

Experiential Graphic Sonification for Visual and Auditory Communication Design and
Musical Expression

Woohun Joo

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Individualized Interdisciplinary PhD

Committee Members

Charles Nichols

Myounghoon Jeon

Patrick Finley

Zachary Duer

May 6, 2022

Blacksburg, VA 24061

Keywords: Human Computer Interaction, Sonification, Communication Design, Sound
Art, Computer Science, Visual Art

Copyright © 2022 Woohun Joo

Experiential Sonification Design for
Visual and Auditory Communication Design and Musical Expression

Woohun Joo

ABSTRACT

This dissertation explains two sonification platforms designed for image-sound association study and art, and the study results. A platform for user study was developed first and an artistic audiovisual platform was derived based on it. First, the five image-sound association studies were conducted to see whether people can successfully associate sounds and fundamental shapes (i.e., a circle, a triangle, a square, lines, curves, and other custom shapes) and the correct answer rate was high. Then, the same sonification platform was transformed by adding colors to the audiovisual platform for artistic/musical expression. A line-by-line sonification method and an object-oriented method were newly developed to sonify the background and the shapes separately. To enhance user experience, the sound of each shape was spatialized in a multi-layer speaker environment or a virtual listening environment.

Experiential Sonification Design for
Visual and Auditory Communication Design and Musical Expression

Woohun Joo

GENERAL AUDIENCE ABSTRACT

This dissertation introduces two sonification platforms designed for image-sound association study and art, and the study results. A platform for user study was developed first and an artistic audiovisual platform was derived based on it. First, the five image-sound association studies were conducted to see whether people can successfully associate sounds and fundamental shapes (i.e., a circle, a triangle, a square, lines, curves, and other custom shapes) and the correct answer rate was high. Then, the same sonification platform was transformed by adding colors to the audiovisual platform for artistic/musical expression. A line-by-line sonification method and an object-oriented method were newly developed to sonify the background and the shapes separately. To enhance user experience, the sound of each shape was spatialized in a multi-layer speaker environment or a virtual listening environment.

Acknowledgements

Throughout my doctoral studies and the writing of this dissertation, I have received a great amount of assistance and support in many ways.

First, I would like to express my deepest appreciation to my PhD supervisor, Professor Charles Nichols, and my doctoral advisory committees, Professor Zachary Duer, Professor Patrick Finley, and Professor Myoungsoon Jeon, whose expertise was invaluable in formulating the research hypotheses, designing the user studies, and the sonification methodology. Their discerning feedback always led me to firm my research direction, broadened my perspective, and stimulated my creativity.

I am extremely grateful to Karen P. DePauw, a former dean for Graduate Education, and Dr. William Huckle, an associate dean of the Graduate School, for their valuable trust and support throughout my doctoral studies from start to finish.

In addition, I would like to thank the VT Center for HCI for funding my user study. I was successfully able to compensate the study participants throughout my user study.

Special thanks to Tanner Upthegrove at iCAT, for caring about many technical issues in the Cube and helping me a lot with spatial audio for my own and group research.

Also, I would like to acknowledge all user study participants (30 in total). They helped me complete my image-sound matching survey online and I successfully gathered data without a single issue.

Lastly, I would like to thank my wife Boyoung Lee for your patient support. You are always there and believe in me. I could not have done it without you.

Table of Contents

List of Figures.....	vii
List of Table	xi
1 Introduction and Background.....	1
1.1 MOTIVATION.....	1
1.2 VISUAL COMMUNICATION DESIGN.....	4
1.2.1 <i>Graphic Design</i>	4
1.2.2 <i>Data Visualization</i>	8
1.2.3 <i>Graphical User Interface</i>	9
1.2.4 <i>Visual Communication Design for Non-Visual Communication</i>	10
1.3 SONIFICATION.....	11
1.3.1 <i>Definition</i>	11
1.3.2 <i>Sonification Types</i>	12
1.3.3 <i>Sonification in Assistive Technology for Visually Disabled People</i>	15
1.4 CROSS-MODAL, SYNESTHETIC ART/DESIGN APPROACHES.....	18
1.4.1 <i>Introduction</i>	18
1.4.2 <i>Synesthesia in Art and Graphic Score</i>	19
SOURCE: ADAPTED FROM [148]	37
1.4.3 <i>Abstract Film and Visual Music</i>	40
1.4.4 <i>Multi-Modal Musical Instrument</i>	46
1.4.5 <i>Audiovisual Art, Multimedia Art, and Design</i>	58
1.4.6 <i>Multimedia Software</i>	68
1.5 THESIS OUTLINE AND CONTRIBUTION	69
2 Literature Review	71
2.1 INTRODUCTION	71
2.2 GRAPHIC SONIFICATION MODELS	72
2.2.1 <i>Graphic Sonification Models in Art</i>	72
2.2.2 <i>Graphic Sonification Models in Science</i>	82
2.2.3 <i>Graphic Sonification Models in Musical Instrument</i>	94
2.3 PROPOSITIONS.....	102
3 Research Hypotheses.....	103
4 Graphic-Sound Sonification, Experiments and Methodologies	104
4.1 INTRODUCTION	104
4.2 USER TEST	105
4.2.1 <i>Purpose</i>	105
4.2.2 <i>Participants</i>	106
4.2.3 <i>Implementation</i>	106
4.2.4 <i>Study Procedure</i>	107
4.2.5 <i>Pre-questionnaire</i>	108

4.2.6	<i>Study1: Task and Sound Design</i>	109
4.2.7	<i>Study 2 and Study 3: Task and Sound Design</i>	114
4.2.8	<i>Study 4: Task and Sound Design</i>	119
4.2.9	<i>Study 5: Task and Sound Design</i>	120
4.2.10	<i>Post-Questionnaire</i>	122
5	Data Analysis	124
5.1	RESULTS (PRE-QUESTIONNAIRE)	124
5.2	RESULTS (STUDY 1)	125
5.2.1	<i>Analysis and Findings</i>	125
5.3	RESULTS (STUDY 2 AND STUDY 3)	133
5.3.1	<i>Analysis and Findings</i>	133
5.4	RESULTS (STUDY 4)	144
5.4.1	<i>Analysis and Findings</i>	144
5.5	RESULTS (STUDY 5)	147
5.5.1	<i>Analysis and Findings</i>	147
5.6	RESULTS (PROBABILITY).....	150
5.7	RESULTS (POST-QUESTIONNAIRE)	151
6	Artistic Audiovisual Expression	154
6.1	INTRODUCTION	154
6.2	IMPLEMENTATION	154
6.3	SONIFICATION METHODOLOGY	156
6.3.1	<i>Overview</i>	156
6.3.2	<i>Background Sonification</i>	157
6.3.3	<i>Shape Sonification</i>	161
6.3.4	<i>Control Interface</i>	163
6.3.5	<i>Outcome</i>	165
7	Conclusion and Future Work	167
7.1	CONCLUSION	167
7.2	FUTURE WORK	168
8	APPENDIX. Test Materials	171
9	References	203

List of Figures

Figure 1.1: Cover art of <i>I don't live here anymore</i> (2021)	2
Figure 1.2: Dynamic Poster Design in Instagram	2
Figure 1.3: Wassily Kandinsky's Questionnaire	7
Figure 1.4: UltraCane.....	16
Figure 1.5: VISION-800 Glasses with vOICe	17
Figure 1.6: Fugue in Red by Paul Klee.....	22
Figure 1.7: Abstract Trio by Paul Klee	22
Figure 1.8: Vision by Carol Steen.....	23
Figure 1.9: Concerto for Four Violins by Anne Salz.....	23
Figure 1.10: Fontana Mix by John Cage.....	25
Figure 1.11: December 1952 by Earle Brown	26
Figure 1.12: Graphic Score of Zyklus by Stockhausen	27
Figure 1.13: Elektronische Studie II by Stockhausen	28
Figure 1.14: MIKROPHONIE I by Stockhausen	28
Figure 1.15: Threnody for the victims of Hiroshima by <i>Penderecki</i>	29
Figure 1.16: Treatise by Cardew.....	30
Figure 1.17: Your Kindled Valors Bend by Escot	30
Figure 1.18: Twelve Fantasy-Pieces after the Zodiac for Amplified Piano by Crumb	31
Figure 1.19: Prima Vista by Kagel	32
Figure 1.20: Stripsody by Berberian.....	32
Figure 1.21: Golden Flower Piece by Philips	33
Figure 1.22: Ornamentik.....	34
Figure 1.23: Music for Airport by Eno	34
Figure 1.24: Towards an Unbearable Lightness by Nelson	35
Figure 1.25: Bird Gong Game by Guy.....	36
Figure 1.26: Composition No. 14 by Brexton.....	37
Figure 1.27: Luminous Axis by Smith.....	37
Figure 1.28: Celestial Spheres Fantasy for Improvisers by Loyato	38
Figure 1.29: Picnic by McQueen	39

Figure 1.30: Pavilion Score by Roden	39
Figure 1.31: Light Play Opus 1 by Ruttman	41
Figure 1.32: Symphonie Diagonale by Eggeling	42
Figure 1.33: Rhythmus 21 by Richter	42
Figure 1.34: Allegretto by Fischinger	43
Figure 1.35: Synchrony by McLaren	44
Figure 1.36: Ocular Harpsochord	47
Figure 1.37: Pyrophone.....	48
Figure 1.38: Sarabet	50
Figure 1.39: Variophone	51
Figure 1.40: ANS synthesizer	51
Figure 1.41: Oramics	52
Figure 1.42: Orbita.....	53
Figure 1.43: Biophilia	55
Figure 1.44: Lumi keys	57
Figure 1.45; Eyesy	58
Figure 1.46: Electronic Television by Paik.....	59
Figure 1.47: Dream House by Young	60
Figure 1.48: Reflection by Eno.....	62
Figure 1.49; Alpha Pulse by Nicolai.....	62
Figure 1.50: Test Pattern by Ikeda.....	63
Figure 1.51: Layer Cubed	64
Figure 1.52: Quantum Memories by Anadol	66
Figure 2.1: UPIC by Xenakis.....	96
Figure 2.2: Metasynth	97
Figure 4.1: Bouba/Kiki Effect.....	105
Figure 4.2: A dot scanning path on a circle	106
Figure 4.3: User Study Implementation	107
Figure 4.4: Four basic shapes for Study 1.....	113
Figure 4.5: Three basic shapes for Study 2.....	117
Figure 4.6: Four lines for Study 3.....	118

Figure 4.7: Three custom shapes for Study 4	119
Figure 5.17: Spectrogram: Sawtooth VS Square Wave in 220Hz	132
Figure 6.1: Implementation for the User Study	156
Figure 6.2: Colormatrices	157
Figure 6.3: Original Image (left) and Sonification Visualization Components (right)...	158
Figure 6.4: Scan line path for the background	159
Figure 6.5: Additive and waveshaping	160
Figure 6.6: Background Color Examples.....	161
Figure 6.7: Shapes and their corresponding channels (except CH6)	162
Figure 6.8: MIRA interface for the six channels	165
Figure 6.9: Two sonification methods for the audiovisual platform	166
Figure 6.10: Screenshot of the audiovisual version	166
Figure 6.11: Virginia Tech’s Perform Studio	167
Figure 8.1: IRB Approval Letter (pg.1)	171
Figure 8.2: IRB Approval Letter (pg.2)	172
Figure 8.3: Recruitment flier (pg.1)	173
Figure 8.4: Recruitment flier (pg.2)	174
Figure 8.5: Consent Form	175
Figure 8.6: Pre-Questionnaire.....	176
Figure 8.7: Main Study Version 1 (Task 1 Intro)	177
Figure 8.8: Main Study Version 1 (Task 1)	179
Figure 8.9: Main Study Version 1 (Task 2 Intro)	180
Figure 8.10: Main Study Version 1 (Task 2)	182
Figure 8.11: Main Study Version 1 (Task 3 Intro)	184
Figure 8.12: Main Study Version 1 (Task 3)	185
Figure 8.13: Main Study Version 1 (Task 4 Intro)	186
Figure 8.14: Main Study Version 1 (Task 4)	188
Figure 8.15: Main Study Version 1 (Task 5 Intro)	189
Figure 8.16: Main Study Version 1 (Task 5)	191
Figure 8.17: Main Study Version 2 (Task 1)	193
Figure 8.18: Main Study Version 2 (Task 2)	195

Figure 8.19: Main Study Version 2 (Task 3)	197
Figure 8.20: Main Study Version 2 (Task 4)	199
Figure 8.21: Main Study Version 2 (Task 5)	201
Figure 8.22: Post-Questionnaire	202

List of Table

Table 4-1: Sound-Shape Mapping in Study1	112
Table 4-2: Sound-Shape Modulation Mapping in Study2 and Study 3	116
Table 4-3: Sound FX-Shape’s Movement Mapping in Study5	122
Table 5-1: Pre-questionnaire 1	124
Table 5-2: Pre-Questionnaire 2 (Left) and 3 (Right)	125
Table 5-3: Pre-Questionnaire 4.....	125
Table 5-4: Answer to question 1 of Study 1	126
Table 5-5: Likert question 1 of Study 1	126
Table 5-6: Answer to Question 2 of Study 1	127
Table 5-7: Likert Question 2 of Study 1	128
Table 5-8: Answer to Question 3 of Study 1	129
Table 5-9: Likert Question 3 of Study 1	129
Table 5-10: Answer to Question 4 of Study 1	130
Table 5-11: Likert Question 4 of Study 1	130
Table 5-12: Confusion Matrix of Study 1	133
Table 5-13: Accuracy, Precision, and Recall for each question	133
Table 5-14: Answer to Question 1 of Study 2	134
Table 5-15: Likert Question 1 of Study 2	134
Table 5-16: Answer to Question 2 of Study 2	134
Table 5-17: Likert Question 2 of Study 2	135
Table 5-18: Answer to Question 3 of Study 2	135
Table 5-19: Likert Question 3 of Study 2	136
Table 5-20: Confusion Matrix of Study 2.....	136
Table 5-21: Accuracy, Precision, and Recall for each question	137
Table 5-22: Answer to Question 1 of Study 3	137
Table 5-23: Likert Question 1 of Study 3	138
Table 5-24: Answer to Question 2 of Study 3	139
Table 5-25: Likert Question 2 of Study 3	139
Table 5-26: Answer to Question 3 of Study 3	140
Table 5-27: Likert Question 3 of Study 3	140

Table 5-28: Answer to Question 4 of Study 3	141
Table 5-29: Likert Question 4 of Study 3	142
Table 5-30: Confusion Matrix of Study 3.....	143
Table 5-31: Accuracy, Precision, and Recall for each question	143
Table 5-32: Answer to Question 1 of Study 4	144
Table 5-33: Likert Question 1 of Study 4	144
Table 5-34: Answer to Question 2 of Study 4	145
Table 5-35: Likert Question 1 of Study 4	145
Table 5-36: Answer to Question 3 of Study 4	146
Table 5-37: Likert Question 3 of Study 4	146
Table 5-38: Confusion Matrix of Study 4.....	147
Table 5-39: Accuracy, Precision, and Recall for each question	147
Table 5-40: Answer to Question 1 of Study 5	148
Table 5-41: Likert Question 1 of Study 5	148
Table 5-42: Answer to Question 2 of Study 5.....	149
Table 5-43: Likert Question 2 of Study 5	149
Table 5-44: Answer to Question 3 of Study 5	149
Table 5-45: Likert Question 3 of Study 5	150
Table 5-46: Summary of significance level tests for all studies	151
Table 5-47: Answer to Post-Questionnaire 1	152
Table 6-1: HSL Mapping for the background.....	160
Table 6-2: Filter-related values for custom shapes in Channel 4.....	162
Table 6-3: Pitch setting for the four channels.....	163

1 Introduction and Background

1.1 Motivation

Today, smartphones are seen as high-spec entertainment devices. How we consume media, services, products, and information has profoundly changed. This ecosystem has become how we express ourselves, artists and designers promote themselves, and businesses expose commercials via social networking services. This environment is actively adopting new mainstream technologies to improve user experience and communication design. Where is graphic design in this technology area? Graphic design plays a key role as a visual communication tool in interface design for all technology platforms. Traditional graphic design techniques have evolved into more dynamic and lively forms in social media and other platforms we commonly use. For instance, in Apple music, album artwork has begun to be animated. The snow in the cover art of The War on Drug's latest album is scattered (Figure 1.1). Almost all newly released album covers have animation. Poster design and other graphic design expressions found in social media platforms animate in smart phone screens like when Flash-based [1] web design was everywhere. Graphic designer Shin Dokho's poster is animated in his Instagram archive (Figure 1.2). Likewise, abstract geometric shapes that decorate poster design have started to move. This technique has become a trend on social media platforms. These platforms create new user experience in graphic design. This lively graphic design form was not previously possible with traditional printed media.



Figure 1.1: Cover art of *I don't live here anymore* (2021)

Source: Adapted from [2]

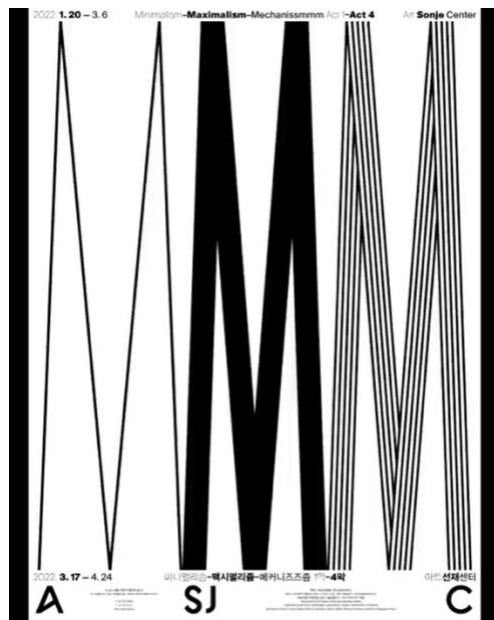


Figure 1.2: Dynamic Poster Design in Instagram

Source: Adapted from [3]

Is there another medium that could enhance user experiences with graphic design? What if sounds represented graphics in poster design? What if we could sense those visual shapes with sounds? A design researcher Dr. Michael Haverkamp emphasizes the importance of the cross-sensory design approach by saying that “Paul Klee and Wassily

Kandinsky argued for correlations in painting and music . . . in spite of the fulfillment of functional requirement, the design of industrial products was mainly focused on the visual form.” [4] Tactility is often considered as one of the most talked-about topics [5][6] as an alternative graphic design for visually impaired individuals, but it is not well fitted for a portable computing environment or everyday use because it naturally requires additional bulky hardware devices [7][8]. Even if we do not necessarily bring art and design examples, as the bouba/kiki effect (introduced in 1929 by a psychologist Wolfgang Köhler) [9] shows, we humans have functions that associate between visual and sound naturally and unconsciously regardless of a social and cultural background. Considering this fact, the bouba/kiki effect can be a compelling scientific resource that inspires sonification, specifically for the graphic-sound conversion before we create a blueprint for a comprehensive relationship between shapes and sounds.

One thing I need to point out is that if sound is applied to enhance visual communication design or to further increase user experience, this does not necessarily mean that sound should always be delivered to viewers. Sound is more intrusive than visual information because visual information can be blocked by closing our eyes, but auditory information cannot be voluntarily shut out because we cannot close the ears. That is, if visual/auditory information packages are shown through mobile phones with private listening devices like headphones, or if the content is provided with interactive interfaces that allows the users to turn the sound on and off, there will be no problem. However, if these types of art must be displayed in public spaces and the sound should be always on, it cannot be said that this design approach always provides a pleasant experience for passersby.

If graphic design can be combined with sonification in a way that can deliver shape and color information, in addition to enhancing the user experience, this graphic-sound combination also can possibly help visually impaired people imagine or feel how the graphic design elements are composed. Even for sighted people, combining sonification and graphic design has great potential for becoming a new type of artistic expression and communication, stretching the boundary of traditional graphic design or media art from a

user experience perspective, and taking a step further towards the next evolution of graphic design and new media art methodology.

The goal of this study is to design sonification-driven audiovisual platforms that convert elementary visual shapes into sounds and examines whether or not people can easily sense the relationship between them. Sonification with sound spatialization techniques will help listeners imagine the moving shapes and feel the sense of space in headphones or speakers to boost the liveliness of auditory expression. This exploration is to show sonification as a potential to enhance visual communication design and create a new type of audiovisual tool. Due to the multidisciplinary nature of this study, the background of a wide range of areas from graphic design, visualization, sonification, musical instrument design, visual music, and multimedia art will be covered in this chapter to explain why the study is significant.

1.2 Visual Communication Design

1.2.1 Graphic Design

According to the Merriam-Webster dictionary, “the definition of graphic design is the art or profession of using design elements (such as typography and images) to convey information or create an [visual] effect [10].” This clear-cut definition describes what graphic design is today. In college graphic design classes, it is common for students to be asked to define graphic design by themselves. For example, must it include the fine arts or paintings? This question does not require a clear answer because the definition of graphic design varies depending on what age we are living or how we utilize the designs. In this respect, graphic design has quite a broad spectrum in visual art forms and can be interpreted in various ways.

The term “graphic” includes the meaning of both *writing* and *drawing* [11][12]. Two different media serve as the same visual communication tool. From a pragmatic point of view, the word “design” implies problem-solving processes [13][14], and graphic design is one type of these processes that utilizes both writing (typography) and drawing (other visual media such as illustration and photography). And since graphic

design is dependent on our sense of sight, the term “visual communication design” is considered as a synonym for graphic design. Generally, graphic designers create symbols, icons, texts, and images to form visual representations of ideas and messages. The main role of graphic design is to communicate/convey information as speedily and effectively as possible in a visual form.

In a broad sense, the origin of graphic design lives together with the history of human existence [15]. From prehistoric times, cave paintings all over the world serve as a type of visual communication that contains various subjects, stories, and themes. History scholars or anthropologists can determine a great deal of information about a living environment from cave paintings regardless of the presence or absence of written language: our ancestors communicated by drawing on the walls and ceilings of caves.

One of the oldest human communication media are spoken languages. Spoken languages have inherently extinctive qualities. That is, if no one is present to hear what is spoken, the message is lost. Thus, we use written languages to convey messages disregarding the time and place. Written languages are a symbolic system to be recorded and can be understood as a type of visual communication tool [15]. The earliest writing systems, such as cuneiform characters, are logographic scripts, an iconic and symbolic presentation of words. Chinese is a typical example of an ideogram [16]. In an ideogram, an idea, concept, and object are symbolized with a visual resemblance, thus an ideogram itself does not convey how to pronounce the symbol. In contrast, Hangul (Korean) and English are phonograms meaning that they indicate a pronunciation key [17][18]. More specifically, the shapes of consonants and vowels in Hangul were created based on shapes of human vocal organs and resultant sound [19], so Hangul phonograms are visual icons of pronunciation. Despite their differences, human writing systems form the most common visual communication design in history.

The invention of printing had great significance and making books became more popular and easier than ever before. Since printing facilitated mass production and reduced the cost of the product, education and knowledge, which were limited to the upper class, were efficiently and effectively made available to the public. Diffusion of knowledge provided the main impetus for reform of our society and this change was in line with communication design history. Early printing technologies from the advent of woodblock printing, movable wood types to movable metal type blocks were great historical events. Once the printing industry started to use logotypes and mass-produced print advertisements in the 17th century, modern graphic design entered a new era. It has been said that graphic design truly starts with the advent of printing [15][20]. Printing can be understood as an early type of communication media.

In general, the 20th century is viewed as an important period in the explosion of graphic design [21]. The Bauhaus movement has become one of the most influential icons of modern design [22]. Graphic designers in any generation consider Bauhaus as the apex of modern and contemporary design and even now the design philosophy of Bauhaus continues as the most up-to-date design methodology. With the introduction of a personal computer, graphic design underwent a rapid change. The design process, education, and working environment were dramatically changed and simplified by computer graphics software such as Adobe Illustrator [23], Photoshop [24], and QuarkXPress [25]. Lately, graphic design has been faced with the convergence of computational design methodology. The role of graphic design and methodologies continue to evolve.

At Bauhaus, Kandinsky proposed universal relevance [22] between three basic forms - triangle, square, circle- and colors. In this reference, the triangle is yellow, the square is red, and the circle is blue. The test participants were asked to fill in these three shapes with the primary colors (Figure 1.3). Kandinsky wanted to seek an ideal universal visual language that can be free from the cultural barriers of alphabetical writing. Kandinsky's passion for those elementary shapes became a representative symbol of Bauhaus and today these shapes are the basis of all graphic design products

in signages, posters, and even typography, serving as an elemental material of the visual language. The significance of Kandinsky's equation implies modernism, post-modernism, universal visual form, logo, transhistorical script, utopian idealism, and vision as a realm of expression [22]. At its simplest form, those three shapes represent the foundation of visual communication. I took Kandinsky's universal visual forms as my sonification metaphor and interpreted them as sound.

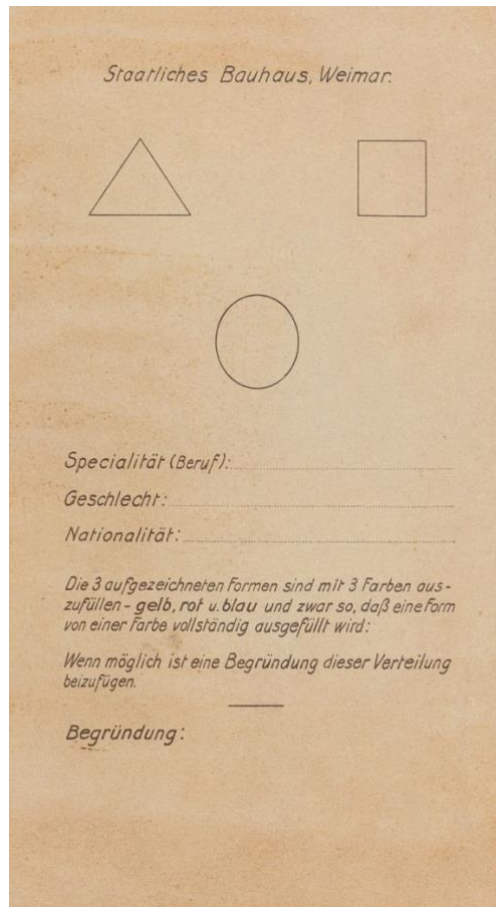


Figure 1.3: Wassily Kandinsky's Questionnaire

Source: Adapted from [26]

Graphic design innovation is consistently driven through technological improvement. The following sections in this chapter cover data visualization, graphical user interfaces, and visual communication as an assistive device. These areas serve as contemporaries of visual communication design.

1.2.2 Data Visualization

Like written language, data visualization is also a form of visual communication. The definition of data visualization is simply a graphical representation of data and information [27][28]. It uses two- or three-dimensional figures, lines, colors, and sizes to illustrate non-visual data in the form of charts, graphs, and maps. Therefore, data visualization is a form of visual communication. The ideal purpose of traditional data visualization is to design easily accessible data, so its viewers can understand and learn from it [28]. Since the visual symbols are used to transform number-based data, this visual design method is great for expressing complex numeric forms, data variances and changes over time as well as data comparison [27][28][29].

William Playfair and Edward Tufte studied practical applications of data visualization. Playfair developed systematic visual representations to replace conventional number-based data [30], and Tufte, who is known as an early pioneer in data visualization, supplemented his empirical theory and practice of the statistical infographics to comprehend these visual expressions in his book *The Visual Display of Quantitative Information* [29]. Tufte then provided guidelines on how to universally communicate information through visual shapes, words, and color. This visual communication design approach enhanced the readability of traditional text/data tables as well as opened a new paradigm of the data communication methodology [29][31][32]. As such, a visual representation has become the norm and has been commonly used as the form of data information communication in an efficient and precise way.

Computer technology and the advent of personal computers now make the data visualization process easier and simpler. Microsoft PowerPoint [33] and Apple Keynote [34] are examples of data visualization software. A code-based drawing tool, Processing [35], offers a more flexible and malleable visualization environment that allows a computational design approach. This approach allows visual designers to customize the visualization tool according to their preference and to utilize data for artistic expression.

The primary goal of data visualization is information delivery. Therefore, its successful implementation requires intuitiveness and practicality. In contrast, due to its inherent flexibility, with no limitation of expression, and a host of possibilities for artistic merit, it has become common to see artists and designers use it as a medium of creative expression as well as an art/design education tool. Now data visualization does not necessarily mean the practical side of data presentation. Instead, it covers both art and information science.

The sonification system in this dissertation has a visualization component that display the sonification process in real time. This visualization component will lead the listener to synchronize the sound and visual information more closely.

1.2.3 Graphical User Interface

For another example of visual communication, Graphical User Interface (GUI) – also called User Interface (UI)- exists specifically for the purpose of facilitating better communication between humans and software/hardware, that is, the usability of products [36]. UI design is a creative design process for building an interface in a computerized system in software and/or hardware [37]. This does not solely focus on visual aesthetics, rather, a great deal of importance is given to intuitiveness, effectiveness, and efficiency as well as usability, satisfaction, and likeability [36][37].

The main role of a GUI/UI is to enhance user experience design and augment the quality of communication strategies between users and devices. For instance, the first generation of iPod was operated by a scroll wheel and a few buttons around it. Later, these control interfaces were replaced with a single touch-based scroll wheel. This was a turning point for UI design, as the wheel began to grab the attention of designers across the world. Those minimalistic design approaches inspired by Dieter Rams continue to have a great effect on GUI/UI design to this day.

Specifically, GUI design refers to the use of visual representation to control devices with which users interact. For example, computer Operating Systems (OS), such as Windows

[38], Android [39], Ubuntu [40], and MacOS [41] provide their own visual user interface for interaction: mouse cursors, icons, folders, buttons, and menu items on the screen.

GUI/UI design is an ubiquitous visual communication design and a pivotal part of the modern HCI field because it determines how to use tools effectively. Likewise, since sonification itself plays an important role as an auditory user interface, the purpose of this study can be seen to share the same significance as visual GUI/UI design. The user study in Chapter 4 can be extended to be an important future research topic to study how the shape-sound association can be used in user interface design.

1.2.4 Visual Communication Design for Non-Visual Communication

The process of visual communication design places emphasis on support of the graphical appearance of communication ideas. This includes drawing, text, images, and graphic design [42]. Sight, smell, taste, touch, and hearing are our five basic senses, and visual design heavily relies on vision, which is our dominant sense [43][44]. Traditionally, it has been acknowledged that “vision and thinking are one process” [44]. Since a lot of sensory information is conveyed by sight, product design traditionally focuses on the visual appearance of objects [4]. Research shows that 80 to 85% of our perception and cognitive activities are mediated through vision [45]. For instance, sighted people can easily and rapidly see the fonts, photos, colors, and layout of each page in this dissertation. This process is almost simultaneous. Therefore, we may conclude that visual communication design is not fit for people with visual impairment because blind people do not have access to visual information. There have been challenges in overcoming inherent limitations of visual design in combination with tactile sensation [5][7][8][46]. Coupling with sounds can be another design solution. Sonification refers to this approach (sonification will be discussed in Chapter 1.3). When other forms are added, it can no longer be considered visual information design. It is simply and more broadly communication design.

A photographer, Yann Arthus-Bertrand needed to explain his *Earth from Above* exhibition [46][47] to some blind children. Alain Mikli, who is well-known for designer

glasses, created a tactile version of Bertrand's photography, transferring his high-definition images to eight level reliefs, according to pixel color, and layering them on an 8mm cellulose acetate plate [7]. Later, Mikli planned a tactile art project for blind and partially sighted people in museums in France. The same tactile design method is still adopted today. This design approach seeks an optimal connection between visual information and tactile perception. Naturally, it is not possible to completely replace our sight for other sensory modalities due to the nature of each sense. For example, a tactile perception has a sequential character [49]. With a finger touching the objects, it takes time until mental images of the objects are formed based on the size of fingertips [7][49]. In contrast, visual recognition involves fast recognition [7][49][50]. With one's eyes open, objects can be seen simultaneously and the perceptive area is wider than tactile perception [49]. In sonification, sound can be heard, causing an instantaneous reaction [43]. However, there are perception limits on classifying many kinds of different sounds at once [43][51].

In an assistive model of visual communication design, the role of one sensory modality can be enhanced, expanded, or substituted by another counterpart. Even though an approach does not exist for perfect replacement for each mode, they act as a supporting aid or assisting tool. The examples above seek to illustrate the possibility and feasibility of alternative communicative means. Successful outcomes can include an innovative cross-modal sensory approach to communication design.

1.3 Sonification

1.3.1 Definition

According to the NSF (National Science Foundation) Sonification Report of 1999, "Sonification is the use of nonspeech audio to convey information" [52]. Another definition proposed by a sonification researcher, Thomas Hermann is that "Sonification is the data-dependent generation of sound, if the transformation is systematic, objective and reproducible so that it can be used as scientific method" [53][54]. Sonification aims to seek the audibly comprehensive relationship between data and sound [43] and the nature of sonification is inherently interdisciplinary [43], because sonification is the utilization

of sounds to replace non-sound sources and those non-sound sources are not specific to academic fields. The interdisciplinary nature of sonification is not fully explored, because sonification continues to support new fields of study today. Specifically, sonification is the main component of auditory display, which communicates data to users by using sound. Auditory display covers a broad scope for human-machine interaction as a whole [43]. Considering that data visualization is categorized under the visual display, sonification can be seen as the counterpart. However, sonification uses a different sensory modality. The sonification design process can generally be summarized in five steps: perception, data gathering or analysis, sound synthesis, applying the sonification technique, and central application, in that order [43].

The primary community in sonification is the International Community for Auditory Display (ICAD) founded in 1992. Sonification has been studied for the human perceptual system to understand complex and big data heard by the human ear, data monitoring processes such as earthquake sensing, and to create a human-machine interface [43]. Since sonification research is relatively new, there is little information on thoroughly understanding its functionality, practicability, and accessibility. For example, there are only three sonification books [43][55][56] currently in print. Since *Sonification Handbook* [43] tends to be solely used as a central reference for this field, this sound-side communication camp is not systemically as well-organized as Tufte's visualization camp. Despite this lack of definition, sonification has a great potential for ubiquity in communication, just as similar trends, such as technology development and neuroscience and human sensory modality studies, have the potential to expand with more time spent researching. Sonification remains underutilized [43][52][57][58] despite prior and ongoing efforts to broaden human cognitive bandwidth when combined with other senses [43][4][59]. Therefore, other closely aligned contexts and further studies are warranted.

1.3.2 Sonification Types

Sonification maps data to various sound properties to convey information that is inherently inaudible [43][58]. Auditory displays have been used to create assistive devices [58][59][60][61] for visually impaired individuals as well as to analyze data

[58][62], augment [58][63][64][65] [66], and entirely replace visual displays [58][67]. This shows that sonification has the potential to naturally complement and cooperate with visual communication.

According to the *Sonification Handbook*, the primary function of sonification can be divided into four categories: 1) alarms, alerts, and warnings; 2) status, process, and monitoring; 3) data exploration; 4) art, entertainment, sports, and exercise [43]. The first three groups were classified by Buxton and others [68][69][70], while the fourth group was added by Hermann and others [43]. First of all, alerting functions demand listeners' immediate attention based on the occurrence of an event [43]. For instance, a phone's ringtone tells listeners to answer the phone, but the ringtone itself does not contain any information. The alerting sound is information-poor [43]. Secondly, status, progress, and monitoring functions are particularly suited for a status check when listeners need to inspect the ongoing process of systems [43]. For example, the sound of electrocardiogram monitors in a hospital indicates a patient's stability. Third, data exploration functions are closer to what sonification is supposed to be [43]. While alert and monitoring functions of sonification capture momentary situations, the sound used for data exploration depicts an entire snapshot of data [43]. Auditory graphs (equivalent to visual graphs), charts, and plots, that are usually developed as an assistive tool for visually impaired people [43] are good examples of sonification for data exploration. Lastly, sonification can be used for musical expression. For example, Quinn composed music based on data sonification [71][72]. Unlike the previous sonification purposes, the mapping between data and music must be systematic for musicality but the sound herein does not necessarily have a purpose to make an audience understand data information.

These are five types of sonification: 1) audification, 2) parameter mapping sonification, 3) model-based sonification, 4) auditory icons, and 5) earcons. Audification is a technique for rendering sound from data [43]. Audification directly translates data into sound samples [43]. It is useful for observing patterns or characteristics of data flow when data is not obvious in visual forms, so ideally the sound for audification must be resampled within the range of human hearing [43]. Examples of audification include

seismology, astronomy, and EEG (electroencephalogram). Parameter mapping sonification (PMSon), often the technique that the term “sonification” refers to, changes sound parameters using the information of data [43]. For example, data can control timbre, amplitude, or other acoustic features, or data can modulate filter parameters such as frequency, bandwidth, and Q factor. For PMSon, a mapping between data and sound parameters must be considered in a meaningful way. Model-based sonification mainly focuses on creating a virtual model that requires users’ input to explore data structures. A virtual model can be understood as an instrument or tool with which the user can interact. Hermann describes that “the model-based sonification is the reaction of the data-driven model to the actions of the user” [73].

Auditory icons and earcons have a similar role to that of GUIs, in that they use brief and iconic sounds for communicating events. The difference between auditory icons and earcons is that the former uses non-speech sounds that have an analogous relationship with the objects that the sound describes. Analogous sound has a direct relationship or ecological closeness with its reference [43]. For instance, the sound of paper crumpling can be heard when a user clicks on the recycle bin (or “empty trash”) icon on the computer. Sounds for auditory icons mimic sounds that we are familiar with in our everyday lives. Earcons use symbolic auditory representations where the sound has no resemblance to its reference [43][69][74]. Various kinds of abstract sounds can accompany common commands such as “save,” “open,” or “delete files.” Yet these commands are physically non-existent and inherently have no natural counterpart. Therefore, earcons are metaphorical.

Sonification can be categorized based on its function and technique. However, most importantly, sonification is flexible and different types can be used in conjunction. Moreover, Hermann and others said that sonification names, or taxonomic boundaries, are not crucial for successful sonification design [43]. However, the classification of sonification can become guidance and reference, and understanding theoretical definition can be critical in implementing effective sonification.

1.3.3 Sonification in Assistive Technology for Visually Disabled People

Sonification has a great potential as an assistive technology for visual impairment where sounds are meant to replace visual information. Visually impaired individuals do not have access to visual data, so sonification may be used to convey visual information. Auditory perception offers unique affordances. Sound has an enormous amount of information that a human brain can interpret. Blind people can imagine or describe the outside scenery by listening to the sound of rain dropping on their surroundings [75]. Despite this potential, we don't understand enough about perfect utilization of non-speech sounds to convey the same amount of information [43]. Though universal devices (commercially available or successful products) do not yet exist in assistive technologies, we can, at least, understand the major difference between sight and hearing, the fundamental limitations for substituting visual information, and the nature of sonification. The bandwidth problem [43] is an exemplar of the capacity of our cognition.

Computers or cellphones have a potential to provide auditory interfaces for blind users because of their GUIs. For people who have a visual impairment, this visual interface needs to be substituted by an auditory or tactile interface. Braille is the most widely used type of tactile communication [76]. Screen readers were developed when personal computers became widespread. The first generation of a computer operating system did not provide color images. It was completely text-based. Thus, screen readers relied on a speech-based approach. The next generation of screen readers started to have GUIs requiring more complicated communication methods for blind users. WindowsBridge provides sound feedback based on mouse positions and Mercator/Guib [77] uses non-speech sound samples. Jaws [78] uses auditory icons or earcon-style samples for more complex control commands. However, none of them were commercially successful. Microsoft's Narrator [79] and Apple's VoiceOver [80] are the latest types (but they are still speech-driven) of screen readers, and with the spread of portable computing devices such as smartphones and tablet PC, the number of users is growing. The efficacy of screen readers has not proven to be useful and is still controversial [43].

Humans inevitably move, so mobility is an important matter in addition to computer access. There are two types of assistive technology using sound, obstacle detection in short-range and navigation aid for longer distance travel [43]. Electronic Travel Aids (ETA) refers to these technologies. The most popular tool for obstacle detection is a white cane (or guide cane). When a tip of a guide cane touches the objects, sound and vibration are delivered through haptic and auditory senses, so that blind users can detect a nearby environment when moving. The detectable range of traditional guide cane is somewhat narrow and limited to vertical spots, so more advanced models started to use sensors and headphones for better communication and wider detection ability. For examples of commercially available products, *K Sonar-Cane* [81] and *UltraCane* [82][83] (Figure 1.4) use a laser beam and transfer obstacle information by using pitched sounds that respond to the distance between a user and nearby objects.



Figure 1.4: UltraCane

Source: Adapted from [82]

Auditory display in assistive technologies can be exploited for a long-distance navigating system for blind users, This approach is a combination of speech-driven screen readers and GPSs. NavBelt [84] is a more advanced type of navigation tool, because it uses sonar

and provides two different modes: guidance mode and the image mode. Guidance mode mainly detects obstacles in a user's path. Image mode transfers the entire scene near a user into an audible scene. However, the high cost of the product became an obstacle to popularization. Similar to the image mode of NavBelt, vOICe [85][86] attempted to substitute for visual images. vOICe scans a gray-scaled image horizontally, and the brightness and positions of vertical pixels are mapped into amplitude and pitch. However, it is said that there has not been a considerable amount of evaluation of these technologies [43] (Figure 1.5).



Figure 1.5: VISION-800 Glasses with vOICe

Source: Adapted from [87]

There is a ubiquitous sonification type that we can commonly see every day. When parking and obstacles behind get closer, the tempo of the warning signal increases. Cars now have advanced safety systems such as Lane Keeping Assist System or Front/Rear Collision Warning System, so we can easily experience auditory assistive systems in our daily life. Sonification delivers information as an effective means of communication.

1.4 Cross-Modal, Synesthetic Art/Design Approaches

1.4.1 Introduction

Cross-modal perception is an integration of sensory awareness that relates interaction between two or more different sensory modalities [4]. Synesthesia, a unique neurological phenomenon where stimulation to one sensation triggers a secondary sensory channel, is a representative example of cross-modality. For instance, sound can additionally stimulate a sense of vision, or color and visual features can provoke auditory sensations. When this idea, considering multi-sensory perception as a process of expression and communication, is combined with the art and design-making process, it is characterized as synesthetic art or synesthetic design. The term “synesthetic” is not referring to genuine synesthesia [4].

The integral role of synesthetic approaches in art is to expand traditional art forms and provide unique experience to audiences. In visual music and audiovisual art, music (or sounds) and animated videos are closely interconnected, one art form corresponds to another one or one substitutes for the other [88], so it naturally generates augmented experience to audiences. In interactive art, audiences can be directly involved in art pieces by touching or communicating with it. Conventional art forms rely on a single primary cognitive pathway, but that does not mean that only one sensation will be used, because humans instinctively use all senses all the time. Rather, this means that visual features are mainly pronounced in painting and acoustic features are the base modality in music. Whereas, different modalities in synesthetic art carry the same weight and involve more direct stimulation to two or more sensory modalities including vision, tactility, smell, audition, and taste [4].

Synesthetic design pursues a systematic approach to seek ideal correlations between human sensory modalities [4]. Conventional product design approaches mainly focus on their looks [4]. In contrast, synesthetic design considers all possible strategies of all of our perceptual sensory connections before developing product concepts, technologies, materials choices, and construction [4]. A design researcher, Neil Haverkamp, said, “the aim of synesthetic design is to coordinate all sensations stimulated by an object in a

manner that results in a pleasant, harmonious overall appearance while coinciding with the particular functions desired” [4]. He gave a car design example from a human-machine interaction viewpoint, mentioning that not only the overall sound quality but also the noise behavior of all individual components of a car influences customers’ experience, impression, and purchasing. Therefore, those correlations must be considered and optimized in the car design process [4].

Let's take modern-day theaters as an example of multimodality that we can commonly experience every day. In movie theaters, a traditional communication pathway between viewers and movies is a screen (vision) and speakers (sound). Now, our imaginations interacting with the world of cinema becomes reality: 4D films provide physical effects to viewers such as vibrating seats, fog, light effects, mist, temperature changes, and scent. There is a fast-growing movement to invest in developing sound spatialization techniques in audio technology, and augmented reality and virtual reality in the computer game and IT industry. This shows that technologies face a turning point in implementing the process of human cognitive ability for a more realistic, augmented, and immersive experience. Strictly speaking, this cinema example is not intended to mimic synesthetic experience or simulate real synesthesia (as if music videos were not intended to provide synesthesia). Instead, it can be considered that the purpose of 4D cinema design is to provide a metaphorical synesthetic experience by providing an additional and coercive sensory experience to tow the one-dimensional signal pathway of conventional cinemas to the next level. Obviously, modern multimedia technology is not designed for synesthetic experiences, but in terms of the aspect of augmenting the user experience, modern multimedia art or design technologies can be seen or studied as a synesthetic design approach.

1.4.2 Synesthesia in Art and Graphic Score

Synesthesia in art and graphic scores are separate topics to be dealt with, mainly because graphic scores are not necessarily representative of synesthetic art. That is, graphic scores are a different kind of musical notation. As can be seen from the examples below, graphic scores are usually used for performances by artists, bands, or orchestras, or as artistic

expressions through abstract expression of sound, so they are not directly exposed to the audience along with music. However, sometimes graphic scores are created for synesthetic expression or inspired by synesthesia like in Messiaen's work (this will be mentioned below). So, synesthesia in art and graphic scores were introduced together in this chapter.

In art, Kandinsky believed that music and painting share compatible components. In music, Scriabin, Rimsky-Korsakov, and Rachmaninov discussed the relationship between color and key [89]. John Cage, György Ligeti, and Cornelius Cardew succeeded in those multi-modal interests with a graphical expression of sound. As such, the fusion of sight and hearing (or synesthetic phenomenon) has long been a fascinating subject within not only science but art, playing a role as a stimulant to exceed the limit of artistic expression. Visual music masters, such as Oskar Fischinger and Norman McLaren, used cross-modal art approaches in their work and had a huge influence on contemporary new media artists. These pioneers used a concept of synesthesia as an artistic metaphor to extend their creativity. This idea turned visual music into invaluable artistic assets. This chapter will introduce synesthesia in art and graphic score, and briefly list relevant examples. These examples are indeed ancestors of my sonification/audiovisual approach.

In the early 1900s, Wassily Kandinsky explored space and rhythmic tension of painting in connection with music, showing his interconnection between painting and music [90]. In his *Reminiscences*, Kandinsky described the moment when he was fascinated by Richard Wagner's *Lohengrin* at the Bolshoi Theater in Moscow in 1896 [91]. This event changed Kandinsky's career. He said, "I saw all my colors in my mind, they stood before my eyes", and realized that music can arouse visual images, colors, and moods [92]. These musical connections influenced the titles of his paintings (e.g., *Improvisation*, *Fugue*, *Compositions*). His book, *Point and Line to Plane* (1926), tore down the barrier between music and painting. He mentioned that his questions and goals were to "pass beyond the boundaries of painting" [93]. In *Concerning the Spiritual in Art* (1912), he divided colors into warm-light, warm-dark, cold-light, cold-dark based on yellow, blue, white, and black, and correlated with musical instruments (e.g., blue instruments: flute,

cello, double bass, organ; green instruments: violin; warm red instruments: trumpet; vermilion instruments: drum; orange instruments: old violin; violet instruments: English horn, wood instruments) [90]. Later, Kandinsky founded *Ber Blaue Reiter* (The Blue Rider), an avant-garde and expressionist group, and Arnold Schoenberg joined as a painter. He tried to combine his synesthetic art with dance in *The Yellow Sound* [94].

Paul Klee was not a genuine synesthete, but he exceeded the boundary between painting and music because he was also a skillful violinist. He mainly explored the time element in music and painting saying “Certainly both arts are temporal. This could be proved easily” in his diary written in 1918 [95]. He regarded polyphony in music and harmonized musical forms as a primary subject of his painting. This harmonic nature is depicted as gradient colors and repetitions of shapes [96]. For example, in *Fugue in Red* (1921) (Figure 1.6), he expressed an inherent structure of music composition in a visual form and time progressions with objects floating on air. The objects include familiar objects (e.g., leaf, vase) and abstract figures (e.g., rectangle, triangle, circle) [97]. In his painting *Abstract Trio* (1923) [98], Klee described the sound of an instrument and shapes of players (Figure 1.7). Klee also transferred music notation of *Bach’s Sonata in G Major* with graphical expressions [99]. In the 1920s, Klee, along with Kandinsky and a famous artist, Josef Albers, taught art theory and music at Bauhaus. As this shows, studying both art and music can be thought of as a complete subject, or “a total work of art”, which is a combination of different art forms [100].



Figure 1.6: Fugue in Red by Paul Klee

Source: Adapted from [101]

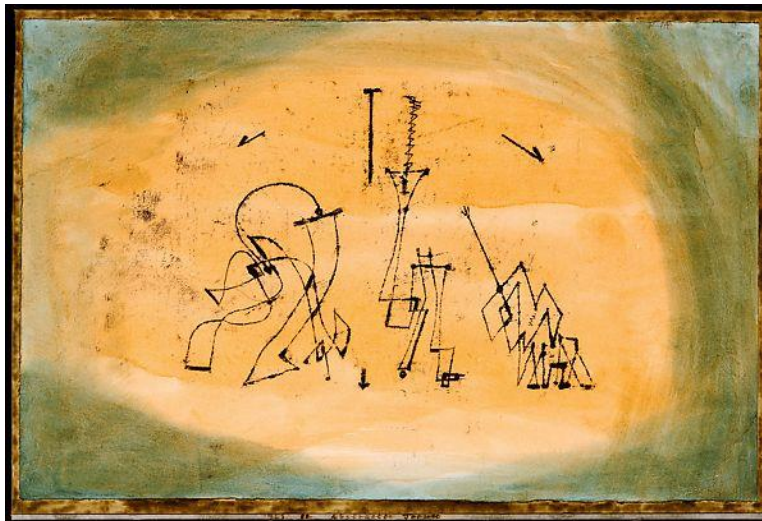


Figure 1.7: Abstract Trio by Paul Klee

Source: Adapted from [98]

Back in 1913, a Russian artist, Léopold Survage, tried to overcome the immobile limitation of traditional art, saying “an immobile abstract form does not say much” [102]. He weighed the “mode of succession of the elements in time” [103] and created more than two hundred watercolors to animate them. *Colored Rhythm: Study for the Film*

(1913) was his first attempt to create an abstract film, but due to the limitation of the technology at the time, it never came to life [104].

In contemporary painting, Carol Steen, an American painter and sculptor, said that she saw colors and shapes with her eyes closed [105]. She depicted her synesthetic perception through her painting *Vision* (1966) (see Figure 1.8), Her visual experience was derived from her sense of touch, hearing, and other sensory absorption [105]. Anne Salz, a Dutch musician and artist, recognized music as a colored pattern [106]. She drew *Vivaldi* (2016) while listening to the opening of *Concerto for Four Violins* to express her synesthesia (see Figure 1.9). She said that when music gives a happy feeling, she perceives red, yellow, and orange, and when she sees a combination of colors, it becomes a musical chord [106].

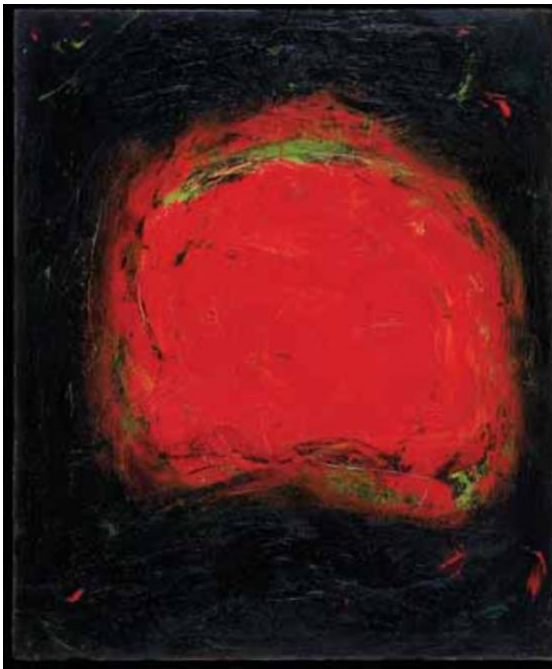


Figure 1.8: Vision by Carol Steen
Source: Adapted from [107]



Figure 1.9: Concerto for Four Violins by
Anne Salz
Source: Adapted from [82]

In music, it is known that Alexander Scriabin's later works were heavily influenced by the association between color and music. He believed that unification between colored light and his symphony creates "a powerful psychological resonator for the listener [108]". One of his symphonic works, *Prometheus: The Poem of Fire* (1910), used a color organ he invented. This instrument is called *Clavier à Lumières*, which means a "light keyboard", and was implemented by Scriabin's subjective color-sound association. Scriabin only recognized red, yellow, and blue corresponding to C, D, and F# [109], though he was not known as a genuine synesthete [110]. But interestingly, his perception of D pairing with "yellow" matched that of Nikolai Rimsky-Korsakov's, a well-known synesthete. However, according to Leonid Sabaneev, a Russian composer and scientist who studied Scriabin as well as other composers whose works are closely related to a color-sound association, including Wagner, Liszt, Chopin, and Debussy, the same result was founded by 78% of his test subjects, because many music scores for brass instruments were written in D. Therefore, Scriabin and Korsakov's agreement on D may stem from their subconscious [108].

These synesthetic approaches in art and music have broadened interests in a graphic music score for many renowned contemporary artists. Graphic notations express sound and music with abstract visual elements. One of the early examples is *Projection I* by Morton Feldman. It was written in 1950 for solo cello. Graphic notations of *Projection I* look like minimal art filled with moderated graphical expression consisting of lines, dotted lines, and hollow rectangular shapes.

John Cage published a book, *Notations* (1969) in company with an American artist, Alison Knowles [111]. *Notations* contained a huge collection of visual scores, manuscripts, and text phrases (ranging from one to sixty-four words) from 269 contemporary composers, showing various methods for music scoring. Cage's first theater piece, *Water Music*, written in 1952, marked a total of 41 musical events with texts and pictogram-like icons on a 55x34 inch large poster [112]. The texts indicate piano notes, radio frequencies, actions, and timings with decimal precision [113]. In *Aria* (1958), Cage notated wiggly lines with texts with ten different colors and sixteen black

squares that present non-musical vocal sounds [114]. The use of different colors were associated with distinct vocal styles [115]. A similar approach can be found in *Fontana Mix* (1958) (Figure 1.10). This piece is comprised of six different curved lines, straight lines, dots distributed, and tilted grids on ten regular sheets and twelve transparent sheets [116]. Earle Brown (collaborated with John Cage) wrote a visual score for *December 1952* (Figure 1.11). This piece has a series of vertical and horizontal lines (that could be seen as thin rectangles) and rectangles on paper. Each element is positioned at right angles varying in thickness. He is also known as an early pioneer of graphic music notations.

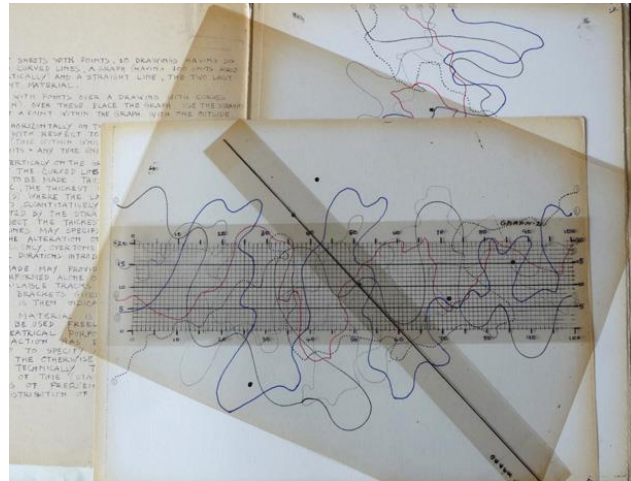


Figure 1.10: Fontana Mix by John Cage

Source: Adapted from [117]

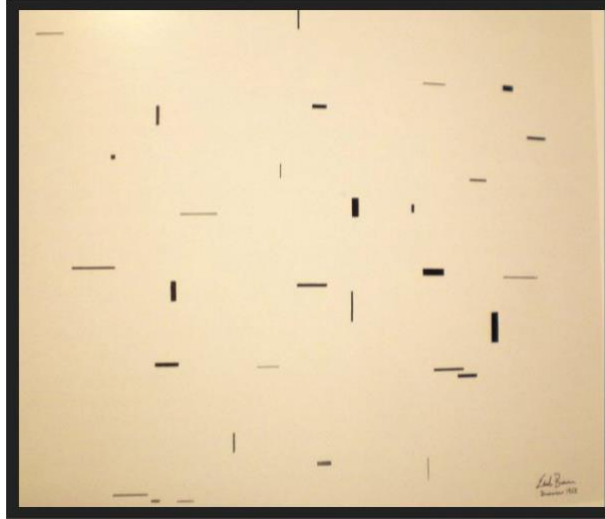


Figure 1.11: December 1952 by Earle Brown

Source: Adapted from [118]

Olivier Messiaen, is known as a synesthete. Color is a core of Messiaen's musical inspiration. He said that he can sense colors clearly when he hears music [119]. Starting in 1948, Messiaen wrote his compositional practices and theories in his *Traité de rythme, de couleur, et d'ornithologie*, which means "Treatise on rhythm, color, and ornithology, describing his transition to colors from chords and modes". For example, in volume 5, he writes that the 15th chord is A major and blue, G sharp adds a bit of gold, and F adds pale green [120]. For more information, please refer to the English translation version [121]. In his orchestral piece, *Couleurs de la Cité Céleste*, Messiaen declared that "melodic, rhythmic themes, sound complexity, timbres evolve based on his color perception" [122].

Karlheinz Stockhausen attempted to describe sounds with various forms such as graphic notation, verbal descriptions, time lines, along with a traditional music notation [123]. A solo percussion work *Zyklus* (1959) is his first graphic score. Thirteen instruments were assigned to particular graphic symbols (Figure 1.12). The symbols consist of 17 sections, called periods, and each period has sixteen pages. The performers were told that they could begin on any page or play everything successively [124]. In his *Elektronische Studie II* (1954), he depicted the frequencies and durations of five sine tones with box and

triangular shapes (Figure 1.13). In *MIKROPHONIE I* (1964), the graphic score is divided into 4 rows (

Figure 1.14). The top row shows a pitch (high, middle, and low) and intensity. The second row indicates the microphone position from the sound source. The third and fourth row present the frequency bandwidth and the gain level [125]. Similarly, Krzysztof Penderecki's graphic notation *Threnody for the victims of Hiroshima* (1961) is a well-known example (Figure 1.15). It does not include any time signature or specify musical notes in a conventional way. Abstract shapes, black rectangles, dots, and lines denote sound dynamics, duration, and timbral characters [126].

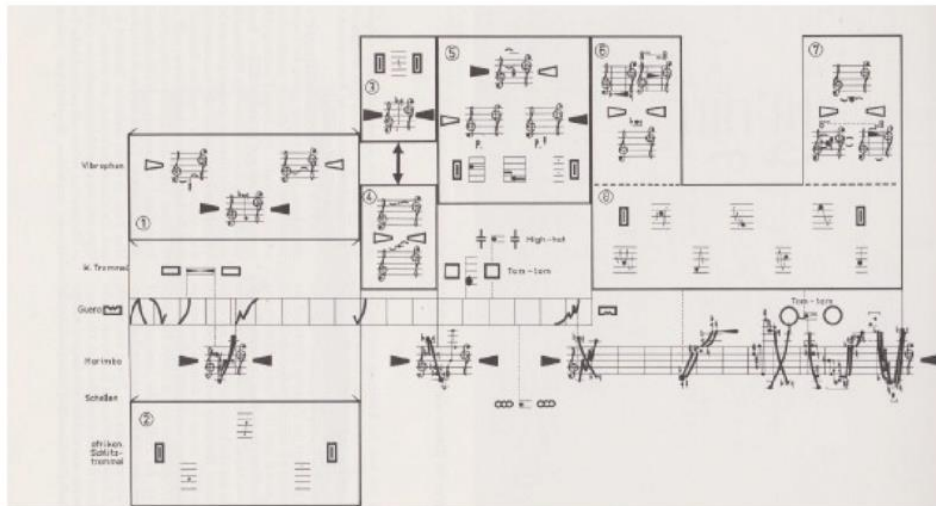


Figure 1.12: Graphic Score of Zyklus by Stockhausen

Source: Adapted from [127]

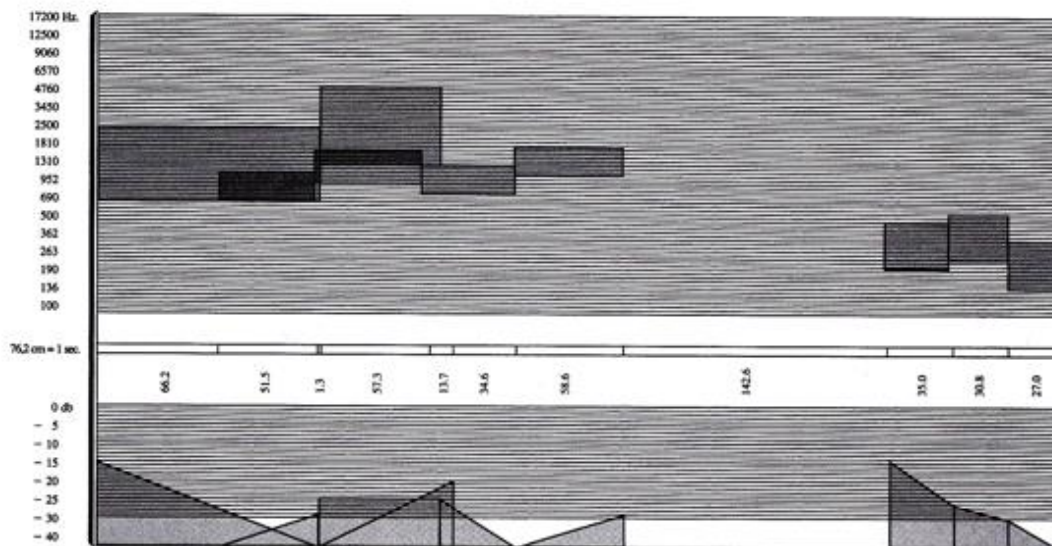


Figure 1.13: Elektronische Studie II by Stockhausen

Source: Adapted from [128]

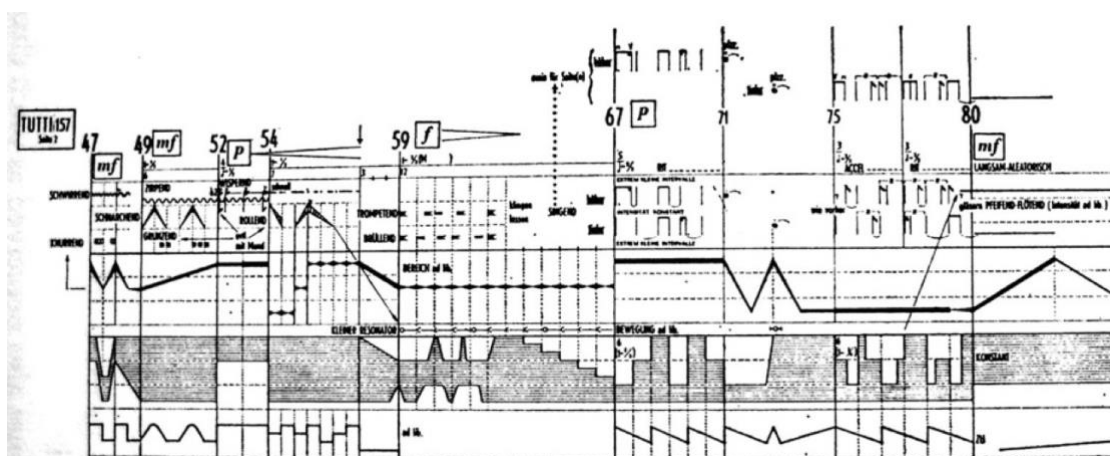


Figure 1.14: MIKROPHONIE I by Stockhausen

Source: Adapted from [125]

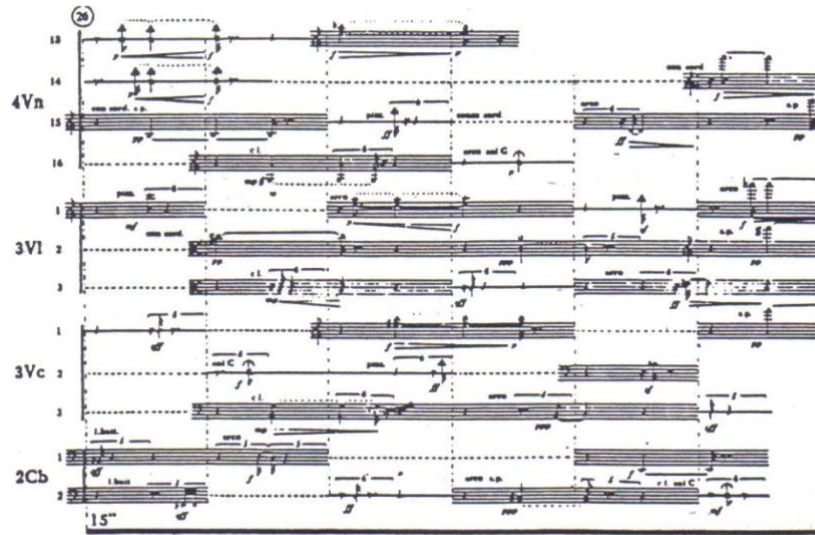


Figure 1.15: Threnody for the victims of Hiroshima by *Penderecki*

Source: Adapted from [129]

Cornelius Cardew wrote a visual score *Treatise* [130] between 1963 and 1967. *Treatise* was visually depicted with straight and diagonal lines, circular shapes, curves, abstract figures, and many other geometric shapes in 193 pages (Figure 1.16). Cardew (as well as Karlheinz Stockhausen) was an assistant of György Ligeti who is one of the most influential figures in contemporary music. Ligeti composed an electronic music piece *Artikulation* in 1958 and later in 1970, a graphic designer, Rainer Wehinger visualized it with a combination of colored dots, lines, oval shapes, and other graphical symbols [131]. Specifically, the visual symbols are horizontally arranged like a traditional music score, so a viewer can trace it from left to right while following the music.

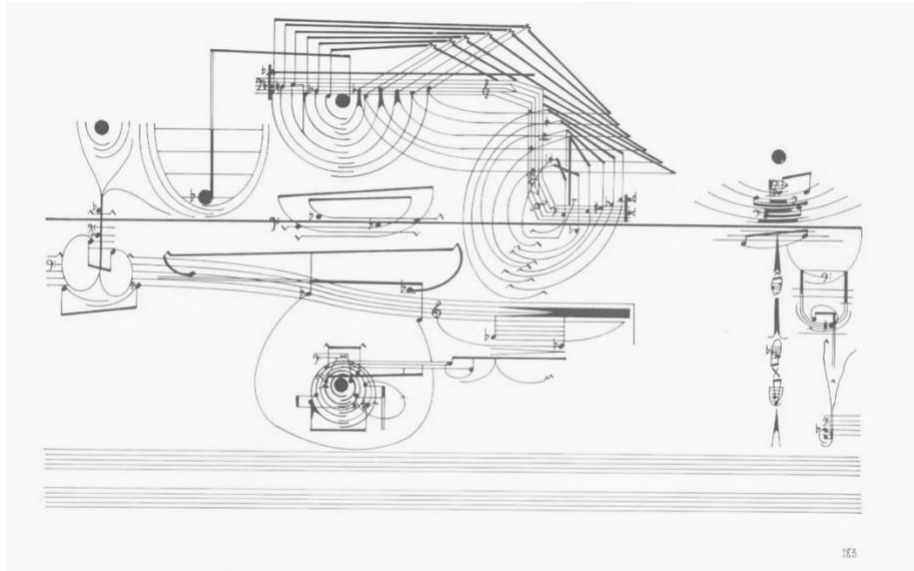


Figure 1.16: Treatise by Cardew

Source: Adapted from [132]

Another work that bears resemblance to traditional music scores is *Your Kindled Valors Bend* (1933), created by a Peruvian composer named Pozzi Escot. Instead of using abstract symbols, conventional music symbols (e.g., five-line score, musical notes) are mainly used, where the music staves are the shape of a triangle, circle, and straight line (Figure 1.17).

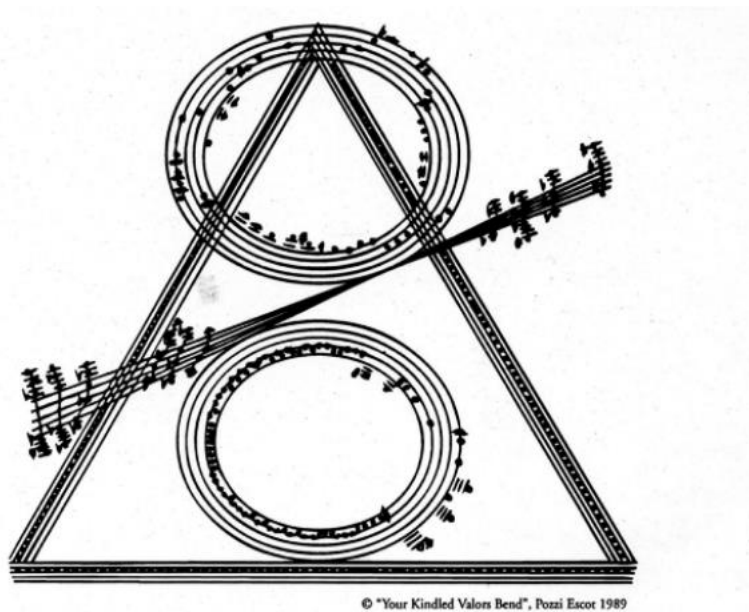


Figure 1.17: Your Kindled Valors Bend by Escot

Source: Adapted from [133]

Similarly, an American composer, George Crumb's score for *Twelve Fantasy-Pieces after the Zodiac for Amplified Piano* (1972-1979) is analogous in his use of shapes (Figure 1.18). Being influenced by American experimentalism, Mauricio Kagel's *Prima Vista* (1962-1963) is comprised of a series of visual scores that look like traditional 2D data visualization such as bar charts or line graphs (Figure 1.19). Cathy Berberian's *Stripsody* (1966) is a graphical vocal score that contains texts and hand-drawn icons arranged on three lines, in which the vertical space functions the same as regular music staves. The pitch of her voice sweeps up and down based on the height position. In the performance, Berberian speaks words and onomatopoeic words commonly found in comic strips (e.g., coughing: [comic depiction], clock ticking: [comic depiction], airplane: [comic depiction], and more) [134] (Figure 1.20).

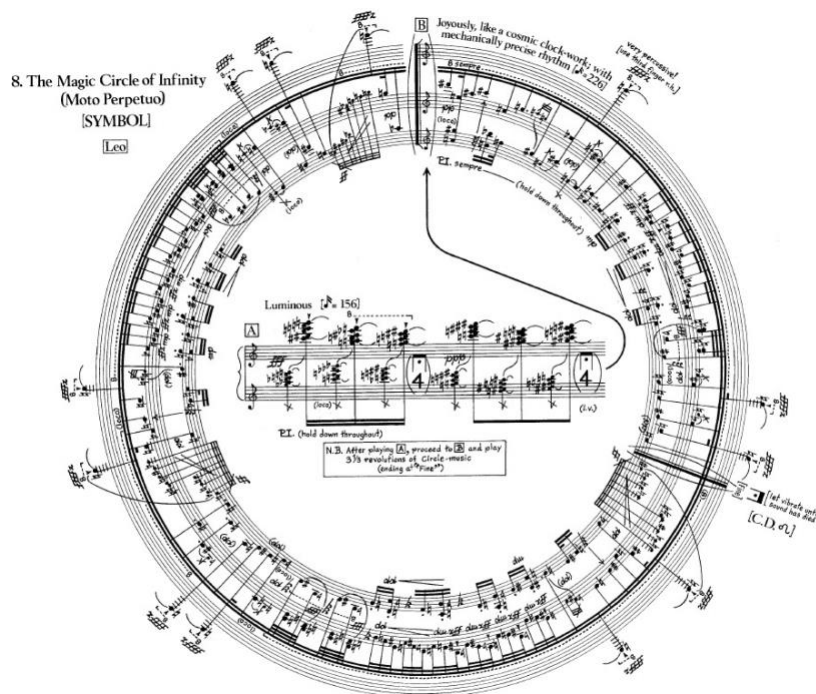


Figure 1.18: Twelve Fantasy-Pieces after the Zodiac for Amplified Piano by Crumb

Source: Adapted from [135]

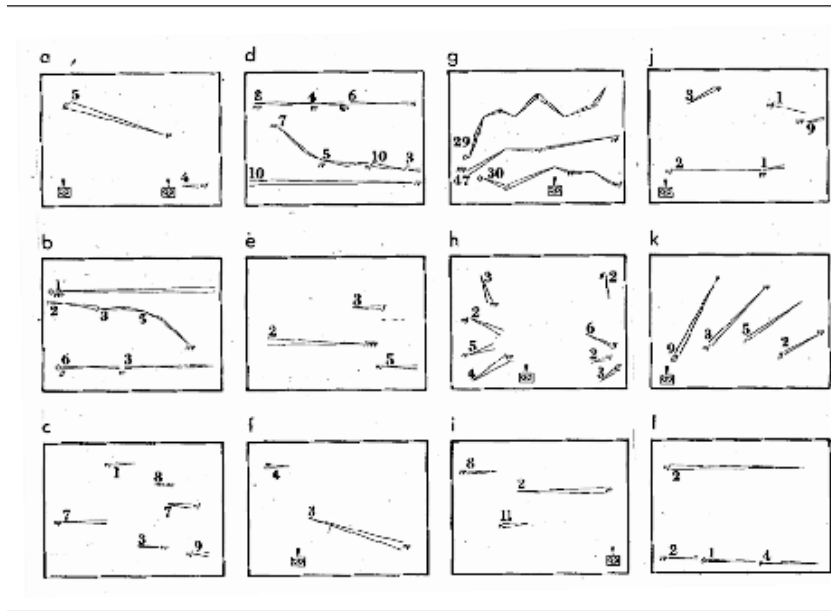


Figure 1.19: Prima Vista by Kagel

Source: Adapted from [136]

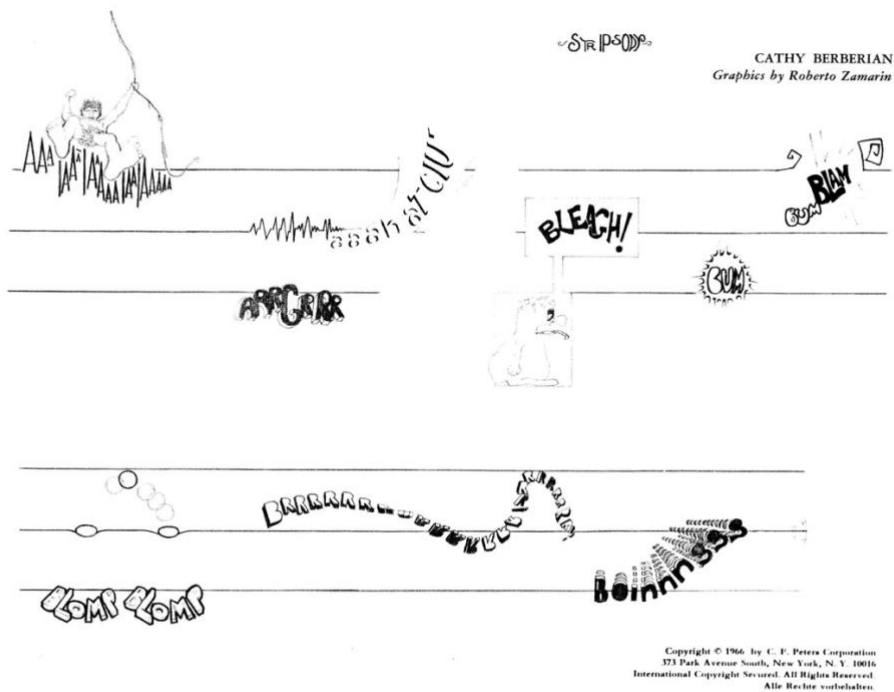


Figure 1.20: Stripsody by Berberian

Source: Adapted from [137]

Tom Phillips, an English artist, also music scores. In *Golden Flower Piece* (1966), cloud-like images or blurry marks surrounded the letters (Figure 1.21). According to the official description, the letters can be read as either in natural keys or flat/sharp keys. The uppercase letters should be in bass and the lowercase in a higher register. The micro grains around the letter indicate loudness and duration of the notes [138]. In his *Ornamentik* (1968), a score for trombone, abstract symbols that look like pictograms or signage icons are listed. These icons represent held, sustained notes, chords and sounds [139] (Figure 1.22). Brian Eno's famous work, *Music for Airports* (1978) received attention in part because of interesting graphic notations on the back cover of the album, four total, one for each track (Figure 1.23). These do not represent the performance or instructions for performers like John Cage's graphic notation. Instead, his visual score notates rhythmic structure, phrases and tape loops [140]. A Danish composer, Carl Bergstrom-Neilsen expressed a process moving from heavy, dark sounds to light sounds with a spiral visual score in *Towards an Unbearable Lightness* (1992). This notation can be read by any number of players and played at different times [141] (Figure 1.24).

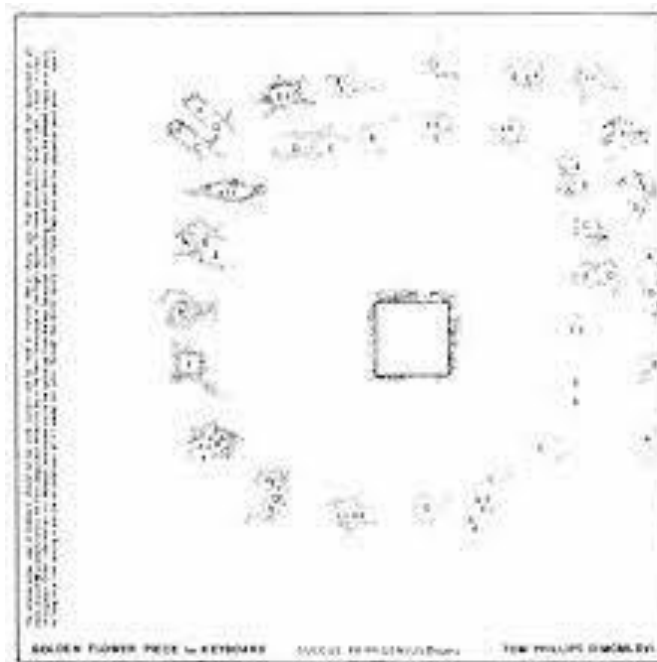


Figure 1.21: Golden Flower Piece by Philips

Source: Adapted from [142]

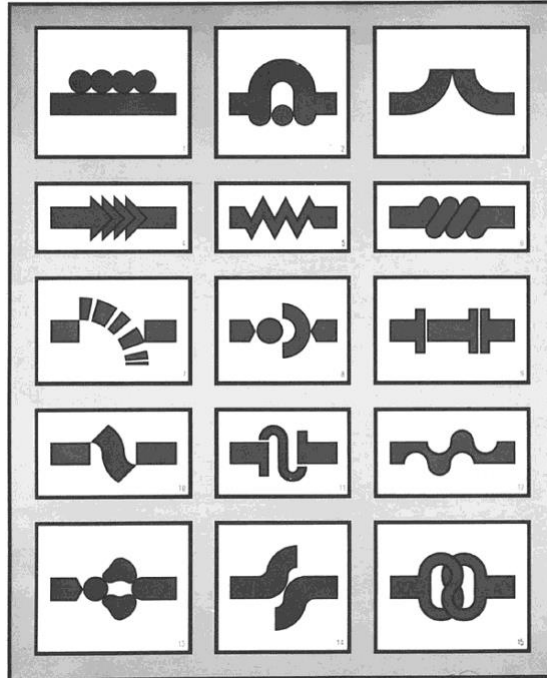


Figure 1.22: Ornamentik
 Source: Adapted from [139]

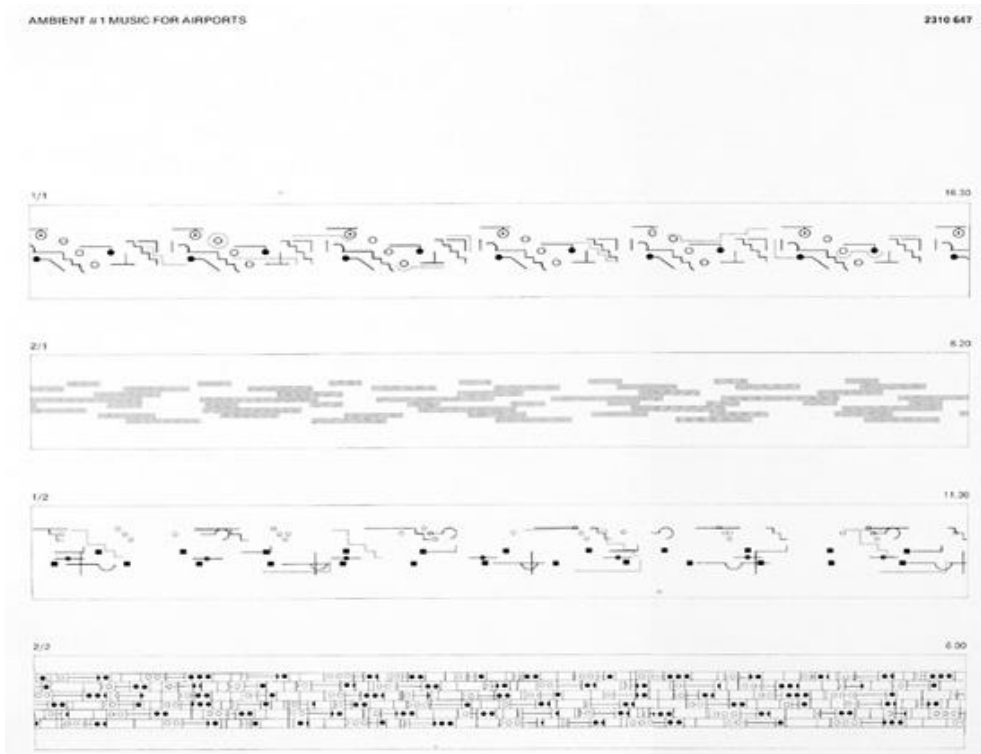


Figure 1.23: Music for Airport by Eno

Source: Adapted from [143]

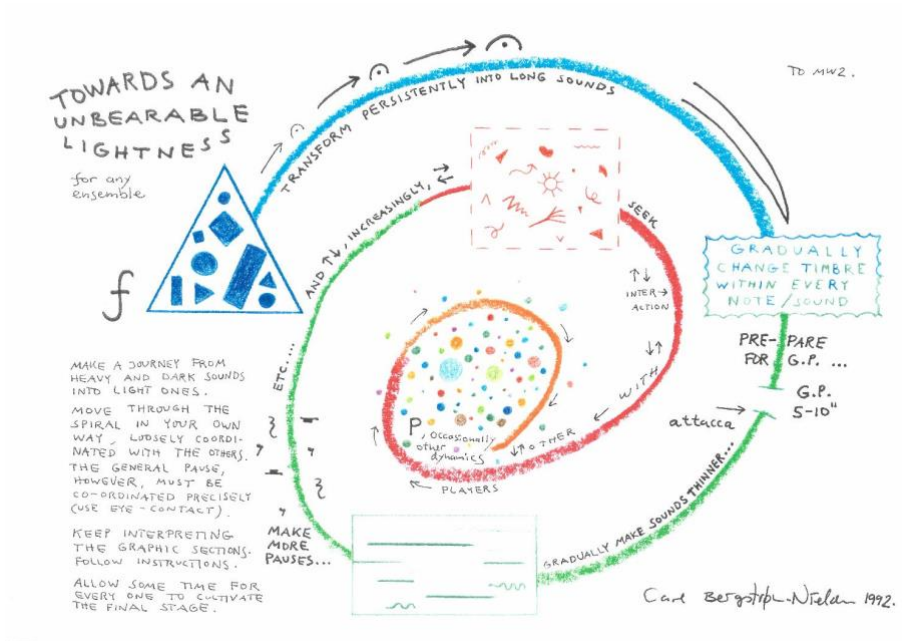


Figure 1.24: Towards an Unbearable Lightness by Nelson

Source: Adapted from [144]

An English bass player, Barry Guy desired to express music with visuals. In his album *Bird Gong Game* (1994), Guy used his graphic notation like a card game. Each card has a unique symbol that came from a Scottish visual artist Alan Davie's work and every time the cards are shown to the performers, they formulate a performance method leading the players [145] (Figure 1.25). Makoto Nomura's *Shogi composition* (1999) is an interesting approach in the way that people (not players) can participate. With color pencils, people draw something from their imagination, and when it is registered, players can start to play the score until the next one is assigned. Each player is supposed to select one color at a time [146].

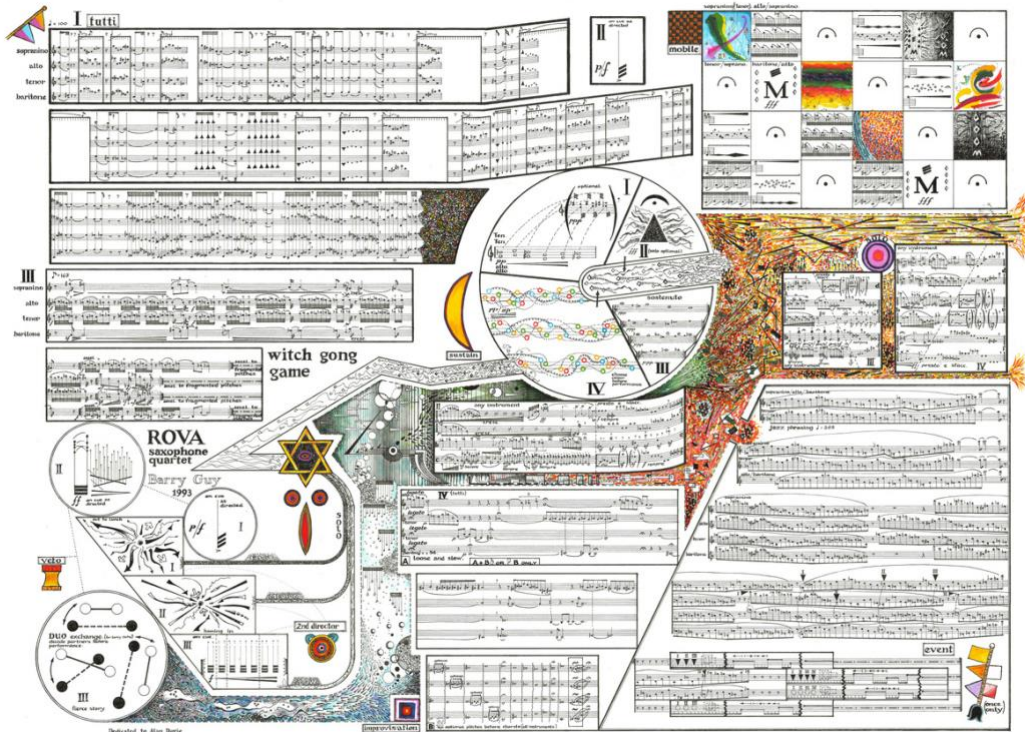


Figure 1.25: Bird Gong Game by Guy

Source: Adapted from [147]

To list examples after the 2000s, a jazz musician, Anthony Braxton is dedicated to graphic score. In *Composition No. 14*, basic figures such as straight lines, dotted lines, arrows, rectangles, and parentheses are arranged on paper (Figure 1.26). A jazz band, Thumbscrew, interpreted this visual score logically and recorded the album *The Anthony Braxton Project* (2000) [148]. Jazz trumpeter Wadada Leo Smith's album *Luminous Axis* (2002), written for trumpet and electronics, included a realization of his visual score (Figure 1.27). "An Electronic Sonic Garden of Delights and Transformations" is the subtitle of his album and describes what the drawing implies. In his *Sonic River* (2014), collaborating with an avant-garde jazz musician John Zorn, a similar drawing is found on the cover art. Argentine composer Martin Loyato's *Celestial Spheres Fantasy for Improvisers* (2005) was graphically notated for fourth performers. This visually appealing piece looks like posters found in space science museums with fifty symbols and eleven instruction pages. The artist used the universe as a musical metaphor. The line types and colors indicate tempo and dynamics of sounds, and each planet is assigned to a

different musical instrument (e.g., mellophone-Venus: piccolo-Uranus: flugelhorn-Mars)
The score was designed for performers to play in a circular way. This circular type is commonly found in the examples listed above [149] (Figure 1.28).

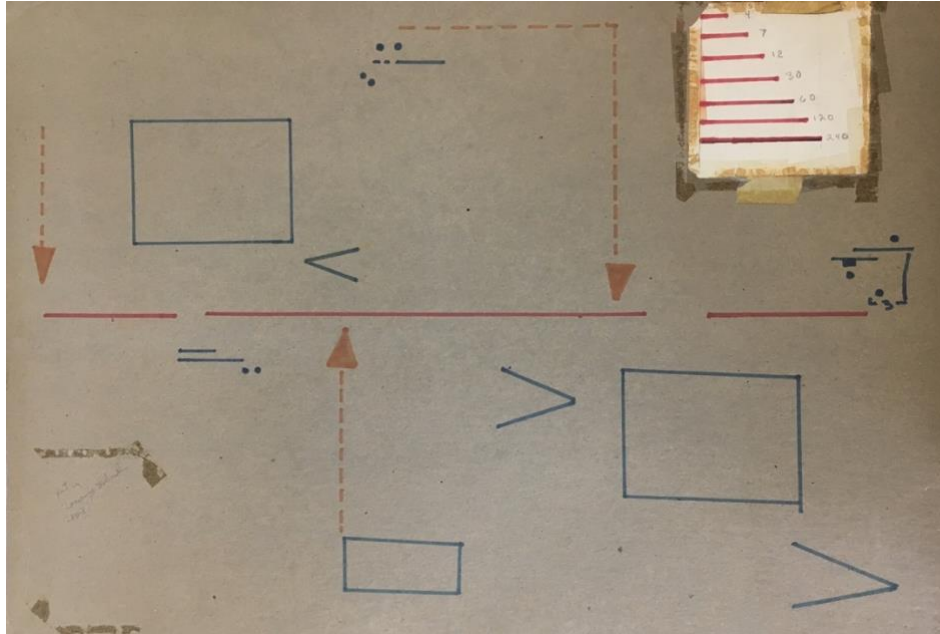


Figure 1.26: Composition No. 14 by Brexton

Source: Adapted from [148]



Figure 1.27: Luminous Axis by Smith

Source: Adapted from [150]

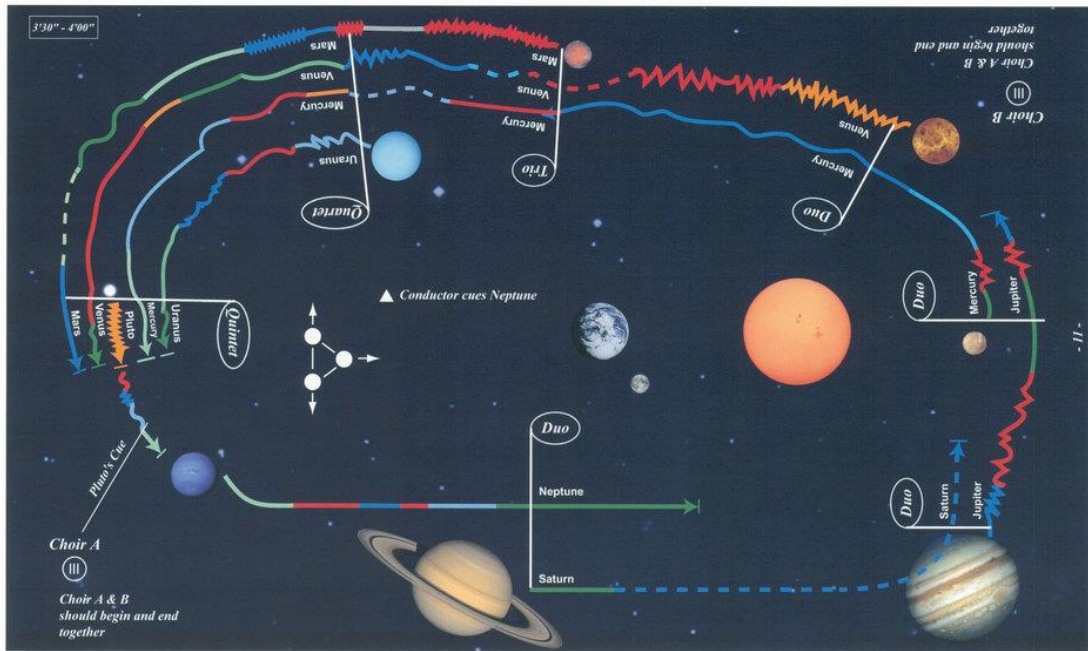


Figure 1.28: Celestial Spheres Fantasy for Improvisers by Loyato

Source: Adapted from [151]

Cilla McQueen’s *Picnic* (2006) was written for violin, oboe, and bass, and in terms of its visual aesthetic, seems to have a similar look to Smith’s visual score. Each line indicates a different musical instrument, and colors and shapes tell how the music should be played. The description of *Picnic* is very poetic: “Two violins, the first smelling of roses, the second holding a sword. The setting sun glints on the edge of the blade. Both violins describe an arc: the first like lips, the second like a slice. An enquiring sound upwardly inflected” [149] (Figure 1.29). In a detail of Steven Roden’s *Pavilion Score* (2005), Roden drew an exhibition site, a pavilion that the gallery owned, and filled it with colors, mapped into musical notes (e.g., big C is red, small C is pink) (Figure 1.30). This score takes the form of planning drawings that architects create [152]. In *Chollobhat* (2007), a jazz musician, Daniel Schnee, notated his musical score to look like an alien’s pictograph in a sci-fi movie. He described his idea as “cognitive ecology”, which links the visual score and music [153].



Figure 1.29: Picnic by McQueen

Source: Adapted from [154]

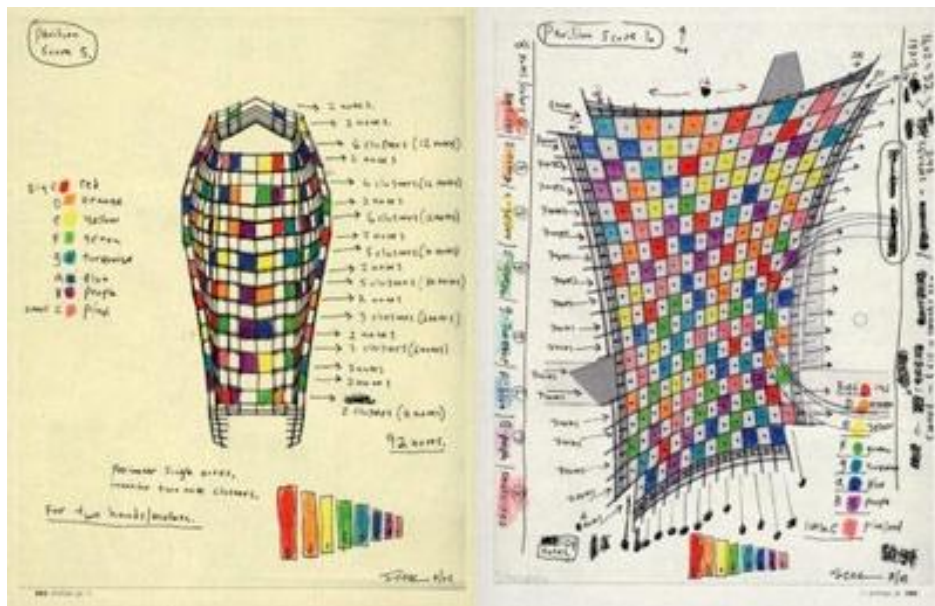


Figure 1.30: Pavilion Score by Roden

Source: Adapted from [155]

Graphic notations exemplified in the previous works are mostly printed or hand drawn—they are not digital. Last but not least, Stephen Malinowski’s Music Animation Machine (MAM) shows an example of a graphic score generated by computer technologies. MAM

(2005) took information from MIDI messages and simply created visual scores filled with basic shapes with colors moving horizontally (like a timeline of DAWs). MAM was featured in Bjork's *Biophilia* album project in 2011 [156]. *Biophilia* will be mentioned in Chapter 1.4.4.

Graphic score in general does not refer to direct synchronization between visuals and music. Rather, it can be seen as an indirect art description through artists' inspiration. In the next chapter, visual music and abstract film will be covered and will direct mergence between visual and sound.

1.4.3 Abstract Film and Visual Music

Animation and film as a medium for synesthetic art was the ultimate dream for synesthetic art for hundreds of years. Although the convenient creative tools like the current digital technology did not exist, there were significant movements among film and animation makers to turn imagination into reality. Through this achievement, synesthetic art that had never been possible before was created, opening many creative possibilities. Visual music has had a very significant impact on art history and can be seen as a genre that is the progenitor of contemporary audiovisual art.

Arnaldo and Bruno Ginanni Corradini are forerunners of abstractionism and the abstract film movement. In *Chromatic Music* (1912) [157], they attempted to paint directly on 16mm film [158]. The red star rotates, trembles, and is gradually zoomed in, and the pen-drawn green background continues to shake based on the sound. A painter, Leopold Survage presented *Colored Rhythm* (1913) [104] which was considered a new form of art: painting in movement. Hans Richter advanced a traditional form of painting, which emphasized rhythm and the movement of an object by creating abstract films [159].

Walter Ruttmann, a German painter and film director, colored film by hand and paved a way for experimental film or visual music going deep into the possibilities of film technology at that time. In 1919, he wrote manuscripts for *Malerei mit Zeit* (painting with time), dreaming of making something between painting and music [160], and this idea

was realized in *Light Play Opus 1* (1921) (Figure 1.31). This work is known as the first abstract film screened in public and the series continued to *Opus IV* until 1925. Oil painted curtain-like shapes filled with mystical gradient colors flutter along with violin music of Max Butting [161]. Presented in Frankfurt, this is the first colored abstract animation in film history [162]. A series of his work mainly focused on the movement of light patterns and music. Later, this creative achievement led to the work of Viking Eggeling, Hans Richter, and Oskar Fischinger.



Figure 1.31: Light Play Opus 1 by Ruttmann

Source: Adapted from [163]

Viking Eggeling, a Swedish filmmaker, made an abstract film *Symphonie Diagonale* (1924). He drew abstract figures with a pencil on paper and used tinfoil masks to partially or entirely hide and show the abstract figures, and then photographed every frame. In *Symphonie Diagonale* (Figure 1.32), a black and white film, a pencil color was reversed, shown as white. Eggeling collaborated with Hans Richter to make abstract films using horizontally or vertically long canvases and exploring motion sequences of the shapes [88]. Eggeling was fascinated by the visual form and how they can evolve over time. To the contrary, Richter focused more on the rhythm and the movement of basic forms such as lines, squares, and rectangles [162]. In his *Rhythmus 21* (1921-1924) (Figure 1.33),

various movements of white squares and rectangles are animated expressing a sense of rhythm with the elements. Richter described it as “a rhythmical movement regulated by itself in which variations and pulsations form a part of the artistic design” [164].



Figure 1.32: Symphonie Diagonale by Eggeling

Source: Adapted from [165]

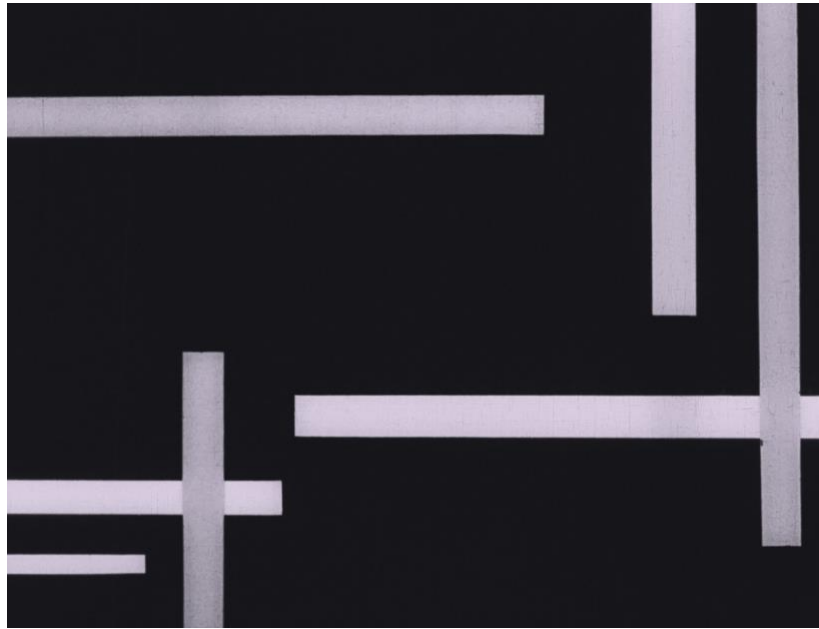


Figure 1.33: Rhythmus 21 by Richter

Source: Adapted from [166]

When Walter Ruttmann’s *Light Play Opus 1* was premiered in 1921, Oskar Fischinger witnessed the show and was greatly influenced by it [167]. He synchronized images with phonographs and produced numerous works of art using various techniques. For example, he created a wax-slicing machine for his film *Experiment with Wax* (1921-1926) [168]. The camera in his machine took a photo of layered wax once every time a

slicer cut the layered wax blocks [169]. In his three minutes short film *Allegretto* (1936-1943) (Figure 1.34), oval shapes spread out in the background and rhombus shapes, curved rectangles, graphic equalizer-like shapes, and wiggly lines filled with colors fly around the screen. Each graphic symbol dances along with Ralph Rainger’s music [170]. In 1936, Fischinger moved to Hollywood to avoid the Nazi regime and was given a chance to work for MGM studio and Paramount Pictures, eventually making an animation sequence for Disney’s *Fantasia* [171]. He made a huge impact on early abstract film by inventing devices and implementing innovative techniques. He said “Visual Music emphasizes the effect of music. It is to music what wings are to birds” [171]. He was a visual music master whose works interplay between shapes, colors, and musical accompaniment.



Figure 1.34: Allegretto by Fischinger

Source: Adapted from [172]

Disney’s *Fantasia* (1940) was influenced by the flow of the visual music and abstract film movement, and it connoted synesthetic associations that are commonly found in synesthesia. Famous orchestral music pieces such as *The Nutcracker Suite* (Tchaikovsky), *Toccata and Fugue in D Minor* (Bach), *The Rite of Spring* (Stravinsky), *The Pastoral Symphony* (Beethoven) were adopted for the *Fantasia*’s opening scenes and various colors illuminated behind the shadow of the orchestra. To be more detailed,

different color codes fit with music and different timbres/instruments and a luminous perception (photism) connected music and visuals (e.g., smooth lines with graceful music, bright colors with high pitch sound and fast tempo), demonstrating *Fantasia*'s synesthetic approach [89]. However, no clear evidence was found that *Fantasia* was created by artists who experience synesthesia [173].

Like Fischinger, Norman McLaren's experiments with drawing on film to create a soundtrack for his animation and motion pictures was groundbreaking [174]. His work *Begone Dull Care* (1949) is highly respected as a new level of direct painting on celluloid film [175]. He said that "animations express the spirit of music" [162]. One of his early works *Dots* (1940) shows that he dreamt of making a 1:1 relationship between image and sound. The movement of blue dots on a red background follows in accordance with the sound. His *Synchromy* (1971) (Figure 1.35) is a great example of how McLaren used graphical sound and produced visual music that shows precise synchronization between sound and image [176]. He drew rectangular and line shapes on short strips of card. The width and distance between shapes created a different pitch when they were photographed directly onto the film soundtrack [177]. The sound was created through optical soundtracks. He found a way to express different timbres, portamento, glissando, and vibrato with them [162].

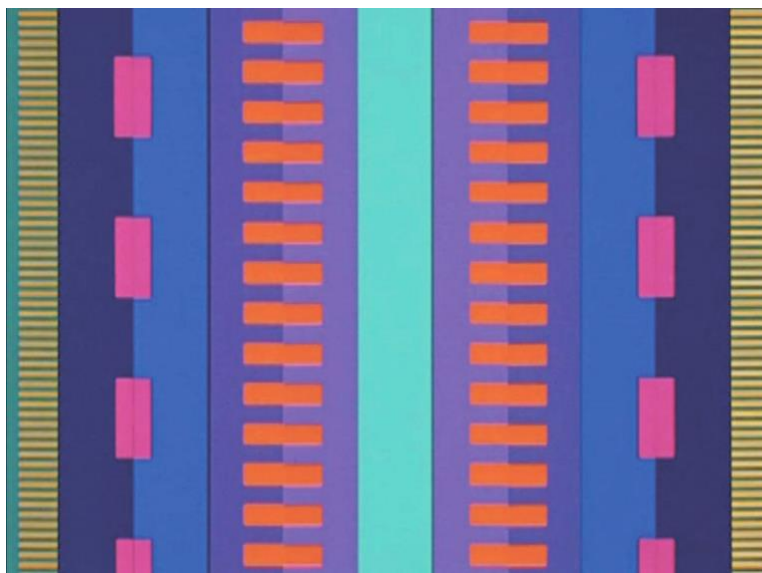


Figure 1.35: Synchromy by McLaren

Source: Adapted from [178]

John and James Whitney produced their visual music pieces in both analog and digital realms. They invented a novel sound experiment system using pendulums. The movement of pendulums (naturally sinusoidal) was recorded onto a film soundtrack and generated sounds up to five octaves [162]. Pendulums worked as a sine wave oscillator. Their serial work *Five Film Exercises* (1943-1944) created the visual images by using light through masks, so the camera captured direct light on a film [179]. John Whitney is considered the “father of computer graphics” [180]. He collaborated with Saul Bass for Alfred Hitchcock’s 1958 movie *Vertigo* [181] and their kinetic typography and animated figures were groundbreaking, becoming a symbol of the movie. His work also affected “The Stargate Corridor” movie sequence in Stanley Kubrick’s classic masterpiece *2002: A Space Odyssey* (1968) [182]. When he worked with IBM later, he brought his know-how to the digital world. He emphasized great potentials for the relationship between visual and auditory domains [183]. He left analog machines and moved to digital processes for better and faster production of his complex work. This area included digital projection films such as *Side Phase Drift* (1965) showing the harmony of abstract and fundamental shapes and *Arabesque* (1976) with the visual of Lissajous forms. Whitney’s works are highly regarded as “visual music brought into a new place because of all the visual and auditory elements (e.g., shape, color, movement, sound) closely synchronized [88].

In another case, from 1934 to 1953, Mary Ellen Bute created a series of her visual music *Seeing Sound*, including her fourteen short films [184]. Inspired by Whitney, Hy Hirsh used oscilloscopic images with his homemade optical printer [185] in his films such as *Divertissement Rococo* (1951), *Eneri* (1952), and *Come Closer* (1953). Len Lye reinvented a drawn-on-film technique using etching stencils applied to pigment onto a celluloid film, before it completely dried. With this technique, he produced *Colour Box* (1935) [186]. Like Whitney’s approach, Jordan Belson’s and Harry Smith’s works were inspired by Asian metaphysics and meditation. In Belson’s *Allure* (1961), he used both traditional and new animation techniques, such as optical printing and lasers [187].

Belson tried to capture natural phenomena of geometric transformation and Smith attempted to present improvisation in jazz music visually [188].

Examples mentioned above continue to influence contemporary visual music artists now. Canadian film maker Steven Woloshen's work has been produced without a camera, mainly based on his musical inspiration from classical music and jazz – Erik Satie [189] and Dave Brubeck [190]. A German artist, Robert Siedel used a wide range of data such as 3D scans, VR-modeler designed for cars, and MRI data from his body to create his immersive audiovisual/data sculpture projection *_grau* (2004) [191]. A Japanese film director, Keita Onishi created a set of twelve minimalistic animations comprised of a black background, white dots, lines, and 3D cubes. His *Forest and Trees* (2011) was displayed on small digital monitors in the exhibition space. He expressed the impression of the sound with the animated visualization [192]. Robbie Lynn Hunsinger showed a world that visual music can be mixed with computer technology and interactive art. In *Constellation* (2016), participants can control six animated geometric shapes mapped with sounds -voice, synth, triangle wave, oboe, and kalimba- to body movement. Today, this cross-genre (the fusion of different art forms) continues to be an interesting subject, widely used for creative expression, coupled with fast changing computer technologies.

Visual music is probably the most accurate and comprehensive term to define my audiovisual work, which will be introduced in Chapter 6. Through sonification, my art form pursues the same artistic significance that early pioneers tried in the analog way.

1.4.4 Multi-Modal Musical Instrument

A curiosity in the relationship between color and sound can be seen through not only an art form but also musical instruments. These ideas have largely been attributed to artists' eagerness for experimentalism, a driving force that expands conventional art forms. In the early to mid 1770s, a French scientist, Louise Bertrand Castel, invented *Clavecin Oculaire* (*the Ocular Harpsichord*). The *Ocular Harpsichord* has a keyboard in front and sixty windows covered by colored glasses (Figure 1.36). Each glass has a curtain and each window has light behind, so when a player pushes a key, a curtain opens, a note

rings, and light illuminates from one of the sixty glass covers. Unfortunately, no detailed examples are left [193]. The inventor claimed that even a deaf audience enjoyed music with this instrument [194]. In 1873, a French scientist, Frédéric Kastner, put flames inside a pipe organ. Hydrogen flames emitted through the glass tubes along with a musical note produced [195]. This device was called the *Pyrophone* (Figure 1.37).

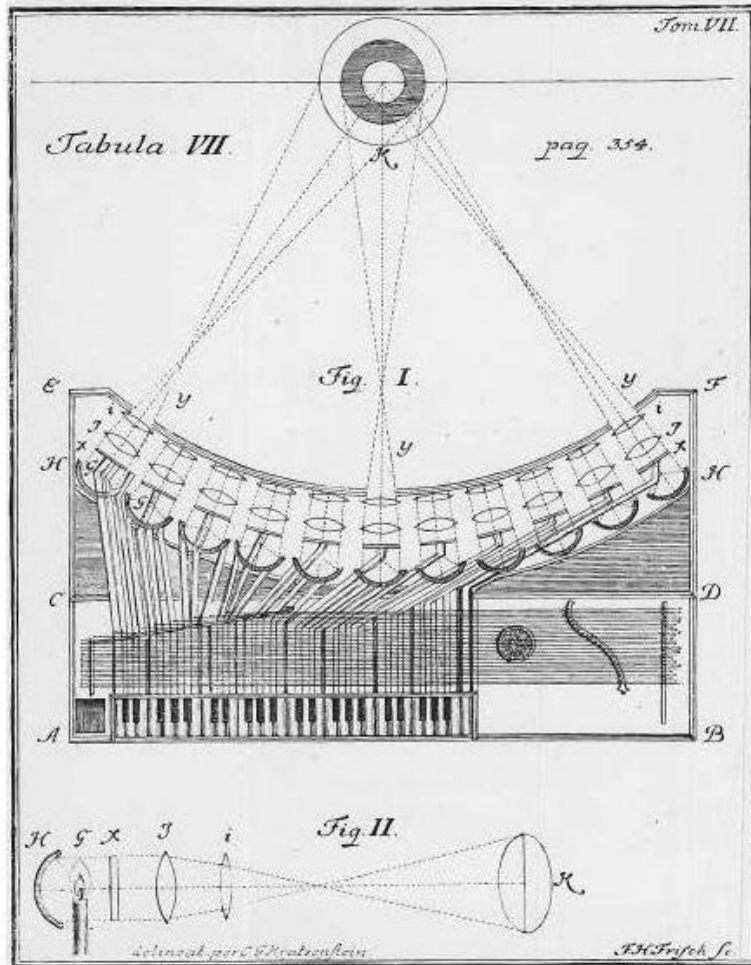


Figure 1.36: Ocular Harpsichord

Source: Adapted from [196]

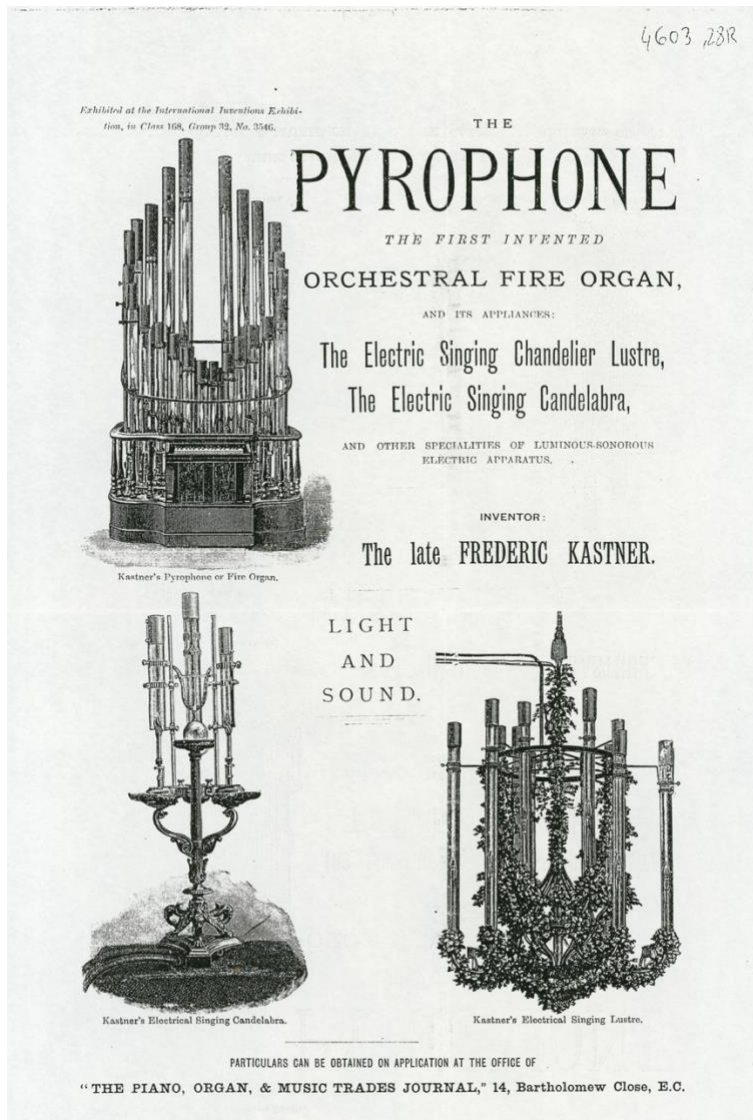


Figure 1.37: Pyrophone

Source: Adapted from [197]

Early multi-modal musical instruments are divided into two types: 1) drawing-based, and 2) light, disc-based. Like the sound-on-film techniques previously mentioned, an early glass disc instrument was the *Piano Optophonique* [198] in 1910. Glass discs were read by passing white light through lenses [199][200]. Mary Hallock-Greenewalt created a color organ called *Sarabet* [201] (Figure 1.38) that was driven by seven-colored lights, receiving eleven patents in the 1920s [202] [203]. Kurt Schwerdtfeger, who studied at Bauhaus under famous artists, such as Johannes Itten and Paul Klee, realized a projection device *Reflektorische Farblichtspiele* (Reflecting Colored Light Plays) with Josef

Hartwig [204]. This device lit movable cardboard, cut in various shapes. It was introduced in 1922 at Kandinsky's loft [205]. Later, together with Ludwig Hirschfeld Mack, they continued to develop *Farblichtspiele* (Colored Light Plays). This projection device projected colored lights through moving stencils. In the book *Color Cinematography* [206], Adrian Bernard Leopold Klein described his instrument that emits colored lights triggered by keyboards in two octaves. Thomas Wilfred, a pioneer of light art called Lumia, introduced *Clavilux* in 1922, which translates to "Light played by key" [207]. Later, his work was used for Rimsky-Korsakov's *Scheherazade* [208]. In 1925, Hungarian musician Alexander László created *Sonchromatophone* in collaboration with Oskar Fischinger, which generated colored light [209]. These types of instruments that make sound with visual media are called "Color Organs" [210]. In 1930, Evgeny Sholpo developed a photosynthesizer (an optical synthesizer) called *Variophone* [5] (Figure 1.39), with Rimsky-Korsakov's support. Sound waves were carved into cardboard tone discs and by rotating the discs, the *Variophone* created sounds in tandem with a 35mm movie film. Under the Miltzvuk group in Soviet Russia, ornamental soundtracks were researched. For example, Nikolai Voinov developed his "paper sound technique" and created a short animation for Rachmaninov's *Prelude* in 1932 [211] and *The Dance of the Crow* in 1933 [212]. From 1937 to 1957, these movements merged into the *ANS synthesizer* [213] (Figure 1.40), invented by Evgeny Murzin. This instrument visualizes sound waves and synthesizes sounds from the spectrogram. In the same period in Germany, a study on hand-drawn sound synthesis was conducted by Rudolf Pfenninger. *Oramics* [214] (Figure 1.41), invented by Daphne Oram in 1957, had a similar approach as *Variophone* because it turned a particular shape into pitch, timbre, and intensity of sound through 35mm film. Gordon Pask's *Musicolour System* [215] in 1958 was also an exceptional piece and took a completely different route than previous examples because his system generated light initiated by sounds. Later, his interactive art installation took part in a famous multimedia, cybernetic art exhibition, "*Cybernetic Serendipity*," in London in 1968. As time passed, entering the age of the computer, this area naturally moved into generative, digital, multi-media, and interactive art.

Feb. 6, 1934.

M. H. GREENEWALT

1,945,635

LIGHT COLOR INSTRUMENT

Filed Jan. 29, 1927

6 Sheets-Sheet 4

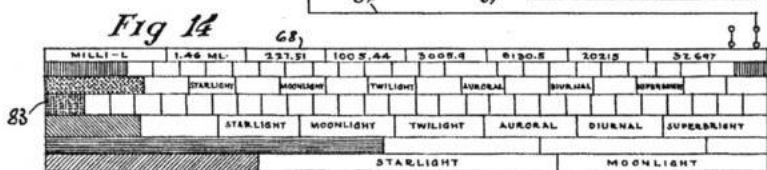
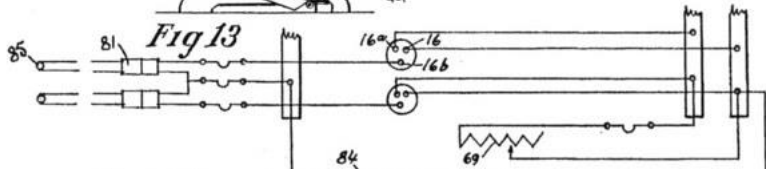
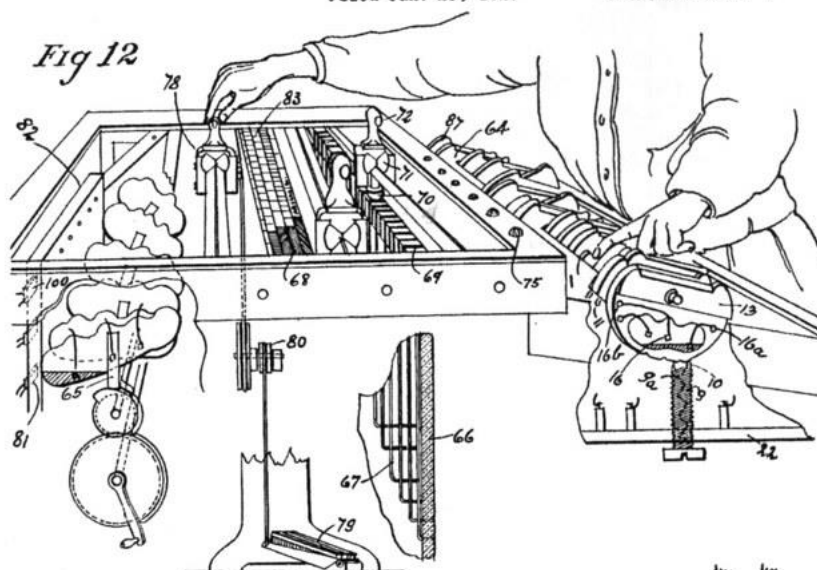


Fig 15



Mm.....=76
 By *Mary Wallocke Greenewalt*
 Inventor.

Fig 16

Figure 1.38: Sarabet

Source: Adapted from [216]

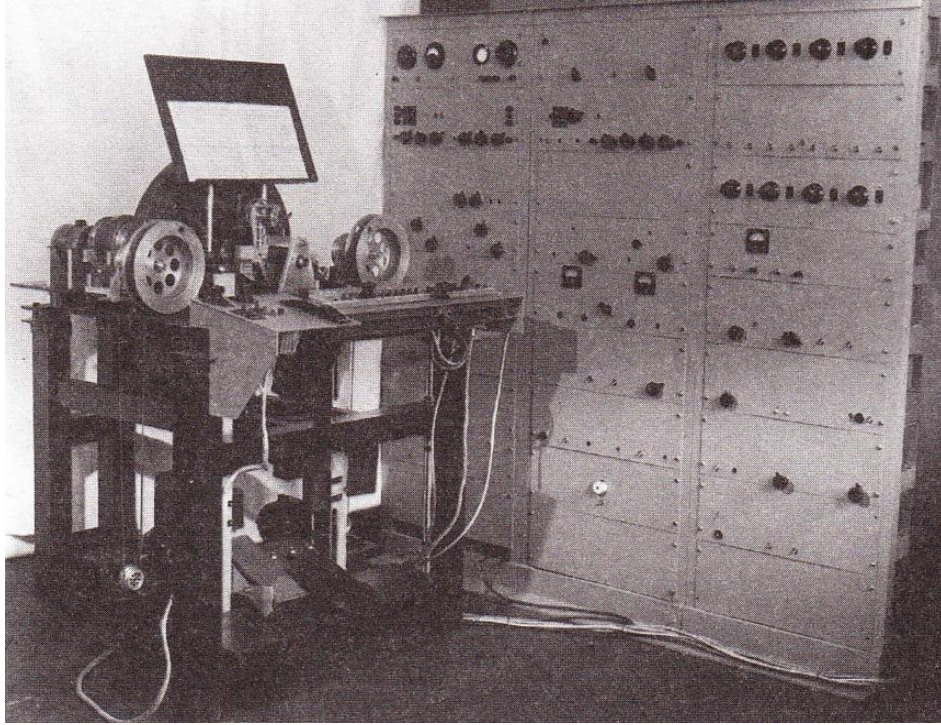


Figure 1.39: Variophone
Source: Adapted from [217]



Figure 1.40: ANS synthesizer
Source: Adapted from [218]

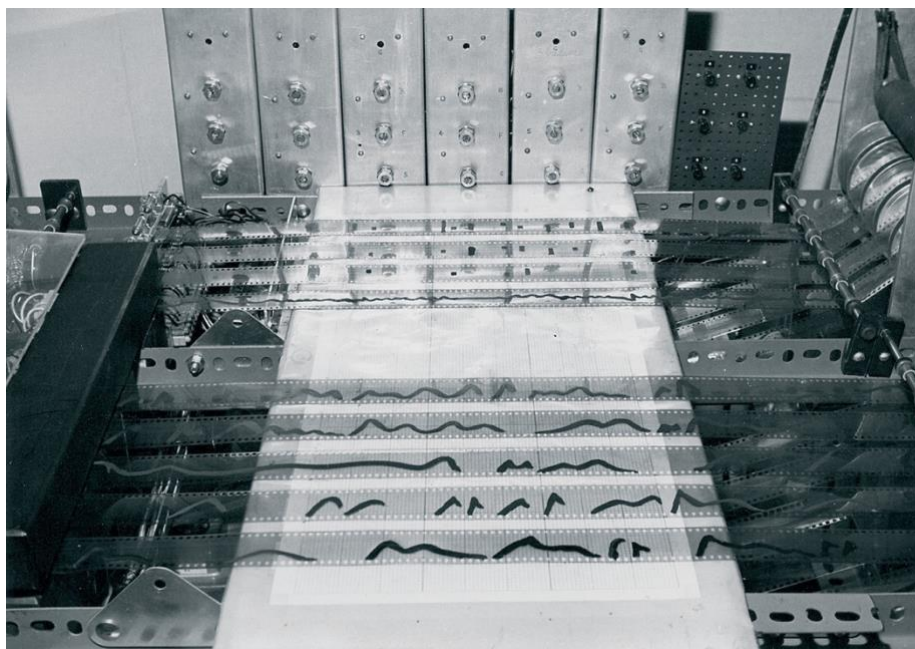


Figure 1.41: Oramics

Source: Adapted from [219]

Now it is common to see many multi-sensory software/hardware instruments that accommodate multimedia technology to stimulate our creativity and provide diverse user experiences. A sound designer, Yuri Suzuki, created the wheeled cubic vehicle to which the micro speaker is attached. This vehicle looks like a miniature toy car. In his work, *Colour Chaser* (2010), the user drew a line with a regular black marker on white paper, and this little vehicle followed this hand-drawn line making a continuous beeping sound [220]. When the user added other colors across the black line with sharpies and the vehicle passed by, RGB data turned into musical notes. The later version came with more cars with different sound types that corresponded to their names: the Basscar, the Arpeggiocar, the Glitchcar, and the Drumcar. A design studio, Playtronica [221], focused on creating musical instruments combining various different media. Their core design concept was to transform anything into sound. For example, in their *Vegetable Orchestra and Jelly Jam* [222], vegetable, fruit, and giant jellies played a role as a touchable MIDI key instrument working in a similar way to a fruit battery. A fruit battery naturally generates an electric current. Based on these ideas, Playtronica created installation art and interactive musical gadgets for musical expressions and educational purposes. Playtronica

widely adopted synesthetic design approaches transforming color into sound. *Orbita* [223] (Figure 1.42) is their new MIDI instrument that is currently planning a Kickstarter campaign. *Orbita* exactly resembles turntables and its spinning wheel has four white circular lines called “tracks”. When the color magnets are placed on the tracks and read by this device, MIDI data are sent to the DAWs connected via USB. The colors painted on the magnets send musical notes and each track on the spinning wheel can be mapped to any musical instruments inside DAWs, assigned to different MIDI channels. The innermost track only takes a drum channel in a fixed manner. *Orbita* is a turntable-like musical sequencer where the colors and the arrangement of the magnets turn into music based on the speed of the spinning wheel.



Figure 1.42: Orbita

Source: Adapted from [223]

There is an example of a multi-modal instrument that combines touch-based interactive visual art. *Aeolian Harp* [224] is a visual music-inspired music app that automatically generates ambient and relaxing sounds and motion graphics by tracking finger touch. The

gesture is input by touching the smartphone screen. The touched areas cycle in random or timed order, and thin strokes, which connect the positions of the finger touch, are formed and deformed. The sound loops in time with the cycling. Scott Snibbe's *Gravilux* [225] simulates a gravitational pull influenced by Newton's equations. The name came from the light instrument *Clavilux* that was previously mentioned in this chapter. The starting screen has a Cartesian grid system filled with micro particles. Through a finger touch, these microparticles pull or push each other, like dust or stars in galaxies. The particles also respond to music that users can play along with.

Musicians release multimedia apps that supplement their studio albums to provide an extended art expression and experience that traditional music media cannot accommodate. When mobile phone app stores opened, they offered a favorable environment for developers to sell their music apps. They provide easy accessibility to art content because anyone can instantly download apps to see or hear artwork, instead of going to art galleries or concert halls. One famous example is Björk's seventh studio album *Biophilia* (2011) (Figure 1.43). This album was released along with a multimedia app developed in collaboration with interactive designer Scott Snibbe. The *Biophilia* app consists of nine songs that each have their own unique visual music theme inspired by the relationship between musical structure and natural phenomena [226]. The graphic design elements of each visual music theme vary depending on the songs, so the music can be explored through the visual information. Each song of *Biophilia* provides a series of educational artworks with interactivity implementation so that users can interplay with the songs, manipulate the visual compositions with a finger touch, and learn about the music. *Biophilia* is considered the first "app album" and the visual aesthetic of the *Biophilia* app is reminiscent of early visual music.

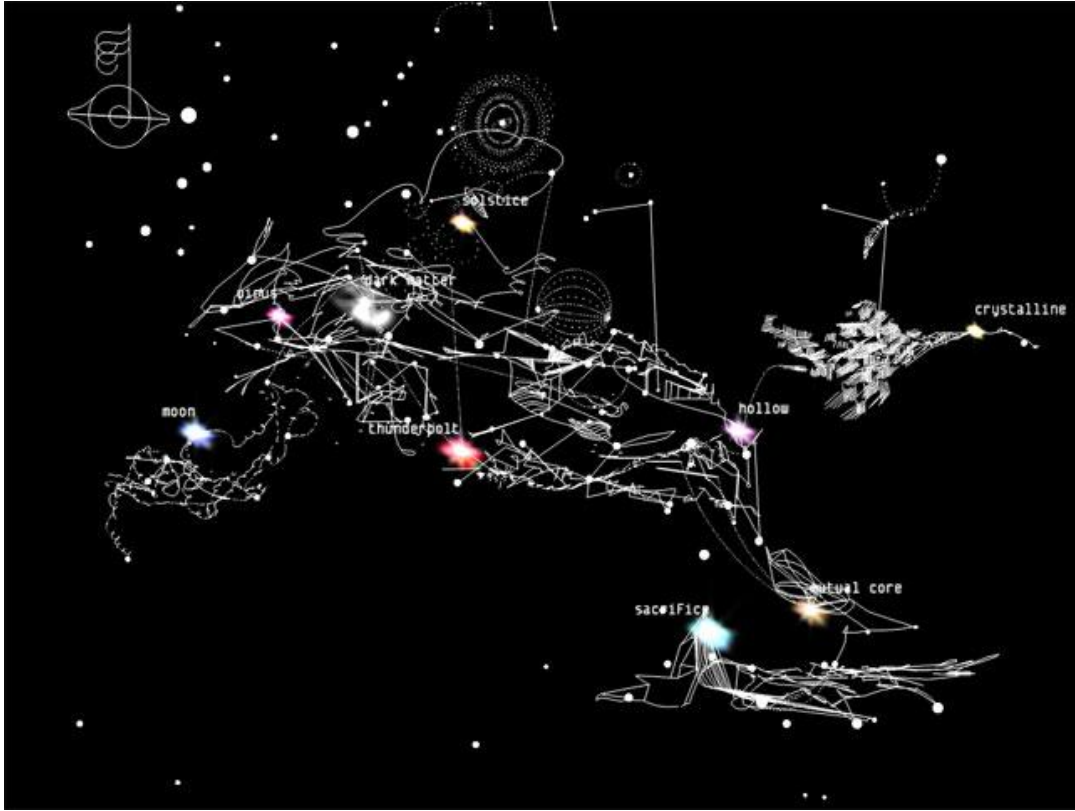


Figure 1.43: Biophilia

Source: Adapted from [227]

Polyfauna [228] is Radiohead's take on a visual music app. The first version of *Polyfauna* was released in 2014 by a design studio, Universal Everything [229], and the music was based on the band's *The King of Limbs* (2011) album. *Polyfauna V2* came out next and the audiotrack was later used for Thom Yorke's *Tomorrow's Modern Boxes* (2014) album. *Polyfauna* presents a series of audiovisual works in 3D that offer an augmented reality environment. This app starts with randomly selected songs and corresponding 3D backgrounds. As the songs flow, the virtual scenery moves forward as if the user is walking forward. The user's point of view changes when the smartphone is tilted and abstract graphic patterns can be drawn over the background with a fingertip. However, in contrast with *Biophilia* where music and animation are fully synchronized, *Polyfauna* does not offer interplay between the visual and sound. *Polyfauna*'s design approach is similar to interactive music videos where users can take part in the experience.

A South Korean music producer and singer-songwriter, Sumin, released her single as an interactive music app called *Fightman* [230] in 2021. Like *Polyfauna*, users' perspective in the *Fightman* app moves forward as the soundtracks play. However, a moving character in a fluctuating oval shape appears in the center of the background and advances through a virtual environment while the user follows on screen. This virtual character reacts to a finger touch. Interestingly, every time the character is tapped, it randomly generates pre-sampled sounds which are mixed into the original music, so users are involved in music-making like a DJ. The background graphics are interactive too. Users can find hidden tracks and features by swiping or touching them. In addition, the virtual character loops the sound repeatedly, when users shake their phones. According to the positions of this character, a noise or low-pass filter can be introduced, and users can tear the character apart with two fingers. By doing so, different sound effects such as sound delay can be applied to the original music

A multi-modal design approach increases the convenience of the musical product usability or provides a multi-sensory user experience. The musical instrument company Roli introduced LUMI keys [231] in 2019 (Figure 1.44). Unlike conventional MIDI keyboards, LUMI keys has illuminated keys like color displays. Interestingly, this product is similar to Scriabin's *Clavier à lumières* (a light keyboard). However, Lumi keys' light keys do not stem from synesthetic phenomena between color and sound, where Scriabin's musical metaphor originated. The distinguishing factors of Lumi keys enhance the functionality because the light keys are designed for user interface. For example, music performers on the dark stage can easily see the keys and the vivid and showy key colors provide a perfect foil for the live performance. Not only that, but Lumi keys offers convenience for better keyboard playing performance, music composition, and education because the various color modes of the illuminating keys help users identify root keys, scales, and chords by following the keys that light up. Lumi keys' color-assisted design approach acts as visual guidance for musicians by lighting up specific keys on the keyboard. New York-based musical instrument company Critter and Guitari's EYESY [232] is a video synth with a little orange box that contains a series of

8-bit graphic-style visual patterns (Figure 1.45). When music or any kind of sound is input to the device, it immediately makes the graphic elements move in response to the intensity of the sound, acting like a music visualizer. Beyond those cited, there are more multi-modal instruments than are possible to be listed in this introductory chapter. Multi-modal instruments are ubiquitous and have become a new genre of instrument design distinguished from typical musical instruments. Later in Chapter 2.2.3, examples of musical instruments that are specifically relevant to visual-sound translation, starting with the classic UPIC, will be discussed.

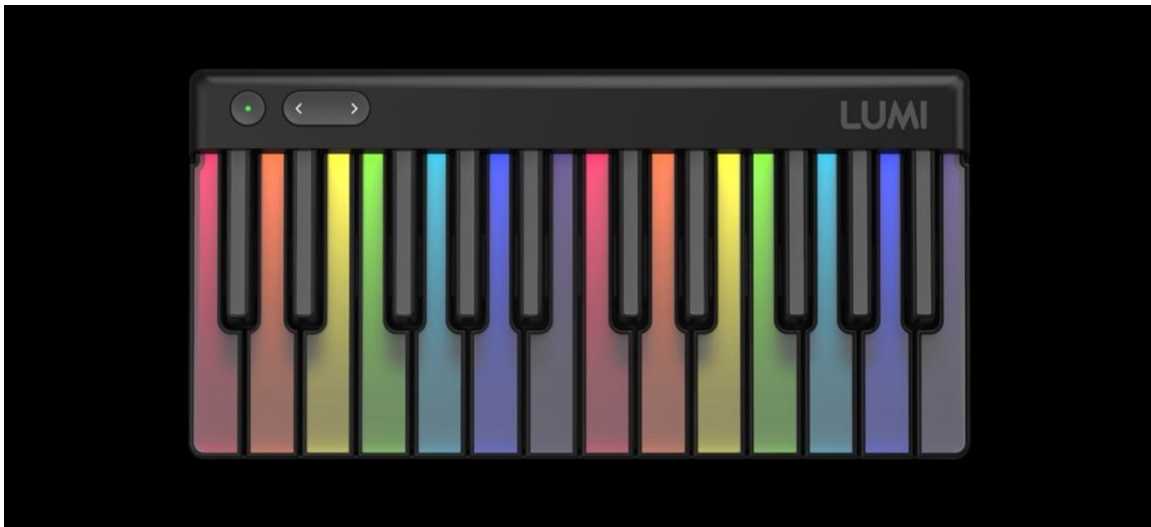


Figure 1.44: Lumi keys
Source: Adapted from [231]



Figure 1.45: Eyesy

Source: Adapted from [232]

1.4.5 Audiovisual Art, Multimedia Art, and Design

Computer science, technology, music, and art are the most active ingredients in audiovisual art, multimedia art, and design now. It is worth briefly mentioning generative art, electronic art, and light art, because audiovisual art and multimedia art stemmed from those fields and encompassed them.

In generative art, an artist makes the use of an autonomous [self-evolving] system [233] that is not under their direct control [234]. Interactive art is a form of digital art that involves audience participation [235]. Multi-media art uses electronic media [236], but generally includes generative, electronic, computer, and interactive art [237].

Specifically, new media art refers to the use of computer technology [238] because new media art existed before computers. Therefore, the term “media art” in this dissertation can refer to both multi-media art and new media art. In my audiovisual work, described in Chapter 6, each shape moves automatically and its color changes constantly. This can be seen as generative art.

Electronic art can be seen as a precedent art form or an alternative, having originated in the sixties [239]. Fluxus, an interdisciplinary artist group, was at the center of this movement. Nam June Paik's *Electronic Television* [240] allows an audience to control screen images with a microphone (Figure 1.46). He used various objects including fabric, a video synth, custom-made electronic devices, and conventional musical instruments. In my work, the visual does not respond to sound from microphones like Nam June Paik's work, but rather the user can control the visual-related parameters (i.e., size, color, position, etc.) with a touch-based control interface.



Figure 1.46: Electronic Television by Paik

Source: Adapted from [240]

Sound has also been expressed by light. La Monte Young's light art installation, *Dream House* [241], generates a sine tone and emits magenta-colored light (Figure 1.47). Scott Arford's *Static Room* [242] makes a flickering sound from the moiré patterns on TV, and the installation *Lux* [243] creates light-sound synchronization. Andy Warhol is regarded as a pioneer who mixed fine art, media art, and pop music production together. His *Exploding Plastic Inevitable* [244] included video installations, projection and light art, and music created by Nico and Velvet Underground. In media art, Steina and Woody Vasulka are highly acclaimed figures for their experiments with electronic creation to

manipulate sound and images [245]. The color-sound mapping in my work is based on digital color values and is represented by digital monitors. Although the means of expression are slightly different, since digital monitors also consist of light-emitting diodes (LEDs), my audiovisual work can be seen as digital light art.



Figure 1.47: Dream House by Young
Source: Adapted from [241]

Generative, interactive, and new media art are contemporaries that make greater use of the computational, multimodal, and multi-disciplinary approaches. As computer technology rapidly develops, traditional design fields such as graphic design have evolved into motion graphics and generative, automate, and code-based approaches have converged. As such, various forms of art started to emerge together and produce new art forms.

Audiovisual art accompanies both visual and sound representation. Brian Eno, a pioneer of the ambient music genre, expressed his generative ambient music along with visual art, working with visual artist and software engineer Peter Chilvers. His latest studio album, *Reflection* (2017) [246] (Figure 1.48), was turned into the iOS app where the sound and the visual endlessly generated, harmonized and shifted according to time-of-day, freed from the size limitation of physical music storage formats like CD, cassette, and vinyl.

The visual in the app stemmed from Eno's light box, *Arduino*-driven [247] LED lights with various colors illuminating the translucent board behind [248]. If I migrate my work in Chapter 6 into iPad/Android apps, it will have a form similar to Eno's work. Users will be able to control various shapes and their properties and immediately appreciate the corresponding sound.

An electronic musician, visual artist, and a founder of the record label Raster-Norton, Carsten Nicolai (a.k.a. Alva Noto), has made a great impact on contemporary visual art, combining synchronized electronic music and abstract motion graphics. For example, in *unidisplay* (2012) or *unicolor* (2014), Nicolai displayed a series of visual components, which were mainly influenced by chromatology and psychology, onto a long immersive projection wall. In his work, *alpha pulse* (2014) (Figure 1.49), pulsating light sources were projected onto Hong Kong's iconic International Commerce Centre, using the light system of the building. In a similar manner, a co-founder of Raster-Norton, Olaf Bender (a.k.a. Byetone), used sine waves to create sounds and plugged them into his flickering visual expression while performing. Likewise, another notable Raster-Norton artist, Ryoji Ikeda, is a pioneer of modern audiovisual art. His installations translate computer data into a visceral audiovisual experience. In his work, *test pattern*, which is a part of his long-term project *datamatics* [249], he uses binary numbers that are converted from text, sounds, movies and photography to create the fast-changing black and white (or zero and one) barcode-like visual component synchronized with sine tones and glitch noises [250] (Figure 1.50). Similarly, a South Korean audiovisual performance team, Tacit Group, performs live concerts from laptops on a table with a huge video projection at their back, like Kraftwerk used to perform in former days. Their works mostly seek to find a humorous and witty presentation, inspired by the Hangul writing system (the Korean alphabet) or the algorithm of the classic video game TETRIS.

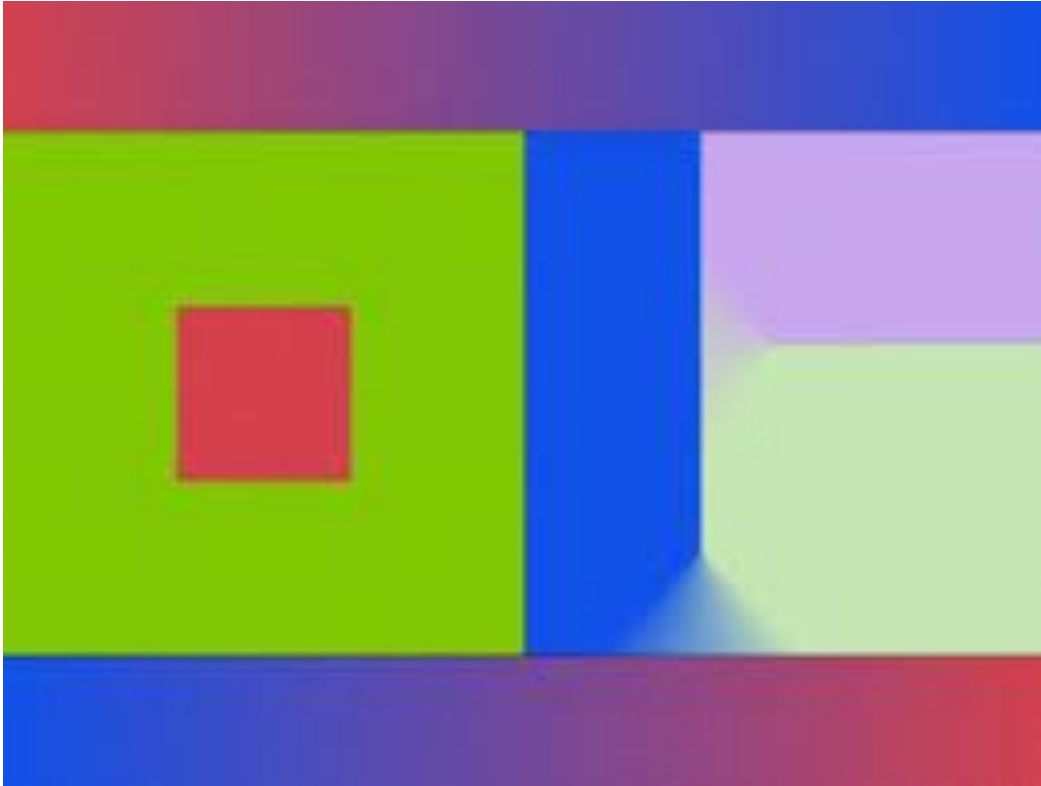


Figure 1.48: Reflection by Eno
Source: Adapted from [246]



Figure 1.49: Alpha Pulse by Nicolai

Source: Adapted from [251]



Figure 1.50: Test Pattern by Ikeda

Source: Adapted from [250]

Electronic musicians (by themselves or collaborating with other visual artists) have been using audiovisuality to express their own unique artistry in a visual form to audiences. The visuality of Brian Williams' (a.k.a. Lustmord) works are a vital part of his dark ambient music, and his work is gradually magnified by visuality. His experimentation across audiovisual art and music becomes clearer through visual aspects [252]. An electronic music legend, Kraftwerk, made the most of visuals that perfectly synchronized with their music. The group played a series of live performances and released the *3-D The Catalogue* (2017) album that showcases their entire music career using 3D projection visualization. This album is a collection of shows from around the world, starting with MoMa in 2012 and continuing to Guggenheim Museum Bilbao in 2016. Tom Jenkinson of Squarepusher is well-known for using massive and splendid audiovisual displays in his show. For the stage of the *ufabulum* tour, the entire stage was covered by LED screens and Jenkinson even made a face covering with an LED. Each visual theme dynamically changed based on music when his custom-made music sequencer sent a signal to control the visuals. Similarly, Deadmau5 wears the large LED mouse helmet for his iconic symbol during his live performance. For his live stage, he implanted a giant 3-sided LED

cube, called Cubev3 [253], that mainly displays the visuals for his show. This massive cube is designed to freely rotate, tilt, and give audiences an overwhelming audiovisual experience. This high-tech structure shows how a modern live stage can look. Likewise, Flying Lotus used a similar cube-shape projection sculpture called Layer³ [254] (Figure 1.51). A media production company Strangeloop studios and a visual artist Timeboy created the visual concepts and designed this mind-blowing stage for Flying Lotus. Continuous fractal or kaleidoscopic graphics were displayed onto the surface of Layer³, and a silhouette of Flying Lotus appeared in the middle of this hypercube. *Vice* magazine said that this projection technology transformed the live experience from virtual 3D to actual 3D [255].

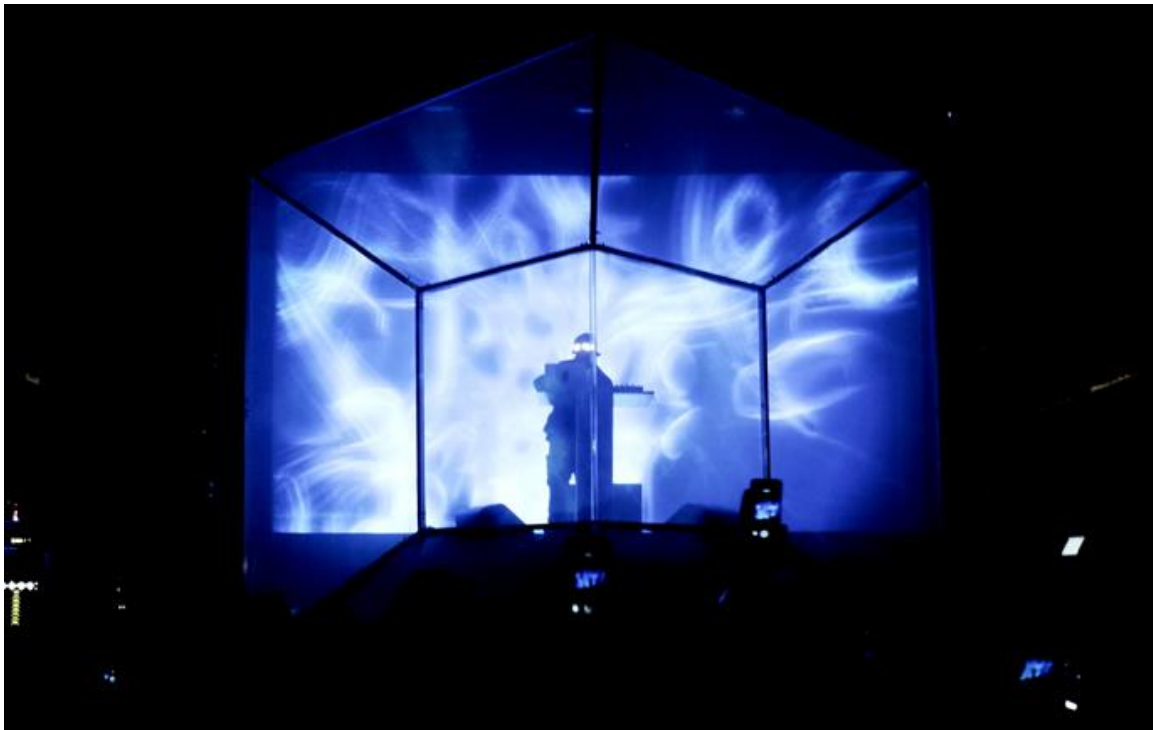


Figure 1.51: Layer Cubed

Source: Adapted from [256]

Amon Tobin teamed up with a multimedia design studio Xite labs [257] for the world tour of his eighth studio album *ISAM* (2011). The *ISAM* tour specially paid attention to audiovisuality setting up a large 3D projection mapping on a custom built cubic-shape sculpture. This massive symbolic structure performed a role as a multidimensional

projection screen that gave audiences an explosive multi-sensory experience. An Australian audiovisual artist, Tom Hall, built his own system. His work explores an unconventional way of showing abstract, geometric, and picturesque visuals that respond to his noisy and glitchy music, entirely derived from his custom-designed Max-based music production software. Besides, as the simplest form, waveforms can be utilized as a primary visualization subject. A French electronic musician and a new media artist, Fraction, in his work *vector field*, created visual effects with a large number of waveforms combined and layered. The visualization elements rotated, blinked, and glitched along to the sound. My audiovisual work, described in Chapter 6, started from the image-sound association study, detailed in Chapter 4, and now has evolved into one two form factors. It shows the potential for musical expression, however, it has not yet been used for splendid electronic music performances like the musicians mentioned above. My ultimate goal is to further develop my audiovisual platform for real-time electronic music performance.

As new media art technology advances, multimedia artists and multimedia design studios are starting up and rapidly growing. Sound is inevitable in audiovisual art, however these multimedia art examples are not necessarily coupled with sounds. Rather, it may embrace audiovisual art or other mixed media forms. A Turkish-American media artist, Refik Anadol, uses data or machine learning algorithms to create his mesmerizing visualization works. His latest work, *Quantum Memories* reflects Google AI quantum computation data, machine learning, and probability as tiny particles displayed on a giant 10x10 meter LED screen (Figure 1.52). In *Seoul Haemong*, machine learning reprocessed the eleven million historical photos of Seoul provided by the Seoul Metropolitan Government and taken by Seoul citizens. Anadol projected it onto the entire wall of the DDP (the Dongdaemun Design Plaza) building, one of the major landmarks in Seoul designed by Zaha Hadid. Similarly, in *Depictions:San Francisco*, he employed data of the story of the city, and the unique curved media wall in 350 Mission Street in San Francisco was used to project this work. This style of data visualization is called data sculpture, a representation of data using physical objects in the real world instead of using conventional flat 2D mapping data [258]. In a similar vein, Ouchhh studio visualized

NASA's Kepler data in conjunction with machine learning. In *Say Superstrings*, the company detected electroencephalogram (EEG) from a classical music trio and visualized the musicians' brain waves in real-time at a concert in Ars Electronica, one of the biggest new media art festivals. Lastly, a Russian multimedia artist duo, 404.zero, created four editions of 3000x3000 pixel images, derived from a code-based random generation algorithm. Since data sculpture in these examples are basically digital image projection consisting of large numbers of micro 3D particles, the primary data presenting objects can be regarded as virtual 3D objects.

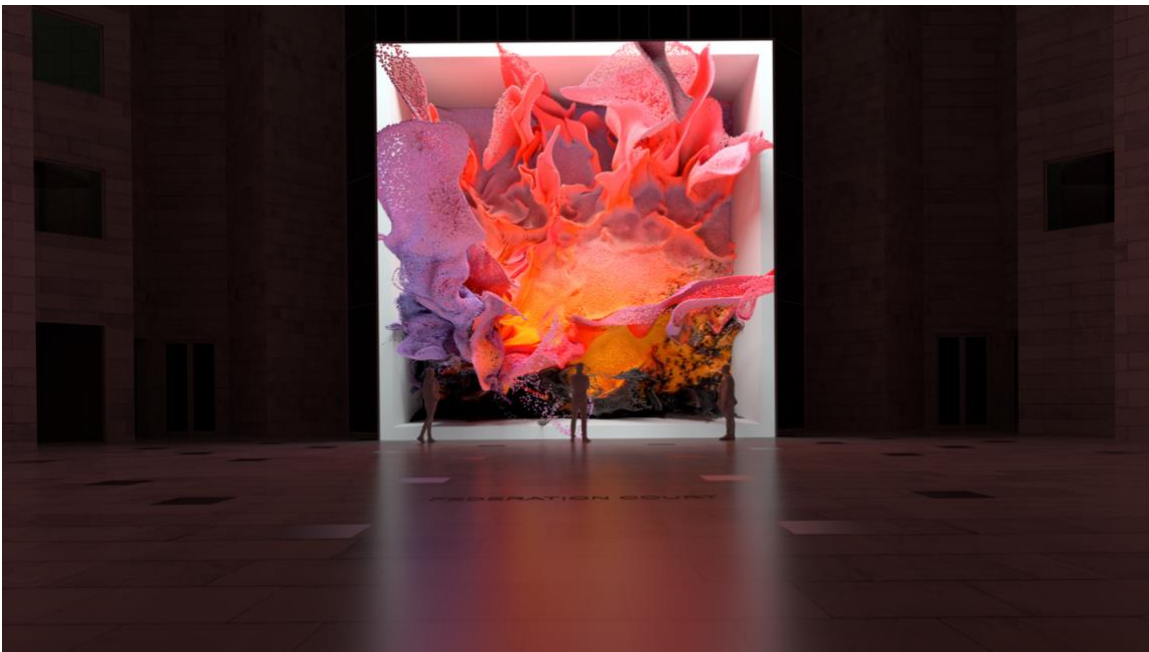


Figure 1.52: Quantum Memories by Anadol

Source: Adapted from [259]

The elements of multimedia art are boundless. Nonotak Studio in France, founded by an illustrator and an architect/musician, primarily creates light art and sound installations in immersive exhibition spaces. Their use of the exhibition space and the interpretation of the sound naturally blended with their geometrically arranged light art expression. ART+COM studios in Germany incorporates architectural or sculptural approaches into their media installation. The work, *RGB/CMY Kinetic*, offers a poetic experience from the combination of kinetic art and light art. Five flat discs float around the space and reflect

color light. This reflection makes a colored shadow on the floor. The five discs are mapped into five different digital instruments and based on the height of the discs, each sound (created by an Icelandic musician, Ólafur Arnalds) modulates. In their work, *Reactive Sparks*, the energy of passing cars near the installation venue is expressed as color light and is shown on seven metal LED steles in Munich, Germany. An Italian media artist, David Quayola, created custom software titled *Partitura* to migrate sound into visual forms in real-time in collaboration with the music group, *Abstract Birds*. This music visualization responds to the structure of music over time. They utilized this system for their audiovisual performance, playing György Ligeti's *Sonate pour alto* in France. Later, *Partitura* was further exploited to create the visuals for an audiovisual performance project with electronic musician Jamie XX in London.

The examples described above show the generalized forms of audiovisual art or multimedia art which means that they are not necessarily related to sonification. Specific to sonification-driven audiovisual works, Golan Levin's audiovisual performance tool *AVES* (Audiovisual Environment Suite) [260] was designed based on free-form image sonification for real-time performance and can be understood as a more advanced version of UPIC. This system incorporates five different graphic-sound mappings with an interactive control environment for audiences. More detailed examples will be explained in Chapter 2.2.1.

Lastly, I would like to briefly introduce festivals and conferences for ideas of synesthetic art. First of all, the Center for Visual Music (CVM) in Los Angeles houses the archives of visual music masters such as Oscar Fischinger and Jordan Belson, and also sells archived videos and screens films. Punto y Raya, which was launched in Barcelona in 2007, also provides a festival and educational content with a focus on visual music, abstract film, and audiovisual art. In the UK, Seeing Sound focuses on multimedia works that deal with sound and images and Audio-Visual-Art (AVA), a festival focused on visual arts and electronic music, is worth mentioning.

1.4.6 Multimedia Software

Software platforms that are widely used in multimedia art and multidisciplinary art/design will be introduced in this chapter. Typically, music/art software focuses on one task: Illustrator [23] for a vector-based drawing, Photoshop [24] for a pixel-based drawing and photo editing, Final Cut Pro [261] and Logic Pro [262] for video and sound production. However, the best features of multimedia software are the expandability and flexibility of uses, they perform as multi-function packages providing graphic, sound, and video tools together, so users can easily utilize each or use them in combination, based on their preferences.

Processing [263] and openFrameworks [264] are the most well-known code-based graphics and visual design tools. Processing is the open-source and Java programming language specifically targeted for artists, designers, and academic researchers who are non-programmers. Since Processing comes with a simple and quick text editor and does not require a profound level of computer programming skill, it provides a fast-prototyping environment. Processing has become famous for educational purposes, teaching the basics of computer programming for creative coding in the context of visual art and data visualization. As a result, it is now popularly adopted among graphic designers and visual artists for professional design tasks. Processing reduces barriers to entry for non-computer science majors. As an alternative, openFrameworks offers a similar type of creative coding platform, but the major difference is that it is based on C++ and is preferred by anyone who already has computer programming experience.

Max [265] is a giant of new media art software. Started by Miller Puckette in 1985, Max was first developed as a programming language, which carries a graphical user interface, for interactive and generative computer music. It has a longer history than other multimedia software introduced in this section. Max is widely used among musicians, sound designers, audiovisual artists and researchers. Its essence of Max is a node-based graphical coding interface that enables easy prototyping for users who do not have text-based coding experience. Max's functionality has been constantly expanded, so it is used not only for music but also for video, visual FX, and 3D graphics. Because Max

interworks with text-based data (e.g., csv or txt files), it has been generally adopted for data visualization and data sonification. As an alternative, Pure Data (a.k.a. Pd) [266] is a primitive version of Max.

Quartz Composer [267], another node-based programming language from Apple, is worth noting considering its popularity. Quartz Composer prioritizes FX and projection mapping. This development tool was popular among video and media artists, however Apple stopped supporting the software and it has been decommissioned. Quartz Composer handed over to the next runner TouchDesigner [268], which is developed by Derivative in Canada. The TouchDesigner community continues to grow centering around multimedia artists. Similar node-based software, vvvv [269], Notch [270], and Vuo [271] are also available.

This multimedia software offers easier and more intuitive accessibility than hard coding, thanks to their graphical coding environment, and provides expandability to other software, making these tools well-suited for multidisciplinary work in graphics, sound design, computational design, algorithmic/generative visual and sound composition, interactive data visualization, and audiovisual performance. For my dissertation work, I mainly used TouchDesigner and Max (see Chapter 4 and Chapter 6).

1.5 Thesis Outline and Contribution

In the previous chapters, an introductory, historical, and background knowledge of the foundational fields to which this research is related were covered. Chapter 2 (the following chapter) will discuss what sonification strategies exist and what corresponding image-sound converting methodologies have been studied so far. It will also delve into how sonification researchers applied their ideas to achieve their goal successfully, what the previous research findings were, how their sonification platforms were tested with the subjects, and where image sonification is now.

In Chapter 3, the research hypotheses will be explained and in Chapter 4, a series of experiments that support the hypotheses will be described in detail. This includes the

relationship between shapes, sizes, positions, motions, and sound characters. This study introduces an innovative approach to convert basic shapes into sounds. Dondis' says that "optimized perception is the basis of abstract design principles" [272]. This research will show how elementary shapes can be interpreted as sounds and how the sonification method efficiently delivers this visual information as aural information.

The main importance of the user test is to introduce the sonification system so the listeners can match the visual shapes by listening to a particular sound. By doing so, the system also provides augmented multi modal experiences to listeners. Considering the size of the sonification community is relatively small across the country, in comparison to other research fields, there are not many precedential sonification works done in connection with elementary graphic design perception and its use in audiovisual composition. Therefore, this research will be a base study for future research in a more complex form of graphic design sonification and its corresponding visual-sound mapping model, specifically for real-time performance and exhibition setup. The proposed sonification described in this dissertation can be utilized as a new vehicle for understanding the overall context of visual-auditory composition, and a new multisensory design tool in which the audiovisual components regain a role and purpose, widening our cognitive awareness.

The methodological contributions of this research will be:

- A fundamental and universal graphic design sonification model to display the shape parameters including colors, shapes, sizes, motions, positions, and the number of shapes.
- A fundamental and universal graphic design sonification model for multi-sensory art expression (audiovisual composition).

2 Literature Review

2.1 Introduction

This dissertation presents sonification models that convert elementary shapes into sound and aims to create functioning audiovisual platforms. The primary goal is to explore how sonification can be approached by graphics, how graphics can be expressed as sounds, and how the merger of the two enhances the functionality of communication design and the multimedia and audiovisual art. How visualization formulates sonification and inversely, how sonification helps clarify the corresponding visualization are important research questions for the next level of communication design. Additionally, an image-sound matching user test was conducted to examine whether the relationship between sounds and shapes can be coherent and logical, and successfully utilized as a new type of visual/auditory (multi-sensory) communication tool to enhance user experience.

This dissertation associates sonification with the transformation of visual information into auditory information. Transitional examples can be found both in the arts and sciences regardless of the functionality, practicality, and methodology. Areas include performance, sound art, multimedia, interactive design, data/cognitive science, assistive devices. Sonification in audiovisual, multimedia, and sound art tends to be more flexible, to secure an aesthetic viewpoint. In contrast, sonification in science (e.g., auditory perception, information delivery) aims for more practically oriented outcomes, such as the development of assistive devices for visually impaired people. To examine the significance of this sonification study, to design a successful sonification model, and to contribute to building the foundation for my proposed sonification approaches, sonification in both science and art will be introduced in this chapter. Specifically, it is necessary to explore (1) image to sound sonification models in art, (2) scientific models developed for visual impairment, (3) software platforms in musical instrument design, and (4) the findings on those preceding models. These examples will be a basis to understand how my sonification method developed, what major differences there are, and how it can be utilized for future sonification purposes.

Since no word exists to describe sonification of graphics, visuals, or images, a concise term needs to be defined. In this dissertation, I would like to coin the term “Graphic Sonification” to specify this sonification type. Namely, graphic sonification means a direct utilization of sounds and sound design techniques derived from graphic design elements.

2.2 Graphic Sonification Models

2.2.1 Graphic Sonification Models in Art

To convert a visual image to a sound, the process of how to read images and employ them must be considered first. In auditory display, time is the primary dimension, thus converting methods that migrate images (space-dependent) into time-domain data need to be determined correctly by sonification designers.

A sonification mapping can be described according to whether or not it probes or scans, its pointer shape, and its path. Yeo and Berger presented an image scanning framework [273] [274] to map still image data to time-dependent (or sound) data. They named time mapping processes *scanning* and *probing*. Scanning applies to a fixed order and a fixed speed of data to be sonified. We can understand this concept through an inverse spectrogram, in which the x-axis is mapped to time and the y-axis is mapped to pitch. The sound is generated from the leftmost column of the spectrogram and the sonification process ends when the scanning column reaches the rightmost part of the spectrogram. While scanning is operated in a predetermined way, probing offers a more fluid interaction environment to users so that the speed, position, and area of sonification can be flexibly determined by viewers.

In addition to scanning and probing, Yeo and Berger classified two concepts called “pointer” and “path”, concerning a reference scanning point of images in time. This can be understood as a type of a mouse cursor we want to use to pick a particular area of pixels. First, the pointer is an area that will be transferred to an auditory domain at the same time. The shape of the pointer can vary between a single point, distributed shapes, line-shapes, or a big circular shape. The pointers’ shapes can be seen as the shape of the

sonification area where the sonification process occurs at the same time. If a pointer has a pre-defined movement or users freely move a pointer, this pointer becomes a path.

Yeo and Berger gave examples of five representative image sonification mappings based on their framework. First, an inverse spectrogram is a scanning processing with a line-shaped pointer, moving along the x-axis. Second, raster scanning is a scanning process with a point pointer. Raster scanning is an image processing technique that can be found in an analog CRT display. This imaging technique works in a line-by-line manner, meaning that it uses sequential progression to display the entire image on a designated screen horizontally from left to right and from top to bottom [275]. In their raster scanning sonification framework, one pixel of a grayscale image takes one audio sample in the wavetable, and brightness is mapped into the audio sample value from -1 and 1. The brighter the pixel, the higher the sample value. If an image has a periodic pattern in a row, naturally the sound will become pitched. The wider the periodic image is, the higher the pitch. In addition, this raster mapping can sonify an image. Third, they introduced two types of virtual paths called “Path on Color Depth” and “Path on Imaginary Axis”. The former does not have a path mapped into the x- or y-axis. Instead, it has a pointer that covers the whole image, and its path is responsible for the z-axis (the 3rd dimension of an image), which is a color value. For example, the darker the pixel is, the shorter the duration is. This model uses a scanning process with a pointer that is of the same size as the image. The “Path on Imaginary Axis” model is a scanning process with a rectangular pointer that moves along a perpendicular direction (or path). Fourth and fifth, the authors introduced SonART [151][152][278] and Viewer.app [276] with an explanation of using the probing process and a combination of probing and scanning processes.

SonART is a multimedia tool where users can designate a scanning zone with a mouse. The opacity and RGB color values of the selected region can be utilized for sonification. That is to say that SonART is based on a probing process with a point pointer on a user-dependent path. Viewer.app is similar to SonART, but the probing mode and scanning mode are interchangeable.

The authors were aware of the drawback of each approach. Scanning lacks the ability to concentrate on partial details of an image at which users want to specifically look. In contrast, probing enables users to do that by focusing on user-defined areas. However, it does not have the feature of overviewing the entire image. Thus, the authors said that combining both would be worth considering.

Yeo and Berger's framework provides a clear classification and description of sonification methods. The main significance is that it can serve as a guideline when an existing sonification model needs to be analyzed, or as a blueprint for future planning where a successful sonification strategy is needed. On top of that, based on those concepts, we can systematically propose a new sonification model according to its purpose. There does not seem to be a huge difference between scanning with a line pointer and raster scanning with a dot pointer, because there exists a particular data processing order that is similar to the one of raster scanning inside a line pointer. Empirical research has not been conducted to compare both methods.

Yeo and Berger mentioned that raster scanning is an intuitive and easy method to understand sonification mappings between image and sound data. From my perspective, raster scanning and a scanning processing with a line-shaped pointer both seem like a natural way to convert two-dimensional visual domain data to time domain data because these approaches are most often found not only in image sonification but also in Digital Audio Workstations (DAWs) such as Ableton Live [279] and Logic Pro [262]. DAWs have their own unique graphic objects that display sound samples or MIDI values, and a line-based pointer triggers the sound.

My master thesis [280] culminated in creating a sonification-driven audiovisual work at the Rhode Island School of Design that used an invisible single line moving horizontally or vertically over abstract graphic patterns. I created an additive synthesis-based sound and corresponding dynamic graphics that present the system's linear scanning process simultaneously. I intuitively came up with a raster scanning-like approach before I investigated Yeo and Berger's paper during my Ph.D. Another example, *Voice of*

Sisyphus, a multimedia installation work by Ryan McGee and others [281], looks as if it was heavily inspired by probing Yeo and Berger's Viewer.app and *Rasterpiece* [282], but the authors said their work was created before they were aware of Yeo and Berger's raster scanning framework. This is not a completely new method, however it is a fact that Yeo and Berger's proposal helps sonification researchers easily classify image scanning frameworks of this kind.

Like my system, *Voice of Sisyphus* scans the image from top to bottom or left to right. The sonification model of *Voice of Sisyphus* uses 8-bit greyscale images for sound synthesis and composition. The major difference from Yeo's models is that the region for sonification is changeable in position and size, and the sonification process operates in real-time. Therefore, this feature allows users to navigate the images more flexibly. Greyscale pixel values within the selected region are converted into floating-point numbers between -1 and 1, and those values are saved as audio samples.

The sonification methodology mentioned above can be found in Marko Ciciliani's Scanline synthesis [283]. Likewise, the fundamental principle is that pixel values (0-255) from images are transported to an audio buffer (ranging between -1 and 1), and this waveshape in the audio buffer serves the role of digital wavetable oscillator. Like the concept of Raster scanning, the image is read by a straight line moving horizontally or vertically. According to Ciciliani, visual music and audiovisual artists often use Scanline synthesis because of the direct connection between the pixel values and the corresponding audio samples in the buffer. The author performed *Formula Minus One* (2014) [284] and *Via* (2015) [285] based on this scanning methodology. In *Formula Minus One*, the F-1 race generates electronic music, becoming a visual score for the violin player standing on the stage. In *Via*, the road trip video is coupled with Scanline synthesis. Each pixel of the video is converted into a waveform and the region for sonification can be selected and controlled via the hand-held device. Up to five different waveforms can be mixed.

Scanline Synthesis can be also found in the following works. Christopher Jette's *SoundLines* (2013) [286] transfers body motion into sound. A live video captures a performer's body and grayscale values in the two horizontal scanlines in the video form sound wave. Dave Poulter's *Scanline Granular Synthesis* [287] uses the same approach for a Max driven granular synthesizer. In Ryoho Kobayashi's *Scanline: Computer Music* [288], the performer searches a keyword "computer music" on Google, and transfers the photo result into sound. The image is read in both a horizontal and vertical way. The brightness of the pixel values creates waveforms. The maximum number of the scanline is up to eight for each direction, and the speed of the scanline can be flexible. Likewise, Scanline Synthesis is quite common in sonification because it is intuitive and natural.

Similarly, Shawn Greenlee created a graphic-sonification platform for his real-time audiovisual performance. He defined his methodology as "Graphic Waveshaping" [289], and this model can be understood as a variant of Scanline synthesis or Scanned synthesis [290][291] because the waveform here can be formed by the visual images. Greenlee's platform uses microscopes with lighting and these microscopes are controlled by the performer to capture the printed or hand-painted images on a paper on the dark stage. This system allows the use of color images, but four channels in RGBA are converted to one luminance channel. The input resolution of the image is limited to 512 x 512 pixels, so 512 data points of the vertical axis can be saved in the audio buffer rescaling the pixel values between -1 and 1. The horizontal length is the time axis.

To be clear, Scanline synthesis and Scanned synthesis are not exactly the same terminology. They both are wavetable synths but *scanned synthesis* does not involve any visual component [283]. To be specific, the waveforms in *scanned synthesis* are determined by a system that scans data with a low-frequency vibration (below 15Hz) periodically. The timbral character of the sound from *Scanned Synthesis* heavily relies on the physical motion of a user's input, and the speed of the scanning function determines pitch. So, the wavetable is animated based on the haptic rates of an object in real-time. Since scanned synthesis is just one of the sound-making techniques such as additive

synthesis and FM synthesis, it does not seem to be directly related to image-sound sonification techniques.

There is another data scanning type that uses a three-dimensional pointer. Wave terrain synthesis [292] uses an orbit (this can be seen as a pointer according to Yeo and Berger's terminology) to create a waveform. The orbit is a closed path traveling around and scanning the surface of three-dimensional data. The 3D data surface can be anything. If an image file can be displayed onto a three-dimensional surface, the wave terrain synthesis technique can be used for sound design.

For more examples of sonification models that create the shape of a waveform directly from color values, in the Monalisa [293] platform, the sonification algorithm uses the image data and produces binary number sequences. This platform comes with 8-bit and 16-bit data transition models. The 8-bit model converts RGB (red, green, and blue) values into separate 8-bit data samples of linear Pulse Code Modulation (PCM) sound data. The order of this data transition is from the top left to the right bottom pixel of the image. For example, the first three arrays of the PCM data represents the RGB values of the topmost-left pixel of the image. In the 16-bit model, the first array of the PCM data contains all RGB values. The art installation *Shadow of the Sound* that is a part of the Monalisa platform allows the audiences to take a picture of themselves and generate the stream of sound. Interestingly, the sound is re-captured by the microphone and reflects the reverberation level of the installation room onto the projection image overlapping with the original portrait image.

Photone (2018) [294] is an interactive photo-sound sonification installation where the authors, Rönnerberg and Löwgren, put an emphasis on musical expression rather than irregular timbre formation or cacophony, which is quite common in this type of sonification. Pitch, amplitude, and timbre are basic elements for this photo sonification and the sound design of *Photone* is inspired by drone music. First, *Photone* determines the dominant hue of the image used, referring to the RGB color system, and the middle values between two primary colors, i.e., yellow, magenta, and cyan are included, while

white and black represent the maximum and minimum light intensity. The predominant hue falls into these seven colors (except black), and the overall harmonicity of music is determined based on which of six musical expressions is assigned, including chord, melody components, bass and high tones, white and black. The chord component has two-notes moving in a five octave range, for example, red takes C and Eb ranging from C2 to C7 for the first note, and Eb2 to Eb7 for the second one. If two primary colors are blended, it will generate four notes, and if white is the dominant color, six notes will be played. Each voice per color contains six triangle wave oscillators, and those six triangle waves are slightly detuned with the distance of -10, -7, -4, 4, 7, and 10 cents. The amplitude attenuation increases or decreases based on each color channel's light intensity. The melody component generates five notes, for example, red consists of G5, C6, Eb6, G6, and C7, however, they are not played at the same time. The color intensity level is divided into five degrees, so according to where the intensity is positioned, one of five notes is played. This harmonic component uses triangle wave oscillators, but with a shorter attack and longer sustain values. Also, the intensity level determines the band-pass filter's cutoff frequency. The bass tone has C2 and C3, and the high tone consists of C7, Eb7, and G7. They both use sawtooth wave oscillators and when the light intensity level is extremely low or high, both components become audible. Lastly, for white and black, a single triangle wave, which resembles a short bell sound on C8, is in charge of the white area, and two sawtooth waves sweeping down from C3 and C2 are mapped to black.

Atau Tanaka's *Bondage* [295][296] adopted image-sonification for a photo and media art exhibition setting. Tanaka projected Japanese bondage photo onto a wall which looks like a door found in traditional Japanese architecture. The photo in the door is divided into four rows and six columns. Each section is scanned from left to right like Yeo's raster scanning. Each row refers to a different range of frequencies. The sonification approach in *Bondage* uses a single luminosity channel of the image, and the overall frequency falls within the human hearing range from 20-20kHz. Sub-low bass frequencies come from the bottommost row; middle-low frequencies from the second bottom row; middle-high from the second top row; and highest treble frequencies from the topmost row. And the

intensity of sound is determined by the luminosity of the image. Moreover, the sound is spatialized with a four-channel sound system. The upper speakers are mapped into the high frequencies, and the lower speakers represent the low frequencies. The six-column zones of the image correspond to the left-right horizontal space mapping. The installation of *Bondage* employs an infrared camera that detects the viewers and reflects a silhouette of them over the projection screen. The intersection between the main subject photo and the viewer's silhouette is visible and sonified.

Golan Levin presented a voluminous sonification series specifically designed for real-time audiovisual performance in his master thesis in 2000 [260]. The main concept of the series is to create a new gesture-based interface to cope with visual-sound sonification playing in an endless loop. In terms of this looping metaphor, Levin's approaches hold some resemblance to Brian Eno's and Peter Chilvers's music apps (described in Chapter 2.2.3). His five audiovisual systems consist of *Yellowtail*, *Loom*, *Warbo*, *Aurora*, and *Floo*.

The first work, *Yellowtail*, is sonification of hand-drawn lines. The author said this work can be understood as an "inverse spectrogram" that is an opposite concept of spectrograms that are supposed to analyze sound. When a user draws a straight or irregular line, the line automatically animates itself by following the order in which it was drawn and moves along the direction from the starting point to the end point of the drawing. If the line reaches to the end of the screen, it reappears on the other side of the screen. This animation is repeating. The behavior of animation is not randomized. Instead, what a user draws and how fast the line was drawn determines the animation behavior. For the sonification process, *Yellowtail* has a square-shaped scanning spot in the center of the canvas, and if the line drawing passes through this area, the sound is triggered. The scanning spot contains a horizontal line that moves repeatedly from bottom to top, and the pixel values of the area where it intersects with this line become the source of sinewave-based additive synthesis. The horizontal axis is connected to pitch, and the brightness of the pixels is mapped into the amplitude of an individual

component of the additive synthesis. This pitch-amplitude mapping also can be found in Iannis Xenakis's UPIC [297], the most famed and historical visual-sound instrument.

The second sonification model is *Loom*, which is an extended version of *Yellowtail*. *Loom* gets rid of the restriction of *Yellowtail* where sonification is performed in the small square area in the center of the canvas. In contrast, in *Loom*, when a user draws a line with a pointing device, the sonification process immediately responds by following the shape of the line, and the number of lines that a user can add is limitless. The line is thickened if a user presses the drawing pen harder. In this case, the sound at the thicker area becomes louder and when the line shape has an abrupt curve, the tone generates brighter timbre temporarily. The line drawing in *Loom* plays the role of a timeline. If the line is curved, the timeline of it is curved too. Every line has its own time-stamped information, so even if the line is severely angled, a time interval in the line is equally divided, and this time stamp also has information of its position, velocity, pen pressure, and curvature. Based on these time stamp values, the system continuously controls the sound parameters of FM synthesis. Gesture-related values including velocity, line pressure, and curvature are mapped into amplitude, depth of vibrato/amplitude, and index of FM modulation, which means the number of sine waves generated.

Levin's third variant, *Warbo*, includes color sonification along with the previous sonification systems shown in *Yellowtail* and *Loom*. *Warbo* supports a two-channel input with a computer mouse and a Wacom tablet. It is worth noting that *Warbo* is not image sonification but sonification of motion. With a mouse, users can create colored spots with circular or polygonal shaped virtual brushes with a popup keyboard on screen. If the color spots are drawn, they animate based on the recorded gestural path information. The color in the spot is a radial gradient getting brighter at the centroid, and if two spots are overlapping, two colors are added and become brighter. Each color spot emits a sine wave, and when a user moves a mouse cursor over the color spots, they build chordal tones. The distance between the centroid of the spot and the mouse cursor affects the volumes of the chord tones. If the cursor is near the centroid, the sound becomes louder.

When users use the other hand with a Wacom tablet, this controls another cursor called the *Streamer* that changes the timbral character of the chord tones.

In *Aurora*, the fourth variation, Levin brought his previous drawing systems that simulate physical properties of hairlike thin filaments. Although this filament algorithm is behind this model, the visual component of *Aurora* does not resemble hair but rather blurry, foggy, murky cloud shapes. Users can start with a three-button mouse by pressing one of the buttons to draw the foggy shape, choosing one of three colors. The available color values have a triadic relationship, spaced out 120 degree-distances on the hue circle. Like *Warbo*, *Aurora* is based on an additive color mixing, so if three colors are equally layered, it produces pure white, an absence of color. What makes *Aurora* differ from the previous models is that *Aurora* uses a granular synthesis technique. The mappings between the control parameters of granular synth and the filament simulation properties include: 1) The mean value of the velocity and the distension of a filament particle controls the grain carrier center frequency. 2) The particle velocity changes the frequency range. 3) The amount of the particles' vigor shortens or lengthens the grain duration floor, which is the shortest duration of the grain, and changes the carrier waveform going back and forth between a sine wave and a square wave. 4) The longer the length of the filaments, the narrower the duration range. 5) Stereo panning is determined by each grain's location. 6) The number of the points in the filament simulation increases or decreases the delay amount between each grain.

The last work, *Floo*, also uses granular synthesis, however, the visual component is based on a Navier-Stokes simulation of fluid flow. Other differences are the number of sound grains designated for sonification simultaneously (*Aurora* takes about ten per each filament, but *Floo* captures one per particle), and the waveform shape. A Shepard tone was used instead of sine and square waves, so the pitch of the grain circulates between C and B. For the mapping between the synthesizer parameters and the particle properties, the particle direction is mapped onto the pitch wheel starting at C and ending at B, and the grain position determines the stereo mapping. The grain amplitude changes according to the speed of each particle. Lastly, *Aurora* has a constant grain duration and a short

delay with a slightly randomized length. Levin's sonification works are rare sonification examples that are designed for artistic expression merged with complex graphics algorithms.

Lastly, Alex Braidwood created a color-based sound generating system for his graduate thesis and installation work, *Synesthetic Din* (2011). Braidwood painted the floor in six different colors that are commonly used in the streets of Los Angeles and installed a metal frame that can hold a mobile phone facing the color floor. Six different sounds were collected from various locations around the city and were assigned to six colors. By moving the mobile phone attached to the movable metal frame horizontally, the phone's camera detects three colors out of six and generates corresponding sound samples. This works the same as having three soundtracks. *Synesthetic Din* does not create timbre or determine tonal characteristics of the sounds, rather it triggers the prepared sounds based on the color information. However, this system shows how visual (color) information can be used for sound compositions. These examples have shown the great potential that sonification can be used for artistic expression. The examples mentioned above mainly convert color information extracted from still photos or video images into sound and use the sounds as an artistic expression. In other words, these sounds are not originally designed for the perceptual image-sound mapping of graphic-sound sonification (which is the main interest of this dissertation). I am yet to find an example of sonification designed first for image-sound association purposes later applied to artistic/musical expression.

2.2.2 Graphic Sonification Models in Science

As Chapter 1.3.3 shows, auditory screen readers, Web navigating frameworks and Electronic Travel Aids (ETA), including auditory GPS, are widely known assistive devices for visually impaired people. Likewise, sonification also has been used as a warning signal or an alerting system to avoid obstacles in one's path and guidance to access a graphical user interface on a computer. Scientific sonification models that particularly convert visuals into sounds aim to replace visual information, and for that reason, practicality and commercial viability are big considerations. This chapter will

introduce specific examples of visual-aid devices that involve image analysis technologies.

An Auditory Visual Sensory Substitution System (AVSS), or Sensory Substitution Device (SSD), converts visual images into sounds and replaces vision with sounds [85][298][299][300][301]. Meijer's vOICe [85][302] is one of the most cited examples of AVSS, and is the most compared by followers [51][300][303][304][305][306][307][308]. The image-sound rendering of vOICe follows an x-frequency and y-time approach. Speaking in Yeo and Berger's term [309], vOICe has a line-based pointer and its path moves from left to right (time-based). The lower part of the stripped line generates a low frequency, the higher part creates a high frequency, and the brightness of each pixel corresponds to volume. The camera input provides 64x64-pixel resolution with 16 degrees of grayscale. The visible angle of the field is 60 degrees.

Many researchers have developed several variations of the vOICe' approach. In 2018, Uno and others developed SIPReS [303], an Android app that converted images into sounds, operating similarly to vOICe. SIPReS analyzed the minimum and maximum brightness of an image and then determined frequency values: the default values were 220Hz and 1760Hz. A user could hear an assigned note or frequency while navigating on a phone screen with a finger. This system was only associated with the brightness of an image and worked even with a polychromatic image. The test subject was given a photo with a banana, an orange, and a blue cloth to make a color contrast between two objects and the background. The subject said that one was vertically long and the other one seemed to be a circular shape, but the subject did not exactly recognize what the two objects were.

In 2016, Wörtwein and others [310] introduced a web and touch-based sonification framework to display a fundamental data visualization of shapes such as bar charts, graphs of a function, and two-dimensional floor maps. A finger touch on a smartphone or tablet surface was required to hear the sound. The authors called it "a tactile graphic." High- and low-pitched sounds differentiate the distance between a user's finger and a

curved line of a math function, the sizes of two bar charts, and the distance between an obstacle, a finger, and the destination in a virtual map. This sonification type, to replace traditional data visualization with sound, can be found in Flowers and others [311] and Sound Graphs [312] for time series data or line graph, and in particular, a model with a touch function, can also be found in Kennel's Audiograf [313]. However, these models aimed for representing bar graph size comparison or vertical line position with sound, so the sonified sound did not directly express the shape of the bar graph itself.

EdgeSonic [304] follows the same x-time and y-frequency mapping of Meijer's vOICe, but it is a smartphone-based visual aid that allows a touch interaction like Audiograf. Unlike vOICe, EdgeSonic provides an edge detection feature to obtain contours of an image. Thus, when a user's finger touches these contours, a corresponding sound is generated. The activating area for sonification using a finger is 30x30 pixels. This model represents a pitch-responsive sound, depending on the position of the image border pixels below the finger position. In other words, users need to imagine the shape as they fumble on the screen. Therefore, the sound does not represent the entire shape. Javier Sanchez's sonification system [314] is similar also, except for the use of mouse and pen tablet input. In this case, both the movement of the pointer and the space between the shape and the pointer provide auditory feedback. Pitch, timbral character, and sound intensity vary based on curve shapes. SoundView [306] does not fully benefit from sonification because its approach is a hybrid of auditory and haptic feedback. SoundView uses a Wacom tablet and Wacom pen, and six elementary shapes such as perfect square, perfect circle, and equilateral triangle. Their corresponding hollow shapes are used for the user test. The test questions have a black background and each shape is displayed in white. Rough texture, in the shape of the six shapes, is printed over a sheet of paper, and the paper is attached on the surface of the tablet. When the Wacom pen navigates the tablet surface and hovers over the rough texture, it gives auditory feedback to the users. The sound used is filtered white noise based on the brightness [305]. SoundView was compared to vOICe, and the authors claimed that SoundView's success rate on the image recognition test doubles the test results of vOICe. SoundView was only tested with sighted participants.

Similarly, Sanchez [314] and SATIN system [315] showed an example of how to convert curve values of a line into sound. Unlike a vOICE-type method, which dealt with black pixel values and their position, Sanchez showed a way to express parametric curve value in sound. In this system, when users explored the line with pointing tools over a pen tablet, the pitch went up and down according to the vertical position of the line. SATIN was a hardware system that represented curve and curvature values as sound when it was attached to curved objects. It reproduced the highest and lowest positions of the curve point in the frequency range from 100Hz to 400Hz. Also, when a particular region of the object was touched, the curvature value of the corresponding position was informed in the same frequency range of the sound. What all of the previous examples have in common is that pixel changes in shapes, that is, the vertical position of the pixels, is expressed as a change in pitch. This is because pitch change is more perceptually clear than loudness change. Loudness has less resolution (which determines the ability to discriminate sound) than frequencies and frequency-related memory is more apparent than loudness-related memory [43].

As camera technologies develop, in addition to using 2D images, adding another mapping dimension in a visual aid system has been examined. SoundView [316] reads barcodes on the objects with a camera and determines the object's location. Brock and others [308] converted three-dimensional space to sounds adopting Microsoft Kinect to sense depth, a distance between a user and an object nearby. The x-axis is mapped to pitch just like other examples above. However, the horizontal axis measures stereo panning as if objects on the right will be heard on the right side of the headphone. Due to the additional z-axis, the sound becomes louder when it comes closer to a user. The use of depth in conjunction with volume is commonly found in similar research [60][61]. Furthermore, various modern sonification techniques, with a hybrid of tactile display and sonification [6], stereo [308][317], binaural sound spatialization technologies [60][318] [319][320][316], and color use [300][318][321] have been explored for better accuracy and enhanced user experience.

There are other types that mainly focus on color-sound mapping. Guido Bologna's SeeColOr [318][322] provides two types of image-sound conversion interfaces with a tablet and a head-mounted camera. These interfaces transfer colored pixels in an image into musical instrument sounds which are spatialized by the standard Head-Related Impulse Response (HRIR). For the color-sound mapping, SeeColOr uses an HSL color system, and seven musical instrument sounds, including viola, oboe, violin, flute, trumpet, piano, and saxophone. Saturation is divided into four sections and mapped to C, G, Bb, and E. If the lightness value is less than 0.5 (where 1.0 is the maximum), a double bass sound enters. In contrast, if the value is larger than 0.5, a singing voice on C, G, Bb, and E is mixed in. According to their test results, it seems that this system's design allows for the participants to memorize the color-instrument relationships, but the resulting sound does not depict what kind of shapes were tested, and, most importantly, memorization of the color-instrument relationship takes a long time. In their prototypes, introduced in 2007, Bologna and others [323] used 320x240 images and simplified them to reduce the amount of pixel information. The experiments were been done with mouse input, a stereoscopic color camera, and the Talking Tactile Tablet. A small portion of the region where the user points at was sonified. This model first used the RGB color system, however, in the case of two similar colors positioned on the RGB color cube at a huge distance, soundwise both colors do not happen to be close to each other. As a result, the HSL color system was adopted. For mapping between colors and sounds, red, orange, yellow, green, cyan, blue, and purple were mapped to oboe, viola, pizzicato violin, flute, trumpet, piano, and saxophone, in order. The five degrees of saturation played the notes C, G, Bb, and E. The four degrees of luminosity presented double bass sounds on C, G, Bb, and E, and another set of the four luminosity degrees overdubed a singing voice playing in the same C, G, Bb, and E.

Likewise, EyeMusic [300], similar to SeeColOr, mapped red, green, blue, and yellow into reggae organ, rapman's reed, brass, and strings, in that order. White matches with a choir sound, and black is coupled with silence. The authors of EyeMusic conducted the test with sighted and visually impaired people. The test consisted of the shape test with no color, and the color only test, however, a combination of both sections has yet to be

implemented and the test required two to three hours of training prior to the real test. EyeMusic was compared to vOICE, examining which model offers a more pleasant listening environment to the participants. According to their test results, blind people performed better than sighted users, and it was reported that EyeMusic is more pleasant than vOICE.

In 1999, Leonard N. Foner developed a wearable color-sound sonification device [324]. This device can be seen as an early type of Eyeborg [325] that does not require an implant surgery. A user can put this device on the side of one's head. Laser spots from the device can detect colors in the surroundings and RGB wavelengths of light are converted into audible wavelengths mapped linearly. Finally, each audible wavelength is summed like additive synthesis, and the resulting sound is transmitted via headphones. Presently, this device is only a prototype, so any experimentations are highly unlikely.

Osmanovic and others [326] used the visible light spectrum for their image sonification model in 2003. Their sonification model uses the RGB color system and grayscale values that range from 0 to 255. Red, green, and blue, which fall into 696nm, 552nm, and 465nm, were mapped into G at 392Hz, B at 494Hz, and D at 587Hz. When each value dropped from 255 to 0, its corresponding musical note of the sampled piano sound shifted -11 semitones (right above -2 octaves). The grayscale of the image controlled the intensity of the overall volume. This sonification model divided the image into 100 pixels and scanned it from the top left to the bottom right corner consecutively. The authors mentioned that their sonification was originally designed to make visually impaired people see artistic content by navigating with a computer mouse, but this sonification model seemed to stay at a primitive level at the time it was published, because it only converts a single pixel into a single sound, it does not support mouse interaction, and no test result was found.

Ausiello and others [327] investigated two color mapping schemes and compared them for a portable device for blind people to see the surrounding scene, by manipulating a virtual rain drop sound. The first is *Colorama* where the hue circle is divided into 64

regions. The perimeter of the hue circle is split into 16 subsets, and each subset has a basic timbre. Each of the 16 subsets has 4 subsets, following the saturation degrees, and if saturation decreases, white noise is added to the original timbre. The author pointed out that the *Colorama* model was taken from the 1976 CIE chromaticity diagram [328][329]. The second is *Georama*, which used the RGB color system, mapped into three different timbres, and the volume of each sound was mixed. For their study result, the Georama model performed better in terms of the color recognition test.

An SSD device called KromoPhone [307] that converted colors into sounds was presented by Capalbo and others in 2009 and the authors compared their sonification model to vOICe. KromoPhone's input source could be a mouse cursor or a head-mounted camera and this device interpreted a single pixel, where the mouse cursor or the camera aimed. With a remote control, the device switched between three different color modes including RGB, HSL, and RGBYW, where yellow and white were added as an additional color dimension. In the RGB mode, each color value was normalized between 0 and 1, and those values were multiplied by three different sine waves assigned to RGB, which were 15.9kHz, 8.4kHz, and 5.4kHz, in order. In addition, each color was spatialized in stereo. For example, red occupied the right stereo channel, blue was displayed in the left channel, and green was positioned in the middle. In the HSL mode, HSL values were mapped to pitch, panning, and volume. RGBYW was the default mode for the device. Each color had a distinctive sine wave and stereo panning just like the RBG mode, but the instrument sounds were newly added. A high-pitched trumpet sound for red displayed in the right ear, a high-pitched ukulele tone for yellow in the left ear, a medium pitched violin sound for green in the right ear, and a low-pitched trumpet sound for blue in the left ear.

Deleflie and others used color information for ambisonic encoded sound spatialization [330]. The image was pixelized in 60x60, so there were 3600 virtual speakers in the listening environment. Granular synthesis was selected for this model, and the grain position corresponded to the HSV hue, the HSV saturation, or the HSV lightness. The

gain volume was mapped to the HSL lightness or the HSL saturation, and pitch took the HSV saturation, but in some cases, pitch values were completely ignored.

Niklas Rönnerberg brought musicality into data sonification for his study [331]. He applied a musical sonification approach to allow the users to distinguish transparency levels of green. Basically, this model works with seven triangle wave oscillators detuned by 2-cent intervals around the fundamental frequency: 0, +2, +4, +6, and their minus counterparts. Each triangle wave moved in 11 keys ranging between C2 and C8, keeping a C major chord. This model had three different sonification settings. The first setting controlled the cutoff frequency of the band-pass filter and the amplitude, and the cutoff frequency swept from 100Hz and 6000Hz, according to the grayscale level of green. The second sonification setting changed the pitch of the sound based on the intensity level of green. As the color brightened, the pitch ascended. In the third setting, the harmonic elements of each sound turned into a dissonance chord relationship when the intensity level of green began to darken. Dylan also introduced how to encode images into musical form. This system basically converted simple geometric shapes such as lines and circles, to black pixels with the same width, just like eight-bit retro video game graphics in black and white and each pixel was mapped to a piano sound. Just like the other examples in this chapter, pitch (of the piano sound) rose and fell according to the pixel height, and the blind users were asked to infer the shape, paying attention to the pitch changes.

There is a case where a graphic sonification for visual substitution was implanted into a body. Neil Harbisson, who was born with achromatopsia, permanently put a device called Eyeborg [325][332] into his head. Eyeborg has a small antenna that detects colors and, through a vibration, enables him to hear the sound of colors using bone conduction. Harbisson defined his color-sound relationship as sonochromatism [59], which basically consists of a microtonal scale with 360 notes, along with a hue circle in a logarithmic way. My first audiovisual sonification work, *the transition between color and sound* (2014), was based on sonochromatism. In contrast, his pure sonochromatic scale includes ultraviolet and infrared rays, mapped between color and sound frequencies is a non-logarithmic scale. The latest version of Eyeborg supports Bluetooth and Wi-Fi, so that

Harbisson can be connected to other Eyeborg devices remotely [333]. Based on his Eyeborg and sonochromatism, Harbisson has performed *Color Concert* (2013) and exhibited *Color Scores* (2008), which is a printed medium filled with various colors layered in a circular or triangular form.

The examples mentioned above show image scanning methods for sonification. A few examples discussed symbolic mappings between colors and musical instruments. However, those examples only describe several ways to convert images (or color values) to audio and they are not designed to convey the image of a visual shape as a sound. It is also worth pointing out that a common issue of sonification models that rely on color-sound mapping is that it takes time to learn (or memorize) the relationship between color and sound. In addition, cognitive meaning and perception of the image-sound conversion has hardly been discussed.

From now on, a few examples, but examples of sonification particularly designed to convey shape will be introduced. In 1994, Ari Hollander [51] conducted experiments that show the effectiveness of the virtual auditory vector display using a virtual listening space (with headphones) that mimics an array of multiple speakers. Basically, the sound traces the edge of the shape used in his study. In his experiment, a sound was given to the test participants, using headphones, and a participant's task was to identify auditory patterns and match the sounds with the images provided. The author used two types of sound displaying techniques. The first technique was that the sound moved quickly within a fixed shaped area in random or serial order. The fixed shaped area is analogous to the way an old electronic calculator displays numbers. The author named this technique "virtual auditory vector displays." The second displaying technique used the audio matrix, which means that seven by seven virtual speaker arrays were employed and the auditory shapes were displayed onto the virtual speakers. The sound itself does not represent each visual shape as a whole. Instead, the sound comes out of a single point of a virtual position moving around a subject drawing a particular shape. It can be understood that each speaker plays a role as a pixel on a computer screen. These types of the study were already examined by Lakatos [334][335] and Békésy [336] but Hollander

reimplemented a similar test and migrated into the virtual auditory display world with headphones.

The sound used for Hollander's study had 12-partial harmonics with an equal amount of amplitude, with fundamental frequencies ranging between 1000 Hz and 1150 Hz. One test only used a sound at 1000 Hz, but later a combination of four different frequencies was implemented. This is because a sine tone, or white noise [337][338], was often used for this type of test, and it was proven that a sine tone is not ideal for a spatialization method [339][340][341]. Moreover, Hollander tried to reproduce Lakatos' test. For the first display type, there were 16 virtual speaker arrays that displayed the alphanumeric characters '3', '6', 'C', 'O', 'P', and 'U'. The sound moved across each virtual speaker in 60 millisecond intervals, where the sound paused on one speaker before it moved to the next position. For the second display technique, six simple geometric shapes were used. The shapes included a horizontal/diagonal line, a rectangle/triangle with no bottom, and an 'L' shape. The same number of virtual speakers (seven) were activated to display those shapes. Later, the author decreased the number of speaker sources from seven to four and used four different fundamental frequencies including 1000 Hz, 1050 Hz, 1100 Hz, and 1150 Hz.

Overall, the results from the participants were not better than Lakatos' results and there was no significant difference in comparison to random accuracy. I believe that there must be some differences in their sonification method and test environment. The author added that the outcome worsened when the increased numbers of virtual speaker arrays were introduced, because of a human's limited cognitive bandwidth. Later, Hollander displayed sounds onto virtual speakers simultaneously rather than playing them sequentially. However, it failed to prove that the participants' performances were clearly better than chance.

Kamel and others' GUESS [342] expressed basic geometric figures such as a square, circle, and triangle based on their three main techniques. Like other scientific sonification examples, it was created for visually impaired people and GUESS used a virtual 2D

sound environment. The three technologies used the virtual tablet, the virtual sonic grid, and the localization technology. The virtual tablet method allowed one to navigate the sonification area with a pointing device. The virtual sonic grid was sound localization that divided the virtual sound environment into a three by three grid. The x-axis matched the clarinet sound and the y-axis matches the vibraphone sound. This served as an indicator of where the location was explored in the whole image. The basic principle of GUESS was similar to the previous study of Hollander, and the shape was inferred by tracking the sound movement in headphones. A left triangle was described by a tone ascending vertically in the right channel and then moving vertically from the top right to the bottom right, and then moving horizontally from the bottom right to the bottom left. And for vertices, where two lines meet, GUESS created a beep sound. When it came to indicating vertices, this system had some similarity to my sonification method. For more information, see Chapter 4.

Discrete Reconfigurable Aural Matrix (DREAM) [343] was a multi-speaker-driven sonification environment to test human aural perception. DREAM was very similar to Kamel's study above, except that it used real speakers and recruited a similar approach to Hollander's, because users were asked to track the sound movement over the sound source (speakers). DREAM consisted of a 6x4 speaker array and the purpose of design was to treat each speaker as pixels on a digital image. The authors used a line, triangle, rectangle, circle, point, and ellipse to test whether subjects could recognize those shapes. After several sessions, the test participants recorded the answers with a Wii Remote controller. The authors used filtered white noise to secure a rich sound frequency spectrum. DREAM had three modes to display the shapes: vector, surface, and highlight. In vector mode, a bandpass filter was applied to white noise, with a 1KHz center frequency and a 1.4 quality factor, that traveled around the perimeter of the shapes. In surface mode, speakers in the region of the shape applied a bandpass filter, with a 1KHz center frequency and a 0.6 quality factor, to white noise. In highlight mode, DREAM used a bandpass filter with a 500Hz center frequency and a 0.6 quality factor, applied to white noise, at each vertex of the shape. Unlike other sounds for vector and surface modes, the sound of highlight had a sharp attack amplitude envelope. According to the

authors, 16 vertices were applied to a circle and an ellipse which did not naturally have vertices. Highlight mode was tested and combined with vector and surface modes. Their test results showed that a combination of vector, surface, and highlight had the lowest error rate in the shape and size recognition test. In contrast, the surface mode had the highest error rate in the same tests, and the combination of all three modes was noticeably better than the other two modes in the shape test, when tested separately.

DREAM was extended to test concurrent shapes with colors [344]. The test subjects were given images that showed simplified landscape scenes including the sun, sky, water, and land colored in red, cyan, blue, and green, respectively. Each color corresponded to the sound of sizzling steak, wind, running water, and forest, respectively. DREAM adopted a symbolic approach (like auditory icons) to represent these four colors. This study showed that most test subjects were able to understand the overall context of the images. However, this second version of the study did not utilize the original sonification method used in their first study. Instead, users had a chance to navigate the screen with a hand-held input device (Wii Remote) to amplify their pointing area: when amplified, the rest of the sounds were attenuated.

Many of the shape perception sonification models are mainly about sound distribution in accordance with speaker arrays with a designated time lag. From my perspective, “shape perception sonification” is about how sonification can illustrate shapes with sounds specifically designed for their corresponding shapes, whereas in Hollander’s, Kamel’s or DREAM’s study, it was processed by simply tracking the trace of a single sound source regardless of the sounds used. As Hollander already pointed out, his study could be seen as an “auditory driven visualization” rather than “auditory shape perception.” Unlike my research goal, this was not about how sounds could be designed to convey shape information. Rather, the main significance of Hollander’s research was to show cognitive bandwidth in fast-changing auditory motion with headphones aside from using speakers. Moreover, this type of sonification is not appropriate to convey more complex variables such as colors, size, and number. However, since this was an early type of sonification study for human auditory perception in a headphone listening environment, it is worth

introducing here because it was one of the few. Hollander anticipated that further studies will be examined as virtual spatialization technology develops. However, regardless of using headphones or speakers, we still do not know what aspects of visual shapes are difficult to recognize and what the best sounds are to represent visual shapes, due to sparse literature.

In this chapter, the use of graphic-sonification for scientific purposes is mentioned. It has been studied for designing SSD devices mainly focusing on how people successfully identify visual information by listening to sonification, without any visual cue. This implies that these scientific approaches do not necessarily need to be musical or artistic. Further, regardless of the color systems used, these tests are obliged to rely on the test participants' memory and good sense of hearing, because sonification alone is nothing like the genuine experience of seeing. Universal sonification methodologies for SSD do not exist, so far. This aspect could be due to a drawback or limitation of sonification. Or, on the other hand, it could be that SSD is an open area of research because of its relatively short history, compared to other research areas.

2.2.3 Graphic Sonification Models in Musical Instrument

Chapter 1.4.4 illustrated musical instruments with an idea of multi-modality. In this chapter, audiovisual instruments that specifically translate visual information into sounds will be described. This type of instrument can be divided into two main categories, 1) utilizing visual information (e.g., shapes, colors, positions) of digital images to change sound parameters, synthesize sound, create new timbral characters and sound effects, or draw waveforms, and 2) assigning pre-recorded or pre-sampled sounds for certain shapes or colors, like a sampler. Since the latter can be interpreted as an auditory icon, which is a subtype of sonification, those examples will be briefly included in this chapter.

One of the most notable former examples that uses hand drawing as an input source is Kim Ryrie and Peter Vogel's digital sampling synthesizer called the Fairlight CMI [345], created in 1979. When they demonstrated the first version of this synthesizer, the term "sampling" was first coined. Now this synthesizer is considered the first commercially

available digital sampler [346]. In terms of sound storage, sampled sounds were stored on 8x8 inch floppy disks. Since the maximum capacity of each floppy disk was about 80 kilobytes, this system had significantly low sound resolution in overall performances, compared to current digital samplers. The Fairlight CMI included a light pen and computer tube display, so in addition to using pre-recorded sound samples, the waveform of the sound could be drawn by using a light pen on the computer screen. The synthesizer provided 32 harmonic voices total, so within this range, users could draw shapes to create sounds. This technique is very much like wavetable synthesis. The waveforms could be chosen to control other sound parameters like Voltage-Controlled Filter (VCF) and Attack/Decay/Sustain/Release (ADSR). Eventually, the development of this synthesizer was suspended after the 3rd iteration. Peter Gabriel was the first owner [347], and Herbie Hancock demonstrated it in an episode of Sesame Street in 1983 [348]. The Fairlight CMI became widely used by famous musicians including John Paul Jones, Jan Hammer, the Smiths, Boz Burrell, Hans Zimmer, Geoff Downes, Nick Rhodes, Alan Parsons, Keith Emerson, Kate Bush, Rick Wright, Joni Mitchell, Stevie Wonder, and more [347].

Ian Xenakis' UPIC [297][349] is one of the most cited examples. UPIC came with a tablet input device connected to a computer (Figure 2.1) and used the line-based scanning process. A user's drawing could be directly converted to a waveform and then this waveform could be arranged as a compositional object such as an oscillator or an amplitude envelope. The horizontal axis of UPIC represented time, and the opposite axis represented pitch. According to Aphex Twin's interview in the Future Music magazine, "UPIC puts almost everything else to shame. It's under 1MB and it shifts on everyone. [350]"

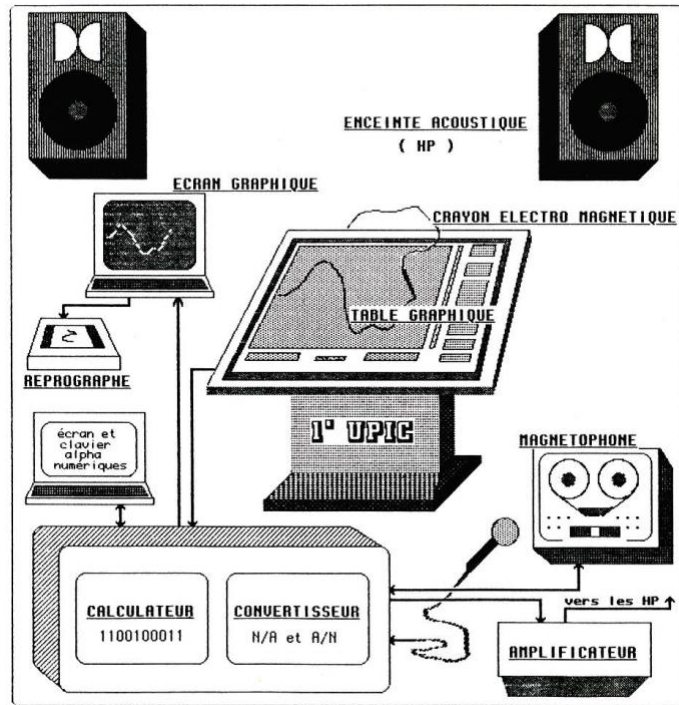


Figure 2.1: UPIC by Xenakis

Source: Adapted from [351]

As computer technology advanced, this type of graphic-sound hybrid instrument evolved into Chris Penrose's Hyperupic [352][353] and other commercial software such as Metasynth [354] (Figure 2.2), Audiopaint [355], Coagula [356], and Photosounder [357]. Metasynth is one of the most popular image synthesizers that ran on the Mac. Metasynth includes a drawing synth toolkit. Like UPIC, the x- and y- axis are coupled with pitch and time, the instrument sounds and the selection of the note scale are changeable depending on users' preference. Metasynth also can transform digital images into spectrograms, and then create sound. Metasynth was commercially used by Aphex Twin on the *Windowlicker* (1999) album [358].

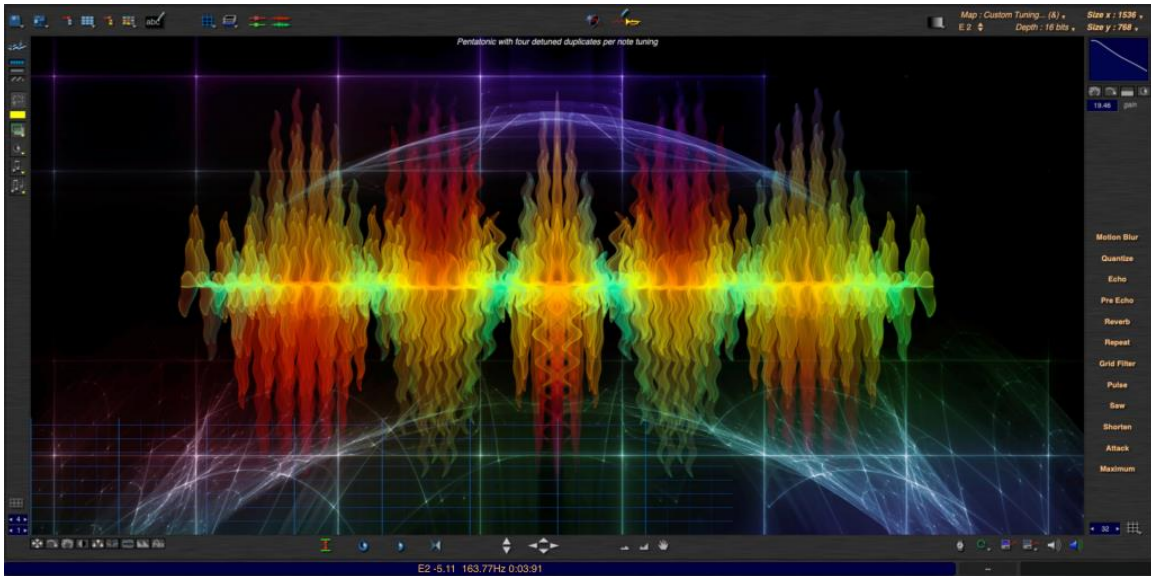


Figure 2.2: Metasynth

Source: Adapted from [354]

Audiopaint uses image files and translates an image's pixel color (RGB or HSB) and position into frequency, amplitude, and stereo panning, using an additive sound synthesis technique. The horizontal axis is mapped to time and the vertical axis corresponds to pitch, so the taller the image, the higher the frequency range. The red and green colors (except blue) of the image manages the amplitude of the left and right audio channels, and overall, the sound gets louder if the pixel values are brighter. However, Audiopaint does not support real-time manipulation.

Coagula is a color organ and an image synthesizer. A drawing is read from left to right, and based on a brush color (red, green, and yellow) is mapped into left, right, and middle stereo panning, in order. In SonicPhoto [359] and Photosounder, the brightness of a pixel affects volume, and a pixel position in the vertical axis changes frequency. Kaleidoscope [360] also takes a similar approach. It scans images from left to right. X-axis is mapped to time, y-axis measures frequency, and the pixel brightness controls amplitude. The graphical manipulation of sound editing and resynthesis in other software such as AudioSculpt [361], QT-Coder in SoundHack [362] and Spear [363] also work in the same way. Similarly, this line scanning pointer also can be found in Adobe Audition

[364]. Like Metasynth, Audition can import 24-bit bitmap images and put them in a spectral display. Identically, the vertical axis indicates frequency and the horizontal axis measures time.

Seiichiro Matsumura's Graph-Sono [365] detected hand drawing lines and converted them into waveshape. An external video camera was attached to a MacBook facing down and captured hand drawing on a paper on the table. Users could save up to three wave shapes and use them to control pitch and amplitude shifting. Likewise, Nakanisynth [366], an iOS synthesizer, created the sound by scanning freehand drawing from the player. The GUI of this synthesizer was simple and straightforward, consisting of two drawing canvases on top and a keyboard in the bottom of the screen. The major principle of Nakanisynth was that freehand drawing was turned into waveforms. The first drawing canvas on the top left of the screen became a sound wave, and the second canvas on the right was used to draw an envelope curve. Each drawing canvas had five empty presets in which the users could save their hand drawing. The keyboard was switchable between a traditional (standard) and ribbon keyboard. Pitch between two notes shifted smoothly with a ribbon keyboard. Nakanisynth was removed from the App Store, so if a drawing-based synth is needed, Shapesynth [367] is an alternative now.

There was an attempt to make a graphic-based musical instrument into a video game. A Japanese interactive media artist, Toshio Iwai, exhibited an audiovisual system *Music Insects* [368] in San Francisco in 1992. A user could select twelve color presets and draw color pixels with its color palette on the bottom of the screen. Those color pixels played a role as a graphic score and when insects crept over those color pixels, it triggered music. The colors played musical notes and each insect had its own pre-sampled instrument sound. This system was initially developed for Nintendo's Sound Fantasy but was canceled [369].

Brian Eno famously merged music into installation and visual art, and later produced interactive visual music iOS apps such as *Bloom* [370], *Trope* [371] and *Scape* [372] in collaboration with a software designer, Peter Chilvers. Later, aside from Eno, Chilvers

also created his own visual music app, *Air* [373], collaborating with vocalist Sandra O'Neill. Conceptually, this app exactly followed Eno's apps. The musical and visual concepts of these apps are self-generating, but users can interact with a finger touch. Eno's and Chilvers's early works, *Bloom* and *Trope*, were created in a similar form in terms of user interactivities. In *Bloom*, users can create ambient music by simply tapping the mobile phone screen. By doing so, circles painted in random colors appear for several seconds, fade away, and then loop infinitely like digital loopers. Sound-wise, it creates a synthesized piano sound. The y-axis of the screen corresponds to pitch. The higher the y position, the higher the pitch. However, the x-axis does nothing, for example, two circles in the same row just have the same color and pitch. *Bloom* has nine moods that come with different color themes. Since only the circles and the background colors are changed based on the color theme settings, the colors do not represent a particular pitch. *Bloom* offers the "listen" mode that automatically creates generative music with infinite visualizations and sounds. Eno said "*Bloom* is an endless music machine, a music box for the 21st Century. You can play it, or you can watch it play itself" [370]. Later in 2018, *Bloom* was expanded to *Bloom: 10 Worlds* [374] with new shapes, sounds, and behavior features.

Next, *Trope* works like *Bloom*, but the main differences are that *Trope* provides twelve color moods and five brush types (e.g., circle, triangle, rectangle, cross, asterisk), which have different tonal characteristics, and support "drag fingers." *Air* is relatively simpler than Eno's other apps because it has eight static triangles, which are adjacent to each other with a fixed position. Each triangle is painted in gradient colors and is randomly assigned to samples of O'Neill's voice and piano sounds. The user's finger touch triggers the samples and changes the colors of each triangle. This app does not incorporate pitch-mapping or allow changing of the sound type. *Air* is a digital sampler that has an abstract visual interface with unique-shaped keys and colors.

Eno's *Scape* is a visual music app album where users can compose music by arranging the prepared shapes on a screen. *Scape* originally comes with nine visual music songs that are pre-composed with abstract visual shapes, that are mapped into twenty-one

musical elements, and five different background colors. Each song has a different visualization and can be played endlessly. Users can listen to the entire album, however, interestingly, users can access these visual shapes through the “create a scape” mode. Users can freely arrange those shapes and compose music on their own. *Scape* is an ambient visual music sampler where the sampled sounds are triggered by predefined abstract visual shapes. All that considered, Brian Eno is fascinated by endless ambient music. His latest album *Reflection* [246] was released in two different forms: a conventional vinyl or CD format, and an iOS app version. The former was pre-recorded for 54 minutes because the storage space was limited. In contrast, the app version of *Reflection* shines because it is pure generative visual ambient music which can be endlessly changing without repeating the same pattern. The music composition varies every time the app is newly played. However, unlike *Bloom*, *Trope*, *Scape*, and *Air*, *Reflection* does not support user interactivity. Eno and Chilvers were able to fully deliver the fluid, flexible, and unpredictable nature of generative visual music into the portable world.

An iOS app, Musical paint [375] is a multi-sensory drawing tool that was originally designed for kids. It consists of a digital drawing canvas and a sound making component, in a simpler form. Eight colored brushes are available on the menu interface, and eight different instrument (pre-sampled) sounds are assigned to each brush. Simply, both x- and y-axis correspond to musical notes. From the bottom-left to the top-left position of the screen, the leftmost column plays all twelve keys as a root key. From there, when a finger moves horizontally, the notes raise in twelve musical notes (e.g., from C2 to C3), so the rightmost positions of the screen play an octave higher than the leftmost positions.

Patatap [376] is a sample-based interactive visual music app that does not require any knowledge or experience in music. Patatap’s screen is divided into twenty-six invisible slots and each slot is assigned to a particular sound sample including percussion, voice, bell, and synth sounds. Each sound is paired with a series of visual effects that consist of abstract and minimalistic animation. When the slots are tapped one by one individually or simultaneously with multitouch, dynamic visual images with effects appear. A line

moving from left to right quickly, a circle growing in size and getting smaller, or multiple dots randomly popping up with various colors are a few examples. This visual-sound kit works similarly to digital music samplers. In a similar context, Takete [377] works as a sampler-based groovebox and audiovisual performance tool to generate visual effects. When the sound preset is triggered, corresponding visual elements appear in the center of the screen, and the visual effects change according to the sound effects applied (e.g., filter, delay, phaser, and reverb).

Some audiovisual apps use math algorithms to generate the visual formation. Scott Snibbe's Bubble Harp [378] forms the visual based on Voronoi diagrams. Voronoi diagrams create the profile of a surface that varies with the distance between the nearest seed points. A user can create music with a finger touch dragging around the touch screen. The shape of bubbles with a white line emerges by following a finger and this slightly animated shape formation generates a harp sound. The line length which connects two seed points represents the pitch of the harp sound.

Musyc [379] combines an icon-based sampler and physics (gravity) together for generative and automated music composition. The sixty-four sampled instrument sounds in the presets are assigned to a perfect circle, a perfect triangle, and a square. The rectangular shape plays a percussive sound. The line worked as a floor (or the ground) that the four shapes fall upon. This line can be drawn tilted or straight. When the line is drawn horizontally and each shape touches the line, the pitch of each sound changes. The right side of the line has a higher pitch, and each shape has a different bouncing behavior. Also, a user can add a black hole that swallows the shapes to remove the sound, the reflection symbol that the shapes could bounce around, and the delay, reverb, overdrive, and compressor effects. Based on what and where the objects were drawn, the sound continuously changes or repeats under the effect of simulated gravity, so Musyc can be seen as a physics-driven music sequencer.

This chapter detailed creative musical instruments for the concept of multimodality in visuals and sound. The design approaches in these examples are mostly based on digital

samplers, color-note mappings and color-musical instrument mappings. Therefore, more possibilities to which various sonification methods can be applied are opening up for future application.

2.3 Propositions

Chapter 2.1 and 2.2 introduced the classification of graphic sonification based on Yeo and Berger's scanning framework, examples of image-sound sonification in a lab environment, and their uses in art, science, and musical instrument design. Judging by the literature review, it is no exaggeration to say that there are no perfect (or universal) models to represent visual images in sounds. As seen in Chapter 2.2, most of sonification conference papers in graphic-sound conversion focus on how to connect colors, pixel values, or other graphic properties to sound parameters. Therefore, sound design approaches based on natural image-sound association (like a bouba/kiki effect) are rare. For artistic/musical examples, sonification has not been used in major electronic music or pop music performances. Above all, shape-driven sonification approaches have never been used for the same purpose either.

Graphic-sound sonification in this dissertation aims to successfully characterize the relationship between graphic design elements and sounds. The graphic design elements refer to basic shapes or a combination of them, which have been traditionally used as the fundamentals of graphic design. The outcomes of this dissertation will culminate in 1) a user study linking the sounds and the images without pre-training (to memorize sound-image relationship) and the reporting of research results, 2) a sonification-driven audiovisual model, partially originate from the user study (mentioned in Chapter 4), and 3) test results that verify how this sonification approach can successfully deliver visual shapes in the form of a sound and how the convergence of graphic design and sonification strengthens communication design.

Due to the short history of sonification, defining a sonification strategy stemming from graphic design language is an unexplored domain, and the use of audiovisuals in this context is even less explored. To my knowledge, there is no similar sonification research,

sonification platform, or audiovisual work to be found using the approach I will propose in Chapter 4. I am not expecting to declare the absolute answer through this study, because of the broad and complex nature of the field. But since exploring the potential of entering a new multimedia design by searching for ideal connections between visual images and sounds is needed, this study is inevitable. I hope that what I explored herein can stretch the possibilities of graphic design, new media art, and sonification. Also, the study of sonification methodology towards the two goals of visual/auditory communication design and artistic expression can be a valuable contribution to the relevant research community.

3 Research Hypotheses

The objectives of this research are to 1) explore the possibility of displaying minimalistic and fundamental visual shapes (commonly found in graphic design) in sonification-driven audiovisual forms, 2) examine whether the relationship between visuals and sound is reasonable and comprehensive to audiences, 3) assess the effectiveness and experience of my sonification system to explore a new approach to visual-auditory communication design, and 4) evaluate the methods to improve future research. The outcome of this sonification gives weight to its artistic expression as audiovisual composition, however, to evaluate the sonification methodologies and examine them objectively, a primary user test was implemented. The sonification platforms and methods (discussed in Chapter 4 and 5) were designed to support each of these hypotheses:

1. Basic waveforms (e.g., sine, triangle, pulse, or a mixture of them) will represent four fundamental visual shapes (circle, equilateral triangle, square, and line), and study participants will be able to match those shapes with the sounds, without pre-training.
2. Sound modulation (e.g., amplitude, pitch, filter, etc.) created from the border of the shapes will describe fundamental shapes (e.g., circle, equilateral triangle, square, and lines) in more detail and will describe more complex shapes (e.g.,

dashed lines, curved lines, and other polygonal shapes), and study participants will be able to match the shapes with the sounds without pre-training.

3. Sound effects (e.g., pitch shift, stereo panning, amplitude changes, etc.) will describe animated expressions of the shapes, and study participants will be able to match the shapes with the sounds without pre-training.

In the next chapter, how sonification was designed to support these hypotheses, how the user tests were conducted, and what the test results were, are described in detail. Based on this sonification method, a new interactive audiovisual instrument was created. This instrument integrates traditional line-by-line sonification (See Chapter 6.3.2) with a newly developed object-oriented sonification. Colors correspond to musical notes, while sound is spatialized based on the location of each shape. Pitch slightly detunes and amplitude varies based on shape rotation and size, and nine voices are added when the brightness increases. Shape movement speed, color change speed, and scanning speed can be controlled. This audiovisual platform is described in Chapter 6.

4 Graphic-Sound Sonification, Experiments and Methodologies

4.1 Introduction

A user study was conducted to support the three research hypotheses mentioned in Chapter 3. In this chapter, how the sound was designed to express four basic shapes (circle, triangle, square, and line) and other customized shapes, and how the user tests were proceeded, and finally, what the test results for each task were will be explained. The sonification methodology for this user study will be the foundation of another sonification platform for audiovisual expression, which will be introduced in Chapter 6.

4.2 User Test

4.2.1 Purpose

In 2001, Ramachandran and Hubbard asked a question about two mimetic words ‘bubba’ and ‘kiki’. Their participant had to guess which of the two shapes represented the words bouba and kiki (Figure 4.1). 95% of the participants chose the left shape for kiki and the right shape for bouba, despite seeing those shapes and words for the first time [380]. The bouba-kiki effect showed sound symbolism of the mapping between speech audio and two shapes. Without any prior memory training, people easily could associate bouba with the round shape and kiki with the spiky shape. The origin of the bouba-kiki effect is the German psychologist Wolfgang Köhler’s experiment which similarly tested the two words ‘Takete’ and ‘Baluba’ in 1929.

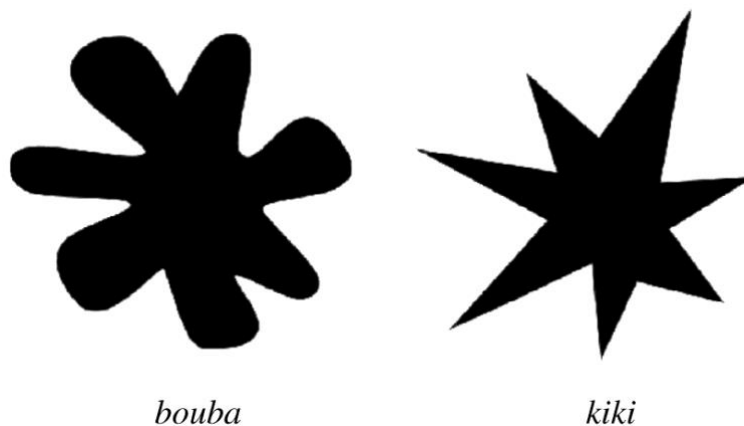


Figure 4.1: Bouba/Kiki Effect

Source: Adapted from [381]

The final goal of this dissertation is to create audiovisual platforms for artistic/musical sonification. However, prior to development, the user study was designed to show that the foundational sonification methodology in the artistic audiovisual platform, derived from the image-sound association test that did not require any pre-training, like the bouba/kiki effect, was effective. This implies that when my sonification is converted into art, even if audiences have not heard the sound-shape relationship before viewing this work, the connection between sounds and shapes can be closely approached unconsciously for the audience.

4.2.2 Participants

This study was open to anyone (including Virginia Tech students, staff, and faculty) above 18 years old who had normal hearing and vision, using headphones and a computer monitor. This study recruited 30 participants in total.

4.2.3 Implementation

The system for the user test mainly consists of three components which are image display, image scanning, and sonification. Three software were used for implementation. For image examples, three basic shapes (i.e., a circle, a triangle, and a square) and three lines (i.e., a sine wave line, a sawtooth wave line, and a pulse wave line) were provided and other customized shapes (i.e., star-like, cloud-like, and stair-like) came after. To draw and display those shapes, Affinity Designer [382] (which is a vector-based drawing tool like Adobe Illustrator) and TouchDesigner were used. Those images feed into the scanning component in TouchDesigner and the image parameters are sent to Max/MSP for sonification. The data from TouchDesigner are transferred to Max/MSP via Open Sound Control (OSC) [383] (Figure 4.3). Basically, the unique timbre for each shape was designed by mixing two of the three basic waveforms (i.e., sine, sawtooth, and pulse). A small dot that acts as a scanning path moves around the shape's border and the x-y coordinates of this point path modulate the sounds that correspond to each shape. For more information, see Figure 4.2.

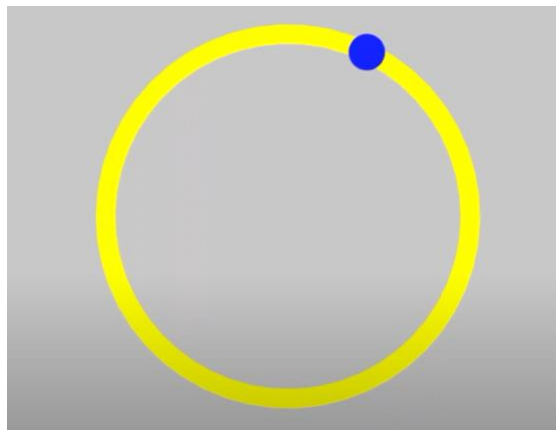


Figure 4.2: A dot scanning path on a circle

The research questionnaires (i.e., pre-questionnaires, task-questionnaires, and post-questionnaires) were implemented via Qualtrics [384], an online research/survey tool. The sonification sounds and their corresponding videos/still images were exported from TouchDesigner and Max. The still images and the sound files were directly shown on Qualtrics, but the videos were uploaded on YouTube first and then embedded on Qualtrics. For more information, see Figure 8.16 and Figure 8.21 in Appendix.

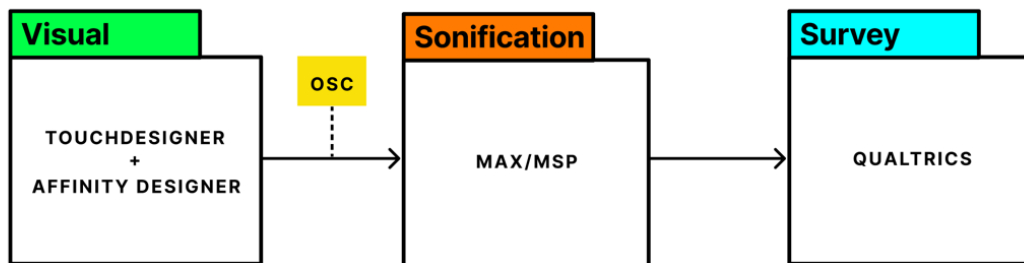


Figure 4.3: User Study Implementation

4.2.4 Study Procedure

The email recruitment flier was sent to Virginia Tech's Graduate School's Web Board, and Center for Human-Computer Interaction (CHCI), School of Visual Art, and School of Performing Arts mailing lists. After appointments were scheduled, the test was held over Zoom and the study questionnaires were provided via Qualtrics online.

This test consisted of the following steps:

1. Participants will have a Zoom link by email.
2. Participants will have a consent information sheet via Qualtrics.
3. Participants will be asked to position their computer monitor or laptop in a way that is comfortable.
4. Participants will answer the pre-questionnaire.
5. Participants will set up headphones (or earbuds).

6. Participants will wear headphones (or earbuds) and adjust volume by finding a comfortable level.
7. When participants feel they are ready, the study will begin (No pre-test is required).
8. Participants will be given a series of test questions. They will hear the sounds paired with the visual images. Multiple-choice and subjective questions will be provided.
9. Participants will answer the post-questionnaire.

When the participant was ready to proceed, the experimenter's Zoom video and voice were completely muted, so as not to disturb participants unwittingly. However, if participants wanted to ask questions in the middle of the study, they were able to talk to the experimenter at any time. When the study was fully completed, the Zoom session was ended and a \$10 Amazon e-gift card was provided to each participant. This study took approximately 35-50 minutes to complete. The screenshots of Qualtrics are attached (see Figure 8.7 in Appendix).

To avoid biased responses from the participants, the study consisted of two versions with a different order of questions in each task. Odd-numbered participants were given version 1 and even-numbered participants were given version 2. Chapter 4 is based on version 1.

4.2.5 Pre-questionnaire

The following pre-questionnaire was designed to collect basic information about the participants. No personal information (i.e., name, gender, exact age) was collected. The questions are as follows:

1. What is your age?
 - 18-24 years old
 - 25-34 years old
 - 35-34 years old
 - 45-54 years old
 - 55 years or older

2. What is the highest level of formal training you have in music?
 - Not at all
 - Middle and high school
 - Undergraduate/Bachelor
 - Graduate/Masters
 - Graduate/Doctoral
 - Self-Trained
 - A Professional
3. What is the highest level of formal training you have in visual art (e.g., studio art, media art, fine art, graphic design, etc.)?
 - Not at all
 - Middle and high school
 - Undergraduate/Bachelor
 - Graduate/Masters
 - Graduate/Doctoral
 - Self-Trained
 - A Professional
4. What is your listening device for this study?
 - Headphones
 - Earphones

4.2.6 Study1: Task and Sound Design

The first research hypothesis is that basic waveforms (e.g., sine, triangle, and pulse) can be used to represent the characteristic of four fundamental visual shapes (circle, triangle, square, and line), and that study participants are able to match those shapes with sounds without pre-training. Study1 is designed to support the first hypothesis and four basic visual shapes and their corresponding four sounds were provided. Study participants were asked to listen to the sounds and choose which sound best describes the image shown.

There are several sound synthesis techniques for sonification. Efficient delivery of data in auditory forms requires a careful and sensible approach. *The Sonification Handbook* introduces some of them: parametric (non-parametric) models, digital audio, Fourier synthesis, modal synthesis, subtractive synthesis, time-domain formant synthesis, waveshaping synthesis, FM synthesis, granular synthesis, physical modeling, and Non-Linear Physical models [43]. The sound synthesis methods used in this dissertation work include additive synthesis (for oscillators), subtractive synthesis (because of the filters) and waveshaping synthesis (this only applies to the artistic sonification model in Chapter 6). Waveshaping synthesis is based on PCM synthesis because waveforms need to be stored as samples. To put this more simply, the sound design approach in this dissertation has the same context as that of traditional synthesizers. The beginning of my sonification design starts with the consideration of musical expression in the future, therefore, the sonification platform itself is intended to be regarded as a musical instrument.

Sine, sawtooth, and square waves are the basic oscillators commonly found in almost all types of synthesizers. Based on them, a sine wave was used to represent a circle, a sawtooth wave to represent a triangle, and a square wave to represent a square. More specifically, a circle is a combination of two sine waves, a triangle is a combination of two sawtooth waves, and a square is a combination of two square waves. The line is based on the sound of the square, but it is designed to give the feeling of a very thin rectangle by excluding the thick and bold feeling of the original square sound, by raising its pitch and using different pulse width settings. The sound synthesis method (for four different timbres) is based on additive synthesis because the sawtooth wave and square wave also are comprised of sine waves. Of course, it cannot be said that a sawtooth wave produces a triangular sound because the sound of the shape does not exist. However, those waveforms were scientifically named after the shape of their physical waveforms in a x-time/y-amplitude plot. So, I wondered if people could unconsciously or consciously associate the relationship between the four fundamental shapes and three basic waveforms when they were paired. It is unlikely that it would be impossible, because each waveform has distinctly different sound characteristics. To recap, the shape-to-sound mapping in Study1 is approached based on the shape of three basic waveforms.

Currently, similar sonification methods for image-sound association tests have not been found, so if this research outcome is successful, it will be a meaningful discovery of this methodology. For the specific mapping information, see Table **4-1**.

The ramp value of a sawtooth wave increases from 0 to 1 linearly and a pulse wave alternates between 0 and 1. A sine wave can be discriminated also because the vertical movement between -1 and 1 does not move linearly. This obviously makes a difference if those waveforms are played at a lower frequency. When playing at a very high frequency, it becomes difficult to distinguish between a sawtooth and a pulse wave. Since we cannot hear frequencies below 20Hz, if those basic waveforms are used as an LFO to modulate the sound, the difference between each waveform becomes more apparent.

To explain the sound design of this study with the concept of earcons and auditory icons which are two common taxonomies of auditory display, earcons use synthesized sound to alert events that needs attention (i.e., sound messages on smartphones) but auditory icons use sounds that we can instantly associate with a real world event (i.e., a paper crumpling sound for “empty trash” on computers) [43]. The sound design approach in this study is characterized by the coexistence of both concepts. For earcons, the sound is based on synthesized tones or sound patterns and does not use realistic sounds like auditory icons, however, the user study does not require the pre-training that earcons would. So, the sound design falls into the earcons category, but the sounds incorporate some characteristics of auditory icons, because this study aims not to have any pre-training.

Another important point is that the sounds are designed by mixing two identical waveforms (with a slightly different setting for each) instead of using a single waveform. This is to create richer frequency components due to the limitations of a sine wave, which has only one frequency component. I ran a pilot test with 5 participants to see if my approach is comprehensible to others. The sounds in Table **4-1** were given and the participants were asked to pair the sounds with their corresponding images (i.e., circle, triangle, square, and line). Four out of five participants paired them correctly. However, when only unmixed waveforms were provided with the images, three out of five

participants were confused about the sawtooth wave and pulse wave pairing with a triangle and a square. All participants associated the sine wave with a circle successfully. Based on this pilot test, I concluded that combining two waveforms could help in describing and emphasizing each soundwave's character (e.g., rough, rounded, and edged) better. For the triangle, the square, and the line, the sounds were also composed of a combination of two corresponding sound waves (i.e., two sawtooth waves and two pulse waves). This approach can be found on traditional synthesizers where one can blend more than two sound oscillators.

The questions in Study 1 cover a total of four images (Figure 4.4) and four sounds. Each question has one image and four sounds (i.e., Sound A, Sound B, Sound C, and Sound D). Sound A, Sound B, Sound C, and Sound D are the sounds for the circle, the line, the triangle, and the square in order. Participants first saw the image, listened to all four sounds, and then chose the answer. The order of the sounds is the same for all four questions in Study 1, but the order of the questions is different in version 1 and version 2. After that, Likert questions were given on a 5-step scale to receive feedback on how close each sound matches with the image. Table 4-1 shows the sound-shape mapping of each basic shape in Study 1. Please see Chapter 5.2 for the study results.

Table 4-1: Sound-Shape Mapping in Study1

Shapes	Circle		Triangle		Square		Line	
	Sine	Sine	Saw	Saw	Pulse	Pulse	Pulse	Pulse
Frequency (Hz)	261.6	130.8	261.6	130.8	261.6	65.41	261.6	523.3
Mix (%)	60	40	70	30	60	40	60	40
Pulse Width	N/A	N/A	N/A	N/A	0.5	0.3	0.1	0.2

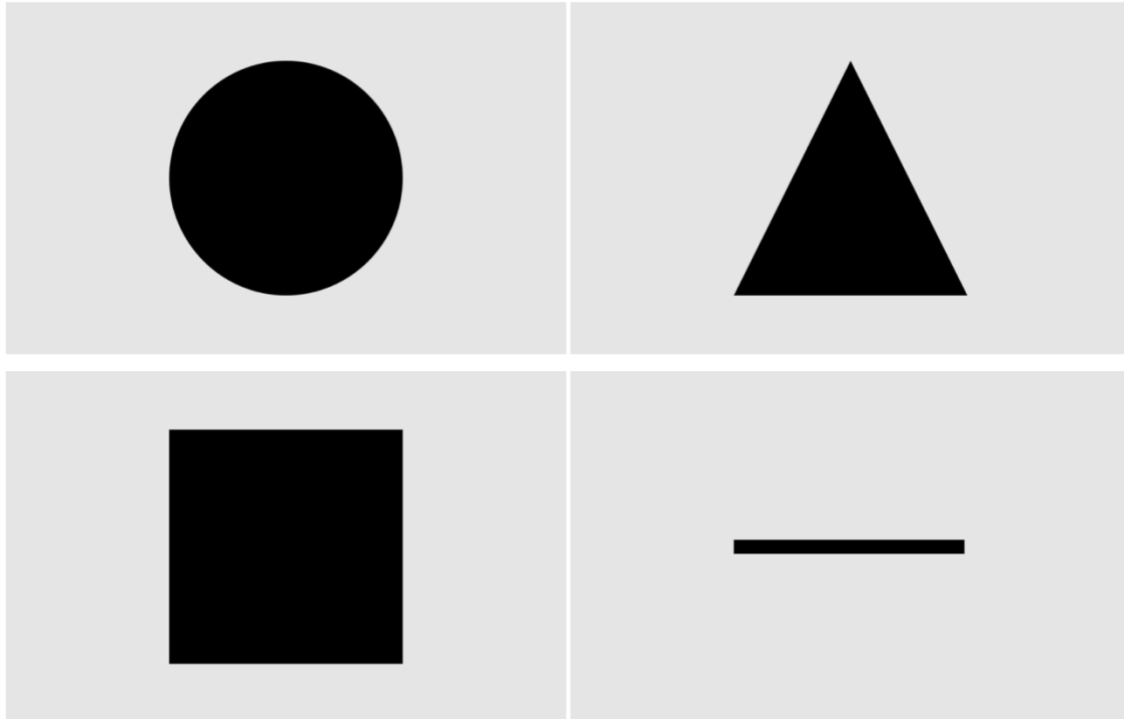


Figure 4.4: Four basic shapes for Study 1

The frequency of the main oscillator (the left column of the waveforms) is middle C (C4). Each shape has an additional waveform that acts as a sub oscillator. The circle is designed to give a mellow, warm, and soft feeling by adding C3 (an octave lower), and the triangle also adds C3 (an octave lower) to emphasize its harsh and rough feeling of a sawtooth wave. The lower the frequency, the clearer the sound character between the waveshapes. In contrast, for example, the higher the frequency, the more difficult it is to distinguish between a sawtooth and a pulse wave. For the square shape, C2 (two octaves lower) is added instead of C3, to make a clearer distinction from the sawtooth wave-based triangle. The sound design for the line was planned to give the feeling of a very thin rectangle by adding C5 (an octave higher). However, the pulse width setting between the square and the line is different to give the square a thick and heavier feeling, and the line a thin and light feeling.

4.2.7 Study 2 and Study 3: Task and Sound Design

The second research hypothesis is that sound modulation (e.g., amplitude, pitch, filter, etc.) driven by tracing the border of the shapes can be used to describe fundamental shapes (e.g., a circle, a triangle, a square, and lines) in more detail and can be extended to describing more complex shapes (e.g., dashed lines, curved lines, and other customized shapes), and study participants are able to match the shapes with the sounds without pre-training. Study 2 and Study 3 were designed to support this hypothesis. Three basic shapes and three sounds were provided for Study 2 and four line types and four corresponding sounds were provided for Study 3. The study participants were asked to listen to the sounds and choose which sound best describes the image shown.

The focus of the second user study was the application of sound modulation along the edges of the shape. This approach can be found in traditional synthesizers, so this idea originally started from my synth learning experience. For example, the ramp value of a sawtooth wave increases from 0 to 1 linearly and a pulse wave only moves from 0 to 1. A sine wave can be discerned because the vertical movement between -1 and 1 does not move linearly. When those basic waveforms were used as an LFO (Low Frequency Oscillation) to modulate sound sources by lowering its frequency (below 5Hz for example), it was possible to infer what waveforms were used. Since we can hear the tonal changes slowly and periodically, it would not be difficult to perceive this change without special training. This approach for Study 2 is unprecedented, so if the test results are successful, it will be a meaningful finding as a preceding study, and it will be an opportunity to develop a more advanced study in the future.

The tonal character (timbre) of the shapes in Study 2 was the same as in Study 1, but the key was that x-y coordinates of the border of each shape modulate the sound to describe the shapes more clearly. Many attempts had been made before ending up with this sound design. The first three modulation mappings I applied were pitch, amplitude, and filter-related values including the center frequency and the quality factor (Q factor). Pitch mapping is a common sonification parameter [385], however, it was not a good choice for this study because heavy reliance on pitch modulation could lead to problems with the

musical expression platform (see Chapter 6). The platform will introduce color-pitch mapping, so this mapping should overlap. Amplitude mapping was not an ideal scenario too because it is known that the perception of loudness is less discriminable in comparison to pitch because loudness does not offer an effective display dimension [385]. Filter modulation was tried next. For stock market and oceanographic data sonification, Ben-Tal and Berger [386] created a vowel-like sound from white noise and a pulse wave by manipulating the center frequency of the filter and varying the bandwidth settings of each individual filter to examine whether participants could detect variable changes in data through sound. As was proven, filter modulation can be used as an important attribute for sonification. When this filter modulation was solely used, a more distinguishable tonal change was given compared to the first two, but the distinction between sine wave and sawtooth wave modulation was not clearly noticeable. This may have something to do with the speed of the modulation, but it seemed difficult for people who were not trained or had less sensitive ears to tell the difference between sine wave modulation and sawtooth wave modulation. I ran a brief pilot test with five participants on these four modulation types (i.e., pitch-only, amplitude-only, filter-only, and all three combined). Interestingly, the participants felt the strongest sound modulation when all three types were combined. However, the distinction between sine wave and sawtooth wave modulation was still the most obscure point. So, I came up with a new idea that emphasizes the shapes more clearly when the scanning path reaches vertices of the shapes. That is, when the vertices are reached, Q-factor dramatically rises. Since sine curves have no vertices, the difference between sine wave and sawtooth modulation will be easier to distinguish.

EdgeSonic [304] (previously mentioned in Chapter 2) detected and sonified image edges. For example, if a mountain picture was given, a black line along the border of the mountain was created and based on the black line, a simple sine wave sound on a x-time/y-frequency mapping was played, like the drawing-based waveshaping synth UPIC. Edge Sonic was designed for the blind and their user study showed that the study participants could infer the shape of the line. A similar approach can be found in bar-chart sonification. Flowers and Hauer [311] depicted line graphs (x-time/y-level) with

sound to examine if their auditory graphs could be used to substitute traditional visual graphs. The line graphs included linear flat, linear positive, linear negative, etc. For another example, DREAM [343] and Kamel's GUESS [342] (introduced in Chapter 2) also put an emphasis on the shapes' vertices. However, those examples are different from my approach because, in their works, another alarm sound (i.e., beep sound) was newly added along with the original sound representing each shape. This alarm sound was just triggered when the vertices were reached.

For one more interesting example, Thoret and others [387] synthesized sounds by using the friction of a pencil touching the paper when a person drew. In other words, when drawing the curvature of a shape (e.g., circles, ellipses, loops, and arches), the varying velocity of the friction was applied to the sound. The study participants tested whether they could relate the sounds to the shapes. The success rate was different for each shape, but it had a similar context to my study in that it linked the border of the shape to the sound.

Study 2 presented a total of three images (Figure 4.5) and three sounds that corresponded to each image. Each question had one image and three sounds (i.e., Sound A, Sound B, Sound C). Sound A, Sound B, and Sound C were the sounds for the triangle, the circle, and the square in order. Participants first saw the image, listened to all three sounds, and then chose the answer. After that, Likert questions were given on a 5-step scale to receive feedback on how close each sound matched with the image. Table 4-2 below shows the sound-shape mapping of each basic shape in Study 2. See Chapter 5.3 for the study results.

Table 4-2: Sound-Shape Modulation Mapping in Study2 and Study 3

Shapes	Circle	Triangle	Square	Line (Sine)	Line (Saw)	Line (Pulse)	Line (Dotted)
Filter Number	1 LPF*	1 HPF 1 Peak*	4 Peak	1 HPF*	1 HPF	1 HPF 1 Peak	NA

Filter Gain	1	HPF: 1.5 Peak: 2	2	1	1	HPF: 0.5 Peak: 2	N/A
Cutoff Frequency Range (Hz)	-100 - 100	-100 - 100	-130 - 130	N/A	N/A	N/A	N/A
Q-Value	1	1.5	2	1.5	1	HPF: 1 Peak: 0	N/A
Q-Value Sweep	-0.3 to 0.3	HPF: 1.3 Peak: 2	3	0.3	1.2	HPF: 1 Peak: 2	N/A
Amp Mod (Dry/Wet)	80/20	50/50	80/20	50/50	50/50	50/50	N/A
Amp Mod Range	-0.5-0.5	-0.5-0.5	-0.5-0.5	-0.5-0.5	-0.5-0.5	-0.5-0.5	N/A
Pitch Mod Range	0-20 cents	0-40 cents	0-50 cents	N/A	N/A	N/A	N/A

*LPF: Low Pass Filter, Peak: Peak Notch Filter, HPF: High Pass Filter



Figure 4.5: Three basic shapes for Study 2

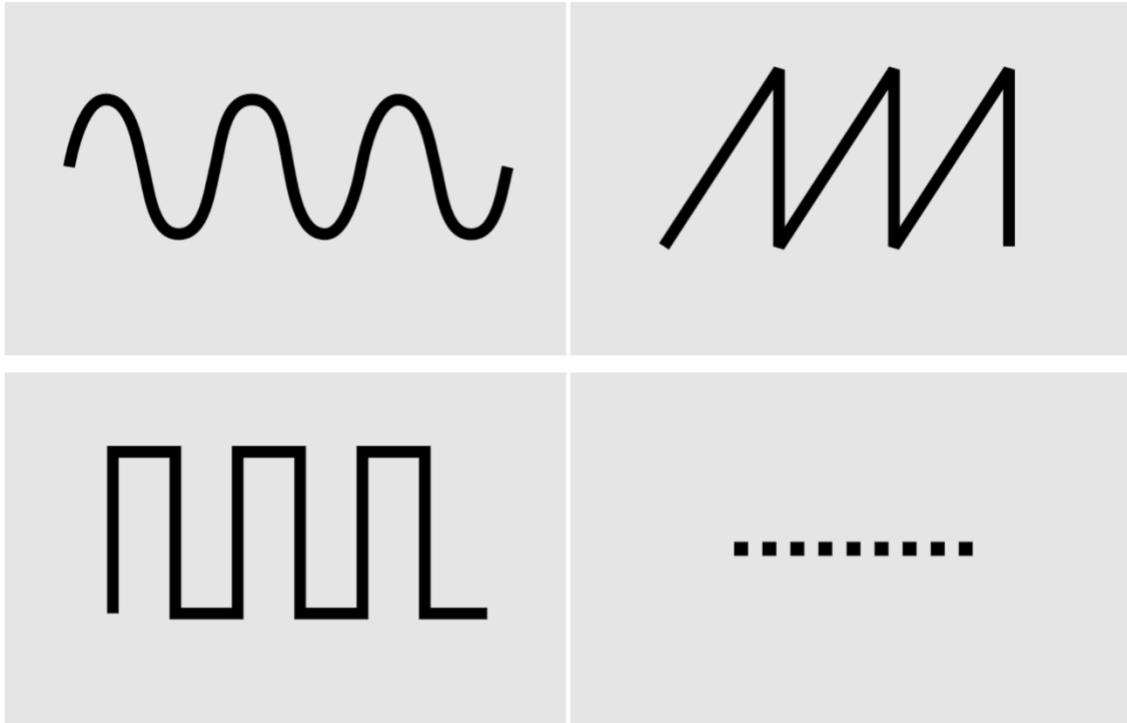


Figure 4.6: Four lines for Study 3

Study 2 and Study 3 introduced a new feature; the timbre of each shape was modulated according to the x-y coordinates of the shape's border. The timbre of each shape was succeeded from Study 1. The sound modulation included amp modulation, pitch modulation, and filter modulation. Also, for filter modulation, the center frequency, the filter gain, the initial Q-value, and the movement of the Q-value were controlled (see Table 4-2 for more information). A different type of filter was applied to each shape because filter types can affect the overall tonal character of each shape, making them sound warm, round, rough, or sharp, for example. The amp modulation was also adjusted so that all the sound could be heard as much as possible, because modulation can result in significant volume loss. The other modulation parameters had slightly different settings according to the shapes, but these different settings were used to fine-tune the sound, to describe each shape more effectively.

The questions in Study 3 were given a total of four images (see Figure 4.6) and four corresponding sounds. Each question had one image and four sounds (i.e., Sound A,

Sound B, Sound C, and Sound D). Sound A, Sound B, Sound C, and Sound D were the sounds for the sine curve, the dashed line, the sawtooth line, and the pulse wave line, in order. Participants first saw the image, listened to all three sounds, and then chose the answer. After that, Likert questions were given on a 5-step scale to receive feedback on how close each sound matched with the image. Table 4-2 above shows the sound-shape mapping of each basic shape in Study 3. See Chapter 5.3 for the study results.

4.2.8 Study 4: Task and Sound Design

Study 4 followed the same sonification mapping except with different visual shapes. Study 4 gave a total of three customized images and three sounds that correspond to each image. Each question had one image and three sounds (i.e., Sound A, Sound B, and Sound C). Sound A, Sound B, and Sound C were the sounds for the star-like shape, the cloud-like shape, and the reversed stair-like shape, in order. The timbre of each shape was succeeded from Study 1 and the modulation for each shape was succeeded from Study 2. That is, the star-like shape used the same timbre and modulation from the triangle, the cloud-like shape used the same timbre and modulation from the circle, and the reversed stair shape used the same timbre and modulation from the square, in the previous studies. Participants first saw the image, listened to all three sounds, and then chose the answer. After that, Likert questions were given on a 5-step scale to receive feedback on how close each sound matches with the image. See Chapter 5.4 for the study results.



Figure 4.7: Three custom shapes for Study 4

4.2.9 Study 5: Task and Sound Design

The third research hypothesis is that sound effects (e.g., pitch shift, stereo panning, amplitude change, etc.) can be used to describe animated expressions of the shapes, and study participants are able to match the shapes with the sounds without pre-training. Study 5 was designed to support this hypothesis and three moving shapes and three sounds corresponding to each shape were provided. The study participants were asked to watch the video examples first and then choose which sound best describes the moving image shown.

Visual arts are not always static, so movement is a major consideration when sonifying visual arts. Michel Chion described the role of sound and movement in motion pictures by saying that "...one perception influences the other and transforms it. It is not uncommon in the film industry for people to view film sound as an 'add-on'." [388]. Iyamita and others [389] investigated congruency between sound effects and switching visual patterns. For example, an ascending musical scale can follow a shape that increases its size in the center of the screen, and a descending scale can be mapped with a shape that decreases its size. Also, pitch rising with bigger images and falling with smaller images can be a reasonable mapping. Walker [385] mentioned that dynamic auditory stimuli seems to work similarly to dynamic visual stimuli, so for better sonification strategies, it would be important for sonification designers to understand the interaction between those two stimuli.

Cooley [390] tested the interaction between sound and moving images. His user study consisted of experiments that presented one image and asked participants to choose various sounds. Another option presented similar sounds and asked participants to choose a different image. Listeners responded to the test questions by combining different variations of images and sounds pairs. The result was that not only do different sounds affect the image, but the image can also change the perception of sound. Their first experiment showed an animation where two squares collide and let the listeners see the collision while hearing different sounds. The second experiment showed a triangle passing through a screen and, depending on the sound, the triangle looked like a person

with personality, attitude, and emotions. The third experiment was used to determine the distance between the listener and the sound source according to the size of the image. Their third experiment has a context similar to my Study 5.

Mingham and others [391] mentioned three types of classification for information delivery. 1) Symbolic makes an association between events and data, and it requires a pre-test to memorize. This tends to use conventional meaning. 2) Nomic has a physical association between data and events. 3) Metaphorical does not use physically related information to present but rather it uses similarities (i.e., increasing pitch for climbing a hill). Based on this classification, the significance of study 5 is to examine the movement of the shape and its metaphorical sound effects. The main purpose of Study 5 is to see how users can associate the relationship between the moving shape and its corresponding sound. The same triangle shape from the previous study was used and three sound effects (i.e., stereo panning, pitch change, and volume change) were applied to three different movements: left-right, rotation, and size.

Study 5 was designed to prepare for the artistic sonification platform (See Chapter 6). For the art platform, all objects will not be statically positioned. The shapes will move up and down and left and right, they will increase and decrease in size, and they will rotate as well. Study 5 aimed to examine whether congruency between the sounds and the movements in my sonification can be comprehensive or universal. Table 4-3 below shows the mapping between the movement and the sound-FX.

In Study 5, a triangular sound (with modulation) was used as a representative shape. The triangle with the left and right movement matches the sound with stereo panning. The size of the triangle changes volume and modulation intensity, i.e., the larger the size, the larger the amplitude, and the smaller the size, the smaller the amplitude. It was natural for the volume and the modulation intensities to decrease because the smaller the shape size, the smaller the range of sound modulation derived from the border of the shape. Finally, the rotating triangle was pitched up when it turned to the right and pitched down when it turned to the left.

Study 5 gave a total of three moving shape images (i.e., Movement A, Movement B, and Movement C) and three sound effects that corresponded to each moving image. Each question had one YouTube video that showed one movement synced with three different sound effects (i.e., volume intensity, stereo panning, pitch shifting). Movement A was moving left to right, Movement B was rotating, and Movement C was zooming in and out. The first question showed Movement A along with volume intensity, stereo panning, and pitch shifting, in order. The second question showed Movement B along with pitch shifting, volume intensity, and stereo panning, in order. The third question showed Movement C along with pitch shifting, volume intensity, and stereo panning, in order. Participants first saw the moving image, listened to all three image-sound pairs, and then chose the answer. After that, Likert questions were given on a 5-step scale to receive feedback on how closely each sound matched with the image. Table 4-3 below shows the sound-shape mapping in Study 5. See Chapter 5.5 for the study results.

Table 4-3: Sound FX-Shape’s Movement Mapping in Study5

Movement	Moving left to right	Zooming in and out	Rotating
Sound FX	Stereo Panning	Volume and Modulation Intensity	Pitch Shifting (-200 to 200 cents)

4.2.10 Post-Questionnaire

The post-questionnaire collected feedback from the participants after the completion of the study. The questions were as follows:

1. I could easily match the sound with the visual.
 - Strongly agree
 - Somewhat agree

- Neither agree nor disagree
 - Somewhat disagree
 - Strongly disagree
2. The study was fun and interesting.
 - Strongly agree
 - Somewhat agree
 - Neither agree nor disagree
 - Somewhat disagree
 - Strongly disagree
 3. The sounds enhanced the visual information and provided multi-sensory experience (i.e., you experienced visually and auditory).
 - Strongly agree
 - Somewhat agree
 - Neither agree nor disagree
 - Somewhat disagree
 - Strongly disagree
 4. What do you like the most about the study?
 5. Do you think this sonification (the use of sound to convey non-audio information) method has the potential in communication design, art, or education? If so, which areas would you see it being most useful (Select all that apply)?
 - User Interface Design
 - Interaction Design
 - Assistive Technology
 - Human Computer Interaction
 - Education
 - Art
 - Music
 - Other
 6. Why do you think so?

7. What task was the most challenging to you and why? (You can skip this question if the study questions were easy to answer).
8. Please share any other thoughts if you have them. This will help me improve this sonification system for future use.

Please see Chapter 5.7 for the answers.

5 Data Analysis

5.1 Results (Pre-Questionnaire)

The age distribution of the 30 participants was 8 (26.67%) between the ages of 18-24, 15 (50%) between the ages of 25 and 34, 5 (16.67%) between the ages of 35-44, 1 (3.33%) between the ages of 45-54, and 1 (3.33%) over the age of 55. When recruiting participants, their specific age was not considered. See Table 5-1 for more information.

Table 5-1: Pre-questionnaire 1

What is your age?

Age	Freq	Perc(%)
18-24 years old	8	26.67
25-34 years old	15	50.00
35-44 years old	5	16.67
45-54 years old	1	3.33
55 years or older	1	3.33

Next, music and visual art education levels were asked. Regarding music training, middle and high school had the highest number of 10 (33.33%) students, and no visual art training was the most with 12 (40%) people. See Table 5-2 for more information. The participants used their own headphones or earphones as listening devices, and there were 18 (60%) headphone users and 12 (40%) earphone users. See Table 5-3 more

information. However, this study was not designed to find out whether these listening devices affect test results.

Table 5-2: Pre-Questionnaire 2 (Left) and 3 (Right)

What is the highest level of formal training you have in music?			What is the highest level of formal training you have in visual art?		
Level	Freq	Perc(%)	Level	Freq	Perc(%)
Not at all	9	30.00	Not at all	12	40.00
Middle and high school	10	33.33	Middle and high school	7	23.33
Undergraduate/Bachelor	4	13.33	Undergraduate/Bachelor	2	6.67
Graduate/Masters	5	16.67	Graduate/Masters	6	20.00
Graduate/Doctoral	1	3.33	Graduate/Doctoral	0	0.00
Self-trained	1	3.33	Self-trained	3	10.00
A Professional	0	0.00	A Professional	0	0.00

Table 5-3: Pre-Questionnaire 4

What is your listening device for this study?

Device	Freq	Perc(%)
Earbuds	12	40
Headphones	18	60

5.2 Results (Study 1)

5.2.1 Analysis and Findings

The answer to the first question of Study 1, which was the sound corresponding to the circle, was Sound A. 26 participants selected the correct answer. The correct answer rate was 86.67%. See Table 5-4 for more information. For the Likert question following every first question, the participants scaled how well each sound (from Sound A to Sound D) fit with the image. The answer that Sound A was “very likely” or “likely” was the most at 76.67% and 20%, respectively. But interestingly, 40% of the participants chose “likely”

for Sound D, which was the sound of the square (see Table 5-5). This was probably because the frequency of the square's sub oscillator was the lowest at C2, so the participants may want to associate the low-pitched sound with the rounded shape of the circle.

Table 5-4: Answer to question 1 of Study 1

Task 1.1

Sound	Freq	Perc(%)
Sound A	26	86.67
Sound B	0	0.00
Sound C	0	0.00
Sound D	4	13.33

Table 5-5: Likert question 1 of Study 1

Task 1.1: Sound A

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	1	3.33
Neutral	0	0.00
Likely	6	20.00
Very likely	23	76.67

Task 1.1: Sound B

Degree	Freq	Perc(%)
Very unlikely	14	46.67
Unlikely	13	43.33
Neutral	3	10.00
Likely	0	0.00
Very likely	0	0.00

Task 1.1: Sound C

Degree	Freq	Perc(%)
Very unlikely	9	30.00
Unlikely	14	46.67
Neutral	6	20.00
Likely	1	3.33
Very likely	0	0.00

Task 1.1: Sound D

Degree	Freq	Perc(%)
Very unlikely	3	10.00
Unlikely	5	16.67
Neutral	8	26.67
Likely	12	40.00
Very likely	2	6.67

The answer to the second question of Study 1, which was the sound corresponding to the triangle, was Sound C. 14 participants selected the correct answer. The correct answer rate was 46.67%. See Table 5-6 for more information. For the Likert question following every first question, the answer that Sound C was “very likely” or “likely” was the most at 33.33% and 36.67%, respectively. And 26.67% of the participants chose “very likely” for Sound D, the second highest number. See Table 5-7 for more information.

Table 5-6: Answer to Question 2 of Study 1

Task 1.2

Sound	Freq	Perc(%)
Sound A	4	13.33
Sound B	10	33.33
Sound C	14	46.67
Sound D	2	6.67

Table 5-7: Likert Question 2 of Study 1

Task 1.2: Sound A

Degree	Freq	Perc(%)
Very unlikely	17	56.67
Unlikely	6	20.00
Neutral	2	6.67
Likely	2	6.67
Very likely	3	10.00

Task 1.2: Sound B

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	6	20.00
Neutral	5	16.67
Likely	11	36.67
Very likely	8	26.67

Task 1.2: Sound C

Degree	Freq	Perc(%)
Very unlikely	1	3.33
Unlikely	5	16.67
Neutral	3	10.00
Likely	11	36.67
Very likely	10	33.33

Task 1.2: Sound D

Degree	Freq	Perc(%)
Very unlikely	4	13.33
Unlikely	6	20.00
Neutral	5	16.67
Likely	12	40.00
Very likely	3	10.00

The answer to the third question of Study 1, which was the sound corresponding to the square, was Sound D. 16 participants selected the correct answer. The correct answer rate was 53.33%. See Table 5-8 for more information. For the Likert question following every first question, the answer that Sound D was “very likely” or “likely” was the most at 26.67% and 46.67%, respectively. The sum was the highest. However, 36.67% of the participants chose “very likely” for Sound C. Considering only “very likely”, the selection rate was higher than for Sound D. For more information, see Table 5-9.

Table 5-8: Answer to Question 3 of Study 1

Task 1.3

Sound	Freq	Perc(%)
Sound A	1	3.33
Sound B	1	3.33
Sound C	12	40.00
Sound D	16	53.33

Table 5-9: Likert Question 3 of Study 1

Task 1.3: Sound A

Degree	Freq	Perc(%)
Very unlikely	9	30.00
Unlikely	8	26.67
Neutral	5	16.67
Likely	6	20.00
Very likely	2	6.67

Task 1.3: Sound B

Degree	Freq	Perc(%)
Very unlikely	6	20.00
Unlikely	11	36.67
Neutral	7	23.33
Likely	6	20.00
Very likely	0	0.00

Task 1.3: Sound C

Degree	Freq	Perc(%)
Very unlikely	2	6.67
Unlikely	2	6.67
Neutral	11	36.67
Likely	4	13.33
Very likely	11	36.67

Task 1.3: Sound D

Degree	Freq	Perc(%)
Very unlikely	1	3.33
Unlikely	4	13.33
Neutral	3	10.00
Likely	14	46.67
Very likely	8	26.67

The answer to the fourth question of Study 1, which is the sound corresponding to the line, is Sound B. 13 participants selected the correct answer. The correct answer rate was 43.33%. See Table 5-10 for more information. For the Likert question following every first question, the answer that Sound B was “very likely” or “likely” was the most at 33.33% and 23.33%, respectively. The sum was the highest. For more information, see Table 5-11.

Table 5-10: Answer to Question 4 of Study 1

Task 1.4

Sound	Freq	Perc(%)
Sound A	7	23.33
Sound B	13	43.33
Sound C	4	13.33
Sound D	6	20.00

Table 5-11: Likert Question 4 of Study 1

Task 1.4: Sound A

Degree	Freq	Perc(%)
Very unlikely	3	10.00
Unlikely	6	20.00
Neutral	6	20.00
Likely	10	33.33
Very likely	5	16.67

Task 1.4: Sound B

Degree	Freq	Perc(%)
Very unlikely	2	6.67
Unlikely	2	6.67
Neutral	9	30.00
Likely	7	23.33
Very likely	10	33.33

Task 1.4: Sound C

Degree	Freq	Perc(%)
Very unlikely	5	16.67
Unlikely	4	13.33
Neutral	9	30.00
Likely	9	30.00
Very likely	3	10.00

Task 1.4: Sound D

Degree	Freq	Perc(%)
Very unlikely	8	26.67
Unlikely	9	30.00
Neutral	3	10.00
Likely	4	13.33
Very likely	6	20.00

The question about the circle was 86.67% accurate, while triangles, squares, and lines were somewhat less accurate. As observed in the pilot test mentioned in Chapter 4.2.6, the matching between sine wave and the circle showed a 100% accuracy rate even with a single sine wave. This was probably because a sine wave produces a soft sound and conveys the feeling of round shapes, whereas a sawtooth and a pulse wave seem to give a rather vague impression to intuitively relate to the characteristics of the triangle and the square, due to the complexity of the frequency component. The harmonic diagram (see Figure 5.1) shows that there is no significant difference visually, as well. Musically trained people were accustomed to distinguishing between sawtooth and square waves, but inexperienced participants may find it ambiguous. Although the correct answer rates for questions 2, 3, and 4 decreased, most participants picked the correct answer. Since these issues were anticipated prior to designing this study, the use of sound modulation was introduced as a better way to increase accuracy regardless of the complexity of each waveform. See Chapter 5.3, Chapter 5.4, and Chapter 5.5 for more information.

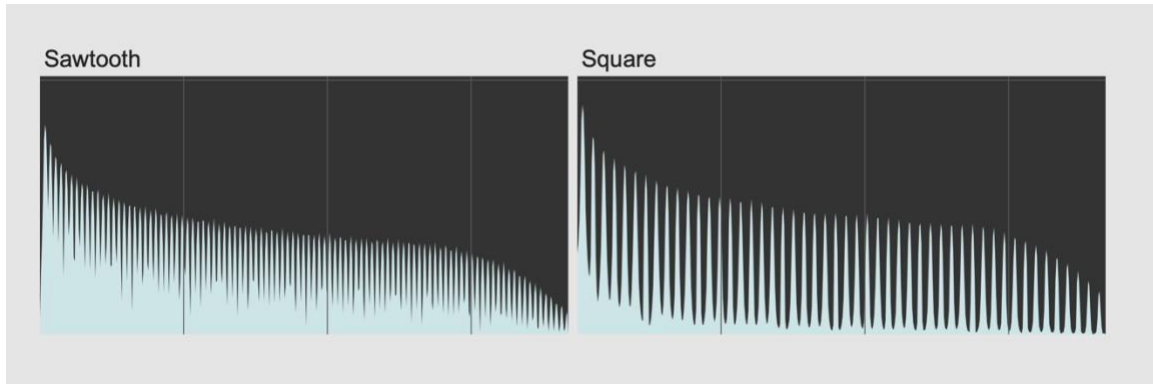


Figure 5.1: Spectrogram: Sawtooth VS Square Wave in 220Hz

Table 5-12 is the representation of the confusion matrix of Study 1: four questions in total. This provided a comparison between Predicted Values (x-axis) and Actual Values (y-axis). For the question in the A row, the TP (True-Positive) value was 26, the FN (False-Negative) value was 4 (0+0+4), the FP (False-Positive) value was 12 (7+4+1), and the TN (True-Negative) value was 79 (13+4+6+10+14+2+1+13+16). For the question in the B row, the TP (True-Positive) value was 13, the FN (False-Negative) value was 17 (7+4+6), the FP (False-Positive) value was 11 (0+10+1), and the TN (True-Negative) value was 79 (26+0+4+4+14+2+1+12+16). For the question in the C row, the TP (True-Positive) value was 14, the FN (False-Negative) value was 16 (4+10+2), the FP (False-Positive) value was 16 (0+4+12), and the TN (True-Negative) value was 74 (26+0+4+7+13+6+1+1+16). For the question in the D row, the TP (True-Positive) value was 16, the FN (False-Negative) value was 14 (1+1+12), the FP (False-Positive) value was 12 (4+6+2), and the TN (True-Negative) value was 78 (26+0+0+7+13+4+4+10+14).

For the performance measures for Study 1, accuracy, precision, and recall were shown (See Table 5-13). Accuracy means how many correct predictions were made among total values. Precision shows the percentage of positive predictions to the total amount of true positives. Recall is the correct answer rate per question and an indicator of how well the true positive rate was predicted. For those three values, the higher the percentage, the better the system. In Study 1, the sound of the circle (answer A) had the highest percentage of accuracy, precision, and recall.

Table 5-12: Confusion Matrix of Study 1

		Predicted Values			
		A	B	C	D
Actual Values	A	26	0	0	4
	B	7	13	4	6
	C	4	10	14	2
	D	1	1	12	16

Table 5-13: Accuracy, Precision, and Recall for each question

Answers	Accuracy	Precision	Recall
A	0.87	0.68	0.87
B	0.77	0.54	0.43
C	0.73	0.47	0.47
D	0.78	0.57	0.53

NOTE. Accuracy= $(TP+TN)/(TP+TN+FP+FN)$

Precision = $TP/(TP+FP)$

Recall = $TP/(TP+FN)$

5.3 Results (Study 2 and Study 3)

5.3.1 Analysis and Findings

The answer to the first question of Study 2, which was the sound corresponding to the circle, was Sound B. 22 participants selected the correct answer. The correct answer rate was 73.33%. See Table 5-14 for more information. For the Likert question following every first question, the participants answered how well each sound (from Sound A to Sound C) fit with the image. 63.33% of the participants chose “very likely” for Sound B. This showed quite a big difference compared to the second ranked Sound A. For more information, see Table 5-15.

Table 5-14: Answer to Question 1 of Study 2

Task 2.1

Sound	Freq	Perc(%)
Sound A	5	16.67
Sound B	22	73.33
Sound C	3	10.00

Table 5-15: Likert Question 1 of Study 2

Task 2.1: Sound A

Degree	Freq	Perc(%)
Very unlikely	6	20.00
Unlikely	7	23.33
Neutral	7	23.33
Likely	6	20.00
Very likely	4	13.33

Task 2.1: Sound B

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	1	3.33
Neutral	3	10.00
Likely	7	23.33
Very likely	19	63.33

Task 2.1: Sound C

Degree	Freq	Perc(%)
Very unlikely	13	43.33
Unlikely	5	16.67
Neutral	6	20.00
Likely	4	13.33
Very likely	2	6.67

The answer to the second question of Study 2, which was the sound corresponding to the triangle, was Sound A. 23 participants selected the correct answer. The correct answer rate was 76.67% (see Table 5-16). For the Likert question following every first question, the answer that Sound A was “very likely” or “likely” was the most at 50.00% and 36.67%, respectively. This showed quite a big difference compared to the second ranked Sound C. For more information, see Table 5-17.

Table 5-16: Answer to Question 2 of Study 2

Task 2.2

Sound	Freq	Perc(%)
Sound A	23	76.67
Sound B	2	6.67
Sound C	5	16.67

Table 5-17: Likert Question 2 of Study 2

Task 2.2: Sound A			Task 2.2: Sound B			Task 2.2: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	1	3.33	Very unlikely	12	40.00	Very unlikely	2	6.67
Unlikely	1	3.33	Unlikely	8	26.67	Unlikely	9	30.00
Neutral	2	6.67	Neutral	7	23.33	Neutral	7	23.33
Likely	11	36.67	Likely	2	6.67	Likely	9	30.00
Very likely	15	50.00	Very likely	1	3.33	Very likely	3	10.00

The answer to the third question of Study 2, which was the sound corresponding to the square, was Sound C. 21 participants selected the correct answer. The correct answer rate was 70.00% (see Table 5-18). For the Likert question following every first question, the answer that Sound C was “very likely” or “likely” was the most at 50.00% and 26.67%, respectively. This showed quite a big difference compared to the second ranked Sound B. For more information, see Table 5-19.

Table 5-18: Answer to Question 3 of Study 2

Task 2.3

Sound	Freq	Perc(%)
Sound A	2	6.67
Sound B	7	23.33
Sound C	21	70.00

Table 5-19: Likert Question 3 of Study 2

Task 2.3: Sound A			Task 2.3: Sound B			Task 2.3: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	6	20.00	Very unlikely	3	10.00	Very unlikely	4	13.33
Unlikely	6	20.00	Unlikely	10	33.33	Unlikely	1	3.33
Neutral	11	36.67	Neutral	5	16.67	Neutral	2	6.67
Likely	6	20.00	Likely	9	30.00	Likely	8	26.67
Very likely	1	3.33	Very likely	3	10.00	Very likely	15	50.00

Table 5-20 below is the representation of the confusion matrix of Study 2: three questions in total. For the question in the A row, the TP (True-Positive) value was 23, the FN (False-Negative) value was 7 (2+5), the FP (False-Positive) value was 7 (5+2), and the TN (True-Negative) value was 53 (22+3+7+21). For the question in the B row, the TP (True-Positive) value was 22, the FN (False-Negative) value was 8 (5+3), the FP (False-Positive) value was 9 (2+7), and the TN (True-Negative) value was 51 (23+5+2+21). For the question in the C row, the TP (True-Positive) value was 21, the FN (False-Negative) value was 9 (2+7), the FP (False-Positive) value was 8 (5+3), and the TN (True-Negative) value was 52 (23+2+5+22).

For the performance measures for Study 2 (see Table 5-21), the sound of the triangle (answer A) had the highest percentage of accuracy, precision, and recall, however, all values recorded similarly high percentages. This implies that Study 2 was well designed and more stable than Study 1.

Table 5-20: Confusion Matrix of Study 2

		Predicted Values		
		A	B	C
Actual Values	A	23	2	5
	B	5	22	3
	C	2	7	21

Table 5-21: Accuracy, Precision, and Recall for each question

Answers	Accuracy	Precision	Recall
A	0.84	0.77	0.77
B	0.81	0.71	0.73
C	0.81	0.72	0.70

NOTE. Accuracy= $(TP+TN)/(TP+TN+FP+FN)$

$$Precision = TP/(TP+FP)$$

$$Recall = TP/(TP+FN)$$

The answer to the first question of Study 3, which was the sound corresponding to the sine curve, was Sound A. 28 participants selected the correct answer. The correct answer rate was 93.33%. See Table 5-22 for more information. For the Likert question of Study 3 following every first question, the answer that Sound A was “very likely” or “likely” was the most at 70.00% and 30%, respectively. Only one participant chose Sound D as “very likely” and no one chose Sound B and Sound C. For more information, see Table 5-23.

Table 5-22: Answer to Question 1 of Study 3

Task 3.1

Sound	Freq	Perc(%)
Sound A	28	93.33
Sound B	0	0.00
Sound C	1	3.33
Sound D	1	3.33

Table 5-23: Likert Question 1 of Study 3

Task 3.1: Sound A

Degree	Freq	Perc(%)
Very unlikely	0	0
Unlikely	0	0
Neutral	0	0
Likely	9	30
Very likely	21	70

Task 3.1: Sound B

Degree	Freq	Perc(%)
Very unlikely	20	66.67
Unlikely	7	23.33
Neutral	3	10.00
Likely	0	0.00
Very likely	0	0.00

Task 3.1: Sound C

Degree	Freq	Perc(%)
Very unlikely	6	20.00
Unlikely	11	36.67
Neutral	6	20.00
Likely	7	23.33
Very likely	0	0.00

Task 3.1: Sound D

Degree	Freq	Perc(%)
Very unlikely	5	16.67
Unlikely	8	26.67
Neutral	9	30.00
Likely	7	23.33
Very likely	1	3.33

The answer to the second question of Study 3, which was the sound corresponding to the sawtooth curve, was Sound C. 26 participants selected the correct answer. The correct answer rate was 73.33% (see Table 5-24). For the Likert question following every first question, the answer that Sound C was “very likely” or “likely” was the most at 46.67% and 36.67%, respectively. This showed quite a big difference compared to the second ranked Sound D. See Table 5-25.

Table 5-24: Answer to Question 2 of Study 3

Task 3.2

Sound	Freq	Perc(%)
Sound A	3	10.00
Sound B	0	0.00
Sound C	22	73.33
Sound D	5	16.67

Table 5-25: Likert Question 2 of Study 3

Task 3.2: Sound A

Degree	Freq	Perc(%)
Very unlikely	10	33.33
Unlikely	5	16.67
Neutral	10	33.33
Likely	2	6.67
Very likely	3	10.00

Task 3.2: Sound B

Degree	Freq	Perc(%)
Very unlikely	15	50.00
Unlikely	8	26.67
Neutral	4	13.33
Likely	2	6.67
Very likely	1	3.33

Task 3.2: Sound C

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	3	10.00
Neutral	2	6.67
Likely	11	36.67
Very likely	14	46.67

Task 3.2: Sound D

Degree	Freq	Perc(%)
Very unlikely	2	6.67
Unlikely	4	13.33
Neutral	6	20.00
Likely	15	50.00
Very likely	3	10.00

The answer to the third question of Study 3, which was the sound corresponding to the pulse curve, was Sound D. 16 participants selected the correct answer. The correct answer rate was 53.33% (see Table 5-26). For the Likert question following every first question, the answer that Sound A was “very likely” or “likely” was the most at 36.67% and 40%, respectively. 33.33% of the participants also chose “very likely” for Sound C, which was the sound of the sawtooth line. For more information, see Table 5-27.

Table 5-26: Answer to Question 3 of Study 3

Task 3.3

Sound	Freq	Perc(%)
Sound A	2	6.67
Sound B	2	6.67
Sound C	10	33.33
Sound D	16	53.33

Table 5-27: Likert Question 3 of Study 3

Task 3.3: Sound A

Degree	Freq	Perc(%)
Very unlikely	10	33.33
Unlikely	10	33.33
Neutral	6	20.00
Likely	4	13.33
Very likely	0	0.00

Task 3.3: Sound B

Degree	Freq	Perc(%)
Very unlikely	8	26.67
Unlikely	8	26.67
Neutral	7	23.33
Likely	5	16.67
Very likely	2	6.67

Task 3.3: Sound C

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	7	23.33
Neutral	6	20.00
Likely	7	23.33
Very likely	10	33.33

Task 3.3: Sound D

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	6	20.00
Neutral	1	3.33
Likely	12	40.00
Very likely	11	36.67

The answer to the fourth question of Study 3, which was the sound corresponding to the dashed line, was Sound B. 30 participants selected the correct answer. The correct answer rate was 100.00% (see Table 5-28). For the Likert question following every first question, 96.67% of the participants chose “very likely” for Sound B. For more information, see Table 5-29.

Table 5-28: Answer to Question 4 of Study 3

Task 3.4

Sound	Freq	Perc(%)
Sound A	0	0
Sound B	30	100
Sound C	0	0
Sound D	0	0

Table 5-29: Likert Question 4 of Study 3

Task 3.4: Sound A

Degree	Freq	Perc(%)
Very unlikely	19	63.33
Unlikely	9	30.00
Neutral	2	6.67
Likely	0	0.00
Very likely	0	0.00

Task 3.4: Sound B

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	0	0.00
Neutral	0	0.00
Likely	1	3.33
Very likely	29	96.67

Task 3.4: Sound C

Degree	Freq	Perc(%)
Very unlikely	3	10.00
Unlikely	8	26.67
Neutral	8	26.67
Likely	11	36.67
Very likely	0	0.00

Task 3.4: Sound D

Degree	Freq	Perc(%)
Very unlikely	4	13.33
Unlikely	11	36.67
Neutral	8	26.67
Likely	7	23.33
Very likely	0	0.00

In the results of Study 2 and Study 3, most participants chose the correct answer. The accuracy rate was very high compared to Study 1. This showed that sound modulation describes the sounds of the shapes more accurately. However, the accuracy rate of the third question of Study 3 was somewhat less accurate than the others. The similarity of the sound pattern between the pulse line and the sawtooth line may cause confusion to the participants.

Table 5-30 below was the representation of the confusion matrix of Study 3: three questions in total (see Table?). For the question in the A row, the TP (True-Positive) value was 28, the FN (False-Negative) value was 2 (0+1+1), the FP (False-Positive) value was 2 (0+0+2), and the TN (True-Negative) value was 85

(30+0+2+0+22+10+0+5+16). For the question in the B row, the TP (True-Positive) value was 30, the FN (False-Negative) value was 0 (0+0+0), the FP (False-Positive) value was 2 (0+0+2), and the TN (True-Negative) value was 88 (28+1+1+3+22+5+2+10+16). For the question in the C row, the TP (True-Positive) value was 22, the FN (False-Negative) value was 8 (3+0+5), the FP (False-Positive) value was 11 (1+0+10), and the TN (True-Negative) value was 79 (28+0+1+0+30+0+2+2+16). For the question in the D row, the TP (True-Positive) value was 16, the FN (False-Negative) value was 14 (2+2+10), the FP (False-Positive) value was 6 (1+0+5), and the TN (True-Negative) value was 84 (28+0+1+0+30+0+3+0+22).

For the performance measures for Study 3 (Table 5-31), the sound of the dotted line (answer B) had the highest percentage of accuracy, precision, and recall. However, all values recorded similarly high percentages. Answer B had a 100% correct answer rate. This implies that Study 3 was well designed and as stable as Study 2.

Table 5-30: Confusion Matrix of Study 3

		Predicted Values			
		A	B	C	D
Actual Values	A	28	0	1	1
	B	0	30	0	0
	C	3	0	22	5
	D	2	2	10	16

Table 5-31: Accuracy, Precision, and Recall for each question

Question	Accuracy	Precision	Recall
1	0.97	0.93	0.93
2	0.98	0.94	1.00
3	0.84	0.67	0.73
4	0.83	0.73	0.53

NOTE. Accuracy= $(TP+TN)/(TP+TN+FP+FN)$, Precision = $TP/(TP+FP)$,

$$Recall = TP/(TP+FN)$$

5.4 Results (Study 4)

5.4.1 Analysis and Findings

The answer to the first question of Study 4, which was the sound corresponding to the cloud shape, was Sound B. 30 participants selected the correct answer. The correct answer rate was 100.00%. See Table 5-32 for more information. For the Likert question following every first question, the participants answered how well each sound (from Sound A to Sound C) fit with the image. The answer that Sound B was “very likely” or “likely” was the most at 76.67% and 23.33%, respectively. For more information, see Table 5-33.

Table 5-32: Answer to Question 1 of Study 4

Task 4.1

Sound	Freq	Perc(%)
Sound A	0	0
Sound B	30	100
Sound C	0	0

Table 5-33: Likert Question 1 of Study 4

Task 4.1: Sound A			Task 4.1: Sound B			Task 4.1: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	18	60	Very unlikely	0	0.00	Very unlikely	7	23.33
Unlikely	9	30	Unlikely	0	0.00	Unlikely	8	26.67
Neutral	3	10	Neutral	0	0.00	Neutral	11	36.67
Likely	0	0	Likely	7	23.33	Likely	4	13.33
Very likely	0	0	Very likely	23	76.67	Very likely	0	0.00

The answer to the second question of Study 4, which was the sound corresponding to the spiky shape, was Sound A. 27 participants selected the correct answer. The correct answer rate was 90.00% (see Table 5-34). For the Likert question, the answer that Sound A was “very likely” or “likely” was the most at 63.33% and 33.33%, respectively. This

showed quite a big difference compared to the second ranked Sound C. See Table 5-35, **Error! Reference source not found.**, **Error! Reference source not found.**, and **Error! Reference source not found.** more information.

Table 5-34: Answer to Question 2 of Study 4

Task 4.2

Sound	Freq	Perc(%)
Sound A	27	90
Sound B	0	0
Sound C	3	10

Table 5-35: Likert Question 1 of Study 4

Task 4.2: Sound A

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	0	0.00
Neutral	1	3.33
Likely	10	33.33
Very likely	19	63.33

Task 4.2: Sound B

Degree	Freq	Perc(%)
Very unlikely	22	73.33
Unlikely	6	20.00
Neutral	2	6.67
Likely	0	0.00
Very likely	0	0.00

Task 4.2: Sound C

Degree	Freq	Perc(%)
Very unlikely	0	0.00
Unlikely	9	30.00
Neutral	8	26.67
Likely	10	33.33
Very likely	3	10.00

The answer to the third question of Study 4, which was the sound corresponding to the reversed stair shape, was Sound C. 25 participants selected the correct answer. The correct answer rate was 83.33% (see Table 5-36). For the Likert question following every first question, the answer that Sound A was “very likely” or “likely” was the most at 40.00% and 40.00%, respectively. This showed quite a big difference compared to the second ranked Sound A. For more information, see Table 5-37.

Table 5-36: Answer to Question 3 of Study 4

Task 4.3

Sound	Freq	Perc(%)
Sound A	5	16.67
Sound B	0	0.00
Sound C	25	83.33

Table 5-37: Likert Question 3 of Study 4

Task 4.3: Sound A			Task 4.3: Sound B			Task 4.3: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	3	10.00	Very unlikely	12	40.00	Very unlikely	1	3.33
Unlikely	6	20.00	Unlikely	12	40.00	Unlikely	2	6.67
Neutral	7	23.33	Neutral	5	16.67	Neutral	3	10.00
Likely	9	30.00	Likely	1	3.33	Likely	12	40.00
Very likely	5	16.67	Very likely	0	0.00	Very likely	12	40.00

The accuracy of Study 4 was very high compared to the others. This showed that sound modulation can be successfully applied to other customized shapes. The results of Study 1 show the limitations of sound without modulation. However, other studies showed that it was possible to sonically describe different types of lines and other free-form shapes by using sound modulation.

Table 5-38 below is the representation of the confusion matrix of Study 4: four questions in total. For the question in the A row, the TP (True-Positive) value was 27, the FN (False-Negative) value was 3 (0+3), the FP (False-Positive) value was 5 (0+5), and the TN (True-Negative) value was 55 (30+0+0+25). For the question in the B row, the TP (True-Positive) value was 30, the FN (False-Negative) value was 0 (0+0), the FP (False-Positive) value was 0 (0+0), and the TN (True-Negative) value was 60 (27+3+5+25). For the question in the C row, the TP (True-Positive) value was 25, the FN (False-Negative)

value was 5 (0+5), the FP (False-Positive) value was 3 (3+0), and the TN (True-Negative) value was 57 (27+0+0+30).

For the performance measures for Study 4 (see Table 5-39), the sound of the cloud shape (answer B) had the highest percentage of accuracy, precision, and recall. However, all values recorded similarly high percentages. Answer B had a 100% correct answer rate. This implies that Study 4 was well designed and as stable as Study 2 and Study 3.

Table 5-38: Confusion Matrix of Study 4

		Predicted Values		
		A	B	C
Actual Values	A	27	0	3
	B	0	30	0
	C	5	0	25

Table 5-39: Accuracy, Precision, and Recall for each question

Question	Accuracy	Precision	Recall
1	0.91	0.90	0.84
2	1.00	1.00	1.00
3	0.91	0.83	0.89

NOTE. Accuracy = $(TP+TN)/(TP+TN+FP+FN)$

Precision = $TP/(TP+FP)$

Recall = $TP/(TP+FN)$

5.5 Results (Study 5)

5.5.1 Analysis and Findings

The answer to the first question of Study 5, which was the sound corresponding to the triangle shape moving left to right, was Sound B (stereo panning). 15 participants selected the correct answer. The correct answer rate was 50.00% (see Table 5-40). For

the Likert question following every first question, the participants scaled how well each sound (from Sound A to Sound C) fit with the image. 33.33% of the participants chose “very likely” for Sound B. This showed quite a big difference compared to the second ranked Sound A and Sound C. For more information, see Table 5-41.

Table 5-40: Answer to Question 1 of Study 5

Task 5.1

Sound	Freq	Perc(%)
Sound A	10	33.33
Sound B	15	50.00
Sound C	5	16.67

Table 5-41: Likert Question 1 of Study 5

Task 5.1: Sound A			Task 5.1: Sound B			Task 5.1: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	4	13.33	Very unlikely	4	13.33	Very unlikely	2	6.67
Unlikely	5	16.67	Unlikely	7	23.33	Unlikely	10	33.33
Neutral	10	33.33	Neutral	5	16.67	Neutral	7	23.33
Likely	7	23.33	Likely	4	13.33	Likely	7	23.33
Very likely	4	13.33	Very likely	10	33.33	Very likely	4	13.33

The answer to the second question of Study 5, which is the sound corresponding to the triangle shape rotating, was Sound A (pitch shifting). 19 participants selected the correct answer. The correct answer rate was 63.33% (see Table 5-42). For the Likert question following every first question, the answer that Sound A was “very likely” or “likely” was the most at 30.00% and 43.33%, respectively. 26.67% of the participants chose “very likely” for Sound C, which is the sound with stereo panning. For more information, see Table 5-43.

Table 5-42: Answer to Question 2 of Study 5

Task 5.2

Sound	Freq	Perc(%)
Sound A	19	63.33
Sound B	1	3.33
Sound C	10	33.33

Table 5-43: Likert Question 2 of Study 5

Task 5.2: Sound A			Task 5.2: Sound B			Task 5.2: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	0	0.00	Very unlikely	6	20.00	Very unlikely	2	6.67
Unlikely	3	10.00	Unlikely	8	26.67	Unlikely	5	16.67
Neutral	5	16.67	Neutral	13	43.33	Neutral	8	26.67
Likely	13	43.33	Likely	3	10.00	Likely	7	23.33
Very likely	9	30.00	Very likely	0	0.00	Very likely	8	26.67

The answer to the third question of Study 5, which was the sound corresponding to the triangle shape zooming in and out, is Sound B (volume). 21 participants selected the correct answer. The correct answer rate was 70.00% (see Table 5-44). For the Likert question, the answer that Sound B was “very likely” or “likely” was the most at 53.337% and 30.00%, respectively. This showed quite a big difference compared to the second ranked Sound A. For more information, see Table 5-45.

Table 5-44: Answer to Question 3 of Study 5

Task 5.3

Sound	Freq	Perc(%)
Sound A	9	30
Sound B	21	70
Sound C	0	0

Table 5-45: Likert Question 3 of Study 5

Task 5.3: Sound A			Task 5.3: Sound B			Task 5.3: Sound C		
Degree	Freq	Perc(%)	Degree	Freq	Perc(%)	Degree	Freq	Perc(%)
Very unlikely	2	6.67	Very unlikely	2	6.67	Very unlikely	12	40.00
Unlikely	8	26.67	Unlikely	3	10.00	Unlikely	10	33.33
Neutral	5	16.67	Neutral	0	0.00	Neutral	3	10.00
Likely	8	26.67	Likely	9	30.00	Likely	5	16.67
Very likely	7	23.33	Very likely	16	53.33	Very likely	0	0.00

It is worth noting that Study 5 did not have absolute answers. When the shapes were rotating to the right, some thought that it was natural for the musical pitch to rise, and some thought that it was natural for the sound position to move to the right. The interpretation of the sound here was open and the choices were not unique for each question. For example, Sound A in the first question was about amplitude changes but Sound A in the second question changed pitch. Considering those aspects, the confusion matrix was excluded for Study 5. But interestingly, the correct answer scored the highest probability.

5.6 Results (Probability)

We have looked at the test results from Study 1 to Study 5, so far. The questions about age and educational background were given in the pre-questionnaire test, so I wanted to examine if these additional variables affected the test result. Table 5-46 shows the significance level based on age, educational background, and combination of the two. I set significance level α to 0.05 in this test ($p < .05$). JMP Pro 16 [392] was used for this statistical analysis with the logistic regression model. Entries were ChiSq (p-value). As shown in Table 5-46, it can be seen that most studies were not affected by the categorical variables (i.e., age, educational background). Only age, visual art background, and both visual art and music background were significant for Study 2. These results showed that this sonification can be used as a universal auditory symbol. Please note that since there were not enough data to estimate an interaction effect between these categorical variables, combined variables like A+M, A+V, M+V were not shown.

Table 5-46: Summary of significance level tests for all studies

	Study 1	Study 2	Study 3	Study 4	Study 5
A	5.7809778 (0.2161)	7.9443903 (0.0472)	1.7655926 (0.7788)	0.2330511 (0.8900)	0.5528657 (0.9071)
M	5.4774517 (0.2417)	5.4974039 (0.2400)	4.5181965 (0.3404)	0.5503918 (0.7594)	1.4143782 (0.9227)
V	4.039838 (0.4006)	6.1248589 (0.0468)	3.1602069 (0.3676)	1.4021033 (0.4961)	1.0023625 (0.9094)

NOTE. A = Age; M = Music; V = Visual Arts

5.7 Results (Post-Questionnaire)

The post-questions were a mix of Likert and open-ended questions for subjective feedback. The questions were as follows:

1. I could easily match the sound with the visual.
2. The study was fun and interesting.
3. The sounds enhanced the visual information and provided multi-sensory experience (i.e., you experienced visually and auditory).
4. What do you like the most about the study?
5. Do you think this sonification (the use of sound to convey non-audio information) method has the potential in communication design, art, or education? If so, which areas would you see it being most useful (Select all that apply)?
6. Why do you think so?
7. What task was the most challenging to you and why? (You can skip this question if the study questions were easy to answer).
8. Please share any other thoughts if you have them. This will help me improve this sonification system for future use.

To question 1, when asked whether I could easily match the sound with the visual, 22 (73.33%) participants chose “Somewhat agree” and 4 (13.33%) participants chose “Strongly agree”. To question 2, when asked whether the study was fun and interesting, 27 (90.00%) participants chose “Strongly agree”. To question 3, when asked whether the sounds enhanced the visual information and provided multi-sensory experience, 20

(66.67%) participants chose “Strongly agree” and 7 (23.33%) participants chose “Somewhat agree”. See Table 5-47 for more information.

Table 5-47: Answer to Post-Questionnaire 1

Post-Q1

Sound	Freq	Perc(%)
Strongly agree	4	13.33
Somewhat agree	22	73.33
Neither agree nor disagree	2	6.67
Somewhat disagree	2	6.67
Strongly disagree	0	0.00

Post-Q2

Sound	Freq	Perc(%)
Strongly agree	27	90.00
Somewhat agree	1	3.33
Neither agree nor disagree	1	3.33
Somewhat disagree	0	0.00
Strongly disagree	1	3.33

Post-Q3

Sound	Freq	Perc(%)
Strongly agree	20	66.67
Somewhat agree	7	23.33
Neither agree nor disagree	2	6.67
Somewhat disagree	1	3.33
Strongly disagree	0	0.00

When asked about what you like the most about the study, one of the most common feedback from the participants was about the concept behind this study. Many participants mentioned that “the sound with animation was fun”, “associating sound with shape and animation was fun and enjoyable”, “this is an interesting way to think about sound and shape”, “the whole experience was immersive.” Some said, “this study was new”, “the concept was innovative”, “it effectively conveyed the sense of shape”, and “the sound was designed very thoughtfully.” One participant said, “it is like flexing a sort of creative muscle that one does not use often.”

When asked about whether this sonification method has the potential in communication design, art, or education, if so, which areas would you see it being most useful, the participants said, “everyday life”, “children’s education”, “limitless applications”, “language making”, “captioning for music”, “interface/interaction design”, “human

computer interaction”, and “marketing”. The reasons for those answers include “it helps visual recognition”, “it could enhance a new experience to users for website and game interfaces”, “it will be useful in art and music”, “it will help data sonification”, “useful to create keys or symbols for people who may be blind or have vision problems”, “it was synesthetic to some degree, so it can help bridge communication gap”, “it has a lot of potential for people who can hear and see”, and “the synchronization of audio and visual content is much better than any one of them individually.” One said, “the animation part seems to have many potentials to build better UX experience because there are a lot of motion effects, such as falling, bumping(collision), and fade in-out.” However, one participant pointed out that “sonification is not good for conveying a lot of data, it may be better at a simple piece of information.”

For the question about what the most challenging task was, the most common feedback was that Study 5 (the moving triangle) was the most challenging because of the movement and two participants said they felt like the sound and the movement of the shape were not exactly synced up. One said none of the sounds fit any moving images well. Another one said, “sometimes I didn’t think any sound really fit the shape.” One said that dotted line vs square wave line was difficult because they sounded similar. Another one said, “sawtooth line and square wave line were somewhat interchangeable and ambiguous perception for me.” Interestingly, three participants said, “none of the tasks was more challenging than the others”, “none”, “none of them were difficult.”

The participants also added that “I have never heard of a concept like this. I think its innovative and great!” and “it is very unique and interesting.” Many participants mentioned that they want to see the relationship between color and sound, more dynamic shapes with the sound, and another version with more complex sounds/images in the near future. Some were curious about what the study would be like to use natural sounds like chimes, bells, and horns. One said that this study helped create images in their mind. And another said it would be better if some basic knowledge about sonification was shared prior to the study. However, as previously mentioned, the most important condition of this study is that no information is given.

6 Artistic Audiovisual Expression

6.1 Introduction

In Chapter 4 and Chapter 5, the sonification studies to support three research hypotheses were explained, and the research results about how correctly the test participants matched the sounds with the shapes were analyzed. However, the original idea of this research was to see whether image-sound sonification could be used as an audiovisual platform for musical expression. This artistic sonification approach would not necessarily require a user study and it would be acceptable if it relies entirely on the artist's personal interpretation for image-sound mappings. The main reason that the user tests were conducted prior to the introduction of the artistic sonification was to show that the foundation of this work was designed to confirm objectivity what makes listeners associate the sounds with the shapes.

In this chapter, how the same sonification method from the user studies was transformed into artistic audiovisual platforms will be discussed. The key difference is that this new version adds color variables such as a single color and multicolors (i.e., gradient colors). The background of the image and the shape objects use different sonification methods and different sound synthesis techniques. The number of shapes/objects to be displayed is from zero to five and those graphic objects can harmonize with each other. Compared to the user study version, this version introduces more controllable variables (i.e., size, angle, position, scanning speed, etc.), so there is a possibility of using those variables for future user studies. Future research is mentioned in Chapter 7.

6.2 Implementation

The artistic audiovisual platform has a structure similar to the user study version in many respects. However, this platform has some differences as it incorporates visualizing components (to show the sonification process in real-time) and a different number of variables for sonification methodologies (i.e., size, angle, position, scanning speed, etc.). The audiovisual platform is divided into four components: the image creation component, the image scanning component, the sonification component, and the sonification-

visualizing component. The image creation components (i.e., image display, color change, shape change, image movement) were built in TouchDesigner. TouchDesigner also handled two image scanning components that read the image background and the border of the objects. The sonification-visualizing component visualizes the line scan path for the background. For the objects, scanning dots indicate where the scan path is located. For more information, see Chapter 6.3.

The x-y coordinate of the objects' scanning dot and background color values are sent to Max in two different ways. The x-y coordinates of the objects are transmitted to Max via Open Soun Control (OSC). On the other hand, background color values are directly sent to Max's jit.matrix through the video frame sharing software Syphon and then the color values are extracted in Max. For example, if the image resolution of the background is 1000x1000, the quantity of color information that needs to be sent will be $1000 \times 1000 \times 3$ (3 means three channels for HSL/RGB excluding alpha values) = 3,000,000. Therefore, sending images to Max via Syphon is a more efficient method because it uses fewer computer resources than sending integer or float numbers via OSC. Lastly, the sonification component was built in Max (Figure 6.1).

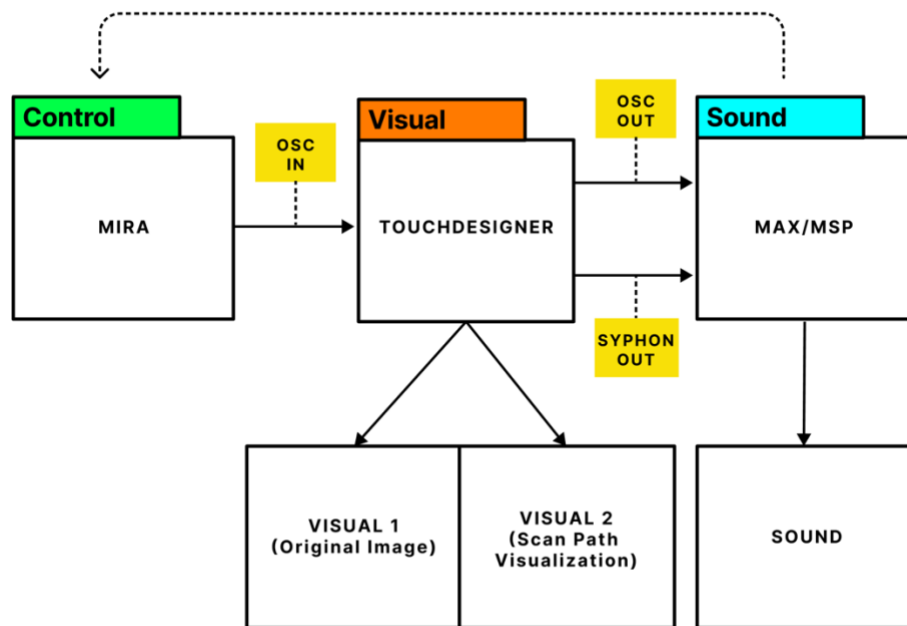


Figure 6.1: Implementation for the User Study

6.3 Sonification Methodology

6.3.1 Overview

In this chapter, artistic audiovisual works will be introduced. In this version, colors were newly added, and each shape automatically moves on the canvas and the colors change as well. In addition, this audiovisual version allows interactivity. Through MIRA [393] for iPad, the background color mode, background color change speed, shape color mode, shape color change speed, shape movement speed, shape size, shape angle, and scanning point/path speed can be controlled. Not only that, but the two sonification methods are integrated in this version. The first is the line-by-line method and the second is the object-oriented method (see Chapter 6.3.3). The line-by-line method scans the image horizontally or vertically, so this method is useful for sensing overall color changes rather than depicting shapes [394]. On the other hand, the object-oriented sonification successfully delivers the shape information sonically, as shown in the user studies in Chapter 5. Based on my previous research using both ways, the line-by-line method was

determined to sonify the background of the visual information and the object-oriented method was determined to sonify the shapes displayed over the background.

6.3.2 Background Sonification

The line-by-line sonification method was developed for the background sonification. The line-by-line sonification is one of the most common approaches found in image-to-sound sonification, as shown in the examples in Chapter 2. Apparently, this approach of reading the entire image area from left to right or top to bottom seems to be a very intuitive and natural method for transferring images into time-domain forms. The line-by-line method was the first sonification method I tried. I wrote papers and created several works based on this method during my doctoral research [394][395][396][397] (Figure 6.2). In my experience, this method was impossible for describing shapes because the line-by-line method only uses pixel information from a single row or a single column. Instead, it can be used as a method to observe the overall color change of the entire image, and it was also useful for artistic expression. Considering those aspects, I concluded that the line-by-line method would best fit for background sonification for my audiovisual version because the background has no shapes but only colors.



Figure 6.2: Colormatics

Visual presentation of the background sonification consisted of two screens. The first screen showed the original background image and the second screen showed the sonification process. The second screen displayed where the line path of the sonification was positioned. All pixels in the scan line path were stretched out and fit into the entire canvas. That is, this additional visualization component allowed the audience to naturally synchronize the sound and visual (see Figure 6.3). The scan line path of the line-by-line sonification moves over the image from left to right or from top to bottom, or can be fixed in the center of the image horizontally or vertically (see Figure 6.4).

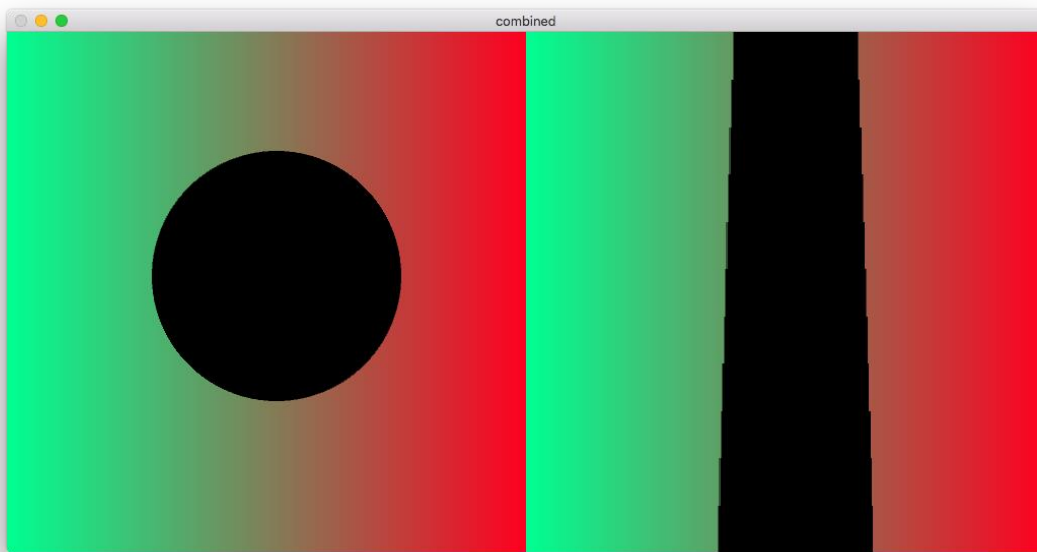


Figure 6.3: Original Image (left) and Sonification Visualization Components (right)

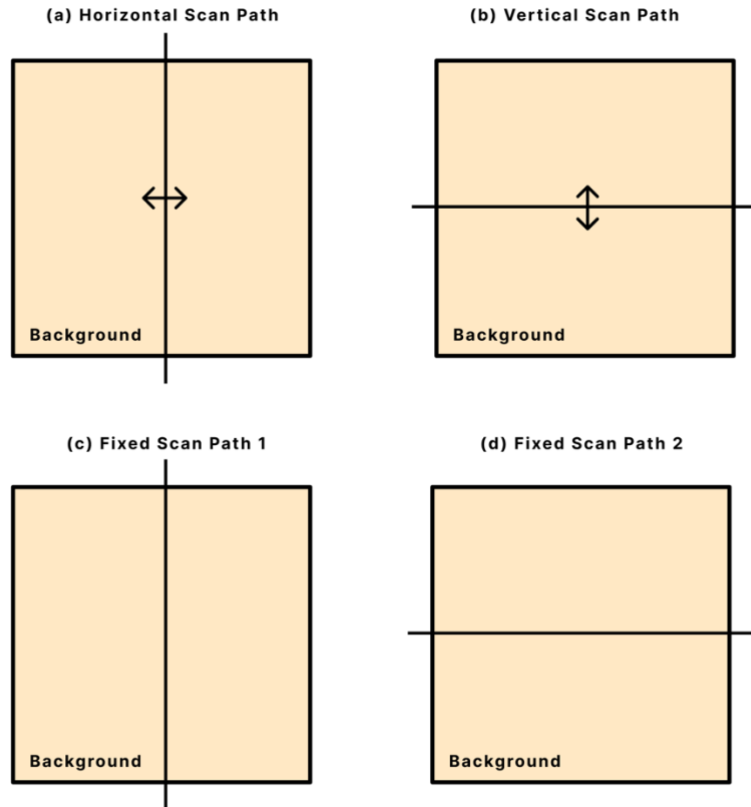


Figure 6.4: Scan line path for the background

The sound synthesis techniques for the background sonification (see Figure 6.6) are driven by additive synthesis and waveshaping synthesis. Either one or a mix of both can be used. Additive synthesis and waveshaping synthesis are two of the most intuitive methods in terms of creating sound from all the pixels located on the scan path at once. The additive synth part used the HSL value of each pixel, where hue was mapped to pitch, saturation was for detuning, and lightness was for amplitude level (see Figure 6.5 and Table 6-1). The minimum and maximum frequency range of the pitch can be set by users. That is, if the image width or height is 1000 pixels, 1000 sine waves are added to create the sound. Naturally, this method produces a rich timbre, especially when the colors are rich, such as in gradient colors. On the other hand, the waveshaping synth only uses hue values. Wave tables are drawn by hue values (see Figure 6.5). The hue values ranging from 0 to 360 are scaled between -1.0 and 1.0 to draw the waveshape in the audio buffer. That is, if the image width or height is 1000 pixels, 1000 data points can be

transferred into the audio buffer. The audio buffer will contain 1000 samples and the waveshape will be drawn according to the hue value. The fundamental oscillator frequency follows the same color-pitch mapping used in the additive synth. If the background has a single color, the sound becomes inaudible because the waveshape will only contain a single straight line.

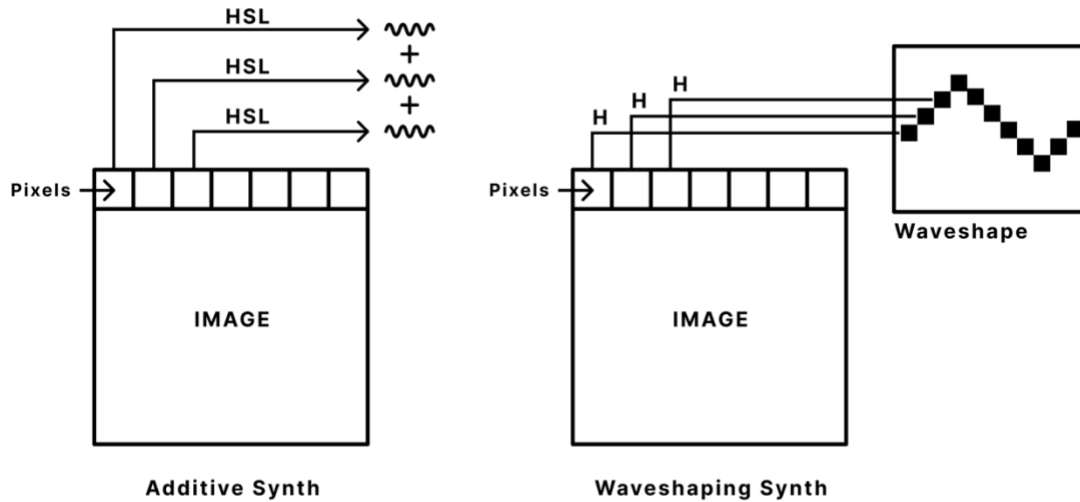


Figure 6.5: Additive and waveshaping

Table 6-1: HSL Mapping for the background

Hue: 0-360	Saturation: 0.0-1.0	Brightness: 0.0-0.5
Frequency: 165-440Hz	Detune: -100-0 cents	Amplitude: 0.0-0.5

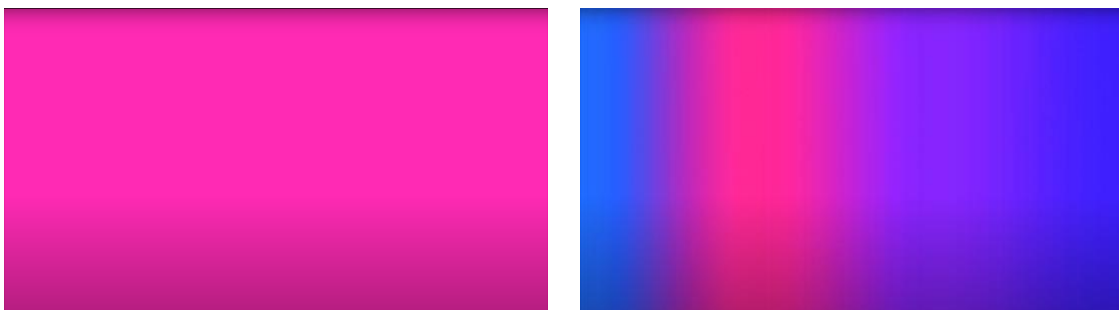




Figure 6.6: Background Color Examples

6.3.3 Shape Sonification

An object-oriented sonification method was developed for shape sonification, based on the method used by the user study (see Chapter 4). The key difference is that this audiovisual version has colors. Like the background sonification, there is no absolute mapping between color and pitch, instead the mapping can be flexible depending on users' choices. The minimum and maximum frequency can be set up by users. The audiovisual version allows five shape sonification channels and one background sonification channel at once.

The shapes are also based on the same basic shapes used in the user study, such as circles, triangles, squares, and lines. The lines include a straight line, dotted line, sine wave line, sawtooth wave line, triangle wave line, and pulse wave line. In addition, new custom shapes are added. See Figure 6.7 for what shapes were used.

The audiovisual platform supports six channels including the background. Different types of shapes can be displayed for each channel. Compared to traditional synthesizers, it is equivalent to having 6 voices. Three basic shapes with fill colors are assigned to channel 1, three basic shapes with a stroke without fill colors are assigned to channel 2, four curves are assigned to channel 3, six customized shapes are assigned to channel 4, a straight line is assigned to channel 5, and the background is assigned to channel 6. For each channel, the shapes can be switched through the MIRA interface. See Figure 6.7 for the shapes supported by each channel. To minimize overlapping frequency regions for each channel, the color-pitch mapping was set differently. The frequency range of one

channel is mostly one octave. See Table 6-3 for more information. For Channel 4, the round shapes are based on the circle's timbre, the pointed shapes are based on the triangle's timbre, and the rhombus shape is based on the square's timbre. In most cases, the filter is controlled in the same way as the user study version, but in some cases, it has been slightly improved. See Table 6-2 for more detailed information. According to the brightness, new voices are added. In the darkest setting, only 30% of the original color is shown and only one voice is played. For example, a single voice for the triangle is a combination of two sawtooth waves. When the brightness reaches 100%, nine additional voices are added. Channel 1 and channel 4 are based on a minor pentatonic scale and channel 2, channel 3, and channel 5 are based on a major pentatonic scale.

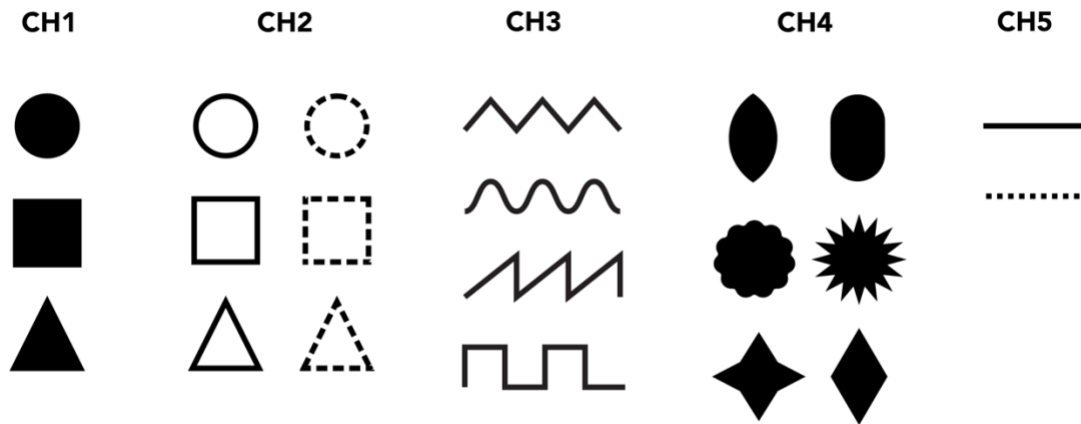


Figure 6.7: Shapes and their corresponding channels (except CH6)

Table 6-2: Filter-related values for custom shapes in Channel 4

Shapes	1 (Top-Left)	2 (Top-Right)	3 (Middle-Left)	4 (Middle-Right)	5 (Bottom-Left)	6 (Bottom-Right)
OSC	Circle's	Circle's	Circle's	Triangle's	Triangle's	Square's
Filter	Circle's	Circle's	Circle's	Triangle's	Square's	Square's

Additional Changes	NA	NA	1 Low Pass Filter Added	3 Peak Notch Filters Added	N/A	N/A
Filter Gain	Circle's	Circle's	2	Triangle's	Square's	Square's
Cutoff Frequency Range (Hz)	Circle's	Circle's	-30.0 – 30.0	-10.0 – 10.0	Square's	Square's
Q-Value	Circle's	Circle's	1.5	Triangle's	0	0
Q-Value Sweep	Circle's	Circle's	1.3	1.5	Square's	Square's

Table 6-3: Pitch setting for the four channels.

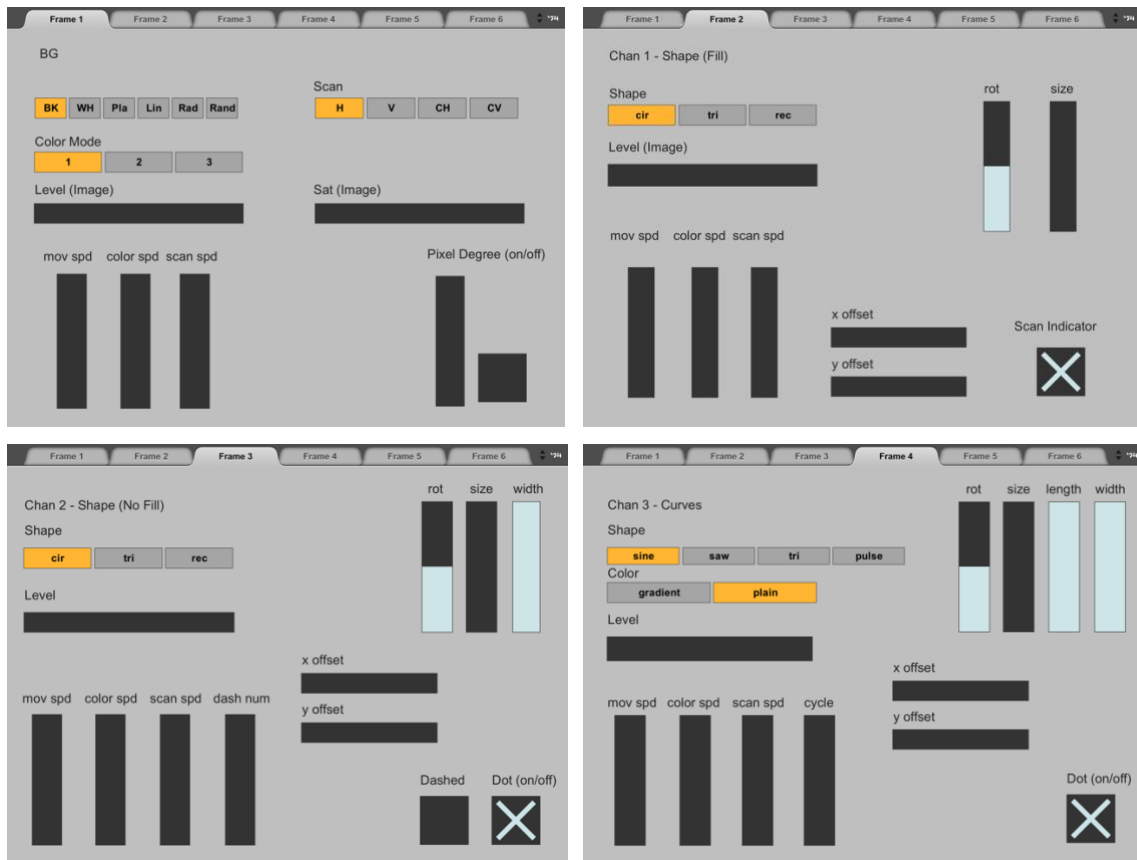
CH1	CH2	CH3	CH4	CH5	CH6 (BG)
87-175Hz	263-523Hz	659-1329Hz	220-440Hz	362-784Hz	165-440Hz

6.3.4 Control Interface

Changes in color and movement of shapes has some degree of autonomy. However, the touch-based control interface is provided so that users can interact with other controllable variables. To do so, the audiovisual version implements a MIRA control interface that can be quickly synchronized with Max/MSP. In the background sonification component, the color mode, scanning mode, scanning speed, color change speed, gradient color change speed, brightness, and saturation can be adjusted. The color mode supports black, white, other plain colors, linear gradient colors, and radial gradient colors. In the gradient color mode, two or three colors can be mixed, The scanning mode can be switched between H (horizontal), V (vertical), CH (center-horizontal), and CV (center-vertical)

(see Figure 6.4 and Figure 6.8). Also, the background image can be pixelated to degrade the overall audio quality. This can be seen as using a bitcrusher effect.

In the shape sonification component, brightness, the size of the shapes, the angle of the shapes, shape movement speed, color change speed, scanning speed, and x- and y-position of each shape can be controlled. Also, the shape types in each channel can be switched. Additionally, for Channels 2, 3, and 5, which contain the stroke-based shapes, the line width can be adjusted. The thicker the line, the more the center frequency of the filter is boosted. For Channels 2 and 5, users can switch between regular lines and dotted lines, and the number of the dots can be controlled. When the dotted lines are activated, the sawtooth wave-driven tremolo effect intervenes, and as the number of dots increases, the tremolo speed increases. Also, lines on Channel 3 and Channel 5 support plain color mode and gradient color mode. For the control interface, see Figure 6.8 for more information.



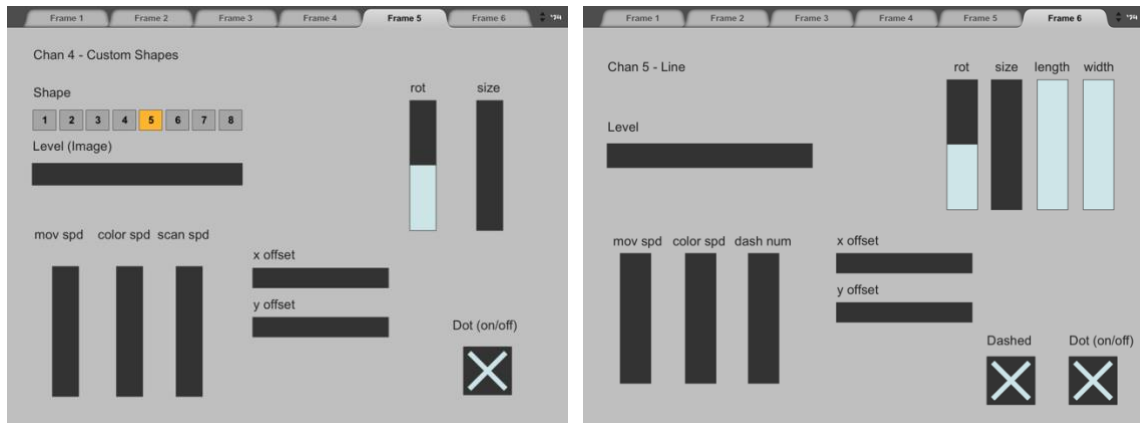


Figure 6.8: MIRA interface for the six channels

6.3.5 Outcome

For the final system, the background sonification and the object-oriented sonification components were integrated. The implementation and the sample images are as follows (See Figure 6.9 and Figure 6.10). To further enhance user experience, five shapes were spatialized using a Resonance Audio Monitor ambisonics plugin [398]. The audio of the background channel was exactly centered. This allowed audiences to feel the movement of each shape in a virtual 3D space when listening with headphones. Two versions were recorded for the project documentation. The first version was directly recorded on a MacBook Pro 2019 using a Resonance Audio Monitor ambisonics plugin so that it is compatible with stereo speakers or headphones. The second version (exhibition version) was displayed at Virginia Tech's Perform Studio (Figure 6.11) using the ICST ambisonics plugin [399]. The Perform studio has 24 speakers in two levels (12 speakers are positioned near eye-level and the rest are above them).

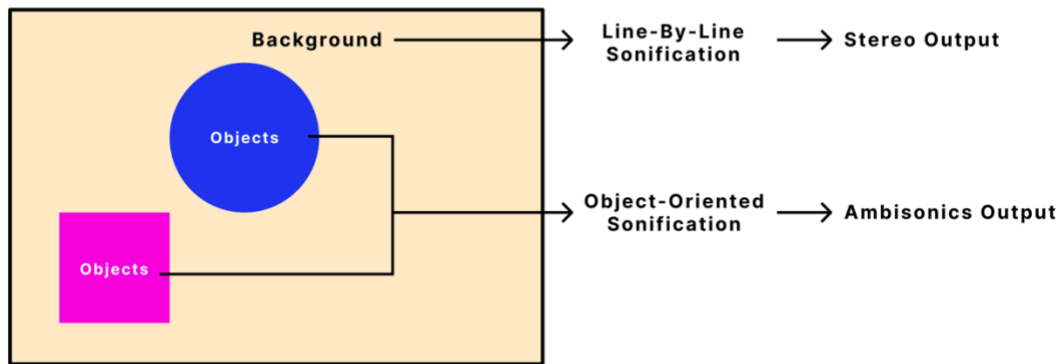


Figure 6.9: Two sonification methods for the audiovisual platform

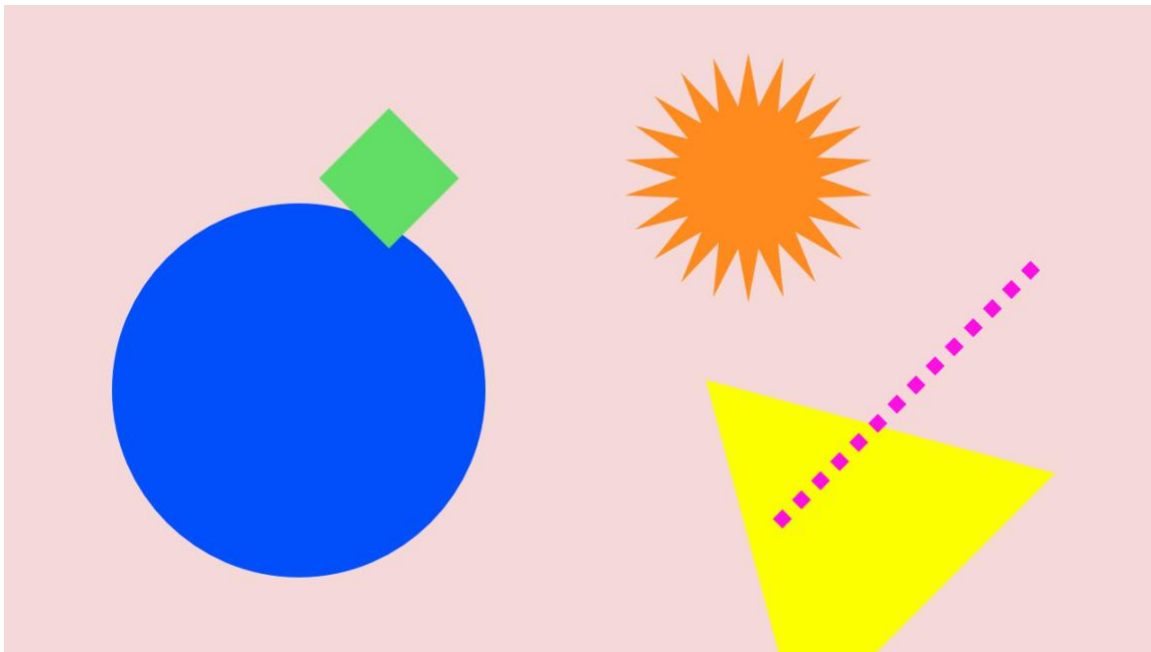


Figure 6.10: Screenshot of the audiovisual version



Figure 6.11: Virginia Tech's Perform Studio

7 Conclusion and Future Work

7.1 Conclusion

Chapter 4 and Chapter 5 showed that a proper combination of basic waveshapes (e.g., sine, sawtooth, and pulse wave) with the sound modulation based on the border of the shape can describe basic shapes (e.g., a circle, a triangle, a square, lines) as well as even more complex shapes such as the spiky shape, stair shape, and cloud shape. Further, the study result showed the possibility to expand for more customized shapes for future research.

In Chapter 6, the same sonification methodology, which was used in the user study, was migrated into the art platform in conjunction with the color-pitch mapping and two different sonification scanning models: object-oriented and line-by-line. Although color variables were not adopted in the user study because the user study mainly focused on timbres and sound modulation, using colors is also meaningful to show its potential as a cornerstone for artistic/musical expression of graphic-sound sonification that can be used in new media art, sound art, audiovisual art, electronic music, and musical instrument.

Considering the sound synthesis methods used in this dissertation and the type of the basic shapes used in the user study, the possibilities for its application to broader and more detailed research are limitless. This will be discussed in the next chapter.

7.2 Future Work

Future work can be divided into two main directions. The first is the utilization of image-auditory design in the same context as the user study described in Chapter 4. As mentioned by the participants in the post-questionnaire, this approach can be used for communication design for human-machine interface. The second direction is the use of the musical (audiovisual) instrument mentioned in Chapter 6, which means the adoption of the sonification methodology focusing on artistic (and musical) expression. In addition, this artistic approach always includes real-time visualization showing the sonification process, encompassing the characteristics of both sound art and visual art.

First, for the communication design approach, the user study version can be extended to more scientific and detailed research. For example, in the sonification methodology used in the user study, pitch modulation is applied when the scanning path is moving along the border of the shapes and the pitch modulation is linearly mapped. However, it is known that humans perceive changes differently in high pitches and low pitches [400].

Considering this factor, focusing more on psychoacoustics is a possible area of exploration. Stereo panning is also applied linearly as the shape moves in the user study. According to my other PhD research [401][402][403], sound recognition accuracy for a sound moving horizontally and vertically is obviously different. More advanced sound spatialization algorithms focusing on psychoacoustic perception can be applied for future work. In addition, it is possible to design more detailed user studies based on the previous studies. For example, for the circle sound in Study 1, two sine waveforms at 261.6Hz and 130.8Hz were mixed at a ratio of 60:40. The sounds for each shape were first determined by the research examiner and then the participants were asked to compare one with the other. For future research, it would be interesting to set those values differently and examine what ratios best evoke a circle.

Color variables were excluded from the user test, but the use of colors can be considered in future research. According to my literature review in Chapter 2, color-sound sonification studies necessarily require a long pre-test session to allow participants to memorize the relationship between color and sound. In fact, long pre-training sessions are a very counterproductive process. For research not to depend on memorization, better sonification strategies that reduce time for the pre-test need to be considered.

Alternatively, it will be necessary to carefully consider whether a color-related study can be designed without pre-tests by applying psychological color associations such as red-warm and blue-cold. Such sonification research would greatly contribute to the work of the sonification community, which is rapidly growing, and could be applied various multidisciplinary fields.


For artistic expression, I mainly used additive, waveshaping, and subtractive synthesis techniques in this dissertation. Applying other types of sound synthesis methods (e.g., granular, FM) could be considered. For example, hardware synthesizers from each manufacturer have different sound characteristics future audiovisual works could be approached with different sound synthesis methods for each work. In addition to focusing on the shape-timbre relationship shown in this dissertation, various sound effects such as reverb, echo, and phaser and their corresponding visualization effects could also be applied. Then, along with shape-timbre, shape-modulation, and color-pitch expressions, the current platform could be expanded for a richer and more sophisticated outcome.

The audiovisual platform introduced in Chapter 6 is mainly designed for musical expression. If this platform is compatible with hardware musical instruments (e.g., modular synthesizer, effects, MIDI instruments), my work could naturally have more potential as a musical instrument. This can be seen in a similar vein to the fact that hardware synthesizers, samplers, and drum machines are compatible with other instruments through MIDI or CV ports regardless of manufacturer. If CV in-out functions are applicable for my work, the audiovisual platform could be used to drive external hardware synthesizers and interfacing with modular synthesizers (which are even more

customizable) could be an option. Through this plan, the current sonification platforms could be expanded and become more flexible for electronic music performance.

Finally, the current platform, regardless of the direction of art or academic research, commonly operates based on 2D projection or computer screens, however, if other multimedia platforms such as VR, AR, and XR could be combined, these new interactive devices would further increase the user experience and could also lead to other types of studies in related fields.

8 APPENDIX Test Materials

	<p>Division of Scholarly Integrity and Research Compliance Institutional Review Board North End Center, Suite 4120 (MC 0497) 300 Turner Street NW Blacksburg, Virginia 24061 540/231-3732 irb@vt.edu http://www.research.vt.edu/sirc/hrpp</p>
---	---

MEMORANDUM

DATE: November 9, 2021

TO: Patrick Finley, Woohun Joo

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: See the sounds and listen to the shapes

IRB NUMBER: 21-097

Effective November 8, 2021, the Virginia Tech Human Research Protection Program (HRPP) determined that this protocol meets the criteria for exemption from IRB review under 45 CFR 46.104(d) category(ies) 2(ii),3(i)(B).

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit an amendment to the HRPP for a determination.

This exempt determination does not apply to any collaborating institution(s). The Virginia Tech HRPP and IRB cannot provide an exemption that overrides the jurisdiction of a local IRB or other institutional mechanism for determining exemptions.

All investigators (listed above) are required to comply with the researcher requirements outlined at: <https://secure.research.vt.edu/external/irb/responsibilities.htm>

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Determined As: **Exempt, under 45 CFR 46.104(d) category(ies) 2(ii),3(i)(B)**
Protocol Determination Date: **November 8, 2021**

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.

Figure 8.1: IRB Approval Letter (pg.1)

SPECIAL INSTRUCTIONS:

*Please note: The HRPP office has stopped stamping documents for Exempt protocols. It is your responsibility to maintain these documents and make current versions available on request.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.

Figure 8.2: IRB Approval Letter (pg.2)

Subject: User Study

Study Title: See the sounds and listen to the shapes

Hello,

My name is Woohun Joo. I am an iPhD student in the Graduate School. I am looking for participants for my user study. I have designed an audiovisual system that converts elementary visual shapes into sounds, and am about to conduct a user study (Virginia Tech IRB #21-097) to collect data that will help me examine what the ideal sound for each shape and movement will be and whether the system can provide multi-modal user experience.

Participants will listen to the sounds and provide answers to multiple-choice questions and subjective questions regarding whether or not sonification reasonably depicts the shapes, and the user's overall experience. This study is designed to provide a fun multi-sensory experience.

You will hear this!



Or this!



Who can participate?

Anyone (including Virginia Tech students, staff, and faculty) above 18 years old who has normal hearing and vision, and headphones is welcome to participate.

How to participate?

You can sign up for this user test here at [Appointment Calendar](#). The test will be held over Zoom and questions will be given via Qualtrics XM, an online survey platform. Both links will be provided after the test session is scheduled. The approximate time required to complete this

Figure 8.3: Recruitment flier (pg.1)

assessment will be 50-60 minutes. A \$10 Amazon gift card will be given at the end of successful completion of the assessment.

If you have any questions about this research, please do not hesitate to contact me. Thank you so much for considering participating in this study!

Best regards,
Woohun Joo <joowh@vt.edu>

Figure 8.4: Recruitment flier (pg.2)



Information Sheet for Participation in a Research Study

Principal Investigator: Patrick Finley

IRB# and Title of Study: IRB-21-097 - See the Sounds and listen to the Shapes

Sponsor: N/A

You are invited to participate in a research study. This form includes information about the study and contact information if you have any questions.

I am an iPhD (individualized PhD) student in the Graduate School at Virginia Tech, and I am conducting this research as part of my dissertation. The main goal of this research is to assess the effectiveness and experience of a newly developed method for converting elementary visual shapes (e.g., line, circle, rectangle, triangle) into sounds and, by doing so, explore a new way of visual-auditory communication design.

WHAT SHOULD I KNOW?

If you decide to participate in this study, you will complete a pre-questionnaire, the main survey, and a post-questionnaire. As part of the study, you will be asked to choose which sound fits better to each corresponding visual shape. Overall, study participants will be asked to comfortably experience the audiovisual contents that my system creates. This study is seeking to collect data that will help us examine whether this sonification approach can successfully deliver visual shapes in the form of a sound. After reading the consent form, the participants will participate in the main study by answering a series of multiple-choice and subjective questions. No pre-test is required.

The study should take approximately 60 minutes of your time.

We do not anticipate any risks from completing this study. However, you will be using your headphones through the entire study process and you will be exposed to the sound modulating in some cases. Although unlikely, you may experience brief and temporary side effects (lasting anywhere from 1-2 minutes) such as:

- Dizziness
- Residual pulsation in the ears or head from exposure to auditory stimuli

You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

CONFIDENTIALITY

We will do our best to protect the confidentiality of the information we gather from you, but we cannot guarantee 100% confidentiality. Your responses are anonymous, so no one can associate your answers back to you. Please do not include your name or other identifying information in your responses that can identify you.

WHO CAN I TALK TO?

If you have any questions or concerns about the research, please feel free to contact Woohun Joo at 401-580-3031 (joowh@vt.edu). You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research participant, contact the Virginia Tech HRPP Office at 540-231-3732 (irb@vt.edu).

Please print out a copy of this information sheet for your records.

If you would like to participate in this survey, click yes to begin or no to exit

- I consent, begin the study
- I do not consent, I do not wish to participate



Figure 8.5: Consent Form

Pre-Questionnaire

What is your age?

- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55 years or older

What is the highest level of formal training you have in music?

- Not at all
- Middle and high school
- Undergraduate/Bachelor
- Graduate/Masters
- Graduate/Doctoral
- Self-trained
- A professional

What is the highest level of formal training you have in visual art (e.g., studio art, media art, fine art, graphic design, etc)?

- Not at all
- Middle and high school
- Undergraduate/Bachelor
- Graduate/Masters
- Graduate/Doctoral
- Self-trained
- A professional

What is your listening device for this study?

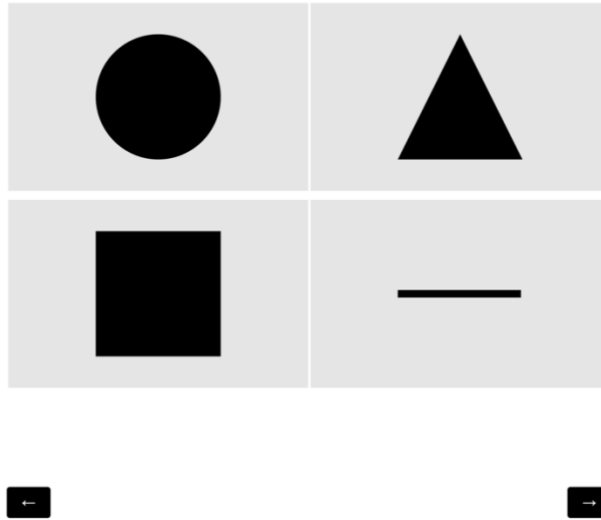
- Headphones
- Earbuds



Figure 8.6: Pre-Questionnaire

Task 1: Tonal Character + Four Basic Shapes

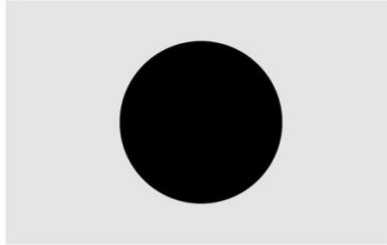
Each question will show one basic shape that comes with four different sounds. The four sounds are specially designed for four basic shapes (circle, square, triangle, and line). Please see the image and answer the question.



Powered by Qualtrics

Figure 8.7: Main Study Version 1 (Task 1 Intro)

Image: Circle



- Sound A ▶ 0:00 / 0:10 🔊 ⓘ
- Sound B ▶ 0:00 / 0:10 🔊 ⓘ
- Sound C ▶ 0:00 / 0:10 🔊 ⓘ
- Sound D ▶ 0:00 / 0:10 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Triangle



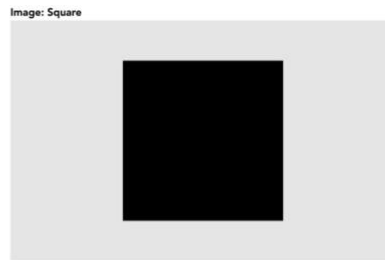
- Sound A ▶ 0:00 / 0:10 🔊 ⓘ
- Sound B ▶ 0:00 / 0:10 🔊 ⓘ
- Sound C ▶ 0:00 / 0:10 🔊 ⓘ
- Sound D ▶ 0:00 / 0:10 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Sound A ▶ 0:00 / 0:10 ———— 🔊 |

Sound B ▶ 0:00 / 0:10 ———— 🔊 |

Sound C ▶ 0:00 / 0:10 ———— 🔊 |

Sound D ▶ 0:00 / 0:10 ———— 🔊 |

Q1. From sounds A through D, which do you think describes the image the best?

Sound A

Sound B

Sound C

Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Sound A ▶ 0:00 / 0:10 ———— 🔊 |

Sound B ▶ 0:00 / 0:10 ———— 🔊 |

Sound C ▶ 0:00 / 0:10 ———— 🔊 |

Sound D ▶ 0:00 / 0:10 ———— 🔊 |

Q1. From sounds A through D, which do you think describes the image the best?

Sound A

Sound B

Sound C

Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 8.8: Main Study Version 1 (Task 1)



Task 2: Shape + Modulation

For each question, you will be presented with an image accompanied with three different sounds. The three sounds are specially designed for three basic shapes (circle, triangle, and square). Choose the sound response that best identifies the image. To start Task2, please move to the next page.

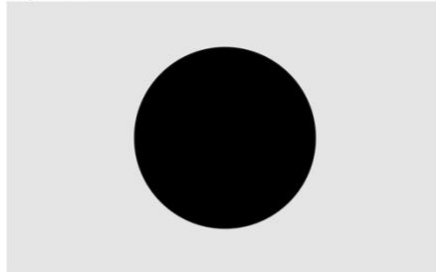


Powered by Qualtrics

Figure 8.9: Main Study Version 1 (Task 2 Intro)



Image: Circle



Sound A ▶ 0:00 / 0:14 ⋮

Sound B ▶ 0:00 / 0:13 ⋮

Sound C ▶ 0:00 / 0:14 ⋮

Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Triangle



Sound A ▶ 0:00 / 0:14 ⋮

Sound B ▶ 0:00 / 0:13 ⋮

Sound C ▶ 0:00 / 0:14 ⋮

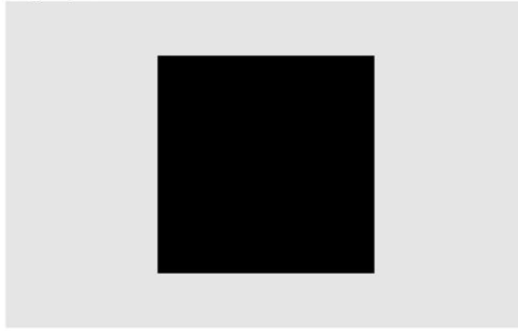
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Square



Sound A ▶ 0:00 / 0:14 ——— 🔊 ⋮

Sound B ▶ 0:00 / 0:13 ——— 🔊 ⋮

Sound C ▶ 0:00 / 0:14 ——— 🔊 ⋮

Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Powered by Qualtrics LT

Figure 8.10: Main Study Version 1 (Task 2)



Task 3: Line + Modulation

For each question, you will be presented with an image accompanied with four different sounds. The four sounds are specially designed for four line shapes (sine, sawtooth, pulse, and dashed line). Choose the sound response that best identifies the image. To start Task3, please move to the next page.

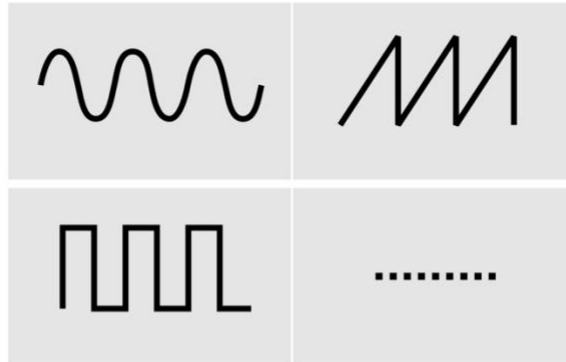
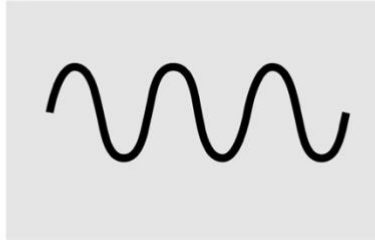


Image: Sine



- Sound A ▶ 0:00 / 0:18 ———— 🔊 ⓘ
- Sound B ▶ 0:00 / 0:15 ———— 🔊 ⓘ
- Sound C ▶ 0:00 / 0:13 ———— 🔊 ⓘ
- Sound D ▶ 0:00 / 0:12 ———— 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Sawtooth Line



- Sound A ▶ 0:00 / 0:18 ———— 🔊 ⓘ
- Sound B ▶ 0:00 / 0:15 ———— 🔊 ⓘ
- Sound C ▶ 0:00 / 0:13 ———— 🔊 ⓘ
- Sound D ▶ 0:00 / 0:12 ———— 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 8.11: Main Study Version 1 (Task 3 Intro)

Image: Pulse Line



Sound A ▶ 0:00 / 0:18 ———— 🔊 ⓘ

Sound B ▶ 0:00 / 0:15 ———— 🔊 ⓘ

Sound C ▶ 0:00 / 0:13 ———— 🔊 ⓘ

Sound D ▶ 0:00 / 0:12 ———— 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Dashed Line



Sound A ▶ 0:00 / 0:18 ———— 🔊 ⓘ

Sound B ▶ 0:00 / 0:15 ———— 🔊 ⓘ

Sound C ▶ 0:00 / 0:13 ———— 🔊 ⓘ

Sound D ▶ 0:00 / 0:12 ———— 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 8.12: Main Study Version 1 (Task 3)



Task 4: Other Shapes + Modulation

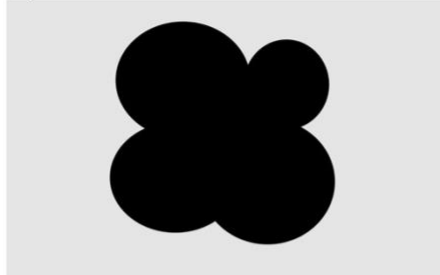
For each question, you will be presented with an image accompanied with three different sounds. The three sounds are specially designed for three shapes (bubble, spike, and stair). Choose the sound response that best identifies the image. To start Task4, please move to the next page.



Powered by Qualtrics

Figure 8.13: Main Study Version 1 (Task 4 Intro)

Image: Bubble



Sound A ▶ 0:00 / 0:12 🔊 ⋮

Sound B ▶ 0:00 / 0:12 🔊 ⋮

Sound C ▶ 0:00 / 0:12 🔊 ⋮

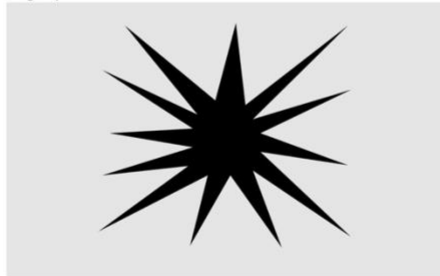
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Spike



Sound A ▶ 0:00 / 0:12 🔊 ⋮

Sound B ▶ 0:00 / 0:12 🔊 ⋮

Sound C ▶ 0:00 / 0:12 🔊 ⋮

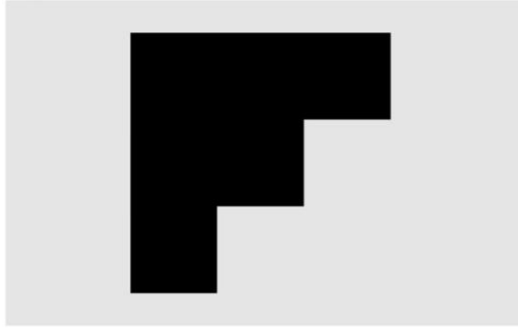
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Stair



Sound A ▶ 0:00 / 0:12 ———— 🔊 ⋮

Sound B ▶ 0:00 / 0:12 ———— 🔊 ⋮

Sound C ▶ 0:00 / 0:12 ———— 🔊 ⋮

Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Powered by Qualtrics

Figure 8.14: Main Study Version 1 (Task 4)



Task 5: Dynamic shapes

Each question will show a dynamic triangle shape that comes with three different sound effects. The three sound effects are specially designed to describe three dynamic shapes (tilting, zooming in and out, and moving left to right). Please watch the following videos and answer the questions. To start Task5, please move to the next page.

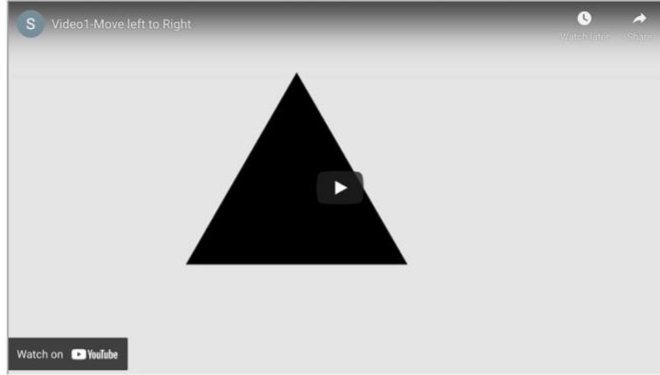


Figure 8.15: Main Study Version 1 (Task 5 Intro)



Video 1

Q1. From sounds A through C, which do you think describes the image the best?



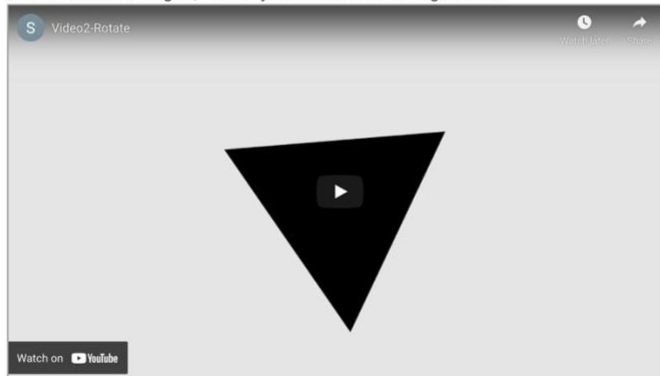
- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Video 2

Q1. From sounds A through C, which do you think describes the image the best?



- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Video 3

Q1. From sounds A through C, which do you think describes the image the best?



- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

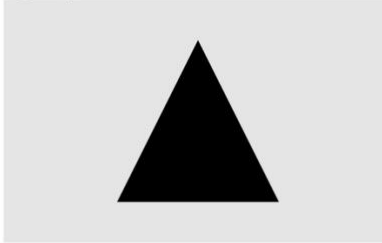


Powered by Qualtrics IQ

Figure 8.16: Main Study Version 1 (Task 5)



Image: Triangle



- Sound A ▶ 0:00 / 0:10  |
- Sound B ▶ 0:00 / 0:10  |
- Sound C ▶ 0:00 / 0:10  |
- Sound D ▶ 0:00 / 0:10  |

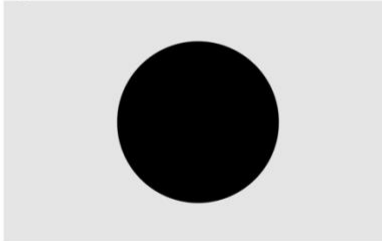
Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Circle



- Sound A ▶ 0:00 / 0:10  |
- Sound B ▶ 0:00 / 0:10  |
- Sound C ▶ 0:00 / 0:10  |
- Sound D ▶ 0:00 / 0:10  |

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Sound A ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Sound B ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Sound C ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Sound D ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best

Sound A

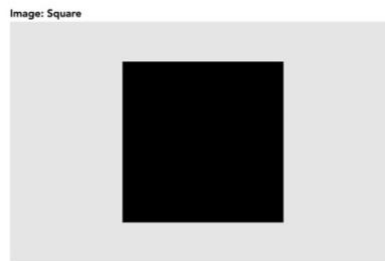
Sound B

Sound C

Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Sound A ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Sound B ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Sound C ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Sound D ▶ 0:00 / 0:10 ———— 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

Sound A

Sound B

Sound C

Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

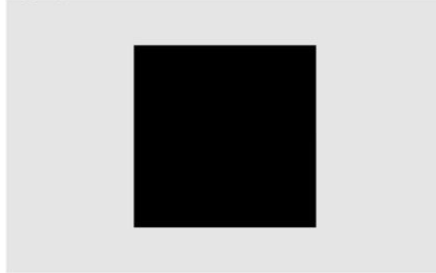
	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 8.17: Main Study Version 2 (Task 1)



Image: Square



▶ 0:00 / 0:14

▶ 0:00 / 0:13

▶ 0:00 / 0:14

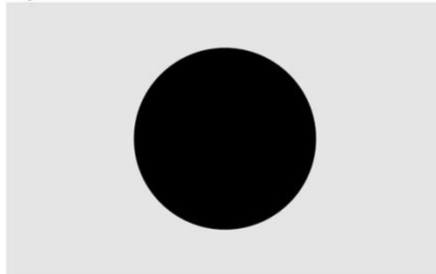
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Circle



▶ 0:00 / 0:14

▶ 0:00 / 0:13

▶ 0:00 / 0:14

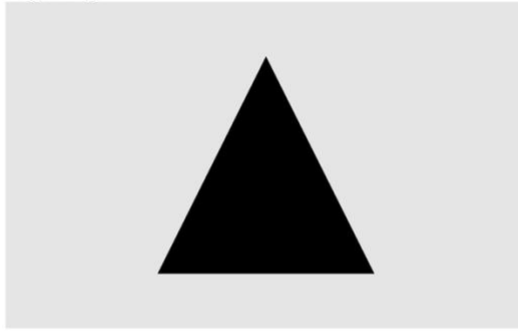
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Triangle



Sound A ▶ 0:00 / 0:14 ——— 🔊 ⋮

Sound B ▶ 0:00 / 0:13 ——— 🔊 ⋮

Sound C ▶ 0:00 / 0:14 ——— 🔊 ⋮

Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

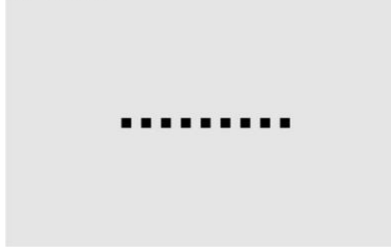


Powered by Qualtrics LT

Figure 8.18: Main Study Version 2 (Task 2)



Image: Dashed Line



Sound A ▶ 0:00 / 0:18 🔊 ⓘ

Sound B ▶ 0:00 / 0:15 🔊 ⓘ

Sound C ▶ 0:00 / 0:13 🔊 ⓘ

Sound D ▶ 0:00 / 0:12 🔊 ⓘ

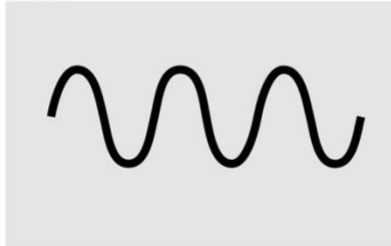
Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Sine



Sound A ▶ 0:00 / 0:18 🔊 ⓘ

Sound B ▶ 0:00 / 0:15 🔊 ⓘ

Sound C ▶ 0:00 / 0:13 🔊 ⓘ

Sound D ▶ 0:00 / 0:12 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Pulse Line



Sound A ▶ 0:00 / 0:18 🔊 ⓘ

Sound B ▶ 0:00 / 0:15 🔊 ⓘ

Sound C ▶ 0:00 / 0:13 🔊 ⓘ

Sound D ▶ 0:00 / 0:12 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Sawtooth Line



Sound A ▶ 0:00 / 0:18 🔊 ⓘ

Sound B ▶ 0:00 / 0:15 🔊 ⓘ

Sound C ▶ 0:00 / 0:13 🔊 ⓘ

Sound D ▶ 0:00 / 0:12 🔊 ⓘ

Q1. From sounds A through D, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C
- Sound D

Q2. On a scale from "very unlikely" to "very likely", please rate the four sounds based on how closely they match the image.

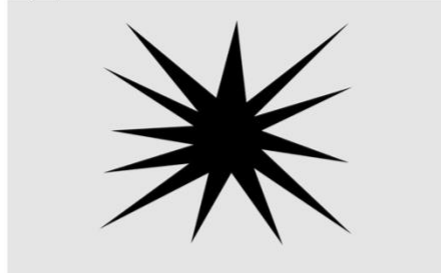
	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 8.19: Main Study Version 2 (Task 3)



Image: Spike



Sound A ▶ 0:00 / 0:12 ——— 🔊 ⓘ

Sound B ▶ 0:00 / 0:12 ——— 🔊 ⓘ

Sound C ▶ 0:00 / 0:12 ——— 🔊 ⓘ

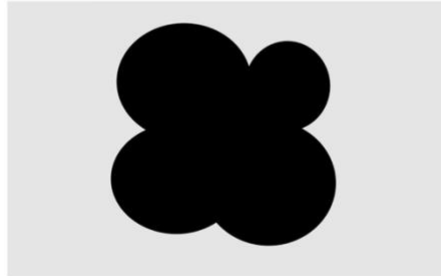
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Bubble



Sound A ▶ 0:00 / 0:12 ——— 🔊 ⓘ

Sound B ▶ 0:00 / 0:12 ——— 🔊 ⓘ

Sound C ▶ 0:00 / 0:12 ——— 🔊 ⓘ

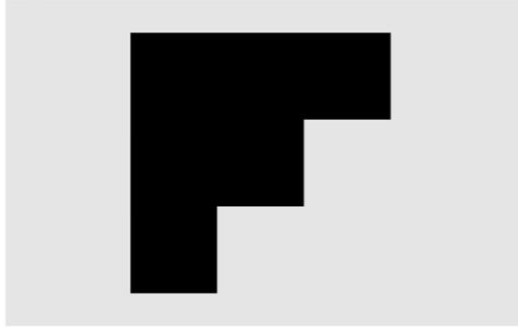
Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Image: Stair



Sound A ▶ 0:00 / 0:12 ——— 🔊 ⋮

Sound B ▶ 0:00 / 0:12 ——— 🔊 ⋮

Sound C ▶ 0:00 / 0:12 ——— 🔊 ⋮

Q1. From sounds A through C, which do you think describes the image the best?

- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



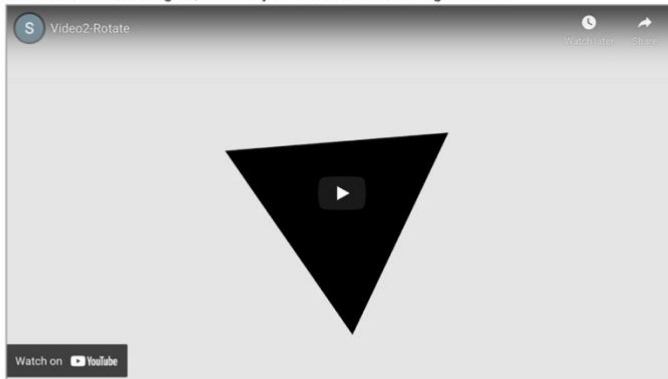
Powered by Qualtrics

Figure 8.20: Main Study Version 2 (Task 4)



Video 1

Q1. From sounds A through C, which do you think describes the image the best?



- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q1. From sounds A through C, which do you think describes the image the best?



- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Video 3

Q1. From sounds A through C, which do you think describes the image the best?



- Sound A
- Sound B
- Sound C

Q2. On a scale from "very unlikely" to "very likely", please rate the three sounds based on how closely they match the image.

	Very unlikely	Unlikely	Neutral	Likely	Very likely
Sound A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sound C	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Powered by Qualtrics iQ

Figure 8.21: Main Study Version 2 (Task 5)



Post-Questionnaire

I could easily match the sound with the visual.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

The study was fun and interesting.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

The sounds enhanced the visual information and provided multi-sensory experience (i.e., you experienced visually and auditory)

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

What do you like the most about the study?

Do you think this sonification (the use of sound to convey non-audio information) method has the potential in communication design, art, or education? If so, which areas would you see it being most useful (Select all that apply)?

- User interface design
- Interaction design
- Assistive technology
- Human computer interaction
- Education
- Art
- Music
- Other:

Why do you think so?

What task was the most challenging to you and why? (You can skip this question if the study questions were easy to answer).

Please share any other thoughts if you have them. This will help me improve this sonification system for future use.



Figure 8.22: Post-Questionnaire

9 References

- [1] “Adobe Flash Player End of Life.” [Online]. Available: <https://www.adobe.com/products/flashplayer/end-of-life-alternative.html>. [Accessed: 11-Sep-2021].
- [2] “Music - The War On Drugs.” [Online]. Available: <https://www.thewarondrugs.net/music/i-dont-live-here-anymore>. [Accessed: 25-Feb-2022].
- [3] “Dokho Shin on Instagram: ‘전시 미니멀리즘-맥시멀리즘-메커니즈즈즘 1막-2막을 위한 포스터 a poster for an exhibition Minimalism-Maximalism-Mechanissmmm Act 1-Act 2.’” [Online]. Available: https://www.instagram.com/p/CY4_JHoFyeF/. [Accessed: 25-Feb-2022].
- [4] M. Haverkamp, *Synesthetic design: Handbook for a multi-sensory approach*. Walter de Gruyter, 2012.
- [5] S. Ina, “Computer graphics for the blind,” *ACM SIGCAPH Comput. Phys. Handicap.*, no. 55, pp. 16–23, 1996.
- [6] J. Su, A. Rosenzweig, A. Goel, E. De Lara, and K. N. Truong, “Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices,” *ACM Int. Conf. Proceeding Ser.*, pp. 17–26, 2010.
- [7] E. Costes, J.-F. Bassereau, O. Rodi, and A. Aoussat, “Graphic design for blind users:an industrial case study,” *3rd IASDR World Conf. Des. Res.*, 2009.
- [8] S. Shimada, S. Yamamoto, Y. Uchida, M. Shinohara, Y. Shimizu, and M. Shimojo, “New design for a dynamic tactile graphic system for blind computer users,” *Proc. SICE Annu. Conf.*, pp. 1474–1477, 2008.
- [9] W. Köhler, “Gestalt psychology,” *Psychol. Forsch.*, vol. 31, no. 1, pp. XVIII–XXX, 1967.
- [10] “Graphic Design | Definition of Graphic Design by Merriam-Webster.” [Online]. Available: [https://www.merriam-webster.com/dictionary/graphic design](https://www.merriam-webster.com/dictionary/graphic%20design). [Accessed: 03-Jul-2020].
- [11] “Graphic | Definition of Graphic by Merriam-Webster.” [Online]. Available: <https://www.merriam-webster.com/dictionary/graphic>. [Accessed: 01-Feb-2021].
- [12] P. B. Meggs, *Type and image: The language of graphic design*. John Wiley & Sons, 1992.
- [13] IDEO, *The field guide to human-centered design*. Design Kit, 2015.
- [14] R. Buchanan, “Wicked Problems in Design Thinking Author (s): Richard Buchanan Published by : The MIT Press,” *Des. Issues*, vol. 8, no. 2, pp. 5–21, 1992.
- [15] P. B. Meggs and A. W. Purvis, *Meggs’ history of graphic design*. John Wiley & Sons, 2016.
- [16] G. Ambrose and P. Harris, *The Visual Dictionary of Graphic Design*. 2017.
- [17] K.-M. Lee and S. R. Ramsey, *A history of the Korean language*. Cambridge University Press, 2011.
- [18] A. M. Callan, D. E. Callan, and S. Masaki, “When meaningless symbols become letters: Neural activity change in learning new phonograms,” *Neuroimage*, vol. 28, no. 3, pp. 553–562, 2005.
- [19] Y.-K. Kim-Renaud, *The Korean alphabet: Its history and structure*. University of Hawaii Press, 1997.
- [20] S. J. Eskilson, *Graphic design: A new history*. Yale University Press, 2019.
- [21] “Graphic design | art | Britannica.” [Online]. Available: <https://www.britannica.com/art/graphic-design>. [Accessed: 01-Feb-2021].

- [22] E. Lupton and J. A. Miller, *ABC's of the Bauhaus:: The Bauhaus and Design Theory*, vol. 5. Princeton Architectural Press, 1991.
- [23] "Industry-leading vector graphics software | Adobe Illustrator." [Online]. Available: <https://www.adobe.com/products/illustrator.html?promoid=PGRQQLFS&mv=other>. [Accessed: 01-Feb-2021].
- [24] "Photo, image & design editing software | Buy Adobe Photoshop." [Online]. Available: <https://www.adobe.com/products/photoshop.html?promoid=PC1PQQ5T&mv=other>. [Accessed: 01-Feb-2021].
- [25] "QuarkXPress 2020 | Layout design and Desktop Publishing Software." [Online]. Available: https://www.quark.com/products/QuarkXPress/?utm_source=google&utm_medium=cpc&gclid=Cj0KCQiA6t6ABhDMARIsAONiYyxle3nQ4dRmUUz6M_F-YmvnaAVramkQ1iW8qx7k5pvm2hbzchW3TmcaAt9JEALw_wcB. [Accessed: 01-Feb-2021].
- [26] "Primary colors and shapes at the Bauhaus - Narratives - documenta archiv." [Online]. Available: <https://www.documenta-bauhaus.de/en/narrative/476/primary-colors-and-shapes-at-the-bauhaus>. [Accessed: 31-Dec-2021].
- [27] C. Chen, W. Härdle, A. Unwin, and M. Friendly, "A Brief History of Data Visualization," *Handb. Data Vis.*, pp. 15–56, 2008.
- [28] S. Few, "Data visualization for human perception," *Encycl. Human-Computer Interact. 2nd Ed.*, 2013.
- [29] E. R. Tufte, *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 2001.
- [30] P. J. Fitzpatrick, "Leading British Statisticians of the Nineteenth Century," *J. Am. Stat. Assoc.*, vol. 55, no. 289, pp. 38–70, 1960.
- [31] S. Heller and V. Vienne, *100 ideas that changed graphic design*. Laurence King London, 2012.
- [32] E. R. Tufte, N. H. Goeler, and R. Benson, *Envisioning information*, vol. 126. Graphics press Cheshire, CT, 1990.
- [33] "Microsoft 365 with Office apps | Microsoft 365." [Online]. Available: <https://www.microsoft.com/en-us/microsoft-365>. [Accessed: 02-Feb-2021].
- [34] "Keynote - Apple." [Online]. Available: <https://www.apple.com/keynote/>. [Accessed: 02-Feb-2021].
- [35] "Processing.org." [Online]. Available: <https://processing.org/>. [Accessed: 02-Feb-2021].
- [36] R. Oppermann, "User-interface Design BT - Handbook on Information Technologies for Education and Training," H. H. Adelsberger, B. Collis, and J. M. Pawlowski, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 233–248.
- [37] "What is User Interface Design? | Interaction Design Foundation (IxDF)." [Online]. Available: <https://www.interaction-design.org/literature/topics/ui-design>. [Accessed: 02-Feb-2021].
- [38] "Explore Windows 10 OS, Computers, Apps, & More | Microsoft." [Online]. Available: <https://www.microsoft.com/en-us/windows>. [Accessed: 02-Feb-2021].
- [39] "Android | The platform pushing what's possible." [Online]. Available: <https://www.android.com/>. [Accessed: 02-Feb-2021].
- [40] "Enterprise Open Source and Linux | Ubuntu." [Online]. Available: <https://ubuntu.com/>. [Accessed: 02-Feb-2021].

- [41] “macOS Big Sur - Apple.” [Online]. Available: <https://www.apple.com/macOS/big-sur/>. [Accessed: 02-Feb-2021].
- [42] K. L. Smith, S. Moriarty, K. Kenney, and G. Barbatsis, *Handbook of visual communication: Theory, methods, and media*. Routledge, 2004.
- [43] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. 2011.
- [44] D. Sless, *Learning and visual communication*. Halsted Press, 1981.
- [45] D. L. Ripley and T. Politzer, “Vision disturbance after TBI,” *NeuroRehabilitation*, vol. 27, no. 3, pp. 215–216, 2010.
- [46] R. R. and S. B. Wai Yu, “Haptic Graphs for Blind Computer Users,” *J. Clin. Ultrasound*, vol. 11, no. 9, pp. 507–509, 1983.
- [47] Y. Arthus-Bertrand, L. R. Brown, and N. Columbus, *Earth from above*. HN Abrams, 2002.
- [48] “YANN ARTHUS-BERTRAND PHOTO | From above | Official.” [Online]. Available: <https://www.yannarthusbertrandphoto.com/categorie-produit/from-above/>. [Accessed: 09-Feb-2021].
- [49] Y. Hatwell, A. Steri, and E. Gentaz, *Touching for knowing: cognitive psychology of haptic manual perception*, vol. 53. John Benjamins Publishing, 2003.
- [50] F. Richard, C. Vaz-Cerniglia, and S. Portalier, “Évolution des procédures d’exploration haptique chez des sujets voyants, aveugles tardifs et aveugles précoces,” *Eur. Rev. Appl. Psychol.*, vol. 54, no. 4, pp. 227–236, 2004.
- [51] A. J. Hollander, “An Exploration of Virtual Auditory Shape Perception,” 1994.
- [52] Kramer G, Walker B, Bonebright T, Cook P, and Flowers J.H, “Sonification Report: Status of the Field and Research Agenda,” *Fac. Publ. Dep. Psychol.* 444, no. March, 2010.
- [53] T. Hermann, “Taxonomy and definitions for sonification and auditory display,” 2008.
- [54] “definition | sonification.de.” [Online]. Available: <https://sonification.de/son/definition/>. [Accessed: 20-Aug-2020].
- [55] G. J. Sandell, “Auditory Display: Sonification, Audification, and Auditory Interfaces.” JSTOR, 1996.
- [56] D. Worrall, *Sonification Design: From Data to Intelligible Soundfields*. Springer, 2019.
- [57] S. Barrass, “Sonification Design Patterns,” *Int. Conf. Audit. Disp.* 2003, no. July, pp. 170–175, 2003.
- [58] S. Barrass and G. Kramer, “Using sonification,” *Multimed. Syst.*, vol. 7, no. 1, pp. 23–31, 1999.
- [59] A. L. Alcaraz, “Cyborgs’ Perception, Cognition, Society, Environment, and Ethics: Interview with Neil Harbisson and Moon Ribas, 14 October 2016, Ace Hotel, New York City,” *J. Posthuman Stud.*, vol. 3, no. 1, pp. 60–73, 2019.
- [60] P. Revuelta Sanz, B. Ruiz-Mezcua, and J. M. Sánchez Pena, “A sonification proposal for safe travels of blind people,” 2012.
- [61] T. Pun, B. Deville, and G. Bologna, “Sonification of colour and depth in a mobility aid for blind people,” 2010.
- [62] S. Saue, “A model for interaction in exploratory sonification displays,” 2000.
- [63] J. Tardieu, N. Misdariis, S. Langlois, P. Gaillard, and C. Lemercier, “Sonification of in-vehicle

- interface reduces gaze movements under dual-task condition,” *Appl. Ergon.*, vol. 50, pp. 41–49, 2015.
- [64] D. Di Censo and S. Marti, “Spatial sonification of accelerating objects.” Google Patents, 21-Feb-2017.
- [65] M. Jeon, “Advanced vehicle sonification applications,” in *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2014, pp. 1–5.
- [66] N. Ronnberg, “Musical Elements in Sonification Support Visual Perception,” pp. 114–117, 2019.
- [67] M. Watson and P. Sanderson, “Sonification supports eyes-free respiratory monitoring and task time-sharing,” *Hum. Factors*, vol. 46, no. 3, pp. 497–517, 2004.
- [68] W. Buxton, “Introduction to this special issue on non-speech audio.,” *Human-Computer Interact.*, vol. 4, pp. 1–9, 1989.
- [69] G. Kramer, “An introduction to auditory display,” *Audit. display-Sonification, audification Audit. interfaces*, pp. 1–77, 1994.
- [70] B. N. Walker and G. Kramer, “Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making,” *Ecol. psychoacoustics*, pp. 150–175, 2004.
- [71] M. Quinn, “Research set to music: The climate symphony and other sonifications of ice core, radar, DNA, seismic and solar wind data,” 2001.
- [72] M. Quinn, W. Quinn, and B. Hatcher, “For those who died: A 9/11 tribute,” 2003.
- [73] T. Hermann, “Sonification for exploratory data analysis.” 2002.
- [74] M. M. Blattner, D. A. Sumikawa, and M. Robert, “Human – Computer Interaction Earcons and Icons : Their Structure and Common Design Principles,” vol. 0024, no. August 2014, pp. 37–41, 2009.
- [75] J. M. Hull, *On sight & insight: a journey into the world of blindness*. Oneworld publ., 1997.
- [76] J. Rantala *et al.*, “Methods for presenting braille characters on a mobile device with a touchscreen and tactile feedback,” *IEEE Trans. Haptics*, vol. 2, no. 1, pp. 28–39, 2009.
- [77] E. D. Mynatt and G. Weber, “Nonvisual presentation of graphical user interfaces: contrasting two approaches,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1994, pp. 166–172.
- [78] “JAWS® – Freedom Scientific.” [Online]. Available: <https://www.freedomscientific.com/products/software/jaws/>. [Accessed: 12-Feb-2021].
- [79] “Complete guide to Narrator.” [Online]. Available: <https://support.microsoft.com/en-us/windows/complete-guide-to-narrator-e4397a0d-ef4f-b386-d8ae-c172f109bdb1>. [Accessed: 12-Feb-2021].
- [80] “Accessibility - Vision - Apple.” [Online]. Available: <https://www.apple.com/accessibility/vision/>. [Accessed: 12-Feb-2021].
- [81] “Sonar Curriculum Handbook - American Printing House for the Blind.” [Online]. Available: <https://www.yumpu.com/en/document/view/52138365/sonar-curriculum-handbook-american-printing-house-for-the-blind>. [Accessed: 12-Feb-2021].
- [82] “UltraCaneputting the world at your fingertips.” [Online]. Available: <https://www.ultracane.com/>. [Accessed: 12-Feb-2021].

- [83] J. Borenstein, I. Ulrich, and S. Shoval, "Computerized obstacle avoidance systems for the blind and visually impaired," no. March, 2000.
- [84] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the navbelt-a computerized travel aid for the blind," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.)*, vol. 28, no. 3, pp. 459–467, 1998.
- [85] P.B.L. Meijer, "An experimental system for auditory image representations."
- [86] P. Meijer, "The vOICE-seeing with sound." February, 1996.
- [87] "The vOICE for Android on smart glasses." [Online]. Available: <https://www.seeingwithsound.com/android-glasses.htm>. [Accessed: 31-Dec-2021].
- [88] T. Adajian, "Visual Music: Synaesthesia in Art and Music Since 1900 edited by brougher, kerry, olivia mattis, jeremy strick, ari wiseman and judith zilczer," *J. Aesthet. Art Crit.*, vol. 64, pp. 488–489, Sep. 2006.
- [89] K. Peacock, "Synesthetic Perception: Alexander Scriabin's Color Hearing," *Music Percept.*, vol. 2, no. 4, pp. 483–505, 1985.
- [90] W. Kandinsky, *Concerning the spiritual in art*. Courier Corporation, 2012.
- [91] M. Dabrowski, "Kandinsky compositions: The music of the spheres," *MoMA*, no. 19, pp. 10–13, 1995.
- [92] W. Kandinsky and M. Press, *A Passage from Wassily Kandinsky's Reminiscences*. Moonkosh Press, 1987.
- [93] W. Kandinsky and H. Rebay, *Point and line to plane*. Courier Corporation, 1979.
- [94] R. R. Pevitts, "WASSILY KANDINSKY'S" THE YELLOW SOUND": A SYNTHESIS OF THE ARTS FOR THE STAGE.," 1981.
- [95] P. Klee, *The Diaries of Paul Klee, 1898-1918*. Univ of California Press, 1964.
- [96] A. Sheldon and S. Exhibition, "Painting Music :," *Structure*, pp. 1–65, 2007.
- [97] H. Dichtung, *Paul Klee: Painting Music*. Prestel, 2004.
- [98] "Paul Klee | Abstract Trio | The Metropolitan Museum of Art." [Online]. Available: <https://www.metmuseum.org/art/collection/search/483154>. [Accessed: 12-Feb-2021].
- [99] R. Verdi, "Musical influences on the art of Paul Klee," *Art Inst. Chicago Museum Stud.*, vol. 3, pp. 81–107, 1968.
- [100] D. Roberts, *The total work of art in European modernism*. Cornell University Press, 2011.
- [101] "Paul Klee: Fugue in Red : Interlude." [Online]. Available: <https://interlude.hk/paul-klee-fugue-red/>. [Accessed: 12-Feb-2021].
- [102] M. Graves, "The art of color and design," *J. Aesthet. Art Crit.*, vol. 2, no. 5, 1942.
- [103] S. Putnam, "The Glistening Bridge: Leopold Survace and the Spatial Problem in Painting," 1929.
- [104] "Léopold Survage. Colored Rhythm: Study for the Film. (1913) | MoMA." [Online]. Available: <https://www.moma.org/collection/works/85487>. [Accessed: 12-Feb-2021].
- [105] C. Steen, "A R T I S T ' S Visions Shared : A Firsthand Look," vol. 34, no. 3, pp. 203–208, 2021.
- [106] C. Van Campen, *The hidden sense: Synesthesia in art and science*. Mit Press, 2010.
- [107] F. McDonald, "Bringing out the contours," *Aust. Art Rev.*, no. July-October, pp. 76–78, 2006.

- [108] K. Peacock, "Synesthetic Perception: Alexander Scriabin's Color Hearing," *Music Percept. An Interdiscip. J.*, vol. 2, no. 4, pp. 483–505, 1985.
- [109] L. Sabaneev and S. W. Pring, "The relation between sound and colour," *Music Lett.*, vol. 10, no. 3, pp. 266–277, 1929.
- [110] B. M. Galeyev and I. L. Vanechkina, "Was Scriabin a synesthete?," *Leonardo*, vol. 34, no. 4, pp. 357–361, 2001.
- [111] J. Cage, "Notations," *J. Aesthet. Art Crit.*, vol. 28, no. 4, 1970.
- [112] A. Mount, "Testing Art with a Live Studio Audience : John Cage at an Intersection of Popular Television and the Avant-Garde Author (s): Andre Mount Source : Music and the Moving Image , Vol . 4 , No . 3 (Fall 2011) , pp . 31-56 Published by : University of Illino," vol. 4, no. 3, pp. 31–56, 2011.
- [113] "John Cage Complete Works." [Online]. Available: https://johncage.org/pp/John-Cage-Work-Detail.cfm?work_ID=241. [Accessed: 30-Aug-2021].
- [114] "John Cage Complete Works." [Online]. Available: https://johncage.org/pp/John-Cage-Work-Detail.cfm?work_ID=29. [Accessed: 30-Aug-2021].
- [115] R. Haskins, *John Cage*. Reaktion Books, 2013.
- [116] "John Cage Complete Works." [Online]. Available: https://johncage.org/pp/John-Cage-Work-Detail.cfm?work_ID=79. [Accessed: 30-Aug-2021].
- [117] "BBC - Radio 3 Blog: 433 words about John Cage's 4'33" - and the first ever radio realisation." [Online]. Available: <https://www.bbc.co.uk/blogs/radio3/2011/11/433-words-about-john-cages-433-and-the-first-ever-radio-realisation.shtml>. [Accessed: 14-Apr-2022].
- [118] E. Brown, "On December 1952," *Am. Music*, vol. 26, no. 1, pp. 1–12, 2008.
- [119] C. Samuel and O. Messiaen, *Conversations with Olivier Messiaen*. Stainer & Bell, Limited, 1976.
- [120] "Visualizing Visions: The Significance of Messiaen's..." [Online]. Available: <https://www.musicandpractice.org/volume-2/visualizing-visions-the-significance-of-messiaens-colours/>. [Accessed: 29-Aug-2021].
- [121] M. Baggech, "An English Translation of Oliver Messiaen's *Traité de Rythme, de Couleur, et D'ornithologie* Volume I," 1998.
- [122] M. T. King and S. Weiss, "OLIVIER MESSIAEN ' S COULEURS DE LA CITÉ CÉLESTE : by," 2015.
- [123] B. Ferneyhough, *Karlheinz Stockhausen*. 2018.
- [124] M. Williams, "Stockhausen: 'Nr. 9 Zyklus,'" *Elements*, no. 9, pp. 60–69, 2001.
- [125] "Stockhausen: Sounds in Space: MIKROPHONIE I." [Online]. Available: <http://stockhausenspace.blogspot.com/2014/05/opus-15-mikrophonie-i.html>. [Accessed: 30-Aug-2021].
- [126] M. E. Rörich, "Penderecki's "Threnody for the Victims of Hiroshima": a case study in the dangers of expressive aesthetics and the limits of formal analysis," *Pap. Knowl. . Towar. a Media Hist. Doc.*, 2014.
- [127] "Score of Karlheinz Stockhausen's Zyklus with eligibility in the... | Download Scientific Diagram." [Online]. Available: https://www.researchgate.net/figure/Score-of-Karlheinz-Stockhausens-Zyklus-with-eligibility-in-the-different-instrumental_fig4_339988722. [Accessed: 15-Apr-2022].

- [128] “Stockhausen: Sounds in Space: ETUDE, STUDIE I & II.” [Online]. Available: <http://stockhausenspace.blogspot.com/2014/12/opus-3-studie-i-studie-ii-and-etude.html>. [Accessed: 15-Apr-2022].
- [129] “The Graphic Notation of Krzysztof Penderecki’s ‘Threnody for the Victims of Hiroshima’ | Music History Collaborative Blog.” [Online]. Available: <https://musichistoryfsu.wordpress.com/2015/04/06/the-graphic-notation-of-krzysztof-pendereckis-threnody-for-the-victims-of-hiroshima/>. [Accessed: 15-Apr-2022].
- [130] C. Cardew *et al.*, *Treatise*, no. 7560. Gallery Upstairs Press, 1967.
- [131] G. Ligeti and R. Wehinger, *Artikulation*. Schott, 1970.
- [132] “‘Treatise’: A Visual Symphony Of Information Design | by Jason Forrest | Nightingale | Medium.” [Online]. Available: <https://medium.com/nightingale/treatise-a-visual-symphony-of-information-design-2ced33ef01a0>. [Accessed: 15-Apr-2022].
- [133] “Escot - Your Kindled Valors Bend - Art and music collide in these 20 stunning... - Classic FM.” [Online]. Available: <https://www.classicfm.com/discover-music/latest/graphic-scores-art-music-pictures/escot-your-kindled-valors-bend/>. [Accessed: 15-Apr-2022].
- [134] “Stripsody: Transforming Comics into Vanguard Art - Interdisciplinary Italy.” [Online]. Available: <http://www.interdisciplinaryitaly.org/stripsody-transforming-comics-into-vanguard-art/>. [Accessed: 31-Aug-2021].
- [135] “Critics At Large : American Maestro: George Crumb Interpreted by Yoshiko Shimizu.” [Online]. Available: <https://www.criticsatlarge.ca/2019/12/american-maestro-george-crumb.html>. [Accessed: 15-Apr-2022].
- [136] “Score components (Slides a. through l.) of Mauricio Kagel: Prima Vista... | Download Scientific Diagram.” [Online]. Available: https://www.researchgate.net/figure/Score-components-Slides-a-through-l-of-Mauricio-Kagel-Prima-Vista-1962-3-Excerpt_fig1_290428734. [Accessed: 15-Apr-2022].
- [137] “‘Stripsody’ by Cathy Berberian – Houston Public Media.” [Online]. Available: <https://www.houstonpublicmedia.org/berberian-stripsody-pg4/>. [Accessed: 15-Apr-2022].
- [138] “Tom Phillips - Golden Flower Piece.” [Online]. Available: <https://www.tomphillips.co.uk/works/music-scores/item/5418-golden-flower-piece#>. [Accessed: 01-Sep-2021].
- [139] “Tom Phillips - Ornamentik Opus IX.” [Online]. Available: <https://www.tomphillips.co.uk/works/music-scores/item/5281-ornamentik-opus-ix>. [Accessed: 01-Sep-2021].
- [140] J. T. Lysaker, *Brian Eno’s Ambient 1: Music for Airports*. Oxford University Press, 2018.
- [141] “Review: New Art Music Ensemble | Music Durham.” [Online]. Available: https://www.musicdurham.org/review-new-art-music-ensemble/?__cf_chl_jschl_tk__=pmd_UbupaV72aYh9vnemyj.mI5rud6BGPH.MyUAYh7R.Wd8-1630528598-0-gqNtZGzNArujcnBszQh9. [Accessed: 01-Sep-2021].
- [142] “Tom Phillips - Golden Flower Piece.” [Online]. Available: <https://www.tomphillips.co.uk/works/music-scores/item/5418-golden-flower-piece>. [Accessed: 15-Apr-2022].
- [143] “Music for Airports Soothes the Savage Passenger | Arts & Culture| Smithsonian Magazine.” [Online]. Available: <https://www.smithsonianmag.com/arts-culture/music-for-airports-soothes-the->

- savage-passenger-116079759/. [Accessed: 15-Apr-2022].
- [144] “Carl Bergström-Nielsen: Towards an Unbearable Lightness (1992) by Edition·S – music–sound–art - Issuu.” [Online]. Available: https://issuu.com/edition-s/docs/towards_an_unbearable_lightness_a4land_/3?e=3221088/81280694. [Accessed: 15-Apr-2022].
- [145] “Graphic Scores by Barry Guy.” [Online]. Available: <https://www.pointofdeparture.org/PoD38/PoD38Guy.html>. [Accessed: 01-Sep-2021].
- [146] “MAKOTO NOMURA: SHOGI COMPOSITION (2002).” [Online]. Available: https://intuitivemusic.dk/iima/nomura_shogi.htm. [Accessed: 01-Sep-2021].
- [147] “Maya Recordings, Barry Guy, Maya Homburger.” [Online]. Available: https://mayarecordings.com/graphic_scores. [Accessed: 15-Apr-2022].
- [148] “The Music of Anthony Braxton: A System of Possibilities — Tri-Centric Foundation.” [Online]. Available: <https://tricentricfoundation.org/carl-testa-essay>. [Accessed: 01-Sep-2021].
- [149] T. Sauer, *Notations 21*. Mark Batty Publisher, 2009.
- [150] “Graphic music scores - in pictures | Music | The Guardian.” [Online]. Available: <https://www.theguardian.com/music/gallery/2013/oct/04/graphic-music-scores-in-pictures>. [Accessed: 01-Sep-2021].
- [151] “Music To Be Heard — And Seen : Deceptive Cadence : NPR.” [Online]. Available: <https://www.npr.org/sections/deceptivecadence/2011/06/14/136822623/music-to-be-heard-and-seen>. [Accessed: 15-Apr-2022].
- [152] “steve roden . work . sounding architecture (and pavilion scores).” [Online]. Available: <http://www.inbetweennoise.com/works/sounding-architecture-and-pavilion-scores/>. [Accessed: 01-Sep-2021].
- [153] “TEDx YorkU: 2010 | Daniel Schnee.” [Online]. Available: <https://danielpaulschnee.wordpress.com/2014/10/28/6215/>. [Accessed: 01-Sep-2021].
- [154] “Graphic scores: lively alternative ways of writing down music | Financial Times.” [Online]. Available: <https://www.ft.com/content/7b14f656-243d-11e3-8905-00144feab7de>. [Accessed: 15-Apr-2022].
- [155] “NIU Today | Music to my eyes: NIU Art Musuem to unveil three spring exhibitions concerned with sounds.” [Online]. Available: <https://niutoday.info/2011/03/18/music-to-my-eyes-niu-art-musuem-to-unveil-three-spring-exhibitions-concerned-with-sounds/>. [Accessed: 15-Apr-2022].
- [156] “Music Animation Machine.” [Online]. Available: <https://www.musanim.com/>. [Accessed: 02-Sep-2021].
- [157] G. Bendazzi, “Chromatic Music by Bruno Ginanni Corradini (1912),” in *A Moving Subject*, CRC Press, 2020, pp. 11–20.
- [158] G. Bendazzi, “The Italians Who Invented the Drawn-On Film Technique,” *Animat. J.*, vol. 4, pp. 69–77, 1996.
- [159] “Hans Richter - Rhythmus 21 (Rhythm 21).” [Online]. Available: <https://www.museoreinasofia.es/en/collection/artwork/rhythmus-21-rhythm-21>. [Accessed: 12-Feb-2021].
- [160] M. Cowan, “Absolute Advertising: Walter Ruttmann and the Weimar Advertising Film,” *Cine. J.*, pp. 49–73, 2013.

- [161] M. Le Grice, "Abstract Film and Beyond.," 1977.
- [162] A. Abbado, *Visual Music Masters: Abstract Explorations of Past and Present Artists*. Skira, 2014.
- [163] "See this Sound (<http://see-this-sound.at>)." [Online]. Available: <http://www.see-this-sound.at/print/work/175.html>. [Accessed: 12-Feb-2021].
- [164] "Rhythmus 21, an Early Abstract Film by Hans Richter (1921) – SOCKS." [Online]. Available: <https://socks-studio.com/2014/07/30/rhythmus-21-an-early-abstract-film-by-hans-richter-1921/>. [Accessed: 04-Sep-2021].
- [165] "Still images from Viking Eggeling's *Symphonie Diagonale* (1924). It is... | Download Scientific Diagram." [Online]. Available: https://www.researchgate.net/figure/Still-images-from-Viking-Eggelings-Symphonie-Diagonale-1924-It-is-by-now-no-longer_fig4_261377470. [Accessed: 15-Apr-2022].
- [166] "Hans Richter. *Rhythmus 21*. 1921 | MoMA." [Online]. Available: <https://www.moma.org/collection/works/309358>. [Accessed: 15-Apr-2022].
- [167] R. Russett and C. Starr, *Experimental Animation*. Da Capo Press, 1988.
- [168] "Importance of Being Fischinger." [Online]. Available: <http://www.centerforvisualmusic.org/library/ImportBF.htm>. [Accessed: 12-Feb-2021].
- [169] "CVM Fischinger pages - Spiritual in Art - Fischinger excerpt." [Online]. Available: <http://centerforvisualmusic.org/Fischinger/AbFilmColorMusic.htm>. [Accessed: 03-Sep-2021].
- [170] "Oskar Fischinger | The Guggenheim Museums and Foundation." [Online]. Available: <https://www.guggenheim.org/artwork/artist/oskar-fischinger>. [Accessed: 03-Sep-2021].
- [171] O. Fischinger, C. Keefer, J. Guldemon, and E. Y. E. F. I. Nederland, *Oskar Fischinger 1900-1967: Experiments in Cinematic Abstraction*. EYE Filmmuseum, 2013.
- [172] "Oskar Fischinger's *Allegretto*, from 1936 (left) and *AVOL* (right) | Download Scientific Diagram." [Online]. Available: https://www.researchgate.net/figure/Oskar-Fischingers-Allegretto-from-1936-left-and-AVOL-right_fig4_236898246. [Accessed: 15-Apr-2022].
- [173] D. Cai, S. Goto, T. Shinohara, N. Nagata, J. Kurumisawa, and A. Fukumoto, "Synesthetic color scheme in *Fantasia*," *ACM SIGGRAPH 2010 Talks, SIGGRAPH '10*, p. 4503, 2010.
- [174] N. McLaren and W. Jordan, "Notes on Animated Sound," *Q. Film Radio Telev.*, vol. 7, no. 3, pp. 223–229, 1953.
- [175] P. Melançon, "Begone Dull Care," *Senses Cine.*, 2005.
- [176] N. McLaren, *Synchromy*. National Film Board of Canada, 1971.
- [177] N. Danino, M. Mazière, and U. magazine, *The Undercut Reader: Critical Writings on Artists' Film and Video*. Wallflower, 2003.
- [178] "How to write a film on a piano: Norman McLaren's visual music | Sight & Sound | BFI." [Online]. Available: <https://www2.bfi.org.uk/news-opinion/sight-sound-magazine/features/how-write-film-piano-norman-mclaren-s-visual-music>. [Accessed: 15-Apr-2022].
- [179] W. Moritz, "Who's Who in Filmmaking: James Whitney." *Sightlines*, 1985.
- [180] V. Sorensen, "Sound, Media Art, and the Metaverse," *array. J. ICMA*, vol. 2016, no. August 2015, pp. 5–8, 2020.
- [181] A. Hitchcock *et al.*, *Vertigo*. Universal, 1958.
- [182] G. Youngblood, "The stargate corridor," in *Expanded Cinema*, Fordham University Press, 2020,

pp. 151–156.

- [183] J. Whitney, “How Shall Motion Pattern Time?,” *Digital Harmony. On the Complementary of Music and Visual Art*. pp. 37–45, 1980.
- [184] “CVM - Mary Ellen Bute.” [Online]. Available: <http://www.centerforvisualmusic.org/ButeRetrospective.htm>. [Accessed: 03-Oct-2021].
- [185] J. Powers, “A DIY Come-On: A History of Optical Printing in Avant-Garde Cinema,” *Cine. J.*, vol. 57, no. 4, pp. 71–95, 2018.
- [186] L. Lye, J. Annear, and N. Bullock, *Len Lye*. Art Gallery of New South Wales, 2000.
- [187] S. MACDONALD, “A Critical Cinema 3: Interviews with Independent Cinema.” Berkeley, 1998.
- [188] K. Brougher, O. Mattis, C. . Museum of Contemporary Art (Los Angeles, and H. M. and S. Garden, *Visual Music: Synaesthesia in Art and Music Since 1900*. Thames & Hudson, 2005.
- [189] T. McSorley, “Short Takes-Bru Ha! Ha!,” *Tak. One Film Telev. Canada*, vol. 12, no. 43, 2003.
- [190] “Cameras Take Five | Canadian Filmmakers Distribution Centre.” [Online]. Available: <https://www.cfmdc.org/film/2135>. [Accessed: 08-Sep-2021].
- [191] “Robert Seidel — _grau.” [Online]. Available: <https://robertseidel.com/grau/>. [Accessed: 07-Sep-2021].
- [192] “ICC | ‘Forest and Trees’ - ONISHI Keita (2011-).” [Online]. Available: <https://www.ntticc.or.jp/en/archive/works/forest-and-trees/>. [Accessed: 08-Sep-2021].
- [193] R. B. Elder, *Harmony and dissent: Film and avant-garde art movements in the early twentieth century*. Wilfrid Laurier Univ. Press, 2008.
- [194] “CABINET / The Scale and the Spectrum.” [Online]. Available: <https://www.cabinetmagazine.org/issues/22/peel.php>. [Accessed: 20-Aug-2021].
- [195] M. Dunant, “Description of M. Kastner’s new musical instrument, the pyrophone,” *J. Franklin Inst.*, vol. 99, no. 5, pp. 367–375, 1875.
- [196] “Ocular Harpsichord :: Museum of Imaginary Musical Instruments.” [Online]. Available: <http://imaginaryinstruments.org/ocular-harpsichord/>. [Accessed: 15-Apr-2022].
- [197] “The Amazing Adventures Of Kastner’s Miraculous Pyrophone (Part One) - Science Museum Blog.” [Online]. Available: <https://blog.sciencemuseum.org.uk/pyrophone1/>. [Accessed: 15-Apr-2022].
- [198] L. Ribas, “Performativity as a Perspective on Sound-Image Relations and Audiovisuality,” *Mono*, vol. 2, pp. 29–50, 2014.
- [199] “Optophonic Piano, Music, Light, Vladimir Baranoff-Rossiné.” [Online]. Available: <https://baranoff-rossine.com/optophonic-piano/>. [Accessed: 12-Feb-2021].
- [200] J. Bowlt, “Concepts of Color and the Soviet Avant-Garde,” *Struct.*, no. 13, p. 21, 1973.
- [201] K. Peacock, “Instruments to perform color-music: Two centuries of technological experimentation,” *Leonardo*, vol. 21, no. 4, pp. 397–406, 1988.
- [202] M. Wright, “NOURATHAR: An early 20th century color organ ,” *26th Annu. Conf. Soc. Electro-Acoustic Music*, no. September, pp. 1–9, 2011.
- [203] M. Hallock-Greenewalt, *Mary Hallock-Greenewalt: The Complete Patents*. Wildside Press LLC, 2005.

- [204] “Reflecting Light Games.” [Online]. Available: <https://www.bauhauskooperation.com/knowledge/the-bauhaus/works/photography/reflecting-light-games/>. [Accessed: 08-Sep-2021].
- [205] “On the Reconstruction of Kurt Schwertfeger’s ‘Reflektorische Farblichtspiele’ (Reflective Colored Light Plays) from 1922 - Articles – bauhaus imaginista.” [Online]. Available: <http://www.bauhaus-imaginista.org/articles/6211/on-the-reconstruction-of-kurt-schwertfeger-s-reflektorische-farblichtspiele-reflective-colored-light-plays-from-1922>. [Accessed: 08-Sep-2021].
- [206] A. B. Klein, *Colour Cinematography*. Chapman & Hall, Limited, 1940.
- [207] M. Betancourt, *Thomas Wilfred’s Clavilux*. Wildside Press LLC, 2006.
- [208] N. Rimsky-Korsakov, *Scheherazade*. Courier Corporation, 1984.
- [209] W. C. Wees, *Light-Play and the Aesthetics of Avant-Garde Film*. Brill Rodopi, 2007.
- [210] W. Moritz, “The dream of color music, and machines that made it possible,” *Animat. World Mag.*, vol. 2, no. 1, 1997.
- [211] R. S. James, “Avant-Garde Sound-on-Film Techniques and Their Relationship to Electro-Acoustic Music Author (s): Richard S . James Published by : Oxford University Press Stable URL : <http://www.jstor.org/stable/948107> Accessed : 15-06-2016 09 : 33 UTC Avant-Garde Soun,” *Music. Q.*, vol. 72, no. 1, pp. 74–89, 1986.
- [212] “Listen to early Soviet synthesizer music, hand drawn on film and made from cut paper | Dangerous Minds.” [Online]. Available: https://dangerousminds.net/comments/listen_to_early_soviet_synthesizer_music_hand_drawn_on_film_and_made_from_c. [Accessed: 12-Feb-2021].
- [213] S. Kreichi, “The ANS synthesizer: Composing on a photoelectronic instrument,” *Leonardo*, pp. 59–62, 1995.
- [214] P. D. Manning, “The Oramics machine: from vision to reality.,” *Organised sound.*, vol. 17, no. 2, pp. 137–147, 2012.
- [215] U. Haque, “The architectural relevance of Gordon Pask,” *Archit. Des.*, vol. 77, no. 4, pp. 54–61, 2007.
- [216] “Cinegraphic: Mary Hallock-Greenewalt’s Nourathar.” [Online]. Available: <https://www.cinegraphic.net/article.php?story=20110220085223923&mode=print>. [Accessed: 15-Apr-2022].
- [217] “The ‘Variophone’ Yevgeny Sholpo. Russia, 1932 – 120 Years of Electronic Music.” [Online]. Available: <https://120years.net/wordpress/the-variophone-yevgeny-sholpo-soviet-union-1932/>. [Accessed: 15-Apr-2022].
- [218] “The ‘ANS Synthesiser’ Yevgeny Murzin. Russia, 1958 – 120 Years of Electronic Music.” [Online]. Available: <https://120years.net/wordpress/the-ans-synthesiser-yevgeniy-murzin-soviet-union-1958-2/>. [Accessed: 15-Apr-2022].
- [219] “‘Oramics’, Daphne Oram, UK, 1959. – 120 Years of Electronic Music.” [Online]. Available: <https://120years.net/wordpress/oramics-daphne-oram-uk-1959-2/>. [Accessed: 15-Apr-2022].
- [220] “Colour Chaser « YURI SUZUKI.” [Online]. Available: <https://yurisuzuki.com/archive/works/colour-chaser/>. [Accessed: 28-Sep-2021].
- [221] “Playtronica.” [Online]. Available: <https://playtronica.com/>. [Accessed: 28-Sep-2021].
- [222] “Workshops.” [Online]. Available: <https://playtronica.com/workshops>. [Accessed: 28-Sep-2021].

- [223] “Turn colour into music with Orbita.” [Online]. Available: <https://orbita.playtronica.com/>. [Accessed: 28-Sep-2021].
- [224] “Aeolian Harp.” [Online]. Available: <https://aeolusapp.appspot.com/>. [Accessed: 11-Aug-2021].
- [225] “Gravilux App — Scott Snibbe - Interactive Art.” [Online]. Available: <https://www.snibbe.com/gravilux-app>. [Accessed: 28-Sep-2021].
- [226] “Björk: Biophilia on the App Store.” [Online]. Available: <https://apps.apple.com/us/app/björk-biophilia/id434122935>. [Accessed: 28-Sep-2021].
- [227] “MoMA | Biophilia, the First App in MoMA’s Collection.” [Online]. Available: https://www.moma.org/explore/inside_out/2014/06/11/biophilia-the-first-app-in-momas-collection/. [Accessed: 15-Apr-2022].
- [228] “Radiohead app – Universal Everything – Polyfauna.” [Online]. Available: <https://www.universaleverything.com/projects/polyfauna>. [Accessed: 28-Sep-2021].
- [229] “Latest – Universal Everything.” [Online]. Available: <https://www.universaleverything.com/>. [Accessed: 28-Sep-2021].
- [230] “FIGHTMAN on the App Store.” [Online]. Available: <https://apps.apple.com/us/app/fightman/id1558850208>. [Accessed: 28-Sep-2021].
- [231] “LUMI Keys Studio Edition by ROLI - The illuminated MPE controller with per-key pitchbend and polyphonic aftertouch | ROLI.” [Online]. Available: <https://roli.com/products/blocks/lumi-keys-studio-edition>. [Accessed: 28-Sep-2021].
- [232] “EYESY.” [Online]. Available: <https://www.critterandguitari.com/eyesy>. [Accessed: 28-Sep-2021].
- [233] P. Galanter, “What is generative art? Complexity theory as a context for art theory,” *6th Gener. Art Conf.*, p. 21, 2003.
- [234] M. Boden and E. Edmonds, “What is generative art?,” *Digit. Creat.*, vol. 20, pp. 21–46, Mar. 2009.
- [235] “Interactive art – Art Term | Tate.” [Online]. Available: <https://www.tate.org.uk/art/art-terms/i/interactive-art>. [Accessed: 13-Feb-2021].
- [236] “Multi-media – Art Term | Tate.” [Online]. Available: <https://www.tate.org.uk/art/art-terms/m/multi-media>. [Accessed: 13-Feb-2021].
- [237] “Media Art – Media Art.” [Online]. Available: http://www.virtualmuseum.ca/sgc-cms/expositions-exhibitions/arts_mediatiqes-media_arts/art_mediatiqes-media_art-eng.php. [Accessed: 13-Feb-2021].
- [238] L. Manovich, *The language of new media*. MIT press, 2002.
- [239] D. Quaranta, *Beyond New Media Art*. Lulu.com, 2014.
- [240] M. A. Net, “Media Art Net | Paik, Nam June: Exposition of Music – Electronic Television,” Feb. 2021.
- [241] “MELA Foundation Launches Campaign to Save La Monte Young’s New York Dream House - Artforum International.” [Online]. Available: <https://www.artforum.com/news/mela-foundation-launches-campaign-to-save-la-monte-young-s-new-york-dream-house-83568>. [Accessed: 13-Feb-2021].
- [242] “Scott Arford | RECOMBINANT 2020.” [Online]. Available: <http://recombinantfestival.com/2020/project/scottarford/>. [Accessed: 13-Feb-2021].
- [243] “LUX - Granular Synthesis.” [Online]. Available: <https://granularsynthesis.info/LUX>. [Accessed:

- 13-Feb-2021].
- [244] B. W. Joseph, “‘My Mind Split Open’: Andy Warhol’s Exploding Plastic Inevitable,” *Grey room*, pp. 80–107, 2002.
- [245] “VASULKA.ORG home.” [Online]. Available: <http://www.vasulka.org/>. [Accessed: 13-Feb-2021].
- [246] “Brian Eno - Reflection.” [Online]. Available: <https://brian-eno.net/reflection/index.html>. [Accessed: 15-Oct-2020].
- [247] “Arduino - Home.” [Online]. Available: <https://www.arduino.cc/>. [Accessed: 16-Aug-2021].
- [248] “Brian Eno - Works | Paul Stolper.” [Online]. Available: <https://www.paulstolper.com/artists/56-brian-eno/works/73-light-boxes/>. [Accessed: 29-Sep-2021].
- [249] R. Ikeda, K. Abe, M. B. S. de Ibarra, and B. Weil, *Ryoji Ikeda: Datamatics*. Charta, 2012.
- [250] “ryoji ikeda | exhibitions.” [Online]. Available: https://www.ryojiikeda.com/archive/exhibitions/#test_pattern_no3_t2g. [Accessed: 16-Aug-2021].
- [251] “Superblue Presents Artist Carsten Nicolai.” [Online]. Available: <https://www.superblue.com/artists/carsten-nicolai/>. [Accessed: 15-Apr-2022].
- [252] J. Richardson, C. Gorbman, and C. Vernallis, *The Oxford Handbook of new audiovisual aesthetics*. Oxford University Press, 2013.
- [253] “cubev3.com.” [Online]. Available: <https://cubev3.com/>. [Accessed: 29-Sep-2021].
- [254] “flying-lotus - Strangeloop Studios.” [Online]. Available: <https://strangeloop-studios.com/flying-lotus>. [Accessed: 29-Sep-2021].
- [255] “Step Inside Flying Lotus’s Mind-Blowing Performance Sculpture.” [Online]. Available: <https://www.vice.com/en/article/wnpnp/step-inside-flying-lotuss-mind-blowing-performance-sculpture>. [Accessed: 18-Aug-2021].
- [256] “Take a look inside Flying Lotus’ hi-tech performance cube - Fact Magazine.” [Online]. Available: <https://www.factmag.com/2014/11/06/take-a-look-inside-flying-lotus-hi-tech-performance-cube/>. [Accessed: 15-Apr-2022].
- [257] “Creative & Tech - XR, Virtual Production, Motion Design - Xite Labs.” [Online]. Available: <https://xitelabs.com/>. [Accessed: 29-Sep-2021].
- [258] “Thoughts On Designing Data Sculptures – MIT Center for Civic Media.” [Online]. Available: <https://civic.mit.edu/index.html%3Fp=2555.html>. [Accessed: 17-Aug-2021].
- [259] “Quantum Memories - Refik Anadol.” [Online]. Available: <https://refikanadol.com/works/quantummemoires/>. [Accessed: 15-Apr-2022].
- [260] G. Levin, “Painterly Interfaces for Audiovisual Performance,” *MIT Media Lab.*, pp. 1–151, 2000.
- [261] “Final Cut Pro - Apple.” [Online]. Available: <https://www.apple.com/final-cut-pro/>. [Accessed: 14-Aug-2021].
- [262] “Logic Pro - Apple.” [Online]. Available: <https://www.apple.com/logic-pro/>. [Accessed: 19-Sep-2020].
- [263] B. Fry and C. Reas, “Processing. org,” *Process. org*, 2010.
- [264] “openFrameworks.” [Online]. Available: <https://openframeworks.cc/>. [Accessed: 14-Aug-2021].
- [265] “Cycling ’74.” [Online]. Available: <https://cycling74.com/>. [Accessed: 30-May-2020].
- [266] “Pure Data — Pd Community Site.” [Online]. Available: <https://puredata.info/>. [Accessed: 14-Aug-

- 2021].
- [267] “Quartz Composer | Apple Developer Documentation.” [Online]. Available: https://developer.apple.com/documentation/quartz/quartz_composer. [Accessed: 14-Aug-2021].
- [268] “Derivative |.” [Online]. Available: <https://derivative.ca/>. [Accessed: 14-Aug-2021].
- [269] “vvvv - a multipurpose toolkit | vvvv.” [Online]. Available: <https://vvvv.org/>. [Accessed: 14-Aug-2021].
- [270] “Notch - The Real-time Graphics Tool.” [Online]. Available: <https://www.notch.one/>. [Accessed: 14-Aug-2021].
- [271] “Vuo.” [Online]. Available: <https://vuo.org/>. [Accessed: 14-Aug-2021].
- [272] D. A. Dondis, *A primer of visual literacy*. Mit Press, 1973.
- [273] W. S. Yeo and J. Berger, “A framework for designing image sonification methods,” *Proc. 11th Int. Conf. Audit. Disp. (ICAD 2005)*, pp. 323–327, 2005.
- [274] W. S. Yeo and J. Berger, “Application of image sonification methods to music,” *Int. Comput. Music Conf. ICMC 2005*, pp. 1–4, 2005.
- [275] W. S. Yeo and J. Berger, “Raster Scanning-A New Approach to Image Sonification Sound Visualization, Sound Analysis and Synthesis,” *Proc. ICMC 2006*, pp. 309–314, 2006.
- [276] I. Conference and A. Display, “A FLEXIBLE FRAMEWORK FOR REAL-TIME SONIFICATION WITH SONART Woon Seung Yeo , Jonathan Berger , R . Scott Wilson The Center for Computer Research in Music and Acoustics,” *Computer (Long. Beach. Calif.)*, pp. 1–3, 2001.
- [277] W. S. Yeo, J. Berger, and Z. Lee, “SonART: A framework for data sonification, visualization and networked multimedia applications,” *Proc. 2004 Int. Comput. Music Conf.*, pp. 180–184, 2004.
- [278] O. Ben-Tal, J. Berger, B. Cook, M. Daniels, G. Scavone, and P. R. Cook, “SonART: The Sonification Application Research Toolbox,” *Int. Conf. Audit. Disp. 2002*, pp. 1–3, 2002.
- [279] “Music production with Live and Push | Ableton.” [Online]. Available: <https://www.ableton.com/en/>. [Accessed: 19-Sep-2020].
- [280] “RISD/Athenaeum / All Locations.” [Online]. Available: <https://librarycat.risd.edu/record=b1296911~S4>. [Accessed: 15-Apr-2022].
- [281] R. M. McGee, J. Dickinson, and G. Legrady, “Voice of sisyphus: An image sonification multimedia installation,” 2012.
- [282] W. S. Yeo and J. Berger, “Rasterpiece: A cross-modal framework for real-time image sonification, sound synthesis, and multimedia art,” in *International Computer Music Conference*, 2007.
- [283] M. Ciciliani, “Scanline Synthesis As a Sonification Method for Digital Images: Techniques and Aesthetics--a Critical Reflection,” *Proc. Underst. Vis. Music 2015 Symp.*, p. 91, 2015.
- [284] “Formula minus One - MARKO CICILIANI.” [Online]. Available: <http://www.ciciliani.com/formula-minus-one.html>. [Accessed: 08-Aug-2021].
- [285] “Via - MARKO CICILIANI.” [Online]. Available: <http://www.ciciliani.com/via.html>. [Accessed: 08-Aug-2021].
- [286] “_| E -|- -| E.” [Online]. Available: https://cj.lovelyweather.com/_works/sndLines.html. [Accessed: 08-Aug-2021].
- [287] “Scanline Granular Synthesis on Vimeo.” [Online]. Available: <https://vimeo.com/25905946>.

[Accessed: 08-Aug-2021].

- [288] “scanline: computer music from Ryoho Kobayashi on Vimeo.” [Online]. Available: <https://player.vimeo.com/video/143388037>. [Accessed: 08-Aug-2021].
- [289] S. Greenlee, “Graphic Waveshaping,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2013.
- [290] B. Verplank, M. Mathews, and R. Shaw, “Scanned synthesis,” *J. Acoust. Soc. Am.*, vol. 109, no. 5, pp. 2400–2400, 2001.
- [291] R. Boulanger, P. Smaragdis, and J. Ffitch, “Scanned Synthesis: An introduction and demonstration of a new synthesis and signal processing technique,” *Proc. Int. Comput. Music Conf.*, pp. 372–375, 2000.
- [292] J. Bischoff, R. Gold, and J. Horton, “Music for an Interactive Network of Microcomputers,” *Comput. Music J.*, vol. 2, no. 3, p. 24, 1978.
- [293] K. Jo and N. Nagano, “Monalisa: See the Sound, Hear the Image,” *Proc. 8th Int. Conf. NIME*, pp. 6–9, 2008.
- [294] N. Rönnberg and J. Löwgren, “Photone: Exploring Modal Synergy in Photographic Images and Music,” no. *Icad*, pp. 73–79, 2018.
- [295] A. Tanaka, “Biomuse to bondage: Corporeal interaction in performance and exhibition,” in *Intimacy Across Visceral and Digital Performance*, Springer, 2012, pp. 159–169.
- [296] A. Tanaka, “The sound of photographic image,” *AI Soc.*, vol. 27, no. 2, pp. 315–318, 2012.
- [297] G. Marino, M. Serra, J. Raczinski, and G. Marino, “The UPIC System : Origins and Innovations Published by : Perspectives of New Music THE UPIC SYSTEM : ORIGINS AND INNOVATIONS MARIE-HELENE SERRA ,” vol. 31, no. 1, pp. 258–269, 2019.
- [298] C. Capelle, C. Trullemans, P. Arno, and C. Veraart, “A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution,” *IEEE Trans. Biomed. Eng.*, vol. 45, no. 10, pp. 1279–1293, 1998.
- [299] E. Striem-Amit, M. Guendelman, and A. Amedi, “‘Visual’ Acuity of the Congenitally Blind Using Visual-To-Auditory Sensory Substitution,” *PLoS One*, vol. 7, no. 3, pp. 1–6, 2012.
- [300] S. Abboud, S. Hanassy, S. Levy-Tzedek, S. Maidenbaum, and A. Amedi, “EyeMusic: Introducing a ‘visual’ colorful experience for the blind using auditory sensory substitution,” *Restor. Neurol. Neurosci.*, vol. 32, no. 2, pp. 247–257, 2014.
- [301] S. Maidenbaum, S. Abboud, and A. Amedi, “Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation,” *Neurosci. Biobehav. Rev.*, vol. 41, pp. 3–15, 2014.
- [302] M. Auvray, S. Hanne-ton, and J. K. O’Regan, “Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with ‘The vOICE,’” *Perception*, vol. 36, no. 3, pp. 416–430, 2007.
- [303] H. Handa, “SOUND-BASED IMAGE AND POSITON RECOGNITION SYSTEM ~ SIPR E S ~ Shin ’ ichiro Uno , Yasuo Suzuki , Takashi Watanabe Nihon Fukushi University Faculty of Health Sciences Department of Human Care Engineering Miku Matsumoto , Yan Wang Shouei Denshi Kenkyusyo Lt,” no. *Icad*, pp. 67–72, 2018.
- [304] T. Yoshida, K. M. Kitani, S. Belongie, and K. Schlei, “EdgeSonic : Image Feature Sonification for the Visually Impaired Categories and Subject Descriptors,” *Image Rochester NY*, pp. 1–4, 2011.

- [305] K. van de Doel, "Soundview: Sensing color images by kinesthetic audio," 2003.
- [306] K. van den Doel *et al.*, "Geometric shape detection with soundview," 2004.
- [307] Z. Capalbo and B. Glenney, "Hearing color: radical pluralistic realism and SSDs," *Proc. 5th Asia-Pacific Comput. Philos. Conf. (AP-CAP 2009)*, no. Byrne 2001, pp. 135–140, 2009.
- [308] M. Brock and P. O. Kristensson, "Supporting blind navigation using depth sensing and sonification," *UbiComp 2013 Adjun. - Adjun. Publ. 2013 ACM Conf. Ubiquitous Comput.*, pp. 255–258, 2013.
- [309] W. S. Yeo and J. Berger, "A framework for designing image sonification methods," *Proc. 11th Int. Conf. Audit. Disp. (ICAD 2005)*, pp. 323–327, 2005.
- [310] T. Wörtwein, B. Schauerte, K. Müller, and R. Stiefelhagen, "Mobile interactive image sonification for the blind," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9758, pp. 212–219, 2016.
- [311] J. H. Flowers and T. A. Hauer, "Musical versus visual graphs: Cross-modal equivalence in perception of time series data," *Hum. Factors*, vol. 37, no. 3, pp. 553–569, 1995.
- [312] D. L. Mansur, M. M. Blattner, and K. I. Joy, "Sound graphs: A numerical data analysis method for the blind," *J. Med. Syst.*, vol. 9, no. 3, pp. 163–174, 1985.
- [313] A. R. Kennel, "AUDIOGRAF : A DIAGRAM-READER FOR THE BLIND," pp. 51–56, 1996.
- [314] J. Sanchez, "Recognizing shapes and gestures using sound as feedback," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 3063–3067, 2010.
- [315] M. Covarrubias, A. Mansutti, M. Bordegoni, and U. Cugini, "Flexible touch sensor for evaluating geometric properties of virtual shapes through sound: This paper reports a sonification approach to visualise geometric features that are missing in haptic display," *Virtual Phys. Prototyp.*, vol. 10, no. 2, pp. 77–89, 2015.
- [316] Y. Q. Min Nie, Jie Ren, Zhengjun Li, Jinhai Niu, "SoundView: An Auditory Guidance System Based on Environment Understanding for the Visually Impaired People," pp. 7240–7243, 2009.
- [317] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "A stereo image processing system for visually impaired," *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 2, no. 8, pp. 2794–2803, 2008.
- [318] G. Bologna, B. Deville, and T. Pun, "On the use of the auditory pathway to represent image scenes in real-time," *Neurocomputing*, vol. 72, no. 4–6, pp. 839–849, 2009.
- [319] H. Donker, P. Klante, and P. Gorny, "The design of auditory user interfaces for blind users," *ACM Int. Conf. Proceeding Ser.*, vol. 31, pp. 149–155, 2002.
- [320] F. Ribeiro, D. Florencio, P. A. Chou, and Z. Zhang, "Auditory augmented reality: Object sonification for the visually impaired," *2012 IEEE 14th Int. Work. Multimed. Signal Process. MMSP 2012 - Proc.*, pp. 319–324, 2012.
- [321] K. Kahol, J. French, L. Bratton, and S. Panchanathan, "Learning and perceiving colors haptically," *Eighth Int. ACM SIGACCESS Conf. Comput. Access. ASSETS 2006*, vol. 2006, pp. 173–180, 2006.
- [322] G. Bologna, B. Deville, and T. Pun, "Blind navigation along a sinuous path by means of the See ColOr interface," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, 2009, pp. 235–243.
- [323] G. Bologna, B. Deville, T. Pun, and M. Vinckenbosch, "Transforming 3D coloured pixels into musical instrument notes for vision substitution applications," *Eurasip J. Image Video Process.*,

vol. 2007, 2007.

- [324] L. N. Foner, "Artificial synesthesia via sonification: A wearable augmented sensory system," *Mob. Networks Appl.*, vol. 4, no. 1, pp. 75–81, 1999.
- [325] "BBC - Devon Features - New device helps people with colour blindness." [Online]. Available: http://www.bbc.co.uk/devon/news_features/2005/eyeborg.shtml. [Accessed: 21-Sep-2020].
- [326] N. Osmanovic, N. Hrustemovic, and H. R. Myler, "A testbed for auralization of graphic art," *IEEE Reg. 5 - 2003 Annu. Tech. Conf. Conf. Rec. - Pap. Present. 2003 Annu. Meet.*, pp. 45–49, 2003.
- [327] L. Ausiello, E. Cecchetelli, M. Ferri, and N. Caramelli, "Acoustic rendering for color information," *Audio Eng. Soc. - 120th Conv. Spring Prepr. 2006*, vol. 1, pp. 450–454, 2006.
- [328] A. R. Robertson, "The CIE 1976 color-difference formulae," *Color Res. Appl.*, vol. 2, no. 1, pp. 7–11, 1977.
- [329] "The 1976 Revision of the CIE Color System." [Online]. Available: <http://hyperphysics.phy-astr.gsu.edu/hbase/vision/cie1976.html#c1>. [Accessed: 14-Jul-2021].
- [330] E. Deleflie and G. Schiemer, "Images as Spatial Sound Maps," *Proc. Int. Conf. New Interfaces Music. Expr.*, no. Nime, pp. 130–135, 2010.
- [331] N. Rönnerberg, "Musical sonification supports visual discrimination of color intensity," *Behav. Inf. Technol.*, vol. 38, no. 10, pp. 1028–1037, 2019.
- [332] A. Alfaro, Á. Bernabeu, C. Agulló, J. Parra, and E. Fernández, "Hearing colors: an example of brain plasticity," *Front. Syst. Neurosci.*, vol. 9, p. 56, 2015.
- [333] "Antenna implanted in cyborg's skull gets Wi-Fi, color as sound - CNET." [Online]. Available: <https://www.cnet.com/news/cyborg-interview-hear-colors-with-antenna-in-your-skull/>. [Accessed: 02-Oct-2021].
- [334] S. Lakatos, "Recognition of complex auditory-spatial patterns.," *Perception*, vol. 22, no. 3, pp. 363–374, 1993.
- [335] S. Lakatos, "Temporal constraints on apparent motion in auditory space," *Percept. Psychophys.*, vol. 54, no. 2, pp. 139–144, 1993.
- [336] G. von Békésy and W. T. Peake, "Experiments in Hearing," *J. Acoust. Soc. Am.*, vol. 88, no. 6, pp. 2905–2905, 1990.
- [337] D. R. Perrott, "Discrimination Of The Spatial Distribution Of Concurrently Active Sound Sources: Some Experiments With Stereophonic Arrays," *J. Acoust. Soc. Am.*, vol. 76, no. 6, pp. 1704–1712, 1984.
- [338] D. R. Perrott and T. N. Buell, "Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise," *J. Acoust. Soc. Am.*, vol. 72, no. 5, pp. 1413–1417, 1982.
- [339] R. M. Ruff and E. Perret, "Auditory spatial pattern perception aided by visual choices," *Psychol. Res.*, vol. 38, no. 4, pp. 369–377, 1976.
- [340] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.
- [341] D. R. Perrott, "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity," *J. Acoust. Soc. Am.*, vol. 75, no. 4, pp. 1201–1206, 1984.
- [342] R. R. Sinha, H. M. Kamel, and P. Roth, "Graphics and User 's Exploration via Simple Sonics (GUESS): Providing Interrelational Representation of Objects in a Non-visual Environment,"

Seventh Int. Conf. Audit. Disp., pp. 261--266, 2001.

- [343] I. I. Bukvic, D. Gracanin, and F. Quek, "Investigating artistic potential of the dream interface: The aural painting," *Int. Comput. Music Conf. ICMC 2008*, 2008.
- [344] I. I. Bukvic and J. S. Kim, "Perception and interpretation of concurrent aural shapes using dream interface," *Int. Comput. Music Conf. ICMC 2010*, pp. 542–545, 2010.
- [345] J. Grant, "The Fairlight Explained (EMM Aug 1984)," *Electron. Music Mak.*, no. Aug 1984, pp. 27–29, 1984.
- [346] "Journal on the Art of Record Production » Following the Instruments, Designers, and Users: The Case of the Fairlight CMI." [Online]. Available: <https://www.arpjournal.com/asarpwp/following-the-instruments-designers-and-users-the-case-of-the-fairlight-cmi/>. [Accessed: 02-Oct-2021].
- [347] "Fairlight CMI [Retrozone]." [Online]. Available: <https://www.soundonsound.com/reviews/fairlight-cmi-retrozone>. [Accessed: 01-Aug-2021].
- [348] "Herbie Hancock Shows The Sesame Street Kids How A Fairlight CMI Works – Synthtopia." [Online]. Available: <https://www.synthtopia.com/content/2010/07/19/herbie-hancock-shows-the-sesame-street-kids-how-a-fairlight-cmi-works/>. [Accessed: 11-Aug-2021].
- [349] H. Lohner, "the UPIC system: A User's Report," *Comput. Music J.*, vol. 10, no. 4, pp. 42–49, 1986.
- [350] "Aphex At Last..., Future Music Magazine (2006) – Lanner Chronicle." [Online]. Available: <https://lannerchronicle.wordpress.com/2020/08/31/aphex-at-last-future-music-magazine-2006/>. [Accessed: 09-Aug-2021].
- [351] "UPIC – Musica Informatica." [Online]. Available: <http://www.musicainformatica.org/topics/upic.php>. [Accessed: 15-Apr-2022].
- [352] C. Penrose, "Hyperupic." 1992.
- [353] "Hyperupic for OS X - MacMusic." [Online]. Available: <http://www.macmusic.org/news/view.php/lang/en/id/1491/>. [Accessed: 19-Sep-2020].
- [354] "MetaSynth + Xx – U&I Software, LLC." [Online]. Available: <https://uisoftware.com/metasynt/>. [Accessed: 19-Sep-2020].
- [355] "Nicolas Fournel » AudioPaint." [Online]. Available: http://www.nicolasfournel.com/?page_id=125. [Accessed: 27-Mar-2019].
- [356] "Coagula." [Online]. Available: <https://www.abc.se/~re/Coagula/Coagula.html>. [Accessed: 19-Sep-2020].
- [357] "Photosounder.com - Image-sound editor & synthesizer." [Online]. Available: <https://photosounder.com/>. [Accessed: 19-Sep-2020].
- [358] S. Cummings, "Son e (s) t Lumière: Expanding notions of composition, transcription and tangibility through creative sonification of digital images," in *Sound and Image*, Focal Press, 2020, pp. 348–364.
- [359] "SonicPhoto - Convert pictures to sounds!" [Online]. Available: <https://www.skytopia.com/software/sonicphoto/>. [Accessed: 15-Oct-2020].
- [360] "2CAudio - Kaleidoscope | It's A Trip." [Online]. Available: https://www.2caudio.com/products/kaleidoscope#_overview. [Accessed: 15-Oct-2020].
- [361] "AudioSculpt | Ircam Anasynth." [Online]. Available: <http://anasynth.ircam.fr/home/english/software/audiosculpt>. [Accessed: 19-Sep-2020].

- [362] “SoundHack | Happy working for the good of the community.” [Online]. Available: <https://www.soundhack.com/>. [Accessed: 19-Sep-2020].
- [363] “SPEAR Homepage.” [Online]. Available: <http://www.klingbeil.com/spear/>. [Accessed: 19-Sep-2020].
- [364] “Audio recording and editing software | Adobe Audition.” [Online]. Available: https://www.adobe.com/products/audition.html?sdid=KKQPG&mv=search&ef_id=CjwKCAjw2Jb7BRBHEiwAXTR4jWdj9sqdSRr28zZR8YEC8fsAqYGCbwUPqsmS56SfmwAC-O5hTfNDdxoC4CEQAvD_BwE:G:s&s_kwid=AL!3085!3!379243313645!e!!g!!adobeaudition&gclid=CjwKCAjw2Jb7BRBHEiwAXTR4jWdj9sqdSRr28zZR8YEC8fsAqYGCbwUPqsmS56SfmwAC-O5hTfNDdxoC4CEQAvD_BwE. [Accessed: 19-Sep-2020].
- [365] S. Matsumura, “Graph-Sono – Hand Drawing Sound - Graph-Sono system,” pp. 356–357, 2008.
- [366] K. Nakanishi, P. Haimes, T. Baba, and K. Kushiya, “NAKANISYNTH: An Intuitive Freehand Drawing Waveform Synthesiser Application for iOS Devices,” *Proc. Int. Conf. New Interfaces Music. Expr.*, vol. 16, pp. 143–145, 2016.
- [367] “humbleTUNE - shapelynth.” [Online]. Available: <http://www.humbletune.com/shapelynth/>. [Accessed: 08-Aug-2021].
- [368] T. Iwai, “Music insects,” *Perm. Collect. Explor. San Fr.*, 1992.
- [369] “Sound Fantasy (Found Cancelled 1993 SNES Game) | Lost Media Archive | Fandom.” [Online]. Available: [https://lostmediaarchive.fandom.com/wiki/Sound_Fantasy_\(Found_Cancelled_1993_SNES_Game\)](https://lostmediaarchive.fandom.com/wiki/Sound_Fantasy_(Found_Cancelled_1993_SNES_Game)). [Accessed: 09-Aug-2021].
- [370] “Bloom by Brian Eno and Peter Chilvers | GenerativeMusic.com.” [Online]. Available: <http://www.generativemusic.com/bloom.html>. [Accessed: 30-May-2020].
- [371] “Trope by Brian Eno and Peter Chilvers | GenerativeMusic.com.” [Online]. Available: <http://generativemusic.com/trope.html>. [Accessed: 15-Oct-2020].
- [372] “Scape by Brian Eno and Peter Chilvers | GenerativeMusic.com.” [Online]. Available: <http://generativemusic.com/scape.html>. [Accessed: 15-Oct-2020].
- [373] “Air by Sandra O’Neill and Peter Chilvers | GenerativeMusic.com.” [Online]. Available: <http://www.generativemusic.com/air.html>. [Accessed: 30-May-2020].
- [374] “Bloom: 10 Worlds on the App Store.” [Online]. Available: <https://apps.apple.com/us/app/bloom-10-worlds/id1438214016>. [Accessed: 11-Aug-2021].
- [375] “Musical Paint on the App Store.” [Online]. Available: <https://apps.apple.com/us/app/musical-paint/id643923851>. [Accessed: 08-Aug-2021].
- [376] “Patatap.” [Online]. Available: <https://patatap.com/>. [Accessed: 11-Aug-2021].
- [377] “takete ∞ on the App Store.” [Online]. Available: <https://apps.apple.com/us/app/takete/id1012637066>. [Accessed: 11-Aug-2021].
- [378] “Apps — Scott Snibbe - Interactive Art.” [Online]. Available: <https://www.snibbe.com/apps#/bubbleharp-app/>. [Accessed: 11-Aug-2021].
- [379] “Musyc | Fingerlab.” [Online]. Available: <http://www.fingerlab.net/portfolio/musyc>. [Accessed: 11-Aug-2021].
- [380] V. S. Ramachandran and E. M. Hubbard, “Synaesthesia - A window into perception, thought and language,” *J. Conscious. Stud.*, vol. 8, no. 12, pp. 3–34, 2001.

- [381] Ć. Aleksandra *et al.*, “The bouba / kiki effect is robust across cultures and writing systems,” no. Part II, 2021.
- [382] “Affinity – Professional Creative Software.” [Online]. Available: <https://affinity.serif.com/en-us/>. [Accessed: 04-Mar-2022].
- [383] “OSC index.” [Online]. Available: <https://ccrma.stanford.edu/groups/osc/index.html>. [Accessed: 04-Mar-2022].
- [384] “Qualtrics XM // The Leading Experience Management Software.” [Online]. Available: <https://www.qualtrics.com>. [Accessed: 29-Apr-2022].
- [385] B. N. Walker, “Congruency effects with dynamic auditory stimuli,” *Masters Thesis. Rice Univ. Houston, Texas.*, 1997.
- [386] O. Ben-Tal and J. Berger, “Creative aspects of sonification,” *Leonardo*, vol. 37, no. 3, pp. 229–232, 2004.
- [387] E. Thoret, M. Aramaki, and R. Kronland-Martinet, “From shape to sound: sonification of two dimensional curves by reenaction of biological movements,” *9th Int. Symp. Comput. Music Model. Retr.*, no. June, pp. 278–285, 2012.
- [388] M. Chion, *Audio-vision: sound on screen*. Columbia University Press, 2019.
- [389] S.-I. Iwamiya, M. Seki, K. Yoshikawa, and M. Takada, “Congruency between sound effects and switching patterns of images.,” *Japanese J. Ergon.*, vol. 39, pp. 292–299, Jan. 2003.
- [390] M. Cooley, “Sound+ image in computer-based design: learning from sound in the arts,” 1998.
- [391] R. Minghim and A. R. Forrest, “Illustrated analysis of sonification for scientific visualisation,” *Proc. IEEE Vis. Conf.*, pp. 110–117, 1995.
- [392] “New in JMP 16 | Statistical Discovery Software from SAS.” [Online]. Available: https://www.jmp.com/en_us/software/new-release/new-in-jmp.html. [Accessed: 08-Apr-2022].
- [393] “Mira iPad App for Max | Cycling ’74.” [Online]. Available: <https://cycling74.com/products/mira>. [Accessed: 24-Mar-2022].
- [394] W. Joo, “Sonifyd: A graphical approach for sound synthesis and synesthetic visual expression,” 2019.
- [395] “Colormatics – Sonification-driven audiovisual works – CreativeApplications.Net.” [Online]. Available: <https://www.creativeapplications.net/member-submissions/colormatics-sonification-driven-audiovisual-works/>. [Accessed: 09-Apr-2022].
- [396] “cyborg futures - sonifyd.” [Online]. Available: <https://sonifyd.com/cyborg-futures-1>. [Accessed: 09-Apr-2022].
- [397] “spatial visual sound - Woohun Joo.” [Online]. Available: <https://cargocollective.com/joowh/spatial-visual-sound>. [Accessed: 09-Apr-2022].
- [398] “Resonance Audio -.” [Online]. Available: <https://resonance-audio.github.io/resonance-audio/>. [Accessed: 24-Mar-2022].
- [399] “ICST | ICST Ambisonics Site.” [Online]. Available: <https://ambisonics.ch/>. [Accessed: 24-Mar-2022].
- [400] B. N. Walker, “Congruency Effects With.”
- [401] D. Sardana, W. Joo, I. I. Bukvic, and G. Earle, “Perception of spatial data properties in an immersive multi-layered auditory environment,” in *ACM International Conference Proceeding*

Series, 2020.

- [402] I. I. Bukvic, G. Earle, D. Sardana, and W. Joo, “Studies in spatial aural perception: establishing foundations for immersive sonification,” in *International Conference on Auditory Display*, 2019.
- [403] W. Joo, D. Sardana, I. I. Bukvic, and G. Earle, “Introducing locus: a nime for immersive exocentric aural environments,” *NIME 2019*, 2019.

