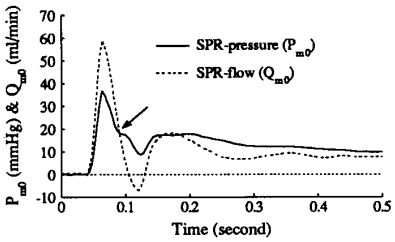

Fig. 7. P_{m0} and Q_{m0} of a rabbit when the backward wave returns in late systole (indicated by arrow). P_{m0} and Q_{m0} have similar waveshapes before the appearance of the backward wave.

summation of odd order reflected waves before they are completely damped, while the forward wave is the summation of the incident wave and even order reflected waves. It is also worth mentioning that most of the net flow (Q_{m0}) occurred in systole and early diastole, while during late diastole the forward wave almost equaled the backward wave, and Q_{m0} decreased to near zero, although there was still a non-zero P_{m0} (Fig. 3). Physiologically, when the heart contracts periodically, the slow decay of P_{m0} maintains the blood pressure in the arterial system even though the blood flow can be very low during diastole.

The assumption of linearity in this study was confirmed by the observation that in some rabbits, when the reflected wave returned to the measured site very late, the difference between the systolic waveforms of P_{m0} and Q_{m0} was small (Fig. 7). The assumption of time-invariance was supported by the fact that the pressure and flow waveforms of the beat preceding the asystole did not change if the R–R interval was constant (Fig. 1). So that, according to our definitions, it is appropriate to calculate Z_0 and RI from the initial P_{m0}, Q_{m0}, P_{f0} and P_{b0}.

In this study, we have proposed a new method to analyze the mechanical characteristics of the carotid artery. This analytic method may be applied to other arterial subsystems, or to the entire arterial system. The method yields more accurate derivations of forward and backward waves, and will be useful in further understanding the mechanical properties of the arterial system and ventricular/arterial interaction.

References

[1] K. P. Brin and F. C. P. Yin, "Effect of nitroprusside on wave reflections in patients with heart failure," *Ann. Biomed. Eng.*, vol. 12, pp. 135–150, 1984.

[2] J.-P. L. Dujardin and D. N. Stone, "Characteristic impedance of the proximal aorta determined in the time and frequency domain: A comparison," *Med. & Biol. Eng.*, vol. 19, pp. 565–568, 1981.

[3] W. K. Laskey and W. G. Kussmaul, "Arterial wave reflection in heart failure," *Circulation*, vol. 75, pp. 711–722, 1987.

[4] J. K.-L. Li, "Time domain resolution of forward and reflected waves in the aorta," *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 783–785, 1986.

[5] W. R. Milnor, *Hemodynamics*, 2nd ed. Baltimore, MD: William & Wilkins, 1989, pp. 170–171.

[6] D. L. Newman, S. E. Greenwald, and N. L. R. Bowden, "An in vivo study of the total occlusion method for the analysis of forward and backward pressure waves," *Cardiovasc. Res.*, vol. 13, pp. 595–600, 1979.

[7] J. P. Murgo, N. Westerhof, J. P. Giolma, and S. A. Altobelli, "Manipulation of ascending aortic pressure and flow wave reflections with the Valsalva maneuver: Relationship to input impedance," *Circulation*, vol. 63, pp. 122–132, 1981.

[8] M. F. O'Rourke and W. W. Nichols, *Blood Flow in Arteries*, 3rd ed. Philadelphia, PA: Lea Febiger, 1990, pp. 285.

[9] N. Westerhof, P. Sipkema, G. C. van den Bos, and G. Elzinga, "Forward and backward waves in the arterial system," *Cardiovasc. Res.*, vol. 6, pp. 648–656, 1972.

[10] E. Wetterer, "Die Wirkung der Herztatigkeit auf die Dynamik des Arteriensystems," *Verhandl. der Deut. Ges. Kreislaufforsch.*, vol. 22, pp. 26–60, 1956.

Optimal Detection, Classification, and Superposition Resolution in Neural Waveform Recordings

Isaac N. Bankman, Kenneth O. Johnson, and Wolfger Schneider

***Abstract*— The effects of noise autocorrelation on neural waveform recognition (detection, classification, and superposition resolution) are investigated in this study using microelectrode recordings from the cortex of a monkey. Optimal waveform recognition is accomplished by passing the data through a whitening filter before matched filtering for detection or template matching for classification and superposition resolution. Template matching without whitening requires about 40% higher signal-to-noise ratio than template matching with whitening for comparable classification and superposition resolution. The comparable difference for detection is 15%.**

I. Introduction

Reliable recordings of the concurrent activity of multiple neurons have become essential for the study of neural systems. Such recordings are obtained with multiple extracellular electrodes, each providing waveforms from one or more neurons. The study presented here is part of an effort to develop a completely automated, optimal system for the recognition of action potential waveforms recorded with high impedance microelectrodes in the mammalian central nervous system. The system is based on template matching using the Euclidean distance and an automatically determined acceptance threshold. The aim of this study is to determine the effect of noise autocorrelation on template matching and to present an optimal approach for recognition of action potential waveforms in colored noise. The contributions of this approach are reported for detecting and classifying individual waveforms as well as for resolving superpositions of two waveforms.

The most widely used technique for detecting action potentials is amplitude threshold crossing. Methods such as principal components [2], and Haar transform [15] have also been used. For optimal signal detection, waveforms should be subject to Bayesian discrimination that separates them from background noise.

Classification methods, ranging from amplitude discrimination to principal components and minimum mean-square-error, have been suggested in the last three decades [1]–[20]. In a comprehensive

Manuscript received December 30, 1992; revised March 29, 1993. This work was supported by the NIH Grant NS 07226, and in part by the Department of the Navy, under Contract No. N00039-91-C-0001, and in part by the WM-Keck Foundation.

I. N. Bankman and W. Schneider are with the Applied Physics Laboratory, The Johns Hopkins University, Laurel, MD 20723.

K. O. Johnson is with the Department of Neuroscience, The Johns Hopkins University, Baltimore, MD 21205.

IEEE Log Number 9209953.

0018–9294/93$03.00 © 1993 IEEE

comparison of separation methods, Wheeler and Heetderks in 1982 evaluated the performance of nine different methods, including spike amplitude, conduction latency, principal components, and template matching using Euclidean distance [7]; the last two methods were found to be the best for spike sorting in noisy data. Because template matching paired with data whitening yields optimal performance and because current technology allows its online implementation, we focused on neural spike sorting by template matching. Optimal and suboptimal (without whitening) template matching methods are compared using action potential waveforms and neural noise recorded from the cortex of an alert monkey.

The simultaneous occurrence of action potentials of two or more neurons causes the waveforms of the individual neurons to be superimposed in extracellular recordings. The probability of three or more action potentials occurring simultaneously is negligibly low in most recordings. However, two action potentials may be superimposed with considerable frequency depending on the number of neurons in the recording, their firing rates, the duration of the action potentials, and the timing relations between the action potentials [1], [11]. Failure to recognize the overlapping spikes can cause underestimation of the firing rates. Solutions for the superposition resolution problem have been suggested using template matching with the Euclidean distance [1] and the city block distance [19] as well as maximum likelihood methods [18]. In this study, we evaluate the gain in superposition resolution performance associated with whitening, which together with template matching is theoretically optimal.

II. Methods

A. Data

Recordings were obtained from the somatosensory cortex of an alert monkey with a filter pass-band of 10 Hz to 10 KHz, a 12-bit A/D converter and a sampling rate of 32 kHz. The standard deviation of amplitude quantization noise was less than 1 μV. The electrodes were quartz glass fibers with tungsten-platinum alloy cores and 3–4 MΩ impedance at 1 kHz. In a recording with relatively low noise levels, five different types of action potential waveforms were isolated by visual inspection, and waveforms of the same type were averaged to form five templates, each represented by 32 samples (1 ms). These templates, shown in Fig. 1(a), constitute the spike test set used in this study. Recorded segments that contained no visually detectable action potentials were used as neural noise segments in this study. The neural noise was stationary and had a normally distributed amplitude with zero mean and a standard deviation of 39.7 μV. Throughout the paper, these data (waveforms and noise) are referred to as raw data to distinguish them from the whitened data obtained by passing the waveforms and the noise through a whitening filter. The signal-to-noise ratio (SNR) was based on raw, not whitened, data and was defined as the rms value of a waveform measured in a window of 1 ms divided by the standard deviation of noise (39.7 μV). The range of SNR's in this study was obtained by scaling the relative amplitudes of templates with respect to noise and by adding scaled templates and noise.

B. Whitening

When the noise involved in the classification is stationary, normally distributed, and colored, optimal Bayesian classification is achieved by using the Mahalanobis distance [21], which is based on the inverse of the noise covariance matrix. The Mahalanobis distance is not suitable for online applications due to the large number of computations that it requires. When the noise is white, i.e., without temporal correlation, the noise covariance matrix is diagonal, and the Mahalanobis distance is reduced to the Euclidean distance. Therefore, when noise is colored, the simpler Euclidean distance metric can provide optimal classification if the data are first filtered with a whitening filter [22].

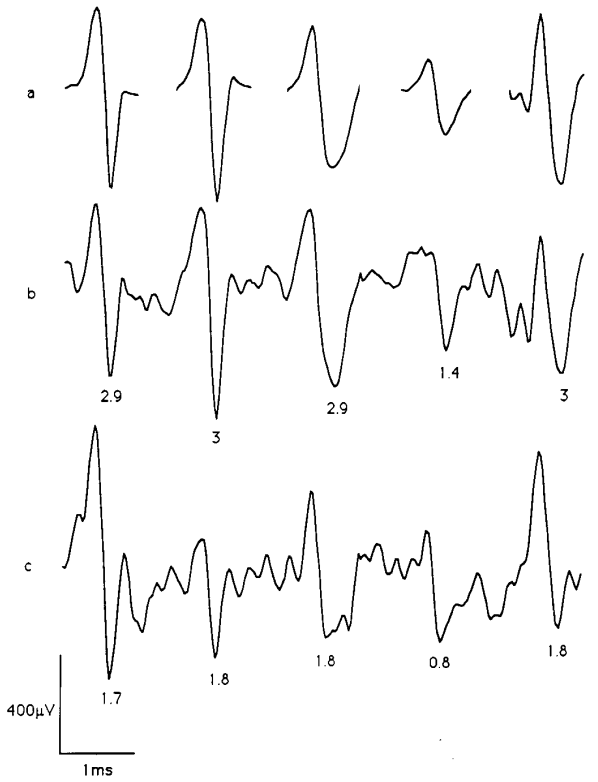

Fig. 1. Spike templates and their noisy versions illustrate some of the noise conditions under which simulations were conducted. (a) Neural waveform templates. Each waveform is obtained by averaging several spike waveforms of the same type taken from the recording. These five waveforms constitute the spike test set. (b) Templates added to neural noise (a scaled part of segment in Fig. 2(a) in which the MSEP is 10. Each template is located at the same position as in (a). The SNR value for each noisy template is shown below the corresponding waveform. (c) Templates added to neural noise with higher variance in which the MSEP is 6.

We computed the autocorrelation function of neural noise using a 16 ms noise segment from the recording mentioned above and observed a significant amount of correlation (see Section III). Similar autocorrelation was observed in several other such segments. The whitening filter was based on an autoregressive moving average (ARMA) model of the neural noise:

$$N(k) = \sum_{i=1}^{u} a_i N(k-i) + \sum_{j=0}^{v} b_j H(k-j)$$

where $N(k)$ is the recorded neural noise, $H(k)$ is white noise, a_i and b_j are corresponding coefficients, and u and v are the model orders of the autoregressive and moving average components, respectively. The coefficients a_i and b_j were computed to minimize the mean-square prediction error of the ARMA model [23]. The inverse of this model provided a whitening filter for the neural noise in this study.

C. Detection

Three techniques for detecting the occurrence of waveforms were studied.

1. Amplitude detection where the decision variable was the sampled voltage.
2. Power detection where the decision variable was the sum of squared samples over a window of 32 samples.
3. Matched filter detection where the decision variable for each waveform type was the inner product of the corresponding template with 32 samples of data.

In each technique, the detection threshold was set at the least value that produced no false positive errors.

D. Classification

A waveform was considered to match a given template when its distance from that template was less than a predetermined acceptance threshold. When a waveform matched more than one template, it was assigned to the closest template.

When the noise is white, the signal-plus-noise distribution for each waveform is hyperspherical, and optimal Bayesian classification can be achieved using Euclidean distance between an unknown waveform and known templates. In some pattern recognition applications, the city block distance is used for its fast implementation. In this study, the classification performance of the squared Euclidean distance (SED) and city block distance (CBD) were compared using data with and without whitening.

The SED is given by:

$$d_s = \sum_{k=1}^{n} [S(k) - T(k)]^2$$

where $S(k)$ is the data vector, $T(k)$ is a template, and n is the number of sample points representing the waveform (32 in this study). For both raw data and whitened data, the acceptance thresholds were set at the lowest possible level that produced no exclusion errors. The whitened noise power had a chi-square distribution, and the acceptance threshold for whitened data corresponded to the 99.8% level of the cumulative chi-square distribution.

The CBD is given by:

$$d_c = \sum_{k=1}^{n} d_k \quad \text{with } d_k = |S(k) - T(k)|.$$

An acceptance threshold for the CBD has been suggested in [6], based on the assumption that d_c has a normal distribution because it is the sum of a large number of d_k's. However, the central limit theorem requires the components of the summation to be independent. We computed the distribution of the random variable:

$$d_{cn} = \sum_{k=1}^{n} |N(k)|$$

for the colored and whitened noise, using 32 for n. In the colored noise, the distribution of d_{cn} was skewed to the left and was significantly different from a normal distribution. In the whitened noise, the distribution of d_{cn} was close to a normal distribution, especially around the mean. However, the upper tail of this distribution was more extended than the normal, and setting the threshold with the assumption of normality resulted in about 5% exclusion errors due to threshold, at all SNR levels. Therefore, we set the acceptance threshold of the CBD at the lowest possible level that resulted in no exclusion errors, in the colored and whitened noise tests.

E. Superposition Resolution

When a waveform is detected but fails to match one of the known templates, the possibility that the waveform is the superposition of two known types embedded in noise has to be tested. The resolution algorithm that we used compares the putative superposition waveform to all possible combinations of each pair of templates with all possible delays in the range of -1 to $+1$ ms between the two templates. The window of this comparison is twice as wide as the individual waveform window; in this study, superpositions are resolved using 64 samples for the waveform and the tested combinations. For each tested combination, a set of SED values is obtained by aligning the start of the combination with several samples surrounding the onset of the unknown waveform. When the minimal SED obtained among all SED values from all combinations is below the acceptance threshold, then the pair of templates corresponding to that combination is considered to indicate the constituent waveform types.

As in classification, the acceptance thresholds for both raw and whitened data were set at the smallest values that caused no exclusion errors.

III. Results

A. Whitening

The neural noise and its autocorrelation function are shown in Fig. 2(a) and (b). The dotted lines in Fig. 2(b) represent the three standard deviation boundaries for ideal white noise. A considerable autocorrelation is evident between lags of 2 and 10 samples. This significant amount of color in neural noise suggests that the use of Euclidean distance with raw data would be suboptimal and that whitening is a prerequisite for optimal Euclidean distance classification.

We tested various combinations for the ARMA model orders u and v in the range of 2 to 10 and evaluated the resulting inverse filter as a whitening filter for neural noise. We obtained the best results with $u = 4$ and $v = 3$. This whitening filter is described by

$$H(k) = (1/b_0) \cdot \left[N(k) - \sum_{i=1}^{u} a_i N(k-i) - \sum_{j=1}^{v} b_j H(k-j) \right]$$

where

$$\begin{aligned} a_1 &= 0.946 \\ a_2 &= 0.106 \\ a_3 &= -0.387 \\ a_4 &= 0.167 \end{aligned} \quad \text{and} \quad \begin{aligned} b_0 &= 1 \\ b_1 &= 1.633 \\ b_2 &= 1.100 \\ b_3 &= 0.335. \end{aligned}$$

The recursive whitening filter designed in this manner is stable with poles inside the unit circle in the complex plane. The neural noise segment of Fig. 2(a) was filtered with the whitening filter (initial conditions set to 0) and produced the whitened noise shown in Fig. 2(c), whose autocorrelation function (Fig. 2(d)) remains in the range expected from white noise.

B. Detection

We compared the effects of whitening on the detection algorithms where signal amplitude, signal power, and output of a matched filter were used. The raw test data comprised the five physiological neural spike waveforms illustrated in Fig. 1 (spike test set) and various levels of additive physiological (colored) noise. Instantaneous power was computed over a 1 ms period. Fig. 3 shows the detection results obtained with the three techniques as a function of SNR.

When matched filtering, power, and amplitude detection techniques were applied to raw (colored) test data, 95% correct detection was obtained above SNR levels of 1.5, 2, and 3, respectively. Whitening improved the performance of matched filter detection by approximately 15%. Neither amplitude nor power detection were improved by whitening. For equal performance, matched filtering, power, and

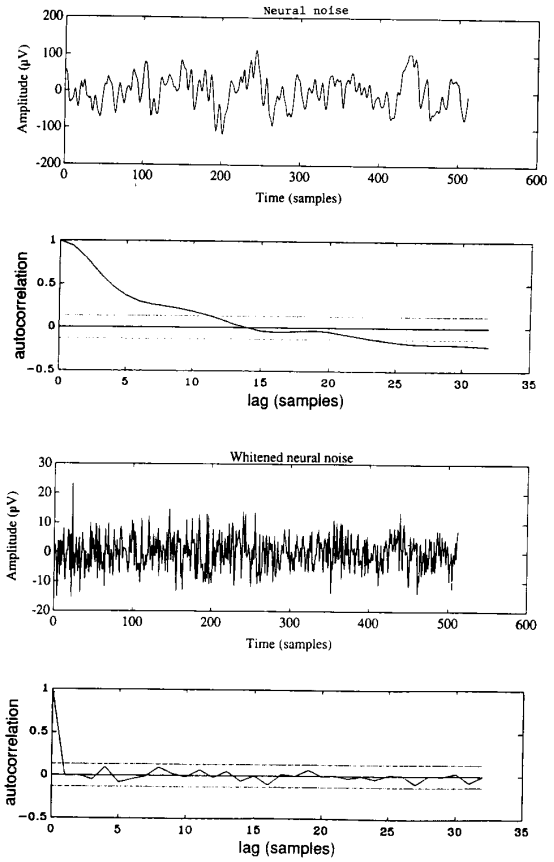

Fig. 2. (a) A 16 ms segment of neural noise digitized at 32 KHz. (b) The autocorrelation function of noise in (a). The dotted lines represent the three standard deviation bounds expected from white noise. (c) White noise obtained by filtering the noise in (a) with the whitening filter. (d) The autocorrelation function of white noise in (c).

amplitude detection based on raw data required approximately 15%, 50%, and 120% more SNR than did matched filtering detection based on whitened data. These ratios of SNR's required for equal performance were similar at all performance levels.

C. Classification

Classification performance depends on the similarity between different spike tests as well as the level of noise in the record. Separability, the Euclidean distance between two spikes measured in units of standard deviation of noise, has been suggested as a parameter that reflects the ease of discriminating two spike waveforms [7]. The minimum separability (MSEP) in a group of spike waveforms is defined as the minimal Euclidean distance between waveforms measured in units of the standard deviation of noise [7] and serves as a measure of classification difficulty. The correspondence between MSEP and SNR at two levels of noise is shown in Fig. 1(b) and (c). In this study, the MSEP value was measured on the raw data.

When classification tests with SED and CBD were applied to the raw data, 95% correct classification performance was obtained at MSEP values of 7.3 and 8.0, respectively (Fig. 4). Whitening improved the performance of both algorithms, yielding 95% correct classification performance at MSEP values of 5.2 for SED and 6.0 for CBD. Approximately 40% more MSEP (or SNR) was required

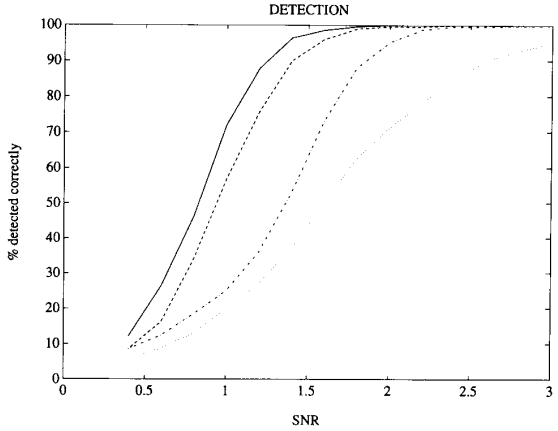

Fig. 3. Detection test results as a function of SNR defined as the ratio of the signal rms value (computed with 32 samples) to the standard deviation of noise (computed on the complete noise record of Fig. 2(a)). Dotted curve, amplitude detection on raw data; dash-dotted curve, power detection on raw data; dashed curve, matched filtering detection on raw data; solid curve, matched filtering detection on whitened data. The performance of each algorithm was assessed with test data containing 100 occurrences of each type, i.e. 500 total spike occurrences at each SNR level in the range of 0.4 to 3 with steps of 0.2.

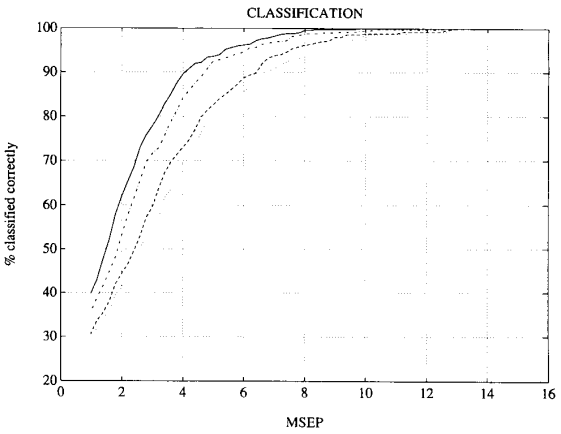

Fig. 4. Classification test results as a function of MSEP, with and without whitening. Dotted and dashed curves show CBD and SED classification respectively, on raw data; dash-dotted and solid curves show CBD and SED classification, respectively, on whitened data. Each curve was obtained by classifying 100 waveforms of each type, i.e., 500 waveforms at each MSEP level in the range of 1 to 16 with steps of 0.2. The MSEP was computed on the raw data.

for SED classification on raw data to obtain the same performance as the optimal approach provided by SED on whitened data. CBD required 10–15% more SNR than SED for equal performance on both raw and whitened data. These percentage differences at 95% correct performance were similar for other performance levels, e.g. perfect classification performance with SED was obtained with MSEP levels of 9 and 13 on whitened and raw data. Classification based on amplitude discrimination of the raw data (not shown in Fig. 4) was considerably worse, requiring SNR values 400% larger than the optimal method for equal performance.

D. Superposition Resolution

Superposition waveforms were simulated by overlapping two templates and embedding the resulting summation waveform into noise

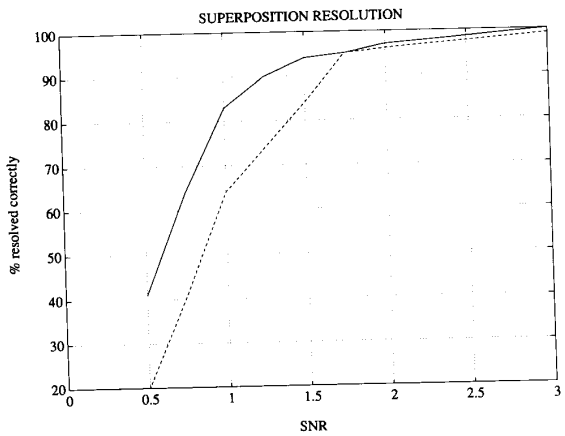

Fig. 5. Superposition resolution performance as a function of SNR of the smallest spike in the test set. The performance in the raw data is shown with the dashed line curve, and that of the whitened data is represented by the solid line curve. Each curve was obtained by resolving 100 superposition waveforms at each SNR level in the range of 0.5 to 3 with steps of 0.25. Each superposition was built with randomly selected pairs of templates with random delay between 0 and 28 samples.

from the recording. Each superposition waveform was built by selecting the two constituent templates randomly from the spike test set and placing them with a random delay between 0 and 28 samples. Resolution tests were run on raw and whitened data using varying levels of additive noise.

A decomposition was considered successful when the algorithm identified the two constituent spike types correctly. The resolution algorithm required an SNR of about 1.6 for 90% correct performance on the raw data. When whitened data were used, the same performance was obtained at an SNR level of 1.25. Below an SNR of 1.25, whitening increased the percent correct resolution by about 20%.

IV. Discussion

Waveform recognition algorithms based on whitened neural recordings performed significantly better than the same algorithms operating on the raw data. Without whitening, classification and superposition resolution algorithms required 40% greater SNR values to provide a performance equal to that obtained on whitened data. The comparable difference in SNR for detection was 15%.

The utility of superposition resolution was also apparent. Superpositions were resolved at the 95% correct level when the SNR was 1.75. Essentially perfect resolution was achieved at SNR of 3. The main source of error in superposition resolution is the similarity between different spike types. Furthermore, even when spike types are not similar, the superposition of two types may be very similar to the superposition of two other types. At relatively low SNR levels, this can cause an additional performance degradation, whose severity depends on the superposition rate and the morphology of individual waveforms.

In an automated system, error sources such as erroneous template estimates, time quantization errors, baseline drifts, changes in amount or nature of noise, gradual changes in spike waveforms, and a sudden decrease in spike amplitude, e.g. when a neuron fires immediately after its previous depolarization, can degrade the classification performance. Some of these error sources can be eliminated; for example, baseline drifts can be compensated by increasing the lower cutoff frequency of the analog band-pass filter, and waveform trends can be tracked with adaptive template updating. Some sources of error, such as time quantization and sudden amplitude drop, can be accounted for by an appropriate modification in acceptance thresholds. Changes in amplitude and nature of noise can affect the threshold estimates and the whitening process; these errors can be prevented by regularly monitoring the autocorrelation of noise and updating the related parameters. Erroneous template estimates can occur when the record has an excessively high level of noise or spikes are similar in shape (very low separability). In such cases, optimal classification cannot be expected due to the difficulty in generating accurate templates.

The whitening filter enhances the high frequency components of noise and reduces the MSEP by about 10%. The seemingly paradoxical improvement in classification under reduced MSEP can be explained by the transformation that the data undergo in multidimensional space. Clusters of each spike type have a hyperellipsoidal distribution due to the autocorrelation of neural noise. The eccentricity of these multidimensional distributions is manifested by the eigenvalue spread of the noise covariance matrix. In our raw neural noise data, these eigenvalues ranged from 2.4 μV^2 to about 16000 μV^2 yielding a 80 : 1 ratio between the lengths of the longest and shortest axes. In such cases, even when the MSEP is moderately low (e.g. 5 to 8) the use of SED without whitening can cause numerous errors. The whitening filter transforms the hyperellipsoidal distribution of each type into a hyperspherical distribution that can be optimally contained within the hyperspherical decision shell provided by the SED threshold.

V. Acknowledgment

The authors would like to thank M. A. Steinmetz, who provided the neural records, R. S. Fisher and A. Cole for assistance in digitization, B. C. Wheeler for useful technical suggestions in the review of a proposal to NIH, and R. J. Johns for his encouragement and advice.

References

[1] V. J. Prochazka and H. H. Kornhuber, "On-line multiunit sorting with resolution of superposition potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 34, pp. 91–93, 1973.

[2] M. Abeles and M. H. Goldstein, "Multispike train analysis," in *Proc. IEEE*, vol. 65, pp. 762–773, 1977.

[3] W. J. Heetderks, "Criteria for evaluating multiunit spike separation techniques," *Biol. Cybernetics*, vol. 29, pp. 215–220, 1978.

[4] E. H. D'Hollander and G. A. Orban, "Spike recognition and on-line classification by unsupervised learning system," *IEEE Trans. Biomed. Eng.*, vol. BME–26, pp. 279–284, 1979.

[5] G. K. Kojima and F. Bracchi, "A microprocessor-based instrument for neural pulse wave analysis," *IEEE Trans. Biomed. Eng.*, vol. BME–27, pp. 515–519, 1980.

[6] G. J. Dinning and A. C. Sanderson, "Real-time classification of multiunit neural signals using reduced feature sets," *IEEE Trans. Biomed. Eng.*, vol. BME–28, pp. 804–812, 1981.

[7] B. C. Wheeler and W. J. Hetderks, "A comparison of techniques for classification of multiple neural signals," *IEEE Trans. Biomed. Eng.*, vol. BME–29, pp. 752–759, 1982.

[8] R. S. LeFever and C. J. DeLuca, "A procedure for decomposing the myoelectric signal into its constituent action potentials—Part I: Technique, theory and implementation," *IEEE Trans. Biomed. Eng.*, vol. BME–29, pp. 149–157, 1982.

[9] K. C. McGill and L. J. Dorfman, "High resolution alignment of sampled waveforms," *IEEE Trans. Biomed. Eng.*, vol. BME–31, pp. 462–468, 1984.

[10] R. M Studer, R. J. P. DeFigueiredo, and G. S. Moschytz, "An algorithm for sequential signal estimation and system identification for EMG signals," *IEEE Trans. Biomed. Eng.*, vol. BME–31, pp. 285–295, 1984.

[11] E. M. Schmidt, "Computer separation of multi-unit neuroelectric data: A review," *J. Neurosci. Meth.*, vol. 12, pp. 95–111, 1984.

[12] J. F. Vibert, J. N. Albert and J. Costa, "Intelligent software for spike separation on multiunit recordings," *Med. Biol. Eng. Comput.*, vol. 25, pp. 366–372, 1987.

[13] M. Salganicoff, M. Sarna, L. Sax, and G. L. Gerstein, "Unsupervised waveform classification for multi-neuron recordings: A real-time

software-based system. I. Algorithms and implementation," *J. Neurosci. Meth.*, vol. 25, pp. 181–187, 1988.
[14] M. F. Sarna, P. Gochin, J. Kaltenbach, M. Salganicoff, and G. L. Gerstein, "Unsupervised waveform classification for multi-neuron recordings: A real-time software-based system. II. Performance comparison to other sorters," *J. Neurosci. Meth.*, vol. 25, pp. 189–196, 1988.
[15] X. Yang and S. A. Shamma, "A totally automated system for the detection and classification of neural spikes," *IEEE Trans. Biomed. Eng.*, vol. 35, pp. 806–816, 1988.
[16] S. R. Smith and B. C. Wheeler, "A real-time multiprocessor system for acquisition of multichannel neural data," *IEEE Trans. Biomed. Eng.*, vol. 35, pp. 875–877, 1988.
[17] D. Stashuk and H. DeBruin, "Automatic decomposition of selective needle-detected myoelectric signals," *IEEE Trans. Biomed. Eng.*, vol. 35, pp. 1–10, 1988.
[18] A. F. Atiya and J. M. Bower, "Optimal neural spike classification," in *Neural Information Processing Systems*, American Institute of Physics, New York, pp. 95–102, 1988.
[19] Y. Wong, J. Banik, and J. M. Bower, "Neural networks for template matching: Application to real-time classification of the action potentials of real neurons" in *Neural Information Processing Systems*, American Institute of Physics, New York, pp. 103–113, 1988.
[20] C. Forster and H. O. Handwerker, "Automatic classification and analysis of microneurographic spike data using a PC/AT," *J. Neurosci. Meth.*, vol. 31, pp. 109–118, 1990.
[21] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[22] G. E. P. Box and G. M. Jenkins, "Identification of transfer function models," *Time Series Analysis: Forecasting and Control*. New York: Holden Day, 1976.
[23] L. Ljung, "Computing the estimate," *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice–Hall, 1987.

An Image Processing Scheme to Quantitatively Extract and Validate Hyoid Bone Motion Based on Real-Time Ultrasound Recordings of Swallowing

Marc A. Cordaro and Barbara C. Sonies

***Abstract*— An image processing technique with associated hardware was developed to quantitatively extract hyoid bone motion from real-time submandibular ultrasound images recorded during the swallowing act. Videofluorographic imaging, the "gold standard" of swallowing studies, was recorded simultaneously and synchronized to the ultrasound. Hyoid position obtained from the ultrasound was validated based on the videofluorography using personal computer-based image processing methods.**

I. Introduction

In the clinical evaluation of swallowing physiology, several instrumental diagnostic techniques are commonly used, including videofluorography (VF), ultrasound (US), manometry, and manofluorography. Of these, the most commonly used technique is videofluorography, which is often used exclusively and repeatedly with minimal consideration of the risks of ionizing radiation. Videofluorography provides a dynamic, complete, pictorial, multiple plane visualization of the oropharynx during swallowing. Submandibular ultrasound provides a partial visualization of the oropharynx, including most of the soft tissues such as the tongue and muscles of the mouth floor, without known harmful effects, but fails to directly visualize the bony structures of the mandible and hyoid bone. Ultrasound does allow dynamic visualization of movement of the tongue, transport of a bolus, and, indirectly, movement of the hyoid bone, and thus captures the oropharyngeal swallow [1].

One parameter of significant clinical interest is the temporospatial pattern of the excursion of the hyoid bone over the course of the oropharyngeal swallow. This pattern is closely related to the ability of the patient to protect the airway during bolus transport into the esophagus, and is of primary clinical significance in the evaluation of dysphagia [2]. Quantitative analysis of hyoid bone dynamics has been performed on fluorographic image sequences [3], but image processing has not been applied to the extraction of hyoid position from ultrasound images of the oropharynx. Such extraction would facilitate the transition of submandibular ultrasound from a subjective visual inspection methodology to a quantitative clinical research tool which may be safely used with children and normal subjects. Such use will allow the development of a quantitative database of swallowing parameters not previously possible with fluoroscopy due to the radiation risk.

This paper describes the application of image processing instrumentation and methods to extract the position of the hyoid bone from ultrasound image sequences. The technique is validated with a frame-by-frame comparison of hyoid bone position extracted from simultaneously recorded ultrasound and videofluorography over the course of a swallow. The complex motions and morphology of the swallowing mechanism make the application of a phantom study for calibration and validation prohibitive; however, a qualitative anatomic correlation has been performed using ultrasound images of a tongue excised from a human cadaver [4]. Therefore, a comparison of ultrasound to the "gold standard" (videofluorography) was chosen for this pilot study.

II. Methodology

The requirements of the system include the ability to record US and VF simultaneously and in synchrony onto video tape, digitize both modalities frame by frame, spatially remap frames to eliminate motion artifact, objectively extract hyoid bone position, and transform the VF coordinate system to perform quantitative comparisons of hyoid trajectory between modalities.

A. Simultaneous Real-Time Ultrasound/Videofluorographic Recording

The real-time video signals from an Advanced Technology Laboratories (Bothell, WA) Ultramark 9 ultrasound scanner with a 5 MHz curved array (IVT) transducer and a standard videofluorography unit were synchronized in time using two For-A video timers (0.01 s resolution), which were slaved together with a common reset switch. The two nonsynchronous video signals were recorded onto separate Sony U-Matic VO-5800 video cassette recorders (see Fig. 1.). This configuration allows both imaging modalities to be represented as full resolution composite video frames with the timer values superimposed. The timer values were used to temporally align sequences of video frames from each medium during digitization and analysis of individual swallows.

Manuscript received October 5, 1992; revised January 4, 1993.
M. A. Cordaro is with ACW Research, Inc., Brooklyn, NY 11204.
B. C. Sonies is with the Department of Rehabilitation Medicine, Warren Grant Magnuson Clinical Center, National Institutes of Health, Bethesda, MD 20892.
IEEE Log Number 9209954.

0018–9294/93$03.00 © 1993 IEEE