Observing the Observers: A New Experimental Paradigm for the Study of Seeing and Drawing

by
Ege Özgirin
B.Arch., Istanbul Technical University, 2010
Submitted to the Department of Architecture and the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of
Master of Science in Architecture Studies and
Master of Science in Electrical Engineering and Computer Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 2018

Author .....
Department of Architecture and the Department of Electrical Engineering and Computer Science
January 18, 2018

Certified by .....
Terry Knight
Professor of Design and Computation
Thesis Supervisor

Certified by .....
Patrick H. Winston
Ford Professor of Artificial Intelligence and Computer Science
Thesis Supervisor

Accepted by .....
Leslie Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, EECS Committee on Graduate Theses

Accepted by .....
Sheila Kennedy
Professor of Architecture
Chair of the Department Committee on Graduate Students

Committee

Thesis Advisor: Terry Knight

Title: Professor of Design and Computation

Thesis Advisor: Patrick H. Winston

Title: Ford Professor of Artificial Intelligence and Computer Science

Thesis Reader: Pawan Sinha

Title: Professor of Vision and Computational Neuroscience

Observing the Observers: A New Experimental Paradigm for the Study of Seeing and Drawing

by

Ege Ozgirin

Submitted to the Department of Architecture and the Department of Electrical Engineering and Computer Science
on January 18, 2018, in partial fulfillment of the
requirements for the degrees of
Master of Science in Architecture Studies and
Master of Science in Electrical Engineering and Computer Science

Abstract

One way to study how people design is to understand how others observe them designing. I take a step towards this understanding by examining how people segment visual design events temporally, in other words, how they divide these events into smaller pieces. I developed a methodology to comparatively study how multiple observers segment design events. In order to test my methodology, I conducted an experiment. In this experiment, I compared different attributes of a design event to see if some attributes communicate more meaning than others. From the results of the experiment, I observed that the segmentation of the design event was affected more by the gestures of the designer than by the produced designs. My observations suggest computational principles that could be used to develop computational design assistants that better understand designers intentions.

Thesis Supervisor: Terry Knight
Title: Professor of Design and Computation

Thesis Supervisor: Patrick H. Winston
Title: Ford Professor of Artificial Intelligence and Computer Science

Acknowledgments

I would like to thank,

Pawan Sinha for his capacity to inspire,

Terry Knight for her guidance and wit,

Patrick Winston for his attention and teaching me how to communicate better.

I am indebted to Cynthia Stewart, for her full support and understanding.

Special thanks to Sharon Gilad-Gutnick and Sarah Wu, for their enthusiasm for finding things out,

Mine Ozkar and Onur Sonmez, for their wisdom and influence,

Arzu Erdem and Zeynep Kuban, for their perception and advice.

Another set of special thanks to Scott P., for all the discussions that are stimulating and fulfilling and Hunmin K., for his empathy and friendship,

Cagri Z. and Nil T., for being the most supportive friends,

and my friends Yasaman, Dishita, Oscar, Jonathan and Gizem for their companionship.

I am grateful, for my sister Ada for her unbounded capacity to ask questions that keeps me curious, and my mom and dad, for being the most inspiring teachers in my life and being the coolest parents ever.

Finally, I am most thankful to Izgi, for her unreserved love, courage and patience.

Contents

Introduction 15

Background 19

Segmentation: A Cause or a Consequence ? 19

We Need Robots that can Learn How to Flip Pancakes from Videos 20

There are Organic Chunks within Design 22

Methodology 25

Proposed Methodology 25

Experiment 28

Analysis 31

Results 31

Discussion 33

Contributions 37

Bibliography 39

List of Figures

1 Regarding the different temporal representations of an event, where does the human interpretation fit in? Imagine a scenario from the field of cognitive neuroscience where a human performance on a specific task is investigated. Rows correspond to (1) the event under study, (2), (3) individual interpretations, (4) Oculomotor data and (5) MEG or fMRI data. 16

2 Still image examples extracted from videos featuring a model performing everyday activities [Baldwin et al., 2008]. 20
3 Warneken and Tomasello [2006] showed that even prelinguistic infants can interpret the intentions of another person. Starting from very early ages people are really good at anticipating the goals of another agent and intervene accordingly. 20
4 Recurrent neural networks are used to capture the statistical regularities in sequential data. 21
5 Visualization of the density outputs of the RNN. Large blobs demonstrate that the predictions at the end of the strokes have high variance. 22
6 Segments from the protocol analysis in Suwa and Tversky study 23

7 A sample of the software user interface. 26
8 (left) Subjects indicate breakpoints by pressing keys. (right) A new breakpoint and a segment defined upon keypress. 26
9 Depending on the protocol, subjects may be instructed to provide labels for the corresponding breakpoints and segments 27
10 In the recursive segmentation protocol, subjects can change the location of breakpoints or the labels of breakpoints and segments. 28
11 A pair of videos in which the gestures are exactly the same but only one contains the marks. 29
12 Experimental procedure for the simultaneous labeling protocol 29

13 Kernel density estimation [Rosenblatt, 1956] centers blocks with a certain bandwith around the data points. Adding up the intersections results in a smoother distribution compared to histograms. Image copyright: Duong. 31
14 Comparison of the lag adjusted responses for both videos. 32
15 Absolute differences between consecutive frames are calculated. 33
16 Comparison of the (left) responses to the video without marks against absolute differences of frame pairs in the video without marks, (right) of the responses to the video with marks against absolute differences of frame pairs in the video with marks. 33

Introduction

I STARTED TO BE INTERESTED in how people design while I was an undergraduate student in architecture. As most architecture students do, I enjoyed that the studio classes provided a class experience very different from most other classes. Generally in studio classes students are encouraged to start a project with an ideation phase, followed by further development of an idea, culminating in a mostly developed design. One thing I am certain of: The studio experience allowed me to reflect on not only what I produced but also *how* I produced it. Very early on, I realized that drawing helped me to resolve any impasses experienced within later phases of design development. Moreover, drawing helped me to trace what I did; it kindled my inspiration to iterate further. Drawing enabled me to shift my attention across the materials I was looking at, to see a detail closer, to change my goals during the activity, and to see new things in my design and manipulate them. I am interested in understanding the act of drawing as a powerful tool in the designer's toolbox.

One way to study design is to understand how people observe the design process. In this study, I explore drawing events as one form of designing. However, I conduct this exploration indirectly, by paying attention to how other people see drawing events. With this, we can rethink drawing activities as a rich and challenging area of study for apprehending how seeing works. Because, when people observe an activity they divide it up into pieces in a specific way. What is meant by dividing up is temporal segmentation and it can be defined as the act of parsing or chunking continuous temporal sequences into discrete parts that are meaningful to the observer.

This segmentation by multiple observers is especially relevant, as drawing activities are dynamic - they unfold in time. Moreover, they involve both perception and action, not as separate faculties but as two systems relying on each other. Therefore, the complexity of the design process demands an approach that should be

different from a single-handed observation.

There is an extensive body of work dedicated to studying the act of seeing but less so for studying the act of drawing. On the one hand, in the scientific domain the study of seeing is subsumed under the study of vision. And for now, let's consider the study of drawing as the study of one form of sensorimotor engagement. Historically, neurophysiological and psychophysical studies have undertaken vision and sensorimotor engagement separately. These fields rely on technical instruments for observing subjects, and the data captured by these methodologies has high temporal resolution [Ullman, 1996, Schiller and Tehovnik, 2015, Goodale and Humphrey, 1998].

On the other hand, several methods such as "Protocol Analysis" and "Think-aloud" protocols provided by artificial intelligence inspired an array of cognitive scientists and design researchers to study seeing and drawing together [Jiang and Yen, 2009, Bayazit, 2004]. These methods operate in relatively coarser time scales and the analysis of the data provided by the subjects depends entirely on qualitative interpretations by the researcher. If we assume the task under study to be a continuous event, these qualitative interpretations take the form of labels and descriptions attached to the specific parts of that event. It is common practice for the researchers within this paradigm to bring their own schemas, concepts and examples in order to conduct their analysis. [Bamberger and Schon, 1983, Schon and Wiggins, 1992, Suwa and Tversky, 1997].

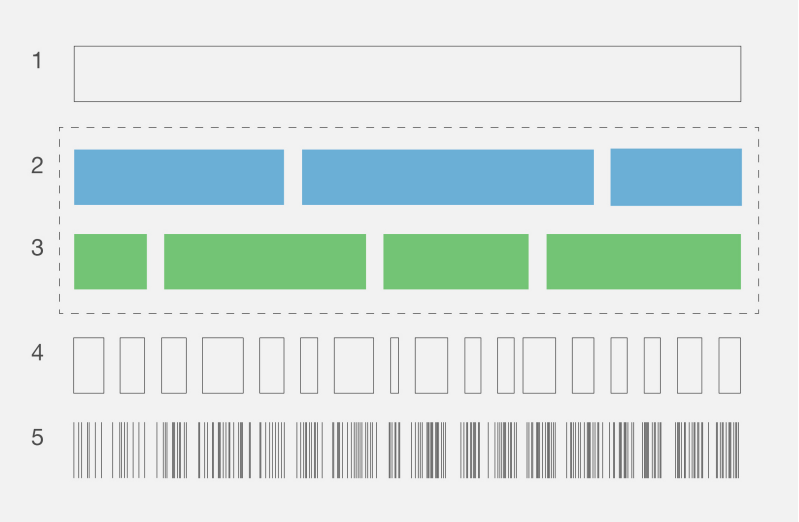

Figure 1: Regarding the different temporal representations of an event, where does the human interpretation fit in? Imagine a scenario from the field of cognitive neuroscience where a human performance on a specific task is investigated. Rows correspond to (1) the event under study, (2), (3) individual interpretations, (4) Oculomotor data and (5) MEG or fMRI data.

It is worth taking time to understand the role of interpretation

in both of these methodologies, which operate in entirely different temporal resolutions. I claim that all of these methodological approaches involve segmentation as a form of interpretation (see figure 1). Imagine a scenario in which subjects are presented with a video depicting people performing everyday tasks, and the subjectsâĂŹ responses are measured using multiple imaging techniques. The layers in this diagram illustrate the amount of granularity that can be achieved with different techniques. For example, fMRI or EEG techniques yield high resolution data across multiple time frames in the order of milliseconds. But in order to make sense of that granularity we need individual interpretations, represented as the middle layers in the figure 1, shown in color. By collecting multiple interpretations together, we can search for regularity in the middle layer. If there are matches between multiple observations, where do they occur? And if there are mismatches, what caused them? This aspect of segmentation provides a critical segue into understanding complex temporal events such as design.

What I propose is that by documenting the act of segmentation - by observing the observer - I can integrate these two different methodological approaches into a successful study of design activities. This approach will allow me to obtain data with high temporal resolution while at the same time documenting individual interpretations.

The segmentation of visual design events by multiple observers have not been comparatively studied. In most of the studies, segmentation has been done by the researcher. I suggest that increasing the number of observers who segment the activity can provide novel insights. Moreover, detailed study of the human segmentation of drawing activities can help to better identify the regularities in the perception of drawing events.

Background

In order to give a brief overview of how the segmentation of events is understood in different domains, I will talk in this section about the methods employed within the fields of cognitive sciences, artificial intelligence and design research. These methods are designed to understand events featuring an agent perceiving and performing actions directed to solve a specific task.

Segmentation: A Cause or a Consequence ?

In cognitive sciences, several researchers have studied the segmentation of events [Newtson, 1976, Tversky and Zacks, 2012, Kurby and Zacks, 2008, Baldwin et al., 2008]. Newtson [1] studied whether segmentations of videos featuring humans performing everyday tasks matched across subjects. Hierarchical organization between the events and their parts have also been studied [Newtson, 1973, Hard et al., 2006]. *Event Segmentation Theory* proposes that segmentation is a side effect resulting from an agent's desire to anticipate the upcoming information [Kurby and Zacks, 2008]. According to these researchers, when the change in the scene is maximal, prediction becomes more error-prone. They believe that these moments correspond with both the implicit and explicit segmentation performed by human observers. Baldwin et al. [2] propose that chains of actions constituting events are like words in a sentence. They claim that implicit segmentation can also occur by taking into account the statistical dependencies between actions. In other words, events are unitized automatically due to the statistical dependencies between sequences.

In psychophysics, measurement of pupil diameter has proven to be important evidence in detecting event boundaries. Increase in pupil diameter provides a measure of the amount of cognitive processing. Zacks and Swallow [2007] showed that this increase is correlated with the boundaries of an event identified by the individuals who are eye-tracked. Furthermore, in neuroscience,

[1] Darren Newtson. The Perceptual Organization of Ongoing Behavior. *Journal of Experimental Social Psychology*, 1976

[2] Dare Baldwin, Annika Andersson, Jenny Saffran, and Meredith Meyer. Segmenting dynamic human action via statistical structure. *Cognition*, 106(3):1382–1407, March 2008

preliminary neurophysiological evidence suggests that brain activity in specific regions is correlated with event boundaries even when the task does not require segmentation [Zacks et al., 2001, Speer et al., 2003, 2007, Zacks et al., 2006]. However, whether neural processing plays a causal or a consequential role in event segmentation is still an open question in neuroscience.

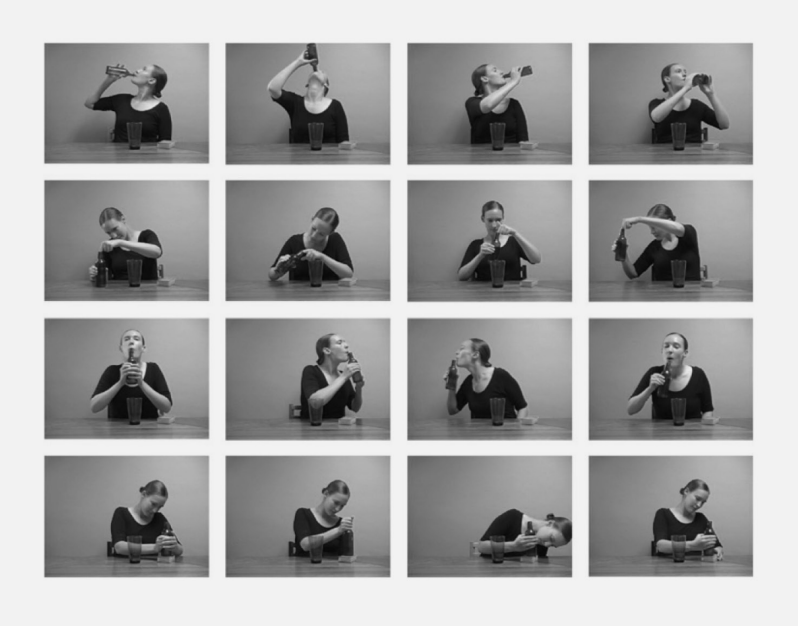

Figure 2: Still image examples extracted from videos featuring a model performing everyday activities [Baldwin et al., 2008].

In terms of how event segmentation relates to the understanding of self-initiated events, the recently developed theory of *Event Coding* suggests that perceived events and yet-to-be-produced events (actions) are equally represented in a common representational medium [Hommel, 2015]. In this account, the events you can produce constrain the events that you can perceive.

Figure 3: Warneken and Tomasello [2006] showed that even prelinguistic infants can interpret the intentions of another person. Starting from very early ages people are really good at anticipating the goals of another agent and intervene accordingly.

We Need Robots that can Learn How to Flip Pancakes from Videos

In artificial intelligence and robotics, designing better learning algorithms is a necessary step if we are to develop systems that can help humans in a variety of tasks. Whether the goal is to create a robot that can recognize human actions in order take care of household tasks, or to design a system that parses videos as humans do, defining sound procedures for segmenting human activity is becoming an important area of research.

In the literature, some researchers have undertaken the problem of segmenting everyday tasks. For example, Bouchard and

Badler [2015] segmented human motions by utilizing a qualitative analysis method called *Laban Movement Analysis* (LMA). Laban Movement Analysis attributes qualitatively determined semantic categories to human motion. The authors trained a supervised learning algorithm with manually segmented motion capture data. The algorithm produced semantic segmentations that were effective when used with general motion capture data, but when compared with human segmentation data it also resulted in many false positives. The authors claimed that this was due to the occurrence of specific motion elements that have no high-level meaning for the human observers. Spriggs et al. [2009] investigated cooking activities performed by humans and performed a segmentation and classification study. They recorded activities by using first-person sensing with accelerometers and inertial measurement units. Although they were able to segment activities in an unsupervised manner, they were not able to achieve their desired level of accuracy in classifying action units. The authors claimed that this was due to high variability in the periodicity of human actions even in a constrained context such as cooking.

Unlike the approaches above that specifically focus on how to segment a video, generative approaches in AI can also give us clues for parsing visual events. For example, inline with the other statistical approaches discussed earlier in the cognitive science literature, there are a number of statistical techniques employed for representing and generating human activities. Graves [2013] studied the statistical features of handwriting samples recorded as a sequence using recurrent neural networks. Recurrent neural networks (RNNs) are a family of neural networks that are used to establish statistical relationships between the parts of any sequential data in the form of $\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)}$. In this approach, a RNN has been trained on handwriting data. Then, the network is used to generate handwriting sequences by predicting a distribution of pen locations. One important consequence of being trained with this unique data is that the network outputs probability distributions with high variance at certain points. This is because uncertainty increases at the end of strokes for generating certain letters due to the statistical properties of the English language.

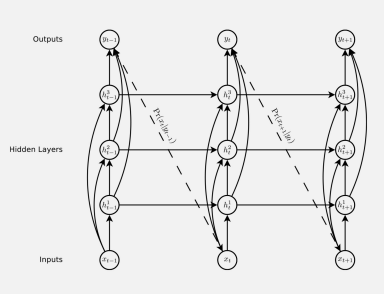

Figure 4: Recurrent neural networks are used to capture the statistical regularities in sequential data.

There are a number of systems that blend recurrent neural networks with convolutional neural networks Krizhevsky et al. [2012] to identify temporal patterns in a video. Convolutional neural networks (CNNs) are another family of neural networks inspired by the structure and arrangement of object recognition pathways in the brains of primates. CNNs form their own internal representations of visual features, and are able to decide whether an

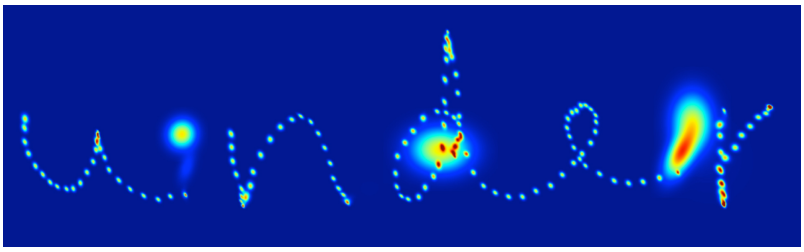

Figure 5: Visualization of the density outputs of the RNN. Large blobs demonstrate that the predictions at the end of the strokes have high variance.

object is present in an image by incorporating elements of local context (e.g. pixels in images). As a result, they demonstrate superior performance at detecting specific objects. When combined with RNNs it becomes possible to keep track of visual features that are computed by the CNNs through time. Furthermore, the organization of RNNs permits the formation of non-linear relations between derived visual features across different timescales. This is key in finding regularities in a sequence in order to predict next instances. For example, [Lotter et al., 2016] proposed a combined network that, when given a set of consecutive video frames, predicts the next few frames.The integration of these two networks produces promising results in prediction tasks, however, their capacity to parse events into coherent pieces by taking into account the prediction errors has not been exploited yet.

There are Organic Chunks within Design

Understanding how people design has been the main area of focus of many design researchers. Inspired by developments in artificial intelligence and cognitive sciences, design researchers have adopted techniques such as "Protocol Analysis" and "Think-aloud" protocols in order to study designers [Bayazit, 2004, Anders and Simon, 1980, Ericsson and Simon, 1993]. Protocol Analysis and "Think-aloud" protocols include prompting the person under study to verbally describe what she is doing during the execution of a cognitive task.

In design research, Suwa and Tversky [1997] studied sketching activity in architectural design. They asked experienced designers and students to design a layout for a museum by sketching for 45 minutes. They conducted a protocol analysis that involved instructing designers to watch a video of their drawing sequence and provide information about what they were thinking at the time. They had the subjects in question divide their own videos into segments. Then, researchers labeled the segments themselves by taking into account the designers' activities in the videos and

the corresponding verbal descriptions of the designers along with their activity in the videos. . They then counted and compared the segments to derive conclusions based on the similarities between expert and novice performances.

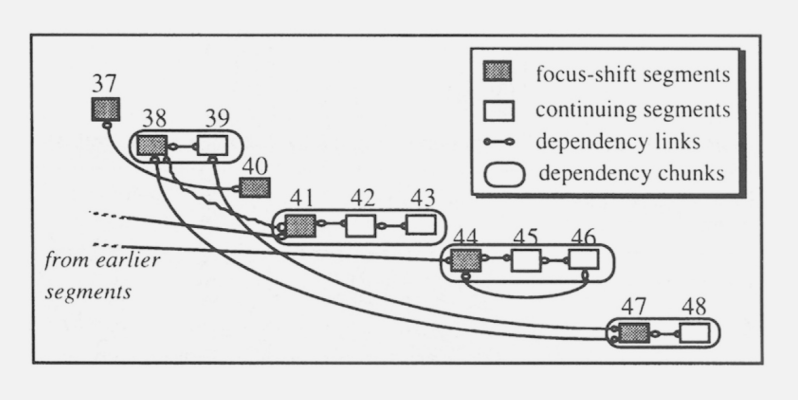

Figure 6: Segments from the protocol analysis in Suwa and Tversky study

Bilda et al. [2006] also studied sketching activity in architectural design. They prompted participants to design a house in 45 minutes, blindfolding half of the participants to force them to rely on their imagery. They further instructed participants to think aloud while designing. After that, they segmented the verbal descriptions of the designers by interpreting their intentions. In order to accomplish this, they used recorded videos to determine the start and end points of segments. They encoded the segments with four action categories: physical, perceptual, functional and conceptual. These were determined on the assumption that design thinking progresses at multiple levels in parallel. They established links between action categories belonging to each segment. Finally, these outcomes are evaluated to derive certain conclusions such as the effects of working memory limitations or implications of design education on the designers' behavior.

Jeanne Bamberger and Donald S. Schon [3] provided an interesting analysis of making events that also inspired the methodology proposed in this work. Schon, a design researcher, and Bamberger, a music educator, studied the process of making a tune out of bells with different pitch. They filmed the making process while subjects tried to find a tune. Researchers then studied the videos and attempted to find important *boundaries* to produce *organic chunks* within the continuous making events. However, their greatest insight was to realize that by repeating this chunking process several passes and also by making other people find their own meaningful chunks in the same events, they were able to find new boundaries

[3] Jeanne Bamberger and Donald A. Schon. Learning as Reflective Conversation with Materials: Notes from Work in Progress. *Art Education*, 36(2):68–73, 1983

and new chunks. This enabled researchers to reflect on and change their own interpretations. Eventually, this study resulted in more discoveries about the individual making processes.

Methodology

In the previous section I provided a compilation of state of the art methodologies from diverse fields aimed at understanding how people perceive both daily events and design events. In this chapter I will share the details of my proposed methodology and the experiments I did to test my methodology.

Proposed Methodology

One of the most important considerations undertaken during the design of the methodology was creating an easy-to-use and accurate segmentation tool. Capturing accurate and detailed data from subjects is an important task, and capturing data that features individual interpretations of events is becoming more important for the fields of artificial intelligence, cognitive neurosciences and design. For this reason, I designed a web-based software that allows researchers to design, analyze and share their segmentation experiments. Moreover, with this software, researchers can collect data from subjects regarding specific moments in events that the subjects identify as important as well as corresponding labels for those moments and durations. This can be done with a millisecond accuracy that is well suited to proper scientific analysis.

A segmentation study starts with researchers selecting which events to focus on. Events under study can be any event featuring specific everyday tasks, improvised dance or design performances or footage of observations of human interaction. The software then lets researchers upload or link their videos into their protocols. A protocol might feature multiple phases. For example, in order to make subjects familiar with the task researchers can introduce a preparatory segmentation phase. This phase can be followed by additional phases for recording baseline measurements or distracting subjects for the purposes of the experiment. Then, the main phases of measurement and collection can be appended to the protocol. Finally, multiple protocols can be appended together to

create more complex experiments or to test new protocols.

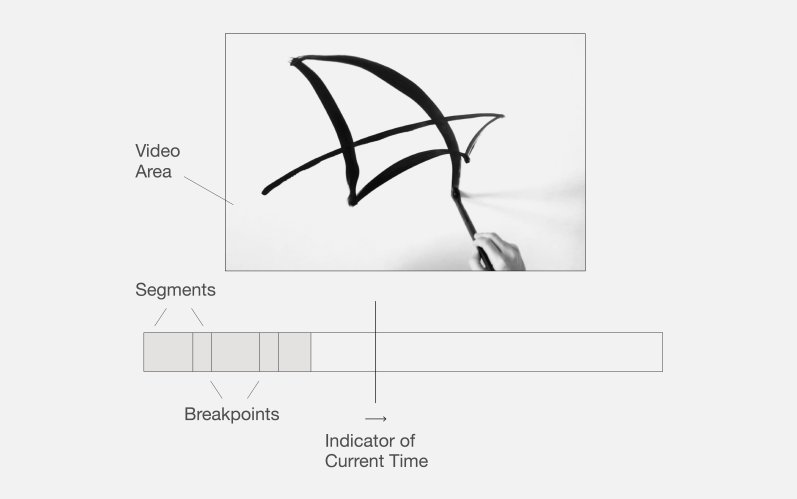

Figure 7: A sample of the software user interface.

The design of the segmentation interface comprises a software screen that contains the design video and an interactive timeline below it. This timeline provides the area where observers indicate event boundaries and segments. An indicator of the current time is shown with a protruded, sliding bar (fig. 7). Subjects press a key to indicate breakpoints. Within the context of this methodology, it is decided to define breakpoints as marks to indicate a specific moment, whereas segments are defined by the interval between consecutive breakpoints.

Figure 8: (left) Subjects indicate breakpoints by pressing keys. (right) A new breakpoint and a segment defined upon keypress.

Depending on the protocol, when subjects press a key, they may be instructed to either continue indicating breakpoints without labeling (see fig. 8) or they may be asked to include their descriptions for the corresponding breakpoint and segment (fig. 10). For the latter, the software produces pop-up forms triggered by the key-press, in which subjects can provide their labels. The video and timer can be stopped in this moment, and whenever subjects are done with labeling they can continue to the video by pressing

a key.

For example, with this software, researchers can develop protocols allowing for retrospective labeling. Retrospective labeling protocol work as follows: Initially, subjects are made familiar with the task and their baseline measurements are recorded. Then, in the first phase of the experiment subjects are presented with the video to be segmented. In this phase they are instructed to provide breakpoints only and they are not able to stop the video. In the second phase, subjects are presented with the same video and are instructed to provide labels for the segments and breakpoints they gave in the first phase. They can freely start and stop the video in this second phase. As subjects will have more time and control in this phase, the labels are subject to more careful interpretation. However, this protocol has its disadvantages. Primarily, a retrospective look on the given breakpoints and segments can contaminate the labels that they originally gave in the first phase.

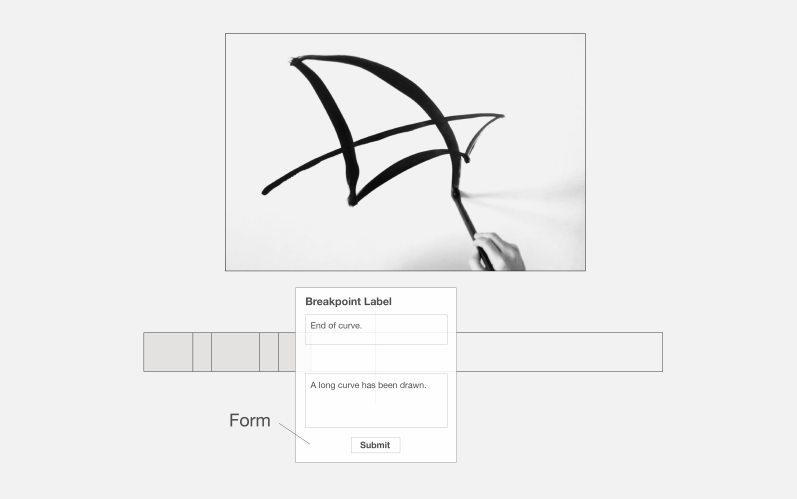

Figure 9: Depending on the protocol, subjects may be instructed to provide labels for the corresponding breakpoints and segments

Additionally, the software enables researchers to produce protocols that let subjects change the locations of breakpoints, change the labels from previous phases, or even introduce new breakpoints or delete existing ones. Responses to this protocol can give clues about how the segmentation strategies of a subject change through multiple iterations. Furthermore, this protocol could help researchers understand what constitutes better information for a particular observer and study how subjects construct hierarchical relationships between coarse and fine segments.

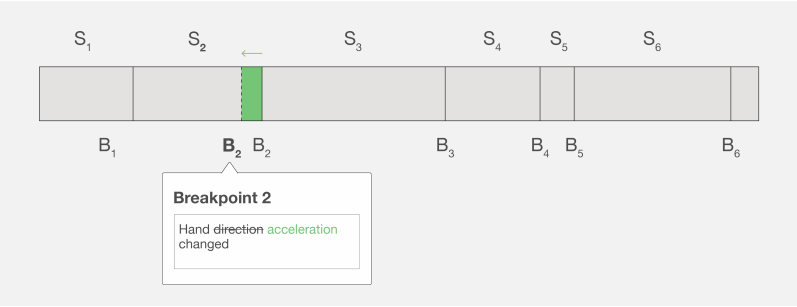

Figure 10: In the recursive segmentation protocol, subjects can change the location of breakpoints or the labels of breakpoints and segments.

Experiment

I conducted an experiment to test the proposed methodology. I formed two groups and showed each group a different drawing performance video. In these videos, the gestures (actions) of the designer are the same but only one video results in marks on the canvas. I asked whether the breakpoints and labels provided by subjects change when they are limited to observing the performance of a designer instead of the gestures alongside the produced design. I hypothesized that the breakpoints and labels provided by the subjects in both groups would match. Furthermore, I expected that in both conditions the activity would be segmented in a way that is primarily governed by the motion produced by the designer. In other words, particular attributes of motion such as changes in speed and direction that also correspond to low-level features of an image sequence (video) consistently result in observer breakpoints. Moreover, collected labels for each breakpoint and segment will be used to evaluate whether the segments provided by a subject relate to the gestures (actions) or to the produced design.

For the performance videos, I produced a series of abstract drawing videos [4]. I claim that abstract drawings in the videos used in the experiment can be thought of as designs. Because design can be defined as a creative act that requires planning, execution, evaluation and improvisation. Moreover, design processes include change of goals or strategy and re-framing of a problem at hand. During the production of these videos, large effort has been made to ensure that the drawing process incorporates all of the attributes above.

The videos were recorded with a Canon 7D DSLR camera at 1920x1080 pixels in 30 fps. I selected a video that has just enough variation in gesture speed, discontinuity, stroke width and length. However, instead of drawing with actual paint I only captured

[4] Available at this address.

Figure 11: A pair of videos in which the gestures are exactly the same but only one contains the marks.

the gestures by drawing on an empty canvas using a brush without paint. I produced two videos by overlaying a digital layer of a drawing sequence on one and keeping only the gestures of the designer on the other. This way I ensured that the gestures were exactly the same but only one of the pair contains the marks (see figure 11). The videos are then encoded to H265 format and reduced to $1080x720$ pixels to be able to be served on a third-party video server. During the experiment, one from the pair is selected randomly for each subject.

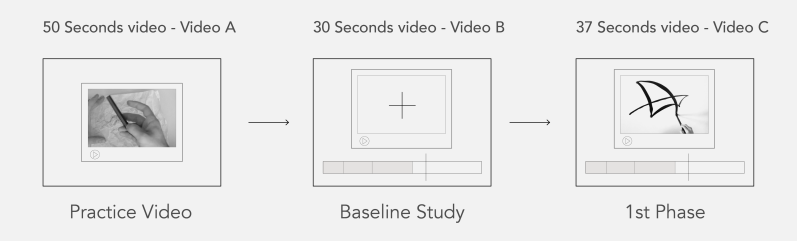

Figure 12: Experimental procedure for the simultaneous labeling protocol

For this experiment, I recruited 70 subjects between ages 18 and 40 through the Amazon Mechanical Turk platform. Subjects were directed to the web-based software. The protocol for this experiment is called simultaneous labeling protocol and it is executed as follows. In order to make subjects familiar with the task, a short practice video is presented. After this practice phase, subjects are directed to the baseline phase. In the baseline phase subjects are asked to provide breakpoints when they see a cross in a video sequence. This way, the amount of lag in the responses of each subject can be measured during data normalization. In the last and main phase of the experiment subjects are presented with a performance video randomly selected from the pair of videos, one of which contains the gestures with marks on the canvas and the other only the gestures. In this phase they are instructed to provide breakpoints and labels for both breakpoints and segments.

Full instructions for the main phase were as follows:

1. "This is the same task as in the preparatory phase but with a different video."
2. "In the next section a drawing performance video will start playing when you press the *Start* button."
3. "Your task is to press the *spacebar* to indicate the important moments in the video. What makes an important moment is up to you."
4. "When you press the key, video will stop and a pop-up form will appear and the video will not continue until you provide your descriptions for both the breakpoints and segments."
5. "Think of a breakpoint as a mark for that important moment. A description for a breakpoint should be about what happened at that exact moment."
6. "Whereas, a segment is defined as the period from the previous breakpoint and the current breakpoint. A description for a segment should be about what happened during that period."
7. "If you produce less than 3 breakpoints, you will have to repeat this task.".

Analysis

In this section I will share the details of the approaches I've taken to interpret the results of the experiment defined in the previous section. Then, I will discuss the implications of my analysis for the larger thesis detailed in this work.

Results

Out of 70 subjects who participated in the experiment, 37 of them segmented the video that contain both gestures and marks and 33 of them segmented the video that contain gestures only. In order to account for the lags in breakpoints and to normalize the breakpoints for each subject the following procedure is used. From the breakpoints of that subject, Gaussian kernel density [Rosenblatt, 1956] is computed. In choosing the appropriate bandwidth for the densities, a search algorithm has been used provided by the scikit-learn package [Pedregosa et al., 2011] in Python. This function computes the best bandwidth value by finding the density that best fits the data. Criteria for the best fit is determined by the maximum likelihood estimation. This way, data received from each subject has been transformed into a probability density. This allows for calculating the correlation of the computed densities with each other. Therefore,

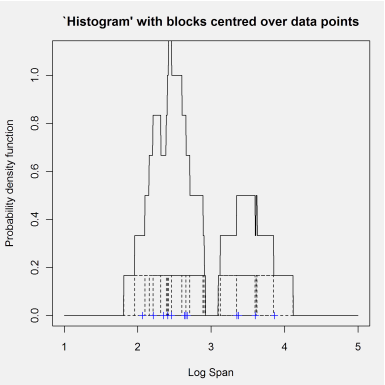

Figure 13: Kernel density estimation [Rosenblatt, 1956] centers blocks with a certain bandwith around the data points. Adding up the intersections results in a smoother distribution compared to histograms. Image copyright: Duong.

- A breakpoint set from a subject that gave the largest average correlation with the signals produced out of of the remaining breakpoint sets is selected as the baseline set,
- Another subject is selected from the group,
- Different lag values ranging between -100ms to 100ms in 10ms intervals are added to the breakpoints of the second subject,
- Gaussian kernel density is computed for each of the new breakpoint sets of that subject,

- The lag value that gave the best correlation with the first density is added to each breakpoint in the breakpoint set of the second subject,
- Adjusted breakpoints are added to the group set.

The procedure above is repeated for all the remaining subjects in the group to generate a new lag-adjusted breakpoints set. Then, for each group, kernel densities are computed from the sets of breakpoints. Optimal bandwidth values for the density estimates for each distribution are calculated with the same procedure described above. The fig. 14) shows both distributions.

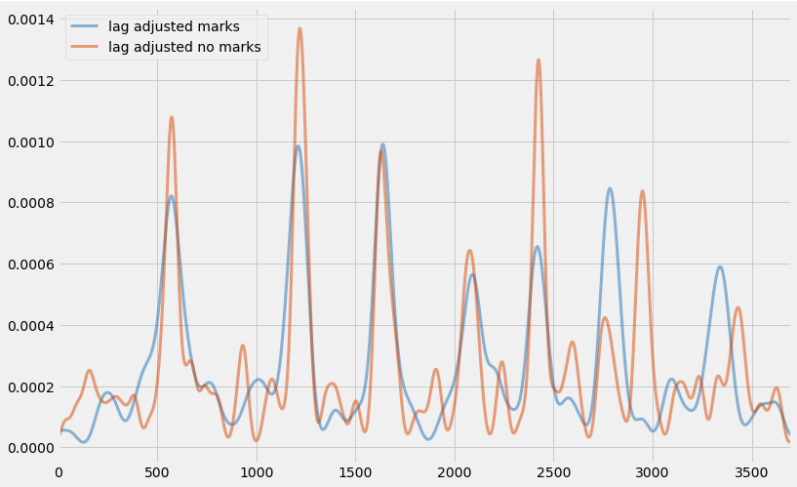

Figure 14: Comparison of the lag adjusted responses for both videos.

Distributions defined by the kernel densities are non-parametric. Therefore, in order to compare the two distributions, two-sample Kolmogorov-Smirnov(KS) test is used. Two-sample Kolmogorov-Smirnov test statistic is given by,

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|, \tag{1}$$

where $F_{1,n}(x)$ and $F_{2,m}(x)$ are the distribution functions of two samples. Calculated test statistic and p value obtained by comparing the two samples are given by $D = 0.054$ and $p = 0.104$. Furthermore, within group distributions are also tested against the uniform distributions in $10,000$ samples by using the KS tests. Where it yield $p < 0.0005$.

Computing the Video Features

Furthermore, subject responses are compared with the features derived from the video. The feature used in the analysis is sum

of absolute differences between the pixel intensities of consecutive video frames (see fig. 15). Therefore, this feature gives the moments where most change occurred. Moreover, as the pixel values in the background of the performance videos were uniform, these changes are more likely to correspond to the motions of the designer.

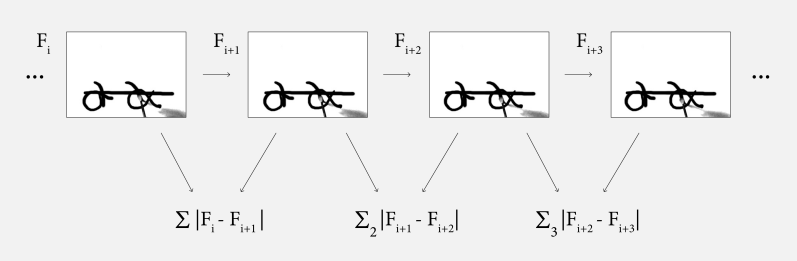

Figure 15: Absolute differences between consecutive frames are calculated.

Before computing the absolute differences, video frames are transformed into jpeg images and resized to 500 x 280 pixels. Then, means of the channels (R, G, B) are subtracted from the corresponding channel values in each image. Finally, the sum of absolute differences between the consecutive frames of videos are mapped to the range between 0 and 1 and plotted against the corresponding lag-adjusted subject responses.

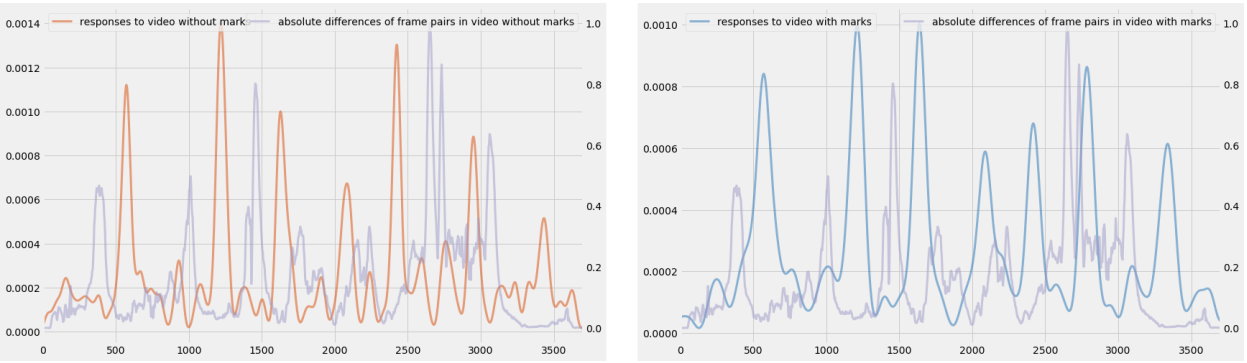

Figure 16: Comparison of the (left) responses to the video without marks against absolute differences of frame pairs in the video without marks, (right) of the responses to the video with marks against absolute differences of frame pairs in the video with marks.

Discussion

Now that the results are presented, we can discuss their implications.

Subjects Responded Similarly to both Videos

Comparing the distributions of the lag adjusted responses to both videos with a KS-test yield a p-value ($p = 0.104$) that lets us to

make the argument that we cannot reject the hypothesis that two distributions are the same. This observation is supported with the visual inspection of the two distributions (see fig. 14). Based on these, it would not be misguided to claim that the aggregate responses in both groups are similar. In other words, a clear signal is present within each group and this signal is correlated between the two groups.

Low-level Video Features Affect Segmentations

When the two distributions are plotted with the differences between frame pairs, it can be seen that the number of breakpoints increases after bursts of motion. This might be due to two reasons. Initially, the bursts of motion correspond to the drawings of shapes followed by the moments where the brush gradually slowed down and lifted off from the canvas. Therefore, subjects chose those moments because of the differential changes in motion such as changes in the speed of the brush or in the direction of the hand during the lift off. This makes clear that the gestures played a larger role in the observers' segmentation of the drawing event than the design itself.

Based on this observation we can propose that the agreement in responses across subjects can be predicted by the features of the gestures computed from the image sequences.

In order to emphasize the importance of this proposal, let's rewind back to my larger thesis. I suggest that the scope of individual interpretations of a design event are based on how people divide up that event. This can be observed from the fact that the labels provided by each subject are determined by the segments produced by the subjects themselves. I believe that segments defined by the breakpoints and the labels together constitute an individual story about how this design process is understood. Therefore, I found that the majority of subjective interpretations of a creative event are based on regularities that are computable from the sequential motion cues.

This finding can be utilized in developing computational design assistants. For example, with this knowledge, we can build computational agents that know which moments to focus on in a design event. Moreover, supplying these agents with the capacity to divide up long sequences into small, manageable and coherent pieces will help them achieve greater computational flexibility. We can then let AI systems learn from videos or by directly observing designers. These learning systems would attain a greater understanding of the design events they observe by segmenting them

based on computational regularities in the sequential visual data. Furthermore, we can reach this design goal without giving away the system's capacity to redefine the pieces altogether. I believe that these developments could help these systems explore new possibilities and eventually create their own designs.

Although I primarily studied the moments with significant agreement across subjects, we can also learn from the moments where there is little agreement. I suspect that these moments would more likely feature labels that indicate different reasoning patterns. The methodology proposed in this work enables researchers to identify these idiosyncratic cases. Hopefully, careful analysis of both convergent and divergent moments can be used to build a better understanding of design processes.

Contributions

During the course of this work, I have:

- compiled previous work on event segmentation and discussed their scope and limitations,
- outlined a methodology enabling the comparative study of how multiple observers observe how people design,

  This is one of the first methodologies that examines how multiple observers segment creative design events.

- developed a software enabling researchers to reliably collect segmentation data from events,

  Researchers can use this software to compose and execute different experimental protocols related to the segmentation of visual events. Furthermore, this software can be leveraged to collect segmentation data from multiple annotators in order to create a dataset comprised of better annotated videos. This can help us produce learning algorithms that are competent enough to deal with events in subtler ways.

- identified regularities in the perception of design events,
- demonstrated that the causes of the perceptual regularities can be computed from the low-level features of image sequences (videos),
- showed that the computational principles suggested by the experimental observation can be used to develop design assistants that better understands designers' actions and intentions.

Bibliography

K. Anders and Herbert A. Simon. Verbal reports as data. *Psychological Review*, 87(3):215–251, 1980. ISSN 1939-1471 0033-295X. DOI: 10.1037/0033-295X.87.3.215.

Dare Baldwin, Annika Andersson, Jenny Saffran, and Meredith Meyer. Segmenting dynamic human action via statistical structure. *Cognition*, 106(3):1382–1407, March 2008.

Jeanne Bamberger and Donald A. Schon. Learning as Reflective Conversation with Materials: Notes from Work in Progress. *Art Education*, 36(2):68–73, 1983.

N. Bayazit. Investigating Design: A Review of Forty Years of Design Research. *Design Issues*, 20(1):16–29, January 2004. ISSN 0747-9360. DOI: 10.1162/074793604772933739.

Zafer Bilda, John S. Gero, and Terry Purcell. To sketch or not to sketch? That is the question. *ResearchGate*, 27(5):587–613, September 2006. ISSN 0142-694X. DOI: 10.1016/j.destud.2006.02.002. URL https://www.researchgate.net/publication/222424553_To_sketch_or_not_to_sketch_That_is_the_question.

Durell Bouchard and Norman I. Badler. Segmenting Motion Capture Data Using a Qualitative Analysis. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, MIG '15, pages 23–30, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3991-9. DOI: 10.1145/2822013.2822039. URL http://doi.acm.org/10.1145/2822013.2822039.

Tarn Duong. An introduction to kernel density estimation. URL http://www.mvstat.net/tduong/research/seminars/seminar-2001-05/.

Karl Anders Ericsson and Herbert Alexander Simon. *Protocol analysis*. MIT press Cambridge, MA, 1993.

M. A. Goodale and G. K. Humphrey. The objects of action and perception. *Cognition*, 67(1-2):181–207, July 1998. ISSN 0010-0277.

Alex Graves. Generating Sequences With Recurrent Neural Networks. *arXiv:1308.0850 [cs]*, August 2013. URL http://arxiv.org/abs/1308.0850. arXiv: 1308.0850.

Bridgette M. Hard, Barbara Tversky, and David S. Lang. Making sense of abstract events: Building event schemas. *Memory & Cognition*, 34(6):1221–1235, September 2006. ISSN 0090-502X, 1532-5946. DOI: 10.3758/BF03193267. URL http://link.springer.com/article/10.3758/BF03193267.

Bernhard Hommel. The theory of event coding (TEC) as embodied-cognition framework. *Front. Psychol.*, 6, 2015. ISSN 1664-1078. DOI: 10.3389/fpsyg.2015.01318. URL http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01318/full.

Hao Jiang and C. Yen. Protocol analysis in design research: a review. *IASDR*, 78(24):16, 2009. URL http://www.academia.edu/download/32378718/Protocol_Analysis_in_Design_Research_-_a_Review.pdf.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Christopher A. Kurby and Jeffrey M. Zacks. Segmentation in the perception and memory of events. *Trends in cognitive sciences*, 12(2):72–79, February 2008. ISSN 1364-6613. DOI: 10.1016/j.tics.2007.11.004. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2263140/.

William Lotter, Gabriel Kreiman, and David Cox. Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. *arXiv:1605.08104 [cs, q-bio]*, May 2016. URL http://arxiv.org/abs/1605.08104. arXiv: 1605.08104.

Darren Newtson. Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1): 28–38, 1973. ISSN 1939-1315 0022-3514. DOI: 10.1037/h0035584.

Darren Newtson. The Perceptual Organization of Ongoing Behavior. *Journal of Experimental Social Psychology*, 1976.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Murray Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.*, 27(3): 832–837, September 1956. ISSN 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177728190. URL https://projecteuclid.org/euclid.aoms/1177728190.

Peter H. Schiller and Edward J. Tehovnik. *Vision and the Visual System*. Oxford University Press, Oxford ; New York, 1 edition edition, August 2015. ISBN 978-0-19-993653-3.

Donald A. Schon and Glenn Wiggins. Kinds of seeing and their functions in designing. *Design Studies*, 13(2):135–156, April 1992. ISSN 0142-694X. DOI: 10.1016/0142-694X(92)90268-F. URL http://www.sciencedirect.com/science/article/pii/0142694X9290268F.

Nicole K. Speer, Khena M. Swallow, and Jeffery M. Zacks. Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4): 335–345, December 2003. ISSN 1530-7026, 1531-135X. DOI: 10.3758/CABN.3.4.335. URL http://link.springer.com/article/10.3758/CABN.3.4.335.

Nicole K. Speer, Jeffrey M. Zacks, and Jeremy R. Reynolds. Human Brain Activity Time-Locked to Narrative Event Boundaries. *Psychological Science*, 18(5):449–455, 2007. ISSN 0956-7976. URL http://www.jstor.org/stable/40064637.

E. H. Spriggs, F. De La Torre, and M. Hebert. Temporal segmentation and activity classification from first-person sensing. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 17–24, June 2009. DOI: 10.1109/CVPRW.2009.5204354.

Masaki Suwa and Barbara Tversky. What do architects and students perceive in their design sketches? A protocol analysis. *Design Studies*, 18(4):385–402, October 1997. ISSN 0142-694X. URL https://keio.pure.elsevier.com/en/publications/what-do-architects-and-students-perceive-in-their-design-sketches.

B. Tversky and Jeffrey M. Zacks. Event perception. In *The Oxford Handbook of Cognitive Psychology*, pages 83–95, Oxford, July 2012. Oxford University Press.

Shimon Ullman. *High-level vision: object recognition and visual cognition*. MIT Press, Cambridge, Mass, 1996. ISBN 978-0-262-21013-3.

Felix Warneken and Michael Tomasello. Altruistic helping in human infants and young chimpanzees. *Science*, 311(5765):1301–1303, March 2006. ISSN 1095-9203. DOI: 10.1126/science.1121448.

Jeffrey M. Zacks and Khena M. Swallow. EVENT SEGMENTATION. *Current directions in psychological science*, 16(2):80–84, April 2007. ISSN 0963-7214. DOI: 10.1111/j.1467-8721.2007.00480.x. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3314399/.

Jeffrey M. Zacks, Todd S. Braver, Margaret A. Sheridan, David I. Donaldson, Abraham Z. Snyder, John M. Ollinger, Randy L. Buckner, and Marcus E. Raichle. Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6):651–655, 2001. ISSN 1097-6256. DOI: 10.1038/88486. URL https://uncch.pure.elsevier.com/en/publications/human-brain-activity-time-locked-to-perceptual-event-boundaries.

Jeffrey M. Zacks, Khena M. Swallow, Jean M. Vettel, and Mark P. McAvoy. Visual motion and the neural correlates of event perception. *Brain Research*, 1076(1):150–162, March 2006. ISSN 0006-8993. DOI: 10.1016/j.brainres.2005.12.122.