Animation, Sonification, and Fluid-Time: A *Visual-Audioizer* Prototype

Taylor Olsen
Advanced Computing Center for the Arts and Design
Department of Design
The Ohio State University
Columbus, Ohio 43210
Olsen.190@osu.edu

ABSTRACT
The *visual-audioizer* is a patch created in Max in which the concept of fluid-time animation techniques, in tandem with basic computer vision tracking methods, can be used as a tool to allow the visual time-based media artist to create music. Visual aspects relating to the animator's knowledge of motion, animated loops, and auditory synchronization derived from computer vision tracking methods, allow an immediate connection between the generated audio derived from visuals—becoming a new way to experience and create audio-visual media. A conceptual overview, comparisons of past/current audio-visual contributors, and a summary of the Max patch will be discussed. The novelty of practice-based animation methods in the field of musical expression, considerations of utilizing the *visual-audioizer,* and the future of fluid-time animation techniques as a tool of musical creativity will also be addressed.

Author Keywords
Animation, computer vision, digital musical instruments, Max, sonification, tracking

CCS Concepts
•Applied computing~Arts and humanities~Sound and music computing

1. INTRODUCTION
"Music informs images just as images inform music." [3] But, which one should come first? Within the current age of audio-visual expression, advancements in technological aspects of CPU/GPU architecture allow the elucidation of visuals from digital audio synthesis. What if instead we interpret audio from visuals? The central aim of this paper is to demonstrate the feasibility, prototyping, and usage of a *visual-audioizer*—a Max patch that translates visual information into digital audio and encourages an audience to consider principles of animated motion as an instrument for musical expression.[1] Beyond this, a method of contributing the status of computer vision techniques as a means of artistic expression within time-based media.

In the current age of audio-to-visual stimuli, methods in which quantifiable visual aspects such as shape, size, color, and opacity, can be directly affected by means of digital audio manipulation. As such, we can imagine the process of quantifying audio information and scaling the data to appropriate numbers suited for animated visualizations. For example: a pitch might delegate a position of a shape, the opacity of the shape, or the color of the shape. While musicians and sound engineers have the means to parse out audio information and map them to animated visualizations—what about the animator? The *visual-audioizer* allows an audience to consider the techniques within animated motion as a tool for musical expression. Considering this method, the main thread of information for this paper stems from a few different questions.

1) From an animator's perspective, how do we turn frame-by-frame methodologies into a real-time instrument for musical expression? While sound designers typically face the challenge of attributing musical attributes to visual manifestations, what principles of animated forms could dictate the creation of digital audio? Per the ability of an animator to create nuanced motion and complex forms, similar sounds to those created by sound designers can be replicated and altered based on animation principles when utilizing the *visual-audioizer* interface.

2) How can computer vision aid in the translation of visuals-to-audio within the modern computing era? With the development and innovations of computer vision software, visually quantifiable elements such as position, scale, color, and elongation can be used to the discretion of an animator's ability to generate audio. This question also carries historical implications of previous attempts at eliciting audio from visuals and will be addressed later.

3) What creative effect does real-time user manipulation of data within the translation of visual-to-audio synthesis demonstrate? The ability of modern tracking methods allows the *visual-audioizer* interface to observe and react synchronously; this question can also pose as a space for an audience to learn about the animated form as a method for musical expression. Having the ability and interface to change how the digitized motion is interpreted allows instantaneous feedback and insight into the motion of the animated form; allowing the creator the ability to experiment, edit, record, and create musical pieces with synchronized visuals.

2. CONCEPTS
2.1 Fluid-time animation
Fluid-time, a concept for the purpose of this patch, is where the process of creating animated loops removes the notion of a starting/ending frame. This concept may also rely on creating hand-drawn animated forms as a real-time gestural interaction, rather than a frame-by-frame one. What makes utilizing the *visual-audioizer*, as well as fluid-time animation, unique is the ability of processing visuals in real-time and creating sound, allowing a non-objective animation workflow as a new method

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).
NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

[1] Example of the *visual-audioizer* sonifying animated forms: https://sites.google.com/view/visualaudioizer/home?authuser=2

for musical expression. The concept was eventually experimented utilizing the fluid-time animation software called *Looom* and the *visual-audioizer.*[2] *Looom* employs the use of a vector-based animation engine, a fluid-time frame looping system, the gestural ability from the user and a human interaction device (iPad and Apple pencil for example), and layering techniques to create various/editable animated loops—much like that of an audio loop sequencer.

2.2 Computer vision

The *visual-audioizer* patch partially relies upon computer vision externals within Max from Jean-Marc Pelletier.[3] "Computer vision (CV) is the field of study surrounding how computers see and understand digital images and videos," [1] as defined from DeepAi.org. The purpose of utilizing the techniques of computer vision within the patch is to extract the positional x/y data, as well as the scale, of recognized forms. One of the numerous ways of utilizing CV, especially in this approach, is to convert imagery/video to pure black & white (no grey). These black and white values, when interpreted by CV methods, can be considered as the values 0 and 1 respectively. The CV method then finds groups of either the white (1) or black (0) values based on a determined distance threshold and is considered as an "object" with a centralized position in an x/y coordinate space. Beyond position values derived from the CV system, considerations of scale (size of the form), elongation (how thin a form is), and orientation (degrees of rotation) of a form can be observed and digitized. Considering the use of animation, this makes content creation a straightforward process by animating the form as a white object against a black background. The advantage of knowing how the system interprets data, in relation to the ability of the animator, allows the artistry (and the motion) of the form to create varying degrees of sound.

2.3 Sonification

"Sonification is the use of nonspeech audio to convey information…the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation." [2] Forms, that don't move and are within our field of view, are generally silent. Only when the form collides, or moves, does the compression of air allow the creation of sound. Advantages of considering sonification to make inaudible data have perceptible means include our auditory ability to distinguish pitch, sound localization (position), and loudness (amplitude). Though sonification is generally a means of turning data into something audibly perceptual, much like data-visualization, the use of sonifying visuals within the field of animation is to provide another layer of depth to the animated form. The animated form, much like a still object, does not make sound on its own. Once audio is produced/recorded, we as creators can sync the visuals to the audio, or vice versa.

An example of visual sonification can be observed from Jean-Marc Pelletier in which the horizontal and vertical axis of a video have been mapped to pitch.[4] The video itself is a still shot of a river with a still branch penetrating the surface but remaining generally within the center of the image. As time goes on, small leaves float down the stream, cascading from left to right and creating a "glissando" from a low to high pitch.[5] We as the audience can follow the motion, as well as make the association to which object has created the sound. And though the sound might not be what an audience expects, it's the consideration that notable changes in visual information is what caused the audio to be generated in the first place. Pelletier states [4], "Since there is no single correct way to sonify an image artistically, the choice of the precise type of sound to use is left to the creator." The *visual-audioizer* interface specializes in this consideration—allowing the user to experiment with the sonification of pre-made animated forms, or through a streamed source of live input.

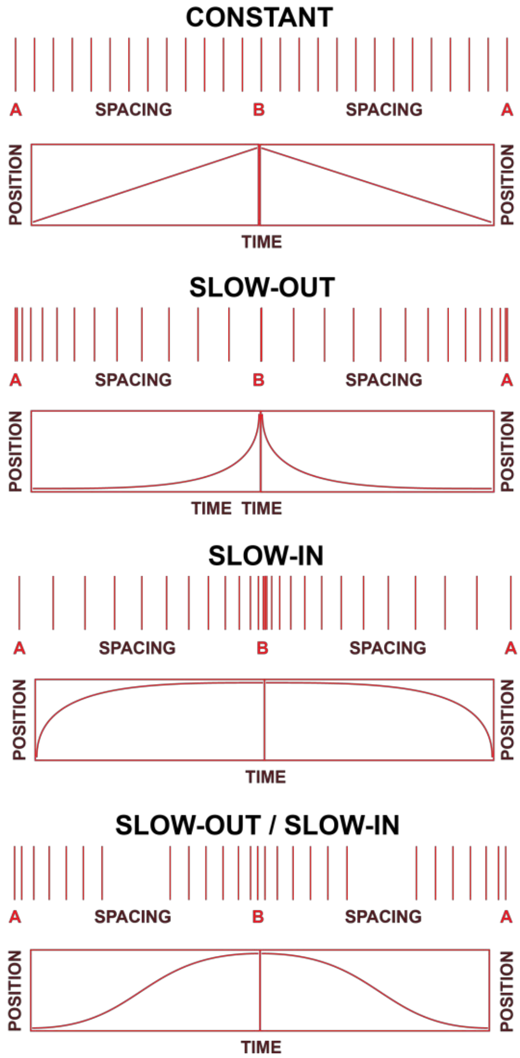

Figure 1: Timing & Spacing

2.4 Timing & spacing

While the main aspect of using this patch as an animator is to allow animated motion to be the sole determinate of digital audio making, the consideration of two basic animation principles make a dramatic difference in the elucidation of audio. As mentioned by Richard Williams, the creative mind behind *Who Framed Roger Rabbit*, animation is all about the timing and spacing of forms [7]. [6] Using the *visual-audioizer,* significant principles in consideration to Williams comment (such as timing & spacing) provide audio-driven insight into the differences in positional coordinates tracked by the patch. With this in mind, we can imagine two points: A and B. If a form was to translate from point A to point B (over a set time period) while the patch simultaneously tracks and outputs audio, the many ways in which the timing & spacing of the animated objects position can be creatively altered allow the animator to create more expressive sound with a the consideration of a glissando sound. An example of different degrees

[2]The app *Looom* from Iorama Studio: https://iorama.studio/
[3]Pelletier's cv.jit library notes: https://jmpelletier.com/cvjit/:
[4]An example of Pelletier's sonification experiments: https://www.youtube.com/watch?v=Z7btudUVT4E
[5]Glissando: a musical slide, either upwards or downwards, from one note to another.
[6]Williams's "Who Framed Roger Rabbit" (1988): https://www.imdb.com/title/tt0096438/

of motion are shown in Figure 1. Beyond frame-by-frame animating the translation of forms, it is up to the animator during a fluid-time performance to complement the *visual-audioizer* with considerations of timing, spacing, framerate, and visual ambiguity to create music.

2.5 Cooperative interaction

The consideration of utilizing the *visual-audioizer* and techniques of the animator as an instrument of musical expression can be supplemented from music psychologist Shinichiro Iwamiya; specifically the interaction between auditory and visual processing. When comparing the complementary aspects of audio and visual cooperative enhancement he proposes a concept called *cooperative interaction* [8], in which "each modality contributes to the evaluation of the other. In audiovisual communication, both modalities work together to make the product more effective." Iwamiya found that working with audio and visual spectrums almost always complemented one another, but noted that if there was a clear association between the causation/timing of the audio and visual stimuli that an enhancement was noticed. For the case of using the *visual-audioizer*, all sound is derived directly from visuals – directly complementing one another on a 1:1 basis. This interplay allows the audio and visuals to be synchronous; providing insight in the experimentation of how animated motion influences the output of digital audio.

3. PRIOR WORKS & INFLUENCES

Most considerations for the use of the *visual-audioizer* comes from multiple sources of audio-visual artists. If we recall from Pelletier about there being no correct way to sonify a visual, there will never be a correct way to visualize a sound. The theoreticians and artists behind the works mentioned henceforth are ones who consider the manifestation of visuals to be complimentary to their works and have influenced the considerations of utilizing fluid-time animation techniques as a means of musical expression.

3.1 Experimental Animation

Paul Wells, an animation theorist, describes the realtionship between experimental animation and music as a relationship of colors and shapes moving and pulsating; rhythmically dynamic in nature and explorative in its form [6]. For the consideration of the Max patch, the *visual-audioizer* allows an animator to influence the emotiveness of the audio output based on the motion of the animated object (among other factors like scale and elongation). While the general populous is more akin to having a story behind an animated film, the purpose of utilizing this patch is to explore beyond the boundaries of animated characters and consider that ways in which the focus is *motion.*

In similar fashion, Wells also states, "It may be said that if orthodox animation is about 'prose', then experimental animation is more 'poetic' and suggestive in its intention.[7] [Experimental] Animation prioritizes abstract forms in motion, liberating the artist to concentrate on the vocabulary he/she is using in *itself* without the imperative of giving it a specific function or meaning." The "vocabulary" that Wells is speaking of on behalf of the artist (animator) includes the principles of animated motion. Taking into consideration the importance of developing an animation skillset for the creation of sound through the *visual-audioizer*, the "poetic" vocabulary that an animator utilizes in their experimental motion studies poses as a means of musical expression when utilized correctly.

3.2 Visual poetry

Oskar Fischinger (1900-1967), a German-American animator, was often opposed to representational imagery. He strayed from the 3-act narrative structure that Disney was dominating at the time (even though he worked for him on a few different films as a cartoon effects animator) and focused on the mental imagery that became an association from the auditory rhythm that music held. The connection that abstracted animation holds in tandem with the rhythm and pitch of a song has an emotional appeal. "Fischinger…was perhaps more than the others committed to preserving film as art, that is to say, in Kandinskyesque terms, as pure form and colour, as a spiritual and emotional experience with the artist as prophet." [7] Considering Fischinger was ahead of his time in the explorative mental imagery that is transposing audio to visual media, his dynamic relationships with music and animated imagery shape and change the way we associate our preliminary viewing and/or listening of the material. His visuals provided another layer of sustainability among past musicians whom already dedicated their life to the shaping of musical expression.

In works such as *An Optical Poem* (1938), Fischinger explored the concept that his works might have been unconditional experiences in and of themselves, much like the music would provide emotional/representational appeal when played alone.[8] The visuals synced up to Franz Liszt's "Hungarian Rhapsody No. 2" with pieces of cut out paper-circles hung delicately by wires.[9] The circles cascade, flow, appear/disappear, and move in fashions that only animated imagery and nuanced motion can capture. With the intentionality of being deliberate in the visual interpretation of the audio, this allows the visuals to establish a deeper connection to the timing, rhythm, and flow of the music. Fischinger's choice colors, shapes, and non-representational imagery attempt to visualize that which only our ears can discern as identifiable – but restrain themselves only to that of what the music has to offer. In a broad sense, he was interested in using the identifiable traits of musical "laws" (rhythm, tone, envelope, harmony, timbre, etc.) as a means of visual expression. His works influenced the prototyping of the *visual-audioizer* by considering the establishment of mental imagery as a direct influence on the audio. It is deliberate, controlled, and explorative; allowing a seamless connection between the desired sounds and the realized motion. The "layer of sustainability" the visuals bring to the audio, as mentioned in the last paragraph, can be considered less heavily – as the *visual-audioizer* provides the animated form with the proposed layer of musical expression.

3.3 Synchronicity of visual-driven audio

Norman McLaren (1914-1987), an influential animator on the National Film Board in Canada for 40+ years, would create sounds in his animations to compliment his visuals. McLaren would "draw" the sounds onto the piece of the film strip itself that coincided with the imagery on the same frame – meaning he would manually put in the marks for a specific frame on the films sound strip.[10] An example of this is visually elongated shapes would sound shrill and high, while large shapes that take up space on the screen would be loud, low, and resonant. McLaren was not using the animated imagery as the sounds that would be made, but rather using this imagery as a basis to what his mind

[7]Paul Wells

[8]Oskar Fischinger's film, "An Optical Poem" (1938): https://archive.org/details/1937OSKARFISCHINGERANOPTICALPOEM

[9]Franz Liszt's song, "Hungarian Rhapsody No. 2" (1847): https://archive.org/details/LisztHungarianRhapsodyNo.2_689

[10]Norman McLaren's film, "Pen Point Percussion" (1951): https://archive.org/details/1951penpointpercussionbynormanmclarennfb1080mp4

interpreted the sound a certain shape would create. Though his technique allowed him complete control over his audio, the amount of time that it would take to both draw the visuals and the sounds themselves was at least doubled.

There was a limitation of not being able to create sounds for each individual shape, but for the temporal moment at hand. His sounds had to achieve an overarching tone if there was many objects on the screen—or would have to prolong themselves to provide more information that a visual could not. Another limitation to consider is that as soon as the mark-making was present, the ability to readdress the pitch or tempo of the audio meant having to re-draw entire sections. Though McLaren could have just recorded different sound strips, this was not in the spirit of the artistry surrounding the simultaneity of the visuals. Most of the time it was just small brush strokes that would create little blips of sound for each frame, but later went on to have long connected strokes on the sound strip for stricter control over the length of a sound. This provided a sense of depth within the animations, allowing the sounds to compliment the change in motion rather than be spontaneous and off-screen.

In an article from Kuihara Utako, an audio-visual theorist, he recalls a quote from McLaren, "What happens between each frame is more important than what happens on each frame. How it moved is more important than what moved". McLaren believed that the still of the frame was more about being a part of whole experience rather than singular. He stressed that animation has an ability to be used as an inquisitive medium rather than a source of reaction; he explored this intention through his film *Rythmetic* (1955).[11] As examined by Utako [4] the said film, "was classified as a film not for "aesthetic pleasure" but for "information and education". I found that McLaren was attempting to express a universal language through his animations—meaning his use of animation and numerical values would be recognized in most (if not all) countries and would provide a shared connection between cultures. Using sounds, visuals and timing, Utako says, "we could be struck with wonder at the well regulated placement of the figures and symbols and numerous calculations on the wall-to-wall screen, in addition to the continuously moving and changing nature of animation…In *Rythmetic*, the orderly enchantment and the ornamental one are organized from confrontation to integration". McLaren knew his visuals and audio would not work without one-another. If someone were to solely watch the visuals of *Rythmetic*, temporal sense of timing and auditory expression would be lost. The counter-part, however, shows that while the temporal aspects of audio are notated with greater esteem, the visuals bolstered and gave meaning to the sounds attempting to mimic the motion. Utako closes his observations with, "*Rythmetic* is an invitation to the aesthetic confusion of order and disorder by figures and symbols; and besides, it is a temporally designed artwork both in the visual and auditory aspects, more so than an educational film or arithmetic lecture".[12]

3.4 Audio-visual cooperative enhancement

More recently, Christian Ludwig (aka Jerobeam Fenderson) has been exploring this medium of image to sound. He uses an oscilloscope and Max (as well as sequencer program Ableton live) to "draw" shapes using sound.[13] What Ludwig does is take sin/cos values of sound and manipulates them out until they're only recognizable as a line on an oscilloscope screen. This line is then used to bend, stretch, and replicate the visuals presented. These manipulations can become imagery based on the frequency, amplitude, and envelope – anything that can be audibly numericized. After using mathematical expressions to create the audio, Ludwig explores the ability of translating and projecting the imagery onto 3D models. In this case, he has managed a way to display sound as direct creative means to a visual manifestation. The visual aesthetic of these auditory explorations is technical and on-beat; the effect is satisfying to watch, knowing the visual representations and audio are directly in-sync with one another. Though we as an audience can find the differences in sounds (similar shapes seem to create different pitches) throughout Ludwig's pieces, it is meant more as a complimentary aspect to the electronic sounds created.

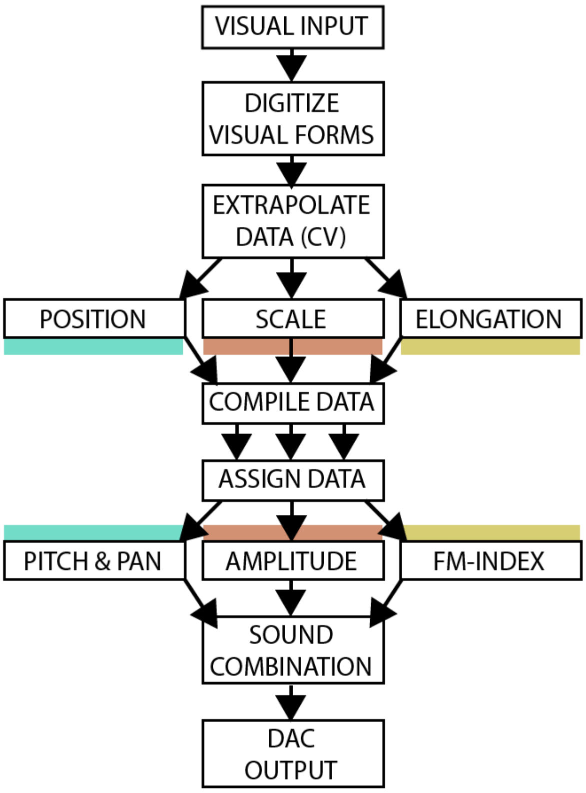

Figure 2: Visual input to DAC output flowchart

4. INTERFACE AND APPLICATIONS

4.1 Interface, mapping, and sound

The interface for using the *visual-audioizer* is within a max patch. For the tracking of the visuals the CV.jit objects from Pelletier are utilized on a two-dimensional axis. Within the interface there are multiple aspects of control for the audio output. The patch gives the user the ability (in real-time) to read fluid-time animated forms from a streamed source, or pre-animated files; allowing the user and an audience to witness the tracking of the animated forms coupled with digital audio output. The ability to speed-up/slow-down the visuals is available both in the *Looom* application, and within the *visual-audioizer* patch via pre-rendered files. Within the pre-animated sequences, considerations of timing & spacing, (as well as squash and stretch—a supplementary animation principle) are represented. There is also the ability to switch between MIDI and Msp sounds – allowing the user to consider the implementation of digitized instruments vs. that of generated audio signals. The Msp objects such as cycle~, rect~, saw~, and tri~ are utilized and easily interchangeable via the interface.

[11]Norman McLaren's film, "Rhythmetic" (1956): https://www.nfb.ca/film/rythmetic/

[12]Kurihara Utako

[13]Christian Ludwig's example of Oscilloscope drawing (2014): https://www.youtube.com/watch?v=rtR63-ecUNo

The ability to alter the pitch is also within the interface, allowing the user to adjust the frequency range of the audio.

The method to interpret the visual data into the various options of visual information include utilizing the position, scale, and elongation of animated forms. Pitch is mapped to the position of the object, the scale is to amplitude (the percentage of screen space the form takes up is proportional to amplitude), and the elongation of the form is mapped to the frequency modulation index of the audio output. The position of the form also dictates how the panning of the pitch is interpreted; meaning forms on the L/R side of the screen directly correspond to the output of audio in a L/R speaker setup. The flow from the visual input, to digitized values, to audio output is represented in Figure 2.

Pitch-mapping of the visuals include various options of reinterpreting the translation of the x/y data into digital audio signals via the interface. From low to high pitch, the various methods of data mapping to pitch within the system can be seen in Figure 3. The methods of reading the visuals via the CV.jit objects include beginning the analysis from the top-left to bottom-right of the screen space. These include x-axis (left to right), y-axis (bottom to top), x- & y- (top-left to bottom-right), x+ & y+ (bottom-right to top-left), x-split (lower pitch in the center, higher on the left/right edges), y-split (lower pitch in center, higher on bottom/top edges), center-out (lower pitch in center, higher in corners), and center-in (higher pitch in center, lower in corners). Positional coordinates for panning remain constant; to reiterate, objects on the L/R side of the screen space dynamically adjust and correspond to the L/R speaker output.

5. DISCUSSION AND FUTURE PROJECTS

For the moment, it is important to remember that the prime method of correctly utilizing the *visual-audioizer* is to animate with black and white forms; future uses of color tracking will be utilized, but the current status of the patch works best with this intentionality. While there is no correct way to map and record the audio from the patch, the considerations of animated motion and the ability of the animator to control these motions in pre-rendered visuals, as well as fluid-time performance, becomes an overarching discussion in the use of CV methods to aid in the sonification of animation. I consider the *visual-audioizer* as a working proof-of-concept; meaning there is room in the future for considerations of color, pictorial ambiguity, and deep-learning techniques to allow visual elements within complex settings to be sonified.[14]

5.1 Multiple forms

A caveat for the animator when using the *visual-audioizer* is to consider the use of multiple forms. The patch parses out the multiple objects within the scene and sends each piece of information to a poly~ object. A max of 255 forms can be recognized. The *Looom* app specializes in the ability to create editable framerates as separate layers; this is great considering how complex imagery and patterns could be digitized and sent to multiple outputs; allowing the simultaneity of polyrhythms to be sonified and interpreted henceforth.[15] Depending on the amount of recognizable forms, an uzi object sends out the grouped data to the poly~ object. And though the *visual-audioizer* can discern multiple objects, it can also be the cause of straining the patch and not being as synchronous as hoped. When playtesting with the number of objects in a scene, I noticed when the number of discernable forms jumped back and forth from 1 to 100, the patch had a hard time interpreting the data quickly enough for the poly~ audio to be simultaneous with the visuals. This could be solved by having the data interpreted, digitized, recorded, and played back on its own—but ultimately removes the notion of the real-time feedback produced by CV methods.

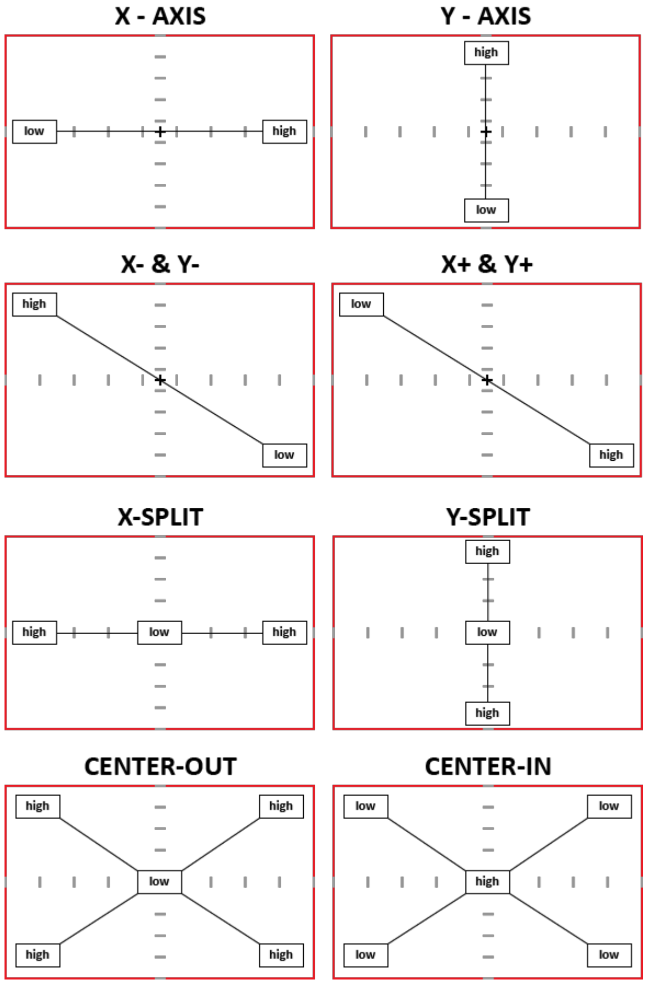

Figure 3: Pitch mapping examples

5.2 Form Proximity

Another consideration for the animator when creating content for the *visual-audioizer* is the proximity of forms. If some forms were too close to one another, the CV method groups them and creates a unified sound, rather than separate pieces of audio. Though this can be solved by applying a threshold to the proximity-grouping method in the CV patch, it often became a headache of tweaking the settings until a 'perfect' scenario was produced. Being flexible with the animated form is a goal of utilizing this patch and knowing how to work within the limitations of the CV methods will yield more results for the creative who enjoys experimentation.

5.3 Framerate and codec

Another caveat for the animator to consider is the framerate of the video input. Considering the range of human hearing can lie anywhere from 20 – 20,000Hz, the human eye is only able to identify framerates anywhere from 1 – 150 frames-per-second (fps), or 150Hz; the standard framerate that an animator works with is 24fps (a visual 24Hz). This in turn can cause some limitations in the process of providing enough visual "depth" within the motion of the animated form to create a desired envelope with audible nuances. In relation to this visual "depth",

[14] Neil Harbisson, an artist who sonifies light/color in real-time: https://www.nationalgeographic.com/news/2017/04/worlds-first-cyborg-human-evolution-science/

[15] Polyrhythm: combining comparative/contrasting rhythms in a musical composition.

specific video codecs should be utilized for the flow of visual information, data extrapolation, and audio output to be as synchronous as possible. Visually dense codecs that utilize "lossless" quality will slow the system, while codecs such as HAP will substantially reduce the amount of CPU usage and allow the patch to produce real-time audio with more synchronicity to the visual input. When animating in fluid-time, the *Looom* app can vary the speed of these framerates on the fly, allowing the framerate to be edited and varied for the consideration of polyrhythms. To retain visual complexity and audible synchronization, the user should lower the resolution of the streamed content into the *visual-audioizer* and stream the full resolution fluid-time visual performance to a separate source.

5.4 Scale and envelope

When designing the scale of the animated form using the *visual-audioizer*, it is intriguing to note that the rate of scale increase can draw parallels to that of synthesized envelopes. For example, if the form dramatically increases in scale over a short period of time this is comparable to the "attack" of the sound. Similarly, if the animator were to diminish the scale of the form over a shorter/longer period, or fall to zero, this is the "decay", "sustain", and/or release of the sound.[16] Experimenting with increase/decrease in scale can lead to dramatically different results in the ADSR spectrum; and coupled with the ability of the animator to control scale over time, this can lead to complex techniques with holds and atypical modulations in the envelope of a sound. This allows the animator to consider the size of their forms as an audible dynamic range. A limitation of envelope "attack" is to consider the previous paragraph speaking about the framerate of an animation; while the animator can control the scale of the form, the shortest amount of "attack" is 1/24th of a second (41.66ms) if the animator is following that of standard animation practices.

6. FUTURE WORK

The potential of the *visual-audioizer* will be utilized in an ongoing MFA thesis project with sounds solely created from the patch. Along with the recorded audio, the visuals that produced such will be synced and manipulated into an audio-visual experience. Using the *Looom* app, a remote-desktop application (Splashtop), and a visual frame-sharing system (Syphon), the *visual-audioizer* will be also utilized during a live audio-visual performance.[17, 18] The patch itself, as a standalone piece, will allow an audience and artist to explore the advantages of the animated form to elicit audio, and will provide a space for reflecting upon the benefits of CV methods in musical expression.

7. CONCLUSION

The *visual-audioizer* points towards the future and adds to the conversation of utilizing visual information as a tool for musical creation. The concept of fluid-time animation and the *visual-audioizer* together provides a new experience of interfacing with digital music creation. Though there is a myriad of ways to consider visual complexity and CV methods to create music, utilizing animation principles provides a clear direction in the creative process for applying animated motion and the *visual-audioizer* as a tool for musical expression. I hope for a surge of CV technologies to be incorporated into the creative process of sonifying visual information to its full extent. For example, there is always room for improvement in the methods of reading visual data; including RGBA values, hue/saturation/opacity, and 3D imagery. Coupling this with the ability of the *Looom* app, colors could then be mapped as different ways of altering digital audio derived from animated motion. For the creatives who are more visually inclined, and who appreciate time-based mediums such as animation, the *visual-audioizer* is a way to push CV methods further into the hands of those willing to create music utilizing the advantages of the animated form and fluid-time techniques. With the rise of new technology and tracking methods, I propose we as both artist and audience will see a growth of fundamental CV aspects be incorporated into the creative process of musical expression—and should welcome these technological/conceptual advancements not as a limitation in the field of musical expression, but as ever-expansive insights into coexisting with modern/future creative techniques. I hope for others to use a finalized version *visual-audioizer* to explore the layers of depth the animated form can hold within the spectrum of audio creation, and welcome those who haven't considered CV methods to elicit audio and create music.

8. ACKNOWLEDGMENTS

I would like to thank my wife, Taylor, for anything and everything you do. Also, my family, advisors, and friends among my place of research which feels like a second home, ACCAD.

9. REFERENCES

[1] Anon. 2019. Computer Vision. (May 2019). Retrieved August 21, 2019 from https://deepai.org/machine-learning-glossary-and-terms/computer-vision

[2] Gregory Kramer, Bruce Walker, and Robin Bargar. 1999. *Sonification report: status of the field and research agenda*, Faculty Publications, Department of Psychology, University of Nebraska, Lincoln, Nebraska: International Community for Auditory Display.

[3] Jean Detheux. 2013. Neither Fischinger nor McLaren, Visual Music in a different key. Retrieved September 9,·2018 from https://blog.animationstudies.org/?p=346

[4] Jean-Marc Pelletier. 2008. Perceptually Motivated Video Sonification. (2020). Retrieved March 11, 2019 from hrrps://jmpelletier.com/sonification-demos/

[5] Kurihara Utako. 2011. Norman McLaren's Animated Film Rythmetic as Temporal Art. *Aesthetics* 15 (2011), 116–124.

[6] Paul Wells. 2015. *Understanding animation*, London: Routledge.

[7] Richard Williams. 2012. *The Animators Survival Kit,* London: Faber and Faber.

[8] Shinichiro Iwamiya. Interactions between auditory and visual processing when listening to music in an audiovisual context: 1. Matching 2. Audio quality. *Psychomusicology : A Journal of Research in Music Cognition* 13, 1-2 (March 1995), 143, DOI:http://dx.doi.org/10.1037/h0094098

[9] Tate. Where abstraction and comics collide: Oskar Fischinger – Tate Etc. Retrieved Augutst 18, 2019 from https://www.tate.org.uk/tate-etc/issue-7-summer-2006/where-abstraction-and-comics-collide

[16] In-depth explanation for ADSR envelopes: https://blog.landr.com/adsr-envelopes-infographic/

[17] Splashtop remote-desktop example: https://www.splashtop.com/demo-videos

[18] Syphon visual streaming information: http://syphon.v002.info/