Borderlands: An Audiovisual Interface for Granular Synthesis

Chris Carlson
CCRMA - Stanford University
600 Lomita Dr.
Stanford, CA 94305
carlsonc@ccrma.stanford.edu

Ge Wang
CCRMA – Stanford University
600 Lomita Dr.
Stanford, CA 94305
ge@ccrma.stanford.edu

ABSTRACT
Borderlands is a new interface for composing and performing with granular synthesis. The software enables flexible, real-time improvisation and is designed to allow users to engage with sonic material on a fundamental level, breaking free of traditional paradigms for interaction with this technique. The user is envisioned as an organizer of sound, simultaneously assuming the roles of curator, performer, and listener. This paper places the software within the context of painterly interfaces and describes the user interaction design and synthesis methodology.

Keywords
Granular synthesis, painterly interfaces, improvisation, organized sound, NIME, CCRMA

1. MOTIVATION
Max Mathews famously posited that any sound can be represented by a sequence of digits, with the consequence that, given a computer with enough power, the entire audible universe may be explored. In recent years, he was careful to clarify his original theorem, pointing out that most of the sounds that can be made by computers are, "uninteresting, horrible, or downright dangerous to (one's) hearing!" [14]. Mathews emphasized the importance of seeking sounds that are beautiful, while recognizing that the perception of beauty is a subjective matter. This line of research goes hand in hand with John Cage's idea of the composer as an organizer of sound [3] or, as Brian Eno recently suggested, the metaphor of the composer-as-gardener [6]. Eno's metaphor likens the creation of music to the act of planting a variety of seeds, observing their growth and evolution, and iteratively pruning the results into a one of many potential forms. Given that the human mind is underequipped to imagine the entirety of the set of all beautiful sounds, let alone the incredibly vast audible universe, a composer who desires to break new sonic ground often engages in this iterative process. The act of listening is then essential to the generation and evolution of new ideas.

The conceptual framework provided by Mathews, Cage, and Eno forms a foundation for *Borderlands*, which stems from the desire to strip away common musical paradigms such as notation, scales, tracks, and arrangements and allow users simply explore and transform sound. The software exists in a space in which composition, performance, visualization, and,

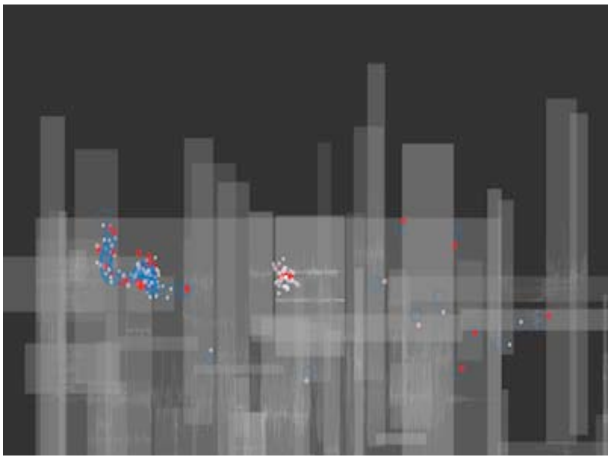

Figure 1. *Borderlands* main interface with multiple grain clouds operating on an overlapping landscape of sounds.

most importantly, listening, overlap. The user is envisioned as an organizer of sound, simultaneously assuming the roles of curator, performer, and listener.

This work is closely related to a class of instruments described in detail by Golan Levin in his Master's thesis, *Painterly Interfaces for Audiovisual Performance.* In his dissertation, Levin enumerates a set of key design goals for the ideal audiovisual interface [quoted from 11]:

- The system makes possible the creation and performance of dynamic imagery and sound, simultaneously, in real-time.
- The system's results are inexhaustible and extremely variable, yet deeply plastic.
- The system's sonic and visual dimensions are commensurately malleable.
- The system eschews the incorporation, to the greatest extent possible, of the arbitrary conventions and idioms of established visual languages, and instead permits the performer to create or superimpose her own.
- The system's basic principles of operation are easy to deduce, while, at the same time, sophisticated expressions are possible and mastery is elusive.

Direct and immediate interaction, expressiveness, and flexibility both sonically and visually – these objectives are difficult to completely realize in every system, but, when considered holistically, the results can be compelling. Tarik Barri's Versum [2], for example, is a hybrid tool in which users compose by building a virtual 3D universe of custom audiovisual objects and perform by flying through the space. In this software, graphics and sound are inextricably linked – the visual is not merely a representation of the sound, it *is* the sound. This system engenders a new compositional process in which the form of a piece is equally influenced by both sound

NIME'12, May 21-23, 2012, University of Michigan, Ann Arbor.

and space. Similarly, Nick Kruge's MadPad [13] is an audiovisual sampler designed for mobile devices. Users are invited to "remix their lives" by recording a set of short video clips and making music with them on a drum machine-like interface. There are exciting possibilities for interaction that result from this approach, but the most interesting outcome is an external one: a refocusing of users' attention on the sounds of daily life - "musique concrete for the masses." Both Versum and MadPad, while aesthetically very different, create new musical experiences through the unification of sound and image into a single audiovisual entity. The mutual reinforcement of aural and visual queues results in deeper, more intuitive interaction for users and provides an entry point into the musical process for observers.

2. GRANULAR SYNTHESIS

In the 1940's, Gabor theorized that sound can be broken down into fundamental particles or *acoustical quanta* [8]. Years later, inspired by Xenakis's description of "grains of sound" in his book *Formalized Music* [20], Curtis Roads wrote some of the first granular synthesis (GS) programs in the 1970's and 80's. [15]. Many techniques emerged from these experiments and the work of others: synchronous GS, asynchronous GS, pulsar synthesis, glisson synthesis, trainlet synthesis, and formant wave function synthesis to name a few. A detailed survey of these and other techniques is available in Roads, 2004 [16]. While these methods have key differences in their execution, all fundamentally involve the generation and playback of amplitude-modulated, or "windowed," fragments of either synthesized or sampled sounds.

The early granular experiments were necessarily performed on non-real time systems, and it was not until the mid 1980's that granular synthesis was realized in a real-time setting with the work of Barry Truax and his piece, *Riverrun* [19]. Today, many real-time implementations exist, including generalized frameworks for building systems in Max [5], CSound [1], and other musical languages as well as fully functional iPhone apps, VST plugins, and various other commercial instruments [9].

Figure 2. Examples of granular synthesis interfaces developed in recent years. Note the prevalence of knobs, dialogs, sliders, XY controllers, and single-waveform displays. Images sourced from [9]

However, of the available software tools that provide graphical interfaces, there are few that enable interactions that go beyond the standard set of knobs, sliders, XY control surfaces, and single-waveform displays. One recent attempt at breaking free of these constraints is the Curtis Heavy iPhone app [4], which maps grain position within a recorded sound to touch input and visualizes the output waveform at the sampling location. Only a single sound can be granulated, however, and rapid sonic shifts are difficult to achieve. IXI Audio's grainBox maps multiple parameters to nodes on a plane that may be repositioned, but the relationship to the source waveform is invisible to the user [10]. Interaction with these and other granular synthesis instruments can be problematic for first time users as it is based on an understanding of the underlying synthesis abstractions. In the context of live performance, interaction with these interfaces is generally hidden from the audience behind the laptop screen. The intricacies of this technique are lost on the uninitiated.

3. BORDERLANDS

Borderlands presents a new interactive visualization of granular synthesis. By directly displaying grains in relation to their sampled waveforms, the software enables direct deconstruction and transformation of sound. From the perspective of the audience, *Borderlands* is capable of providing a link between perceived sounds and the actions of a performer if the interface is displayed.

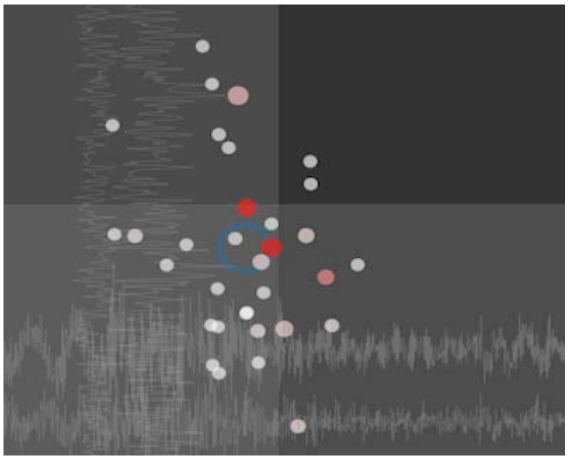

Figure 3. Close-up of the *Borderlands* interface. A cloud of grain voices samples two overlapping audio files.

3.1 User Interaction

Interaction with the current implementation of *Borderlands* is primarily mouse and keyboard based and is designed for efficient modulation of multiple granular sampling entities. Prior to starting up the software, users place any number of .wav and .aif files in a "loops" directory contained in the source distribution. When launched, the software randomly distributes these files, forming a two-dimensional landscape on the screen. Each waveform is constrained to a rectangle, forming a "sound quad" that is oriented either vertically or horizontally. Both stereo and mono waveforms are represented. These rectangles may be selected, moved, resized, and flipped between vertical and horizontal orientations. Two overlapping sound files with different orientations can be seen in Figure 3. Given that it is possible for a large number of overlapping sound files to be present in the interface, a user may cycle selection of all rectangles under the current cursor location (e.g. by pressing the *tab* key).

Mini-granular sampling engines, or "grain clouds" may be added to the screen under the current mouse position with the press of a key. The user has control over the number of voices in a cloud and their positioning as well as various parameters relating to the synthesis. See Figure 3 for an example cloud. In the current implementation, keys for selecting parameters and editing their values are efficiently arranged under a single hand position to enable quick access. By selecting a cloud and moving it over a rectangle, the sound contained in the rectangle will be sampled at the relative position of each grain voice as it is triggered. Visually, the cloud consists of a small blue ring (green when selected) and small, white circular areas representing each voice. The ring surrounds the central point of the cloud and provides an area for selection. It also gives rough feedback regarding the granulation rate through a proportional pulsing along its radial dimension. Each grain voice also pulses, turning red when triggered over a waveform and fading back to white over its duration. The position of each voice is randomized each time it is triggered. The boundaries for this randomization along the X and Y axes are specified by selecting a cloud, selecting the key for the appropriate axis (*x* for X, *y* for Y, and *r* for both), and moving the mouse a distance away from the center of the grain cloud.

Users are able to quickly construct rhythmic and ambient soundscapes by creating and destroying new clouds, modifying the synthesis parameters and voice positions, and dynamically moving each cloud around the audio landscape. Alternatively, the clouds may be statically positioned and the audio files may be moved and reoriented instead. With practice, various performance actions and associated geometries emerge and may be incorporated into the user's performance vocabulary.

3.2 Audiovisual Elements

Borderlands is written in C++, using OpenGL for graphics and RtAudio [17] for the real-time audio processing. As mentioned in the previous section, three key classes of audiovisual entities exist within *Borderlands*: sound quads, grain voices, and grain clouds. These objects are the sole elements with which users interact and directly represent the hierarchy of components involved in the granular sampling process. Waveforms bounded by sound quads are sampled by grain voices, which are controlled and triggered in turn by a parent grain cloud. The number of each of these entities that may be instantiated is limited only by the processing power of the user's computer, allowing performers to build highly dense sonic environments.

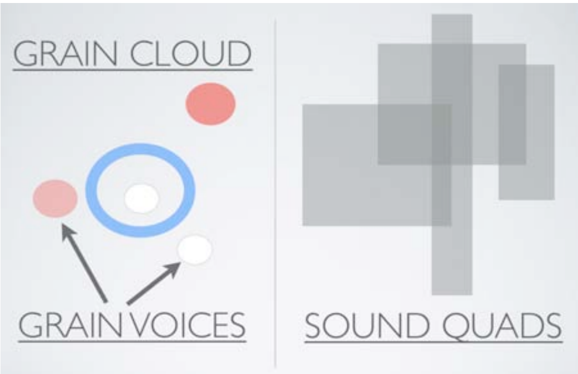

Figure 4. Audiovisual entities in *Borderlands*

3.2.1 Sound Quads

Each transparent, gray rectangle represents a single mono or stereo audio file chosen by the user. When the program is launched, the waveforms are loaded into memory using the libsndfile library, an open-source toolkit for multi-format audio input and output [12]. The length of the audio file influences the sound quad shape. In general, however, the initial dimensions, positioning, and vertical/horizontal orientation are randomized to provide a fresh sense of exploration on successive launches of the software. Each waveform is plotted inside its container, allowing users to take locations of transient events, silence, and soft, and loud passages into consideration when positioning the grain clouds. The ability to adjust the size of the sound quads by clicking, dragging, and holding a key allows users to zoom in on specific events in the audio file, enhancing the temporal precision of the granular sampling. Control over orientation and positioning of the quads enables dynamic shifts and expressive gestures. Users can quickly reconfigure the sonic terrain being sampled by a collection of grain clouds.

3.2.2 Grain Voices

A grain of pre-recorded sound consists of an amplitude-modulated fragment of a source sound. The duration of the grain and the type of modulating window have a significant impact the timbral characteristics of the output. Grains in *Borderlands* are limited to 1 millisecond or greater in duration. A variety of window types are selectable and are shown in Figure 5. The reader is referred to [18] for a detailed treatment of the spectral properties of audio windows.

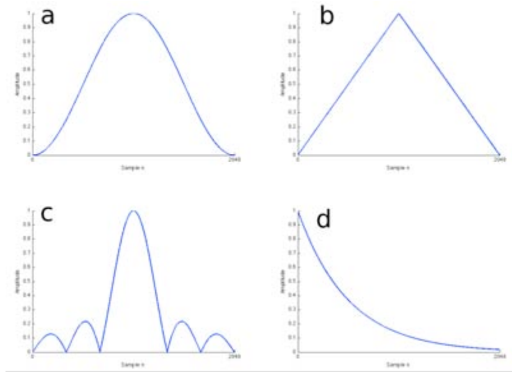

Figure 5. Window types (a) Hanning, (b) Triangle, (c) Sinc, (d) EXPDEC. A time-reversed EXPDEC is also available.

When triggered, a voice will determine its relative location in both dimensions of all sound quads underneath its current position. The location along the time dimension determines the starting point within each file. The location along the orthogonal dimension determines the amplitude of each waveform as it is "polyphonically" mixed. Linear interpolation is implemented for fractional sample indices – this allows the user vary the playback rate or "pitch" of each grain voice. The playback direction may also be changed via keyboard control. As mentioned earlier, each grain voice is visualized as a small white circular that pulses red when triggered over one or more sound quads.

3.2.3 Grain Clouds

Each cloud is represents a miniature granular sampler that manages the temporal and spatial locations of a variable number of voices. The synthesis implementation is an extension of a Pure Data example from Andy Farnell's book, *Designing Sound* [7]. It utilizes a technique called synchronous granular synthesis, in which grains are played in sequence at regular time intervals and may overlap by up to half of their duration. It is at the cloud level that all of the synthesis parameters are exposed and controlled. The number of grains and their duration, window, pitch, playback direction, overlap, and position are determined through by interacting with the selected cloud. Parameters are selected for editing by the keyboard. New values may be entered with numeric keypad or incremented by holding down the parameter key (plus shift for decrementing). Changes are applied globally to all voices in the cloud and are queued so that they may be "picked up" on the next trigger signal for each voice.

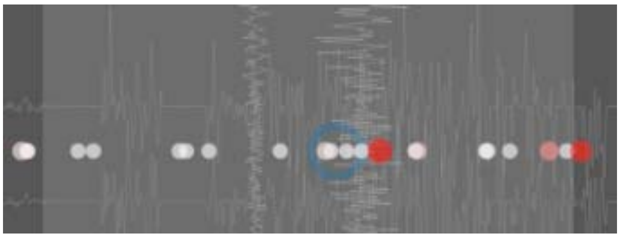

Figure 6. Sampling an instant in time and many locations simultaneously with two overlapping sound quads.

The excitement in performing with *Borderlands* arises from the ability to define a number of grain clouds with different parameter sets and influence their interaction with multiple sound files at once. Suspended textures may be generated using clouds of highly overlapping voices. With one swipe of the mouse, a cloud of voices can be condensed to a single point within a file or expanded to randomly sample the entire visible screen. A single time slice of sound can be sampled by orienting the voices along a one dimension, only to be immediately spread in time by rotating the underlying sound file. Rhythmic sequences can be achieved by using exponentially decaying envelopes, short grain durations with minimal overlap, and random positioning over a sound files and empty space. These are just a few of the sonic gestures made possible by the audiovisual interaction design.

4. FUTURE WORK

From a design perspective, *Borderlands* successfully incorporates many of the criteria envisioned by Golan Levin in his classification of painterly audiovisual interfaces. Real-time dynamic image and sound are central to the software's functionality, and the use of sampled audio and granular synthesis ensures that the space of sonic possibilities is large. Traditional musical structures are abandoned in favor of pure engagement with sound, yet the interface remains direct and flexible. The simplicity of the audiovisual entities along with the unique perspective they provide into the underlying synthesis algorithms immediately focuses users on exploring sound and bypasses the learning curve associated with typical granular synthesis instruments. These factors not only make *Borderlands* a viable new instrument for real-time composition and performance of experimental music, but also position it as a useful pedagogical tool. The wide variety of rhythmic and spectral effects that may be achieved with synchronous granular synthesis is readily demonstrated. Multiple parameter sets are easily configured and may be situated side-by-side to illustrate their differences. Most importantly, the relationship between clouds, grains, and source sound material is made explicit through the audiovisual interface.

Borderlands originated as a class project at Stanford University's Center for Computer Research in Music and Acoustics in the fall of 2011. The software is still in its infancy, and many new features are planned. Usability improvements such as audio export, set saving capabilities, runtime loading and deletion of sound files, and midi and OSC support are envisioned. Enhancements to the overall expressiveness of the instrument are in the works as well. Gesture recording and looping will allow animation of cloud and landscape elements for more intricate, evolving compositions. Tempo-synced grains, both to a master clock and locally between clouds, will enable much more precise sequencing of rhythmic events. Control over the stereo mix of grain voices, clouds, and sound quads will expand the sonic image provide another expressive dimension. We hope to extend this work into the domain of mobile music with the development of a collaborative, multi-touch application. A participatory sound installation is also planned in which gallery visitors assume the role of a single grain cloud as they move through space.

Borderlands is available for download at:
http://ccrma.stanford.edu/~carlsonc/256a/Borderlands/

5. ACKNOWLEDGEMENTS

Many thanks to Jorge Herrera, Mike Rotondo, Hunter McCurry, and Tarik Barri for early feedback on design ideas.

6. REFERENCES

[1] Brandtsegg, O., Johansen, T., Saue, S. Particle synthesis – a unified model for granular synthesis. *Proceedings of the 2011 Linux Audio Conference*. (*LAC '11)*(Maynooth, Ireland, May6-8, 2011).

[2] Barri, T. Audiovisual composing in 3d. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '09)*(Carnegie Mellon University, June 4-6, 2009).

[3] Cage, J. *Silence: Lectures and Writings*. Hanover, NH: Wesleyan University/University Press of New England, 1961.

[4] Curtis Heavy iPhone App. The Strange Agency: http://thestrangeagency.com/products/curtis-heavy/

[5] Eckel, G. Iturbide, M. R., Becker, B. The development of GiST, a Granular Synthesis Toolkit Based on an Extension of the FOF Generator. *Proceedings of the 1995 International Computer Music Conference (ICMC '95)*(San Francisco, CA, 1995).

[6] Eno, B. Composers as Gardeners. http://edge.org/conversation/composers-as-gardeners. 2011.

[7] Farnell, A. *Designing Sound*. MIT Press, 2010. 310-312.

[8] Gabor, D. Acoustical quanta and the theory of hearing. *Nature 159 (4044)*, 1947, 591-594.

[9] Granular Synthesis Software List: http://www.granularsynthesis.com/software.php.

[10] IXI Audio. http://www.ixi-audio.net/content/about.html.

[11] Levin, G. *Painterly Interfaces for Audiovisual Performance*. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[12] libsndfile: http://www.mega-nerd.com/libsndfile/

[13] Kruge, N. Wang, G. MadPad: A Crowdsourcing System for Audiovisual Sampling. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '11)* (Oslo, Norway, May 30 - June1, 2011).

[14] Mathews, M. Quote from presentation at Stanford. http://soundcloud.com/cloudveins/max. Stanford University, October 2010.

[15] Roads, C. Introduction to Granular Synthesis. *Computer Music Journal 12, 2.* Cambridge, MA, 1988.

[16] Roads, C. *Microsound*. MIT Press, Cambridge, MA, 2004.

[17] RtAudio: http://www.music.mcgill.ca/~gary/rtaudio/

[18] Smith, J. *Spectral Audio Signal Processing*. W3K, 2009.

[19] Truax, B. Real-Time Granular Synthesis with a Digital Signal Processor. *Computer Music Journal 12,2.* Cambridge, MA, 1988.

[20] Xenakis, I. *Formalized Music*, Bloomington: Indiana University Press, 1971.