HARMONIC SHUFFLING: WAVEFORM AND IMAGE RESYNTHESIS IN *THAT WHICH PULLS*

Alexander Dupuis
Independent researcher
Hanover, New Hampshire
Alexander.P.Dupuis@gmail.com
alexanderdupuis.com

ABSTRACT

Harmonic shuffling is a new method for dynamically remixing audiovisual data. Derived from John Whitney's principles of differential motion, harmonic shuffling redistributes a source's data values at harmonically related distances within an array, harnessing the embedded structural relationships to generate new content. This technique can be applied to a wide variety of musical and visual sources, and its usefulness for animating images and resynthesizing audio is examined. The application of harmonic shuffling in the creation of the video piece *That Which Pulls* is given as an example of its uses in a multimedia compositional process.

1. INTRODUCTION

The musical insights of John Whitney have been largely overshadowed by his well-acknowledged successes as an animator. Whitney's theories on music unapologetically champion elementary harmonic and rhythmic principles at the expense of much avant-garde composition [3], suggesting that despite his revolutionary visual work his musical convictions are at odds with modern theoretical discourse. Nonetheless, his animation strategies represent a remarkable success of the translation of his idealized musical principles to the visual realm. By applying simple harmonic relationships to the speeds and trajectories of moving elements, formal patterns of tension and release emerge that are analogous to development and resolution within Western harmonic music. Whitney recognized that we do not experience music as a moment-to-moment apprehension of its constituent elements, but rather as a constantly evolving set of hierarchical relationships that can enhance our perception of organized form. By constructing motion patterns based on the fundamental relationships underlying his preferred musical forms rather than attempting to represent the musical components themselves, Whitney established a set of principles and techniques that he hoped would become part of a revolution in audiovisual music [2].

Whitney admitted that his own applications of harmonic motion principles encompass a fraction of their possible artistic uses. While his visuals are created using a generative approach for positioning graphical elements in two-dimensional space, the underlying concept of harmonically structured motion can be applied to data of any dimension organized within a finite space. Expanding these ideas to include the restructuring and resynthesis of existing data opens up numerous avenues of content generation and composition, approaches that can be equally applied to images, movies, audio files, or digital musical scores. This paper focuses specifically on a method I've adapted from Whitney's ideas, dubbed "harmonic shuffling", and its applications as a method of resynthesizing existing image and audio data. The results of the initial experimental stages are described, as well as the use of harmonic shuffling within the composition of an audiovisual piece, *That Which Pulls*.

2. METHODS

2.1. Harmonic Shuffling

Whitney's animations exploited the dynamic forms that emerge from a technique he calls "differential motion." He defines differential motion as the advancement of a set of graphical elements using a motion function, moving successive elements at iteratively larger speeds which are multiples of a common fundamental rate. A simple example might posit that if the first element moves at a given rate, the second moves twice as fast, the third three times as fast, and so on. By combining precisely related speeds with a shared motion function, order/disorder dynamics develop as perceivable forms coalesce and disintegrate. Within a complete cycle, elements align and separate at different nodes based on subdivisions of the path's periodic length, creating an effect of tension and release that Whitney calls "harmonic resonance" [2].

"Harmonic shuffling" is a particular instance of this sort of differential motion applied to an array of data. While Whitney's work differentially moves graphical elements within a two-dimensional space, harmonic shuffling moves the elements of a one-dimensional data set within the space of its array. This allows for the manipulation of the elements of a pre-existing audio or visual object, such as the pixels of an image or the samples of an audio waveform. Harmonic shuffling thereby differs from harmonic motion in that it reconfigures the components of an existing perceptual object rather than forming one from scratch. The process becomes one of remixing or resynthesis, breaking down the original objects into noise and coalescing their data into new percepts.

The method for generating harmonically shuffled data is as follows: assume an array S of length L containing the source data, and an array T of equal length that will hold the altered material. For each element S_i, where i is the index from 1 to L, generate a target index j using the following formula:

$$j = i \cdot (aL + b) \pmod{L} \tag{1}$$

Following this calculation, set T_j to S_i to shuffle the datum.

The variables a and b, both non-negative real numbers, represent the points of control over the shuffling process. As the value of a approaches a simple integer ratio, the values of the resulting target array T will align along equally spaced harmonic nodes, with the number of nodes being equivalent to the denominator of a. The variable b determines the spread of the resulting data, with lower values clustering the data around the harmonic nodes and higher values spreading the data out. Incrementing values of a and/or b by small amounts to approach or leave these nodes generates points of harmonic resonance, with the specific perceptual effects varying based on the medium and source.

2.2. Implementation

The audio processing software was implemented in the Max/MSP programming environment. An audio sample is loaded into a source buffer, and several buffers of equivalent length are initialized to hold the shuffled samples. Samples from the source buffer are inserted at new positions in a given target buffer based on the results of Equation 1, with the modulo wrapping samples back to the beginning of the buffer when their calculated position exceeds the buffer's length. Each target can be played back in real-time, and all buffers can be synchronized at harmonically related speeds. The a and b values for each target can also be manipulated or iterated in real-time, and the resulting audio can be computed in the background and swapped in at the moment when the buffer playback loops. Continuous updating and playback of the target buffers based on changes in the a and b values is also implemented as an alternate option to the swapping method, though it was not used in the resulting piece.

The code for harmonic shuffling of an image was written using Processing. Although an image is rendered onscreen as a two-dimensional grid of pixels, the pixel data is actually stored as a one-dimensional array and can therefore be manipulated using the harmonic shuffling method. The array is structured such that a pixel located at the onscreen coordinate (x, y) is stored at the array index $y \cdot w + x$, where w is the width of the image in pixels. The array can therefore be thought of as the rows of the image placed end to end in a straight line, and progressing consecutively through the array indices produces a zig-zagging pattern starting from the upper left of the image and ending in the lower right. Shuffling the pixel data moves each pixel along this path, looping from the bottom right to the top left when the array's bounds are exceeded.

The Processing sketch creates a target image buffer to hold this reshuffled data and draws the target to the screen at a designated frame rate. As with the audio, the a and b values can be modified in real-time by the composer or rendered using iterative functions of the composer's choosing. These values can be linked between the audio and image software via Open Sound Control to create parallel audiovisual shuffling.

The wrapping of large target indices with the harmonic shuffling function creates the situation where multiple shuffled data might occupy the same position within the output array. In the case of an image or an audio file, only one value can be output for each index of the shuffled array, so a choice must be made about how to resolve any conflicts with overlapping content. Both the audio and video patches have options to choose between outputting the largest of the conflicting values or the value from the largest source index i. The two results produce, at times, quite different output depending on the source materials, and the latter option proved important in shaping the final audio content of the piece.

2.3. Findings

Preliminary experiments were conducted by generating shuffled audio and video from a variety of sources, with each source shuffled using multiple methods for iterating the a and b values. The first approach incremented the a values from zero to one at a constant rate. This was followed by iterating a through a function that would ease in and out of selected resonant nodes (given as whole-number ratios) falling between zero and one. Finally, audiovisual content was generated by skipping directly between consecutive resonant nodes based on a given denominator. These incrementation strategies were run multiple times using various whole-number ratios between the a and b values. Each value of a produced a single shuffled source image and audio buffer, and an audio stream recording the real-time shuffling over the course of each iteration was output.

Although these experiments were not remotely exhaustive of the vast number of possibilities available, the generated results revealed significant differences in the effectiveness of these approaches when used on different types of audiovisual sources. The video output, though generated through image remixing rather than animation of objects, retains many of the perceptual characteristics of animated differential motion and its two-dimensional patterns of tension and release. Along the expected nodes (whole-number ratios of a) the target indices align into columns, much like objects in Whitney's animations will align at certain spatial locations when traversing identical paths. Other points of alignment occur between these nodes, such as a values that form a whole number ratio with the width w in the denominator. These latter points of resonance can render variations on the original content of the image, with the lowest fraction $1/w$ regenerating the original content.

Rendering audio in this manner leads to decidedly different perceptual results. While the image sequences exhibit overlapping structural evolutions that manifest at different speeds, constant iteration of the *a* value for an audio sample of several seconds results in a target buffer that quickly devolves into unintelligible noise as *a* moves away from nearby resonant nodes. Part of this perceptual disjunction might be that the localized motion exhibited in shuffled video does not translate meaningfully at the audio sample scale, passing too quickly for the listener to hear. Another cross-modal difference is the time needed to apprehend the total result of a shuffle: a sonic object is comprehended over the length of its playback, while an image can be viewed in its totality almost instantly. If we reshuffle a waveform mid-playback, any part of the previous waveform that was not played is lost to the listener. For a waveform with a length of several seconds, shuffling at the same frequency as the video framerate (anywhere from ten to sixty times a second) misses the vast majority of the shuffled data. These two restrictions led me to believe that the most intriguing sonic results would be attained by focusing the audio shuffling on the resonant nodes, rendering out the complete target buffers for the lower whole-integer ratios of *a*.

However, when the value of *b* is zero these whole-integer ratio values of *a* cluster the audio data into columns, much like the spatial alignments within the video. While this clustering is part of the tension-release dynamic that works so effectively for the visuals, within a static audio buffer the effect is the removal of almost all the sonic information, reducing the output to a series of clicks along the nodes. Increasing the *b* values reduces this clustering by spreading the sample information away from the nodes, revealing the shuffled waveform structures to the listener. As the value of *b* approaches that of *a*, the samples become redistributed to form multiple degraded copies of the source buffer, with the number of copies equivalent to the number of nodes. The higher the denominator of *a* and *b*, the greater the number of nodes, the higher the pitch and speed, and the greater the degradation as the original samples increasingly separate and become discontinuous.

As the value of *b* increases beyond that of *a*, these nodal reductions begin to stretch and overlap one another. When the denominators of *a* and *b* are multiples of a common fundamental, the resulting overlapping sections temporally align at evenly spaced positions. The overlapping of the samples within these sections reduces the data within the reshuffled buffer, creating syncopated rhythms as well as variations in timbre and amplitude. Incrementing through whole-number numerators of *a* steps through the various shuffles that align around these nodes, leading to different rhythms as the samples overlap in unique ways - see Figure 1 for a hypothetical example. These possible combinations of *a* and *b* values thereby allow for a rich set of options for generating harmonically related variations on a single audio source.

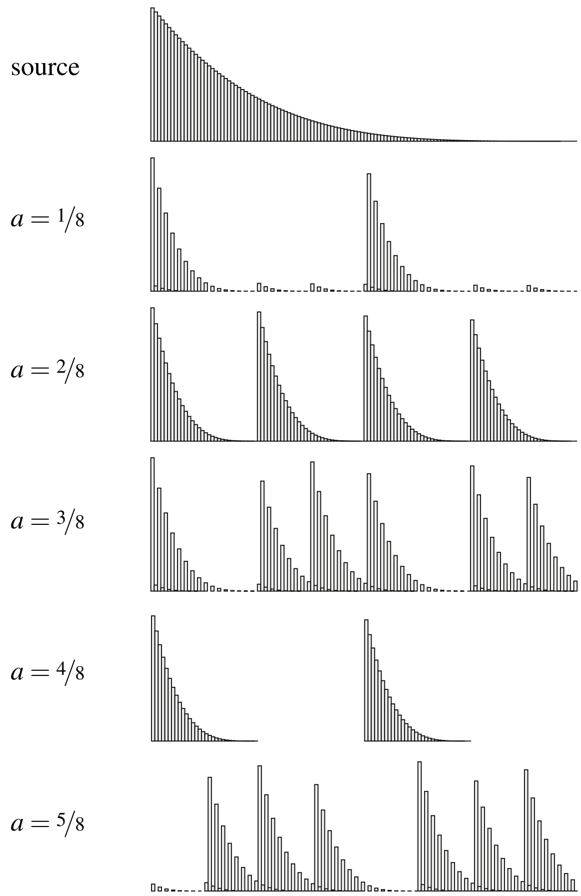

Figure 1. Example of the syncopative effects of shuffling where the denominators of *a* and *b* share a common power. $b = 1/4$ for all examples, save the first which is given to show the structure of the source data.

2.4. Composition

The results of the experimental stage informed the generation and organization of material for the piece *That Which Pulls*. The piece is constructed from two simple sources chosen based on their success within the experimental stage: a photograph of a car taillight and a recording of a single note from a wind chime. A sample of a single note is an ideal basis for demonstrating the increasing rhythmic and timbral complexity of the shuffling process, and the pitched nature of the wind chime helps the harmonic changes of the shuffles stand out. The percussive quality of its sound also allows the timbre to degrade gracefully as the shuffled discontinuities become more pronounced, taking on a more distorted, enharmonic tone. The photograph contains a high-contrast color palette and simple initial shape, characteristics that help emphasize the increasing complexity over the course of the shuffling process. The image was downsampled to further focus on the motion by abstracting the source image while maintaining the color scheme and shape of the original.

The previous experiments revealed that the audio and video benefitted from different shuffling approaches, precluding the use of linked *a* and *b* values as a unifying compositional strategy. Instead, I generated the audiovisual materials separately, using the approaches found to work best for each respective modality in the course of the experimentation stage. For the visuals this meant creating an automation envelope for altering the a value over time, structuring the visuals around the temporal locations of the resonant nodes. For the audio, I chose to generate each shuffled buffer in its entirety, allowing the listener to fully grasp the resultant audio while highlighting the harmonic, rhythmic, and timbral developments of the process that manifest sonically.

The audio was generated as a layered pyramid, with each successive layer's nodes increasing by a power of two. At the base, the a and b values equal 1, generating the source waveform. The upper layers enter in stepwise fashion, with each layer l shuffled upon each repeated loop. By using a values of $r \cdot 2^{-l}$, where r is the current loop iteration, the number of nodes doubles for each successive layer, and the shuffled syncopations change on each repeat as r increases. The value of b for these layers is set to 2^{2-l}, maintaining a constant ratio of $4:1$ between the denominators of a and b. As noted in the findings, the various ratios between the denominators of a and b generate patterns of differing rhythmic complexity and timbre, with larger distances resulting in stretched, distorted tones and lower distances producing more percussive sounds with faster rhythms – see Table 1 for an example. Maintaining a constant ratio of this size allowed the resulting syncopations and timbral subtleties to manifest at proportional scales for each doubling of the nodes.

l	a	b	
0	1	1	
1	2^{-1}	2^{1}	
2	2^{-2}	2^{0}	
3	2^{-3}	2^{-1}	

Table 1. Shuffled results for the first four audio layers on their first repeats (where $r = 1$), showing the syncopations and change in speed that occur as a result of the shuffling approach.

Each buffer was generated at four synchronized octaves, highlighting the increasing rhythmic interplay of the layers and extending the harmonic range. Layer entrances were rhythmically aligned in a stepped pattern, timed so the ninth and final layer entered at the golden mean of the piece's length. After reaching this apex the layers drop out in decreasing order to return to the source sound. The video's envelope was similarly shaped, beginning from the 0^{th} node at $a = 0$, aligning at the source image for the first major entrance of the audio, crossing the fifth harmonic node at the piece's golden mean, and resolving back to the source image when the sound returns to the source wind chimes.

Once the major points of alignment were set, the audio and video content went through alternating stages of editing and refinement to accentuate the cross-modal interplay as well as the dramatic pacing of the piece. Entrances and exits within the audio were adjusted to focus on the developments within the shuffling process, and the approaches and departures from the resonant nodes were adjusted to better emphasize the timing of the tension and release. The dramatic apex of the piece at the six minute mark was highlighted by a negative acceleration of the visuals' a value towards the fifth resonant node, as well as the coincidence of multiple entrances and exits in the audio. The local directions of visual movement were shifted using textural displacement throughout the course of the piece, increasing local disjunction to complement the timbral degradations developing in the audio. Finally, the visuals build towards a lesser peak at the third harmonic before resolving to the source image, with a slowly accelerating overdub of the shuffled audio mirroring the visuals' change in speed. This final revised structure of the piece can be seen in the score excerpt in Figure 2.

By aligning resonant nodes of the video's a envelope with critical points within the audio, the tension-release dynamics within the graphics and sound serve to mutually enhance the strongest dramatic moments. This temporal alignment of audiovisual resolutions perceptually fuses the audio and video in accordance with Chion's idea of *synchresis* [1], establishing moments of cross-modal linking through concurrence of audiovisual events. While synchresis is a powerful effect of our audiovisual perception, strict temporal linking of all audiovisual events can diminish its power within the narrative of a piece and trivialize the sonic and visual relationship. By using the resonant nodes as structural guideposts, intermittent visual resolutions interweave with the sonic pacing rather than exactly mimicking its form, creating a propulsive counterpoint between the two modalities as their internal developments leapfrog each other.

3. DISCUSSION AND FUTURE WORK

The aesthetic and perceptual results of *That Which Pulls* offer a promising look at the applicability of harmonic shuffling as a resynthesis and remixing tool while touching on the disparities than can surface when applying identical shuffling processes to audio and image sources. Sounds and images reveal different insights into harmonic shuffling through their perceptual manifestations of its changes, and in turn the shuffling process offers new perspectives on the information inherent in these sources as well as the communicative advantages and disadvantages of the two mediums.

The tests and discoveries of the composition phase unearthed a number of different applications and directions

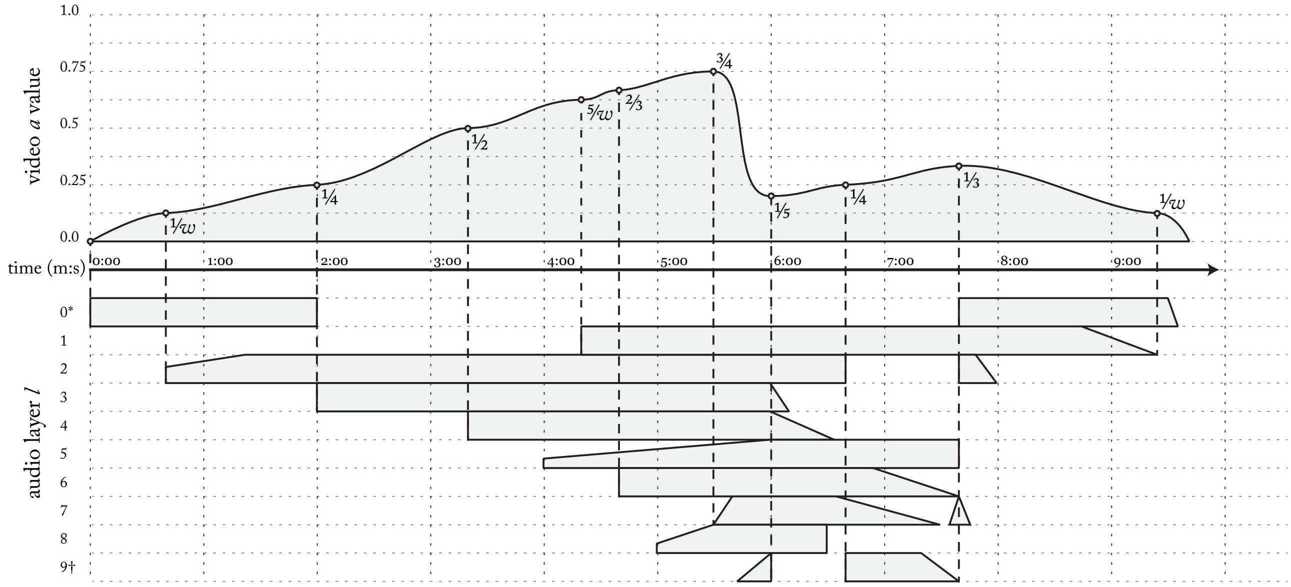

Figure 2. Diagram of the structure of *That Which Pulls*, taken from the score.

that remain to be explored. Harmonic shuffling might be applied to shorter waveforms as a more direct audio synthesis method, or to video data as a visual restructuring tool in the temporal domain. Should either of these experiments prove fruitful, they might open the door towards audiovisual works which can effectively utilize the same harmonic shuffling processes on both the sound and image data, possibly as a realtime performance system. Harmonic shuffling might also be applied to higher-order elements, such as sequences of a waveform or notes of a piece of music. *That Which Pulls* represents a first step towards these applications of harmonic shuffling as a compositional tool.

4. REFERENCES

[1] M. Chion, *Audio-vision: Sound on Screen*. Columbia University Press, Jan. 1994.

[2] J. H. Whitney, *Digital harmony: on the complementarity of music and visual art*. Peterborough, USA: Byte Books, 1990.

[3] G. Youngblood, *Expanded cinema*. New York: P. Dutton & Co., Inc, 1970.