Found Data: Generating Natural Looking Shapes by Appropriating Scientific Data

Andres Wanner
andres_wanner@sfu.ca
Simon Fraser University, Vancouver, Canada

Ruth Beer
rbeer@ecuad.ca
Emily Carr University of Art and Design, Vancouver, Canada

Keywords: Installation Art, Generative Art, Interactive Art, Data Visualization, Public Data, Appropriation, Oceanography, Sustainability.

Abstract: The installation *Breathe/Live/Speak* utilizes oceanic data to generate an organic distribution of screen elements.

This paper describes the installation as part of the *Catch and Release* research/creation project. We introduce our approach of Found Data, derived from artistic practices of Found Object and Readymade, as an alternative to the widely used Perlin Noise for generating natural looking shapes.

The approach is demonstrated in detail, and some examples are presented. We outline how these are implemented in the installation, and conclude by arguing for the relevance of this method in a time of increasingly available data.

xCoAx2013
Computation Communication Aesthetics and X. Bergamo, Italy. xcoax.org

1. Introduction

Practitioners in Generative Art have access to a wide range of techniques and methods for generating organic, natural looking shapes. In *The Nature Of Code*, Shiffman presents Cellular Automata, Koch-Curve Fractals, L-systems and Genetic Algorithms as a variety of methods for representing nature computationally (Shiffman). Our approach is different: we represent nature by directly using data from nature.

Manovich has suggested an analogy between visual arts and information visualization, and compares the choice of data with an artist's selection of a visual motif: "Figurative artists express their opinions about the world by choosing what they paint… Now artists can also talk about our world by choosing which data to visualize." (Manovich, 13)

Our aim in this paper lies in presenting our Found Data approach producing natural looking patterns. We refrain from offering a metric or criterion to evaluate how 'natural' a pattern looks, and leave this to the reader to assess. Instead we will discuss a thematic motivation for using a non-computer-generated randomness, and present the results we achieved.

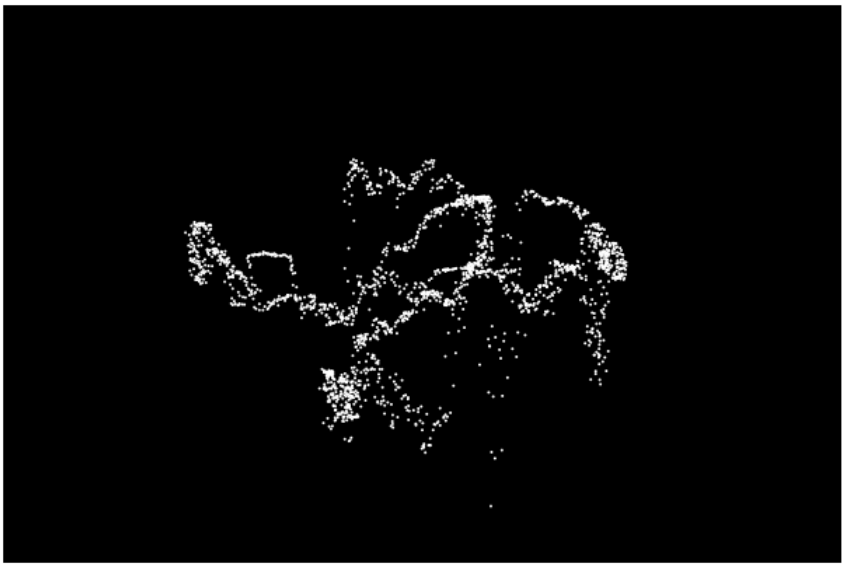

Fig 1. A Found Data plot with forms based on scientific data.

2. Catch and Release and the installation Breathe/Live/Speak

Catch and Release: Mapping geographic and cultural transitions is a research/creation project with the goal of raising awareness about current issues of cultural and environmental sustainability. This government-funded 3-year initiative acts as an umbrella for interdisciplinary art projects — interactive storyscapes that engage viewers with these issues through the immersive experience mediated by multimedia installations.

The interactive installation *Breathe/Live/Speak* is one of these projects — a dynamic composition of oceanic elements on projection screens. These elements — plankton organisms, bubbles, and typographic content — represent oceanic life, their motions suggest particles floating in underwater currents. The installation visualizes numerical empirical data from the *NEPTUNE Canada* regional cabled ocean network, which gathers live data from undersea environments in the Northeastern Pacific. An online platform makes these marine data publicly available — users find data on oxygen concentration, salinity, temperature, current, and other variables (Neptune).

By referencing the oceanic context, and through the subtle interactivity with the organisms and other screen-elements, the installation aims at raising awareness and reminding viewers of their impact on a fragile oceanic ecosphere.

3. Context

3.1. Situation between Generative Art and Data Visualization

Our work implements aspects of both generative art and data visualization. We borrow the autonomous generative process from generative art (Galanter), and the emphasis on representation of abstract data from data visualization. However, in contrast to the "clear and effective communication" Friedman demands from visualizations (Friedman), as artists we want to leave room for ambiguity and interpretation. We are not concerned with scientific semantics of our data, but will use it to generate aesthetic forms, in agreement with Manovich:

> The intent of these projects is not to reveal patterns or structures in data sets but to use information visualization as a technique to produce something aesthetically interesting. (Manovich, 13)

3.2. Perlin Noise

Shiffman observed that "In a computer graphics system, it's often easiest to seed a system with randomness." — and simultaneously pointed out problematics of this approach (Shiffman, 7). While different strategies of randomness and noise are prevalent in Generative Art, one approach is that of Perlin Noise, originally developed by Ken Perlin for textures in computer-generated animations. It is suited and widely used for generating unpredictable and "naturally looking" patterns featuring "the subtle irregularities of real objects" (Perlin, 12). Thus it would be a common approach for our purpose of generating a natural distribution of screen-elements.

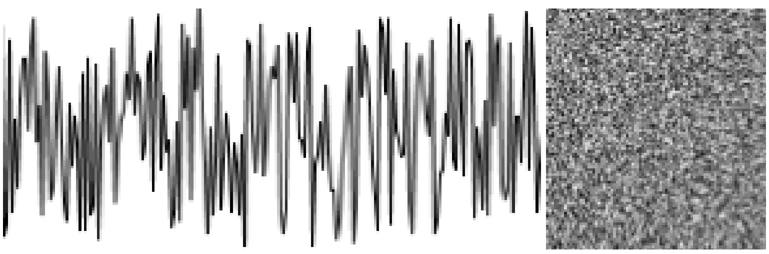

Fig 2a. The random-function: function graph and 2D pattern of grey values.

Fig 2b. The Perlin-Noise function: function graph and 2D pattern of grey values.

In contrast to a standard random function, Perlin Noise is coherent, i.e. two neighbour points will have a similar noise value. Perlin specified that "all the apparently random variations be the same size and roughly isotropic" — they will look similar in all directions and positions. (Perlin, 5).

3.3. Found Object and Found Data

Early twentieth century artists have introduced the Found Object into art history, an everyday object that obtains its state as a work of art through the selection and introduction into a new context. More particularly, with his Readymades, Marcel Duchamp

> … took an ordinary article of life, placed it so that its useful significance disappeared under the new title and point of view — and created a new thought for that object. (Duchamp)

Discussing the implications on authorship, Irvin suggests that

> Holding the artist responsible for a work means, in part, holding the artist responsible for having released it into a context where particular interpretative conventions and knowledge are operative. (Irvin)

The conventions of the new context are as important as the actual fabrication of the artifact.

In an analogy to a Found Object, we would like to suggest the term Found Data, in which data is 'released into a context', and a 'new thought' for that data is created. We present the use of Found Data as an alternative to implementing Perlin Noise.

We appropriate data for their formal qualities, and deliberately ignore their scientific denotations. The installation thus re-contextualizes scientific data and uses it for the creation of natural looking distributions of elements.

Representing physical real-world quantities, the number sequences in Found Data are mostly continuous and unpredictable as Perlin Noise, but variations may not be 'same size' and 'isotropic'. As a shared objective however, we hope to produce natural looking patterns.

3.4. Our Motivations for using Found Data

The *Catch and Release* project intends to raise awareness about oceanic life. Our aim is to generate irregular patterns, which give the impression of elements being exposed to oceanic currents and turbulences. The patterns have to be capable of engaging the viewer's capacity for seeing meaningful patterns in random data, and thus of opening an interpretive space for imagination.

The use of technical random numbers to raise awareness for nature seems to be contradictory rhetoric. Shiffman is not the only one to observe that

> Defaulting to randomness is not a particularly thoughtful solution to a design problem — in particular, the kind of problem that involves creating an organic or natural-looking simulation. (Shiffman, 7)

Randomness may produce counterproductive connotations of human non-involvement and technological arbitrariness. For this reason, we prefer a non-computer-driven positioning algorithm that bears a relation with the thematic concern of our installation — awareness for the ocean. By using the ocean as a data source, we provide a self-referential dimension to the work, and align the form with the content.

4. Research: Use of Data

4.1. Method: Drawing scatter plots based on Found Data

In this section, we are going to discuss in detail, how the data in the *Breathe/Live/Speak* installation is used to generate distributed positions of screen elements. Technically, our method consists of generating scatter plots of two variables. Such scatter plots are used to view and analyze a correlation between two variables. Strongly correlated variables will result in a diagonal linear distribution, whereas uncorrelated variables will not show a diagonal pattern and produce a distribution spreading over a wide range of the graph area.

For distributions with interesting and surprising shapes to arise, we look for a pair of uncorrelated variables to be plotted against each other. We choose measurements from different locations and times, to minimize interdependencies.

Fig 3. Scatter plot of two uncorrelated variables (right).

Figure 3 illustrates this method in detail. Data of *oxygen concentration* in Barkley Canyon is plotted against *salinity* from the Folger Deep sensor station. These locations are about 100km from each other, and the measurements are separated by a 14-day period, and taken at a slightly different time of day therefore we can hope they are not strongly correlated. The plot displays 2000 data points, taken in 1-minute intervals. Values of both datasets fluctuate in intervals between 10 minutes to several hours. This is long enough to generate surprising patterns, but not too long to result in a completely uniform distribution.

The graph bears minimal scientific meaning (if any at all), however we begin to see patterns, that we think have potential to be considered *natural looking*.

4.2. Requirements on the data

We impose minimal requirements on the data. We ask that data is continuous, fluctuating in intervals of the order of 100 data points, and that the two datasets are not strongly correlated. In this section we show some counter examples of data that are not suited for this method:

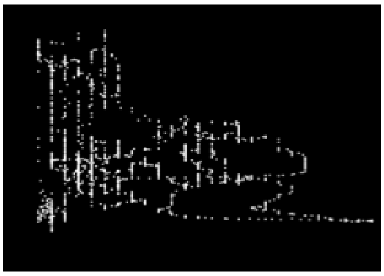

Fig 4a. Horizontal data is not continuous, but in discrete steps.

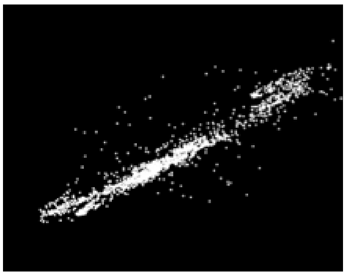

Fig 4b. Horizontal and Vertical data are correlated.

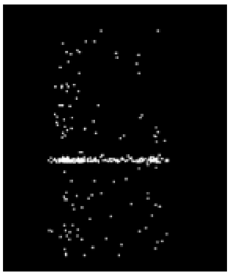

Fig 4c. Vertical data is not continuous.

We demonstrate three cases with data that are not suited for our method. In figure 4a, horizontal data is not continuous, but in discrete steps. These lead to regular gaps in the resulting pattern, which we want to exclude for aesthetic reasons.

In figure 4b, data series are taken from the same time and location — salinity and temperature from the Folger T station. The plot approximates a diagonally ascending line that we would expect from a scatter plot of partially correlated variables.

In figure 4c, the values of the vertical variable — the focus of an underwater video camera — are discontinuous and differ heavily between subsequent measurements. The data-points fill the entire space. Their distribution is mostly uniform apart from clustering vertically around an average value.

4.3. Gallery: Some examples

Fig 5. Barkley Canyon Oxygen vs. Folger Salinity.

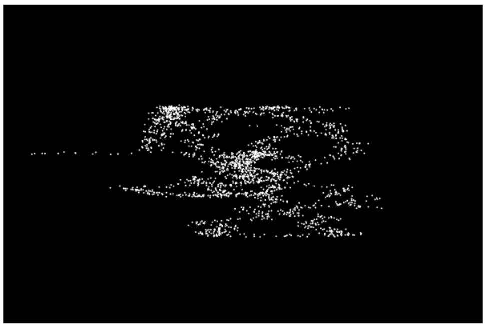

Fig 6. Barkley Canyon Oxygen vs. Folger Pressure

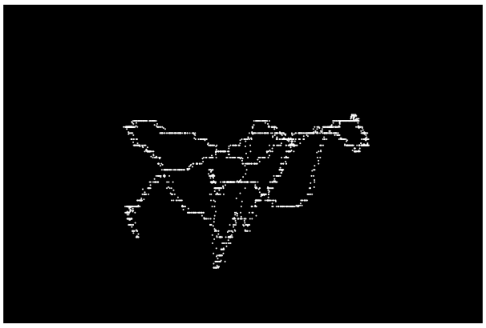

Fig 7. Folger Pressure vs. Barkley Canyon Temperature (bordeline case with discrete values).

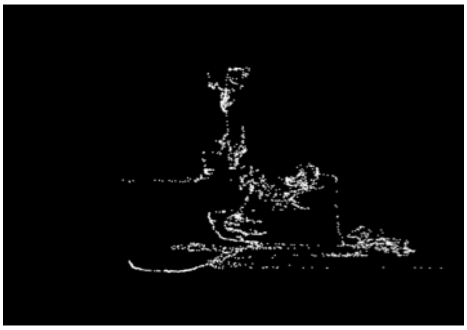

Fig 8. Folger Temperature vs. Folger Salinity.

4.4. Correlating a data series with itself, shifted in time

Fig 9. Barkley Canyon Oxygen correlated with itself, shifted by 13782 minutes.

Fig 10. Folger Temperature correlated with itself, shifted by 5461 minutes.

We also correlated data series with themselves, but shifted in time. Figures 9 and 10 display plots of data series that are correlated with series later in time, but from the same respective sensor. Our observations show, that more interesting patterns arise, once the offset between the two series is over some 1000 minutes.

4.5. Comparison to Perlin Noise

Fig 11. Perlin noise, correlated to other Perlin noise.

When applying our method to two series of Perlin Noise values, we obtain a similar pattern. However the plot looks more regular and less organic than our Found Data plots: the points are more uniformly distributed over the screen, and are lined up in quite regular distances along continuous lines. We speculate that the 'same size' and 'isotropic' properties are counterproductive for the use we have in mind.

4.6. Scaling patterns to cover the full screen

We earlier assumed that Found Data patterns may not necessarily be 'same size' and 'isotropic', thus at times they result in localized and off-centered clusters on the screen. In contrast, with our goal of creating an immersive experience, we wanted to situate the viewer *within* the data environment, rather than having her look *at* an object with finite contours. For our application in the interactive installation, we thus dynamically scaled the data to center patterns extending them to cover the entire screen area. This provides more immersion, although it compromises the density and conciseness of the patterns.

Fig 12. Off-centered plot.

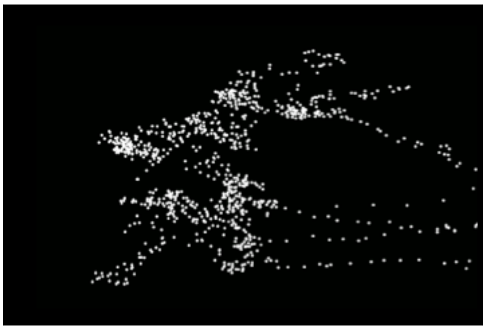

Fig 13. The same data scaled to fit screen.

5. Application

In the *Breathe*/*Live*/*Speak* installation, our method is used to position elements on interactive screens. Timing is chosen in a way so the elements appear to move as in an oceanic current, and the colour scheme further underlines this oceanic connotation. The colours, opacities and orientation angles of the individual elements are calculated using the same two datasets that were used to determine their position. We chose the range of visual parameters so as to create an evocative pseudo-spatiality contributing to the immersive aspect of the work. Viewers interact with the elements and distort their arrangement as a Kinect camera captures their body motions. Without their interference, elements will bounce back to their original, data-directed positions.

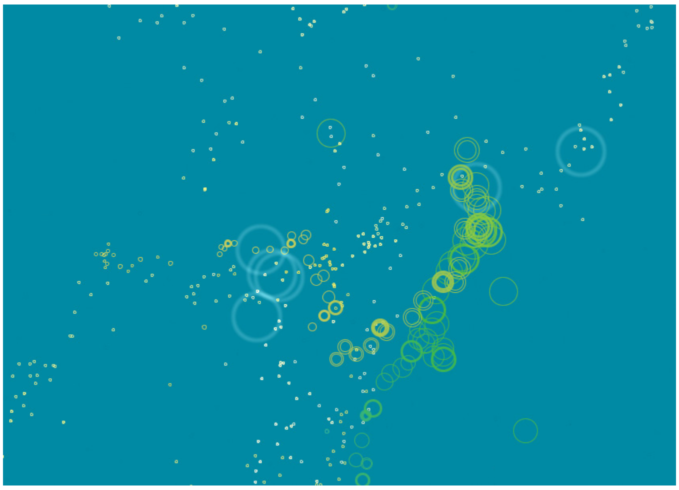

Fig 14. Installation — Screenshots: *Breathe* (detail).

Fig 15. Installation — Screenshots: *Speak* (right).

Three thematic screens illustrate different themes with varying choice of elements: *Breathe* with air-bubble elements, *Live* with plankton organisms, and *Speak* with typographic content.

6. Conclusion

We demonstrate how we use scientific data to generate natural looking patterns on aesthetic scatter plots. We chose to use empirical Found Data instead of Perlin Noise as a generative principle for positioning screen elements, as we find the method shares a thematic relation with our subject matter. Results show that our method is equally suited to produce natural looking patterns.

Our research and creative production suggests that *Found Data* may be a useful concept for directly linking the two closely related fields of Generative Art and Data Visualization. We think of it as an approach that offers a method with low predictability to enhance possibilities within Generative Art, and releases Data Visualization from the expectation of literal interpretation.

While this paper focuses in presenting the method used in our examples of installation artwork, we recognize that the method will benefit from more research to systematically clarify what kind of data will lead to interesting patterns.

Scientific and public data is increasingly accessible, and many observe an "outbreak of public visualization projects" (Lima, 97). With this paper we offer an approach for artistic use of data, and thus hope to inspire others to work with this method and to develop it further.

References

Duchamp, Marcel, Wood Beatrice, and/or Roché, H.P. "Anonymous article." *The Blind Man* # 2. (1917).

Friedman, Vitaly. "Data Visualization and Infographics" *Smashing Magazine*, *Monday Inspiration* (January 14, 2008), accessed Jan 17, 2013, http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics/.

Galanter, Philip. "What Is Generative Art? Complexity Theory as a Context for Art Theory," Paper presented at GA2003 — 6th Generative Art Conference 2003, 2003.

Irvin, Sherri. "Appropriation and Authorship in Contemporary Art." *The British Journal of Aesthetics* 45, no. 2 (April 1, 2005): 123–137.

Lima, Manuel. *Visual Complexity: Mapping Patterns of Information.* Princeton Architectural Press, 2011.

Manovich, Lev. Foreword to *Visual Complexity: Mapping Patterns of Information*, by Manuel Lima, 11–13. Princeton Architectural Press, 2011.

Neptune Canada. "Neptune Canada, Transforming Ocean Science", accessed Jan 17, 2013, http://www.neptunecanada.com/.

Perlin, Ken. "MAKING NOISE." *Based on a talk presented at GDCHardCore* (Dec 9, 1999), accessed Jan 17, 2013, http://www.noisemachine.com/talk1/.

Shiffman, Daniel. *The Nature of Code: Simulating Natural Systems with Processing.* The Nature of Code, 2012.