

TRACES OF MODAL SYNERGY: STUDYING INTERACTIVE MUSICAL SONIFICATION OF IMAGES IN GENERAL-AUDIENCE USE

Niklas Rönnerberg

Linköping University
Media and Information Technology
SE-581 83 Linköping, Sweden
niklas.ronnerberg@liu.se

Jonas Löwgren

Linköping University
Media and Information Technology
SE-581 83 Linköping, Sweden
jonas.lowgren@liu.se

ABSTRACT

Photone is an interactive installation combining color images with musical sonification. The musical expression is generated based on the syntactic (as opposed to semantic) features of an image as it is explored by the user's pointing device, intending to catalyze a holistic user experience we refer to as modal synergy where visual and auditory modalities multiply rather than add. We collected and analyzed two months' worth of data from visitors' interactions with Photone in a public exhibition at a science center. Our results show that a small proportion of visitors engaged in sustained interaction with Photone, as indicated by session times. Among the most deeply engaged visitors, a majority of the interaction was devoted to visually salient objects, i.e., semantic features of the images. However, the data also contains instances of interactive behavior that are best explained by exploration of the syntactic features of an image, and thus may suggest the emergence of modal synergy.

1. INTRODUCTION

Photone is an interactive installation combining photographic images and musical sonification [1]. In Photone, an image is displayed and a dynamically changing musical score is generated based on the overall color properties of the image and the color value of the pixel under the touch-point on the touch-screen. Consequently, the music changes as the user moves the finger exploring the image and simultaneously using the image to explore the music.

When we developed the first version of Photone, we found the quality of *modal synergy* to be potentially relevant when designing multimodal interaction, such as interactive sonification. Modal synergy refers to how two or more modalities fuse in interaction to create a user experience that goes beyond the simple sum of the parts, forming expressions that are not easily predictable, and thus stimulates engagement driven by ludic motivation and the curiosity of exploration. Our work is specifically oriented towards the visual and auditory modalities in the forms of images and musical sonification, even though we feel that the concept of modal synergy might be generative also in designing for other multimodal combinations.



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

In our previous work, the extent to which modal synergy actually manifests itself in the general use of Photone, and the occurrence and nature of actual explorative interaction, were left as more or less open questions. Our purpose here is to start addressing these questions by *studying actual use of Photone by a general audience in the context of a public exhibition*. We hope that this represents a worthwhile contribution to the sonification research community by providing design ideas on the interactive generation of musical elements based on syntactic image elements, as well as insights into the experiential qualities of multimodal interaction in which sonification plays a constituent role.

2. USING PHOTONE

To make the arguments and discussions presented in this paper easier to understand, let us try to convey a sense of the synergistic interaction experience we are talking about. A short video demonstration of Photone to complement the following vignette can be found here: <https://vimeo.com/322740494>.



Figure 1: One of the images used in Photone. Image courtesy Norrköping Visualization center C.

Consider the image from Neonland Experience (Figure 1), entering the image with the finger on the touch-screen in the top right corner. The music is a bit muffled, quite low in its amplitude and with the high frequencies attenuated. The impression is that it is dark, both the music and the image, and the music is experienced as somewhat anticipatory with the ominous red sky and futuristic

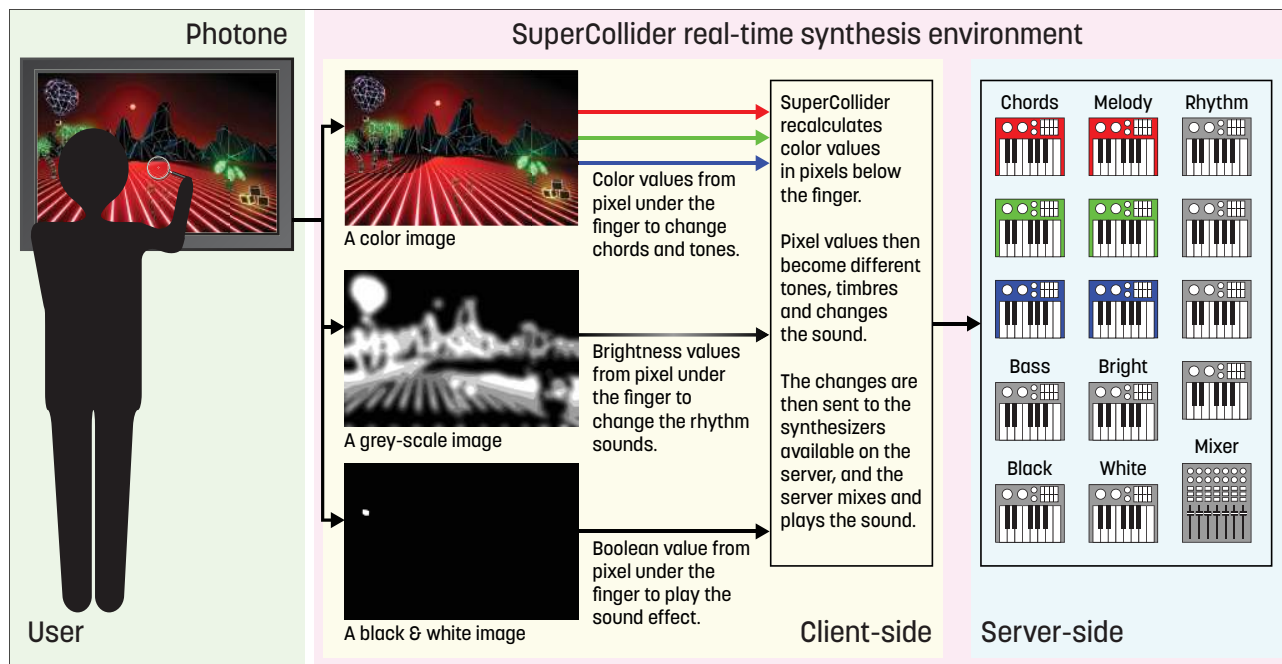


Figure 2: Schematic of the structure in Photone showing the user, the client-side, and the server-side. The interactive sonification is created in SuperCollider by three images, and the color values in these images are mapped to different musical parameters. The server runs the synth definitions for the harmonic ambience (Chords), the melodic components (Melody), the low bass tones (Bass), the high light intensity chord (Bright), the bell-like sound (White), the low frequency sweep (Black), the rhythmic instruments (Rhythm), and the output mixer with reverberation and low-pass filter (Mixer).

digital landscape. As the finger moves towards the brighter area close to the mountain range, the amplitude of the music increases and rising melody keeps the explorer company. If you go back towards the darker area, the melody changes direction and goes downwards in pitch while the background harmonics become more muffled again. When you come close to the mountains, a rhythmic beat accompanies the melody and the harmonic background. As you enter the mountain area most of the musical elements die away, apart from the rhythmic instruments that play along emphasizing the contrasts, the white lines on the black background of the mountains. These bright lines in the mountains now feel like the keys on a piano keyboard, creating outbursts of high pitch tones when you cross over each one of them. As you continue down from the mountains and enter the area of the bright green plant, new harmonic content comes to the fore with new melodic tones that mix with the harmony and tones from the red. The bright lines in the mostly red gradient to the left of the green plant also act like piano keys, playing rising and falling melodic movements as the red color changes from darker red in the middle between the bright lines to almost white in the middle of each line.

With this example we hope to give a passing acquaintance with the interaction and user experience in Photone. We believe that it is rather fruitless to discuss whether one modality augments or supplements the other in Photone. The interaction with image and music has holistic qualities that combine into what we call modal synergy, creating an experience that is larger than its individual components. The example is also meant to show that the image is considered a collection of pixels with specific color values, and that the temporal trajectories in the music are formed by spatial

movement across the surface of the image.

3. DESIGN OF PHOTONE

In Photone, color values in the image at the specific pixel under the touch-point on the touch-screen are read and mapped to different musical elements in the sonification.

3.1. Musical elements

The composition in Photone consists of seven musical elements (see Figure 2). These elements are 1) the overall harmonic ambience, 2) melodic components, 3) two low bass tones, 4) a high light intensity chord, 5) a bell-like sound for pure white, 6) a low frequency sweep for pure black, and 7) rhythmic instruments to highlight areas in the image that are rich in contrasts. The synthesis method used for the musical elements 1 to 6 are described in [1], and the following text will describe the composition of Photone used in the current study.

Similar to the previous version of Photone, the overall harmonic ambience and the melodic components are composed with the three color channels of red, green, and blue (RGB) in mind. The harmonic ambience has been slightly simplified compared to the previous version, and now consists of two-tone intervals multiplied over five octaves, creating a harmonic ambience with ten tones for each color channel. Depending on the pixel value (i.e. the color) at the touch-point under the finger the harmonic ambience varies from a two-tone interval (when information in only

one-color channel is present) to a complex chord (when information in all three color channels are present). The light intensity of the individual color channel determines the amplitude of the components in the harmonic ambience, reflecting how the perception of loudness is closely linked to the perception of brightness [2]. The light intensity is also mapped to the cut-off frequency of a second order band-pass filter between 100 and 4000 Hz for each channel. This makes the harmonic ambience louder and with more high frequency content in bright areas in the image compared to darker areas.

The number of tones for the melodic components is in the current version increased to ten tones for each color channel which are played one tone at a time. The intensity level in each color channel is divided into ten steps and one of the tones is used accordingly. This creates an upwards going melodic movement when intensity in that specific color channel increases, and a downwards going melody when intensity decreases. There is an association between pitch of tones and colors where, for example, higher pitched tones are associated with lighter and brighter colors (see for example [3, 4, 5]). Similar to the harmonic ambience, the melodic components also vary in amplitude and in band-pass filter cut-off frequency according to the intensity level in the pixel value at the touch-point under the finger.

As in the previous version of Photone the two low bass tones are only present when the overall intensity level is low, to emphasize the impression of darker colors. The high light intensity chord is composed with three tones and is only present when the overall light level is high to create an airy and high-intensity feeling. The short bell-like sound is used to further accentuate the dazzling intensity of white, and the downwards sweeping low frequency sound is used to emphasize the change in intensity from different shades of color to darkness.

3.2. Rhythmic instruments

In the current version of Photone rhythmic instruments are added. These rhythmic instruments are synthesized to mimic congas, triangle, and hi-hat sounds, and are used to rhythmically emphasize the amount of contrasts in the images. A script in Matlab divides each image used in Photone into 8 levels of contrast (see Figure 2), where contrast level 0 has no rhythmic instruments but higher levels of contrasts are sonified with increased level of rhythmic sounds. The levels of contrast are chosen as discrete values, so there are clearly defined levels of rhythmic components (see Table 1).

Table 1: The amplification levels of the four rhythmic instruments in the eight levels of contrast.

	0	1	2	3	4	5	6	7
Instrument 1	0	0.5	1	1	1	1	1	1
Instrument 2	0	0	0	0.5	1	1	1	1
Instrument 3	0	0	0	0	0	0.5	1	1
Instrument 4	0	0	0	0	0	0	0.5	1

3.3. Overall image color

Similar to the previous version of Photone each image is determined to have an overall color that affects the composition and musical expression. However, in the current version of Photone the color of an image is determined by a human rather than by

weighted means in the RGB color channels. As the RGB color model is not well adapted to the human color perception [7], this method of choosing the overall color in an image should better correspond to a users impression of an image. The colors available to choose from are yellow, orange, red, white, purple, green, and blue (see Table 2). Based on psychology of colors [6] the composition, as well as the synthesis of the sounds, are adapted to better fit, not to mimic but rather to complement, the impression of the overall color in the image. The reason for adjusting the composition according to the overall color is to attract the user to continued exploration of Photone and to vary the musical expression between images.

A number of musical elements are adjusted according to the selected overall color (see Table 2). The harmony of the harmonic ambience is changed due to the color, where major chords are used for the warmer colors while minor chords are used for the colder colors. Colors with more positive impressions are thus accompanied by chords in major, which in turn might be experienced as more positive [8]. Furthermore, the complexity of the chords is chosen to correspond, at least to some degree, to the energy and the complexity in the colors. The complexity of the chords is connected to the experience of the musical sounds, as a more complex harmonic sound is more captivating for a listener compared to a simpler harmonic sound [9]. However, it is important to keep in mind that the impression of the musical sonification is created by the combination of musical elements, and not by the selection of chord alone.

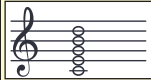



























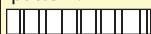
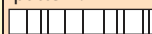
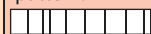
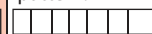
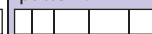
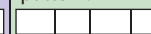

In Photone the dissonance of each tone, i.e. the spread in frequency of the pitches creating each tone, used in the harmonic ambience varies in relation to the impression of the colors. Colors with more energy have a greater dissonance, creating tones with more energy, while colors with less energy have more unison and relaxing tones. The timbre of the harmonic ambience and the melodic components is changed by altering the pulse-width of the square wave forms, where a deviation from 50% pulse-width creates more harmonics compared to 50%. Colors associated as more positive with more energy have pulse-widths creating more harmonics. A softer timbre is experienced as more negative compared to a brighter timbre [10]. By changing the pulse-width the sound changes from rich and prominent sound at 80% pulse-width to a simpler and more hollow sound at 50% pulse-width. The musical sonification is output through a low-pass filter and the cut-off frequency is adjusted according to the overall color, where the sonification for the more positive colors has more high frequency content while the less positive colors have their high frequencies attenuated. The tempo of the rhythmic composition is also generally faster and the rhythm is more complex for the colors with more energy. These musical elements together create a difference in the musical expression between the different colors.

4. IMPLEMENTATION

Photone (see Figure 2) is implemented in SuperCollider 3.10, which is a real-time audio synthesis programming environment [11, 12].

At the request of the science center where Photone is part of the public exhibition, an Easter egg is implemented in each image. This Easter egg is a sound sample that is played back, and mixed with the musical sonification, if the user happens to explore a certain small area of the image (see Figure 2). These sound samples are sound effects that are connected to motif of the image, for

Table 2: Each image used in Photone is determined to have an overall color that affects the composition and musical expression. The table shows the colors used in the current version of Photone, the impression of the color according to Cleary [6], and the musical elements affected of these colors.

	Yellow	Orange	Red	White	Purple	Green	Blue
Color impression	Happiness & vitality	Energy & joyfulness	Passion & intensity	Purity & cleanliness	Creativity & loveliness	Nature & serenity	Calm & sadness
Harmony	Major, ninth 	Major, minor seventh 	Major 	Suspended 2nd & 4th 	Major, major seventh 	Minor 	Minor, minor seventh 
Dissonance	+/- 30 cents	+/- 22.5 cents	+/- 15 cents	+/- 1 cent	+/- 10 cents	+/- 10 cents	+/- 2.5 cents
Timbre (PW)	80% 	70% 	65% 	50% 	60% 	60% 	55% 
LPF cutoff	14kHz 	12kHz 	10kHz 	8kHz 	6kHz 	4kHz 	3kHz 
Tempo (BPM)	98 	96 	94 	96 	94 	92 	90 
Rhythm	Most complex and dense rhythmic pattern. 	Slightly less complex and dense rhythmic pattern. 	Less complex and dense rhythmic pattern. 	Less complex and dense rhythmic pattern. 	Not that complex and dense rhythmic pattern. 	Not complex and sparse rhythmic pattern. 	Least complex and dense rhythmic pattern. 

example the imaginary sound of a science fiction hot air balloon (see Figure 1). The science center argued that searching for these Easter eggs would engage and motivate the user to further explore Photone.

The previous version of Photone was explored with the mouse cursor, however the present version is implemented in a touch-screen environment. To avoid covering the area of interest with the finger, an overlay of an image showing a magnifying glass is implemented with a position offset compared to the actual touch-point (see Figure 2 and Figure 3). The image of the magnifying glass is chosen since the normal use of a magnifying glass is to look into the glass while holding the magnifying glass with an offset to the area of interest.

5. SONIFICATION, MUSIC, AND IMAGES

Even if we claim that the scope of Photone is something new, the combination of images and music is nothing new in itself. The history is full of interesting examples of composed music for images and motion pictures. In film, music is used to help the audience to realize the meaning of the film, to guide the audience to understand the films dramatic and emotional value [13]. Music is also used to create a convincing atmosphere of time and place [14]. The list of the roles of music as a creator of emotions and continuity can be made very long [15], but the use of musical elements in film music differs from Photone. In most cases where music is used as a complement to an image, the music is composed to images based on their denotative and connotative meaning. For example, a sad im-

age is mirrored by sad sounding music, maybe with minor chords and a slow tempo, while a happy image might be reinforced by positive rising melodies, major chords, and a more forward-going tempo. The aesthetics of the music and of the intended meaning is then a psychological analysis, an interpretation and explanation of the experience of the music for the composer as well as the listener [16]. Our artistic intention in Photone is another: By building the sonification upon pixel values of hue and brightness, that is, syntactic rather than semantic properties of an image, we aim to cut through conventional ways of seeing to a more foundational level. Instead of adding sad music to a sad image, Photone elaborates the small elements that create an image, where dark areas in the image are more attenuated and have a more dull timbre compared to brighter areas, where a gradient creates rising and falling melodies, where clear and sharp contrasts between different hues create rhythmic patterns, and where the finger on the touch-screen becomes the conductor’s baton.

As motion pictures, film music, and technology evolved, artists discovered the new tool of optical sound for music production and aesthetic sound synthesis [17]. The artists manipulated the image input to the photocell of the image-to-sound converter [18], unlike film music using the image in syntactic rather than denotative or connotative ways. However, these early artistic explorations and expressions lacked the interaction that unites the visual and auditory modalities. With the development of computers and the use of these as a dynamic medium the artistic emphasis shifted towards the interactive experience of audiovisuality [17]. In an aesthetic experience, in a fully interactive installation, both sound

and image are the means through which the user interacts, and the products of interaction [19]. In Photone the visual expression on the display is not affected by the interaction, and thus the outcome of the interaction, the modal synergy, is formed by nonlinear exploration of the static image and the variable musical sonification. Superficially, the concept of Photone might be described as musicalization of a visual image, as the interaction incorporates musical aspects into the image making it possible to listen to the music of the colors [20]. However, in Photone image, sound, and interaction are tightly integrated, and we argue that the three elements are aspects of the same emergent experience in use [1].

Tanaka [21] converted photographic images to sound. The idea was to create a musical work that replaced the image evoking the same emotional response as the image. This was done by converting the image data, scanning pixels in the image, to sound in different ways, for example brighter values of gray became sound samples of higher amplitudes. But, even if this approach was similar to Photone in that it sonified pixel values rather than the motif in the image, it was not interactive. For sonification to be useful for data exploration, and we would like to argue that sonification of image elements to some extent could be seen as data exploration, dynamic human interaction is necessary [22, 23]. There have been different approaches towards interactive sonification for images apart from Photone (see some examples, discussions, and variations in [24, 25, 26, 27]). O’Neill and Ng [28] presented an interactive sonification to provide feedback in exploration of structure in images. In the user testing O’Neill and Ng found that the sonification supported the foundation of a higher level of understanding of the image structure. Similarly to Photone, O’Neill and Ng used different parameters in the sonification, for example modulation of timbre and pitch. In Photone the sonification is adjusted to every individual pixel in the image, while O’Neill and Ng rather used sonification to differentiate between segments in the image. Heath and Gordon [29] suggested a “primitive” sonification of an image, where an input audio signal was transformed according to statistical interpretation of the average intensity levels in the RGB color channels of an image, in regard to semitone, scale, and chord. This sonification approach has similarities to Photone, even though the proposed sonification was not reported with a user study exploring the user experience or the sonification. Furthermore, it is not clear what the aim was with the suggested sonification: to support visual perception, to evaluate a multi-modal user interaction, or to provide an artistic experience. There has also been examples of sonification of images for the visually impaired (see examples in [30, 31, 32, 33]), even if these show interesting sonification approaches they aim to achieve something slightly different than Photone. These examples try to define the environment for a visually impaired individual, for example in object or obstacle detection, rather than providing a musically interesting multi-modal user experience that aims for modal synergy.

6. METHOD

Photone is exhibited in the science center at Campus Norrköping, Linköping University. It is part of the exhibition Decode the code that aims to explain computer graphics and visualization techniques to the general public (see Figure 3). The science center has about 100,000 visitors per year of all ages, even though school classes of 10- to 14-year-old kids are particularly frequent visitors. In Photone, the user can select from twelve different images to explore and experience. These images come from the different

installations in the science center. All images in the exhibition are synthetic and used to explain volumes, voxels, triangles, shaders and similar computer and visualization concepts for the general public. For the use in Photone, the images were somewhat enhanced in terms of richness of colors, and in range between darker and brighter areas in the image.



Figure 3: Photone being explored by a visitor in the science center, with the ultrasonic distance sensor visible in the dark space between the purple and blue area below the touch screen.

The data from the interactions were collected for two months where roughly 8,000 persons visited the center. The data saved consists of the image that is shown on the display, the position of every pixel explored and the timestamp in milliseconds for each pixel. The update frequency of the data recording is 120 Hz, and the data is saved as a log file named with the date and time for the interaction.

An ultrasonic distance sensor is used to determine if a user is present in front of Photone. When the detected distance exceeds 1 meter the system interprets this to mean that the user has left Photone and saves the data. There are multiple challenges in determining if a user is in front of Photone, for example, if a user leans too far to one side this might be interpreted as the user has left Photone and the interaction will end up in two shorter log files. Moreover, if a user replaces another user too closely, this might go undetected with one log file ending up containing two actual sessions. These challenges could have been overcome by video recording of Photone and the environment around it. However, the distance sensor was deemed to be the best tradeoff between accuracy and privacy.

The data has been analyzed in terms of the total amount of interaction time for each log file. Log files shorter than 30 seconds or longer than 50 minutes have been discarded as either being too short to be interesting for the analyzes or too long to be considered interactions from only one user. The remaining interaction files ($n = 233$) have been analyzed and visualized with a histogram.

The longest 10% of all interactions ($n = 23$) have been further analyzed to study the interaction behavior for users who interacted a long time with Photone. The analysis of these has explored the interaction patterns within each image. The interaction patterns for each image have been analyzed using a heat map where the number of visits to any pixel in the image gets a higher value in the heat map. Based on the heat maps the images have been studied to

identify the image features that seem to attract the users.

7. RESULTS

The level of engagement, as measured in interaction session times, is a long-tail distribution. Photone is not engaging for everyone, and it is clear that most users only interact with Photone for 5 minutes or less, and that the most frequent time interval is 0.5 to 1.5 minutes (see Figure 4). We will focus on the long sessions here, assuming they represent more engaged use.

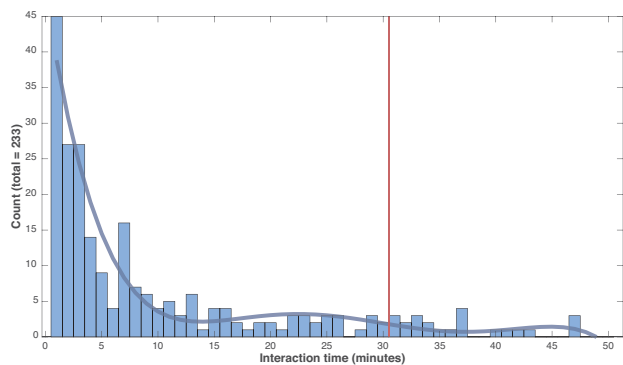


Figure 4: Histogram displaying the interaction time for all interactions (n = 233), with a trend line in dark blue. User interactions to the right of red vertical line (n = 23) are used in the detailed level of the analysis.

The analysis of the 10% longest interactions (n = 23) was done with heat maps displaying the most visited areas in each image. Most of the engaged use seems to follow (semantic) visual saliency, showing a pattern very similar to what conventional eye-tracking data of image viewing would look like (see, for example, [34, 35]). Figures 5 and 6 illustrate this type of distribution.

However, the data also reveals some indications of interaction engagement that are not as clearly similar to visual saliency effects in regular image viewing. Some examples are shown in Figures 7 and 8.

8. DISCUSSION

Already in our previous study [1], our conjecture was that only a limited proportion of a general audience would find interacting with Photone engaging beyond the cursory examination and superficial poking. The distribution of the log data in our present work (Figure 4) confirms this conjecture.

What we did not expect was that among the engaged users (longest 10% of recorded sessions), a majority of the interaction engagement follows a pattern that closely resembles the results we would expect from a conventional eye-tracking study of image viewing. In other words, it would seem like the visually salient objects in an image guide the attention as well as the interactive exploration of the image to a great extent. These areas in the images generally contain more clearly defined visual contrasts resulting in more rapid changing melodies and differences in tonal qualities, as well as more rhythmic instrumentation compared to the background. A plausible scenario could be that a user’s attention is first attracted by the salient visual object, and then the user stays in

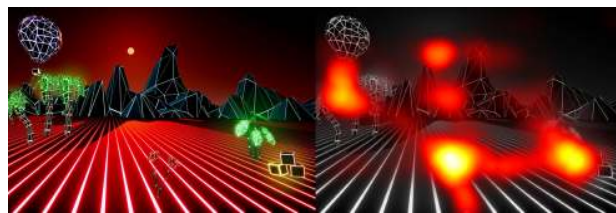


Figure 5: The heat map (to the right) clearly shows that the users mainly explored visual objects in the image, such as the flowers in the foreground, the bush to the right, the palm trees and the hot air balloon to the left, and the sun in the middle.

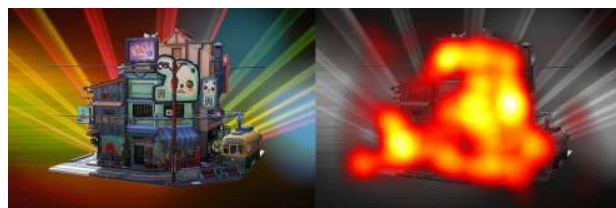


Figure 6: The heat map (to the right) suggests that the block of houses in the middle of the image got much more attention from the user than the background and the color bars emitting from behind the block.



Figure 7: The heat map (to the right) shows that the rabbit was getting more explored than the areas around it. However, for this image the users also explored the shadow beneath the rabbit as well as the reflection in the checkered marble floor.



Figure 8: The heat map (to the right) suggests an interaction pattern where the visually salient objects as well as parts of the non-depictive background were explored.

that area of the image for further exploration of dramatic musical effects and satisfying rhythm.

However, our data does contain examples of interaction engagement also in image areas that are not visually salient in the conventional sense of the word. Looking more closely at those areas, we find that they contain textures and gradients that may yield

engaging visual-auditory synergistic effects in the current design of Photone. For example, in the shadows in front of the rabbit (see Figure 7), the checkers pattern creates clear differences between darker, more sparse musical expression and brighter, more energetic music. The colored reflections from the rabbit also add to the timbre and color the tonal content, as the area is rich in contrasts which creates a quite complex rhythmic instrumentation. A similar experience can be found in the long tails of the cyan shapes (see Figure 8), where the bright tails and darker background create similar experiences as the checkers in the previous example. Also here the area is quite rich in contrasts creating an interesting rhythmic instrumentation.

These are examples of exploration of image areas with high syntactic complexity, from the point of view of the sonification algorithm, but without strong semantic salience in the conventional image-viewing sense. As such, these examples do suggest that there is some amount of modal synergy at work in the experience guiding the user's exploration, albeit less than we had expected when starting our data collection.

One potential reason for the discrepancy is that the first version of Photone, forming the basis for the concept of modal synergy, used proper photographic images. The current version, for which the general audience data were collected, contains synthetic images such as 3D-renderings and even a few flat 2D vector graphics with solid colors as the images used are drawn from an exhibition about computer graphics and visualization. There are significant visual differences between these two versions, in that the photographs have much more nuance, texture and detail than the synthetic images. The sonification mechanism of Photone was originally designed to reward exploration of high-spatial-frequency and visually intricate image areas, which are less prevalent in synthetic images. We suspect that testing a version of Photone with proper photographs would yield slightly different data in general-audience use, possibly showing a more even distribution between visually salient semantic objects and visually-auditively interesting syntactic features.

Another option could be to use less depictive images, where the lack of clearly identifiable visual objects might mitigate the power of visual salience. It would be tempting to test a version of Photone with less-figurative paintings, chosen from e.g. expressionism or pointillism, where a relative lack of visually salient semantic objects is combined with visually complex and engaging syntactic features resulting from the artists craft skills and personal techniques in using brushes and paints with different properties on a textured canvas.

9. CONCLUSION AND FUTURE WORK

Photone does not attract everyone in a public exhibition space, but some of the visitors engage more deeply in the interaction. In general, visually salient objects in an image guide the attention of an engaged user, and exploring them may or may not represent an experience of modal synergy for the user. However, our data also shows interaction engagement in image areas which do not contain visually salient objects in the conventional, semantic sense but rather syntactic features that may yield engaging visual-auditory synergistic effects. Consequently, we claim that there may be traces of modal synergy at work in the general-audience use of Photone. We find these results encouraging enough to warrant further exploration.

A necessary first step would be to validate the method used

in the present study by combining log data with qualitative data to start unpacking the nature of the user experience while interacting with Photone, and specifically how subjectively perceived modal synergy relates to log data as represented in our heat maps above. Self-reflection through prompted recall would be a suitable approach to collect qualitative experiential data than can be correlated with previously collected logs, and think-aloud protocols during use can also be considered even though there is always the danger that verbalization during the interaction changes the nature of the experience significantly.

Such validation can be expected to provide guidelines for more robust interpretation of future interaction logs. Next, a more systematic quantitative study of different types of images, as discussed in section 8 Discussion, all using the same sonification algorithm would be most enlightening. More generally, these considerations form the basis for a slightly more systematic approach to visual image space where candidate images for Photone sonification can be placed along two dimensions: level of complexity in syntactic visual features, and level of semantic figurativeness. Empirical testing with the same sonification algorithm applied to images from all four quadrants of this space could yield two major insights: (1) a better understanding of which kinds of images work best with the current sonification algorithm to evoke modal synergy experiences, and (2) inspiration for re-designing sonification algorithms for each of the four quadrants towards more modal synergy. Such a qualitative study would also answer questions about the experience of the sonification in itself to provide insights in, for example, the quality of the synthesis or the use of musical elements.

A third step in this progression, assuming that we find differences between image types, would be to re-design sonification algorithms specifically to engender modal synergy for each of the main types of images, and to assess them through further rounds of quantitative evaluation.

10. REFERENCES

- [1] N. Rönnerberg and J. Löwgren, "Photone: Exploring modal synergy in photographic images and music," in *Proc. International Conference on Auditory Display (ICAD 2018)*, 2018, pp. 73–79.
- [2] R. W. Pridmore, "Music and color: Relations in the psychophysical perspective," *Color Research & Application*, vol. 17, pp. 57–61, 1992.
- [3] W. G. Collier and T. L. Hubbard, "Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode," *Musicae Scientiae*, vol. 8, pp. 151–173, 2004.
- [4] L. E. Marks, "On cross-modal similarity: Auditoryvisual interactions in speeded discrimination," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13, pp. 384–394, 1987.
- [5] J. Ward, B. Huckstep, and E. Tsakanikos, "Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?" *Cortex*, vol. 42, pp. 264–280, 2006.
- [6] S. P. Cleary, "Using the psychology of color schemes to create an appreciative advising environment," *Journal of Appreciative Education*, vol. 2, pp. 24–38, 2015.

- [7] C. Ware, *Information Visualization: Perception for Design*, 3rd ed. San Francisco: Morgan Kaufmann Publishers Inc., 2013.
- [8] E. G. S. Patrick G. Hunter and U. Schimmack, “Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions,” *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, pp. 47–56, 2010.
- [9] S. A. Iakovides, V. T. Iliadou, V. T. Bizeli, S. G. Kaprinis, K. N. Fountoulakis, and G. S. Kaprinis, “Psychophysiology and psychoacoustics of music: Perception of complex sound in normal subjects and psychiatric patients,” *Annals of General Hospital Psychiatry*, vol. 3, pp. 1–4, 2004.
- [10] P. Juslin and P. Laukka, “Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening,” *Journal of New Music Research*, vol. 33, pp. 217–238, 2004.
- [11] J. McCartney, “Supercollider: A new real-time synthesis language,” in *Proc. International Computer Music Conference (ICMC)*, 1996, pp. 257–258.
- [12] —, “Rethinking the computer music language: Supercollider,” *IEEE Computer Graphics & Applications*, vol. 26, pp. 61–68, 2002.
- [13] R. M. Prendergast, *Film Music: A Neglected Art*. W.W. Norton, 1992.
- [14] C. Gorbman, *Unheard melodies: narrative film music*. London: Bloomington: BFI Pub.; Indiana University Press, 1987.
- [15] B. Langkjær, *Filmyd & filmmusik: fra klassisk til moderne film*. Museum Tusulanum Press, 1997.
- [16] C. E. Seashore, *Psychology of Music*. New York, US: Dover, 1967.
- [17] L. Ribas, “Sound and image relations: a history of convergence and divergence,” *Divergence Press*, 2016.
- [18] T. Y. Levin, “Tones from out of nowhere: Rudolph Pfenninger and the archaeology of synthetic sound,” *Green Room*, vol. 12, pp. 32–79, 2003.
- [19] G. Levin, “Audiovisual software art,” in *Audiovisuology: Compendium*, D. Daniels and S. Naumann, Eds. Köln: Verlag Walther König, 2010, pp. 271–283.
- [20] S. Naumann, “The expanded image: On the musicalization of the visual arts in the twentieth century,” in *Audiovisuology, A Reader, Vol. 1: Compendium, Vol. 2: Essays*, D. Daniels and S. Naumann, Eds. Köln: Verlag Walther König, 2015, pp. 504–233.
- [21] A. Tanaka, “The sound of photographic image,” *AI & Society: Knowledge, Culture and Communication*, vol. 27, pp. 315–318, 2012.
- [22] T. Hermann and A. Hunt, “The discipline of interactive sonification,” in *Proc. of the Int. Workshop on Interactive Sonification Workshop (ISON-2004)*. Germany: Bielefeld University, 2004, pp. 1–9.
- [23] A. Hunt and T. Hermann, “The importance of interaction in sonification,” in *Proc. of the 10th Meeting of the International Conference on Auditory Display (ICAD 2004)*, Sydney, Australia, 2004, pp. ICAD04–1–ICAD04–8.
- [24] R. Sarkar, S. Bakshi, and P. K. Sa, “Review on image sonification: A non-visual scene representation,” in *Proc. 1st International Conference on Recent Advances in Information Technology (RAIT)*, Dhanbad, India, 2012, pp. 86–90.
- [25] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*, 1st ed. Berlin, Germany: Logos Publishing House, 2011.
- [26] T. Pinch and K. Bijsterveld, *The Oxford Handbook of Sound Studies*. Oxford University Press, 2012.
- [27] K. Franinovic and S. Serafin, *Sonic Interaction Design*. MIT Press, 2013.
- [28] C. O’Neill and K. Ng, “Hearing images: Interactive sonification interface for images,” in *Proc. of the International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution*. Florence, Italy: IEEE Computer Society, 2008, pp. 25–31.
- [29] M. D. Heath and G. Hunter, “Listen to the picture! the statson sound sonification system, using vst and dsp,” in *Proc. of the Acoustics 2012 Nantes Conference*. Nantes, France: Société Française d’Acoustique, 2012, pp. 3861–3866.
- [30] M. Banf and V. Blanz, “A modular computer vision sonification model for the visually impaired,” in *Proc. of the 18th International Conference on Auditory Display (ICAD 2012)*, Atlanta, GA, USA, 2012, pp. 121–128.
- [31] —, “Sonification of images for the visually impaired using a multi-level approach,” in *Proc. of the 4th Augmented Human International Conference (AH13)*, Stuttgart, Germany, 2013, pp. 162–169.
- [32] S. Cavacoa, J. T. Henriquesb, M. Menguccia, and F. M. Nuno Correiaa, “Color sonification for the visually impaired,” *Procedia Technology*, vol. 9, pp. 1048–1057, 2013.
- [33] P. Skulimowski, M. Owczarek, A. Radecki, M. Bujacz, and P. Strumiłło, “Interactive sonification of the u-disparity maps of 3d scenes,” in *Proc. 5th Interactive Sonification Workshop (ISON-2016)*. Germany: CITEC, Bielefeld University, 2016, pp. 18–22.
- [34] D. Massaro, F. Savazzi, C. D. Dio, D. Freedberg, V. Gallese, G. Gilli, and A. Marchetti, “When art moves the eyes: A behavioral and eye-tracking study,” *PLoS ONE*, vol. 7, pp. 1–16, 2012.
- [35] D. Villani, F. Morganti, P. Cipresso, S. Ruggi, G. Riva, and G. Gilli, “Visual exploration patterns of human figures in action: an eye tracker study with art paintings,” *Frontiers in Psychology*, vol. 6, pp. 1636 1–10, 2015.