# The Table is The Score: An Augmented-Reality Interface for Real-Time, Tangible, Spectrographic Performance

Golan Levin
School of Art, Carnegie Mellon University
golan [at] andrew.cmu.edu

## Abstract

*Real-time performance instruments for creating and sonifying spectrographic images have generally taken the form of stylus-based drawing interfaces, or camera-based systems which treat a live video image as a spectrogram. Drawing-based approaches afford great precision in specifying the temporal and pitch structures of spectral events, but can be cumbersome, as they only accept input from a single point; camera-based approaches offer quick flexibility in all-around image improvisation, but poor compositional precision because of inadequate visual feedback to the user. In this paper, I present a camera-based spectrographic performance instrument which affords both compositional precision and improvisatory flexibility. This is made possible through an augmented reality (AR) projection overlaid onto and carefully aligned with a dry-erase performance surface.*

## Keywords

Audiovisual performance instrument, augmented reality, spectrographic performance, graphic sound synthesis.

## 1 Introduction

*Spectrograms*, or diagrams which depict the frequency content of sound over time, are a basic visualization tool in computer music and acoustics. Ordinarily, spectrograms are used to analyze pre-existing sounds. Nevertheless, the concept of a composition and performance tool with a spectrographic *input* interface – capable, in theory, of allowing a musician to construct sound entirely from the bottom up – is a recurring one in computer music.

Attempts to build interfaces for spectrographic performance instruments have generally elected to prioritize either compositional precision (with cursors) or improvisatory freedom (with cameras). In this paper, I introduce a solution which I believe offers a good measure of both. To accomplish this, I use techniques borrowed from the field of "augmented reality", which Lev Manovich has defined as the "overlaying of dynamic and context-specific information over the visual field of a user" [8].

In my system, objects placed on a table are interpreted as sound-producing marks in an active spectrographic score. Video projections cast onto this table transform the instrument into a simple augmented reality, in which the users' objects are elaborated through colorful and explanatory graphics. Every point on the table's surface, and each pixel in the camera's view, corresponds to a unique time/frequency possibility, and is performable as such.

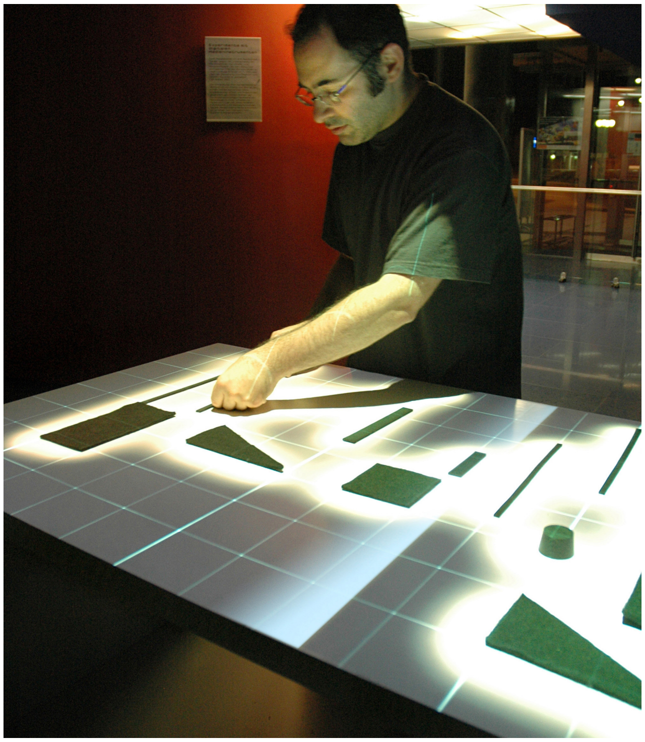

Figure 1. The *Scrapple* spectrographic instrument in use. On the table are a variety of dark rubber and felt objects. The table is also a dry-erase surface and can be scribbled on with conventional whiteboard markers. Note the real-time video projection, from overhead, of various augmented-reality (AR) information layers: a grid representing subdivisions of time and pitch; a "Current-Time Indicator," which scans the table lengthwise; and glowing haloes around the physical objects, indicating successful detection.

## 2 Background

Various implementations of spectrographic sequencers have been created over the past 60 years. In this section I briefly survey a selection of these systems, with an eye towards better understanding the tradeoff between compositional precision and real-time instrumentality.

## 2.1 Background: Performing Spectrograms

The first machine for reconstructing sound from spectrographic images appears to be the *Pattern Playback* machine built by speech researcher Franklin S. Cooper at Haskins Laboratories in the late 1940s. In this system, spectrographic sound patterns are hand-copied in white paint onto an acetate belt, and then conveyed at seven inches per second past a photoelectric sensor. Simultaneously, an intense slit of light from a mercury-arc lamp is focused onto a rapidly rotating "tone wheel." This disk, which has 50 concentric variably-spaced apertures, admits light at a variety of periodic intervals (ranging from 120 to 6000 Hz) onto the belt. Light modulated by the wheel and directed onto the spectrogram belt thus reflects to the photocell only those portions of the light which carry the frequencies corresponding to the painted pattern [2,10]. Signals from the photocell are then amplified and directed to a loudspeaker.

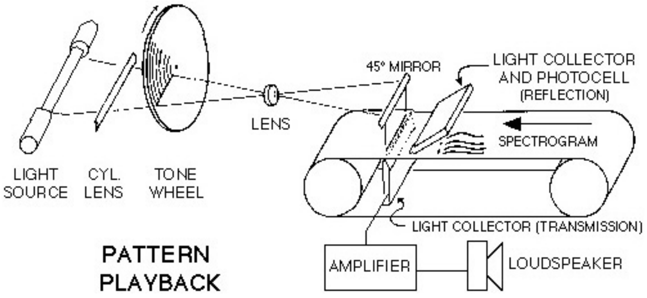

Figure 2. Cooper's 1951 *Pattern Playback* system. From [10].

Cooper's *Pattern Playback* machine continued to find use in audio perception studies as late as 1976; the original device, which is still operational, now resides in the Haskins Laboratories Museum in New Haven, Connecticut [10].

Figure 3. Cooper's 1951 machine as seen today. From [10].

A significant limitation of this optomechanical device is that it could only be used, as Cooper's title suggests, for spectrographic *playback*. With the introduction of real-time digital audio synthesis, two main interface paradigms have arisen to enable *live* improvisation with spectrographic images: *drawing-based* and *camera-based* interfaces.

Iannis Xenakis' *UPIC* system, first realized in 1977, is emblematic of the former. Consisting of a graphics tablet interfaced to an HP computer, users of the *UPIC* could gesturally create, edit and store spectral events with unprecedented precision. By 1988, a version developed by Raczinski, Marino and Serra allowed users to draw and listen to spectrograms simultaneously and in real-time [9]. The core *UPIC* interface concept has been maintained in the popular *Metasynth* software [12], and extended in my own *Yellowtail* [7], wherein the user can draw procedurally *animated* marks into a real-time spectrographic score.

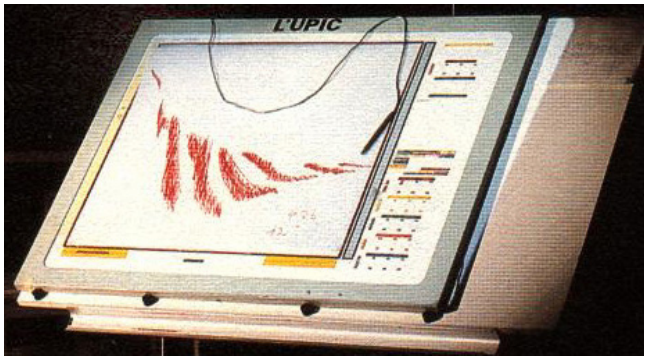

Figure 4. Iannis Xenakis' 1977 *UPIC* system. From [11].

The use of a camera to interactively 'perform an image' – rather than a single-point cursor – forms the second main interaction paradigm for live spectrographic sequencers. An early real-time implementation of this was developed by Finnish artist-researcher Erkki Kurenniemi in his 1971 *DIMI-O* ("Digital Music Instrument, Optical Input") system, which simply treated a live video image as if it were a spectrogram. In this system, a graphical "current time indicator" scanned the live video image from left to right; when this indicator overlapped a sufficiently dark or light video pixel, a synthesizer generated a chromatic tone whose pitch was mapped to the vertical coordinate of the pixel [5]. A modern implementation of this concept can be found in the *Additive Synthesis* demo patch which ships with Cycling74's *Jitter* toolkit [4]. The project described in this paper is related to these priors, but uses an AR projection to provide precise visual feedback to the user.

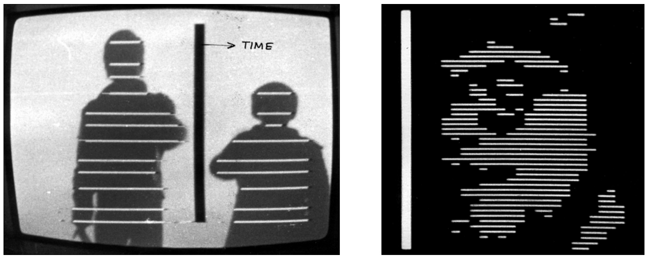

Figure 5. Erkki Kurenniemi's 1971 *DIMI-O* system. From [6].

# 3 The *Scrapple* Instrument

## 3.1 Overview: The Table is the (Active) Score

The spectrographic performance instrument described in this paper, *Scrapple,* consists of a Windows PC, custom software, a 2-to-3m long table covered with a dry-erase board (which serves as the primary user interface), and a digital video camera which observes the table from above. Users perform the instrument by drawing or erasing marks on the table's whiteboard surface, and/or by rearranging a collection of variously shaped tactile objects on the table. Scanning the table lengthwise, the system synthesizes sound in real-time by interpreting all objects or marks as sound-events in a periodically looping spectrographic score.

Objects arranged from left to right (along the length of the table) are sonified sequentially in time, while objects arranged across the width of the table generate tones whose frequencies span eight octaves from low to high. Darker objects produce louder tones; objects covering a larger area generate a correspondingly wider range of frequencies, producing e.g. dissonant chord clusters or noise bursts. *Scrapple* makes use of a variety of playful objects; for example, flexible shape-holding architects' curves allow for the creation of easy-to-transpose melodic patterns, while small wind-up toys yield ever-changing rhythms.

*Scrapple* scans the table every few seconds, producing a looping sound pattern. The loop's tempo can be varied (by means of a separate knob device) between 1 and 1000 bpm.

### 3.2 An Augmented Reality Interface

The core innovation of the *Scrapple* instrument is the use of an “augmented reality” (AR) technique to provide essential visual feedback to the user about their actions and the state of the system. The AR takes the form of a layer of real-time computer graphics projected onto the table from above, in a conjoined camera/projector configuration which John Underkoffler has termed an “I/O Bulb” [13]. *Scrapple's* video projection, which is carefully calibrated and registered with its table, delivers three kinds of *in-situ* contextual information to the user:

1. **A visualization of the position of the instrument's *Current Time Indicator.*** This graphic (a sliding, glowing bar) is precisely coupled to the virtual “play head” of the additive synthesizer which scans across the spectrogram. As a result: at the exact moment when this glowing bar passes over a mark on the table, the corresponding sound of that mark is produced and heard.
2. **A grid which marks off helpful subdivisions of pitch and time.** The default grid indicates octaves and $32^{nds}$ of the table's loop period, but this can be exchanged with other grids according to user preference (e.g., in order to represent triplet meters or pentatonic scales). Because this grid is only a software projection rather than a hardware constraint, users can choose whether or not to use it as a guide for positioning their marks.
3. **A glowing halo around each mark or object on the table.** The presence of this halo indicates that the mark or object has been successfully detected by the software. This is important because light-colored or very tiny marks (e.g. smaller than ~3mm square) may elude the vision system's noise-thresholder (described below). An object's halo fades away gradually after it is sonified; thus the intensity of haloes across the table is also a visual cue about the system's timing and tempo.

*Scrapple*'s augmented reality projection permits a very broad range of interactions, from the compositional precision of drawing-based systems to the improvisatory freedom of camera-based interfaces. The system's projected grids, for example, allow for very careful estimation and placement of detailed spectral markings. At the same time, it is possible to perform *Scrapple* intuitively and effectively by simply passing one's hands in the path of its projected Current Time Indicator.

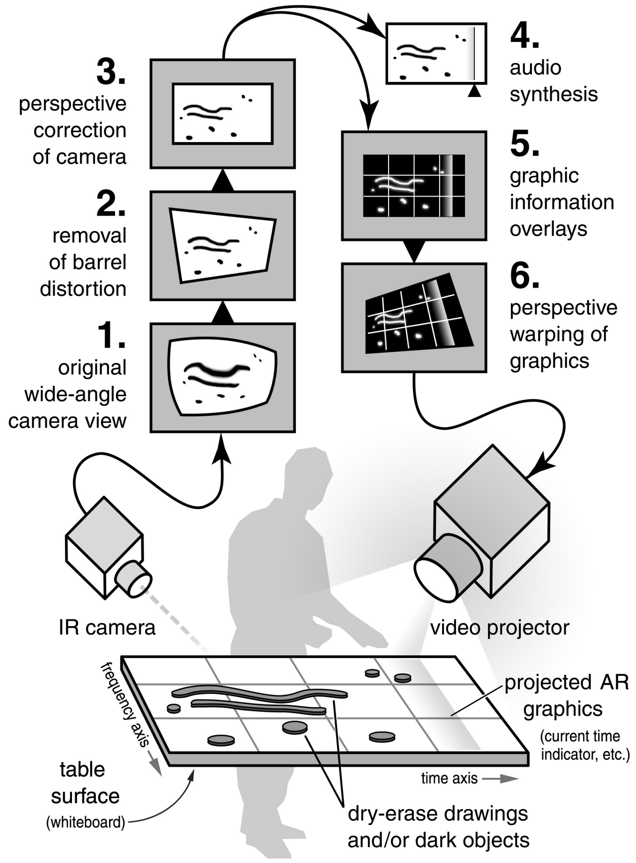

Figure 6. Overview of the *Scrapple* installation.

## 4 Implementation

In the implementation of *Scrapple*, special attention has been given to image processing in order to ensure that the captured spectrographic image is spatially regular and clean. The first few steps, illustrated above, attempt to compensate for inevitable real-world distortions in the captured score due to perspectival offsets and imperfect optics in the camera. These steps are essential in order for the projected AR graphics to coincide accurately with the objects and marks on the table. The first step eliminates radial “barrel” distortion caused by the camera's lens (using Bourke’s method from [1] ); after this, an inverse perspective warp is used to derive a purely rectangular (perspective-free) score image. All warping is performed with bicubic interpolation in order to reduce the effects of pixel aliasing. The final sampling resolution for a 200x50cm table is roughly 4 to 5 camera-pixels per table-centimeter, which is sufficient to capture the marks created by standard whiteboard markers.

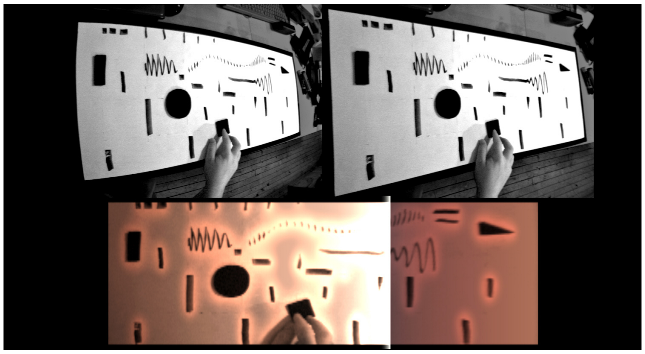

Figure 7. Image processing in *Scrapple*.

Since every pixel observed by the video camera is potentially a sound generator, some form of noise suppression becomes necessary to reduce the sonic consequences of video noise, and to insulate the system against fluctuations in environmental light. To accomplish this, *Scrapple* uses an adaptive thresholding algorithm as described in [3], with the tradeoff that very pale or small objects may elude this thresholder (as mentioned earlier).

Spectrographic synthesis is accomplished with an additive synthesizer, which sums a series of equal-tempered sine waves spanning the 8 octaves from 32 to 8192Hz. The number of frequency bins is adjustable, ranging from a familiar-sounding 12 tones per octave to a practical limit of about 64. The amplitude of each tone is exactly governed by the darkness of a (bicubically interpolated) row of pixels. To appeal better to the ear, the sine tones are pre-scaled by the Fletcher-Munson isoloudness contours, and those below 100Hz are brightened with a tiny bit of harmonic coloration.

The projected computer graphics are based on an inverted version of the score image, as depicted in Figure 6. The "haloes" are produced by blurring the score; to these are added the grid and Current Time Indicator overlays. The projected graphics are once again perspectivally warped to compensate for any off-axis projection errors and to ensure accurate registration between the real and virtual scenes.

In order to prevent the vision system from becoming confused by its own video projections on the table surface, the system's camera is fitted with an infrared (IR) filter optimized to pass IR light beyond ~750nm. This takes advantage of the coincidence that most video projectors cast very little IR light. In this way, machine vision and human perception are respectively segregated into the IR and visible spectra, and the possibility of a video feedback cycle is avoided. The *Scrapple* table is illuminated by an overhead IR source in order to enhance its visibility for the camera.

The *Scrapple* software is written in C++, and makes use of OpenGL, the *PortAudio* library for real-time sound, and the Intel IPP libraries for accelerated signal processing. *Scrapple* captures and processes 640x480-pixel video at 60fps with a monochromatic Firewire video camera.

# 4 Conclusions

I present a real-time, camera-based spectrographic performance instrument with a tangible interface. Unlike previous camera-based systems, I use an 'augmented reality' (AR) overlay to provide the user with *in-situ* visual feedback regarding the state of the system. This feedback helps the user to more accurately predict the effects of their actions (such as placing marks into the score), thus affording the compositional precision of stylus-based systems, while preserving the possibility for coarse body-based improvisation. I describe techniques, such as the correction of radial lens distortion, which I believe are essential for camera-based systems to achieve accurate spectral synthesis and AR image registration.

# 5 Acknowledgments

The creation of *Scrapple* was made possible through a commission from the Ars Electronica Center in Linz, Austria. Many thanks to Gerfried Stocker, Horst Hörtner, and Nicoletta Blacher of this organization for their generous support. Thanks to Scott Ritter for the interior design of the first *Scrapple* installation; to Zachary Lieberman for helpful software suggestions; to Scott Snibbe for photography; and to Sue Costabile and Gregory Shakar for their compositional contributions to the performance version of the project.

# References

[1] Bourke, Paul. "Lens Correction and Distortion." Web site: <http://astronomy.swin.edu.au/~pbourke/projection/lenscorrection/index.html> , 2002.

[2] Cooper, F.S., Liberman, A. M., & Borst, J. M., "The interconversion of audible and visible patterns as a basis for research in the perception of speech." *Proceedings of the National Academy of Science*, 1951, 37, 318-325.

[3] Fisher, Robert, et al. "Adaptive Thresholding", in *The Hypermedia Image Processing Reference*, Web site, http://homepages.inf.ed.ac.uk/ rbf/HIPR2/adpthrsh.htm

[4] Florin, Adam. *Jit.peak. additive-synthesis.pat*. Demo module in Cycling '74 *Jitter* software. <http://www.cycling74.com>

[5] Kurenniemi, Erkki. "DIMI-O", featured in: Taanila, Mika. *Tulevaisuus ei ole entisensä* ("The Future Is Not What It Used To Be") DVD. MEGO Records, Austria, 2002.

[6] Kurenniemi, Erkki. Personal communication.

[7] Levin, Golan. "Painterly Interfaces for Audiovisual Performance." M.S. Thesis, MIT Media Laboratory, 2000.

[8] Manovich, Lev. *The Language of New Media*. MIT Press, 2001.

[9] Marino, G., Serra, M.-H., and Raczinski, J.M. "The UPIC system: Origins and innovations." *Perspectives of New Music*, 31(1), 1993.

[10] Rubin, Philip and Goldstein, Louis. "The Pattern Playback". <http://www.haskins.yale.edu/featured/patplay.html>, 2004.

[11] Timmermans, Paul. "The UPIC System." Web site, <http://membres.lycos.fr/musicand/>

[12] UI Software Inc. *Metasynth* music software. 1998-2006.

[13] Underkoffler, J. and Ishii, H., "Illuminating Light: An Optical Design Tool with a Luminous-Tangible Interface", in *Proceedings of Conference on Human Factors in Computing Systems* (CHI '98), ACM Press, pp. 542-549.