

## The sound of photographic image

Atau Tanaka

Received: 18 April 2011 / Accepted: 9 August 2011 / Published online: 26 August 2011  
© Springer-Verlag London Limited 2011

This text presents three works produced in the period 1998–2004 that are based on the sonification of photographic images. Two of the works, *9m14s Over Vietnam*, and *Bondage.rmx*, were compositions for recorded CD release, while the third, *Bondage*, was an interactive installation. All three utilized similar processes to translate visual information to sound, and to allow musical composition with imagery.

Strong, striking images were chosen for each of the works. In *9m14s Over Vietnam*, the source photograph was Nick Ut's Pulitzer Prize winning photograph of Kim Phuc and her siblings running to escape a napalm attack in 1972 during the Vietnam War. The image in many ways represents the human horror of the war and is one of the iconic images deep in the public collective conscience. For this, the musical work evokes the image in its absence. The music was not meant to accompany the photograph, but to be the photograph perceived through sound. I felt that it was a strong enough and well known enough image that the original did not need to be explicitly displayed.

If perception of the work did not depend on the explicit presence of the source image, the compositional process did. The musical structure was created by working directly with the image, manipulating brightness and color, focusing on specific elements of the image through cropping and zooming. Traditional musical compositional techniques of motivic development, repetition, variation, and orchestration were carried out by cutting, pasting, cropping, layering, and juxtaposing elements from the original

photograph. So, while listening to *9m14s Over Vietnam* taking place in the absence of the image, the composition of the work took place confronting the image in a prolonged and profound way. It is through the intensity of this interaction that I hoped to convey the power of the image.

I used two methods for translating scanned image data to sound in *9m14s Over Vietnam*: temporal mapping and additive synthesis. In the first, the raw image data are read in a brute force manner as sound. The scanned grayscale image was saved in an uncompressed Baseline Tagged Image File Format (TIFF). Each byte of the data file then became the grayscale value of an individual pixel in the image. By modifying the file header and extension, these data were directly converted into a sound file. The grayscale values of the image, scanning the left to right, row by row, became an audible sound waveform that progressed in time. Brighter values of gray became higher amplitude sound samples. Alternation between light and dark values in the image became periodic undulation of the audio waveform, producing a brutal, noisy reading of the image. Setting the playback rate of the data stream to the CD sampling rate of 44,100 Hz resulted in an audio file of 9'14'', giving the musical work its title.

Alongside the direct translation of image data to sound samples, I used a second technique based on Inverse Fourier Transforms (IFT). Sonograms are one way to visualize sound where the spectral content of sound is presented in a two-dimensional image where time progresses from left to right and where sound frequency is represented from low to high. The IFT enables sound to be resynthesized from a sonogram. In this case, the input to the IFT was not a sonogram that represented a previously analyzed sound, but was instead the source photograph.

---

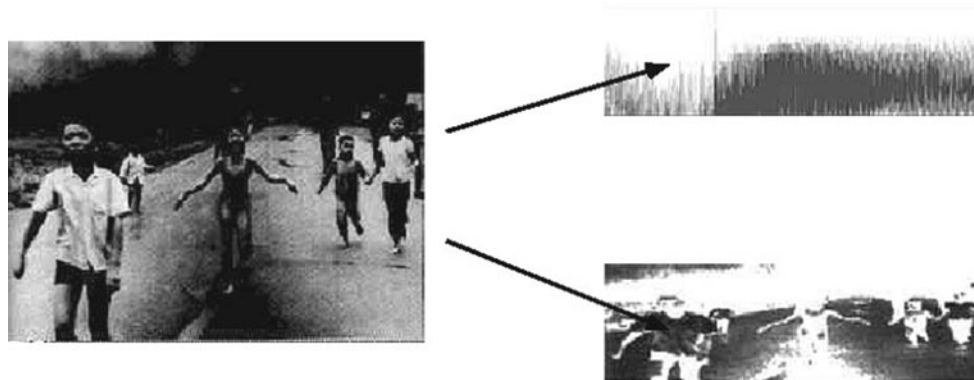
A. Tanaka (✉)  
Culture Lab, Newcastle University,  
Newcastle upon Tyne NE1 7RU, UK  
e-mail: [atau.tanaka@ncl.ac.uk](mailto:atau.tanaka@ncl.ac.uk); [culturelab@ncl.ac.uk](mailto:culturelab@ncl.ac.uk)

Through this process, the image became a sound frequency map, where visual elements in the lower parts of the image represented low-frequency sound content and elements in the upper parts of the image represent higher frequencies. With an image based on a landscape in perspective like this one, the ground would then be sonified as lower-frequency tones and the sky as higher-frequency tones. Each row of pixels in the image is assigned a frequency in a series in the musical spectrum from 32 to 5,000 Hz. As the image is scanned from left to right, luminosity in any given row would translate to amplitude in a sine wave playing at the corresponding frequency. The brighter the image, the louder the corresponding sine wave. Scanning the image from left to right across time, in this case in 9m14s, to correspond to the other, temporal reading of the image, gives a constantly shifting timbre that is the sum of the sine waves playing that modulates with the brightness and curves of the original photograph.

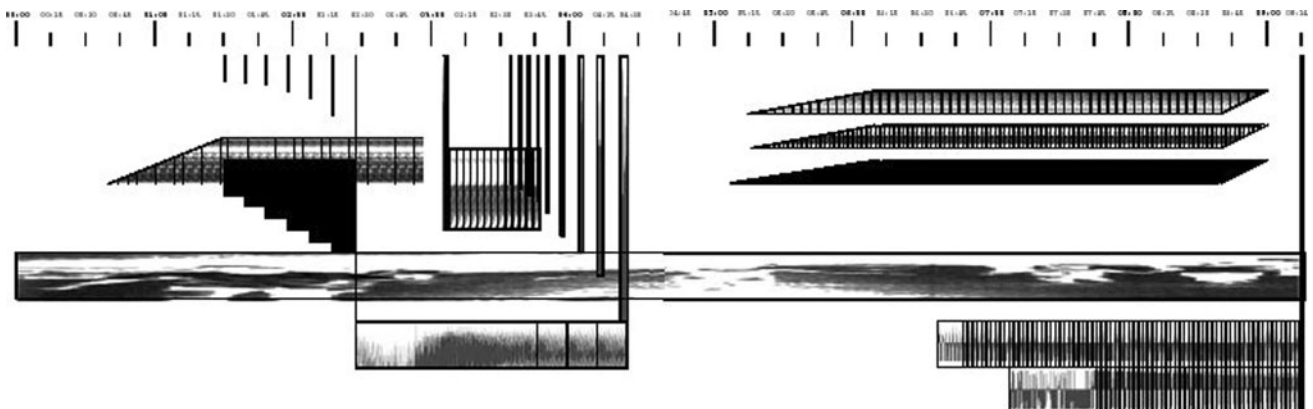
The original source image, its two treatments, and resulting score are seen in Figs. 1 and 2. By inverting the photograph to its negative and using grayscale intensity to encode sound, it is visually clear where (dark pixels) there

will be audible frequencies. By recentering the image with Kim Phuc at the middle of the image, we get a symmetry in which her body and the perspective of the road give an eerie brightness and silence.

Bondage and Bondage.rmx extended the process of 9m14s Over Vietnam, concentrating on spectral resynthesis. They both used the same source image, a photograph by the Japanese photographer, Nobuyoshi Araki. Araki is well known for provocative imagery playing on Japanese iconographies that portray sadomasochistic staging. I was given access to Araki's personal archive of unpublished Polaroids. Based on the compositional experience with 9m14s, I had a sense for the kinds of sounds an image, and characteristics of an image, might give. This was a composer's instinct, not dissimilar to knowing how a certain instrumental combination or orchestration might sound without needing to hear it directly. In this case, the hundreds of Polaroids in the collection all represented variations on a theme—photographs of Japanese women in traditional kimonos staging scenes of bondage. The combination of curves of heads and shoulders, striking diagonal lines of rope, and the



**Fig. 1** Scanned and cropped photograph, its direct translation to sound, and its negative to be used for spectral resynthesis

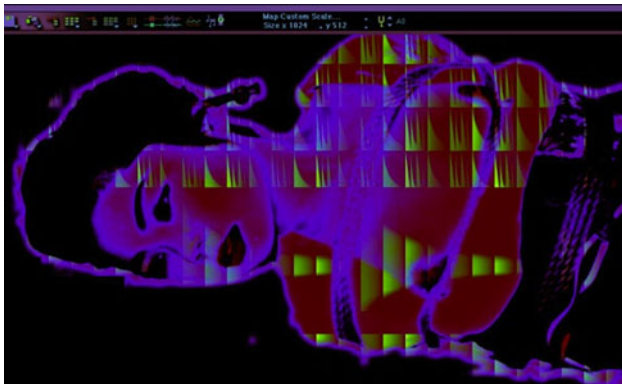


**Fig. 2** Score, with photograph negative stretched across duration in the middle row, and other elements placed, looped, and faded

patterns of the kimono fabric lent to, in my mind's ear, varying orchestrations of smooth glissandi, transient percussive sounds, and intricate timbral detail. I chose an image where these elements were balanced compositionally in the original photograph with the hunch that it would provide me rich musical material.

With *Bondage.rmx*, the process was similar to *9m14s* in working with the image to produce a piece of recorded music for CD release. In this case, I focused exclusively on additive synthesis and did not use a time domain translation of image data to sound (Fig. 3). Instead, I developed a more detailed approach to working with the image as sonogram and with the spectral resynthesis process. The frequency bands for resynthesis could be programmed in a musical fashion and quantized to tonal constraints such as whole tone and pentatonic mappings. This led to an “oriental” tonal sensation, while the overall distribution of tones was shaped by the image as a sonogram. In addition, instead of performing the resynthesis using only sine waves, some frequency bins were assigned sound samples. Here, I drew upon Japanese iconography in a way that Araki does visually. Where his subjects are dressed in kimonos in scenes staged in tatami rooms, I used samples of temple bells and gagaku instruments I had sampled in my travels in Japan. Disruptive visual elements—the rope that bonded Araki's subjects—found gentler reading in the sine waves that occupied frequency bands bounding the samples.

The installation version of *Bondage* was created in response to a residency and commission from Le Fresnoy in northern France and in discussion with curator Kathleen Forde as she prepared a major touring exhibition of sound and image, “What Sound Does a Color Make?” The work was shown at the Panorama 5 exhibition in Tourcoing, La Villette Numérique in Paris, and then toured New York,



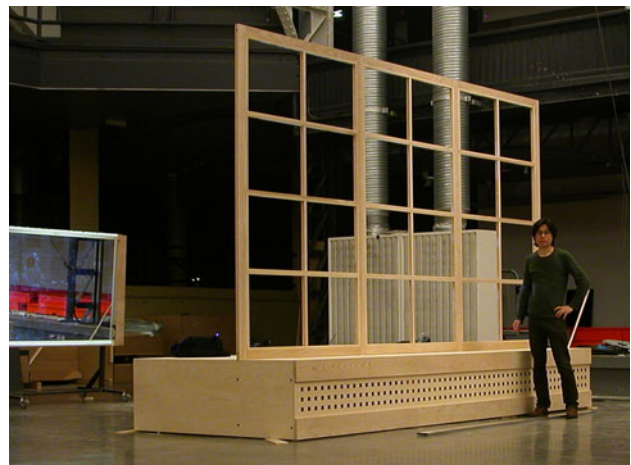
**Fig. 3** Araki's photograph treated and processed inside metasynt software

Pittsburgh, Maryland, Hawaii, and New Zealand with Independent Curators International.

Working with craftsmen in the film set workshop at Le Fresnoy, we constructed an oversized replica of a shoji, the Japanese wood and paper door panel. This 3 m × 4 m structure became a rear projection screen, with layers of translucent paper and fiber cloth stretched across the panes to diffuse the projected image, obscuring the point source nature of the projection bulb and highlighting the rice paper-like fibrous texture. This analog structure of paper and wood became the surface onto which was projected a pixelized interactive digital image of Araki's transformed photograph.

The structure consisted of three independent panels, each with four rows of double panes (Fig. 4), resulting in a 6 × 4 division of the image. Pairs of panes were grouped together, to create a 3 × 4 matrix giving 12 zones that were sonified independently. Loudspeakers were placed in each corner of the screen, creating a vertical quadrasonic space across which sound from each of the 12 zones was panned according to visual spatial distribution.

Araki's image was inverted and projected as an X-ray-like negative onto this surface. A series of 12 red scan lines swept across each pair of panes, acting as the tape head reading luminosity of image and translating that to a spectral orchestration of sine waves and sound samples (Fig. 5). Each row of the shoji represented different frequency ranges, with the bottom row producing sub-low frequencies, the second row bass frequencies, third-row mid-range, and top-row high frequencies. The panning across the four loudspeakers gave a spatial distribution that brought sound into the space according to the vertical topology of the screen, giving the sonified image a form in space that paralleled and abstracted the body of Araki's subject.



**Fig. 4** The wooden shoji structure with base unit, three panels, and 12 pane pairs



**Fig. 5** The negative, pixelized image, and red scan lines

An infrared camera was placed at the top of the screen, capturing human presence in front of the structure. A real-time lumakey process used the silhouettes of people standing in front of the screen as a mask on the source image. Where the photograph in its native state was seen as a negative, the areas under the silhouettes of people beholding the photograph exposed the original black-and-white positive (Fig. 6). As people moved across the image, their own shadows uncovered the body of the subject, her kimono, and the bonding rope. This change in the parts of the image that were negative and the parts that were positive gave dynamic shifts in luminosity that resulted in interactive shifting harmonies in the sound. The bodies of both spectator and photographic subject were mapped to frequencies and entered into interplay (Fig. 7). The silhouette exposing parts of the photograph followed the viewer, but faded away with stillness. This created a voyeuristic tension as the viewer was drawn to inspect the



**Fig. 6** Simulation illustrating lumakey process whereby silhouettes of gallery visitors expose regions of the black-and-white positive photograph



**Fig. 7** Bondage in exhibition at Villettes Numérique, Paris (*photo PER*)

image of the untouchable woman on the other side of the screen. As soon as the spectator settled into focus in, the image faded back to an X-ray negative, ever escaping the voyeur.