall, the installation of video cameras and lighting is varied. The cameras and lights of each piece were differently placed in the performance according to three-point-light technique.

For the first piece, the cameras were placed in three designated spots: (1) the right side of the performer, (2) the left side of the performer, and (3) the left side of the performer to point from drum to torso. The first two cameras capture the close-images of hand movements from approximately a 45-degree angle in front of the performer. The third camera aims at the upper

body of the performer from the side. Figure 9. shows the placement of the cameras in *Piece for Solo Snare Drum and Electronics.*

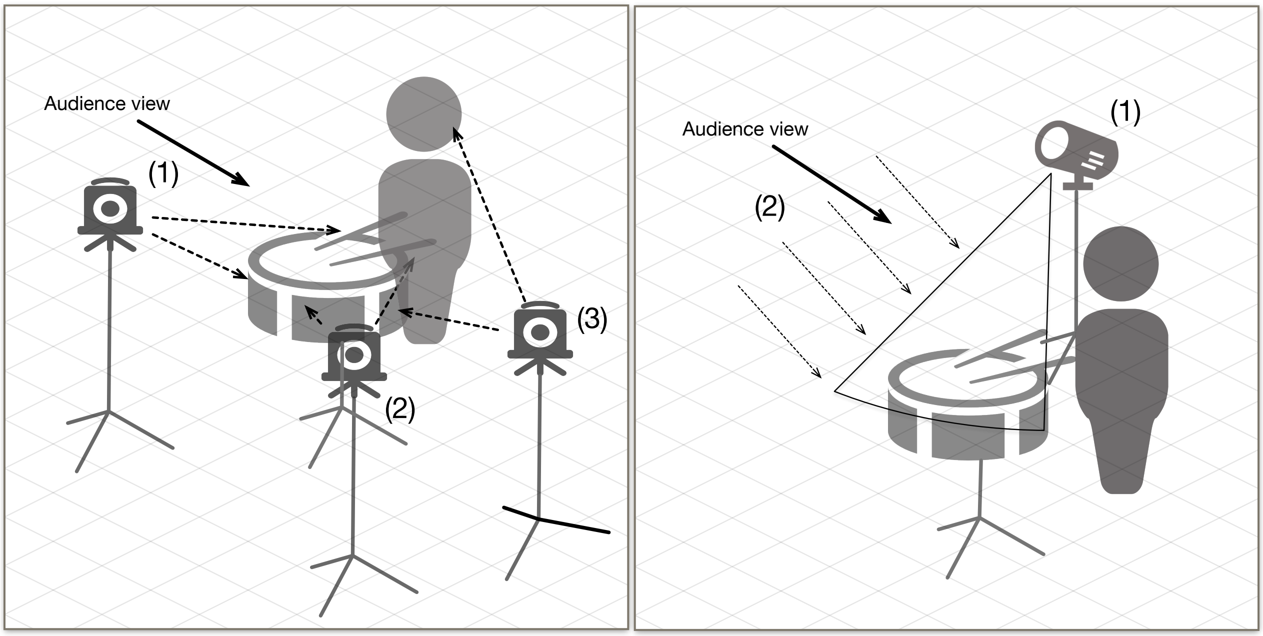



Figure 9. Placement of multiple cameras and lighting for the first piece.
Left : Placement and angle of multiple cameras. (1) Point hands from right side, (2) Point hands from left side, (3) Angle from drum to torso.
Right : Lighting (1) Spotlight from the back, (2) house light from audience side.

The spotlight of the first piece was placed in behind and to the right of the performer. In this case, the spotlight is used as back light, and the house light is employed as the key light and fill light. The lighting of the first piece is illustrated on the right side of Figure 9.

In the second piece, *Memoriam,* three cameras were employed, each in one of three different angles. The first camera angled from the instrument to the legs of the performer. In one version of this piece, the performer crushes the instrument at the end of the piece, completely destroying it, which is not notated in the score. The second and third cameras pointed at the upper body and hands of each performer. (see Figure 10, left)

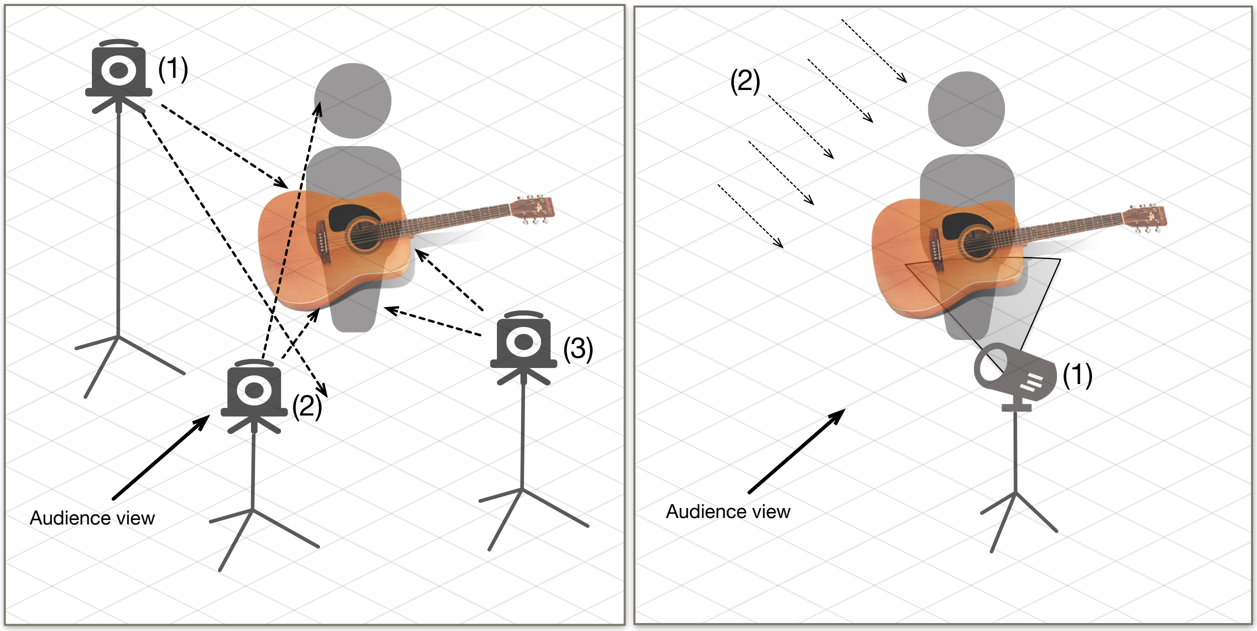



Figure 10. Placement of multiple cameras and lighting for the second piece.
Left : Placement and angle of multiple cameras for the second piece; (1) Angle from guitar to legs, (2) Angle from guitar to face, (3) Angle to hands.
Right : Lighting of the second piece. (1) Spotlight illuminates the performer as Key-light, (2) House light brighten the whole body of the performer from different angle.

The lighting of the second piece has a similar setting as the first piece, but the spotlight was used as a key light to illuminate the expressive body movement and the instrument. The house light was used as back light. (see Figure 10. right)

In the third piece, *Synesthetic Moment,* the placement of the cameras was limited. Because of the open space for the audience's view and the structure of a piano, the cameras were located in similar directions from piano to performer. The first camera pointed at the soundboard of the piano for capturing the hands when the performer plays extended techniques on strings. The second and third were placed on the music rack of the piano. The second camera was angled towards the upper body and face of the performer, and the third one was tilted towards the keyboard for capturing hand gestures. (see Figure 11, left )

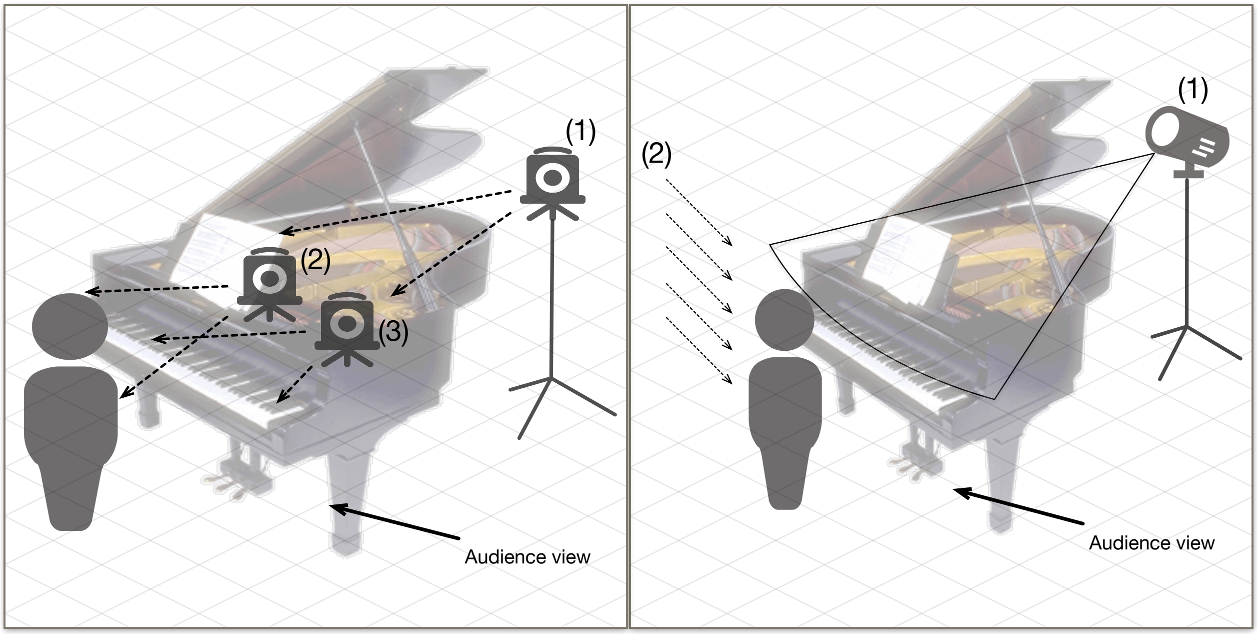


Figure 11. Placement of multiple cameras and lighting for the third piece.
Left : Placement and angle of multiple cameras; (1) Angle to soundboard, (2) Angle from upper body to face, (3) Angle to keyboard. Right : Lighting of the third piece. (1) Spotlight illuminates the performer and piano as key light, (2) House light brightens the whole body of the performer from the back.

Because the top board of the piano blocks the ambient lighting, the spotlight as key light is significant. The key light illuminates the body and face of the performer and the piano as well. The house light as back light decorates the hands and the edge of the performer. (see Figure 11, right)

In each composition, key light illuminated the expressive body movements in the way of three-point-lighting, and house light was used as back light. Three cameras were placed in each composition depending on the instrumentation and performer's position. By applying proper settings of lights and cameras, the visualization of expressive body movements was maximized, and the distraction by equipment on stage was minimized.

# Chapter 4 : Conclusion

## 4.1 Composer's Evaluation of Audio-Based Visualization

From several performances in different venues prior to the doctoral recital, a number of strong and weak aspects were discovered. By addressing the problematic aspects, the system of the compositions was gradually improved. The most recent versions of the three compositions were performed in the doctoral recital. Based on the experience from the performances, various aesthetic values and technological values were obtained.

### 4.1.1 Aesthetic Value

Four advantages of the audio-based visualization are presented which contribute to the aesthetic value of composing with the audio-based visualization. First, the composer does not need to lose the autonomy of the musical content while still augmenting the performance. In other words, the musical content is not changed by the visual component in this method. Because electroacoustic composers are concerned about sound and musical content as the core of the work, it is important that any other component not interfere with the sound of the music itself.

Second, the author believes that the visual property of expressive body movement contributes to convey the performer's intention to the audience, making it more expressive and understandable. By magnifying performance movements, the performer's method of communicating intention can be augmented, hopefully enabling the audience to better understand the musical intention. For example, at the moment the performer is articulating an important phrase with his or her wrists, an audience may see the zoomed-in detail of expressive wrist movements. It has the potential to result in a more understood performance, even if the expressive body movement is small.

Third, the audience can perceive musical aesthetics in the visualization. Since the syntax of tension and release in electroacoustic music has been converted to a visual representation, the audience can potentially appreciate musical aesthetics in two sensations simultaneously. For example, the increasing dynamics in a phrase creates a musical tension, and the dynamics are perceived by increasing the brightness of visuals. As a result, the visual representation of musical contents creates an incorporated sensation.

Fourth, audio-based visualization is adaptable for other compositions. Because the general characteristics of sound are applied in the visualization, the system can be implemented in any composition, and each performance results in different visualization. For example, the source of visuals will be different by providing different stage settings and instrumentation, and the visual effects will be diverse depending on musical content and audio data.

#### 4.1.2 Technical Notes

In the initial trial of the system, two laptop computers were employed: one for audio, and the other for visuals. However, because of the complexity of both hardware and software settings, the system was modified to run on a single laptop computer. Finding effective methods of programming was paramount, and three technical values related to the method were found.

First, an effective method of time synchronization between audio and visuals was achieved. According to the *point of subjective simultaneity* in psychology, the delay time in audio analysis and video processing were tested, and ideal numbers were determined.

Second, an algorithmic method to map musical attributes to visual attributes was introduced. Because music and visuals are two separate sensations, the mapping requires several scalings depending on musical contents. For instance, in the case when there is a loud drone-like

sound in the background with a relatively soft theme, which would otherwise result in the visual brightness of the drone sound overwhelming the visualization, the brightness needs to be additionally scaled in a low range. The same case would apply to a loud theme, in which case the brightness would be scaled to the middle range.

Third, a method to improve visual processing was examined. CPU-based visual processing was converted to GL objects, and unnecessary visual effects were eliminated or replaced. Moreover, the scheme of visual processing was arranged and simplified.

4.2 Results of Survey

The first hypothesis is that the audience's perception of congruency between amplitude and visual brightness is not neutral, and the audience's perception of congruency between noisiness and visual distortion in each composition is not neutral. The second hypothesis is that the audience does not feel neutrally about the visualizations contributing to their appreciation of the music or to their perception of amplification of the performer's body movements. For the first and second hypotheses, we are making an assumption that the audience feels neutral and experiences no congruency on each category of the hypothesis in traditional music performance. Each question in the survey is based on a premise of the hypothesis, and a *one sample t-test* for each question was applied. Collectively, the answer of each question is statistically significant, ($p<0.01$) and the result can be interpreted as the audience perceiving a congruency (see Appendix D). As a result, the null hypothesis, which states there is no significant congruency, is rejected, and as the alternative hypotheses, the first and second hypotheses are accepted: the audience can perceive the congruency between sound and visuals in the audio-based visualization in all three compositions, and the audio-based visualization is helpful to appreciate music performance and

expressive body movement, assuming that the audience would provide neutral ratings for each questionnaire question in the absence of the visualizations.

For the third and fourth hypotheses, the number of participants is questionable and the population is not clearly Gaussian. For better results the satisfaction of two assumptions is required before running two independent sample t-tests: a test for normality and a test for Equal Variance (homogeneity). However, after running a Kolmogrov-Smirnov test (for normality), the distributions did not appear to be normal. To make the distribution normal, several transformations were tried such as log, quadratic, and exponential transformations. Yet, the distributions still were not normal. As a result, one of the nonparametric tests called the *Mann-Whitney U* test was applied, in which the assumption of normality does not need to be satisfied. Depending on the *p-value* of a *Mann-Whitney U* test, whether the third and fourth hypotheses are true can be proved.

For the evaluation we need to separate the participants into two groups depending on musical background, so three factors of musical background were applied: length of musical training years, experience of avant-garde music concerts, and experience of electronic music concerts. There are 17 participants who have musical training of less than 4 years, and 25 people have more than 4 years of musical training, so 4 years is chosen as our threshold to describe musical experience. Additionally, 20 participants have attended an avant-garde concert less than 3 times, and 22 have experienced an avant-garde concert 3 times or more. For the electronic music experience, half of the population has attended the concert 3 times or more, and another half, less than 3 times, so 3 years is chosen as our threshold for experience with avant-garde music.

The third hypothesis is that there are differences between the groups for the perception of the congruency. There were three questionnaires for congruency in the survey, between visuals and loudness, visuals and noisiness, and visuals and body movement. The mean value of most of the answers is around 5 or more degrees (1 = disturbing, 7 = high congruency), but only one response is statistically significant. In the grouping of the experience of avant-garde music concerts, there is a distinction in the congruency between loudness and visuals: group one (less familiar with avant-garde music) has M=5.52, SD =1.2, and group two has M=5.98, SD=1.02, resulting in a *p-value* of 0.012. As a result, the null hypothesis is rejected, that there is no significance between those who have seen several avant-garde concerts and those who have not. The result can be interpreted as people who are more familiar with avant-garde music perceived higher congruency between loudness and visual brightness. (see Table 2.)

Table 2. Means and Standard Deviations for the congruency between sound and visuals in three different groupings.

| | | Musical Training years | | | Familiar with Avant-Garde music Concert | | | Familiar with Electronic Music Concert | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | < 4 years | >= 4 years | p-value | < 3 times | >= 3 times | p-value | < 3 times | >= 3 times | p-value |
| Loudness & Visuals | Mean (SD) | 5.67 (0.93) | 5.81 (1.27) | 0.339 | 5.52 (1.2) | 5.98 (1.02) | *0.012 | 5.74 (1.13) | 5.83 (1.15) | 0.307 |
| Noisiness & Visuals | Mean (SD) | 5.22 (1.22) | 5.14 (1.55) | 0.415 | 5.2 (1.41) | 5.15 (1.42) | 0.452 | 5.27 (1.32) | 4.74 (1.74) | 0.167 |
| Body Movement & Visuals | Mean (SD) | 5.04 (1.51) | 4.79 (1.67) | 0.189 | 4.92 (1.78) | 4.82 (1.43) | 0.416 | 4.97 (1.65) | 4.43 (1.41) | 0.267 |
| *p<0.05 | | | | | | | | | | |

Since the rest of the questionnaire has a *p-value* greater than 0.05 ($p > 0.05$), the responses have a high chance that random sampling would result in the mean randomless being far apart as observed in this experiment. In these cases, more evidence or samples would be demanded to find significance. The fourth hypothesis is that there is a distinction between halves of our three groups (those with and without musical training, avant-garde experience, and electronic music experience) for appreciating the performance and feeling the expressive body movement in the three compositions. Three discriminations were applied, and two questions were asked. Two questions are statistically significant in the grouping of length of musical training. The first question regarded whether the audio-based visualization helped the participant appreciate the music. The group with less than 4 years of musical training has M=5.47, SD=1.76, the other group with more than 4 years has M=5.14, SD=1.38, yielding a p-value of 0.021. The second question is whether the visualization amplify the expressive body movements. The group for less than 4 years has M=5.71, SD=1.22, and the other group with more than 4 years has M=5.13, SD=1.50, yielding a p-value of 0.003. This result can be understood as people who have less musical training years feel more helped by audio-based visualization to appreciate the performance and amplify the expressive body movement. In other words the duration of musical training can affect the degree of the music appreciation of the performance of three compositions. (see Table 3.)

Table 3. Means and Standard Deviations of appreciation the performance or amplifying the expressive body movement.

| | | Length of Musical Training | | | Familiar with Avant-Garde music Concert | | | Familiar with Electronic Music Concert | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | < 4 years | >= 4 years | p-value | < 3 times | >= 3 times | p-value | < 3 times | >= 3 times | p-value |
| Help Appreciate | Mean (SD) | 5.47 (1.76) | 5.14 (1.38) | *0.021 | 5.15 (1.76) | 5.42 (1.3) | 0.333 | 5.3 (1.57) | 5.22 (1.48) | 0.378 |
| Amplify Body Movement | Mean (SD) | 5.71 (1.22) | 5.13 (1.50) | *0.003 | 5.4 (1.46) | 5.35 (1.37) | 0.348 | 5.53 (1.39) | 4.74 (1.32) | 0.195 |
| *p<0.05 | | | | | | | | | | |

In summary, although there are several statistical matters such as populations and normal distribution that could be improved, a couple of valuable results are significant for audio-based visualization  and its application. First of all, audio-based visualization might be useful to help the audience appreciate the music performance, and the method of the audio-based visualization can convey performer's expressiveness between visuals and music. Second, while people who have more experience of avant-garde music perceive congruency between loudness and visuals, these people who have more musical training feel that this method is less helpful to appreciate music performance than those who have less musical training. It is worth noting that the subject test participants were attending the author's thesis defense and may had a non-neutral bias in favor of the author's work. In future work, it would be nice to verify the work by specifically recruiting impartial participants and to carry out control experiments to verify how subjects would answer the questions in the absence of any digital visualizations. Nonetheless, the present survey results appear to positively support the concept of audio-based visualization. In particular,

its advantages seem to include the following: (1) the method of the visualization is helpful to appreciate the music performance, (2) the audience can perceive the expressiveness of performer's body movements through the visualization, (3) the loudness and noisiness of audio can be a persuasive option to visualize music performance.

### 4.3 Future Work

Developing and evaluating audio-based visualization has provoked new compositional ideas and technical plans for future research, deriving from issues and values relevant to the research. These ideas are outlined in the following sections.

#### 4.3.1 Compositional Idea

Although the benefits of audio-based visualization have been demonstrated in the form of three compositions, the idea could be expanded in several ways. First of all, prerecorded video content could be incorporated. By adding pre-recorded videos, the audience could enjoy and appreciate more variety in the visualizations, and at the same time composition could have further options for expressing metaphors in the composition via the pre-recorded video. It contributes to broaden the enormous possibility of audio-based visualization, and to offer a component of *programme music*, which invites imaginative correlation with the music.

Second, a composition for *theme and variations* could give a demonstration of diverse visualizations which gradually change. By developing the musical motif in each variation, the visualization may be gradually changed analogously, and/or by varying visual effects, the different motivic ideas of each variation could be conveyed to the audience. For example, in a rhythmic variation, the time synchronization between audio and sound could be increased by reducing the delay time of processing, and in a harmonic variation, the spectral information of

sound could influence the visual effects. As a result, various musical content of each variation would convert to diverse visual representations.

4.3.2 Technical Setup

Different types of cameras can enlarge the possibility of the visualization. For example, a motorized camera can capture different angles of video depending on the audio or gestural data. The motorized cameras may move around in a designated spot and finds the proper spot for a better view by the data. A high-dynamic-range camera also can be implemented for better images with more details even in the situation of insufficient lighting. A thermographic camera, also called an infrared camera, is another option to experiment in the situation without lighting.

Another plan is to increase the number of lights. In the performance of the three compositions, only one spotlight was employed as key light to illuminate the performer, and house lighting brightened the rest of the subject including instrument and background. However, if other spotlights are added instead of house lighting, the visualization can be more adjustable and satisfying.

The positioning and magnification of the visualization can be dynamically changed by programming. For example, when a sound is shifting to the left channel, the position of the visualization is moving toward the left side of the screen. When the amplitude is decreasing, the visualization is gradually magnifying the gestures to display the small gestures. As a result, additional dimensions of visuals could be amplified for different visual representations.

Finally, various psychological studies of multimedia art are possible. A study of musical noisiness and its perception is rarely found, and a study of realtime responses of audience's

perception that can be realized by using a web-based interface such as NEXUS[56] are important for future research. These studies can broaden the possibility of multimedia art, and the potential of visual-music performance.

### 4.4 Conclusion

Throughout several performances and experiments, several artistic goals were achieved, such as autonomy of music in multimedia art, augmenting expressiveness of performance with visualization, musical aesthetics in visual representation, and broad applicability of audio-based visualization to new compositions. From the result of the survey, one concludes that audiences can perceive the congruencies from the performance, and audio-based visualization can influence the appreciation of live music performance. Moreover, the degree of the congruency can differ depending on musical background, especially on the length of musical training and on the specific implementation of audio-based visualization. Additionally, from these issues and values a more concrete future research can be derived, including a compositional study combining live and pre-recorded video, various visualization methods using different types of cameras and computationally repositioning of each view, or a composition for theme and variations for diverse visual representations, etc. Additionally several psychological studies relating to multimedia art are possible.

The development of multimedia art will continue to evolve and transform into new ideas. It is the responsibility of the composer to lead the evolution of integrating sound in multimedia art to expand the boundaries and meaning of art. The rapid development of technology was

---

[56] Jesse Allison, Yemin Oh, and Benjamin Taylor. “NEXUS: Collaborative Performance for the Masses Handling Instrument Interface Distribution through the Web.” In *Proceeding of New Interface of Musical Expression*, Presented in DaeJeon, South Korea, May 27-30, 2013.

caused an evaluation of the definition of the term multimedia. Powerful mobile devices and small embedded computing devices are ubiquitous, and the usage of audio-visual implements has proliferated. New applications of audio-visual equipment can be found in many contexts, and the complexity of its usage is decreasing. Composers must face this reality and develop new methodologies along with the evolving meaning of multimedia. The author believes the methodology presented will contribute not only to augment the expressivity of the performer in multimedia art, but also to inspire other composers who are defining these multimedia art forms.

## Bibliography


Albarino, Richard. “Goldstin’s LightWorks at Southhampton,” *Variety*, 213, no.12 (August, 1966).

Allison, Jesse, Yemin Oh, and Benjamin Taylor. “NEXUS: Collaborative Performance for the Masses Handling Instrument Interface Distribution through the Web.” In *Proceeding of New Interface of Musical Expression*, Presented in DaeJeon, South Korea, May 27-30, 2013.

Barthes, Roland. “The Grain of the Voice” In *The Responsibility of Forms: Critical Essays on Music, Art, and Representation,* 267-277, trans. by Richard Howard, Berkeley and Los Angeles: University of California Press, 1991.

Barthet, Mathiew, Philippe Depalle, Richard Kronland-Martinet, and Sølvi Ystad. “Acoustical Correlates of Timbre and Expressiveness in Clarinet Performance” *Music Perception: An Interdisciplinary Journal,* Vol. 28, No. 2 (December, 2010): 135-54.

Benitez, Joaquim M. “Avant-garde or Experimental? Classifying Contemporary Music.” *International Review of the Aesthetics and Sociology of Music*, Vol. 9, no. 1 (June, 1978): 53-77.

Bergeron, Vincent, and Dominic M. Lopes. “Hearing and Seeing Musical Expression.” *Philosophy and Phenomenological Research* Vol. 78, No. 1 (January 2009): 2-16.

Boltz, Marilyn G., Brittany Ebendorf, and Benjamin Field. “Audiovisual Interactions: the Impact of Visual Information on Music Perception and Memory” *Music Perception: An Interdisciplinary Journal,* Vol. 26, No. 1 (September, 2009): 43-59.

Bouënard, Alexandre, Marcelo M. Wanderley, Sylvie Gibert, and Fabrice Marandola. “Virtual Gesture Control and Synthesis of Music Performances: Qualitative Evaluation of Synthesized Timpani Exercises” *Computer Music Journal,* Vol. 35, No. 3 (Fall, 2011): 57-72.

Cadoz, Claude. “Instrumental Gesture and Music Composition.” In *Proceeding of the International Computer Music Conference 1988,* (San Francisco, CA: ICMA, 1988): 1-12.

Calvert, Gemma A., Charles Spence, and Barry E. Stein, eds. *The Handbook of Multisensory Processes.* Cambridge, MA: A Bradford Book, 2004.

Camurri, Antonio, and Gualtiero Volpe. “Multimodal Analysis of Expressive Gesture in Music Performance” In *Musical Robots and Interactive Multimodal Systems,* edited by Jorge Solis, and Kia Ng, 47-66, Chennai, India: Springer-Verlag Berlin Heidelberg, 2011.

Carroll, Noël. *A Philosophy of Mass Art.* Oxford: Oxford University Press, 1998.

Chion, Michel. *Audio-Vision: Sound on Screen.* edited and trans. by Claudia Gorbman, New York, NY: Columbia University Press., 1994.

Cook, Nicholas. “Beyond the Notes” *Nature* Vol. 453, No. 26 (June, 2008), 1186-87.

Cooper, Matthew, Jonathan Foote, Elias Pampalk, and George Tzanetakis. “Visualization in Audio-Based Music Information Retrieval” *Computer Music Journal,* Vol. 30, No. 2 (Summer, 2006): 42-62.

Crane, Diana. *The Transformation of the Avant-Garde: The New York Art World, 1940-1985.* Chicago, US:University of Chicago, 1987.

Dahl, Sofia, and Anders Friberg. “Visual Perception of Expressiveness in Musicias’ Body Movements” *Music Perception: An Interdisciplinary Journal,* Vol. 24, No. 5 (June, 2007): 433-54.

Danto, Arthur C.‘‘The Artworld,’’ *Journal of Philosophy* 61 (1964), 571–84.

Davidson, Jane W. “Visual Perception of Performance Manner in the Movements of Solo Musicians.” *Psychology of music* 21 (1993): 103-113.

Davidson, Jane W. “What Does the Visual Information Contained in Music Performances Offer the Observer? Some Preliminary Thoughts” *Music and Mind Machine: The Psychophysiology and Psychopathology of the Sense of Music,* 105-13, edited by Reinhard Steinberg, Berlin, Germany: Springer-Verlag, 1995.

Davidson, Jane W. “Bodily Movement and Facial Actions in Expressive Musical Performance by Solo and Duo Instrumentalists: Two Distinctive Case Studies” *Psychology of Music,* Vol. 40, No. 5 (2012): 595-633.

Delalande, F. “Human Movement and the Interpretation of Music.” presented at *the Second International Colloquium on the Psychology of Music*, Ravello, Italy (1990).

DeWitt, Tom. “Visual Music: Searching for an Aesthetic” *Leonardo, Special Issue: Visual Art, Sound, Music and Technology,* Vol. 20, No. 2, (1987): 115-22.

Fastl, Hugo and Eberhard Zwicker. *Psychoacoustic: Facts and Models.* Berlin, Germany: Springer-Verlag, 2007.

Geringer, John M., Jane W. Cassidy, and James L. Byo. “Effects of Music with Video on Responses of Nonmusic Majors: An Exploratory Study” *Journal of Research in Music Education,* Vol. 44, No. 3 (Autumn, 1996): 240-51.

Greckel, Wil. “Visualization in the Performance of Classical Music: A New Challenge” *The Quarterly,* Vol. 3, No. 4 (March, 2012): 38-48.

Higgins, Dick. “Statement on Intermedia” *In the Spirit of Fluxus,* edited by Joan Rothfuss and Elizabeth Armstrong, Minneapolis: Walker Art Center, 1993.

Hosken, Dan. *An Introduction to Music Technology.* New York, NY: Routledge. 2011.

Iwamiya, Shin-ichiro, Y. Sugano, and K. Kouda. “The Effects of Synchronization of Temporal Structures of Sound and Motion Picture on the Impression of Audio-visual Contents.” In *Proceedings of the 2000 IEEE International Conference on Systems, Man, and Cybernetics,* 1222-225*,* Nashville, 2000.

Jensenius, Alexander R. “Some Video Abstraction Techniques for Displaying Body Movement in Analysis and Performance” *Leonardo,* Vol. 46, No. 1 (2013): 53-60.

Johnston, James D. “Transform Coding of Audio Signals Using Perceptual Noise Criteria” *IEEE Journal on Selected Areas in Communications.* Vol. 6, No. 2 (Feb. 1988): 314-23.

Jones, Randy, and Ben Nevile. “Creating Visual Music in Jitter: Approaches and Techniques” *Computer Music Journal,* Vol. 29, No. 4 (Winter, 2005): 55-70.

Katkere, Arun, et. al. “Multiple Prospective Interactive Video” In *Proceeding of the International Conference on Multimedia Computing and Systems* (1995): 202-11.

Lam, Margaret. “Beyond the Score: Music Visualization and Digital Humanities” *Bulletin of American Society for Information Science and Technology,* Vol. 38, No. 4 (April/May, 2012): 45-47.

Langner, Jörg, and Werner Goebl. “Visualizing Expressive Performance in Tempo-Loudness Space” *Computer Music Journal,* Vol. 27, No. 4 (Winter, 2003): 69-83.

Lorenz, Haik, Matthias Trapp, Jürgen Döllner. “Interactive Multi-Perspective Views of Virtual 3D Landscape and City Models” *In Proceedings of AGILE Conference* (2008), 301-21.

Lovejoy, Margot, Christiane Paul, and Victoria Vesna. eds. *Context Providers: Conditions of Meaning in Media Arts.* Bristol, UK: Intellect. 2011.

Manovich, Lev. “What is Visualisation?” *Visual Studies,* Vol. 26, No. 1 (March, 2011): 36-49.

Maras, Steven. “Convergence/divergence: Ten Definition of Multimedia.” *Metro Magazine: Media& Education Magazine* no.108 (1996): 9-13.

Marinetti, Filippo T., Bruno Corra, Emilio Settimelli, Arnaldo Ginna, Gicacomo Balla, and Remo Chiti. "The Futurist Cinema." *An Futurism: Anthology*, edited by Lawrence Rainey, Christine Poggi and Wittman Laura, New Haven: Yale University Press, 2009.

Marks, Lawrence E. “On Cross-Modal Similarity: Perceptional Structure in Pitch, Loudness, and Brightness.” *Journal of Experimental Psychology: Human Perception and Performance*, 15, (1989): 586-602.

Marks, Lawrence E. “Cross-Modal Interactions in Speeded Classification.” *Handbook of Multisensory Processes*, edited by Gemma Calvert, Charles Spence, and Barry E. Stein, 85-102, US: MIT Press, 2004.

Martino, Gail, and Lawrence E. Marks. “Synesthesia: Strong and Weak” *Current Directions in Psychological Science,* Vol. 10, No. 2 (April, 2001): 61-65.

McAdams, Stephen, and Bruno L. Giordano. “The Perception of Musical Timbre” in *Oxford Handbook of Music Psychology,* edited by Sussan Hallam, Ian Cross, and Michael Thaut, Oxford, New York: Oxford University Press. 2009.

Metzler, Bob. *Audio Measurement Handbook.* Beaverton, Oregon: Audio Precision, Inc.,1993.

Meyer, Richard E. *Multimedia Learning*. 2nd ed., New York, NY: Cambridge University Press. 2009.

Miranda, Eduardo and Marcelo M. Wanderley. *New Digital Musical Instrument: Control and Interaction beyond the keyboard.* Middleton, WI: A-R Editions, Inc., 2006.

Moholy-Nagy, László. “ ‘Theater, Circus, Variety,’ Theater of the Bauhaus” *Multimedia: Wagner to the Virtual Reality,* edited by Randall Packer and Ken Jordan, Chapter 2, New York, NY: W.W. Norton & Company Ltd., 2001.

Palombini, Carlos. “Pierre Schaeffer, 1953: Towards an Experimental Music.” *Music & Letter* vol. 73, no. 4 (Nov. 1993): 542-57.

Parncutt, Richard, Gary E. McPherson, eds. *The Science & Psychology of Music Performance: Creative Strategies for Teaching and Learning.* New York, NY: Oxford University Press. 2002.

Peters, Deniz, “Touch: Real, Apparent, and Absent: On Bodily Expression in Electronic Music.” In *Bodily Expression in Electronic Music,* edited by Deniz Peters, Gerhard Eckel, and Andreas Dorscel, Chapter 1, New York, NY: Routledge Taylor&Francis Group, 2012

Platz, Friedrich, and Reinhard Kopiez. “When the Eye Listens: A Meta-analysis of How Audio-visual Presentation Enhances the Appreciation of Music Performance” *Music Perception: An Interdisciplinary Journal,* Vol. 30, No. 1 (September, 2012): 71-83.

Poynting, John H., and Joseph J. Thomson. *Text-Book of Physics: Sound. Vol. II, 3rd eds.* London: Charles Griffin & Co Ltd., 1904.

Pressnitzer, Daniel, Stephen McAdams, Suzanne Winsberg, and Joshua Fineberg. “Perception of Musical Tension for Nontonal Orchestral Timbres and its Relation to Psychoacoustic Roughness” *Perception & Psychophysics,* Vol. 62, No. 1 (2000): 66-80.

Raichel, Daniel R., *The Science and Applications of Acoustics.* New York, NY: Springer Science-Business Media Inc., 2006.

Rainey, Lawrence, Christine Poggi, and Laura Wittman. *Futurism: An Anthology.* New Haven, CT: Yale University,  2009.

Roads, Curtis. *Microsound*, Cambridge, MA: MIT Press., 2001.

Rodger, Matthew W.M., Cathy M. Craig, and Sile O’Modhrain, “Expertise is Perceived from Both Sound and Body Movement in Musical Performance.” *Human Movement Science* 31 (2012): 1137-50.

Schacher, Jan C., and Philippe Kocher, “Ambisonics Spatialization Tools for Max/MSP.” In *Proceeding of International Computer Music Conference 2006,* 274-77, New Orleans, USA: ICMC, 2006. http://hdl.handle.net/2027/spo.bbp2372.2006.057 (accessed April 25, 2014).

Schaeffer, Pierre. “Introduction à la musique concrète” *Polyphonie*,vi (1950): 48-50.

Schumann, Robert. *On Music and Musicians.* edited by Konard Wolff, trans. by Paul Rosenfeld, Berkeley, US: University of California Press. 1946.

Scruton, Roger. *Aesthetic of Music,* New York: Oxford University Press. 1997.

Sedes, Anne, Benoît Courribet, and Jean-Bapiste Thiébaut. “From the Visualization of Sound to Real-time Sonification: Different Prototypes in the Max/MSP/Jitter Environment” In *Proceedings of International Computer Music Conference 2004,* Miami, US: ICMA, 2004. http://hdl.handle.net/2027/spo.bbp2372.2004.081 (accessed April 25, 2014).

Seller, Graham, Richard S. Wright, Jr., and Nicholas Heamel, *OpenGl SuperBible: Comprehensive Tutorial and Reference.* 6th eds. Upper Saddle River, NJ: Addison-Wesley, 2014.

Serig, Daniel. “A Conceptual Structure of Visual Metaphor” *Studies in Art Education,* Vol. 47, No. 3 (Spring, 2006): 229-47.

Slawson, Brian. “Interactive Multimedia: The Gestalt of a Gigabyte.” *Art Education*, Vol. 46, No. 6 (Nov. 1993): 15-22.

Tan, Siu-Lan, Annabel J. Cohen, Scott D. Lipscomb, and Roger a. Kendall. *The Psychology of Music in Multimedia.* Oxford: Oxford University Press. 2013.

van Eijk, R. L., A. Kohlrausch, J. F. Juola, and S. van de Par.“Temporal Interval Discrimination Thresholds Depend on Perceived Synchrony for Audio-visual Stimulus Pairs.” *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 35, No. 4 (2009): 1254-63.

Vines, Bradley W., Carol L. Krumhansl, Marcelo M. Wanderley, Ioana M. Dalca, and Daniel J. Levitin. “Music to My Eyes: Cross-modal Interactions in the Perception of Emotions in Musical Performance.” *Cognition,* 118 (2011): 157-70.

Wagner, Richard. “ ‘Outlines of the Artwork of the Future,’ The Artwork of the Future.” *Multimedia: From Wagner to Virtual Reality,* Chapter 1, Edited by Randal Packer and Ken Jordan, New York, NY: W.W. Norton & Company Ltd., 2001.

Wood, Albert B.. *Textbook of Sound: Being an Account of the Physics of Vibrations with Special Reference to Recent Theoretical and Technical Developments.* New York, NY: The Macmillan Company, 1930.

Wright, Richard S. Jr., and Michael Sweet, *OpenGL Superbible: The Complete Guide to Opengl Programming for Window NT and Window 95.* United States: Waite Group Press., 1996.

Zicarelli, David. “An Extensible Real-Time Signal Processing Environment for Max” In *Proceedings of the International Computer Music Conference,* San Francisco: ICMA, (1998).

# Appendix A : Open Source Software

Performance-Oriented-Audio-Based-Visualization

https://github.com/cabasa0128/Performance-Oriented-Audio-Based-Visualization.git

Tap.Tools by ElectroTap

http://74objects.com/taptools/

https://github.com/tap/TapTools

SoundHack plus-ins by Tom Erbe

http://www.soundhack.com/freeware/

newverb2~ by CNMAT in UC Berkeley

http://cnmat.berkeley.edu/downloads

noisiness~ by Tristan Jehan

http://web.media.mit.edu/~tristan/maxmsp.html

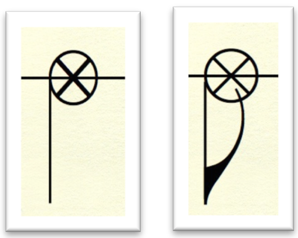
Ambisonics tools by ICST

http://www.icst.net/research/projects/ambisonics-tools/

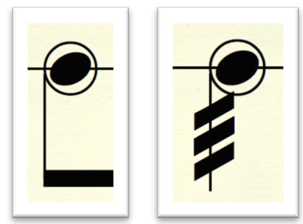
# Piece for Solo Snare Drum and Electronics

*composed by Yemin Oh*

## Notations



Rimshot ( One hand )



Hitting the hoop(rim)

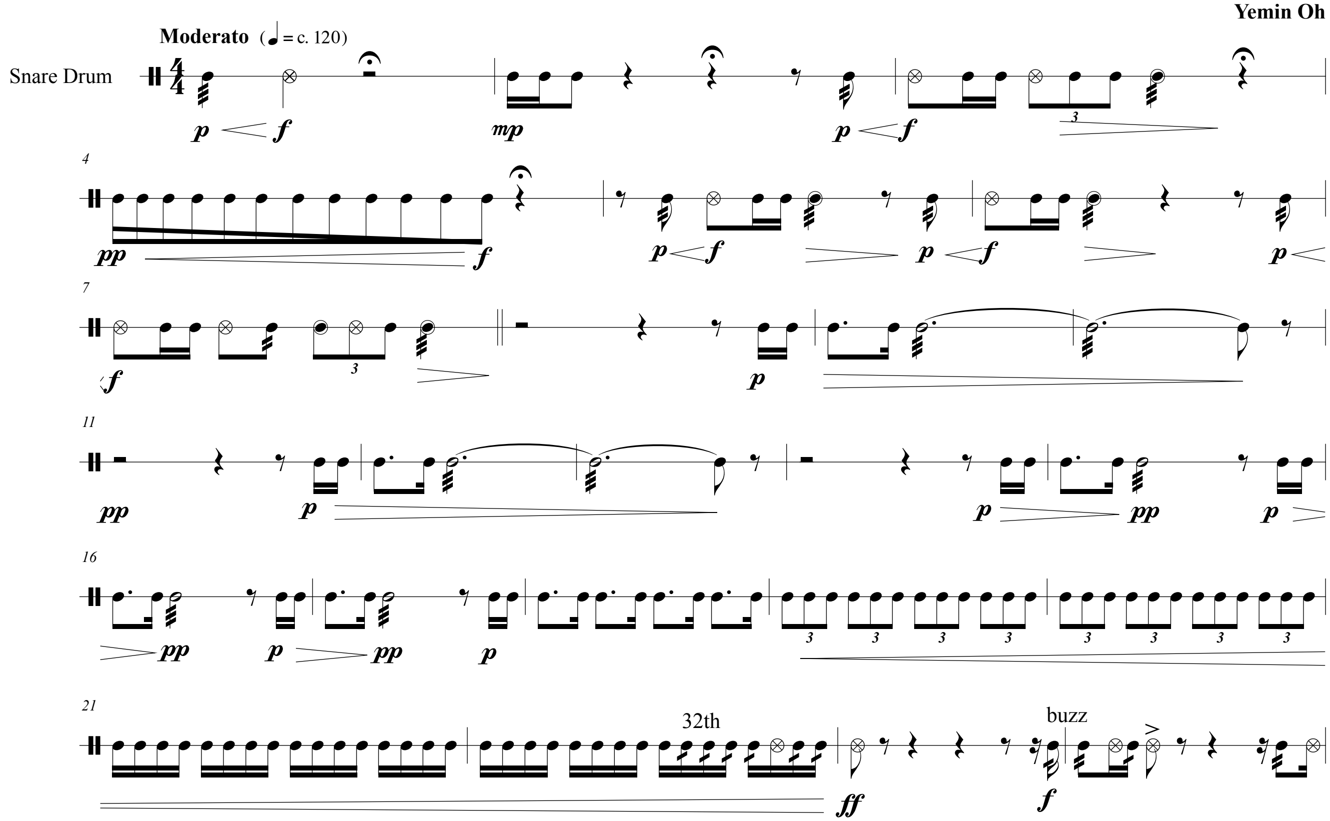
## Program note

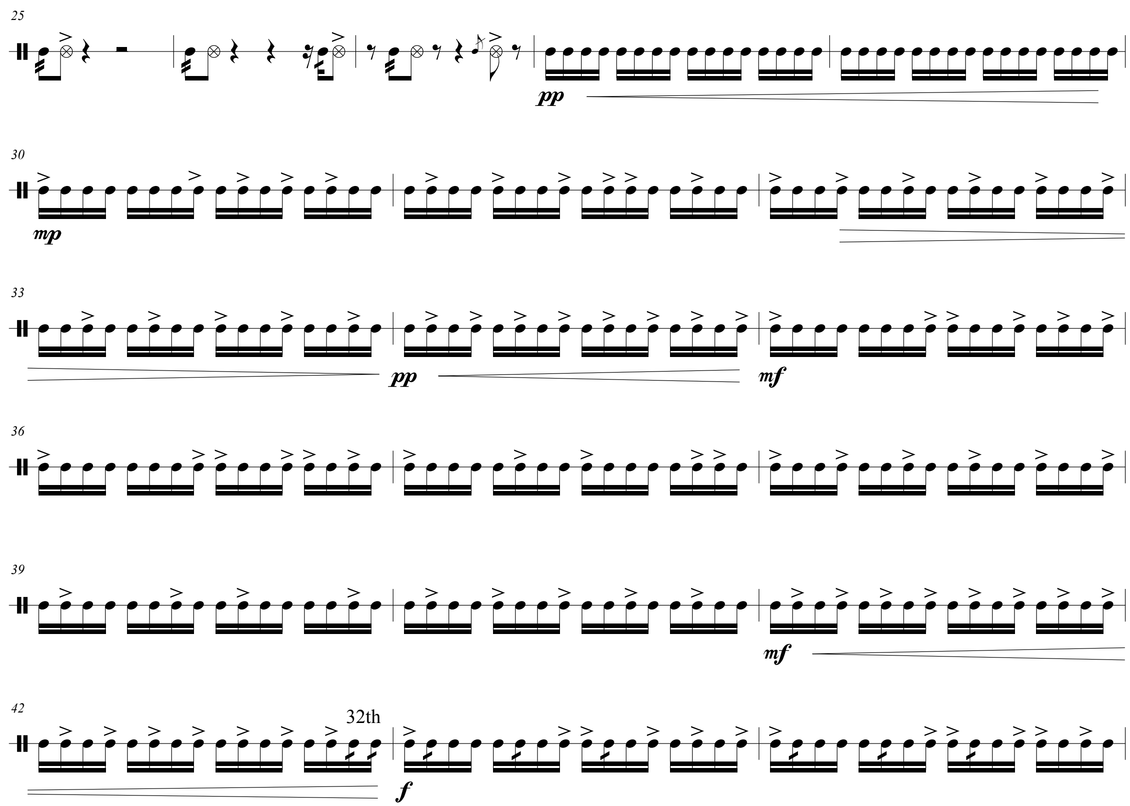
Several physicists have hypothesized existence of an extra dimension of this universe, and speculative theories about dimension of space have been introduced. The endeavor of physicists inspired me to think about an extra dimension of music. In this piece, I tried to create an additional dimension that might be conceptually abstract using a live video camera and four channel sounds as the extra dimensions. Effects of the live video could produce another potential dimension.

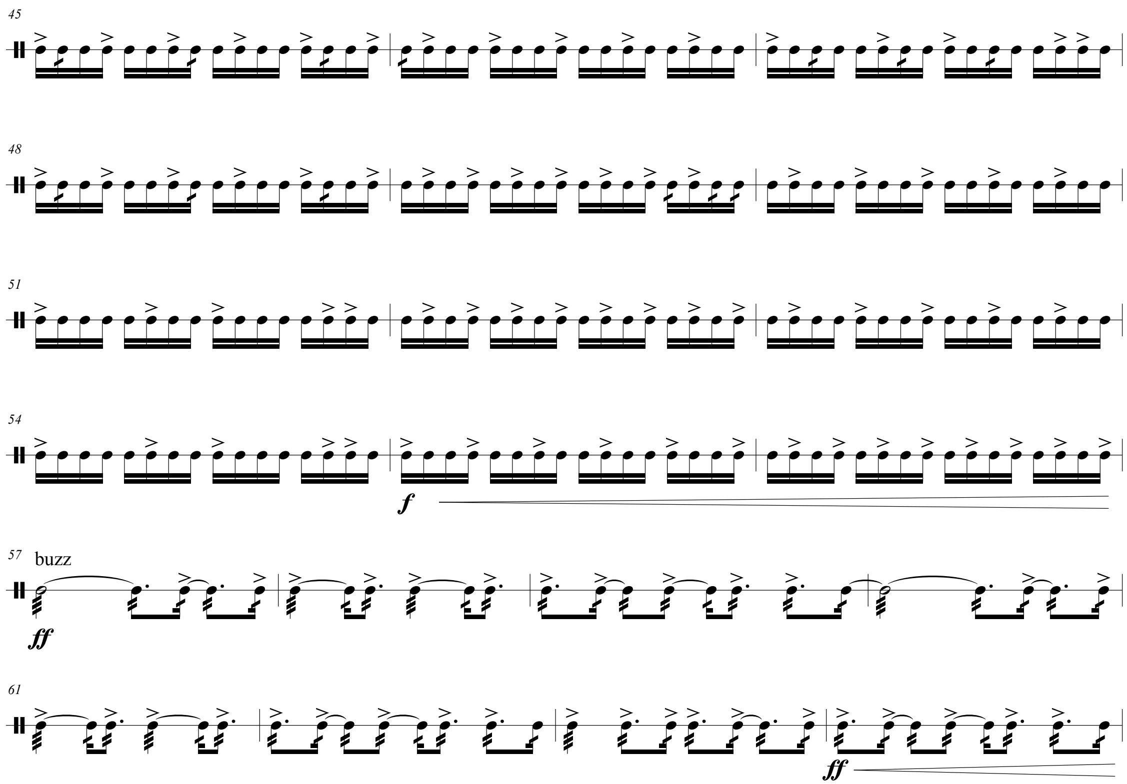
# Piece for Snare Drum and Live Electronics

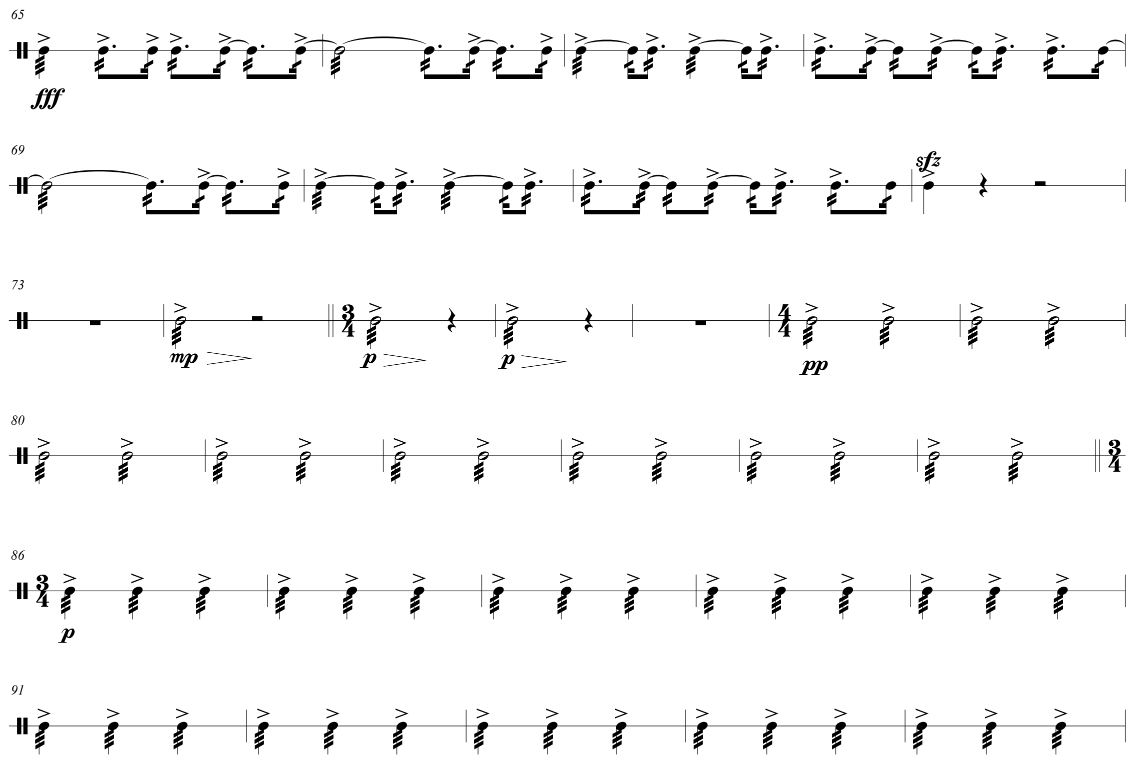
Score

**Yemin Oh**

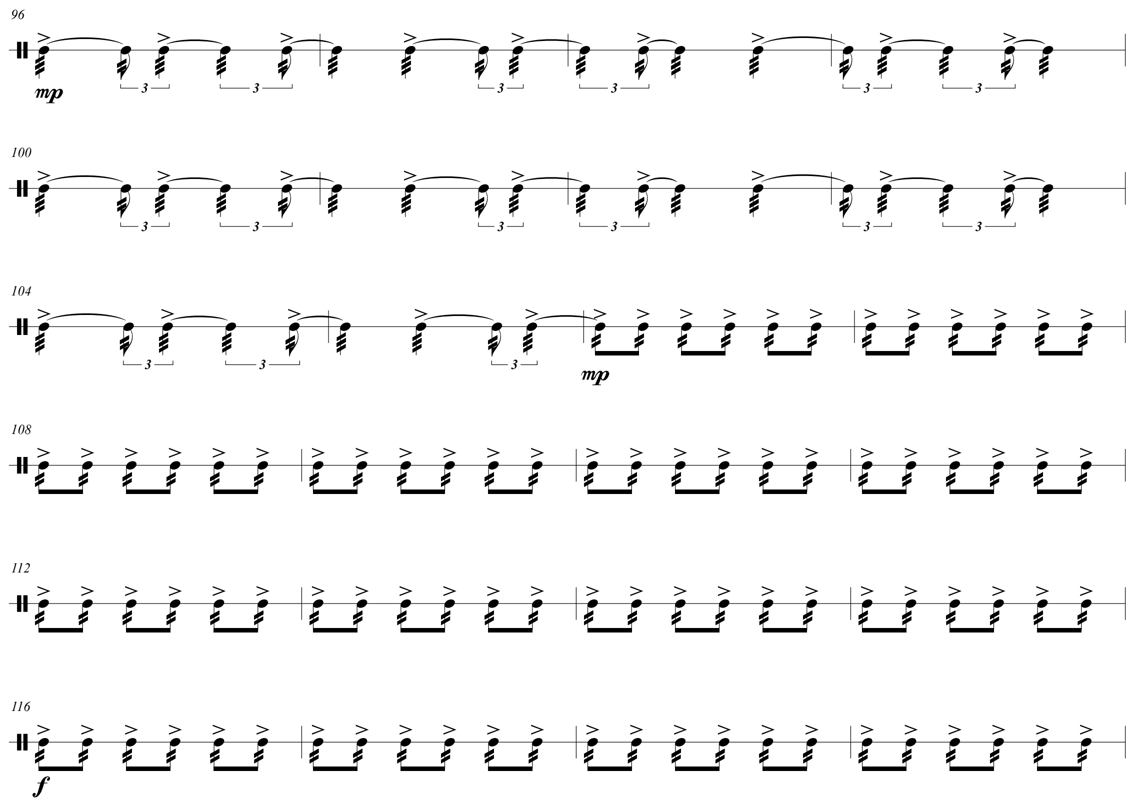


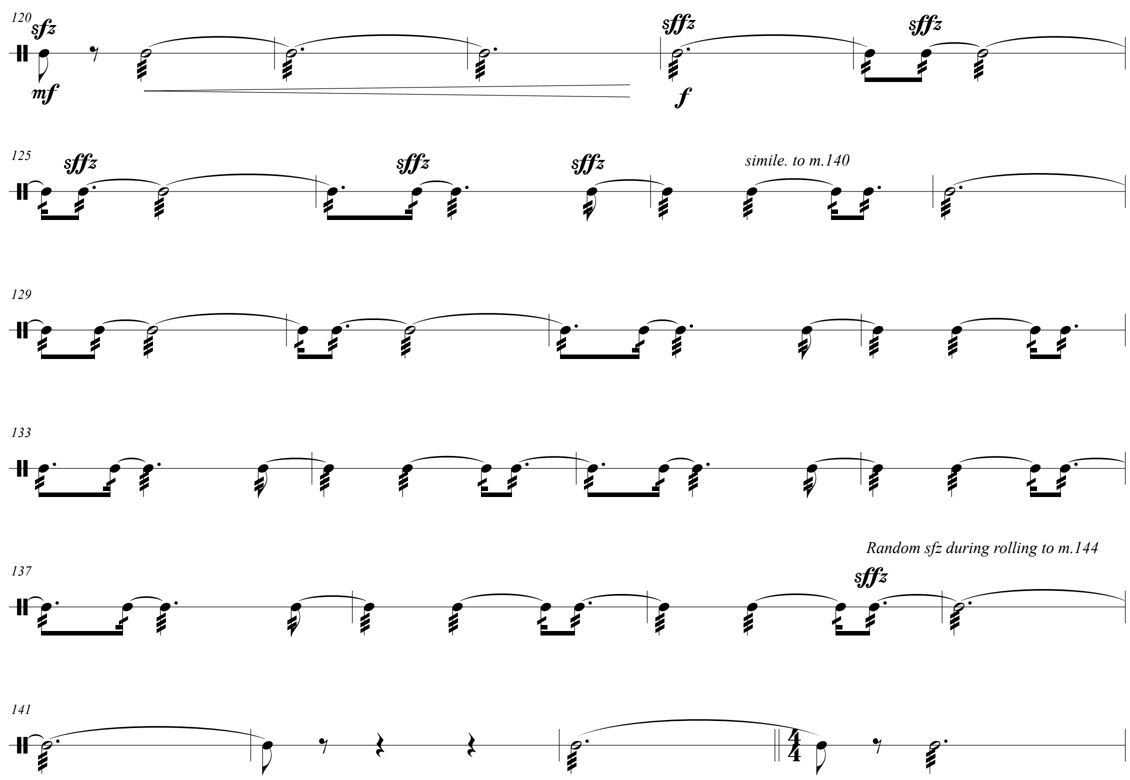

25
pp
30
mp
33
pp
mf
36
39
mf
42
32th
f

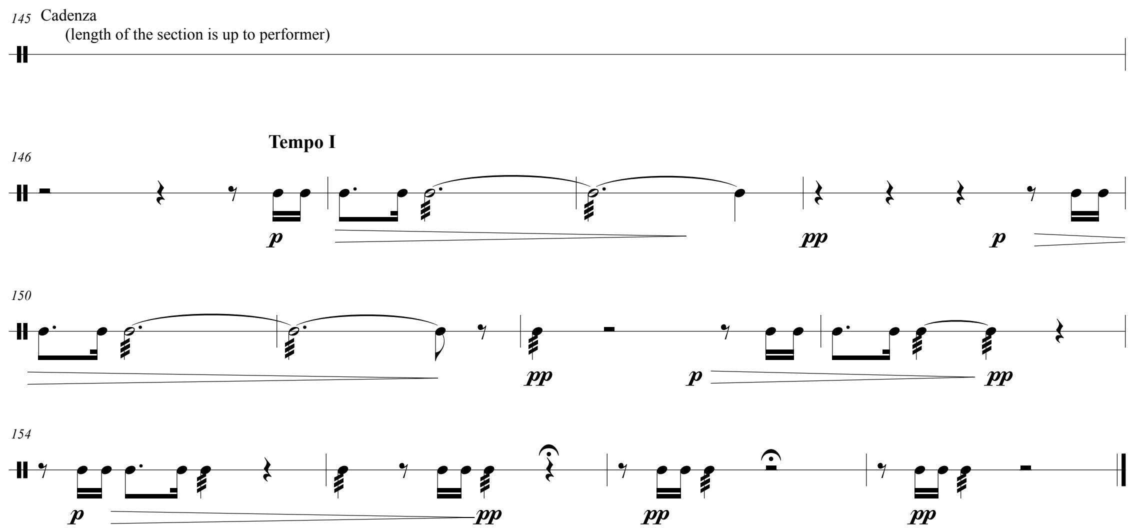

45
48
51
54
f
57
buzz
ff
61
ff


65
fff
69
sfz
73
mp
p
p
pp
80
86
p
91

Piece for Snare Drum and Live Electronics




120
sfz
mf
sffz
f
sffz
125
sffz
sffz
sffz
simile. to m.140
129
133
137
Random sfz during rolling to m.144
sffz
141
4
4


145
Cadenza
(length of the section is up to performer)
146
Tempo I
p
pp
p
150
pp
p
pp
154
p
pp
pp
pp

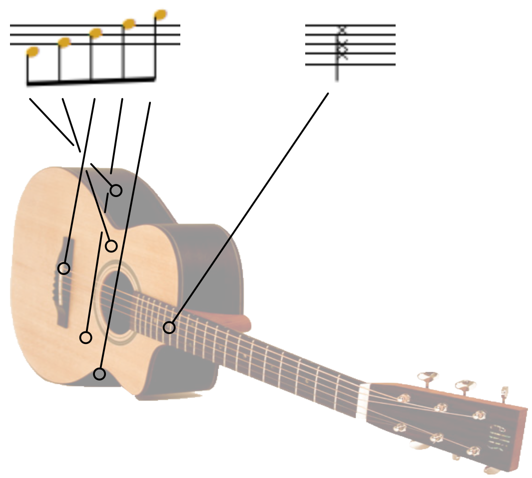
# MEMORIAM

*composed by Yemin Oh*

# Notation

Hit the body with hand

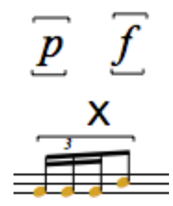
Hit the strings with hand



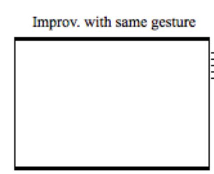
Glissando without plucking


Hold the position without plucking

Combine one from (*p*, *f*) and the patten by performer's randomized choice



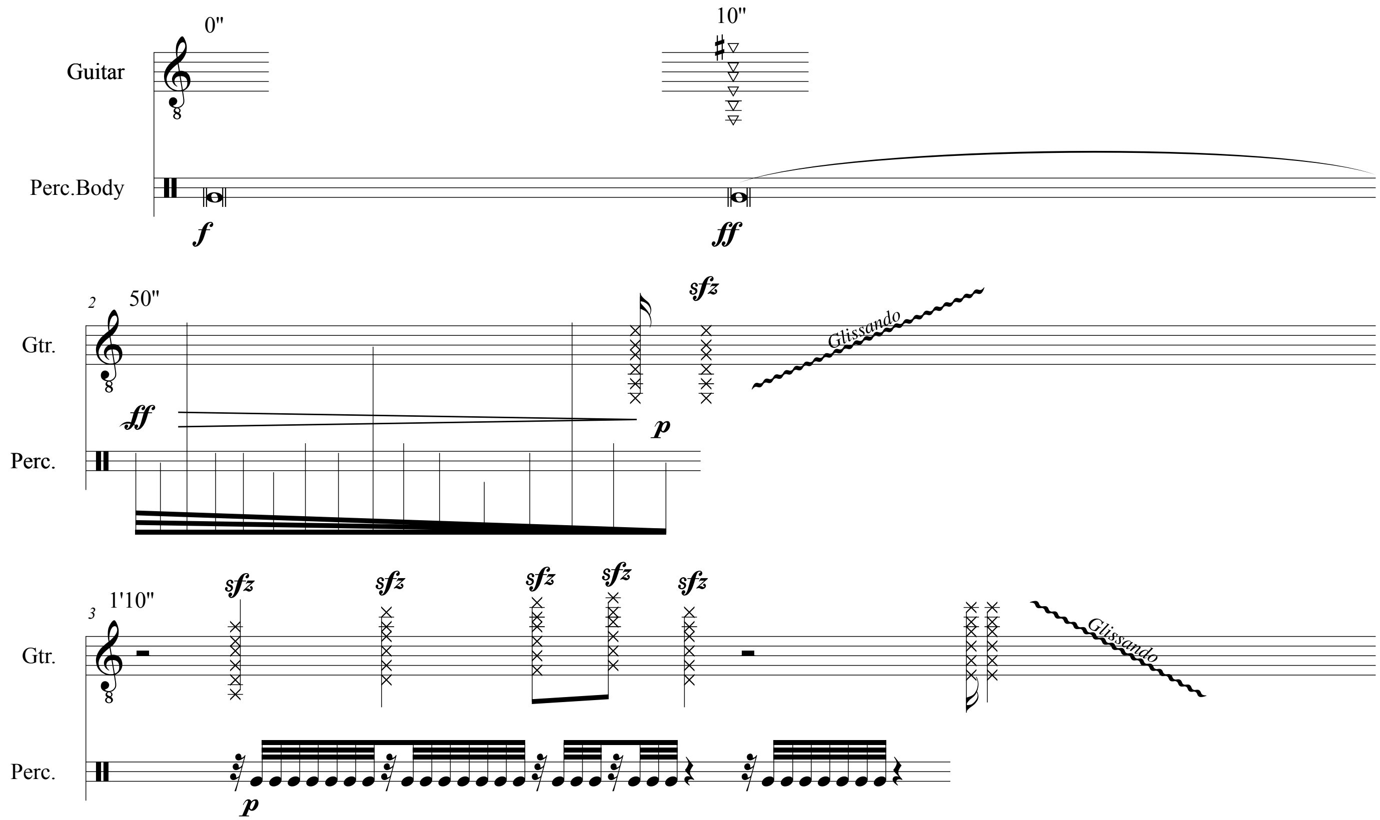

Improvising previous gesture

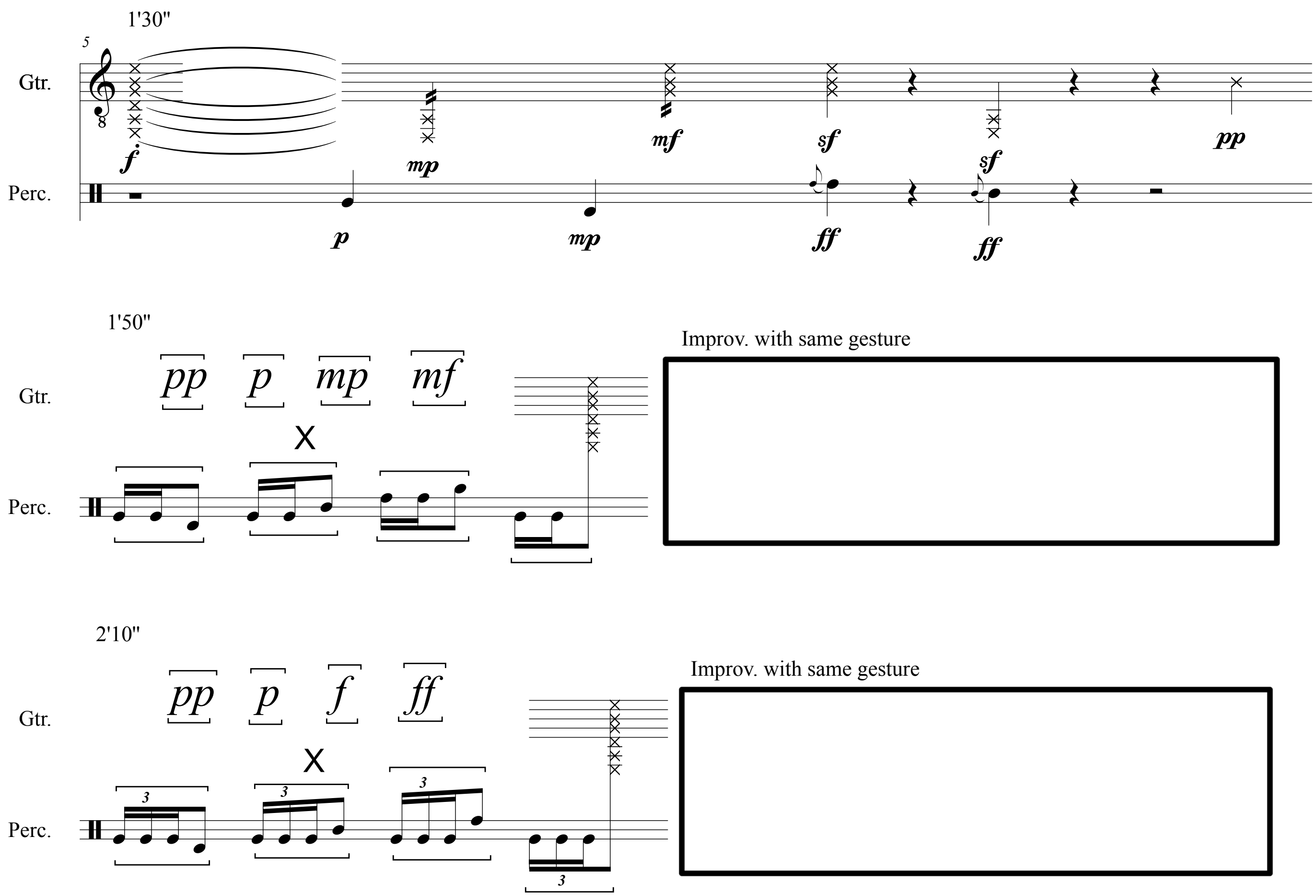


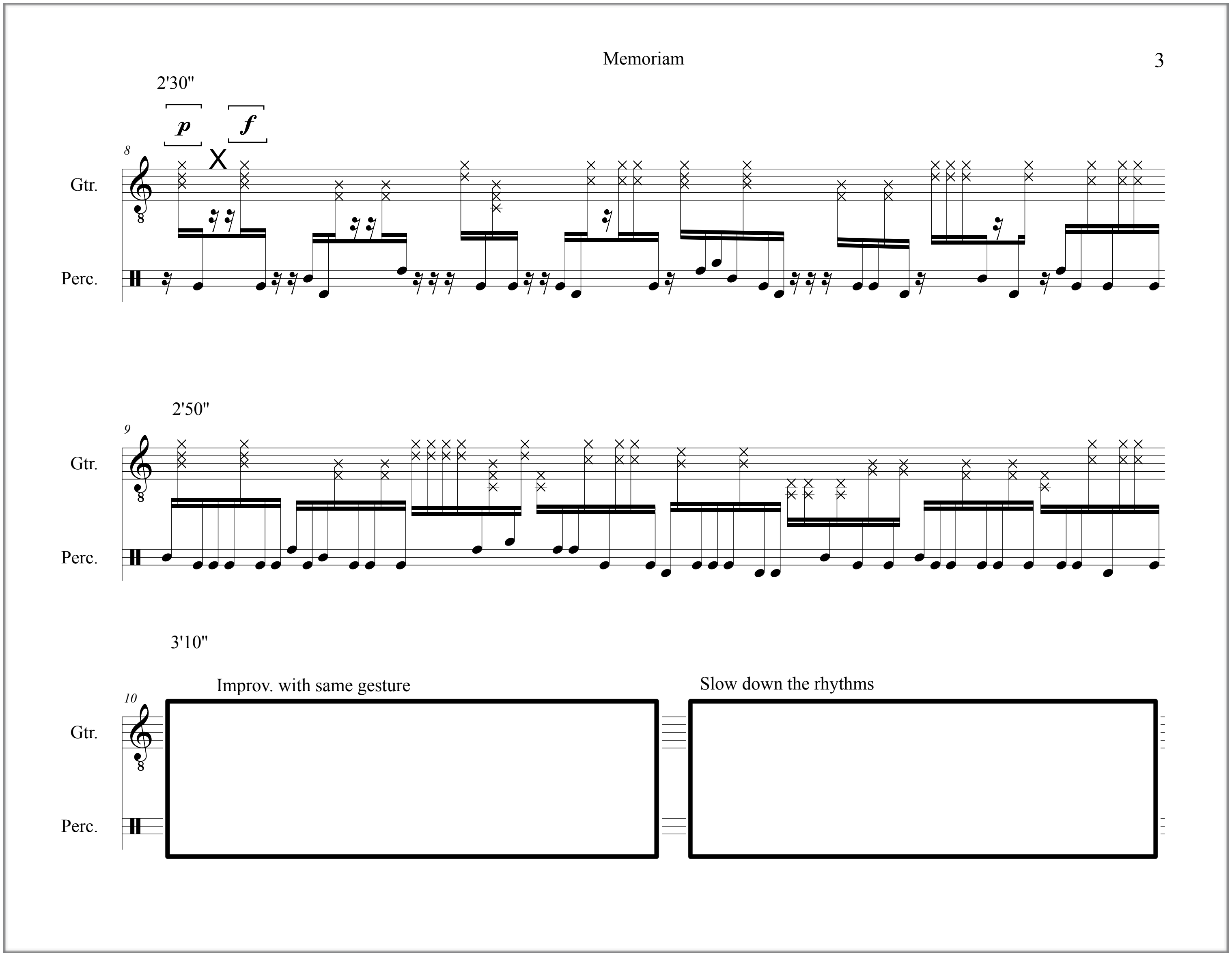
Score

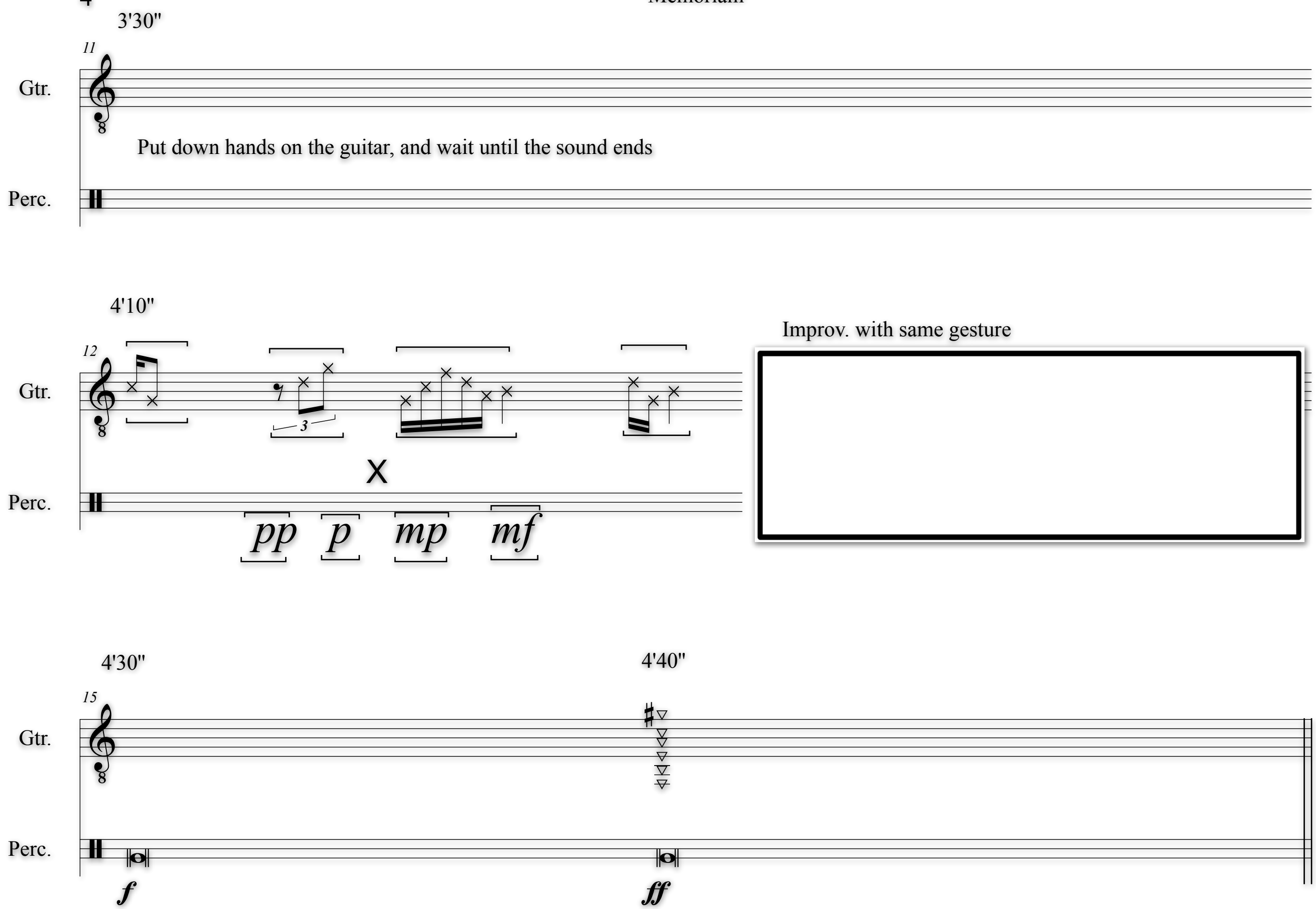
# Memoriam

Yemin Oh



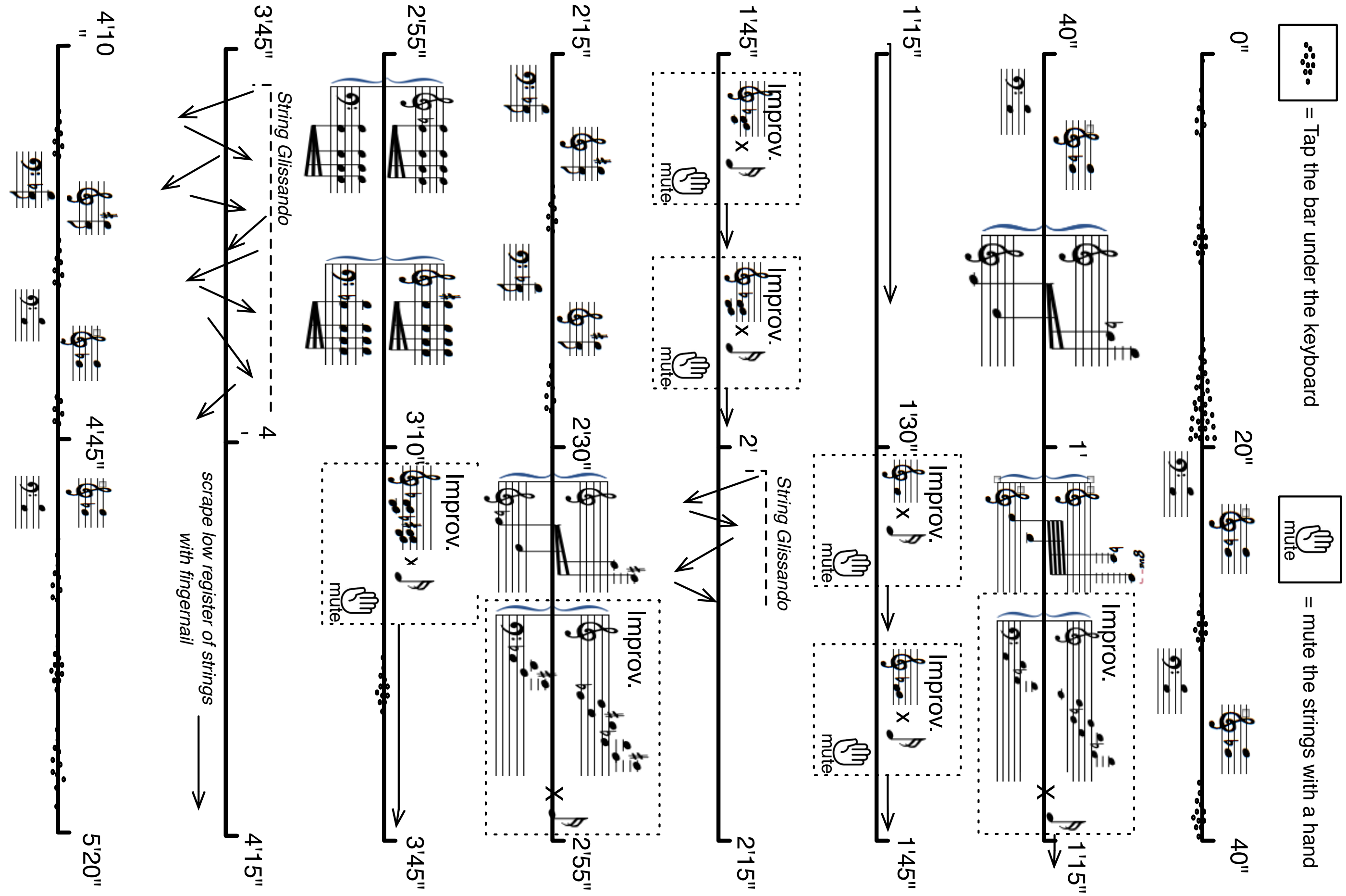

1'30"
Gtr.
Perc.
f
mp
mf
sf
sf
pp
p
mp
ff
ff
1'50"
pp
p
mp
mf
Improv. with same gesture
2'10"
pp
p
f
ff
Improv. with same gesture


Memoriam
3
2'30"
p
f
8
Gtr.
Perc.
2'50"
9
Gtr.
Perc.
3'10"
Improv. with same gesture
Slow down the rhythms
10
Gtr.
Perc.


3'30"
11
Gtr.
Put down hands on the guitar, and wait until the sound ends
Perc.
4'10"
12
Gtr.
Improv. with same gesture
Perc.
pp
p
mp
mf
4'30"
4'40"
15
Gtr.
Perc.
f
ff

# Synesthetic Moment

Composed by Yemin Oh



# Survey

for

## Audio-Based Visualization of Expressive Body Movement in Music Performance

1. *Piece for Solo Snare Drum and Electronics*
2. *Memoriam*
3. *Synesthetic Moment*

composed by Yemin Oh

Please read through the following question sheet
before we begin the performance.

After each piece, fill out the following questionnaire.

This information will be used for dissertation research.

## Information of Participant

(1) Age?

| 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | above 60 |
|---|---|---|---|---|---|
| | | | | | |

(2) Gender?

| Female | Male |
|---|---|
| | |

(3) For how many years have you spent significant time studying music or learning to play a musical instrument?

| None | Less than a year | 1-2 years | 2-4 years | More than 4 years |
|---|---|---|---|---|
| | | | | |

(4) What would you grade your musical skill?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(no skill, no interest) (professional or majored)

(5)How frequently do you go to music concert?

| Every week | Every month | Every 1-6 months | Every 6-12months | Every 1-2year |
|---|---|---|---|---|
| | | | | |

(6) How many times have you attended an avant-garde or experimental music concert before?

| None | 1 | 2-5 | 6-10 | over 10 |
|---|---|---|---|---|
| | | | | |

(7) How may times have you attended an electroacoustic music concert in the past?

| None | 1 | 2-5 | 6-10 | over 10 |
|---|---|---|---|---|
| | | | | |

## “Piece for Solo Snare Drum and Electronics.”

Instrumentation : Snare Drum

**- VISUALS**

(1) How closely do you feel that the loudness of the sound is correlated with the visualizations?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Uncorrelated) (Neutral) (Strongly correlated)

(2) How closely do you feel that the noisiness of the sound is correlated with the visualizations? (Note: For instance, the sound of static is very noisy, whereas the sound of a bell is not noisy at all.)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Uncorrelated) (Neutral) (Strongly correlated)

(3) Did you find that the visualizations help you appreciate the music or distract your attention?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Very Distracting) (Neutral) (Very Helpful)

**- BODY MOVEMENTS**

(4) How obvious is the connection between visualization and body movement?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Not connected) (abstract yet still connected) (Direct connection)

(5) To what extent did the visualizations amplify the expressiveness of the performer’s body movements?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Not at all) (Neutral) (To a great extent)

(6)If you have any other thoughts or opinions on this research, write on back of this page.

## “Memoriam”

Instrumentation : Guitar

**- VISUALS**

(1) How closely do you feel that the loudness of the sound is correlated with the visualizations?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Uncorrelated) (Neutral) (Strongly correlated)

(2) How closely do you feel that the noisiness of the sound is correlated with the visualizations? (Note: For instance, the sound of static is very noisy, whereas the sound of a bell is not noisy at all.)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Uncorrelated) (Neutral) (Strongly correlated)

(3) Did you find that the visualizations help you appreciate the music or distract your attention?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Very Distracting) (Neutral) (Very Helpful)

**- BODY MOVEMENTS**

(4) How obvious is the connection between visualization and body movement?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Not connected) (abstract yet still connected) (Direct connection)

(5) To what extent did the visualizations amplify the expressiveness of the performer’s body movements?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Not at all) (Neutral) (To a great extent)

(6)If you have any other thoughts or opinions on this research, write on back of this page.

## "Synesthetic Moment"

Instrumentation : Piano

**- VISUALS**

(1) How closely do you feel that the loudness of the sound is correlated with the visualizations?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Uncorrelated) (Neutral) (Strongly correlated)

(2) How closely do you feel that the noisiness of the sound is correlated with the visualizations? (Note: For instance, the sound of static is very noisy, whereas the sound of a bell is not noisy at all.)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Uncorrelated) (Neutral) (Strongly correlated)

(3) Did you find that the visualizations help you appreciate the music or distract your attention?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Very Distracting) (Neutral) (Very Helpful)

**- BODY MOVEMENTS**

(4) How obvious is the connection between visualization and body movement?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

(Not connected) (abstract yet still connected) (Direct connection)

(5) To what extent did the visualizations amplify the expressiveness of the performer's body movements?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| | | | | | | |

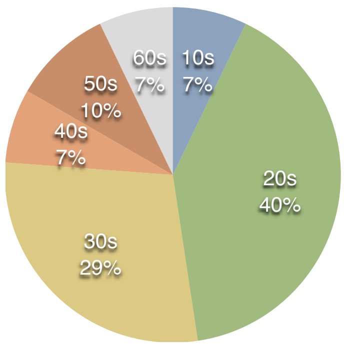
(Not at all) (Neutral) (To a great extent)

(6)If you have any other thoughts or opinions on this research, write on back of this page.

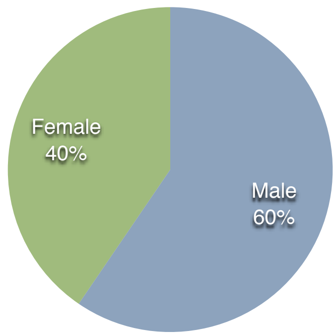
# Appendix D : Tables and Charts of Survey Results

## Background information of Participants

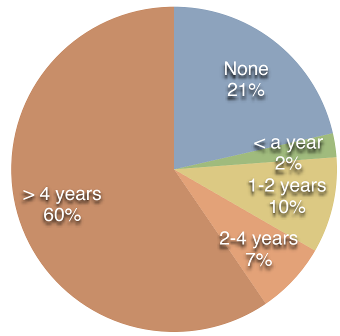
Age



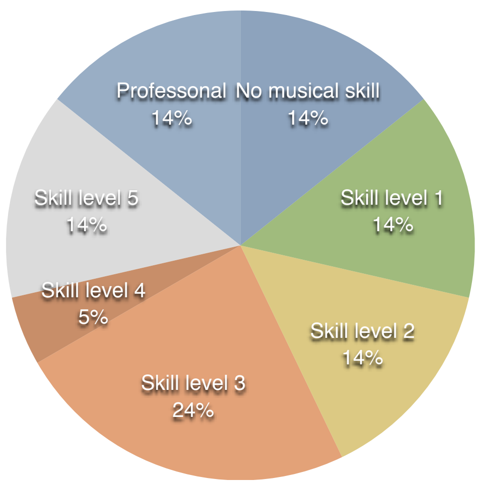

Gender



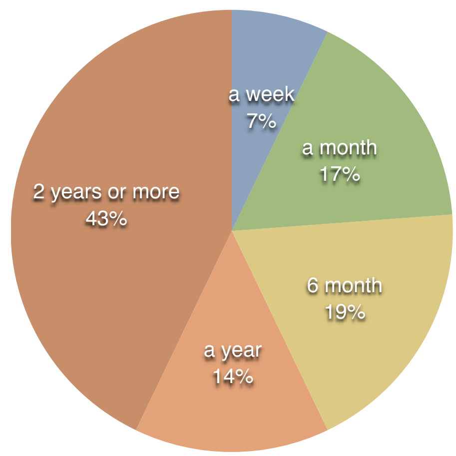

Musical training



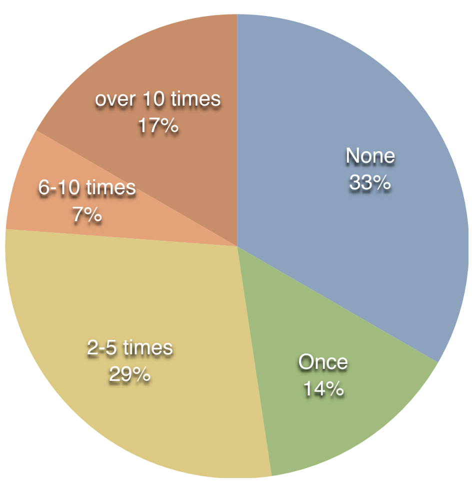

Musical Skill



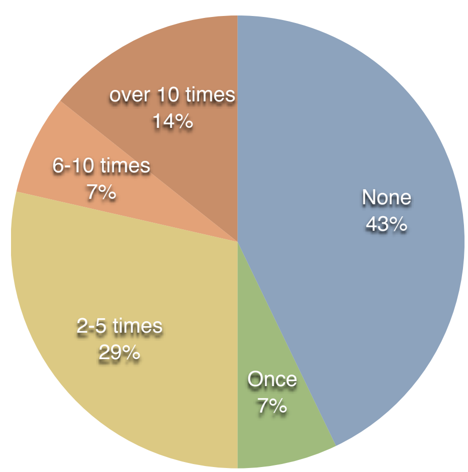

Frequency of attending music concert




Experience with avant-garde music concerts



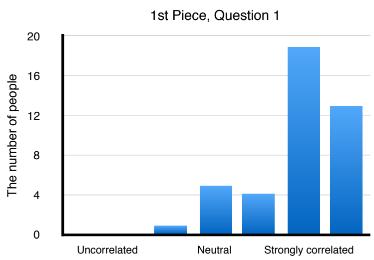

Experience with electroacoustic music concerts

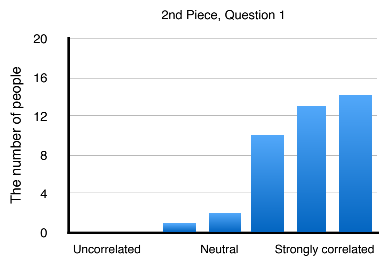


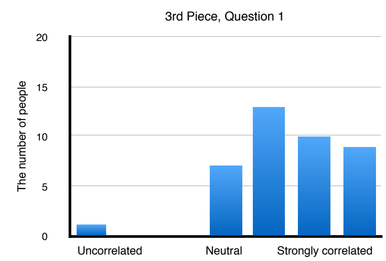
Responses in each questionnaires (One sample t-test)

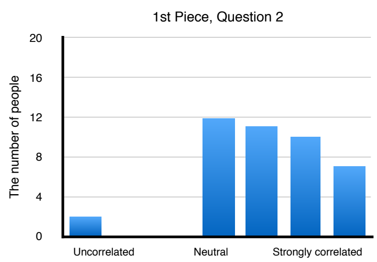
| Piece 1 | Question 1 | Question 2 | Question 3 | Question 4 | Question 5 |
|---|---|---|---|---|---|
| **Mean** | 5.9048 | 5.0952 | 5.2381 | 4.6905 | 5.0952 |
| **SD** | 1.0548 | 1.4109 | 1.6050 | 1.4396 | 1.4109 |
| **SEM** | 0.1628 | 0.2177 | 0.2477 | 0.2221 | 0.2177 |
| **T value (4)** | 11.7026 | 5.0307 | 4.9992 | 3.1083 | 5.0307 |
| **p value (two-tailed)** | 0.0001 | 0.0001 | 0.0001 | 0.0034 | 0.0001 |
| **Piece 2** | **Question 1** | **Question 2** | **Question 3** | **Question 4** | **Question 5** |
| **Mean** | 5.9048 | 5.0952 | 5.2381 | 4.6905 | 5.0952 |
| **SD** | 1.0548 | 1.4109 | 1.6050 | 1.4396 | 1.4109 |
| **SEM** | 0.4205 | 0.3633 | 0.3170 | 0.3243 | 0.3963 |
| **T value (4)** | 12.0547 | 7.0260 | 5.5948 | 3.7010 | 6.9961 |
| **p value (two-tailed)** | 0.0001 | 0.0001 | 0.0001 | 0.0006 | 0.0001 |
| **Piece 3** | **Question 1** | **Question 2** | **Question 3** | **Question 4** | **Question 5** |
| **Mean** | 5.9048 | 5.0952 | 5.2381 | 4.6905 | 5.0952 |
| **SD** | 1.0548 | 1.4109 | 1.6050 | 1.4396 | 1.4109 |
| **SEM** | 0.5109 | 0.4192 | 0.3892 | 0.3707 | 0.4450 |
| **T value (4)** | 7.2499 | 4.2151 | 5.3749 | 3.4986 | 6.9058 |
| **p value (two-tailed)** | 0.0001 | 0.0001 | 0.0001 | 0.0012 | 0.0001 |

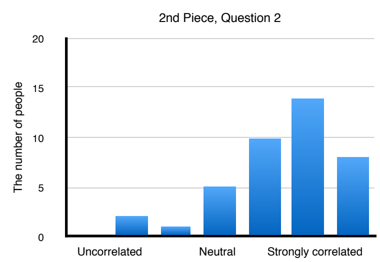
# Charts of Each Questionnaire

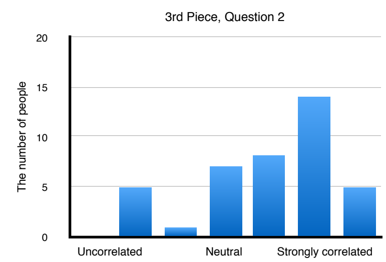


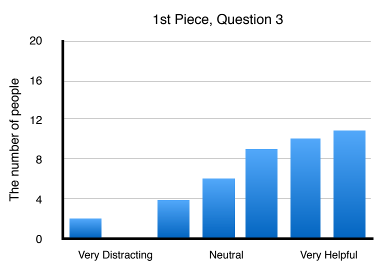



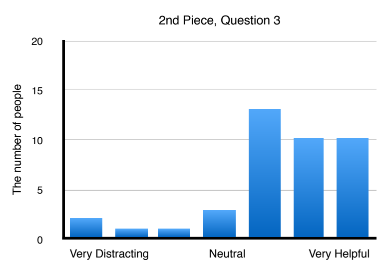



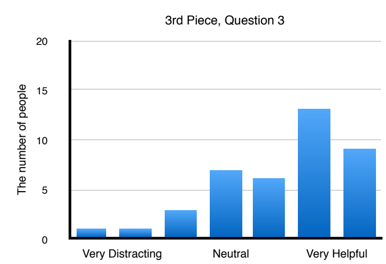



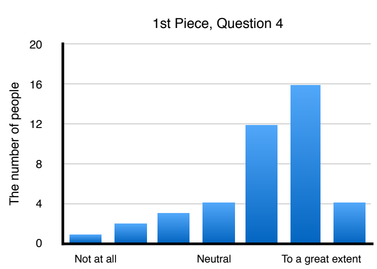



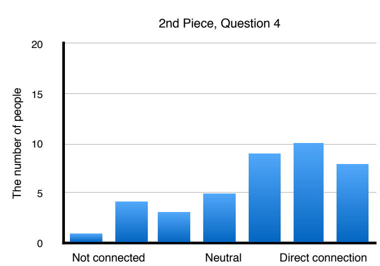



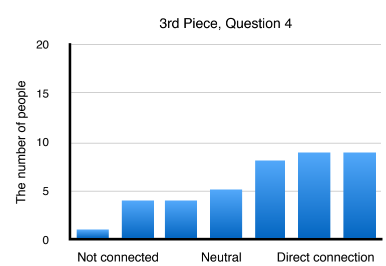



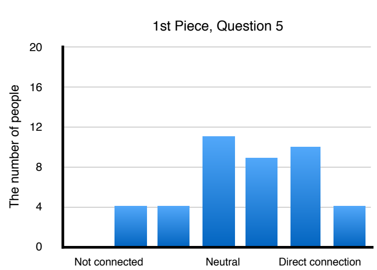



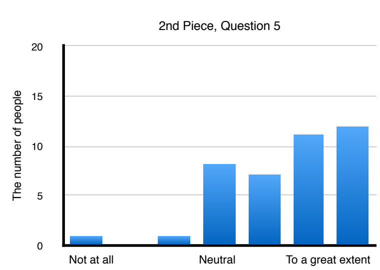



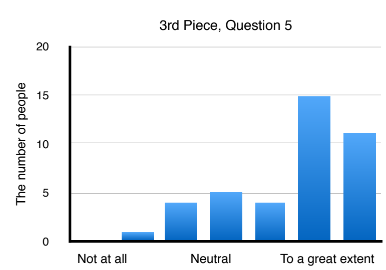



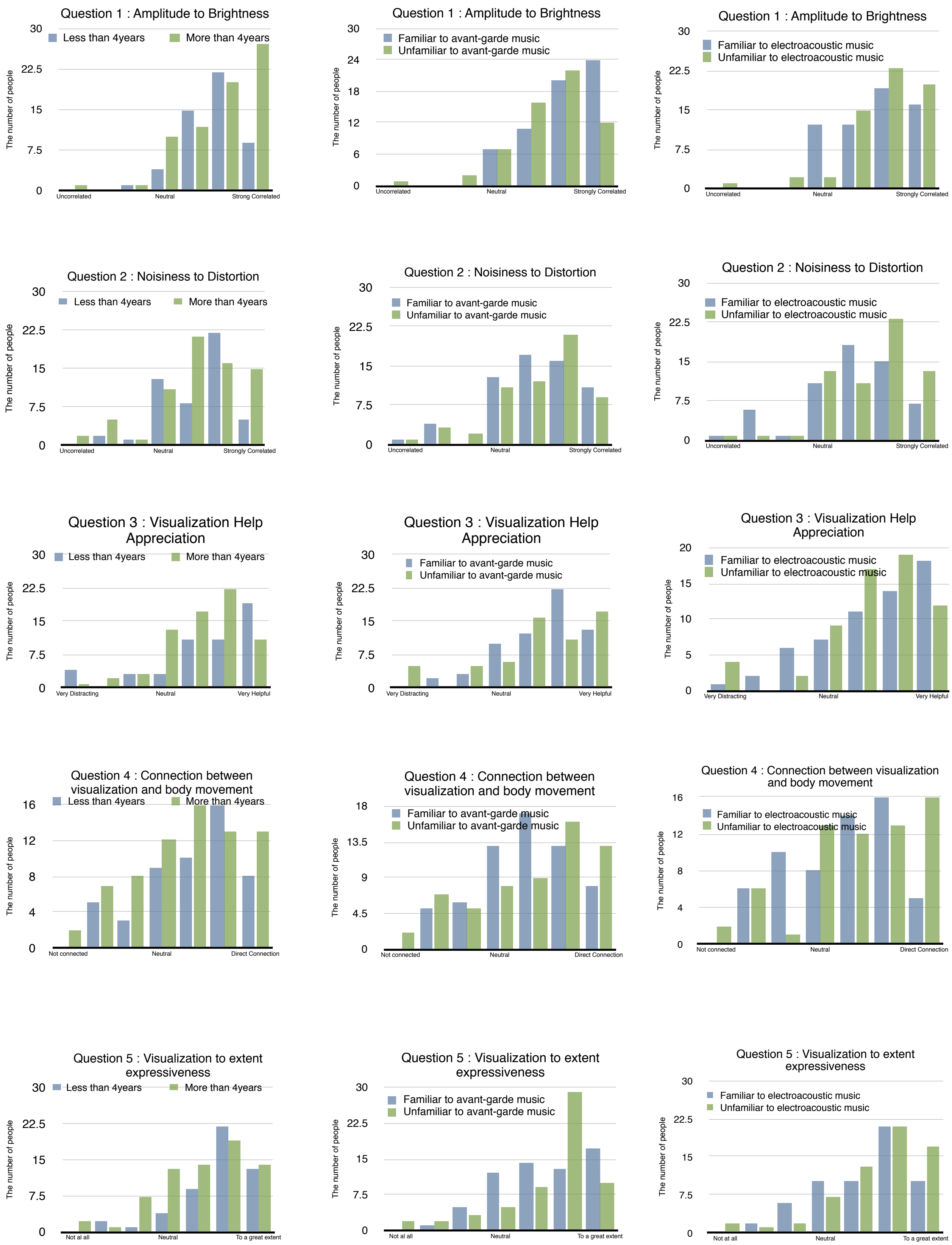



# Charts for Each Questions on Three different Groupings



# Appendix E : IRB Exemption Approval Letter

**ACTION ON EXEMPTION APPROVAL REQUEST**

Institutional Review Board
Dr. Robert Mathews, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu | lsu.edu/irb

**TO:** Yeamin Oh
Music

**FROM:** Robert C. Mathews
Chair, Institutional Review Board

**DATE:** May 7, 2014
**RE:** **IRB#** E8795

**TITLE:** Audio-Based Visualization of Expressive Body Movement in Music Performance: An Evaluation of Methodology in Three Electroacoustic Compositions

**New Protocol/Modification/Continuation:** New Protocol

**Review Date:** 5/7/2014

**Approved** ___X___ **Disapproved** ________

**Approval Date:** 5/7/2014 **Approval Expiration Date:** 5/6/2017

**Exemption Category/Paragraph:** 4

**Signed Consent Waived?:** Yes

**Re-review frequency:** (three years unless otherwise stated)

**LSU Proposal Number** (if applicable): ________

**Protocol Matches Scope of Work in Grant proposal:** (if applicable) _______

**By**: Robert C. Mathews, Chairman ____________________

**PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –**
**Continuing approval is CONDITIONAL on:**

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. SPECIAL NOTE:

**All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb*

# Vita

Yemin Oh (Yeamin Oh) was born in Seoul, South Korea in 1977. He graduated from Kyung-Hee University in 2002 with a Bachelor of Music degree, majoring in music composition. From December 1997 to February 2000, he served in the Korean National Police Band. During that time, he studied and learned music arrangement skills. His duty was to arrange both popular songs and classical music scores, as well as managing a library of musical scores. After he was dismissed from the service, he resumed his degree in music composition, and his interest in electronic music was piqued. He won 1st prize from the Ulsan University Computer Music Festival, and 1st prize from the Kyung-Hee scholarship music competition in music composition. At the same time, he got an honor student scholarship in the school of music. During his college tenure, he studied with Sang-Jik Jun, Gwang-woo Chun, and Dong-hee Woo. After he completed his bachelor's degree, his passion for music composition brought him to the United States, where he attended University of Hartford (Hartt School) from 2004-2007, culminating in a Graduate Professional Diploma (GPD) in music composition. At the Hartt School, he studied with Kenneth Steen, Ingram Marshall, Joseph Turrin, Dr. Stephen Gryc and Dr. Roberts Carl. After his time at the Hartt School, he went back to Korea to marry Yujin Im before returning to the U.S to pursue his Master's Degree in Music Technology at Georgia Southern University, where he studied with Dr. John Thompson. He is currently finishing his Ph.D in Experimental Music and Digital Media at Louisiana State University, studying with Dr. Jesse Allison and Dr. Stephan Beck. During the time at L.S.U., he participated in several projects including the development of NEXUS web music interface library, iCAST sound-diffusion system, and constructing 6-channel hemispherical speakers for the laptop orchestra. Meanwhile, his compositions were invited and

selected to perform in several national or international conferences and festivals: Society of Electro-Acoustic Music in the United States Conference 2010, Society of Electro-Acoustic Music in the United States Conference 2011, Channel Noise Concert 2011 in Georgia Southern University, Electronic Music Concert in Story Brook University 2011, Electronic Music Midwest Festival 2012, National Student Electronic Music Event 2013, New Interface of Music Expression Conference 2013, International Computer Music Conference 2013, STUDIO 300 Media art and Music Festival 2013, Electronic Music Midwest Festival 2013, Electroacoustic Barn Dance Festival 2013, National Student Electronic Music Event 2014, and New York City Electroacoustic Music Festival 2014. Currently, he has two kids and continues to live and compose in Baton Rouge, Louisiana.