Neural Dynamics of Natural Vision

Carl Olsson

4/6/2014

Brown University
Department of Engineering

Contents

Abstract

This project attempts to gain insight into how the cortex functions during natural vision using electrocorticographic (ECoG) data collected from human epileptic patients at Rhode Island Hospital who underwent surgical treatment for their condition. Our approach offers a major departure from existing research in brain sciences. Instead of simplistic, highly constrained stimuli, patients are shown a movie/film of their choosing while we record neural and eye tracking data. Using this data, we can study the brain mechanisms underlying natural everyday vision using complex dynamic visual scenes – when participants are free to move their eyes and shift their attention. Also, since the viewing condition is not cognitively taxing and is actually pleasurable (compared to typical behavioral electrophysiological experiments), the amount of data we can acquire with this experimental approach far surpasses that of a conventional experiment.

1. Introduction

1.1. Motivation

Most previous work in neuroscience has focused on the brain mechanisms underlying the rapid recognition of simple and often artificial, static, isolated stimuli. These experimental paradigms attempt to control as many variables as possible and isolate or randomize potentially confounding variables. Although these experiments have obvious advantages and were effective in the past, such experimental protocols seem to overlook the distinctive complexity of natural, everyday perception. In reality, the world is full of dynamic and highly complex. For example, natural visual scenes usually consist of many objects embedded in background clutter. Despite this, very little is known about how our visual cortex deals with these noisy and ambiguous images.

Our approach offers a major departure from existing research in brain sciences. Instead of simplistic, highly constrained stimuli, patients are shown a movie/film of their choosing while we record neural and eye tracking data. Using this data, we can study the brain mechanisms underlying natural everyday vision using complex dynamic visual scenes – when participants are free to move their eyes and shift their attention. Also, since the viewing condition is not cognitively taxing and is actually pleasurable (compared to typical behavioral electrophysiological experiments), the amount of data we can acquire with this experimental approach far surpasses that of a conventional experiment.

The data for this project was collected from patients who have intractable epilepsy. Using an invasive electrophysiological technique called electrocorticography (ECoG) we were able to collect neural data at much higher temporal and spatial resolution than traditional, non-invasive imaging techniques. This method uses intracranially implanted electrodes which clinicians use during resective surgical treatment for intractable epilepsy.

In this treatment, the patient undergoes a craniotomy and is implanted with intracranial electrodes used to find epileptogenic zones. Typically, the patient is monitored for seizures for over a week. We used this chunk of idle time to collect neural data from patients using the same electrodes they are being monitored with. In addition to neural data, we record eye tracking data using a head-free eye tracking device. The combination of eye tracking data and precise annotation of the videos shown to the patient means we know exactly what they see.

1.2. Goals

The primary goal of the research questions mentioned above is to provide a comprehensive characterization of the functional anatomy of natural vision, which will, in turn, inform the development of computational models of the visual cortex.

The intellectual merits of this project include:

1. Significant advances in our scientific understanding of the computational mechanisms underlying object recognition, scene processing, and visual attention
2. Development of rigorous computational methods to leverage big neuroscience data
3. Development of large-scale annotated stimulus sets and physiological predictions.

2. Methods

2.1. Neural Data Acquisition

In previous work, Prof. Thomas Serre has shown that abstract object category information can be decoded from occipital and inferior temporal areas by using neural decoding techniques (see [ZMB+11, RTS10]). Our previous research experience with this type of patient population has proven that the clinical criteria for electrode placements typically provide sufficient coverage to address the research questions we formulated above.

The experimental data has been recorded at Rhode Island Hospital as part of the NeuroPort project. The NeuroPort project (Investigation of Neuronal Activity in Patients with Epilepsy at Rhode Island Hospital; Potter, PI) is an IRB-approved, active protocol permitting electrophysiologic recording from scalp electrodes, clinical subdural macroelectrodes (grid, strip, and depth electrodes) and research microelectrodes and microwires implanted in patients who are already undergoing craniotomy and implantation of clinical electrodes for epilepsy surgery evaluation. Participants also permit us the use of clinical and research data collected (EEG, ECoG, video, CT/MRI, and medical record).

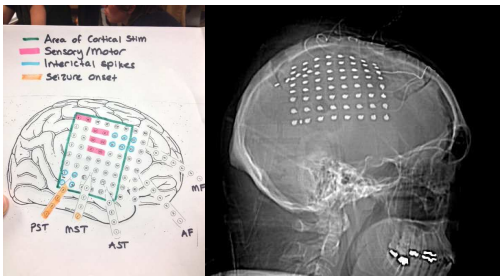

Figure 2.1.: Electrode placements for subjects #1 and #2 respectively

Two patients have been successfully recorded at Rhode Island Hospital. The electrode placements are shown above in Fig. 2.1. The neural data was acquired via the Blackrock Microsystems NeuroPort system. They were pre-processed using a notch filter at 60 Hz to remove line noise.

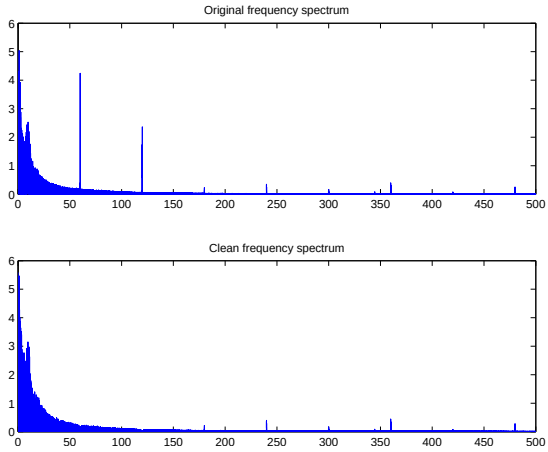

Figure 2.2.: Frequency spectrum before/after neural data pre-processing

2.2. Gaze Data Acquisition

To acquire the gaze data during the experiment, we used the SR Research Eyelink 1000 system. We chose this system because it is head-free. In other words, the patient does not need to wear any equipment and can simply lie in his/her hospital bed. To correct for variations in head pose, the Eyelink system tracks the head pose using a small sticker worn on the forehead. In addition, the Eyelink offers high spatial resolution (sub-degree gaze accuracy), pupilometry data acquistion, and a high sampling rate.

The Eyelink system consists of a host PC running SR Research's custom host operating system and software.

2.3. Data Synchronization

Data synchronization was a big issue in this experiment as we needed extremely high temporal precision in our neural data analysis. Four things needed to be synchronized: visual stimuli, audio stimuli, the neural data , and the eye tracking data. To do this, we custom built a experiment apparatus that delivered stimuli with extremely high temporal precision.

We used a National Instruments Data Acquisition (DAQ) card to send digital triggers into the Blackrock Microsystems NeuroPort system. Triggers contained frame information and included other useful information such as frame refresh, trial onset, movie start, movie paused, movie ended, etc. To synchronize the start of recording with the Eyelink, we had the behavioral computer send messages over ethernet to the Eyelink. In addition, we used the analog output of the Eyelink as extra validation that our data synchronization was accurate.

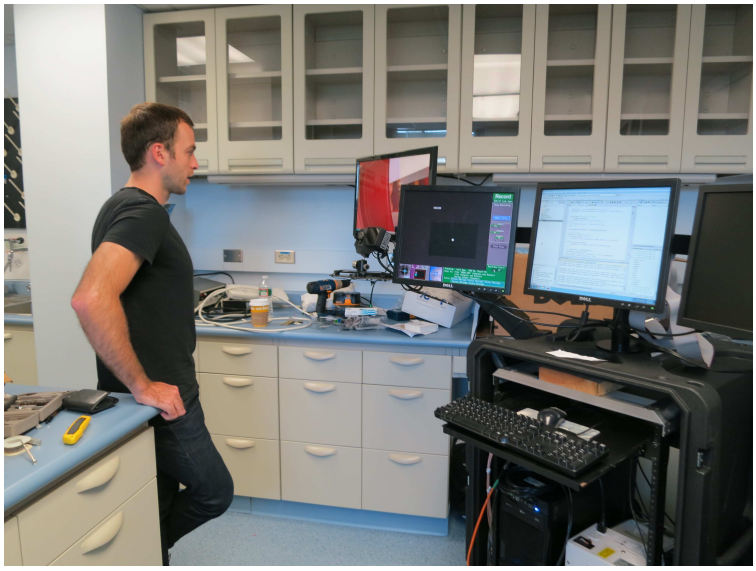

Figure 2.3.: Tracking gaze with the SR Research Eyelink 1000 system

Audiovisual stimuli were presented using gstreamer plugin for Psychtoolbox, a psychophysics toolbox for MATLAB. This toolbox allowed us to present our stimuli with extremely little lag and high temporal precision. We presented the visual stimuli on a 27" monitor at 120 Hz. The high screen refresh rate allows us to play video at higher frame-rates, even though for these experiments the source files were all around 25 Hz.

	Video Title	Duration (min:sec)	Number of saccades	Faces annotated	Shot boundaries
Subject #1	Caso Cerrado #1	8:00	1,185	350	132
	Caso Cerrado #2	8:06	1,087	350	132
Subject #2	Anchorman	97:25	7,382	3,950	1,400
	Parks & Recreation	21:28	1,203	1,050	400

Table 2.1.: Data statistics

After the data acquisition, we analyze the content of the selection of videos that the participant could choose from, to the extent that objects and individuals in every scene are identified and localized. The analysis phase for neural data will commence by aggregating the information from both the experimental and computer vision sides. This aggregate information will reveal the visually driven brain states associated with object and scene recognition during natural viewing.

Figure 2.4.: Building the experimental apparatus

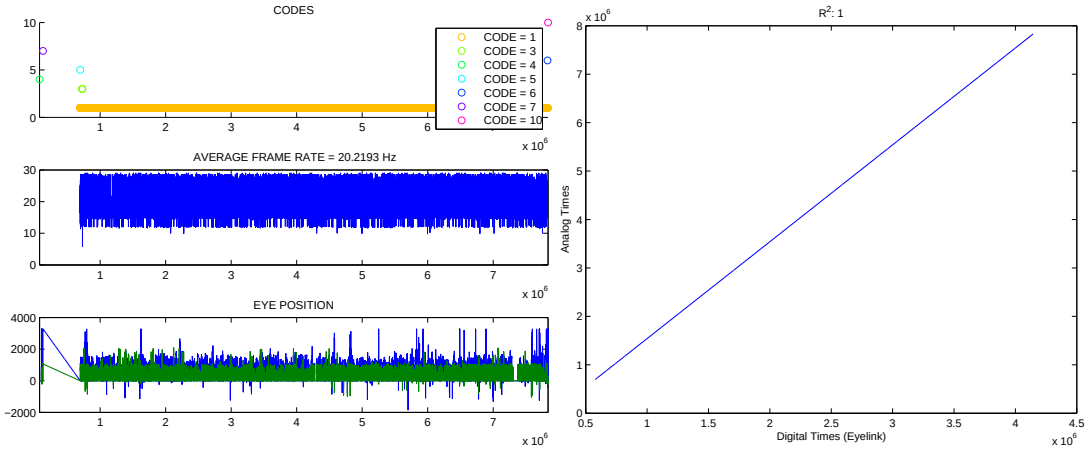

Figure 2.5.: Digital/analog trigger synchronization

2.4. Automated Video Annotation

GOING TO RE-WRITE / PARAPHRASE

On the computer vision side, the primary objective has been to build a computer vision architecture for automated frame-based annotation of the content in the commercial movie and television productions. To leverage the manual annotation of hundreds of thousands of video frames (which would be prohibitively long by hand), we exploit recent innovations in computer vision and machine learning. Our own prior work in automated behavioral analysis of rodent behaviors ([JGY+10]) suggests that computer vision is sufficiently mature to help automate the annotation of objects and actions in videos. This system provides an in-depth analysis of the visual content of the scenes, showing the precise locations and identity of fixated objects in the scene. This analysis will be done off-line and will be supervised by a

human observer who will interact with the system when uncertainty in identification is present, refining the system's predictions for the future.

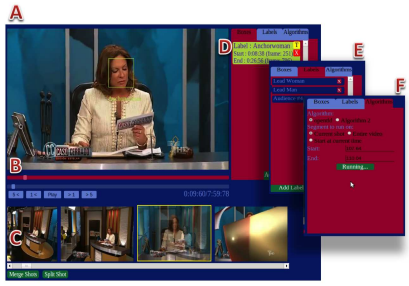

Figure 2.6.: Annotation interface screenshots

At its core the Serre Lab's video annotation interface is a rich HTML5 video player, which streams media from a central server that also hosts the database system. This new video annotation tool allows us to easily and efficiently create accurate annotations from lengthy videos. By harnessing the new and powerful features of HTML5 and CSS3, we were able to develop a computer vision tool that runs in all major web browsers. The process for annotating a video starts with an automatic "shot-detection" that segments the video into discrete shots. The interface then automatically detects all the faces in the video and assigns bounding boxes to them.

The interface can be used for annotation in two ways: Automatic face detection algorithms built-in to the interface can detect the all of the faces in a given movie (or a part of the movie) and the annotator would assign identities to these detections. The user can correct/edit all of the automatic detections and create new ones manually. Moreover, the user can select arbitrary objects in any scene, including but not limited to faces, and track their movement throughout the scene by using an automated video tracker. All of the annotations are stored in one central server in real-time without any possibility of data loss. This workflow lets users easily identify and keep track of detections, something that took many hours of tedious work with previous methods. In fact, with the use of modern user interface concepts, our estimate is that this dataset reduces the frame-by-frame annotation time of videos by hundreds of orders of magnitude. The interface also has built-in tools to facilitate collaborative work. All changes on the annotations are saved to the same central database, allowing for users to quickly and dynamically see what others have done without overwriting.

By integrating a well-established work manager (Asana) in the interface researchers can click a single button and all of the annotators assigned to the project in the work group will receive a task on their todo-list that outlines the specific shots in the video they are required to annotate.

After selecting an algorithm, an annotator can use the current shot, the whole video, or manually enter start and end times to specify the segment of the video to run

the algorithm on. Then all they have to do is click the ‘Run’ button and wait for it to finish processing. That’s it - the interface backend automatically inserts the data into its proper place in the central database and gives the results back to the user in real time, updating the interface display with the new detections.

2.5. Neural Decoding

Acknowledgments

Thanks to Thomas Serre, Ali Arslan, Leigh Hochberg, Wilson Truccolo, Wael Assad, Steve Potter and the rest of Neurosurgery

A. Title of the first appendix chapter

A.1. Overview

A.2. The next section

Bibliography

[JGY+10] Hueihan Jhuang, Estibaliz Garrote, Xinlin Yu, Vinita Khilnani, Tomaso Poggio, Andrew D Steele, and Thomas Serre. Automated home-cage behavioural phenotyping of mice. *Nature communications*, 1: 68, 2010.

[RTS10] Leila Reddy, Naotsugu Tsuchiya, and Thomas Serre. Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, 50: 818–825, 2010.

[ZMB+11] Ying Zhang, Ethan M Meyers, Narcisse P Bichot, Thomas Serre, Tomaso A Poggio, and Robert Desimone. Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108: 8850–8855, 2011.