NeuroImage 70 (2013) 223–232

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Decoding the memorization of individual stimuli with direct human brain recordings

Marcel A.J. van Gerven [a,*,1], Eric Maris [a], Michael Sperling [b], Ashwini Sharan [c], Brian Litt [d], Christopher Anderson [d], Gordon Baltuch [e], Joshua Jacobs [f,**,1]

[a] Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands
[b] Department of Neurology, Thomas Jefferson University Hospital, Philadelphia, PA 19104, USA
[c] Department of Neurosurgery, Thomas Jefferson University Hospital, Philadelphia, PA 19104, USA
[d] Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA
[e] Department of Neurosurgery, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA
[f] School of Biomedical Engineering, Science & Health Systems, Drexel University, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Accepted 22 December 2012
Available online 5 January 2013

Keywords:
Electrocorticography
Machine learning
Visual perception
Working memory

ABSTRACT

Through decades of research, neuroscientists and clinicians have identified an array of brain areas that each activate when a person views a certain category of stimuli. However, we do not have a detailed understanding of how the brain represents individual stimuli within a category. Here we used direct human brain recordings and machine-learning algorithms to characterize the distributed patterns that distinguish specific cognitive states. Epilepsy patients with surgically implanted electrodes performed a working-memory task and we used machine-learning algorithms to predict the identity of each viewed stimulus. We found that the brain's representation of stimulus-specific information is distributed across neural activity at multiple frequencies, electrodes, and timepoints. Stimulus-specific neuronal activity was most prominent in the high-gamma (65–128 Hz) and theta/alpha (4–16 Hz) bands, but the properties of these signals differed significantly between individuals and for novel stimuli compared to common ones. Our findings are helpful for understanding the neural basis of memory and developing brain-computer interfaces by showing that the brain distinguishes specific cognitive states by diverse spatiotemporal patterns of neuronal.

Introduction

Over the last century, neuroscientists have made dramatic progress towards understanding the patterns of human brain activity that encode the properties of cognitive representations. Whereas early work suggested that all memory representations are stored in a fully distributed manner (Lashley, 1950; Přibram, 1991), modern studies show that the human brain has a modular organization, such that neuronal activity in different regions represents separate classes of cognitive information (Farah and McClelland, 1991; Mitchell et al., 2008; Warrington and Shallice, 1984). This work has identified a number of brain regions where large-scale neuronal activations occur when people process specific categories of information, such as faces (Kanwisher et al., 1997), scenes (Epstein et al., 1999), and animals (Martin et al., 1996), among many others (Mitchell et al., 2008).

In addition to category-wide neuronal patterns, an important additional question concerns how the brain differentiates individual memories within a category. Although human neuroscience studies have traditionally not focused on characterizing specific cognitive states, there is emerging evidence that individual neuronal network states can be probed using direct human brain recordings (Chang et al., 2010; Jacobs and Kahana, 2009; Manning et al., 2012; Morton et al., In press; Quian Quiroga et al., 2005). Here our work uses electrocorticographic (ECoG) recordings from electrodes implanted directly on the cortical surface of epilepsy patients undergoing invasive monitoring. ECoG electrodes directly measure the aggregate activity of small neuronal populations with high temporal and spatial resolutions. This makes them useful for measuring neural correlates of specific cognitive states, which might be represented by detailed spatiotemporal patterns of neuronal activity. ECoG is further useful because it simultaneously measures neural activity at multiple frequencies, which is important because neural oscillations at different frequencies are linked to distinct physiological processes (Buzsáki, 2006).

Although research suggests that the brain utilizes distributed patterns, most traditional neuroscience research uses univariate statistical methods, which are incapable of fully quantifying distributed signals. Here we instead use multivariate machine-learning algorithms, which have recently emerged as a powerful technique for identifying and characterizing distributed neural representations. Machine learning methods have most often been used to probe brain data that were recorded non-invasively, such as functional magnetic resonance imaging (Cox and

* Correspondence to: M.A.J. van Gerven, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6525 HR Nijmegen, The Netherlands.
** Correspondence to: J. Jacobs, School of Biomedical Engineering, Sciences, and Health Systems, Drexel University, Philadelphia, PA 19104.
E-mail addresses: m.vangerven@donders.ru.nl (M.A.J. van Gerven), joshua.jacobs@drexel.edu (J. Jacobs).
[1] First and last authors contributed equally.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.12.059

Savoy, 2003; Haxby et al., 2001; Haynes and Rees, 2006; Kamitani and Tong, 2005; Kriegeskorte and Kreiman, 2012; Norman et al., 2006), electroencephalography (Murphy et al., 2011; Simanova et al., 2010) and magnetoencephalography (Chan et al., 2011; Rieger et al., 2008). By utilizing machine-learning techniques in conjunction with ECoG data, our work aims to bring a new level of detail to our understanding of how the brain represents individual cognitive states.

We analyzed ECoG recordings from patients performing a working-memory task where they memorize each of the letters in a short list (Sternberg, 1966). Previously we showed that the amplitude of ECoG activations at 65–128 Hz at individual electrodes distinguished specific memory items (Jacobs and Kahana, 2009). Here, we extend this work to characterize the spectral, spatial, and temporal distribution of each stimulus's ECoG pattern. To foreshadow our main results, we successfully used machine-learning algorithms to decode the identities of individual viewed letters using patterns of neuronal activity distributed across space, time, and frequency. This provides a successful demonstration of using ECoG "mind reading" to decode a person's specific brain state. Further, we scrutinized the machine-learning model computed for each patient to reveal the types of neural signals that distinguish individual cognitive states. Our results show that individual stimuli are represented by ECoG signals at a range of frequencies, with prominent contributions by signals in the theta/alpha (4–16 Hz) and high-gamma (65–128 Hz) bands.

Methods

Patients

We analyzed data from 59 patients undergoing invasive monitoring as treatment for drug-resistant epilepsy. Each patient performed between one and ten testing sessions. We excluded 16 patients where we collected less than 30 trials per stimulus, leaving a total of 43 patients. Our research protocol was approved by the appropriate institutional review boards at several hospitals: Thomas Jefferson University Hospital (Philadelphia, PA), University of Pennsylvania (Philadelphia, PA), University Clinic (Freiburg, Germany), Children's Hospital (Boston, MA), and Brigham and Women's Hospital (Boston, MA). Informed consent was obtained from patients or their guardians.

Data acquisition

In each patient, we collected ECoG recordings from 15 to 160 electrodes. Electrode locations varied across patients due to the variations in each patient's clinical needs, but there were an especially large number of electrodes in temporal cortex. Recording electrodes typically consisted of two configurations: grid and strip electrodes, which are placed on the surface of the neocortex, and depth electrodes, which penetrate through the cortex and record from deep brain structures like the hippocampus. Electrode locations were computed by coregistering a postoperative computed-tomography scan with a higher-resolution magnetic-resonance image and reported in units of Talairach coordinates (Lancaster et al., 2000; Talairach and Tournoux, 1988). ECoG activity was recorded at a sampling rate of 250–1024 Hz using Bio-Logic, XLTek, Neurofile, Nicolet, or Nihon Kohden recording systems, depending on the testing hospital. The recording from each electrode was re-referenced to the average of all signals from electrodes on the same grid, strip, or depth probe. ECoG recordings were synchronized with the patient's task behavior via optically isolated synchronization pulses that were measured on a spare recording channel.

Task

During each ~45-min testing session, patients participated in a working-memory task on a bedside laptop computer (Sternberg, 1966). In each trial of the task, patients were presented with a list of one to six letters. During this presentation portion of the trial, first a fixation cross appeared, and then the letters were displayed sequentially on the computer screen. Each letter was on screen for 700 ms, followed by 275–350 ms (uniformly distributed) of blank screen, for a total of a 975–1050 ms interstimulus interval. Patients were instructed to closely attend to each stimulus presentation and to silently hold the identity of each item in memory. The letter lists included only consonants to prevent patients from using mnemonic strategies, such as treating each list as a single pronounceable word. After the presentation of each list, the response period began when a probe item was displayed after a ~2-s delay. Then patients responded by pressing a key to indicate whether the probe was present in the just-presented list or whether it was absent. After the key press, the computer indicated whether the response was correct, and then a new list was presented. Individual patients participated in different variations of the task, such that they viewed between 8 and 20 consonants. On average, across all sessions and trials, each patient viewed 584 stimulus presentations. In cases when a patient participated in multiple task sessions, we pooled data from multiple sessions together. We were unable to measure patients' eye movements because of the limitations of the hospital testing environment, but we frequently reminded patients to fixate their gaze at the center of the laptop screen.

Data preprocessing

Because our goal was to characterize neural activity related to recognizing and memorizing the currently viewed letter, the data analyses presented here concern the presentation portion of each trial (Jacobs and Kahana, 2009). ECoG recordings were resampled to 500 Hz to provide consistency across different recording systems. We analyzed ECoG activity in the 0–800-ms time period after each letter presentation using the raw ECoG waveforms (time-domain representation) and the Hilbert envelope for different frequency bands and time windows (frequency-domain representation). For the time domain representation, recordings were down-sampled to 200 Hz, high-pass filtered at 5 Hz, and notch filtered using a zero-phase-distortion Butterworth filter at 60 Hz (United States) or 50 Hz (Europe) to remove power-line noise. The data were normalized relative to the 200-ms baseline period before each stimulus appearance and further down-sampled to 100 Hz. For the frequency domain representation, we separately analyzed the amplitude of the signal in the following bands: delta (2–4 Hz), theta (4–8 Hz), alpha (8–16 Hz), beta (16–30 Hz), low gamma (30–60 Hz), and high gamma (60–124 Hz). Amplitude measurements in each frequency band were obtained using bandpass filtering in conjunction with the Hilbert transform (Freeman, 2007). To perform this procedure, first we filtered the raw ECoG signal in each range using a second-order Butterworth bandpass filter. Next, we applied the Hilbert transform, which yields a complex number, and then took the absolute value to extract the instantaneous amplitude. We smoothed the amplitude measurement from each trial with a 100-ms boxcar filter to compensate for trial-to-trial jitter (Jacobs and Kahana, 2009).

Stimulus decoding

We were interested in testing whether we could reliably decode the identity of an individual viewed letter using simultaneous ECoG recordings and, if so, determining the types of brain signals that represent letter-related information. The task of determining which of many letters corresponds to a given neuronal pattern is an example of a multi-class classification problem. We converted this task into a series of two-class classification tasks, as these can be solved straightforwardly with various multivariate algorithms. Our implementation takes as input two possible letter identities and a multichannel ECoG recording from a patient viewing a letter in a trial of the task. The algorithm "decodes" the ECoG signal and outputs a predicted letter, which corresponds to its estimate of the letter that was most likely to be viewed in that recording. That is, given n letters, we solved a binary