Employing a Common Average Reference to Improve Cortical Neuron Recordings from Microelectrode Arrays

Kip A. Ludwig¹, Rachel Miriani¹, Nicholas B. Langhals¹, Michael D. Joseph¹, David J. Anderson²,³ and Daryl R. Kipke¹

¹ Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA
² Professor Emeritus of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA
³ Professor Emeritus of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

E-mail: ludwigk@umich.edu, dkipke@umich.edu
Phone Number: (734) 615-9128
Address:
University of Michigan
1101 Beal Ave
2247 LBME
Ann Arbor, MI 48109

Short title: Common Average Referencing to Improve Neural Recordings
ABSTRACT

In this study, we propose and evaluate a technique known as common average referencing (CAR) to generate a more ideal reference electrode for microelectrode recordings. CAR is a computationally simple technique, and therefore amenable to both on-chip and real-time applications. CAR is commonly employed in electroencephalography (EEG), where it is necessary to identify small signal sources in very noisy recordings. In order to investigate the efficacy of common average referencing, we compared CAR to both referencing with a stainless steel bone-screw, and a single microelectrode site. Data consisted of in vivo chronic recordings in anesthetized Sprague Dawley rats drawn from prior studies, as well as previously unpublished data. By combining the data from multiple studies, we have generated and analyzed one of the more comprehensive chronic neural recording datasets to date. Reference types were compared in terms of noise level, signal-to-noise ratio, and number of neurons recorded across days. Common average referencing was found to drastically outperform standard types of electrical referencing, reducing noise by more than 30 percent. As a result of the reduced noise floor, arrays referenced to a CAR yielded almost 60 percent more discernible neural units than traditional methods of electrical referencing. CAR should impart similar benefits to other microelectrode recording technologies – for example, chemical sensing – where similar differential recording concepts apply. In addition, we provide a mathematical justification for CAR using Gauss-Markov theorem, and therefore help place the application of CAR into a theoretical context.

Keywords: Microelectrode array, chronic neural recordings, ground, common average reference
1. INTRODUCTION

Individual unit recordings from cortical neurons are dependent on separating the recorded extracellular action potential of a neuron from ambient sources of noise. These sources of noise can be correlated or uncorrelated across the electrode array, and include motion artifact, 60 Hz noise, instrumentation noise, thermal noise, and biological sources of noise (Horsch and Dhillon 2004; Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). As cortical recording experiments move away from closed environments designed to reduce noise (e.g. Faraday cages) to more real-world situations (e.g. neuroprosthetic devices), these sources of noise become increasingly problematic (Horsch and Dhillon 2004; Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998).

Recording useful signal is dependent upon minimizing sources of noise through the use of an appropriate reference electrode (Horsch and Dhillon 2004; Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). Typically either an additional microelectrode, or a large electrode such as a stainless steel bone-screw or stripped wire, is placed in a location with minimal cortical activity and used as a reference to subtract out correlated sources of noise (Blanche et al. 2005; Henze et al. 2000; Ludwig et al. 2006; Nelson et al. 2008; Vetter et al. 2004; Webster 1998; Williams et al. 1999). Both microelectrode and large electrode references have their own specific advantages and disadvantages.

A large electrode is often preferred as a reference to minimize impedance (Horsch and Dhillon 2004; Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). Thermal noise is proportional to impedance; therefore, less thermal noise is 'subtracted into' the recordings
when using a low-impedance reference electrode (Horsch and Dhillon 2004; Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). Unfortunately, there is a significant size and impedance mismatch between the reference and the recording sites on the microelectrode array when using a large reference electrode (Horsch and Dhillon 2004; Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). Consequently, the representation of correlated sources of noise (such as motion artifact and 60 Hz noise) is different between the reference and the microelectrode sites, and therefore is not fully removed by reference subtraction (Horsch and Dhillon 2004; Webster 1998).

Due to the sheer size of the large electrode, the reference is typically placed on top of the dural surface, as opposed to implantation in cortical tissue (Ludwig et al. 2006; Vetter et al. 2004). By placing the reference in a location distal from the microelectrode sites, the difference between the voltage representation of correlated sources of noise at the reference and the microelectrode sites is increased (Horsch and Dhillon 2004; Webster 1998). In addition, there remains the possibility of the reference adding an ECoG (electrocorticogram) signal at the dural surface into the recordings.

As opposed to a large reference electrode, the use of an additional microelectrode implanted in cortical tissue as a reference presents alternative problems. Although microelectrode references match recording sites on the array in terms of geometry and impedance, they have greater impedance than their large electrode counterparts, and therefore ‘subtract in’ more thermal noise to the recordings (Horsch and Dhillon 2004; Kovacs 1994; Ludwig et al. 2006; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003). This problem is exacerbated in chronic applications, where the microelectrode reference becomes encapsulated in fibrous tissue, further increasing its impedance (Hochberg et al. 2006; Ludwig et al. 2006; Nicolelis et al. 2003; Otto et al. 2006; Polikov et al. 2005;
When placed on the electrode array itself, a microelectrode reference may actually record neural signal, or uncorrelated biological noise caused by the activity of distal neural sources; this uncorrelated activity is then subtracted into the recordings (Horsch and Dhillon 2004; Webster 1998). If placed in a location to minimize the probability of recording neural activity (e.g. Corpus Callosum), the increased physical separation of the reference electrode from the microelectrode array decreases the correlation between noise at these two locations, and therefore decreases the utility of the reference (Horsch and Dhillon 2004; Webster 1998).

In this study, we introduce a technique known as common average referencing (CAR) to generate a more ideal electrode reference for single unit neural recordings. CAR is commonly employed in electroencephalography (EEG), where it is necessary to identify small signal sources in very noisy recordings (Cooper et al. 2003; Offner 1950; Osselton 1965). Unlike more complex methods of de-noising recorded signals post-hoc (Aminghafari et al. 2006; Bierer and Andersen 1999; Oweiss and Anderson 2001), common average referencing is a computationally simple technique, and therefore amenable to both on-chip and real-time applications. As the name implies, an average of all the recordings on every electrode site is taken and used as a reference (Cooper et al. 2003; Offner 1950; Osselton 1965). Through the averaging process, only signal/noise that is common to all sites (correlated) remains on the CAR\(^1\). Signal that is isolated on one site (single unit activity) does not appear on the CAR, unless the signal is so large as to dominate the average\(^2\) (Cooper et al. 2003; Offner 1950; Osselton 1965). Uncorrelated random noise with a zero mean is minimized through the

---

\(^1\) For example, 60 Hz noise and motion artifact.

\(^2\) Given \(n\) electrodes, this occurs if the signal is \(n\) times larger than the peak-to-peak value of the noise floor.
averaging process\textsuperscript{3}. As the CAR provides an accurate representation of correlated noise at the location of the microelectrode array, but minimizes the contribution of uncorrelated noise sources, we hypothesize that common average referencing will improve neural recording quality with respect to both large and microelectrode references.

In order to investigate the efficacy of common average referencing, we compared CAR to both referencing with a stainless steel bone-screw and a single microelectrode site, using \textit{in vivo} chronic recordings from a prior study (Ludwig et al. 2006) as well as previously unpublished data. By combining the data from multiple studies, we have generated and analyzed one of the largest chronic neural recording datasets to date. Reference types were compared in terms of peak-to-peak noise, signal-to-noise ratio, and number of units recorded across days. Moreover, we provide a mathematical justification for common average referencing based on Gauss-Markov theorem, and therefore build a theoretical context for future CAR applications.

\section{2. METHODS}

\subsection{2.1 Microelectrodes}

Twenty-one male Sprague-Dawley rats were implanted with twenty-six 16-channel chronic silicon ‘Michigan’ microelectrode arrays, using experimental procedures outlined previously (Ludwig et al. 2006; Vetter et al. 2004). Arrays consisted of four shanks, each with four evenly spaced iridium electrodes. Site and shank spacings were sufficient (100 µm or greater) to limit the probability of an individual neuron being recorded from multiple sites (Henze et al. 2000). All of the electrodes on a specific array were the same site size; electrode site sizes on an individual array were either 703 or 1250 µm\textsuperscript{2}.

\textsuperscript{3} For example, thermal noise and distal neural sources.
The data for this study was drawn from previous (Ludwig et al. 2006) and ongoing studies aimed at evaluating the efficacy of the conductive polymer poly(3,4-ethylenedioxythiophene) (PEDOT) for improving neural recording quality. Towards that end, eight of the sites on each array were coated with PEDOT using various deposition methods and counter-ions, the details of which are beyond the scope of this study (Ludwig et al. 2006). The remaining eight sites were left uncoated as controls. Sites were stagger coated to prevent bias due to cortical depth or location (Ludwig et al. 2006).

Overall, PEDOT sites perform similarly to control sites in recording neural activity, with one exception. As noted in other studies, sites coated with PEDOT recorded activity from a slightly larger number of neurons, primarily as a result of reduced thermal noise (Cui and Martin 2003; Ludwig et al. 2006; Yang et al. 2005). These slight differences in recording performance did not affect the results in this paper (See Results and Discussion for details).

2.2 Surgical Techniques

All of the arrays in this study were implanted in motor cortex, targeting cortical layer V. Initial anesthesia was administered via intra-peritoneal injections of a mixture of 50 mg/ml ketamine, 5 mg/ml xylazine, and 1 mg/ml acepromazine at an injection volume of 0.125 ml/100g body weight. Updates of 0.1 ml ketamine (50 mg/ml) were delivered as needed to maintain anesthesia during the surgery. Animals were secured to a standard stereotaxic frame, and three stainless steel bone-screws were inserted into the skull. The electrode connector was grounded to a bone-screw over parietal cortex using a stainless steel wire.

A craniotomy approximately 3 mm by 2 mm was made over the target area (target location 3.0 mm anterior to bregma, 2.5 mm lateral from bregma, and 1.4 mm deep from the surface of the brain).
Two incisions were made in the dura mater to create four flaps, which were subsequently folded back over the edge of the craniotomy. The electrodes were then hand inserted using a microforceps into the approximate target cortical area. Cortical depth was estimated using the known location of the electrode sites on the individual shanks in conjunction with the known length of the individual shanks. Next, the surface of the brain was covered with GelFoam® (Henry Schein, Inc., Miami, FL) for protection. The silicon cable connector was covered with either remaining Gelfoam or Kwik-Sil silicone polymer (World Precision Instruments, Inc). The entire assembly excluding the connector was then enclosed using dental acrylic (Co-Oral-Ite, Dental Mfg. Co., Santa Monica, Ca). Finally, sutures were used to close the skin around the acrylic and triple-antibiotic ointment was applied. All procedures complied with the United States Department of Agriculture guidelines for the care and use of laboratory animals and were approved by the University of Michigan Animal Care and Use Committee.

2.3 Neural Recordings & Data Analysis

For eight of the animals in this study, recorded neural signals were acquired using a Plexon Multi-channel Neural Acquisition Processor (MNAP; Plexon Inc, Dallas, TX). For the remaining animals, signals were acquired using a TDT multi-channel acquisition system (Tucker-Davis Technologies, Gainesville, FL). Neural electrophysiological recordings for all sixteen channels were amplified and bandpass filtered; single and multi-unit recordings were sampled at either 40 kHz (Plexon), or 24414 Hz (TDT), and bandpass filtered from 450-5000 Hz. During recording sessions, animals were placed in an electrically shielded recording booth and multiple 30-second segments of continuous neural recordings were taken.
Neural recording segments were analyzed offline using custom automated MatLab (Mathworks Inc., MA) software, as described in detail elsewhere (Ludwig et al. 2006). In summary, an amplitude threshold window was set 3.5 standard deviations above and below the mean of the sample distribution. For each peak exceeding the threshold window, a 2.4 ms candidate waveform snippet centered on the absolute minimum of the waveform was removed from the recorded segment and stored. The amplitude of the noise voltage for every recording site in each recorded segment was calculated after all candidate waveforms had been removed.

After initial principal component analysis, individual clusters were identified using Fuzzy C-Means clustering (Bezdek 1981; Dunn 1974; Ludwig et al. 2006). When compared to hard clustering, fuzzy clustering reduces classification errors resulting from the synchronous firing of multiple neurons (Zouridakis and Tam 2000). In order to determine the optimum number of clusters, the number of clusters was iteratively increased until the value for the objective function calculated for \( k + 1 \) number of clusters was at least 55 percent of the value for the objective function calculated for \( k \) number of clusters (Karkkainen and Franti August 2002).

After clustering, waveforms with a cluster membership index of greater than 0.8 were used to determine a mean waveform for a cluster. Contributions of white noise and waveforms created by the simultaneous firing of multiple neurons generally do not have a membership index of greater than 0.8 for a particular cluster, and therefore were limited using this procedure (Zouridakis and Tam 2000). An interspike interval histogram for each cluster was generated and visually inspected for an obvious absolute refractory period as an additional measure of noise rejection. Signal amplitude for a cluster was defined as the peak-to-peak amplitude of the mean waveform for each cluster.

The signal-to-noise ratio (SNR) for a given cluster was defined as follows:
SNR = Signal Amplitude / (Peak-to-Peak Amplitude of the Noise Floor)

The peak-to-peak amplitude of the noise on a given site was calculated as six times the standard deviation of the recording after thresholded waveforms were removed, spanning approximately 99.7 percent of normally distributed noise data (Blanche et al. 2005). By using this method, the calculated signal-to-noise ratio and peak-to-peak noise amplitude on a given site was more consistent with a visual inspection of the recorded voltage traces (See Figures 2 and 3). For example, an SNR of 2 would indicate that the mean peak-to-peak amplitude of the signal was twice as large as the peak-to-peak amplitude of the noise floor. As the peak-to-peak amplitude of the noise floor for neural recordings is typically between six and ten times larger than the root mean square (RMS) value of the noise floor (Blanche et al. 2005), SNR calculations for neural recordings based on the RMS of the noise floor raise the SNR with respect to peak-to-peak values.

Clusters with a mean SNR of 1.1 or greater were considered discriminable units, as the signal amplitude of these clusters was sufficient to be reliably differentiated from the noise floor. Conversely, clusters generated by random outlying perturbations from sources of noise had mean SNR values of 0.9 or less. Although normally distributed noise sources will occasionally exceed the 3.5 standard deviation threshold by random chance, the average waveform generated by these noise sources returns to zero after crossing threshold (instead of exhibiting an immediate opposing peak). Consequently, the mean waveform of a noise cluster spans less than six standard deviations of the noise floor, resulting in a calculated SNR of less than 1.

When adjusted for the difference between calculating SNR using peak-to-peak amplitude of the noise floor instead of RMS, an SNR of 1.1 or greater corresponded well with observations of ‘moderate or better’ unit quality based on SNR values from similar recording studies (Henze et al. 2000; Ludwig et al. 2006; Suner et al. 2005).
Isolating action potentials from an individual neuron using an individual recording site is inherently prone to classification errors (Harris et al. 2000; Lewicki 1998). The methodology employed in this study was intended to minimize these errors, and should accurately parallel the true number of underlying neural sources (Ludwig et al. 2006). The sorting routine produces similar results to manual sorting performed by experienced researchers over the same data sets, but with the advantage of being objective and automated (Ludwig et al. 2006).

2.4 Referencing Techniques

As noted previously, all recordings in this study were initially referenced to a stainless steel bone-screw (Screw) located over parietal cortex. Both the microelectrode reference and the common average reference (CAR) were implemented digitally after this initial reference subtraction.

2.4a Common Average Reference (CAR)

The most intuitive implementation of a CAR would be to reference a specific site to the sample by sample average of all of the remaining sites on the array. Unfortunately, this approach presents two problems with respect to real-time cortical recordings. First of all, each of the sixteen sites would have a unique reference, instead of a global reference shared by all sites. This presents an additional layer of complication in translating a CAR from a digital reference to an on-chip analog reference. Second, individual sites on a microelectrode array occasionally fail to function properly. Consequently, these bad sites must be identified and removed from the data set prior to generating a CAR.

To address these problems, the common average reference for each array was generated by taking the sample by sample average of all ‘good’ recording sites, creating one global reference for all sites (CAR-16). For a site to be considered ‘good’, the RMS of the noise floor on the site was required
to be between 0.3 and 2 times the average RMS of the noise floor across all sixteen sites on the array. Sites identified as ‘bad’ through impedance spectroscopy (Ludwig et al. 2006) typically exhibited noise floors with an RMS value three to six times larger than the average RMS of the noise floor on good sites. This simple methodology for eliminating ‘bad’ channels proved sufficient for removing the occasional ‘bad’ site from further analysis in an automated fashion.

As each ‘good’ recording site contributes to the average calculated for the ‘global’ CAR, the amplitude of all samples on a ‘good’ site is slightly decreased when referenced to the CAR. More specifically, given \( n \) good sites, the amplitude of each sample will be \( \left(\frac{n-1}{n}\right) \) times the original value (Cooper et al. 2003; Offner 1950; Osselton 1965). As \( n \) becomes larger, this scale factor approaches a value of 1. Since this scale factor affects both signal and noise equally, it does not affect the calculated signal-to-noise ratio on each site\(^4\). For comparison with other types of referencing, all CAR peak-to-peak noise values denoted in this study have been appropriately adjusted to compensate for this scale factor.

In order to assess the contribution of recording sites electrochemically deposited with a conductive polymer to the overall data, two additional common average references were created and applied to all data sets. One CAR was generated from only the conductive polymer sites on a given array, and the second CAR from only the control iridium sites. The CAR generated from conductive polymer sites was only applied as a reference to the conductive polymer sites; the CAR generated from control sites was only applied as a reference to the control sites. In this paper, data obtained using this secondary method is referred to as CAR-8.

2.4b Single-Best Microelectrode Reference

\(^4\) Although SNR is not affected by this scale factor, CAR improves SNR by removing noise common to all sites without adding uncorrelated noise.
Cortical recordings are often taken in reference to an individual microelectrode site; consequently, we implemented an algorithm to identify the single-best microelectrode reference on an array to compare with common average referencing. As a single microelectrode may pick up single-unit activity, we first used our automated sorting algorithm in conjunction with CAR to identify candidate sites with no discernible unit activity\(^5\). Each of the candidate sites were then employed as a reference for the entire array on the original data set. The candidate microelectrode reference that created the lowest average noise floor across the array on a given day was selected as the single-best microelectrode reference. By identifying the single-best microelectrode reference for each array on a daily basis, we created a conservative comparison for common average referencing.

2.5 Theoretical Justification of CAR Using Gauss-Markov Theorem

Multi-channel neural recordings can be described as an observed noisy signal, \(Y\), which is composed of an unknown true signal, \(X\), and an unknown noise, \(N\). This general situation can be described by the following equation (Albert 1972; Stark and Woods 2002):

\[
Y = AX + N, \tag{1}
\]

where \(Y\) is the \(n \times 1\) vector of observed neural recordings across \(n\) channels, \(X\) is a \(k \times 1\) vector representing the true signal (with \(k\) representing the number of underlying neural sources), \(A\) is an \(n \times k\) matrix (\(n > k\)) which maps \(X\) onto \(Y\), and \(N\) is an \(n \times 1\) normally distributed random noise vector with zero mean and covariance matrix \(K\).

One useful approach for obtaining a ‘good’ estimate, \(\hat{X}\), of \(X\) from the observed values of \(Y\) is to restrict \(\hat{X}\) to be a linear function of \(Y\):

---

\(^5\) Employing a common average reference was necessary to identify sites without unit activity, as the noise floor using the ground screw as a sole reference was sufficient to mask unit activity on many days.
\[ \hat{X} = B Y \]  

(2)

Gauss-Markov theorem states that in a linear model in which the errors have expectation zero and equal variances, a best linear unbiased estimate of the coefficients is given by a least squares estimator (Albert 1972; Stark and Woods 2002). In this formulation, we want to determine the matrix \( B \), a spatial filter that provides the best linear unbiased estimate (BLUE) of \( X \) based on \( Y \).

The filter, \( B_i \), for each channel, \( i \), can be calculated as follows (Albert 1972; Stark and Woods 2002):

\[ B_i = (A_i^T K^{-1} A_i)^{-1} A_i^T K^{-1} \]  

(3)

There are two simplifications that align this general solution with the present study.

1. Site spacing is sufficiently large to ensure that neurons are expressed on only one site. Consequently, \( A_i \) for each channel has the form \([1, 0, 0, ..., 0]\).

2. The noise model for each channel is identical, producing a covariance matrix \( K \) with equal diagonal values \(1\) and equal off-diagonal components \(c\).

Consequently, \( B \) for each channel has the form:

\[ B_i = [1, \alpha, \alpha, ..., \alpha], \]  

(4)

where \( \alpha = -c/(1+(n-2)c) \).

As \( n \) increases, \( \alpha \) approaches \(-1/(n-1)\), leaving \( B_i \) as:

\[ B_i = [1, -1/(n-1), -1/(n-1), ..., -1/(n-1)] \]  

(5)

Placing this result into equation 2, obtaining the best linear unbiased estimate of \( X \) is equivalent to subtracting the average of all other channels from the current channel – in other words, use a common average reference. This formulation is generally true even if the diagonal and off-diagonal
values vary across channels. Note that when $c$ is large (near 1), $c$ approaches $-1/(n-1)$ more quickly. This result is intuitive; when there is high-amplitude correlated noise across channels, referencing is particularly effective. When there is very little or no correlated noise across channels, employing a reference only ‘subtracts in’ uncorrelated noise from the reference to the other channels. Although common average referencing minimizes uncorrelated noise as a function of $n$, it does not remove it entirely. Consequently, referencing becomes counter-productive when there is no correlated noise between the reference and recordings sites.

2.6 Statistical Analysis

Because the four types of referencing in this study were performed digitally on the same data sets, these data sets are highly correlated. To accommodate the correlations between comparison groups, a repeated measures ANOVA was performed; a Bonferroni correction was applied to the tests for each group pair to control for experimentwise error rate. The Bonferroni correction is regarded as a very strict/conservative test of significance when comparing multiple groups. At an alpha value of .01, the Bonferroni correction required a $p$-value of less than .0017 for significance. Equality of variance was verified by the Levene statistic.

3. RESULTS

3.1 Noise

Over the course of this study, the average peak-to-peak noise floor across all sites using CAR was 32.5 µVs (Figure 1), a significant improvement over referencing to either a stainless steel ground screw (45.9 µVs, $p < 10^{-15}$) or to the single-best microelectrode site on the array (40.7 µVs $p < 10^{-15}$). More specifically, every single recording site in the study ($n = 5010$) exhibited a lower noise floor when
referenced to the CAR in comparison to the single-best microelectrode reference. In all but three instances, the noise floor employing CAR was lower than referencing to a ground screw.

Using either CAR or single-best references, the trend in the average noise floor across sites following surgery paralleled the results of other microelectrode recording studies. The noise floor typically decreased over the days immediately following surgery, and then gradually increased over time. An increase in noise over time is expected, as the impedances of electrode sites are known to increase over time due to fibrous encapsulation, increasing the thermal noise on a site (Ludwig et al. 2006; Vetter et al. 2004; Williams et al. 1999).

In contrast to CAR or single-best references, the noise floor using the ground screw as a sole reference was considerably more variable. This variability can be explained by separating the recordings into two distinct cases. In the first case, the recordings made using the ground screw as a reference were contaminated by one or more large sources of correlated noise (Figure 2). Both the CAR and single-best reference minimize the contribution of correlated noise sources, and therefore exhibit lower noise floors and more distinct unit activity than ground screw referencing when large sources of correlated noise are evident. As the ground screw is a poor match for the implanted microelectrode in terms of location, impedance, and geometry, it does not lower the contribution of correlated noise sources as effectively. The noise floor on electrode sites was larger when employing the single-best electrode reference as opposed to the CAR, as the single-best electrode subtracts in the uncorrelated noise from the single electrode site. Conversely, uncorrelated sources of noise with a zero mean tend towards zero when averaged over all the sites on an array using CAR.

In the second case, the recordings were only minimally contaminated by correlated noise (Figure 3). In this case, the ground screw reference actually lowered the noise floor with respect to the single-
best microelectrode reference. Ground screws used in this study had impedances that were multiple orders of magnitude lower than implanted microelectrodes, and consequently subtracted less thermal noise into the recordings\(^6\). As only recordings referenced to the ground screw were significantly altered by the absence or presence of correlated noise, recordings referenced to the ground screw were considerably more variable\(^7\). The CAR, which minimized both correlated and uncorrelated sources of noise, dramatically outperformed the single-best and ground screw references in both cases.

3.2 Signal-to-Noise and Unit Activity

Calculating a meaningful average signal-to-noise ratio (SNR) of unit recordings for comparison between reference types was complicated by the fact that sites registered an overall greater number of discernible units with CAR (See end of Section 3.2). More specifically, the reduction in noise on CAR sites was sufficient to separate previously unapparent signals from the noise floor - instead of just increasing the SNR of already discriminable units. Therefore, comparing the ‘average’ SNR of all discriminable units was not representative of the true difference in SNR between reference types.

In order to calculate the ‘true’ difference in SNR between CAR and other types of referencing, we first used CAR in conjunction with our automated sorting algorithm to identify the location in time of all unit activity. The known location in time of unit activity was then used to calculate the mean

\(^6\) Many students in our lab were surprised to note that digitally referencing an individual electrode site would often increase the noise floor across the array, and initially attributed this result to an equipment malfunction.

\(^7\) Note: Within array variances when employing all types of referencing in this study were equivalent, as verified by the Levene statistic. Ground screw peak-to-peak noise values varied greatly between arrays, depending upon the presence of correlated noise sources. As noted in Methods, both the ‘single-best’ reference and CAR were implemented after initial referencing to a ground screw.
signal amplitude for all reference types. Through this methodology, we were able to identify the location of signal that was previously obscured by the noise floor when not using CAR.

Over the course of this study, sites referenced to CAR exhibited a higher signal-to-noise ratio than sites referenced to either a ground screw or single-best microelectrode reference ($p < 10^{-10}$, See Figure 4). The average SNR on CAR sites was 1.59, in comparison to 1.31 for sites referenced to the single-best microelectrode reference and 1.24 for sites referenced to a ground screw. As CAR does not alter the relative size of signal on a given site, this improvement in performance is attributable to the decreased CAR noise floor.

Referencing with CAR both a) increased the signal-to-noise ratio of units already evident when referencing to the single-best microelectrode reference or the ground screw, and b) decreased the noise floor sufficiently to reveal units that were previously obscured when referencing to the single-best microelectrode or the ground screw. Consequently, there was a greater number of discernible units when referencing to CAR in comparison to a ground screw or single-best microelectrode reference ($p < 10^{-10}$, See Figure 5). Over the course of the study, an average of 46.5 percent of CAR sites exhibited discriminable unit activity, in comparison to 29.6 percent of the single-best sites and 27.46 percent of the ground screw sites. CAR sites recorded an average of 0.842 units per day, whereas single-best sites recorded an average of 0.447 units per day and ground screw sites recorded an average of 0.415 units per day. These daily numbers correspond to 4,045 total units registered on CAR sites during this study, in comparison to 2,152 units registered on single-best sites and 1,998 units registered on ground screw sites. As with the improvement in SNR, this improvement is attributable to the markedly decreased CAR noise floor. Consistent with previous studies (Ludwig et al. 2006; Santhanam et al. 2007; Schwartz 2004; Vetter et al. 2004; Williams et al. 1999), there was a notable decrease in unit activity in the days immediately following surgery, followed by an increase
in unit activity after day fifteen. Although the exact reason for this increase in unit activity at the two
to three week point is unknown, this result has been previously reported in other studies (Ludwig et al.
2006; Santhanam et al. 2007; Schwartz 2004). At two to three weeks post-implantation, the fibrous
encapsulation around an implant becomes more defined, and the immune response begins a
transition into the chronic phase (Turner et al. 1999).

4. DISCUSSION

4.1 Conductive Polymer Common Average Reference

As noted in the methods, two additional common average references were created and applied to
all data sets in order to assess the contribution of recording sites electrochemically deposited with a
conductive polymer to the overall data. One CAR was generated from only the conductive polymer
sites on a given array, and the second CAR from only the control sites (CAR-8). The CAR generated
from conductive polymer sites was only applied as a reference to the conductive polymer sites; the
CAR generated from control sites was only applied as a reference to the control sites.

If the noise between conductive polymer sites was more correlated than the noise between
conductive polymer and control sites, one would expect the noise floor to decrease by generating
separate CARs for conductive polymer and control sites. Instead, creating separate CARs marginally
increased the noise floor (See Figure 1). This increase in noise is a result of using only eight sites to
calculate the CAR, instead of sixteen. As shown in section 2.5, employing CAR more closely
approximates the best linear unbiased estimate of the signal as the number of channels increases.
The average of uncorrelated sources of noise with a zero mean tends towards zero; the contribution
of uncorrelated sources of noise becomes closer to zero when more sites are included to generate
the average. Consequently, decreasing the number of sites used to calculate the CAR increases the contribution of uncorrelated noise sources.

Although the correlation in noise between pairs of sites across a single array was similar on a given day, this correlation value could vary dramatically from day to day. The average cross-channel correlation on an array was 0.66 over the course of this study, but this average value ranged from 0.2 to 0.95 depending on the day. These values are similar to results reported by Rebrik et al, who found that cross-channel correlation coefficients within the same tetrode implanted in cat visual cortex ranged between 0.8-0.92, while the correlation coefficients across channels of different tetrodes ranged between 0.47 and 0.51 (Rebrik et al. 1999).

4.2 Application of CAR to Tetrodes

Under certain circumstances, there is a possibility of registering neural signal on a single site so large that it dominates the average of all sites, and therefore appears on the common average reference (Cooper et al. 2003; Offner 1950; Osselton 1965). Accordingly, this neural signal would appear as a small false unit on all sites referenced to the CAR. For this to occur, given \( n \) recording sites, the average amplitude of the signal on a given site must be \( n \) times larger than the peak-to-peak amplitude of the noise floor (Cooper et al. 2003; Offner 1950; Osselton 1965). In this study, the signal amplitude on one site was never sufficient to dominate the average (See Figure 4); given sixteen recordings sites on every array, the signal on a single site would have to regularly exceed sixteen times the peak-to-peak amplitude of the noise floor (Figures 2 and 3). Based on the largest peak-to-peak amplitudes for action potentials observed in this study in comparison to the peak-to-peak amplitude of the noise floor, a minimum of five electrodes must be used to calculate the CAR to limit the possibility of an action potential dominating the common average reference. As sites were
separated by more than 100 microns in this study, the probability of recording signal from an individual neuron on multiple sites simultaneously was low (Henze et al. 2000).

For some neural recording applications, microelectrodes are manufactured with groups of four recording sites closely spaced to deliberately detect neural activity from an individual neuron simultaneously, known commonly as a tetrode configuration (Gray et al. 1995). As a result, the probability of signal from an individual neuron dominating the CAR average becomes much greater. If \( m \) sites out of \( n \) total electrodes record similar signal from an individual neuron, the average signal across \( m \) electrodes needs only to exceed \( n/m \) times the peak-to-peak amplitude of the noise floor to dominate the CAR average. One solution to this problem would be to increase \( n \), the number of electrode sites used to calculate the CAR. In addition, a site or sites recording aberrantly large signal could simply be removed when calculating the common average reference (Cooper et al. 2003; Offner 1950; Osselton 1965).

When using a CAR, the potential contribution from a neural source on one electrode in a tetrode could diminish the measured potential from the neural source on the three remaining electrodes in the tetrode. This distortion should be constant across all four electrodes, and therefore should not adversely affect standard amplitude-based tetrode clustering techniques (Gray et al. 1995). The distortion could be problematic, however, when employing mathematical techniques to locate the neural source in space.

4.3 Comparing CAR to Alternative Correlation-Based Denoising Methods

A number of post-hoc methods have been previously suggested which identify and utilize the correlated noise across channels to reduce overall noise and improve spike sorting (Bierer 2001; Musial et al. 2002; Rebrik et al. 1999). These post-hoc analysis techniques require initial segments of
data to define the correlation in noise between multiple channels. Consequently, these strategies are computationally intensive in comparison to CAR, and are not easily performed in real-time. Moreover, these methods require additional ‘recalibration’ steps as the correlation in noise across the array changes over time. As a result, these techniques are not easily implemented on-chip. CAR is not a post-hoc analysis technique but instead an alternative electrical reference, which can be implemented in real-time without the need for calibrations to account for transient changes in noise content. In addition, CAR can be implemented with a simple circuit either on-chip or at the recording headstage of the microelectrode array, thereby minimizing the contribution of noise prior to subsequent amplification and digitization.

In order to implement CAR on-chip, it is necessary to ensure that the amplitude of the incoming signal is within the dynamic range of the operational amplifier (op-amp). In some cases, noise transients are large enough to exceed the dynamic range of op-amps due to the large amplification that is traditionally required for front-stage neural acquisition hardware, causing ‘saturation’ or ‘clipping’ of the output signal. To avoid this problem, a large electrode such as a ground screw is normally used as an electrical reference to minimize incoming noise transients. To successfully implement real-time hardware CAR, it would be necessary to either a) increase the dynamic range of the op-amps to exceed maximum values of expected noise transients given the gains necessary for signal reconstruction, or b) first use a ground screw as an electrical reference to prevent amplifier saturation.

4.4 Additional Benefits of Reducing Noise

In some cases in this study, common average referencing provided a significant improvement in recording performance, in other cases CAR salvaged recordings that would have otherwise been
unusable. Decreasing the noise floor not only increases the total number of units recorded across an array, but also decreases the variability of those units. Additional noise can alter the instantaneous shape of an action potential, potentially distorting the recorded waveform in principal component analysis space (PCA), deleteriously affecting common sorting techniques (Lewicki 1998) (Figure 3).

Perhaps even more significant, common average referencing mitigates the contribution of intermittent sources of correlated noise (Figure 2), such as motion artifact. These noise sources distort the relationship between neural firing rates and underlying physiological processes. This is especially problematic in real-time applications such as brain-machine interfaces, where intermittent artifact cannot be removed through post-hoc analysis. Reducing variability in the noise floor can also be critical for comparison studies of electrode performance. Significant differences in electrode performance as a function of electrode design or a drug treatment could potentially be masked by the large variance in noise floor from day to day.

CONCLUSIONS

According to standard differential recording theory, the reference electrode and the recording electrode must be matched as closely as possible in terms of geometry, electrode material, location, and electrical characteristics (Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). These requirements would seem to suggest the use of a single microelectrode as a reference, instead of an unmatched larger electrode. Unfortunately, due to the small size of a microelectrode, coupled with fibrous encapsulation endemic to long-term cortical implants (Edell et al. 1992; Johnson et al. 2005; Ludwig et al. 2006; Polikov et al. 2005; Schmidt et al. 1993; Szarowski et al. 2003; Turner et al. 1999; Vetter et al. 2004; Williams et al. 1999; Xindong et al. 1999), the impedance of a microelectrode reference adds significant uncorrelated thermal noise to cortical recordings.
(Kovacs 1994; Schmidt and Humphrey 1990; Shoham and Nagarajan 2003; Webster 1998). Moreover, the microelectrode reference could potentially register neural signal as well as uncorrelated biological noise – adding both to neural recordings. In contrast, a common average reference minimizes uncorrelated sources of signal and noise through averaging, while eliminating sources of noise common to all sites. Therefore, a common average reference more closely approximates the theoretical differential recording ideal.

Not surprisingly, common average referencing was found to drastically outperform standard types of electrical referencing over the course of this study, reducing noise by more than 30 percent. As a result of the reduced noise floor, arrays referenced to a CAR yielded almost 60 percent more discernible units than traditional methods of electrical referencing. However, these results need to be considered in context of the experimental methodology; recordings in this study were taken from anesthetized animals placed in a Faraday cage designed to reduce ambient noise. Recordings taken from awake and behaving animals in non-shielded environments would have much more noise. Consequently, the drawbacks of either single microelectrode or large electrode references would be further exacerbated. Under these circumstances the common average reference would likely exhibit an even more dramatic increase in recording performance. Common average referencing may impart similar benefits to other microelectrode recording technologies – for example, chemical sensing (Burmeister and Gerhardt 2001; Dressman et al. 2002; Johnson et al. 2007) – where similar differential recording concepts apply.

ACKNOWLEDGEMENTS

The authors of this paper would like to acknowledge all of the members of the Neural Engineering Laboratory at the University of Michigan for their assistance in this study. Specific thanks go to Sarah
Richardson-Burns for coating some of the probes used in this study, and to Erin Purcell, John Seymour, Matt Gibson and Hirak Parikh for invaluable advice and manuscript review.

GRANTS

This work was supported by the NIH P41 Center for Neural Communications Technology (EB002030), Biotectix, LLC and the Whitaker Foundation.

DISCLOSURES

Daryl R. Kipke and David J. Anderson have significant financial interest in NeuroNexus Technologies, a leading supplier of microelectrode technology.
REFERENCES


Xindong L, McCreery DB, Carter RR, Bullara LA, Yuen TGH, and Agnew WF. Stability of the interface between neural tissue and chronically implanted intracortical microelectrodes. Rehabilitation


**Figures**

**Figure 1. Noise Across Days.** Bars denote standard error of the data set on a given block of days. Number of sites on a given block of days, n, is listed on the x-axis. Over the course of the study, sites referenced to a common average reference exhibited 30 percent less noise than standard methods of referencing (p < 10^{-15}). Sites referenced to a ground screw placed over parietal cortex exhibited increased noise floor variability in comparison to referencing with CAR or the single-best microelectrode site. Noise floor amplitude calculated using CAR or single-best microelectrode site references decreased immediately following surgery and then increased afterwards, consistent with the trend in noise level found in prior studies (Ludwig et al. 2006; Schwartz 2004; Vetter et al. 2004; Williams et al. 1999).

**Figure 2. Representative Example of Recordings Contaminated By Correlated Noise.** All columns depict the same two seconds of high speed recordings taken simultaneously across three sites on the same array. Column one data are referenced to a common average reference (CAR). Column two data are referenced to a stainless-steel screw placed above parietal cortex. Column three data are referenced to the single-best microelectrode reference on the array. Sites referenced to CAR exhibit
a lower noise floor and higher signal-to-noise ratio than standard electrical references. When sites are referenced to the ground screw, an intermittent correlated noise source is evident that does not appear when employing either the CAR or single best microelectrode reference. This noise source is sufficient to completely obscure unit activity. Also note that despite the presence of large action potentials on site 2, traces of this signal are not evident on sites 1 and 3 when reference to CAR (See sites 1 and 3 screw and single best reference over the same data set for comparison).  

**Figure 3. Example of Recordings with Low Correlated Noise.** All columns depict the same two seconds of high speed recordings taken simultaneously across three sites on the same array. Column one data are referenced to a common average reference (CAR). Column two data are referenced to a stainless-steel screw placed above parietal cortex. Column three data are referenced to the single-best microelectrode reference on the array. Sites referenced to CAR exhibit a lower noise floor and higher signal-to-noise ratio than standard electrical references. In cases where large sources of correlated noise are not evident, the ground screw reference routinely outperformed the single-best reference in terms of noise level, signal-to-noise, and number of discriminable units. In these cases, common average referencing still outperformed referencing to either a ground screw or the single-best microelectrode site on the array. Arrows denote a neural signal evident on Site 2 when referenced to either CAR or single-best microelectrode reference. Note that the waveform has been distorted on the voltage axis, presumably a result of the increased noise floor when using the single-best microelectrode reference. Even if a neural unit is discernible from the noise, an increased noise floor means more waveform variability, limiting the efficacy of common sorting algorithms (Lewicki 1998).  

**Figure 4. Signal to Noise Across Days.** Bars denote standard error of the data set on a given block of days. Number of units recorded on a given block of days, n, is listed on the x-axis. Over the course
of the study, sites referenced to CAR exhibited a signal-to-noise ratio of 1.59, a significant improvement over referencing to either the single-best microelectrode site (1.31, \( p < 10^{-10} \)) or ground screw (1.24, \( p < 10^{-10} \)). Variability in signal to noise ratio across all types of reference increased towards the end of the study, concurrent with an increase in number of units recorded across all arrays. An increase in number of units recorded starting at three weeks post implantation has been noted in prior studies (Ludwig et al. 2006; Santhanam et al. 2007; Schwartz 2004).

**Figure 5. Percentage of Sites with Units Across Days.** Bars denote standard error of the data set on a given block of days. Number of arrays included in analysis, \( n \), is listed on the x-axis. Over the course of the study, sites referenced to a common average reference yielded almost 60 percent more discernible units than when reference to standard electrical references. This increase in performance is attributable to a reduced noise floor, enhancing signal-to-noise ratio, and therefore increasing the number of discernible units. Unit recordings were strong initially, dipped dramatically in the days following surgeries, and returned to initial levels after the four week point. This trend in recording performance has been noted in previous recording studies (Ludwig et al. 2006; Santhanam et al. 2007; Schwartz 2004).
Footnotes

1. For example, 60 Hz noise and motion artifact.

2. Given n electrodes, this occurs if the signal is n times larger than the peak-to-peak value of the noise floor.

3. For example, thermal noise and distal neural sources.

4. Although SNR is not affected by this scale factor, CAR improves SNR by removing noise common to all sites without adding uncorrelated noise.

5. Employing a common average reference was necessary to identify sites without unit activity, as the noise floor using the ground screw as a sole reference was sufficient to mask unit activity on many days.

6. Many students in our lab were surprised to note that digitally referencing an individual electrode site would often increase the noise floor across the array, and initially attributed this result to an equipment malfunction.

7. **Note:** Within array variances when employing all types of referencing in this study were equivalent, as verified by the Levene statistic. Ground screw peak-to-peak noise values varied greatly between arrays, depending upon the presence of correlated noise sources. As noted in Methods, both the ‘single-best’ reference and CAR were implemented after initial referencing to a ground screw.
CAR Site #1
SNR 2.45
Pk-Pk Noise 25.2

Screw Site #1
SNR 1.07
Pk-Pk Noise 57.7

Single Best Site #1
SNR 1.79
Pk-Pk Noise 34.5

CAR Site #2
SNR 3.65
Pk-Pk Noise 27.7

Screw Site #2
SNR 1.75
Pk-Pk Noise 57.8

Single Best Site #2
SNR 2.83
Pk-Pk Noise 35.7

CAR Site #3
SNR 1.77
Pk-Pk Noise 24.4

Screw Site #3
SNR 0.755
Pk-Pk Noise 57.3

Single Best Site #3
SNR 1.31
Pk-Pk Noise 32.9

Common Average Reference

Voltage (μV)

Seconds
CAR Site #1
SNR 1.65
Pk-Pk Noise 34.6

Screw Site #1
SNR 1.46
Pk-Pk Noise 38.9

Single Best Site #1
SNR 1.32
Pk-Pk Noise 43.2

CAR Site #2
SNR 1.99
Pk-Pk Noise 36.7

Screw Site #2
SNR 1.77
Pk-Pk Noise 41.3

Single Best Site #2
SNR 1.62
Pk-Pk Noise 45.1

CAR Site #3
SNR 3.58
Pk-Pk Noise 34.7

Screw Site #3
SNR 3.14
Pk-Pk Noise 39.5

Single Best Site #3
SNR 2.84
Pk-Pk Noise 43.7

Common Average Reference