

# **Learning and representation of declarative memories by single neurons in the human brain**

Thesis by

**Ueli Rutishauser**

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2008

(Defended May 8th, 2008)

© 2008

Ueli Rutishauser

All Rights Reserved

## Acknowledgments

The research presented in this thesis would not have been possible without the help, advice, and encouragement of a large group of people. While this thesis bears my name, this is partially their work too. I feel exceedingly fortunate in having such a great group of friends, colleagues, and mentors, all of whom contributed to this work.

First, I would like to thank my advisors Erin Schuman, Christof Koch, and Adam Mamelak. Their continued support, guidance, and intellectual freedom allowed me to pursue my interests. The work presented in this thesis was primarily done under the mentorship of Erin Schuman and Adam Mamelak. While initially intimidating, I'm grateful that I was given the opportunity to build, from scratch, a human single-unit recording operation. This was a large endowment of trust, given that I had little experience in neuroscience, let alone electrophysiology or medicine. It was a formative experience, which could not have succeeded without the continued guidance and support of Erin Schuman. The members of her laboratory created a nurturing environment, despite me working on a topic far removed from the Lab's mainstream. Alana Rathbun provided much needed assistance with numerous administrative tasks.

While the work I produced under the mentorship of Christof Koch is not part of this thesis, it had an unquestionable impact on my current thinking as reflected in this thesis. I admire his deep and intense love of science and the search for truth. I would like to thank Christof for all his support, encouragement, enthusiasm, and the nurturing environment that "K-lab" provided for me.

I also received advice from the other members of my committee: John O'Doherty and Gilles Laurent. In addition, I had the pleasure of interacting with Ralph Adolphs, John Allman, Henry Lester, Thanos Siapas and Pietro Perona. My first home at Caltech was Pietro Perona's lab, of which I have fond memories. I'm grateful for the freedom he gave me to explore my interests, which eventually led to this work.

Much of the experimental work I performed took place at Huntington Memorial Hospital (HMH). It was a pleasure to interact with people at HMH, in particular the EBMP (Epilepsy&Brain Mapping) nurses, staff, and physicians. Their help is greatly appreciated—many of these experiments would not have been possible without the continued support. I appreciate help with the Neuropsychological evaluations by Linda Philpott. Most of all, I would like to thank the many patients (whom I not allowed to list by name) for their participation. I'm very grateful for their willingness to participate during a time of great distress and uncertainty in their lives.

Over the years, I have greatly benefited from discussions about my research with Costas Anastassiou, Moran Cerf, Tobi Delbrueck, Daniela Dieterich, Rodney Douglas, Wolfgang Einhaeuser, Michael Fink, Alan Hampton, Jonathan Harel, Alexander Huth, Constanze Hofstoetter, Alex Holub, Hiroshi Ito, Vivek Jayaraman, Andreas Kotowicz, Sally Kim, Gabriel Kreiman, Sotiris Masmanidis, Kevan Martin, Florian Mormann, Dirk Neumann, Zoltan Nadasdy,

Kerstin Preuschoff, Johannes Schwarz, Bryan Smith, Michael Sutton, Chin-Yin Tai, and Nao Tsuchiya (in alphabetical order).

I had the opportunity to mentor and work with Caltech undergraduates, SURF students and master students Arjun Basal, Matt McKinley, Anthony Chong, Davhyd Wing and Andreas Kotowicz. It was a pleasure to work with them.

I would not be at Caltech today without the help and encouragement of my previous advisor, Rodney Douglas. I'm grateful for the opportunity he gave me by exposing me to the most interesting research problems. I would also like to thank the faculty members at my undergraduate institution HSR that have supported me in this endeavor—Josef Joller, Gabor Hardy and Andreas Mueller. Thanks also to Gaston Wolf of ZHW for advice in transferring to graduate school.

My way into graduate school was not as direct as it perhaps could have been—I spent several years in the software industry. Several people from this time have been influential in my personal development and I would not be where I'm today without them. Rene Berchten was my first mentor and taught me many of the fundamental software skills I still use today. Enzo Giannini always supported my efforts and was very helpful in the transition between business and academia. I would also like to thank my business partner Alain Schaefer of many years—It was a pleasure running easc together. The results in this thesis are the product of an estimated 60'000 lines of code (mostly Matlab). The skills I was fortunate to acquire have thus been tremendously useful.

Big thanks go to my friends, both here at Caltech and in Switzerland. I'm grateful for their friendship, support, and time spent away from the lab; much of this work could not have succeeded without them: Maik Berchten, Oliver Bandle, Moritz Delbrueck, Nicolas Foong, Patrick Gaus, Kasia Gora, Jonathan Harel, Alexander Huth, Prashant Joshi, Rene Lavanchy, Christof Moser, Tony Oertli, Bobby Rohrkemper, Bryan Smith, Alain Schaefer, and Aaron Wittmann (in alphabetical order). Thanks to my fellow class-of-2008 members: Anusha Narayan, Hilary Gliden, Matthew Nelson, Dirk Neumann, Ming Gu, Casimir Wierzynski, Will Ford, and Jason Rolfe.

The California Institute of Technology and the Computation & Neural Systems (CNS) program has been an ideal home institution, providing the best possible environment to explore my diverse research interest. Much of my academic thinking has been shaped here. I'm grateful to have been part of Caltech. The group of people that make up the Caltech community is truly remarkable. I specifically would also like to thank International Student Programs, the Office of Technology Transfer, the Center for Neuromorphic Engineering, and the Library for help over the course of my stay. It was very rewarding to be at a place where intense dedication to science and technology is the norm rather than exceptional.

Last, but not least, I would like to thank my family who has continually supported me unconditionally, despite the long distance to Switzerland.

## **Abstract**

Episodic memories allow us to remember not only that we have seen an item before but also where and when we have seen it (context). Neurons in the medial temporal lobe (MTL) are critically involved in the acquisition of such memories. Since events happen only once, the ability to distinguish novel from familiar stimuli is crucial in order to rapidly encode such events after a single exposure. Theoretically, this is a hard learning problem (single-trial learning). Yet, successful detection of novelty is necessary for many types of learning. During retrieval, we can sometimes confidently report that we have seen something (familiarity) but cannot recollect where or when it was seen. Thus episodic memories have several components which can be recalled selectively. We recorded single neurons and local field potentials in the human hippocampus, amygdala, and anterior cingulate cortex while subjects remembered, and later retrieved, the identity and location of pictures shown. We describe two classes of neurons that exhibit such single-trial learning: novelty and familiarity detectors, which show a selective increase in firing for new and old stimuli, respectively. The neurons retain memory for the stimulus for at least 24 h. During retrieval, these neurons distinguish stimuli that will be successfully recollected from stimuli that will not be recollected. Similarly, they distinguish between failed and successful recognition. Pictures which were forgotten by the patient still evoked a non-zero response. Thus, their response can be different from the decision of the patient. Also, we demonstrate that listening to these neurons (during retrieval) enables a simple decoder to outperform the patient (i.e., it forgets fewer pictures). These data support a continuous strength of memory model of MTL function: the stronger the neuronal response, the better the memory (as opposed to a dual-process model). I also describe specific power increases in specific frequencies of the local field potential that are predictive of later retrieval success. These neural signatures, recorded during learning, thus indicate whether plasticity was successful or not.

## Table of Contents

<b>Acknowledgments</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>v</b>
<b>List of Figures</b> .....	<b>ix</b>
<b>List of Tables</b> .....	<b>xi</b>
<b>List of Abbreviations</b> .....	<b>xii</b>
<b>Chapter 1. Introduction</b> .....	<b>1</b>
1.1 Memory.....	1
1.2 Anatomy, connectivity, and function of the medial temporal lobe.....	3
1.3 Mechanisms of plasticity—Circuit and single-cell properties .....	11
1.4 Temporal lobe epilepsy .....	16
1.5 Electrophysiology in epilepsy patients .....	20
1.6 The anterior cingulate cortex.....	26
<b>Chapter 2. Online detection and sorting of extracellularly recorded action potentials 36</b>	
2.1 Introduction.....	36
2.2 Methods .....	40
2.3 Results and Discussion.....	52
2.4 Conclusions and relevance.....	81
2.5 Appendix A — Signal processing and spike detection .....	84
2.6 Appendix B — Online sorting .....	87
2.7 Appendix C — Spike realignment.....	89
2.8 Appendix D — Projection test.....	90
<b>Chapter 3. Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex</b> .....	<b>94</b>
3.1 Introduction.....	94
3.2 Results.....	95
3.3 Discussion.....	110
3.4 Experimental procedures.....	117
3.5 Supplementary material.....	120

3.6 Supplementary figures.....	127
<b>Chapter 4. Activity of human hippocampal and amygdala neurons during retrieval of declarative memories .....</b>	<b>131</b>
4.1 Introduction.....	131
4.2 Results.....	133
4.3 Discussion.....	145
4.4 Methods .....	152
4.5 Supplementary results .....	154
4.6 Supplementary discussion.....	159
4.7 Supplementary methods .....	164
4.8 Supplementary figures.....	172
4.9 Supplementary tables .....	181
<b>Chapter 5. Predictors of successful memory encoding.....</b>	<b>183</b>
5.1 Introduction.....	183
5.2 Methods .....	186
5.3 Results.....	194
5.4 Discussion.....	204
<b>Chapter 6. Summary and General Discussion.....</b>	<b>209</b>
6.1 Novelty detection and single-trial learning.....	209
6.2 The relationship between memory responses and behavior.....	212
6.3 Novelty and familiarity responses in the amygdala .....	213
6.4 Differential response strength in epileptic tissue.....	215
6.5 Predictors of successful learning.....	216
6.6 The value of studying single-unit responses in humans.....	217
6.7 Note on visual tuning of MTL neurons .....	218
<b>Chapter 7. Appendix A: Methods for signal-processing, analysis of spike trains, and local field potentials (LFPs) .....</b>	<b>220</b>
7.1 Signal acquisition .....	220
7.2 The origin and structure of the extracellular signal .....	222
7.3 Signal processing.....	224
7.4 Spike train analysis.....	234
<b>Chapter 8. Appendix B: Population decoding: principles and methods</b>	<b>238</b>
8.1 Motivation and principles.....	238
8.2 Definitions .....	243
8.3 Multiple linear regression.....	244
8.4 Regularized least-square regression (RLSC).....	244

**References ..... 247**



## List of Figures

Figure 2-1. Filtering and detection of spikes from continuously acquired data. ....	42
Figure 2-2. Schematic illustration of spike detection and sorting. ....	45
Figure 2-3. Autocorrelation of raw data. ....	48
Figure 2-4. Simulated raw signal (dataset 1). ....	65
Figure 2-5. Mean waveforms used for simulated dataset 2 and 3. ....	66
Figure 2-6. Performance comparison of different spike sorting algorithms. ....	71
Figure 2-7. Illustration of tools for evaluation of the sorting result, using real data. ....	75
Figure 2-8. Illustration of PCA analysis for one channel of real data. ....	76
Figure 2-9. Population statistics from the 76 neurons obtained from in vivo recordings. ...	77
Figure 3-1. Electrode placement and task design. ....	97
Figure 3-2. Example of a single hippocampal neuron during learning and recognition. ...	102
Figure 3-3. Population summary of all responding neurons. ....	106
Figure 3-4. Population decoding from simultaneously recorded neurons. ....	107
Figure 3-5. Behavioral performance of all subjects. ....	127
Figure 3-6. Population statistics for all neurons. ....	128
Figure 3-7. Single-neuron prediction probabilities. ....	129
Figure 3-8. Example of a novelty-sensitive neuron. ....	130
Figure 4-1. Experimental setup and behavioral performance. ....	135
Figure 4-2. Single cell response during retrieval. ....	139
Figure 4-3. Neuronal activity distinguishes stimuli that are only recognized (R-) from stimuli that are also recollected (R+). ....	143
Figure 4-4. Activity during errors reflects true memory rather than behavior. ....	145
Figure 4-5. Population average of all recorded neurons. ....	172
Figure 4-6. Population response. ....	173
Figure 4-7. A continuous strength of memory gradient exists when the hippocampal neuronal population is considered in isolation. ....	174
Figure 4-8. Whether a stimulus is new or old can be predicted regardless of whether recall was successful or not. ....	175
Figure 4-9. ROC analysis of the neuronal data for all 3 behavioral groups. ....	176

Figure 4-10. Population average of all recorded neurons.....	177
Figure 4-11. Comparison of trial-by-trial response strength for different subcategories of neurons.....	178
Figure 4-12. Sorting quality for the 412 recorded units. ....	179
Figure 4-13. Comparison of response strength across different recording sessions (days).180	
Figure 5-1. Retrieval performance (behavior) shown as a receiver operator characteristic (ROC) curve. ....	192
Figure 5-2. Retrieval performance for all subjects. ....	193
Figure 5-3. Example LFP traces (raw, theta, gamma).....	196
Figure 5-4. Example of LFP power difference due to memory. ....	197
Figure 5-5. Examples of DM effects from all three brain areas as well as different frequency ranges. ....	200
Figure 5-6. Summary of DM effects for all brain areas and frequencies.....	201
Figure 5-7 Example of LFP power difference, shown for learning trials that were retrieved on the same day (green and red) and overnight (magenta and blue). ....	203
Figure 5-8. LFP power can be used to predict overnight memory.....	204
Figure 7-1. Illustration of the complex morlet wavelet. ....	228
Figure 7-2. Illustration of the trade-off between specificity in time and frequency space. ....	229

**List of Tables**

Table 2-1. Simulation 1. ....	56
Table 2-2. Simulation 2. ....	61
Table 2-3. Simulation 3. ....	62
Table 2-4. Comparison of sorting results for the two different threshold estimation methods, as well as other algorithms. ....	68
Table 3-1. Electrode position in stereotactic coordinates (Talairach).....	122
Table 3-2. Location of resected tissue (temporal lobe lobectomy in each case). ....	126
Table 4-1. Neuropsychological evaluation of patients. ....	181
Table 4-2. Number of neurons recorded. ....	182

**List of Abbreviations**

ACC	anterior cingulate cortex
AMPA	alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic
AED	anti-epileptic drug
BOLD	blood-oxygen-level-dependent (see fMRI)
BDNF	brain-derived neurotrophic factor
Ca <sup>2+</sup>	calcium
CA1-3	Cornu ammonis fields (of the hippocampus)
CT	computerized tomography
Cl	chloride
CS	conditioned stimulus
CR	conditioned response
CTA	conditioned taste aversion
CV	coefficient of variation
CMA	cingulate motor area
DA	dopamine
DG	dentate gyrus
DLPFC	dorsolateral prefrontal cortex
EC	entorhinal cortex
EEG	electroencephalography
EPSP	excitatory postsynaptic potential
EPSC	excitatory postsynaptic current
ERP	event-related potential
ERN	error-related negativity

FDR	false discovery rate
FEF	frontal eye fields
fMRI	functional magnetic resonance imaging
Glu	glutamate (neurotransmitter)
GABA	$\gamma$ -aminobutyric acid (neurotransmitter)
IPSP	inhibitory postsynaptic potential
IPSC	inhibitory postsynaptic current
IT	inferotemporal cortex (monkey)
ISI	interspike interval
kstest	Komogorov-Smirnof goodness-of-fit test
K <sup>+</sup>	potassium
LFP	local field potential
LGN	lateral geniculate nuclei
LTP	long-term potentiation
LTD	long-term depression
LIP	lateral intraparietal area
MRI	magnetic resonance imaging
MEG	magneto-encephalographic
MTL	medial temporal lobe
MLREG	multiple linear regression
mPFC	medial prefrontal cortex
MT	middle temporal area of the cortex
MUA	multi-unit activity
MLE	maximum likelihood estimate

NMDA	N-methyl-D-aspartic acid (an amino acid)
Na <sup>+</sup>	sodium
OCD	obsessive-compulsive disorder
OFC	orbitofrontal cortex
PCA	principal component analysis
PET	positron emission tomography
PFC	prefrontal cortex
PPF	paired-pulse facilitation
RLSC	regularized least-square classifier
RT	reaction time
ROC	receiver operator characteristic
RMS	root-mean-square
SNR	signal-to-noise ratio (but also see below)
SNr	nigra pars reticulata, substantia nigra
SVM	support vector machine
STDP	short-time dependent plasticity
SUA	single-unit activity
STD	standard deviation
SE	standard error
STN	subthalamic nucleus
TE	anterior part of inferior temporal cortex (monkey)
TEO	posterior part of inferior temporal cortex (monkey)
TLE	temporal lobe epilepsy
UR	unconditioned response

US	unconditioned stimulus
V1	primary visual cortex, Brodman area 17.
V2	part of the extrastriate visual cortex , Brodman area 17.
V3	part of the extrastriate visual cortex , Brodman area 17.
V4	part of the extrastriate visual cortex , Brodman area 17.
VTA	ventral tegmental area