IMAGE-BASED SPATIALIZATION

Eric Lyon
School of Creative Arts
Sonic Arts Research Centre
Queen’s University Belfast
Belfast BT7 1NN
United Kingdom

INTRODUCTION

An approach to spatialization is described in which the pixels of an image determine both spatial and other attributes of individual elements in a multi-channel musical texture. The application of this technique in the author’s composition Spaced Images with Noise and Lines is discussed in detail. The relationship of this technique to existing image-to-sound mappings is discussed. The particular advantage of modifying spatial properties with image filters is considered.

1. INTRODUCTION

Image-based spatialization involves the mapping of an image to the spatial properties of a sound. Whereas existing literature tends to discuss image-based, or video-based, spatialization systems from the standpoint of introducing compositional tools and software systems, the primary contribution of this paper will be to discuss the specific approach of image-based spatialization to the creation of a recent computer music composition.

2. IMAGE-BASED SPATIALIZATION

In the work described here, spatialization is achieved by mapping pixel values to sound grains. Each horizontal line is virtually wrapped around the perimeter of a surround-sound speaker array. The technique for Spaced Images with Noise and Lines targets an eight-channel surround speaker array, with four virtual sound source locations calculated using power panning across each pair of adjacent speakers. Thus, this configuration calls for images with 32 pixels along the x-axis, to be mapped to 32 virtual sound locations circling the audience. The vertical dimension maps to time, and thus may be arbitrarily determined for each image. The taler the image, the more perimeter lines are converted to sonic moments, each with its own spatial profile. Scanning down the lines of the image advances time at a fixed interval. (The time interval could be warped, rather than constant, if desired.) In a given line, the darkness of each pixel represents the amplitude of an event, or grain, at the virtual spatial perimeter location associated with that pixel. An image with a single black line drawn diagonally downward, with a pixel-width of one, would result in a perimeter-panning spatialization effect.

3. IMPLEMENTATION

A Max patch is used to convert an image to a Csound score, which is then compiled into an 8-channel audio file. (See Figure 1.) Each 32-pixel line is transformed into a list of floating point values from 0-1, which drive a JavaScript program that writes its output directly to a Csound score.

4. SPACED IMAGES WITH NOISE AND LINES

Spaced Images with Noise and Lines was composed in 2011 and premiered at a Spatial Music Collective concert at The Joinery in Dublin. All passages in the work were generated as described above, using images that the author created with the free image-processing software, The Gimp [1]. The images are comprised of noise, generated by The Gimp, and freehand, mouse-drawn lines. Prior to audio conversion, the images are subject to various forms of image processing, such as convolution, erosion, and smoothing. (See Figure 2.) Most passages were generated from a single image, but in one case, sounds generated from two complementary images were superimposed to create a two-voice spatial trajectory counterpart.

5. MAPPING STRATEGIES

The simplest mapping strategy involves routing a single sound to each virtual location, scaled by the amplitude determined by the darkness of its corresponding pixel. This is the duplication technique described by McGee and Wright [6]. This strategy is applied to the first sound heard in the piece, which is a synthetic noise with a percussive envelope and a resonant peak. The changes in the image with its erasures, noise and lines, produces a spatial result that fluctuates between a somewhat indeterminate sense of motion, and clear, dramatic spatial trajectories. The use at the outset of a simple, spectrally stable sound focuses the listener’s attention on spatial attributes. Successive mappings introduce greater variety among the sounds. In one instance, the y-axis is mapped to resonant frequency, so that the pitch descends as the corresponding image is scanned downward. Several sections assign fixed resonant frequencies to different virtual locations, while still implementing the darkness-to-amplitude mapping to articulating sounds. The result is a spatialized chord, when the image is noisy, and melodic when the image is line-based. Both visual elements were generally adjoined in the source images.

In another section, enveloped sine waves are employed. The frequency for each sinusoidal grain is randomly chosen from a fixed, equal-tempered scale. In homage to Karlheinz Stockhausen, the frequency scale is taken from his Studie II. [11] In this section, the pixel amplitude controls two synthesis parameters: the amplitude and duration of a corresponding sinusoidal grain. This section was generated from an image that superimposes two kinds of noise — a low-amplitude, high-density noise, and a high-amplitude, very low-density noise, with just a few pixels activated. The result is a constant, quiet, sort of thudding background articulation (albeit spatially complex) from the low-amplitude pixels, and a sporadic, bell-like surface melody with clear spatial articulation from the high amplitude-pixels. Although sine waves are often difficult to resolve spatially, the single pixel mapping to a fixed virtual location, combined with the percussive attack envelope, locates the individual sinusoids Starkly in different regions of the space.

RELATED WORK

The second section of the piece focuses on sampled drum sounds. The initial treatment of drums starts with a hi-hat, which is reminiscent of the opening filtered noise sound. Very shortly, a full ensemble drum texture emerges, and then gradually destabilizes as individual attack times are subject to increasing random deviation. Individual samples are assigned to fixed virtual locations, which are then articulated according to pixel amplitudes. Random delays, ring modulation and filtering are gradually introduced, complicating the timbral space. The texture gradually thins until only the cowbell sample remains.

The work closes with two contrasting sections. First comes the only slow section, in which pixels articulate sinusoids of 10 seconds duration, with frequencies that glide slowly downward. The time interval between line scans is approximately 500 ms. The concluding section uses a line drawing to articulate enveloped, filtered noise, with resonance frequencies locked to virtual locations. With only lines used in this image, the spatial trajectories of this concluding passage are exceedingly clear to the ear.

The list of related work is very long, and I will try to omit it here. The work described here is related to image-based spatialization, and much of it is focused on automated mapping of video to space. There is also some unpublished work, which I have encountered informally. When I first decided to try these experiments, I discussed the idea with Shawn Greenlee, knowing of his expertise with the image-processing software Jitter. Shawn informed me that not only was what I proposed rather easy to implement, but that he had already done some spatial mapping from video to audio in a stereo field, and reckoned that it would not be difficult to extrapolate his work to a multi-channel context. Shawn graciously shared his patch, which I then modified for the current work. [2] I also recently learned that Augustus Leudar, an MA student at SARC, has worked with video to space in a quadraphonic context, using ambisonics for the spatialization, mapping a single video frame to the geometric square defined by four corner speakers. [3]

Another technique considerably closer to the work described here maps image onto parameters other than space, often to frequency. Here could be mentioned Oramics[7] and Iannis Xenakis’s Upic system [5]. Examples in the digital domain include Christopher Penrose’s HyperUpic [8] and Metasynth by UI Software [12].

7. ADVANTAGES OF IMAGE-BASED SPATIALIZATION

Working from an image provides many of the advantages of a graphical score, including the ability to easily specify pressures and density at multiple time points...
An approach to spatialization is described in which the pixels of an image determine both spatial and other attributes of individual elements in a multi-channel musical texture. The application of this technique in the author’s composition Spaced Images with Noise and Lines is discussed in detail. The relationship of this technique to existing image-to-sound mappings is discussed. The particular advantage of modifying spatial properties with image filters is considered.

1. INTRODUCTION

Image-based spatialization involves the mapping of an image to the spatial properties of a sound. Whereas existing literature tends to discuss image-based, or video-based, spatialization systems from the standpoint of introducing compositional tools and software systems, the primary contribution of this paper will be to discuss the specific approach of image-based spatialization to the creation of a recent computer music composition.

2. IMAGE-BASED SPATIALIZATION

In the work described here, spatialization is achieved by mapping pixel values to sound grains. Each horizontal line is virtually wrapped around the perimeter of a surround-sound speaker array. The technique for Spaced Images with Noise and Lines targets an eight-channel surround-speaker array, with four virtual sound source locations calculated using power panning across each pair of adjacent speakers. Thus, this configuration calls for images with 32 pixels along the x-axis, to be mapped to 32 virtual sound locations circling the audience. The vertical dimension maps to time, and thus may be arbitrarily determined for each image. The taller the image, the more perimeter lines are converted to sonic moments, each with its own spatial profile. Scanning down the lines of the image advances time at a fixed interval. (The time interval could be warped, rather than constant, if desired.) In a given line, the darkness of each pixel represents the amplitude of an event, or grain, at the virtual spatial location associated with that pixel. An image with a single black line drawn diagonally downward, with a pixel-width of one, would result in a perimeter-panning spatialization effect.

3. IMPLEMENTATION

A Max patch is used to convert an image to a Csound score, which is then compiled into an 8-channel audio file. (See Figure 1.) Each 32-pixel line is transformed into a list of floating point values from 0-1, which drive a JavaScript program that writes its output directly to a Csound score.

4. SPACED IMAGES WITH NOISE AND LINES

Spaced Images with Noise and Lines was composed in 2011 and premiered at a Spatial Music Collective concert at The Joinery in Dublin. All passages in the work were generated as described above, using images that the author created with the free image-processing software, The Gimp [1]. The images are comprised of noise, generated by The Gimp, and freehand, mouse-drawn lines. Prior to audio conversion, the images are superimposed on the image with its erasures, noise and lines, produces a percussive envelope and a resonant peak. The changes in the image with its erasures, noise and lines, produces a spatial result that fluctuates between a somewhat indeterminate sense of motion, and clear, dramatic spatial trajectories. The use at the outset of a simple, spectrally stable sound focuses the listener’s attention on spatial attributes. Successive mappings introduce greater variety among the sounds. In one instance, the y-axis is mapped to resonant frequency, so that the pitch descends as the corresponding image is scanned downward. Several sections assign fixed resonant frequencies to different virtual locations, while still implementing the darkness-to-amplitude mapping to articulating sounds. The result is a spatialized chord, when the image is noisy, and melodic when the image is line-based. Both visual elements were generally admired in the source images.

In another section, enveloped sine waves are employed. The frequency for each sinusoidal grain is randomly chosen from a fixed, equal-tempered scale. In homage to Karlheinz Stockhausen, the frequency scale is taken from his Studie II [11]. In this section, the pixel amplitude controls two synthesis parameters: the amplitude and duration of a corresponding sinusoidal grain. This section was generated from an image that superimposes two kinds of noise – a low-amplitude, high-density noise, and a high-amplitude, very low-density noise, with just a few pixels activated. The result is a constant, quiet, sort of thudding background articulation (albeit spatially complex) from the low-amplitude pixels, and a sporadic, bell-like surface melody with clear spatial articulation from the high amplitude-pixels. Although sine waves are often difficult to resolve spatially, the single pixel mapping to a fixed virtual location, combined with the percussive attack envelope, locates the individual sinusoids starkly in different regions of the space.

The work opens with two contrasting sections. First comes the only slow section, in which pixels articulate sinusoids of 10 seconds duration, with frequencies that glide slowly downward. The time interval between line scans is approximately 500 ms. The concluding section uses a line drawing to articulate enveloped, filtered noise, with resonance frequencies locked to virtual locations. With only lines used in this image, the spatial trajectories of this concluding passage are exceedingly clear to the ear.

5. MAPPING STRATEGIES

The simplest mapping strategy involves routing a single sound to each virtual location, scaled by the amplitude determined by the darkness of its corresponding pixel. This is the duplication technique described by McGee and Wright [6]. This strategy is applied to the first sound heard in the piece, which is a synthetic noise with a percussive envelope and a resonant peak. The changes in its erasures, noise and lines, produces a spatial result that fluctuates between a somewhat indeterminate sense of motion, and clear, dramatic spatial trajectories. The use at the outset of a simple, spectrally stable sound focuses the listener’s attention on spatial attributes. Successive mappings introduce greater variety among the sounds. In one instance, the y-axis is mapped to resonant frequency, so that the pitch descends as the corresponding image is scanned downward. Several sections assign fixed resonant frequencies to different virtual locations, while still implementing the darkness-to-amplitude mapping to articulating sounds. The result is a spatialized chord, when the image is noisy, and melodic when the image is line-based. Both visual elements were generally admired in the source images.

In another section, enveloped sine waves are employed. The frequency for each sinusoidal grain is randomly chosen from a fixed, equal-tempered scale. In homage to Karlheinz Stockhausen, the frequency scale is taken from his Studie II [11]. In this section, the pixel amplitude controls two synthesis parameters: the amplitude and duration of a corresponding sinusoidal grain. This section was generated from an image that superimposes two kinds of noise – a low-amplitude, high-density noise, and a high-amplitude, very low-density noise, with just a few pixels activated. The result is a constant, quiet, sort of thudding background articulation (albeit spatially complex) from the low-amplitude pixels, and a sporadic, bell-like surface melody with clear spatial articulation from the high amplitude-pixels. Although sine waves are often difficult to resolve spatially, the single pixel mapping to a fixed virtual location, combined with the percussive attack envelope, locates the individual sinusoids starkly in different regions of the space.

6. RELATED WORK

Relatively little has been published on image-based spatialization, and much of it is focused on automated mapping of video to space. There is also some unpublished work, which I have encountered informally. When I first decided to pursue these experiments, I discussed the idea with Shawn Greenlee, knowing of his expertise with the image-processing software Jitter. Shawn informed me that not only was what I proposed rather easy to implement, but that he had already done some spatial mapping from video to audio in a stereo field, and reckoned that it would not be difficult to extrapolate his work to a multi-channel context. Shawn graciously shared his patch, which I then modified for the current work. [2] I also recently learned that Augustus Leardar, an MA student at SAE, has worked with video to space in a quadraphonic context, using ambisonics for the spatialization, mapping a single video frame to the geometric square defined by four corner speakers. [3]

Another technique considerably closer to the work described here maps image onto parameters other than space, often to frequency. Here could be mentioned Oramics [7] and Iannis Xenakis’s Upic system [5]. Examples in the digital domain include Christopher Penrose’s HyperUpic[8] and Metasynth by UI Software [12].

7. ADVANTAGES OF IMAGE-BASED SPATIALIZATION

Working from an image provides many of the advantages of a graphical score, including the ability to easily specify textures and density at multiple time
advantages over image-based spatialization. For spatialization with video would seem to have certain potential of the technique were shown to be somewhat coherent sonic objects within a desired spatial environment. However, every example I have encountered of video-based spatialization has been presented as an integrated audio-visual composition.) In audio-visual work, the imagery must be as aesthetically strong as the sound, and the two must work well together. This requirement could put constraints on the content of the video. A situation in which both audio and video are presenting the same information introduces the danger of aesthetic redundancy. Applying filters to modify spatial behavior is much easier and more direct for image processing than for video processing. Nonetheless, video-based spatialization can still be quite effective, especially when drawing on the advantages of live performance. In any case, despite some overlap in the manner of data processing, video-based spatialization and image-based spatialization remain two distinctly different approaches to spatialization.

9. AESTHETIC RESULTS

In “Space-form and the Acousmatic Image,” Denis Smalley defines space-form as “An approach to musical form, and its analysis, which privileges space as the primary articulator. Time acts in the service of space” [10]. In the same article, Smalley makes a distinction between interventionist and naturalist approaches. “With the interventionist approach the composer’s hand is in evidence, and the stamp of the technology and techniques is apparent in the kind of material and the way it is manipulated, whereas in the naturalist work there will be some attempt to hide techniques, and avoid exposing technological signifiers” [10]. Spaced Images with Noise and Lines could accurately be described as an interventionist space-form composition. From the outset, the spatial aspect of the work is kept in the compositional foreground through dramatic panning trajectories. Sonic elements are deliberately kept simple throughout in order to highlight the spatial aspect. Different approaches to texture creation, and the admixture of noise and lines in the images, provide a broad variety of spatial experiences to the audience. Perhaps most importantly, the spatial structures are integral to the music. The complexity of the spatial relations is built into the musical structure; it could not have been imposed externally through performed manual diffusion techniques.

10. FUTURE WORK

Only greyscale image processing was used in Spaced Images with Noise and Lines. It would be interesting to use the separate RGB channels to affect different aspects of synthesis, and/or spatialization. The use of image masks could facilitate the creation of various forms of spatial polyphony. The mapping described here makes the most sense in the context of 2D speaker arrays. A rethink of the spatial image mapping strategy may be appropriate for 3D arrays, such as the Klangdon at ZKM, or the Sonic Lab at SARC.

11. CONCLUSIONS

The technique of image-based spatialization was discussed in the context of the author’s composition Spaced Images with Noise and Lines. The creative potentials of the technique were shown to be somewhat different from more traditional trajectory-based spatialization techniques. Given the growing number of multi-channel performance venues, and opportunities for composers and audiences to explore spatial music, techniques such as image-based spatialization may be of increasing interest to composers of computer music.

12. REFERENCES

Spatialization with video would seem to have certain such as Zirkonium \[9\] nonetheless still conceptualize processing software to bear on the creation of spatial to bring the full power of sophisticated image processing to small segments of a larger spatial trajectory. Practically speaking, the relationship between the image and the resulting spatial effect is rarely isomorphic; the cautions about spatial composition provided by Kendall and Cabrera remain in effect \[4\]. However, often enough the results of the process are predictable, and musically effective. More importantly, the spatial results described here would be difficult or impossible to achieve using more traditional control interfaces for spatialization, which tend to focus on the locations or trajectories of coherent sonic objects within a desired spatial environment.

8. COMPARISONS WITH VIDEO-BASED SPATIALIZATION

Spatialization with video would seem to have certain advantages over image-based spatialization. For example, there is a more direct conceptual mapping between an individual video frame, and a surround speaker array, at least for 2D surround sound arrays. Video lends itself to real-time processing in a way that is unavailable for image-based spatialization. However, the temporal association between the video and audio is a mixed blessing. Images used for spatialization are not to be viewed during performance, and therefore their aesthetic qualities need not be considered. (Of course it is possible to create video-based spatial sound in which the video itself is not presented to the audience. However, every example I have encountered of video-based spatialization has been presented as an integrated audio-visual composition.) In audio-visual work, the imagery must be as aesthetically strong as the sound, and the two must work well together. This requirement could put constraints on the content of the video. A situation in which both audio and video are presenting the same information introduces the danger of aesthetic redundancy. Applying filters to modify spatial behavior is much easier and more direct for image processing than for video processing. Nonetheless, video-based spatialization can still be quite effective, especially when drawing on the advantages of live performance. In any case, despite some overlap in the manner of data processing, video-based spatialization and image-based spatialization remain two distinctly different approaches to spatialization.

9. AESTHETIC RESULTS

In “Space-form and the Acousmatic Image,” Denis Smalley defines space-form as “An approach to musical form, and its analysis, which privileges space as the primary articulator. Time acts in the service of space” \[10\]. In the same article, Smalley makes a distinction between interventionist and naturalist approaches. “With the interventionist approach the composer’s hand is in evidence, and the stamp of the technology and techniques is apparent in the kind of material and the way it is manipulated, whereas in the naturalist work there will be some attempt to hide techniques, and avoid exposing technological signifiers” \[10\].

Spaced Images with Noise and Lines could accurately be described as an interventionist space-form composition. From the outset, the spatial aspect of the work is kept in the compositional foreground through dramatic panning trajectories. Sonic elements are deliberately kept simple throughout in order to highlight the spatial aspect. Different approaches to texture creation, and the admixture of noise and lines in the images, provide a broad variety of spatial experiences to the audience. Perhaps most importantly, the spatial structures are integral to the music. The complexity of the spatial relations is built into the musical structure; it could not have been imposed externally through performed manual diffusion techniques.

10. FUTURE WORK

Only greyscale image processing was used in Spaced Images with Noise and Lines. It would be interesting to use the separate RGB channels to affect different aspects of synthesis, and/or spatialization. The use of image masks could facilitate the creation of various forms of spatial polyphony. The mapping described here makes the most sense in the context of 2D speaker arrays. A rethink of the spatial image mapping strategy may be appropriate for 3D arrays, such as the Klangdom at ZKM, or the Sonic Lab at SARC.

11. CONCLUSIONS

The technique of image-based spatialization was discussed in the context of the author’s composition Spaced Images with Noise and Lines. The creative potentials of the technique were shown to be somewhat different from more traditional trajectory-based spatialization techniques. Given the growing number of multi-channel performance venues, and opportunities for composers and audiences to explore spatial music, techniques such as image-based spatialization may be of increasing interest to composers of computer music.

12. REFERENCES