MIDI.CITI: Designing an Experience-oriented Musical Cityscape

Kunwoo Kim
CCRMA, Stanford University
kunwoo@ccrma.stanford.edu

Ge Wang
CCRMA, Stanford University
ge@ccrma.stanford.edu

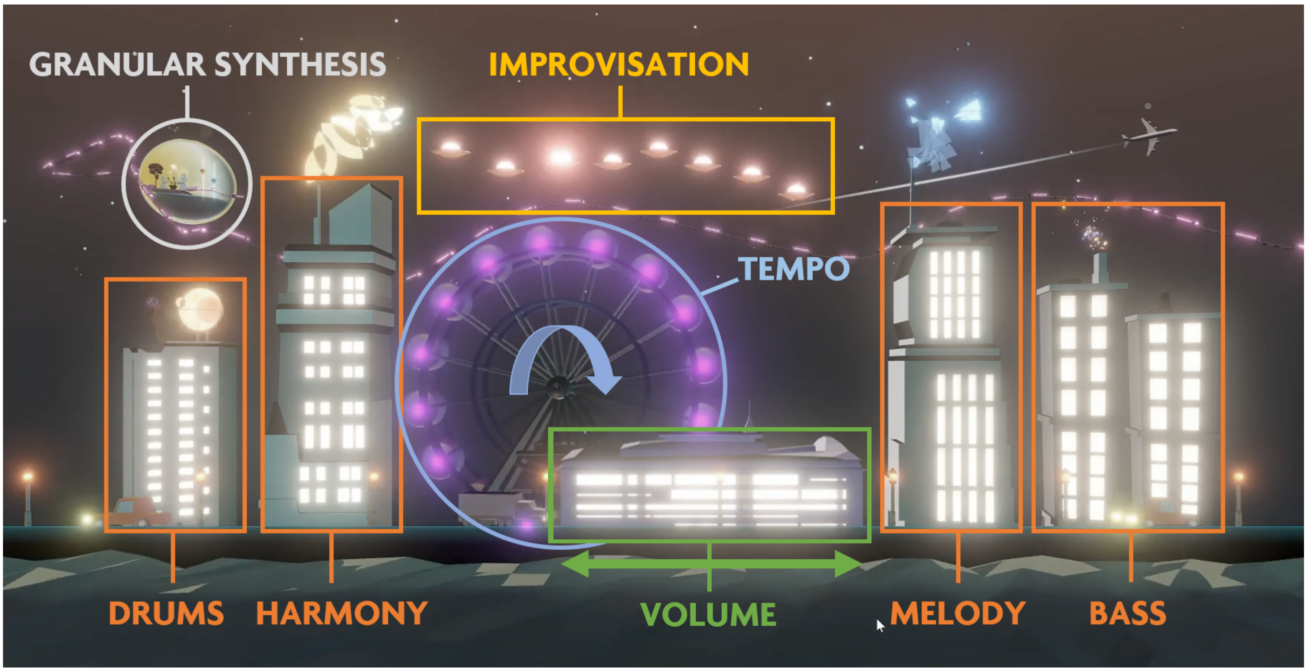

Figure 1: Audiovisual mapping of MIDI.CITI.

ABSTRACT

MIDI.CITI is an interactive audiovisual musical sandbox that offers room for playfulness, expression, and experiential narrative. It contains a real-time algorithmically generated drum machine mapped onto a metaphorical cityscape environment. This paper unpacks the design of MIDI.CITI through the lenses of interaction, play, and designing tools "inside-out", i.e., designing outward from an intended aesthetic experience. We describe its design process as well as its interactions and audio algorithms. Lastly, we put forth a few underlying design principles, including prioritizing the experience, adopting an audio-first approach for tightly-coupled audiovisual correspondence, taking advantage of real-time generative audio, and finding a balance between high-level and direct control. Through these discussions, we aim to provide "things to think with" for creating experience-oriented interactive audiovisual software.

1. INTRODUCTION

MIDI.CITI is an interactive musical sandbox, designed for users to have a playful and expressive experience as they experiment with musical ideas and visual metaphors embedded within a cityscape interface. Despite its name, the software does not use MIDI; it was chosen simply because it rhymes with "city". Instead, MIDI.CITI uses ChucK [1] for generative audio, and is interconnected with graphical and interactive elements using Chunity [2].

Algorithmic processes in MIDI.CITI produce semi-automated musical events that are tightly coupled with the visuals. Chunity's strongly-timed mechanisms allow real-time audio synthesis in tandem with graphics, while precisely governing the audiovisual timing between the two. The synergy of the generative audio, precisely coupled graphics, and interactive elements curates a unique combined experience. MIDI.CITI represents an adaptable design paradigm that can be used for other desktop or VR environments [3].

MIDI.CITI adopts the lenses and design principles of Artful Design [4], aiming to build tools that attend to aesthetic considerations. This translates to critical questions in the design of MIDI.CITI: how might we effectively connect sound, visuals, and interactions to create an expressive audiovisual experience? How can we make use of these underlying mechanics and dynamics to foster playfulness?

In this paper, we unpack the design of MIDI.CITI and its constituent components and a few relevant principles for designing experience-oriented audiovisual tools.

Demo video: https://kunwookim.com/midi.citi

2. MIDI.CITI

The urban elements of MIDI.CITI represent instruments and audio parameters such as tempo and volume. The overall music unfolds as a set of algorithmically-generated audio sequences, programmed in ChucK. The algorithm introduces pseudo-randomization of rhythms and pitches, ensuring sustained engagement across iterations. In tandem with the development of musical layers, the city becomes more visually vibrant, with the addition of lights, cars, planes, trains, and even rabbits on the Moon and UFOs. The rhythm, harmony, and meter are readily modifiable in the ChucK script, facilitating further customization and experimentation.

2.1 The Four high-rise buildings

Users can control the rhythmic complexity of the four different instruments—drums, harmony, melody, and bass—by raising or lowering the amount of window lights in corresponding high-rise buildings (orange in Figure 1). For instance, to intensify drum beats while minimizing melody lines, users can increase the window lights on the drum building, while decreasing lights on the melody building. Moreover, the overall musical sum of the generated loops (e.g., 8 measures, each comprising 12 subdivided beats) maintains uniqueness across iterations through a pseudo-randomized algorithm.

The adjustment of rhythmic complexity is facilitated by the density parameter. First, each instrument is assigned a "trigger array," where each element denotes one of three states: 0 ("do not play"), 1 ("maybe play"), and 2 ("always play"), corresponding to the beat subdivision. Second, the density parameter ranges from 0.0 to 1.0, with unlit windows representing 0.0 and all lit windows indicating 1.0. When the trigger array element is set to 1 ("maybe play"), the density parameter effectively serves as the probability of note generation. Thus, while 0s and 2s establish a fundamental rhythmic structure for an instrument, 1s introduce interesting rhythmic nuances to the sequence.

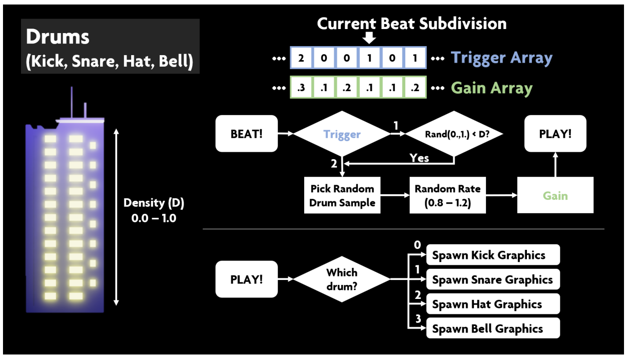

Figure 2. The algorithmic flow of note generation for drums.

The drums consist of kick, snare, hat, and bell, each with a preassigned trigger array (Figure 2). Gain arrays dictate the respective playback volume for each beat subdivision, differentiating between strong and weak beats. Upon receiving a playback signal, the algorithm initiates additional randomization procedures such as randomized sample and playback rate (0.8 to 1.2).

The harmony consists of four `Rhodey` unit generators from ChucK, which establish a four-part chord structure. At the beginning of each measure, it selects one of three trigger array patterns for the measure to maintain rhythmic diversity. With each playback, it generates a chord in either open or closed position, further introducing variability in the accompaniment.

The melody is synthesized using the `ModalBar` unit generator. At the start of each measure, it selects one of five trigger array patterns. Unlike other instruments, the melody exhibits more random rhythmic patterns as its trigger array elements only consist of 0s ("do not play") and 1s ("maybe play"), with no 2s ("always play"). During each playback, the algorithm selects a pitch from a pentatonic scale spanning across three octaves, adapting to the current harmonic context.

The bass is produced using the `Rhodey` unit generator, employing three different trigger array patterns per measure. Unlike the melody, the bass introduces the least amount of randomness in order to uphold the basic structure of the music. Specifically, the first beat of each measure is always played, while its pitch is always the root of the current harmonic context. For subsequent subdivided beats, the pitch is chosen from scale degrees of 1, 3, or 5.

2.2 The Ferris wheel and the low-rise building

Users can rotate the Ferris wheel using their mouse: clockwise to make it go faster and counter-clockwise to make it go slower. This speed of rotation corresponds to the tempo of the whole musical system by changing the beats per minute (bpm) parameter within the main ChucK script that sends global events to each generative instrument process. The tempo can be subtly influenced by rotating the mouse cursor to reinforce or dampen the rotation of the Ferris wheel. In addition, the users can adjust the amount of window lights of the low-rise building, thereby adjusting the overall volume of the music.

2.3 The Moon

Upon flying to the Moon, it is revealed that two rabbits occupy the Moon's dark side, inspired by the East Asian folklore of Moon rabbits. As they pound their mortars and pestles to the underlying rhythm, a granular synthesis train (i.e. grain tram or "tram-ular" synthesis) departs from the Moon back to the city, a musical and symbolic link between the fantastical and the mundane ebb and flow of urban life. The two air balloons on the Moon each control granular synthesis position and volume of the atmospheric soundscape, consisting of four voices of different pitches (Figure 3). The sample used for granular synthesis is a cross-faded interpolation between orchestral and choral sounds, offering a continuous timbre spectrum.

2.4 The UFOs

Lastly, UFOs can be called to appear over the skyline—not to harm anyone, but to play music together! They are the only non-algorithmic, directly playable instruments in MIDI.CITI. Using a computer keyboard, the user can directly improvise different melodies over the algorithmically generated musicscape.

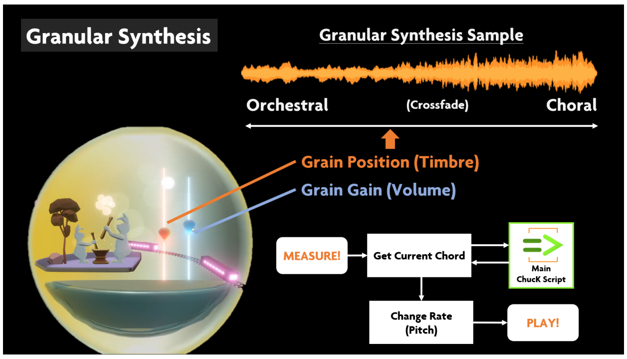

Figure 3. The algorithmic flow of the Moon.

3. DESIGN LENSES

3.1 Interaction

MIDI.CITI's interaction design seeks a balance between automatic generation and direct human control. Principle 5.5 from artful design states, "Have your machine learning—and the human in the loop" [4]. MIDI.CITI uses algorithmically generative audio for its beats, chords, and melodies, but offers a number of simple, yet expressive, metaphorical "knobs" to the user. For example, increasing the density parameter of the building associated with melody introduces additional variations in pitch, rhythm, and dynamics. Similarly, rotating the Ferris Wheel accelerates its rotation speed in tandem with the musical tempo. Together, these simple interactions present many possible recombinations, where the user can create a narrative using the resulting sonic textures. MIDI.CITI adheres to Cook's principle of "instant music, subtlety later" [5], where each interactive urban element allows for instant change in one musical parameter, yet combinations of elements allow for subtle control over the holistic musical outcome.

3.2 Play

Callois states that play is a free, voluntary, uncertain, and unproductive activity, a source of joy and amusement [6]. MIDI.CITI embodies the *paidia* aspect of play, described as "more free-form, expressive, improvisational, recombination of behaviors and meanings" [4]. MIDI.CITI offers iterations of different narratives based on user choice. For example, the densities of different instruments can be maxed out with a fast tempo to achieve exuberance or they can be minimized with a slower tempo to create space for serenity. The users of MIDI.CITI experience open-form aesthetics built on the dynamics of interaction and underlying mechanics of audio algorithms that drive the corresponding graphical output [7].

3.3 Experience

The third lens stems from the idea of designing tools "inside-out" from an intended experience. Artful design's Principle 1.15 states, "Design not only from needs – but from the values behind them", while Principle 2.2 states, "Design inside-out" [4]. MIDI.CITI affords the function of an audiovisual generative drum machine, but was simultaneously designed to prioritize human values such as evoked emotion, playfulness, and expression. As a result, MIDI.CITI aims to follow artful design's principle 1.16, "Design is the radical synthesis of means and ends into a third type of a thing—both useful and beautiful." [4], bringing together complex drum machine algorithms, audiovisual correspondence, and instant yet subtle interactions to ultimately serve an experiential goal.

MIDI.CITI encapsulates its audiovisual interactions within the overarching metaphor of a cityscape, thereby giving form to the experience. It was designed with the conviction that experience-oriented design, with metaphorically aligned audiovisual interactions, have the power to evoke profound reflections and desire for deeper explorations. As one user reflected, it is "a very musical and expressive audiovisual experience that makes me think about the world we live in and how we interact with it." Another user remarked, "I am always left wanting to explore this space more deeply, to go to MIDI.CITI myself and take part in the musical lives of its inhabitants."

4. DESIGN PRINCIPLES

4.1 Don't forget the Experience

Prioritizing experience as the "North Star" of the design is central to value-based design approaches like artful design. This often translates to revisiting a number of questions throughout the design process. What is the end experience? How do we want a person to *feel* within and as a consequence of the experience? How does the system build toward that outcome? How do we invite user participation for play and creative expression?

The metaphor of a musical city provides the framing for MIDI.CITI, where users engage in an experience that is musically creative, playful, whimsical, and expressive, yet uncomplicated and calm. This set of experiential "North Stars" guides the design choices on not only sonic, graphical, and interactive elements, but also the overall balance, narrative, flow, and presentation. It leads to specific design questions that ultimately become the substance of the experience: What if the buildings represented the four instruments in the drum machine? What if the parameters were controlled by window lights? What if a Ferris wheel controlled the tempo? What if we could fly to the Moon? All the while, the "North Star" serves as a set of experiential goals to evaluate each design choice.

4.2 Achieve precise audiovisual correspondence.

A precise correspondence between the audio and the visual promotes functional understanding of the system and enhances the overall aesthetic experience. In MIDI.CITI, every sound the user hears is visualized in the scene with precise audiovisual synchronization (and we can think of the reverse as being true: each visual event is sonified). Whether they are particle effects coming from the rooftops or the animations of rabbits on the Moon, we take the advantage of strongly-timed audio provided by ChucK and use it to precisely drive both audio and graphics. In particular, MIDI.CITI's ChucK code generates audio while sending relevant parameter values and graphical triggers to Unity, enabling a visual element to occur in the nearest frame after a sound has been played. In this way, the audio retains its continuous generation while precisely synchronized with the graphics. The key takeaway here is to synchronize both sound and graphics *from audio*. Indeed, it would have been insufficient here to drive such precise audiovisual synchronization from graphics. A typical video frame rate of 30 fps to 60 fps may not be accurate enough to control a continuous stream of generative audio and its varying nature may result in asynchrony, jitter, or other undesirable artifacts.

4.3 Take advantage of synthesized audio

Synthesized audio has many advantages over pre-recorded audio in interactive software. Various low-level control of audio synthesis parameters including filters and effects, as well as high-level musical parameters such as tempo, dynamics, texture, and rhythmic complexity can be manipulated expressively.

In MIDI.CITI, by introducing pseudo-randomization, the looped sequences of music is rarely repeated in the same manner, yet it can preserve an aesthetic consistency by asserting a prescribed parametric boundary. Also, granular synthesis is used to enrich the overall sonic atmosphere, allowing the user to *resynthesize* a sound stream that interpolates between orchestral and choral timbres.

4.4 Combine high-level and direct control

An important consideration in generative music is finding an appropriate balance between algorithmic processes and meaningful human input. This supports artful design's Principle 7.11A, "That which can be automated should be.", and 7.11B, "That which cannot be *meaningfully* automated should not be". For example, the user has high-level influence over rhythmic variations, musical density, and timbre, but this does not preclude the design from also having direct low-level musical input, such as pitch and note onset.

In MIDI.CITI, the UFOs are mapped in a diatonic scale of the music's overall key and allow for one-to-one action-to-sound interactions. Combined with the semi-automated music environment, this direct-input element enables additional possibilities for play, improvisation, and expression.

4.5 Surprise! Break the established expectation

As users spend time with the design, the mechanics and dynamics of the system get understood and expectations begin to form. An effective way of making the design more interesting is to break such expectations by introducing a novel element that "no one asked for". This can result in a reframing and broadening of the aesthetic experience. For example, in MIDI.CITI, the Moon, initially a visual ornament, serves an unexpected function, wherein the user can fly toward it and interact with Moon rabbits creating grains of sound. Each additional unexpected event builds new layers of mechanics and dynamics, leading to novel ways to aesthetically experience MIDI.CITI.

Breaking the established expectation within the system can be considered a form of *defamiliarization*—a technique that presents a familiar object or situation in an unfamiliar manner to protract the perceptive process and invite fresh context [8]. Such break in expectation sustains the experience, calls attention to its own aesthetics, and invites users to see and hear familiar elements anew.

5. CONCLUSION

As for future work, mechanics for customizing and even programming the musical inner-workings of MIDI.CITI are considered. Currently, the ChucK back-end contains information on probabilities of beats and pseudo-randomized algorithms of pitches. If the user has access to editing these components within the software, it would create another layer for creativity and expression. Moreover, a "free-edit" mode of a city would be an interesting implementation, where the user places different kinds of buildings (or even their own 3D models) and maps them to different audio parameters to customize their own version of a musical city.

In conclusion, we dissected MIDI.CITI as an experience-oriented musical cityscape, unpacking its design through lenses of interaction, play, and experience. We outlined the audio algorithms and pseudo-randomization techniques that aim to create a sustained space for creative engagement. We proposed design principles for using a generative audio-driven system, interconnecting graphics, interactions, and music with the aim of creating a sense of narrative and experiential flow. Overall, if we had a "call to action" in writing this paper, it would be to promote designing more aesthetics-driven audiovisual experiences, as tools for human creativity and expression.

6. REFERENCES

[1] G. Wang, P. R. Cook, and S. Salazar, "Chuck: A strongly timed computer music language," *Computer Music Journal*, vol. 39, no. 4, pp. 10–29, 2015.

[2] J. Atherton and G. Wang, "Chunity: Integrated Audiovisual Programming in Unity," in *NIME*, Conference Proceedings, pp. 102–107.

[3] K. Kim and G. Wang, "VVRMA: VR Field Trip to a Computer Music Center," in *New Interfaces for Musical Expression*, 2024.

[4] G. Wang, *Artful Design: Technology in Search of the Sublime, A MusiComic Manifesto*. Stanford University Press, 2018.

[5] P. Cook, "2001: Principles for designing computer music controllers," *A NIME Reader: Fifteen years of new interfaces for musical expression*, pp. 1–13, 2017.

[6] R. Caillois, *Man, play, and games*. University of Illinois press, 2001.

[7] R. Hunicke, M. LeBlanc, and R. Zubek, "MDA: A formal approach to game design and game research," in *Proceedings of the AAAI Workshop on Challenges in Game AI*, vol. 4. San Jose, CA, Conference Proceedings, p. 1722.

[8] L. Crawford, "Viktor Shklovskij: Diffrance in Defamiliarization," *Comparative Literature*, pp. 209–219, 1984.