On the Variability of Manual Spike Sorting *

Frank Wood† **Michael J. Black**† **Carlos Vargas-Irwin**‡
Matthew Fellows‡ **John P. Donoghue**‡

†Dept. of Computer Science,
‡Dept. of Neuroscience
Brown University, Providence, RI 02912

fwood@cs.brown.edu, black@cs.brown.edu, carlos_vargas_irwin@brown.edu,
matthew_fellows@brown.edu, john_donoghue@brown.edu

December 30, 2003

Abstract

The analysis of action potentials, or "spikes," is central to systems neuroscience research. Spikes are typically identified from raw waveforms manually for off-line analysis or automatically by human-configured algorithms for on-line applications. The variability of manual spike "sorting" is studied and its implications for neural prostheses discussed. Waveforms were recorded using a micro-electrode array and were used to construct a statistically similar synthetic dataset. Results showed wide variability in the number of neurons and spikes detected in real data. Additionally, average error rates of 23% false positive and 30% false negative were found for synthetic data.

Keywords: Spike sorting, neural prosthesis, electrode array, decoding, motor cortex.

---

*Manuscript submitted July 1, 2003 to the IEEE Transactions on Biomedical Engineering, special issue on Brain Machine Interfaces. Support Contributed By NIH-NINDS #NS25074, NIH-NINDS #N01-NS-2-2345, NSF ITR #0113679.

1 Introduction

A common assumption in systems neuroscience is that the brain encodes information in the firing rate of neurons (i.e. the number of action potentials, or "spikes," over a temporal interval). Consequently, finding the spiking activity in eletrophysiological recordings of the brain is seen as a first step in the decoding of neural activity. This often requires making assumptions about the consistency, shape, and individuality of spike waveforms. Analysis of recordings requires first detecting waveforms thought to be action potentials, distinguishing waveforms of true spikes from those that are actually noise and then, in the case where the activity of multiple cells is recorded by a single electrode, classifying which cell, or unit, produced which waveform. The post-detection process is referred to as spike "sorting" and produces a number of "spike trains" corresponding to the temporal sequence of action potentials (see Lewicki [13] for a review). For off-line analysis, these spike trains are typically obtained using manual, or semi-automatic, classification methods while, for on-line decoding, simple thresholding and template matching techniques are employed. In both cases, the quality of the resulting spike trains is dependent on researcher judgment and experience. The variability of human spike sorting performance has been previously noted for recordings from tetrodes and single electrodes [7]. Here we studied this variability for motor-cortical data recorded using chronically-implanted micro-electrode arrays which are increasingly being used for neural prosthetic applications.

Neural prostheses pose special problems with respect to spike sorting. These devices decode the activity of neurons and use this information to generate control signals for the manipulation of the external world. A variety of recording technologies are employed in neural prosthetic applications and, more widely, for the study of neural coding. Tetrodes [19, 20], microwires [27, 28], and micro-arrays [21] have all been exploited to derive control signals. Current implantable prostheses exploit hundreds of electrodes which produce large volumes of data that must be sorted in real-time to achieve continuous neural device control. Regardless of the recording technology, these implanted devices are currently fixed in position and high-quality prosthetic control requires that as much information as possible be recovered from each electrode. Each electrode may have mixed signals coming from multiple neurons, the recorded waveforms may vary markedly in their signal to noise ratio, and this signal to noise ratio may vary over time. These facts make the task of manual or automatic sorting challenging.

In all cases, spike sorting involves converting the raw electrophysiological data into a representation of the neural spiking process. This involves five inter-related tasks:

1. Waveforms of potential spikes must be detected and recorded.
2. The waveforms from each electrode (channel) must be sorted into a set thought to be actual "spikes" and a set thought to be "noise."
3. The number of generating neurons (units) must be determined for every channel, since a given channel might contain the activity of zero or more cells.
4. Each of the spikes on those channels must be attributed to the neuron that generated it.
5. A timestamp must be assigned to mark the occurrence of each spike.

Although detection is itself a potentially large source of error, we did not address it in this study. While various detection methods exist [11], this study used simple thresholding.

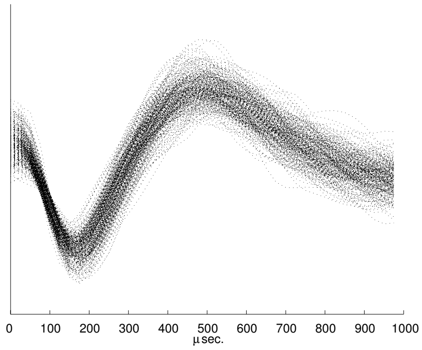
There are also a variety of automated spike sorting methods [3, 7, 8, 12, 13, 16, 20, 24, 26] , yet most off-line research in systems neuroscience has used spike trains that were manually (or semi-automatically) sorted with the aid of various commercial products. On-line applications require automatic sorting but often this involves a manual stage of analysis to establish thresholds or model waveform shapes.

Spike trains often form the basis for both the analysis of neural encoding and the development of decoding algorithms. A variety of decoding methods have been proposed for neural prosthetic applications [21, 22, 27, 28] and many of these methods exploit a rate code in which the discrete spike train data is converted to a rate function by binning or averaging over some temporal window. Misclassification of spikes at the sorting stage can corrupt the resulting rate code in a variety of ways with unknown consequences for decoding performance. Similarly, decoding methods based on point processes [1, 10] exploit the spike trains directly along with a temporal model of the spiking processes. The fundamental assumptions of these methods regarding the statistics of the spiking process may also be violated by sorting mistakes.

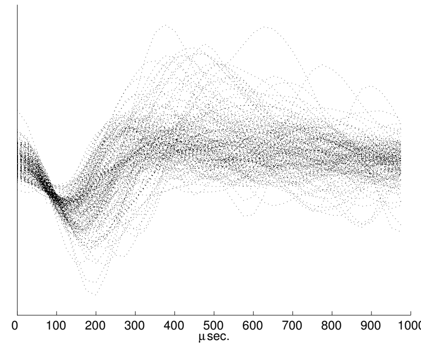
Beyond the problem of decoding for prostheses, the analysis of spike trains and precise spike timing in neural coding often relies on hand-sorted data. An understanding of the variability among sorters and the absolute error rates in the resulting spike trains is critical for understanding and evaluating models of neural coding. Towards that end we studied the performance of expert human spike sorters on natural and synthetic datasets.

Significant variability among human sorters has previously been shown for recordings from tetrodes and single electrodes [7], where error rates were computed by comparing sorted extra-cellular recordings to intra-cellular "ground truth." For neural prosthetic applications, simultaneous intra-cellular recordings are typically not practical. In particular, chronically implanted micro-electrode arrays (such as the Utah intracortical array [14] used in the recordings for the experiment) effectively prohibit both individual electrode placement and simultaneous intra-cellular recording. In addition, in single-unit recording, electrodes are moved to achieve well isolated signals and this is not possible with current array technology. The use of such array's for prosthetic applications is increasing as the technology matures and, consequently, understanding the nature of the signals from such devices is important for the development of automated spike sorting algorithms. In lieu of intra-cellular recordings, we constructed a set of synthetic channels for which we knew the ground truth. These synthetic channels were constructed from a statistical model of the true data making them similar enough to real channels that it was difficult for human subjects to distinguish them from real data.

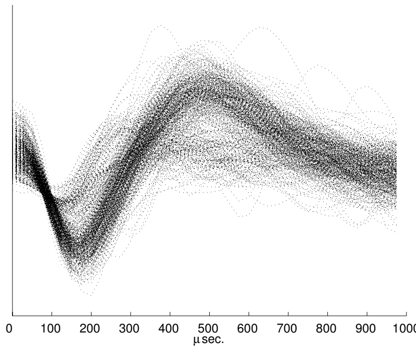
Both the real and synthetic channels were manually sorted by five subjects with commonly used commercial software [9]. The subjects used various techniques such as principal component analysis (PCA) and manual cluster cutting [13]. We found large discrepancies in both the number of units identified and the spikes assigned to each by different subjects. The magnitude of these discrepancies was statistically similar for the real and synthetic data. We calculated a false positive (FP) rate of 23% and a false negative (FN) rate of 30% for the synthetic data and postulate that these rates are similar for real data. The results suggest that neural prosthetic control algorithms



(a) Synthetic spikes



(b) Synthetic noise



(c) Synthetic channel

Figure 1: Generation of synthetic channels: 1(a): Synthetic waveforms sampled from a trained Gaussian model, 1(b): noise sampled from other channels, and 1(c): noise and synthetic spikes combined

could benefit from the development of new statistical techniques for automated spike sorting that account for the inherent ambiguity of the spiking process in the measured waveforms.

2 Methods

To quantitatively assess the subjective variability of sorted spike trains, we asked five expert subjects to sort a set of waveforms recorded from the arm area of primary motor cortex in two different monkeys. Using this data we computed the subjective variability of the spike trains produced by different people. With real data of this type however, there is no principled way to establish "ground truth" and consequently no way to quantitatively measure the error in human sorting performance. To address this, we generated a set of synthetic channels and asked the same subjects to sort them. The synthetic channels were designed to be indistinguishable from the real and allowed us to establish quantitative error rates with realistic waveforms. Details of the methods are described below.

2.1 Recording

In two monkeys, following task training, Bionic Technologies LLC (BTL) 100-electrode silicon arrays [14] were implanted in the arm area of primary motor cortex (MI).

The BTL arrays consisted of 100 platinized tip silicon probes (200-500 kOhms at 1 kHz; [17]), arranged in a square grid (4mm x 4mm, electrode separation 400microns on-center). The electrodes were 1 mm in length, corresponding in MI to recordings near the layer III/V boundary. All procedures were in accordance with Brown University Institutional Animal Care and Use Committee-approved protocols and the Guide for the Care and Use of Laboratory Animals (NIH publication no. 85-23, revised 1985). Signals were amplified and sampled at 40 kHz/channel using a commercial recording system [9]. All events that crossed a manually set threshold were digitized (12-bit voltage resolution) and stored on disk. Waveforms and their corresponding timestamps (relative to the start of the recording session) were saved for each electrode on the array.

The recording setup was similar to that used in [21] for the on-line neural control of 2D cursor motion. In this neural prosthetic task, the animals were trained to move a two-joint manipulandum on a 2D plane to control the motion of a feedback cursor on a computer screen. The simultaneous recording of hand kinematics and neural activity allows the study of motor cortical encoding of hand motion [18] and the training of decoding methods [4, 21, 29]

2.2 Real Data Selection

The arrays in the two animals produced 192 channels of data. To simplify the sorting task we selected a 20 channel subset from the full dataset.

To select those channels, we first asked a single expert sorter to sort all 192 channels. This expert is one of the authors of this paper. From this large set of channels we manually selected a 20 channel subset representative of the whole.

2.3 Synthetic Data Generation

We also constructed 5 synthetic channels such as the one shown in Fig. 1(c) for which we knew the ground truth. Unlike the natural data which could contain multiple units on a single channel, all of the synthetic channels were generated having a single unit for simplicity. The synthetic channels contained both the activity of this single unit and of a realistic noise process.

The synthetic single unit activity was drawn from a Gaussian generative model. To construct such generative models for the synthetic channels we randomly chose 5 out of the 20 natural channels. For each channel we selected one of the sorted units identified by the expert (Fig. 1(a)) and collected the N waveforms corresponding to the spikes of that unit into a matrix $W_c = [\vec{\omega}^1, \ldots, \vec{\omega}^N]$ where c indicates the channel and the waveform is represented by $\vec{\omega}^i = [\omega_1^i, \ldots, \omega_n^i]^T \in \mathbb{R}^n$, where n is the number of time samples in the data corresponding to each waveform. In our case $n = 40$, but in other cases this might vary according to the recording equipment and setup.

For each channel we computed the mean waveform

$$\vec{\mu}_c = \frac{1}{N}\sum_{i=1}^{N} \vec{\omega}^i$$

and the covariance

$$Q_c = (W_c - \vec{\mu}_c)(W_c - \vec{\mu}_c)^T.$$

This mean and covariance define a multi-dimensional Gaussian model of the unit's waveforms.

To generate synthetic waveforms we repeatedly sampled from this distribution in the following manner. Let

$$LL^T = Q$$

be the Cholesky factorization of the covariance matrix where L is a lower triangular matrix [25], and let $\vec{\tau} = [\tau_0, \ldots \tau_n]^T$ where $\tau_i, 1 \leq i \leq n$ are zero mean and identically distributed normal random variables (i.e. $\tau_i \sim N(0,1)$). Then notice that the expected value of $L\vec{\tau}$ is

$$E[L\vec{\tau}] = LE[\vec{\tau}] = 0$$

and the variance of $L\vec{\tau}$ is

$$Var[L\vec{\tau}] = E[L\vec{\tau}\vec{\tau}^T L^T] = LE[\vec{\tau}\vec{\tau}^T]L^T = LL^T = Q.$$

So multiplying the Cholesky factorization of the covariance matrix by a set of such vectors $\{\vec{\tau}_j\}$

$$L_c\,\vec{\tau}_j\; + \vec{\mu}_c \sim N(\vec{\mu}_c, Q_c)$$

produces a set of synthetic waveforms with the same distribution as the training set (Fig. 1(a)). This can be used to generate an arbitrary number of waveforms by generating new random vectors $\vec{\tau}$. To ensure that there were no visible high frequency artifacts in the synthetic waveforms, each one was low-pass filtered using an empirically determined Gaussian kernel.

It is common to use a plot of the inter-spike intervals (ISI) when performing spike sorting; violations of the absolute refractory period indicate misclassification. Consequently, we assigned timestamps to the synthetic waveforms by sequentially drawing inter-spike intervals from an exact, empirically calculated, distribution rather than fitting and sampling from a canonical distribution such as the Poisson or exponential. Figure 2 shows the ISI histogram for a synthetic channel and the ISI histogram for the real channel from which it was generated.

Each of the 5 synthetic channels also included noise. We generated noise by drawing waveforms randomly from every channel except the one used to train the generative model (Fig 1(b)). This process excluded the waveforms that had been classified as spikes in the initial sorting of the 20 channels.

Finally, for realism, the synthetic channels had to exhibit the same recording artifacts the real channels exhibited. Waveforms were captured when they passed through a voltage threshold, and, in the real channels all the captured waveforms were aligned on the threshold crossing. To replicate this easily observed artifact we aligned both the synthetically generated waveforms and the sampled noise waveforms using an appropriate threshold that produced channels that visually matched the real channels.

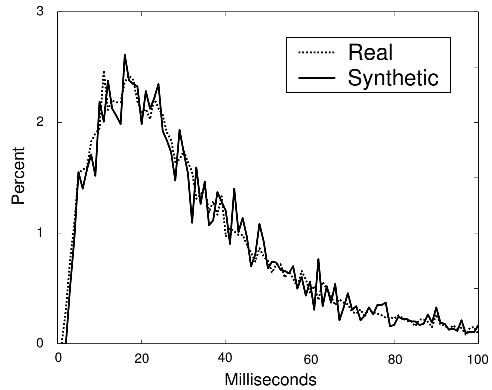


Figure 2: Interspike interval histograms, one for a synthetic channel (solid) and the other for the real channel (dashed) from which it was trained.

2.4 Human Sorting Procedure

The five subjects were graduate students, research assistants, or postdoctoral researchers from the same laboratory; they had significant experience in sorting neural recordings. Some of the authors of this paper were also subjects in this study. The subjects sorted the 25 channels using Plexon's offline spike sorter [9], labeling units and waveforms as they would for their own research. This software provides users with various tools to sort all the waveforms of a particular recording one channel at a time. This is most often achieved by manual cluster selection and refinement in a graphical display constructed by projecting the waveforms onto their first two principal components (see [13] for a review of related techniques). The subjects were given as much time as they liked to sort the data.

3 Results

Realism of Synthetic Data. Prior to sorting, subjects attempted to identify the synthetic channels. The 20 real and 5 synthetic channels were permuted randomly before presentation and the subjects were explicitly told that there were 5 synthetic channels to be found. The subjects were allowed to use any software tool at their disposal and were given unlimited time to come to their conclusions. The expected number of synthetic channels that would be correctly identified by picking 5 channels uniformly at random follows the hypergeometric distribution and is 1 ± 0.82 (mean $\pm$ std.). Our subjects correctly identified 1.3 ± 1.53 which is better than chance but well below correctly identifying all five. In fact, none of the subjects correctly identified all the synthetic channels.

While the Gaussian generative model $\vec{\omega} \sim N(\vec{\mu}, Q)$ for a neuron spike shape distribution is a simplification, on average the expert sorters did not differentiate synthetic channels from real channels at rates much better than chance. This gives some confidence that quantitative error rates for the synthetic data may be indicative of the error rates for real data.

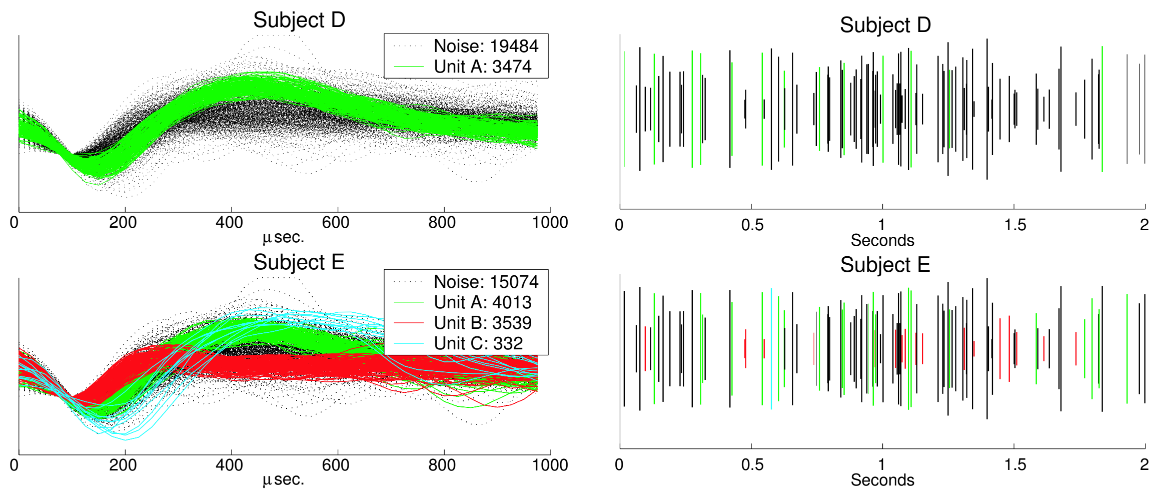


Figure 3: Classification of the same waveforms by two different subjects on a real dataset. **Left:** Waveforms. Subject D (top) classified the data as containing a single unit while subject E (bottom) found three units. While only a small faction of the actual waveforms are shown here to simplify the figure, the subjects had access to the full set of waveforms. **Right:** Two-second segment from the channels (left) shown as spike trains. Vertical bars indicate detected waveforms and the color corresponds to the classification by the subjects (left). Bar height indicates the maximal amplitude of the corresponding waveform.

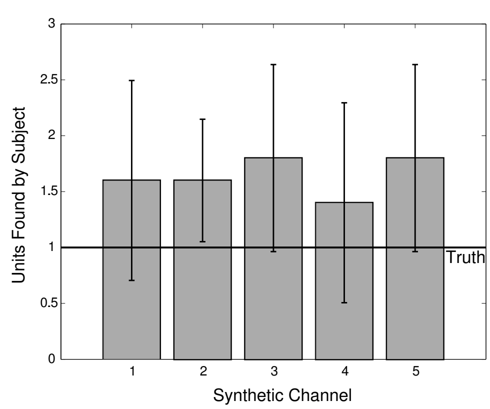


Figure 4: Mean ± std. dev. of the number of units identified by all subjects per synthetic channel. Synthetic channels had only one synthetic unit which we indicated by the black horizontal line.

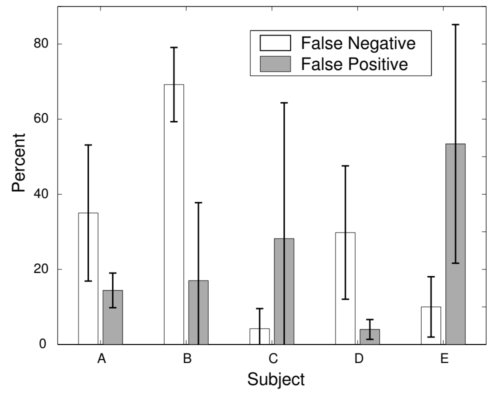


Figure 5: Mean $\pm$ std. dev. false positive (FP) and false negative (FN) error rates over all synthetic channels per subject. FP's occurred when a subject inappropriately counted noise waveforms as having come from the single generating unit FN's occurred when a subject miss-classified a true spike as noise.

Performance on Real Data. Figure 3 illustrates the subjective variability we observed with real data. The figure shows the classification results for two subjects (D and E), where for simplicity only a subset of the waveforms from one of the real channels in the dataset is shown. The green waveforms indicate some agreement between subjects on the presence of a waveform with a particular shape. The subjects disagree however on the number of units present with subject E hypothesizing two additional units in the data. These results are typical of what we observed throughout the study.

Figure 3 provides another view of the same dataset. Here a two-second segment of the data is viewed as a spike train. Each vertical line corresponds to a spike while the color corresponds to the classification in Figure 3. Black lines correspond to waveforms that were treated as noise. Note the level of disagreement in the classified spike trains. Even for the green spikes which have similar waveforms, the two subjects included very different numbers of spikes. We posit that differing views such as this might lead to quite different models of encoding in terms of interspike intervals or synchronous firing.

Table 1 shows the number of units and number of spikes detected by each of the subjects. The subjects agreed on the number of units in the real channels only 25% of the time, and most of these consensus channels either contained no neural activity or were extremely well isolated. The number of units detected varied by roughly a factor of two (from a low of 18 to a high of 35) while the total number of spikes varied even more, with subject E finding approximately four times as many as subject B. Even when the subjects agreed on the number of units in a given channel, quite often they disagreed about how many spikes each generated.

Performance on Synthetic Data. For each synthetic channel there was only one true unit present. Despite this all subjects over segmented these channels as shown in Figure 4.

Figure 5 shows how the subjects performed on the task of segmenting the spikes from the noise. On average the subjects had overall 23% FP and 30% FN error rates for the synthetic channels. The

Subject	A	B	C	D	E
Spikes	99160	50796	150917	77194	202351
Units	28	32	27	18	35

Table 1: Totals for each subject, all real channels combined. Spike counts include all identified spikes; unit counts include all neurons each subject found.

data illustrated in Figure 5 also suggests that the subjects employed individual sorting strategies; this phenomenon was also observed by Harris *et al.* [7]. Despite large variability, it seems that subjects A,B, and D used a sorting strategy that consistently worked to minimize false positives while subjects C and E chose one that worked to minimize false negatives. It also suggests that it might not be possible to overcome the trade-off between FP's and FN's using the sorting tools employed. This might be due to inherent similarities between spike shapes or between spike and noise waveforms. It is also possible that the tools our subjects employed restricted them from being able definitively segment and classify the activities of individual neurons; for example, the software only allows users to view 2D projections onto the principal components (cf. [7]).

Anecdotally, one of the individuals who served as a subject in this study was the same person who served as the expert who sorted the training channels. This person sorted the same channels twice, once to provide the training data and then once a month later as a subject in this study. This individual determined that the real channels contained 12% more neurons the second time around (25 $\rightarrow$ 28) yet classified 8% fewer of the waveforms as neural activity (108073 $\rightarrow$ 99160 spikes). Although this demonstrates the kind of subjective variability we found, it also affects the analysis of our results for synthetic data. While the reported subject-to-subject variability for synthetic data would remain unaffected, the FN and FP rates we reported for synthetic data could vary depending on the way the training channels were initially sorted by our expert.

4 Discussion

Micro-electrode arrays are an important recording technology for neural prosthetic applications. Additionally, recordings from these and other related recording devices are used to model and understand neural coding. Often these analyses rely upon manually (or semi-automatically) sorted spike trains and only rarely is the uncertainty of the underlying data reported. We observed that expert human spike sorters had widely varying performance on both real and synthetic neural datasets. On real data, subjects differed not only in what constituted a spike versus noise but even in the number of units present in the data. To quantify this variability we developed a realistic synthetic dataset where the "ground truth" was known. On average, subjects identified noise as signal 25% of the time while they treated the signal as noise 30% of the time. Moreover, the data suggests that researcher intent plays a large role in the interpretation of recorded data.

For on-line prosthetic applications, careful, manual, spike sorting is not possible. Current techniques for on-line detection are fairly crude and also involve human judgment. For example, experimenter-determined thresholds are used to select waveforms with particular properties. This approach was used in the work of Serruya *et al.* [21] for the real-time control of cursor motion. All detected activity on a channel was binned every $70ms$ and a linear filter was used to model the relationship between this activity and hand position. Since experts sorters often detect more than one unit per channel, treating this binned data as single unit firing "rates" could introduce decoding errors as each channel most likely contains multiple units.

It is interesting to note that Serruya *et al.* achieved good neural control (from 42 channels) without precise on-line spike sorting. This suggests that coarse electrical activity (that may combine units) may be sufficient for neural control applications. Recent work in decoding from local field potentials also suggests that this might be the case [2, 15, 19, 23]. In general, however, such a situation violates the underlying assumptions of most decoding methods. For example, in population vector methods [5, 6, 27], the combination of units with different directional tuning properties will result in a "fictitious" cell whose tuning properties may be very different from the true cells. Similar issues may exist for other decoding methods.

A number of issues emerge for research on neural coding. Theories of neural coding that are based on hand-sorted spikes must be evaluated with respect to the variability of the spike data. In particular, encoding models that rely on precise spike timing, or synchrony, may be affected by the bias of the sorter. For example, a conservative sorter (such as subject D), with a low false positive rate, could miss-classify enough true spikes to remove any evidence of excess synchrony. Alternatively, a sorter with a high-false positive rate could add enough extraneous events to obscure properties of the true firing. The actual effect of this on encoding models deserves further study.

There are two possible solutions. The first would be to build models and test theories using datasets sorted by multiple people. Variability in the performance of the encoding or decoding method could then be reported. Given the time consuming nature of spike sorting this approach may be infeasible. An alternative is to employ an automated spike sorting algorithm. In this case, encoding/decoding performance could be evaluated with respect to a particular sorter with known properties. If the sorting process were consistent, then observed variability across training sets or methods could be more easily evaluated.

A number of issues remain open. This was a fairly small study with all subjects coming from the same laboratory. Given that many groups use the same sorting software, we posit that similar variability would be seen across laboratories, but this remains to be tested.

This work suggests the importance of good, widely available, automated spike sorting methods. The development of such tools requires datasets for evaluation and comparison. The variability of human sorters on the data presented here suggests that it may be difficult to establish the accuracy of automated techniques and in future work we plan to compare human performance with a variety of automated methods (cf. [7] for such a comparison in the case of tetrodes and intra-cellularly recorded ground truth). For this analysis we can exploit one of the main technical contributions of this paper which is the generative waveform model. The synthetic waveforms were shown to be similar to real data as judged by human experts and this suggests that such synthetic data sets may be of value for evaluating automated sorting algorithms.

5 Conclusions

We showed the variability of human spike sorters on waveforms recorded with a micro-electrode array. We also developed a probabilistic model of waveforms that was used to synthesize realistic datasets. Human performance on both real and synthetic data varied widely and suggests that the intent, or "style," of the experimenter influences the resulting spike trains. The results suggest the need for both new ways of evaluating theories of encoding and also algorithms for decoding that take into account spike train variability. The results also point to the need for automated spike-sorting algorithms that provide consistency across experiments.

Acknowledgments

We thank Elie Bienenstock for discussions regarding the statistical evaluation of spike sorting performance. We also thank the reviewers for their valuable feedback.

References

[1] E. Brown, L. Frank, D. Tang, M. Quirk, and M. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neuroscience*, 18(18):7411–7425, 1998.

[2] C. A. Buneo, M. R. Jarvis, A. P. Batista, and R. A. Andersen. Properties of spike train spectra in two parietal reach areas. *Exp. Brain Res.*, 2:134–139, 2003.

[3] M. S. Fee, P. P. Mitra, and D. Kleinfeld. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-gaussian variability. *J. Neuroscience Methods*, 69:175–188, 1996.

[4] Y. Gao, M. Black, E. Bienenstock, S. Shoham, and J. Donoghue. Probabilistic inference of arm motion from neural activity in motor cortex. *Advances in Neural Information Processing Systems 14*, 2002.

[5] A. Georgopoulos, A. Schwartz, and R. Kettner. Neuronal population coding of movement direction. *Science*, 233:1416–1419, 1986.

[6] A. Georgopoulos, A. Schwartz, and R. Kettner. Directional tuning profiles of motor cortical cells. *Science*, 36:73–79, 2000.

[7] K. D. Harris, D. A. Henze, J. Csicsvari, H. Hirase, and G. Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *Journal of Neurophysiology*, 81(1):401–414, 2000.

[8] E. Hulata, R. Segev, and E. Ben-Jacob. A method for spike sorting and detection based on wavelet packets and Shannon's mutual information. *Journal of Neuroscience Methods*, 117:1–12, 2002.

[9] Plexon Inc. http://www.plexoninc.com/ofs.htm. Dallas, TX, 2003.

[10] R. E. Kass and V. Ventura. A spike train probability model. *Neural Computation*, 13:1713–1720, 2001.

[11] K. H. Kim and S. J. Kim. Neural spike sorting under nearly 0-db signal-tonoise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Trans. on Biomedial Engineering*, 47:1406–1411, 2000.

[12] M. S. Lewicki. Bayesian modeling and classification of neural signals. *Neural Computation*, 6:1005–1030, 1994.

[13] M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network*, 9(4):R53–78, 1998.

[14] E. Maynard, Nordhausen C., and R. Normann. The utah intracortical electrode array: A recording structure for potential brain-computer interfaces. *Electroencephalography and Clinical Neuophysiology*, 102:228–239, 1997.

[15] C. Mehring, J. Rickert, E. Vaadia, S. C. De Oliveira, A. Aertsen, and S. Rotter. Inference of hand movements from local field potentials in monkey motor cortex. *Nature Neuroscience*, 153(2):1253–1254, 2004.

[16] M. A. L. Nicolelis, A. A. Ghazanfar, B. M. Faggin, S. Votaw, and L. Oliveira. Reconstructing the engram: Simultaneous, multisite, many single neuron recordings. *Neuron*, 18:529–537, 1997.

[17] C. T. Nordhausen, P. J. Rousche, and R. A. Normann. Optimizing recording capabilities of the Utah intracortical electrode array. *Brain Research*, 637:27–36, 1994.

[18] L. Paninski, M. Fellows, N. Hatsopoulos, and J. Donoghue. Temporal tuning properties for hand position and velocity in motor cortical neurons. *submitted, J. Neurophysiology*, 2003.

[19] B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, and R. A. Andersen. Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nature Neuroscience*, 5(8):805–811, July 2002.

[20] M. Sahani, J. S. Pezaris, and R. A. Andersen. On the separation of signals from neighboring cells in tetrode recordings. In *Advances in Neural Information Processing Systems 10*. MIT Press, 1998.

[21] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue. Brain-machine interface: Instant neural control of a movement signal. *Nature*, 416:141–142, 2002.

[22] K. V. Shenoy, D. Meeker, S. Cao, S. A. Kureshi, B. Pesaran, C. A. Buneo, A. P. Batista, P. P. Mitra, J. W. Burdick, and R. A. Andersen. Neural prosthetic control signals from plan activity. *NeuroReport*, 14(4):591–597, March 2003.

[23] K. V. Shenoy, D. Meeker, S. Cao, S. A. Kureshi, B. Pesaran, C. A. Buneo, A. P. Batista, P. P. Mitra, J. W. Burdick, and R. A. Andersen. Neural prosthetic control signals from plan activity. *Neuroreport.*, 14(4):591–596, 2003.

[24] S. Shoham. Robust clustering by deterministic agglomeration em of mixtures of multivariate t-distributions. *Pattern Recognition*, 35(5):1127–1142, 2002.

[25] G. Strang. *Introduction ot Applied Mathematics*. Wellesley-Cambridge Press, 1986.

[26] S. Takahashi, Y. Anzai, and Y. Sakurai. Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. *Journal of Neurophysiology*, 89:2245–2258, 2003.

[27] D. Taylor, S. Tillery, and A. Schwartz. Direct cortical control of 3d neuroprosthetic devices. *Science*, 296(5574):1829–1832, 2002.

[28] J. Wessberg, C. Stambaugh, J. Kralik, P. Beck, M. Laubach, J. Chapin, J. Kim, S. Biggs, M. Srinivasan, and M. Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408:361–365, 2000.

[29] W. Wu, M. J. Black, Y. Gao, E. Bienenstock, M. Serruya, A. Shaikhouni, and J. P. Donoghue. Neural decoding of cursor motion using a kalman filter. In *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.