DIRTI — Dirty Tangible Interfaces

Matthieu Savary
USER STUDIO
181 rue des Pyrénées
75020 Paris, France
savary@userstudio.fr

Diemo Schwarz
UMR STMS
Ircam–CNRS–UPMC
Paris, France
schwarz@ircam.fr

Denis Pellerin
USER STUDIO
181 rue des Pyrénées
75020 Paris, France
pellerin@userstudio.fr

ABSTRACT
Dirty Tangible Interfaces (DIRTI) are a new concept in interface design that forgoes the dogma of repeatability in favor of a richer and more complex experience, constantly evolving, never reversible, and infinitely modifiable. We built a prototype based on granular or liquid interaction material placed in a glass dish, that is analyzed by video tracking for its 3D relief. This relief, and the dynamic changes applied to it by the user, are interpreted as activation profiles to drive corpus-based concatenative sound synthesis, allowing one or more players to mold sonic landscapes and to plow through them in an inherently collaborative, expressive, and dynamic experience.

Keywords
Tangible interface, Corpus-based concatenative synthesis, Non-standard interaction

1. INTRODUCTION
Dirty Tangible Interfaces (DIRTI) belong to a new generation of complex controllers that take advantage of the finest changes of the environment they are analyzing: for example the very refined movements that the hand can transmit to a Wiimote (Nintendo), or full body motion that the Kinect (Microsoft) can help interface with machines.

This generation of user interfaces (UI) is especially emancipated from the traditional keyboard, mouse, joystick or even graphics tablet that all rely on the boolean and/or analog transduction of a small number of buttons or potentiometers.

In particular, we call *dirty tangible interface* a tangible user interface that bears the following features:

- The return to its neutral position is artificial, in the sense that it is only achieved when the user decides so (ie. via the software that grabs the information from the interface).
- The interface is constantly evolving, and changes that happen, as little as they may be, are not reversible. Only a high enough software threshold set on change detection could reduce the variations, and artificially enable more discrete changes.
- The interface is infinitely customizable by choosing a different interaction material, e.g. grains, liquid, balls.

We implement this principle in a prototype interface, explained in section 3, based on granular or liquid interaction material placed in a glass dish (see Figure 1), the image of which is captured by a camera and translated into a 3D depth image that is then used to activate a corpus-based sound synthesis system and to generate related, visual behaviors on screen. Thus, the most minute physical interaction with the material will alter the produced sound, as can be seen in the accompanying example video at http://vimeo.com/topophonie/dirti. More videos, images and applications can be found at http://smallab.org/dirti.

Figure 1. More or less dirty interaction materials.

2. RELATED WORK
SandScape[1] [1], [2] is a tangible interface for designing and understanding landscapes and drainage aspects of them through a variety of computational simulations using sand. Users view these simulations as they are projected on the surface of a sand model that represents the terrain. The users can alter the form of the landscape model by manipulating sand while seeing the resultant effects of computational analysis generated and projected on the surface of sand in real-time.I HA

The *Relief*[2] [3], [4] and *Recompose*[3] [5] interfaces are actuated tabletop displays, which render and animate three-dimensional shapes with a malleable surface. They allow users to experience and form digital models like geographical terrain in an intuitive manner. The tabletop surface is actuated by an array of

[1] http://tangible.media.mit.edu/project/sandscape
[2] http://tangible.media.mit.edu/project.php?recid=132
[3] http://tangible.media.mit.edu/project/recompose

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

motorized pins. Each pin can be addressed individually and senses user input like pulling and pushing.

The above interfaces are neither aimed at expressivity, nor made for musical purposes, and the precise (re)configuration of the interface is here the aim, contrary to our *dirty* principle. The *Splash Controller* organic UI [6] is closer, detecting manipulation of water in a gaming context. The only slightly *dirty* musical interfaces are *PebbleBox* and *CrumbleBag* [7]. Those examples of a granular interaction paradigm are based on the analysis of the sounds resulting from the manipulation of physical grains of arbitrary material. This analysis extracts parameters as grain rate, grain amplitude and grain density, that are then used to control the granulation of sound samples in real-time. This approach shows a way of linking the haptic sensation and the control of granular sounds. However, this interface focuses on the interaction sound and forgoes to extract information from the configuration of the material.

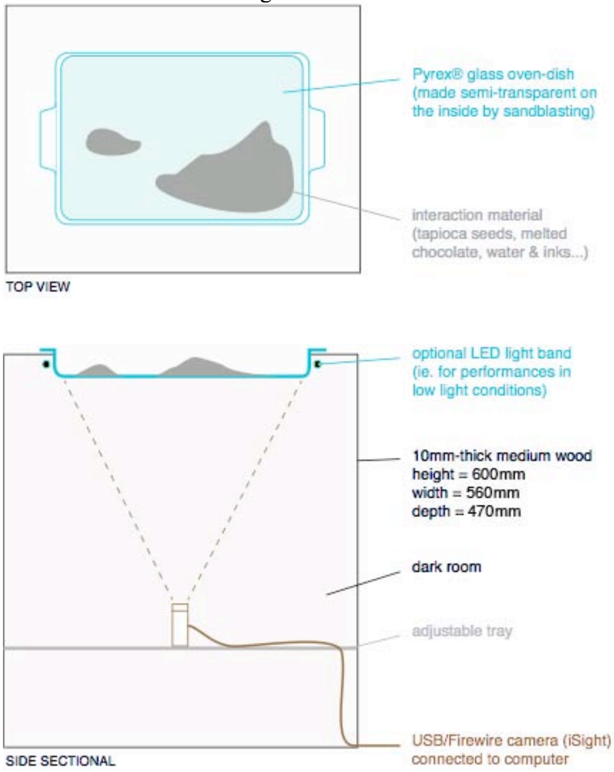

Figure 2. DIRTI hardware schema.

Figure 3. DIRTI prototype, open.

3. PROTOTYPE

The prototype instrument, see Figures 2 and 3, consists of a dark box containing a video camera, a semi-transparent glass dish, surrounded by optional LED lights to perform in darker conditions, and containing the interaction material. Several kinds of interaction materials can be used: dry grains (plastic granulate, tapioca grains, peas, marbles), plastic balls, water with ink(s), ice cream, soft chocolate... depending on the desired expressivity, precision/randomness ratio, and inertia of movement wanted. Movement and density of material in the dish are captured from below, thanks to a camera placed underneath in order to obtain a gray scale image of the interaction material. This image is then converted by the analysis software into a 3D depth image that activates the sound synthesis, as described in the following

4. DETECTION AND ANALYSIS

The grayscale camera image is the source of detection of various parameters:

- density of interaction material
- motion, quantity of movement applied to it
- colors

Based on the grayscale image, first a blurring filter is applied in order to obtain smoother contours in the later analysis stages, and to avoid flicker. Then, OpenCV's hierarchical blob detection algorithm is applied to estimate *n* levels of contours of iso-luminance blobs in the image, where *n* is a parameter that determines the depth resolution of the analysis, usually set to 20. These contours are then interpreted as a 3D relief of the material: each subsequent level is assigned a depth coordinate, which is a simple and sufficiently precise way to estimate the density and thus the height of the interaction material (see Figure 4). However, for dynamic gestural control, our approach is to detect where there is movement in the material. Therefore, a moving blob detection was implemented by first creating the difference image from the blurred grayscale image camera, and then detecting the hierarchical blobs on this. This means that fast movements will automatically result in a deeper blob being detected, because the difference is greater

4.1 Profiles

Both, the depth of the 3D moving blob, and the depth derived from the background image, are then interpreted as *profiles*, i.e. 2D fields carrying a parameter value, that are in the following applied as *activation profiles* to a sound process [8].

4.2 Analysis Software Environment

The Dirty Tangible Interfaces analysis software uses Cinder creative coding framework[4] and its OpenCV Block. The CCGL (CocoaCinderGL) wrapper[5] eases the use of the Cinder framework from within a typical Cocoa project under MacOS X. It allows C++ creative coders to build quick prototypes with several windows and a native Graphical User Interface on the Macintosh. This is especially handy in the context of building DIRTI, where several windows are needed to handle the detection and visualize the interaction.

5. DETECTION AND ANALYSIS

There are two visualization algorithms, helping in developing the interface, and analyzing the interaction modes. First, the visualization of a 3D image derived from the blobs' assigned depth allows to see the relief of the interaction material. See Figure 4 and Figure 5 for examples of the 3D visualization:

- the blobs are drawn with a single pixel outline, from dark red (least light blob) to bright purple (most light blob)

---

[4] http://www.libcinder.org

[5] http://www.smallab.org/code/ccgl

- the user can rotate the blobs in 3D virtual space using the mouse
- the user can grow the blobs away from the same 3D plane by adding z-value according to their “grayscale level” via a slider called “3D effect”

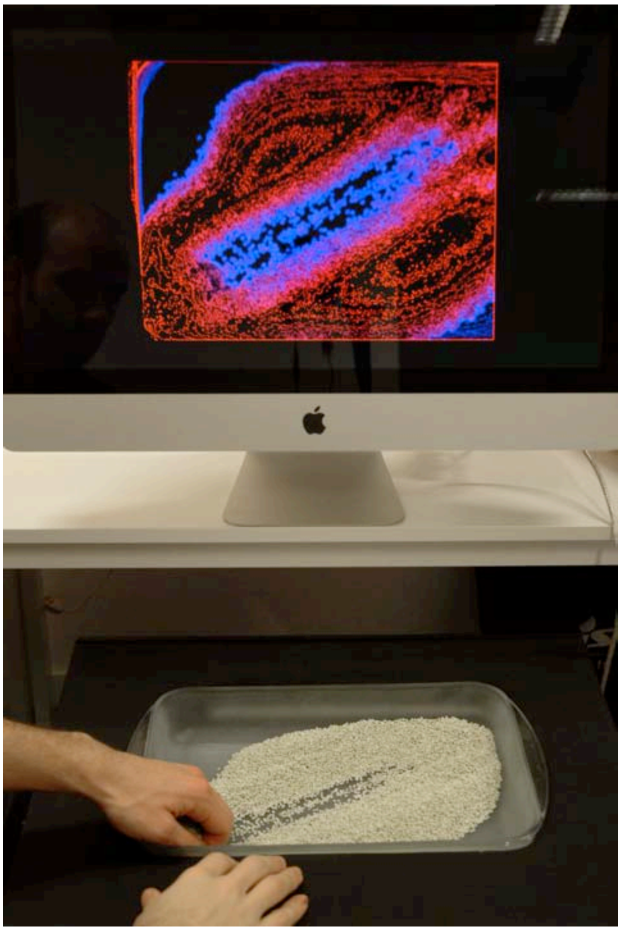

Figure 4. DIRTI detection example.

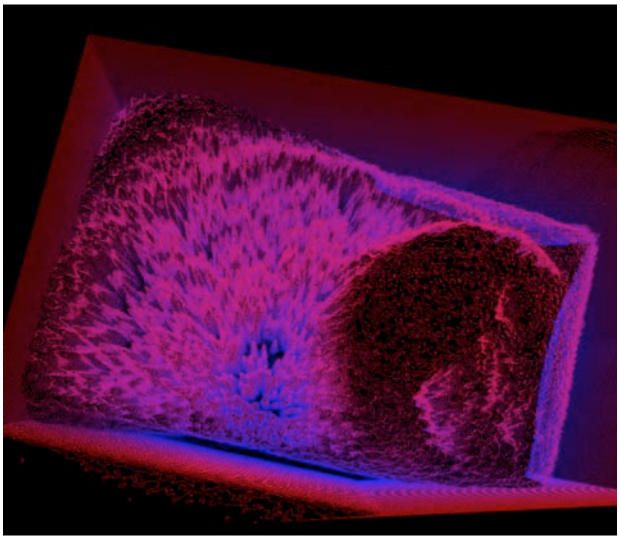

Figure 5. Screenshot of 3D visualization.

The second visualization includes the points corresponding to active sound segments for corpus-based audio synthesis. It includes feedback of the points’ activation: Quantity of movement is mapped to the size of the point, the background grayscale level is mapped to inverse color saturation, i.e. light background is dark green, dark background gives light green. See Figure 6 and the accompanying video for an example.

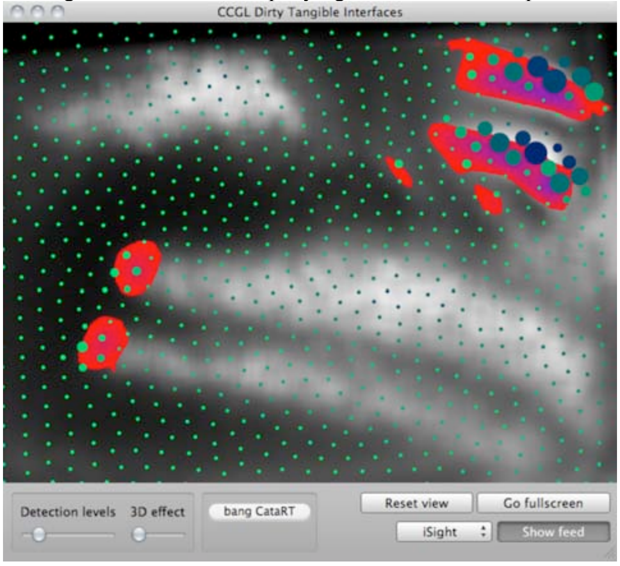

Figure 6. Screenshot of audio activation visualization.

6. INTERPRETATION

6.1 Audio

The audio process is based on corpus-based concatenative synthesis (CBCS) **Erreur ! Source du renvoi introuvable.** as implemented in the CataRT system. CBCS makes it possible to create sound by selecting segments of a large database of pre-recorded audio (the corpus) by giving a target position in a space where each segment is placed according to its sonic character in terms of audio descriptors, which are characteristics extracted from the source sounds such as pitch, loudness, and brilliance, or higher level meta-data attributed to them.

For Dirty Tangible Interfaces, we project the corpus onto the 2D interaction surface by choosing two descriptors as its axes. Each segment then has a 2D coordinate and can be visualized as a point on the detection visualization (see Figure 6).

To play the segment associated to a point, we determine if it lies within a blob, in which case the segment is triggered (if it is not already playing). The depth of the containing blob is mapped to the playback gain, so that fast movements play loud, slow movements play softly.

The background profile can be mapped to a sound transformation parameter, e.g. segment fade-in and out times, reverse probability, filter. In our experiments, we obtained musically interesting subtle effects by mapping the background to a little amount of transposition randomization (maximally +/-2 semitones). This means that at the beginning, with a thick layer of material, sounds play untransposed, but when digging deeper and exposing the bottom of the dish, chorusing effects can be deliberately produced for specific sound segments only.

6.1.1 Optimizing the Navigation Space

While a direct projection of the high-dimensional descriptor space to the low-dimensional navigation space has the advantage of conserving the musically meaningful descriptors as axes (e.g. linear note pitch to the right, rising spectral centroid upwards), we can see in [11] that sometimes the navigation space is not optimally exploited, since some regions of it stay empty, while other regions contain a high density of units, that are hard to access individually. Much of the interaction surface can remain unexploited.

Therefore, we apply the distribution algorithm *Unispring* [10] that spreads the points out using iterative Delaunay triangulation and a mass–spring model, while keeping similar sounding points close together. The results of the algorithm can be seen in Figure 6.

6.1.2 Implementation

Thanks to the underlying FTM&Co. extensions providing optimized data structures and operators in a real-time object system and arbitrary-rate overlap–add granular synthesis to Max/MSP, CataRT can play all activated sound grains in parallel, limited only by the CPU speed of the machine. (In practice thousands of segments playing in parallel are possible.) The detection software communicates with CataRT via OSC, initiating a dump of the corpus segment's positions, and sending back the activation and background levels of all points. CataRT and FTM&Co are released as free open source software at http://imtr.ircam.fr and http://ftm.ircam.fr.

6.2 Graphics

Several projects of graphical interpretation of the dirty interaction are under work. One of them, *Dirti Traces* (see Figure 7), consists of tracking the blobs and using them to represent traces of the movements that get eroded and displaced through time, symbolizing the attack, sustain and decay of the sounds produced by the interaction.
Another, *Dirti Terrain Editor* (see Figure 8), makes use of the density of the interaction material in the dish to edit a 3D terrain on screen, allowing the performer to draw islands, lakes and paths while playing sounds with Dirti.
For the future, a graphical feedback similar to that of the *Parametropophonics*[6] audio–graphic, parametric 3D models (formerly *Swirls*[7]) is planned. Here, the audio descriptors for each segment determine the expression of a parametric 3D shape, and their activation will animate parts of the parameters, or interpolate between models.

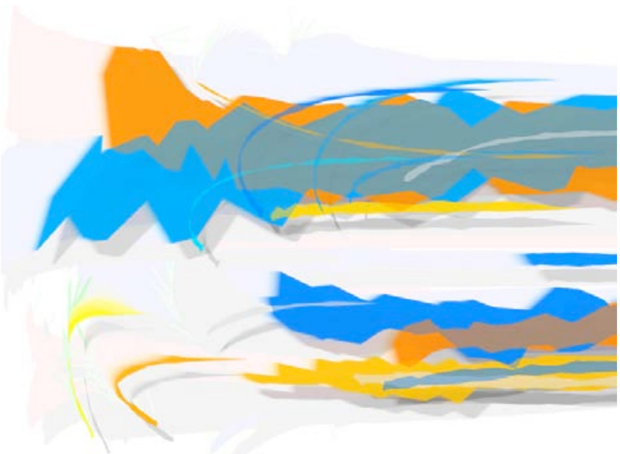

Figure 7. Screenshot of *Dirti Traces* graphical interpretation.

7. CONCLUSION

As can be seen in the example video, also visible online at http://vimeo.com/topophonie/dirti, dynamic and expressive musical play is possible, matching the dynamics of the manipulation of the interaction material. Thanks to the mapping of the space of sound characteristics to the interaction space, timbral evolutions can be purposefully controlled. Even multi-player interaction is possible in this inherently collaborative interface.

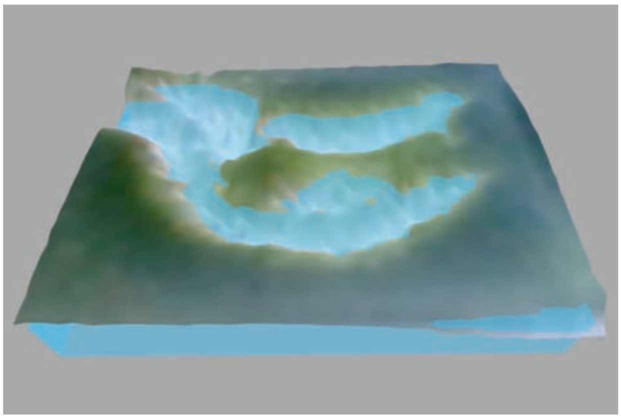

Figure 8. Screenshot of *Dirti Terrain Editor* graphical interpretation (see http://vimeo.com/37313858).

8. ACKNOWLEDGMENTS

Thanks to Romain Pascal for his help on the physical prototype, Roland Cahen, and ENSCI–Les Ateliers.
The work presented here is funded by the Agence Nationale de la Recherche and Cap Digital within the project *Topophonie*, ANR-09-CORD-022 (see http://www.topophonie.fr).

9. REFERENCES

[1] Ishii, H., Ratti, C., Piper, B., Wang, Y., and Biderman, A., Bringing Clay and Sand into Digital Design — Continuous Tangible User Interfaces. In *BT Technology Journal*, 22, 2004.

[2] Piper, B., and Ratti C., Illuminating Clay: a 3-D Tangible Interface for Landscape Analysis. In *SIGCHI Conference on Human factors in computing systems (CHI '02)*, 2002.

[3] Leithinger, D., Lakatos, D., DeVincenzi, A., Blackshaw, M., and Ishii, H., Direct and Gestural Interaction with Relief: a 2.5 D Shape Display. In *ACM Symposium on User Interface Software and Technology*, 2011, 541–548.

[4] Leithinger, D., and Ishii, H., Relief: a Scalable Actuated Shape Display. In *International Conference on Tangible, Embedded, and Embodied Interaction*, 2010, 221–222.

[5] Leithinger, D., Lakatos, D., DeVincenzi, A., and Blackshaw, M., Recompose: Direct and Gestural Interaction with an Actuated Surface. In *ACM SIGGRAPH 2011 Emerging Technologies*, ACM, 2011, 13.

[6] Geurts, L., and Van den Abeele, V., Splash Controllers. In *Tangible, Embedded and Embodied Interaction (TEI '12)*. New York, NY, 2012.

[7] Sile O'Modhrain, M., and Essl, G., Pebblebox and Crumblebag: Tactile interfaces for granular synthesis. In *NIME*, 2004, 74–79.

[8] Schwarz, D., Cahen, R., Brument, F., Ding, H., and Jacquemin, C., Sound Level of Detail In Interactive Audiographic 3D Scenes. In *International Computer Music Conference (ICMC)*, 2011.

[9] Schwarz, D., Corpus-based concatenative synthesis. In *IEEE Signal Processing Magazine*, *Special Section: Signal Processing for Sound Synthesis*, Mar. 2007, 24(2):92–104.

[10] Lallemand I., and Schwarz, D., Interaction-optimized sound database representation. In *Conference on Digital Audio Effects (DAFx)*, Paris, France, 2011.

[11] Schwarz, D., The Sound Space as Musical Instrument: Playing Corpus-Based Concatenative Synthesis. In *New Interfaces for Musical Expression (NIME)*, 2012.

---

6 see http://vimeo.com/37967817
7 see http://vimeo.com/21339248 and http://smallab.org/swirls