Also UCB Statistics Technical Report No. 382 February 1993 pp. 108–126

MOMENTS, CUMULANTS AND SOME APPLICATIONS TO STATIONARY RANDOM PROCESSES

BY DAVID R. BRILLINGER*

University of California, Berkeley

The paper ranges over some basic ideas concerning moments and cumulants, focusing on the case of random processes. Uses of moments and cumulants in developing large sample approximate distributions, in system identification and in inferring causal connections of a network of point processes are presented.

1. Introduction. Moments and cumulants find many uses in main stream statistics generally and with random processes particularly. Moments reflect the parameters of distributions and hence, as via the method of moments, may be used to estimate distributional parameters. Moments may be employed to develop approximations to the statistical distributions of quantities, such as sums in central limit theorems and associated expansions. Moments may be used to study the independence of variates. Moments unify diverse random processes, such as point processes and random fields, and diverse domains, such as the line or space-time.

2. Ordinary case. One can begin by asking: What is a moment? To provide an answer to this question, consider the case of the 0–1 valued

*Research partially supported by NSF Grant DMS-8900613

AMS 1980 subject classifications. Primary 62M10, 62M99.

Key words and phrases. Coherence, cumulant, moment, partial coherence, point process, system identification, time series.

variates X, Y, Z. For these variates

$$E\{XYZ\} = Prob\{X = 1, Y = 1, Z = 1\}$$

This provides an interpretation for a (third-order) moment in terms of a quantity having a primitive existence, namely a probability. Higher-order moments have a similar interpretation. One can proceed to general random variables, by noting that these may be approximated by step (or simple) functions, see eg. Feller (1966), page 107.

Next one can ask: What is a cumulant? One answer is to say that it is a combination of moments that vanishes when some subset of the variates is independent of the others. Suppose for example that X is independent of (Y, Z). The third order joint cumulant may be defined by

$$cum\{X, Y, Z\} = \tag{1}$$

$$E\{XYZ\} - E\{X\}E\{YZ\} - E\{Y\}E\{XZ\} - E\{Z\}E\{XY\} + 2E\{X\}E\{Y\}E\{Z\}$$

By substitution one quickly sees that this last expression vanishes in the case that X is independent of (Y, Z).

Expresion (1) gives one definition of a joint cumulant. An alternate way to proceed is to state that that cumulant is given by the coefficient of $i^3\alpha\beta\gamma$ in the Taylor expansion of

$$\log\left[E\{e^{i(\alpha X+\beta Y+\gamma Z)}\}\right]$$

supposing one exists.

Taking the log here converts factorizations into additivities and one sees immediately why the joint cumulants vanish in the case of independence.

Streitberg (1990) sets down a sequence of conditions that actually characterize a cumulant. These are:

1. Symmetry

$$cum\{X_1, X_2, \cdots\} = cum\{X_2, X_1, \cdots\}$$

2. Multilinearity

$$cum\{\alpha X_1, X_2, \cdots\} = \alpha cum\{X_1, X_2, \cdots\}$$

$$cum\{X_1+Y_1, X_2, \cdots\} = cum\{X_1, \cdots\} + cum\{Y_1, \cdots\}$$

3. Moment property, if the moments of **X** and **Y** are identical up to order k

$$cum\{\mathbf{X}\} = cum\{\mathbf{Y}\}$$

4. Normalization, in the expansion in terms of moments

$$cum\{X_1, \cdots, X_k\} = E\{X_1 \cdots X_k\} + \cdots$$

5. Interaction, if a subset is independent of the remainder

$$cum\{X_1, \cdots, X_k\} = 0$$

Cumulants provide a measure of Gaussianity. If the variate X is normal, then

$$cum_k\{X\} = 0 \tag{2}$$

for $k > 2$. (Here cum_k denotes the joint cumulant of X with itself k times.) Putting (2) together with the fact that the normal distribution is determined by its moments, provides a particularly brief proof of the central limit theorem. Namely suppose that $X_1, X_2, \cdots$ are independent and identically distributed with $E\{X\} = 0$ and $var\{X\} = 1$. Suppose all moments exist for X. Consider

$$S_n = (X_1 + \cdots + X_n)/\sqrt{n} \tag{3}$$

Then

$$cum_k\{S_n\} = n \; cum_k\{X\} \; / \; n^{\frac{k}{2}}$$

which tends to 0 for $k > 2$, as n tends to infinity, and in consequence S_n has a limiting normal distribution.

An error bound may be given for the degree of approximation of the distribution of a random variable by a normal, via bounds on the cumulants. In Rudzkis *et al. (1978)* the following result is developed. Consider a variate Y with mean 0 and variance 1. Suppose that

$$|cum_k\{Y\}| \le \frac{H(k!)^{1+\nu}}{\Delta^{k-2}}$$

for some $\nu \ge 0$, $H \ge 1$, then in the interval $0 \le u \le \delta/H$

$$\sup_u |Prob\{Y < u\} - \Phi(u)| \le \frac{18H}{\delta}$$

where

$$\delta = \frac{1}{7}\left[\frac{\sqrt{2}\Delta}{6}\right]^{1/(1+2\nu)}$$

In the case of a sum, such as (3), one can take $\Delta = \sqrt{n}$ for example.

3. Time series case. Consider a stationary time series $X(t)$ with domain $t = 0, \pm 1, \pm 2, \cdots$. If the $k-th$ moment exists, from the stationarity, the moment function

$$E\{X(t+u_1) \cdots X(t+u_{k-1})X(t)\}$$

will not depend on t, nor will the associated cumulant function

$$c_k(u_1, \cdots, u_{k-1})$$

$$= cum\{X(t+u_1), \cdots X(t+u_{k-1}), X(t)\} \tag{4}$$

The Fourier transforms of these $c_k(.)$ give the higher-order spectra of the series. These functions may be estimated given stretches of data.

It was indicated, by property 5 above, that a joint cumulant measures statistical dependence. This suggests formalizing the intuitive notation that values at a distance in time are not strongly dependent via

$$\sum_{u_1} \cdots \sum_{u_{k-1}} |c_k(u_1, \cdots, u_{k-1})| < \infty \tag{5}$$

for $k = 2, \cdots$. It is now direct to provide a central limit theorem for sums of values of a stationary time series. One has

$$cum_k \{ \sum_1^T X(t) / \sqrt{T} \}$$

$$= \sum_{t_1} \cdots \sum_{t_k} c_k(t_1 - t_k, \cdots, t_{k-1} - t_k) / T^{k/2}$$

$$= \sum_t \left[\sum_{u_1} \cdots \sum_{u_{k-1}} c_k(u_1, \cdots, u_{k-1}) \right] / T^{k/2}$$

$$\approx \sum_u c_2(u) \qquad k = 2$$

and

$$\to 0 \qquad k > 2$$

following (5), giving the limit normal distribution.

Another aspect of the use of cumulants is that a calculus exists for manipulating polynomials in basic variates. Suppose that

$$\begin{aligned} Y &= g(X_1, \cdots, X_L) \\ &= \sum_{\mathbf{i}} \alpha_{i_1 \cdots i_L} X_1^{i_1} \cdots X_L^{i_L} \end{aligned} \tag{6}$$

One has directly from (6) that

$$E\{Y^k\} = \sum_{\mathbf{m}} \beta_{m_1 \cdots m_L} E\{X_1^{m_1} \cdots X_L^{m_L}\}$$

but perhaps more usefully, there are rules due to Fisher, see Leonov and Shiryaev (1959), Speed (1983), providing an expression

$$cum_k\{Y\} = \sum_{\sigma} \gamma_\sigma \, cum\{X_j : j \in \sigma_1\} \cdots cum\{X_j : j \in \sigma_p\}$$

where $\sigma = (\sigma_1, \cdots, \sigma_p)$ is a partition of subscripts into blocks and the γ_σ are coefficients.

A time series analog of an expansion, like (6) for ordinary variates, is provided by the Volterra expansion

$$Y(t) = a_0 + \sum_u a_1(t-u)X(u) + \sum_{u_1,u_2} a_2(t-u_1,t-u_2)X(u_1)X(u_2) + \cdots \quad (7)$$

Using the Cramer representation of the process, namely

$$X(t) = \int e^{it\lambda} dZ_X(\lambda)$$

(7) may be written

$$a_0 + \int e^{it\lambda} A_1(\lambda) dZ_X(\lambda) + \iint e^{it(\lambda_1+\lambda_2)} A_2(\lambda_1,\lambda_2) dZ_X(\lambda_1) dZ_X(\lambda_2) + \cdots$$

in terms of the Fourier transforms of the $a_1(.)$, $a_2(.)$, $\cdots$. This form often simplifies the development of particular analytic results.

Consideration now turns to the use of moments and cumulants in the identification of nonlinear systems. In the case of a polynomial system like (7), Lee and Schetzen (1965) develop estimates of the functions $a_1(.)$, $a_2(.)$, $\cdots$ via empirical moments of the form

$$\frac{1}{T} \sum_{t=0}^{T-1} X(t+u_1) \cdots X(t+u_k) Y(t)$$

for the case that the input, $X(.)$, is Gaussian white noise.

For the case of stationary Gaussian input and a quadratic system

$$Y(t) = a_0 + \sum_u a_1(t-u)X(u) + \sum_{u_1,u_2} a_2(t-u_1,t-u_2)X(u_1)X(u_2) + noise$$

Tick (1961) developed an estimation procedure as follows. Define the cross-spectrum and cross-bispectrum via

$$cum\{dZ_X(\lambda), dZ_Y(\mu)\} = \delta(\lambda+\mu) f_{XY}(\lambda) d\lambda d\mu$$

$$cum\{dZ_X(\lambda_1), dZ_X(\lambda_2), dZ_Y(\lambda_3)\} = \delta(\lambda_1+\lambda_2+\lambda_3) f_{XXY}(\lambda_1,\lambda_2) d\lambda_1 d\lambda_2 d\lambda_3$$

respectively. One has

$$f_{YX}(\lambda) = A_1(\lambda) f_{XX}(\lambda)$$
$$f_{XXY}(-\lambda_1,-\lambda_2) = 2A_2(-\lambda_1,-\lambda_2) f_{XX}(\lambda_1) f_{XX}(\lambda_2)$$

relations from which estimates of the transfer functions, A, may be

$\sum_j M_j \; \delta(t-\tau_j)$, a hybrid process $X(\tau_j)$ and a line process, for example.

6. An example. In this section second-order moments and cumulants are employed to infer the causal connections amongst some contemporaneous point processes.

Consider the stationary bivariate point process (M, N) with points τ_k and γ_l respectively. In what follows an estimate of the product density of order 2 will be needed. The parameter is defined via

$$p_{MN}(u)\; dudt \;= E\{dM(t+u)dN(t)\}$$

$$= Prob\{dM(t+u) = 1,\; dN(t) = 1\}$$

This last suggests basing an estimate on the count

$$\#\{\,|\tau_k - \gamma_l - u\,| < \frac{h}{2}\} \tag{8}$$

for some small binwidth h. Details are given in Brillinger (1976). One result is that it appears more pertinent to graph the square root of the estimate. In the case that the processes M and N are independent, one will have $p_{MN}(u) = p_M p_N$, which possibility may be examined via the statistic (8).

The suggested estimate will be illustrated with some neurophysiological data. Concern in the experiment was with auditory paths in the brain of the cat. To collect data, microelectrodes were inserted with location tuned to sound response. Data was recorded when the neurons were firing spontaneously. Also responses were evoked experimentally by 200 msec. noise bursts, that were applied every 1000 msec., via speakers inserted in the ears. The firing times of 8 neurons were recorded. Figure 1 provides the data itself for 4 selected cells, 2 in the case with stimulation, 2 when the firing is spontaneous. Each horizontal line plots firings as a function of time since stimulus initiation in a 1000 msec. time period. The stimulus was applied 505 times in these examples. In the stimulated case one notices vertical darkening corresponding to excess firing just after the stimulus has been applied. Neurophysiologists speak of locking. In the spontaneous case no locking is apparent. There is some evidence of non-stationarity in this case.

Figure 2 provides the square root of a multiple of (8). The horizontal dashed lines are ± 2 standard errors about a horizontal line corresponding to independence in the stationary case. One infers that the cell pairs are associated in each case. However in the stimulated case one has to wonder if the apparent association of units 6 and 7 is not due to the fact that the cells are being stimulated at the same times.

Fourier techniques provide one means to address this concern. Write

$$d_M^T(\lambda) = \sum_k e^{-i\lambda\tau_k}$$

$$d_N^T(\lambda) = \sum_l e^{-i\lambda\gamma_l}$$

for the data $0 \le \tau_k, \gamma_l < T$. For $\lambda \neq 0$ one has

$$E\{d_M^T(\lambda)\overline{d_N^T(\lambda)}\} \approx 2\pi T\, f_{MN}(\lambda)$$

with $f_{MN}(.)$ the cross-spectrum given by

$$f_{MN}(\lambda) = \frac{1}{2\pi}\int e^{-i\lambda u}\, q_{MN}(u)\, du$$

A useful quantity for measuring the association of M and N may now be defined. It is the coherence,

$$|R_{MN}(\lambda)|^2 = |f_{MN}(\lambda)|^2 \,/\, f_{MM}(\lambda) f_{NN}(\lambda)$$

with the interpretation

$$\lim_{T\to\infty} |corr\{d_M^T(\lambda),\, d_N^T(\lambda)\}|^2$$

It satisfies $0 \le |R_{MN}(\lambda)|^2 \le 1$, with greater association corresponding to values nearer 1. Figure 3 provides coherence estimates for the cell pairs of Figure 2. This evidence of association is in accord with that of Figure 2. The dashed horizontal line provides the 95% point of the null distribution of the coherence estimate.

To return to the driving question of how to "remove" the effects of the stimulus, one can consider the partial coherence. This has the interpretation

$$\lim_{T\to\infty} |corr\{d_M^T - \alpha d_S^T, d_N^T - \beta d_S^T\}|^2$$

with α, β regression coefficients and S referring to the process of stimulus times. Suppressing the dependence on λ the partial coherence is given by $|R_{MN|S}|^2$ where

$$R_{MN|S} = \frac{R_{MN} - R_{MS}R_{SN}}{\sqrt{(1-|R_{MS}|^2)(1-|R_{NS}|^2)}}$$

Figure 4 provides the estimated partial coherence of neurons 6 and 7 in the stimulated case. The level apparent in the top graph of Figure 3 has fallen off substantially suggesting that the association evidenced in Figures 2 and 3 is due to the stimulus.

For interests sake Figure 5 provides the coherence estimate for neurons 3 and 4 in the case of applied stimulation. One might wonder if they would become more strongly associated in the presence of stimulation. The results do not suggest that this has happened.

7. Conclusions. In summary, moments and cumulants may be employed to develop approximations to distributions, approximations such as the normal or the Poisson. They may be employed in system identification. They may be used to infer the "wiring" diagram of a collection of interacting point processes.

The approach presented is nonparametric, not based on special stochastic processes described by finite dimensional parameters. Brillinger (1991) provides a variety of references concerning the work pre 1980 on higher moments and spectra.

Acknowledgements. The neurophysiological data were provided by Alessandro Villa. Terry Speed mentioned the Streitberg (1990) result.

REFERENCES

BRILLINGER, D. R. (1976). Estimation of the second-order intensities of a bivariate stationary point process. J. Roy. Statist. Soc. B 38, 60-66.

BRILLINGER, D. R. (1977). The identification of a particular nonlinear time series system. Biometrika 64, 509-515.

BRILLINGER, D. R. (1991). Some history of the study of higher-order moments and spectra. Statistica Sinica 1, 465-476.

BRILLINGER, D. R. (1992). Nerve cell spike train data analysis: a progression of technique. J. Amer. Statist. Assoc. 87, June.

FELLER, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II. New York, Wiley.

KORENBERG, M. J. (1973). Cross-correlation analysis of neural cascades. In Proc. 10th Annual Rocky Mountain Biomed. Symp. 47-51. Instrument Society of America, Pitttsburgh.

LEE, Y. W. and SCHETZEN, M. (1965). Measurement of the Wiener kernels of a nonlinear system by crosscorrelation. Internat. J. Control 2, 237-254.

LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. Theor. Prob. Appl. 4, 319-329.

RUDZKIS, R., SAULIS, L. and STATULEVICIUS, V. (1978). A general lemma on probabilities of large deviations. Liet. Mat. Rink. 18, 99-116.

SPEED, T. P. (1983). Cumulants and partition lattices. Austral. J. Statist. 25, 378-388.

STREITBERG, B. (1990). Lancaster interactions revisited. Ann. Statist. 18, 1878-1885.

TICK, L. (1961). The estimation of transfer functions of quadratic systems. Technometrics 3, 563-567.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720

Figure Legends

Figure 1. Rastor plot of the firing times of 4 neurons in successive 1000 msec. periods. There are 505 horizontal lines of firing times.

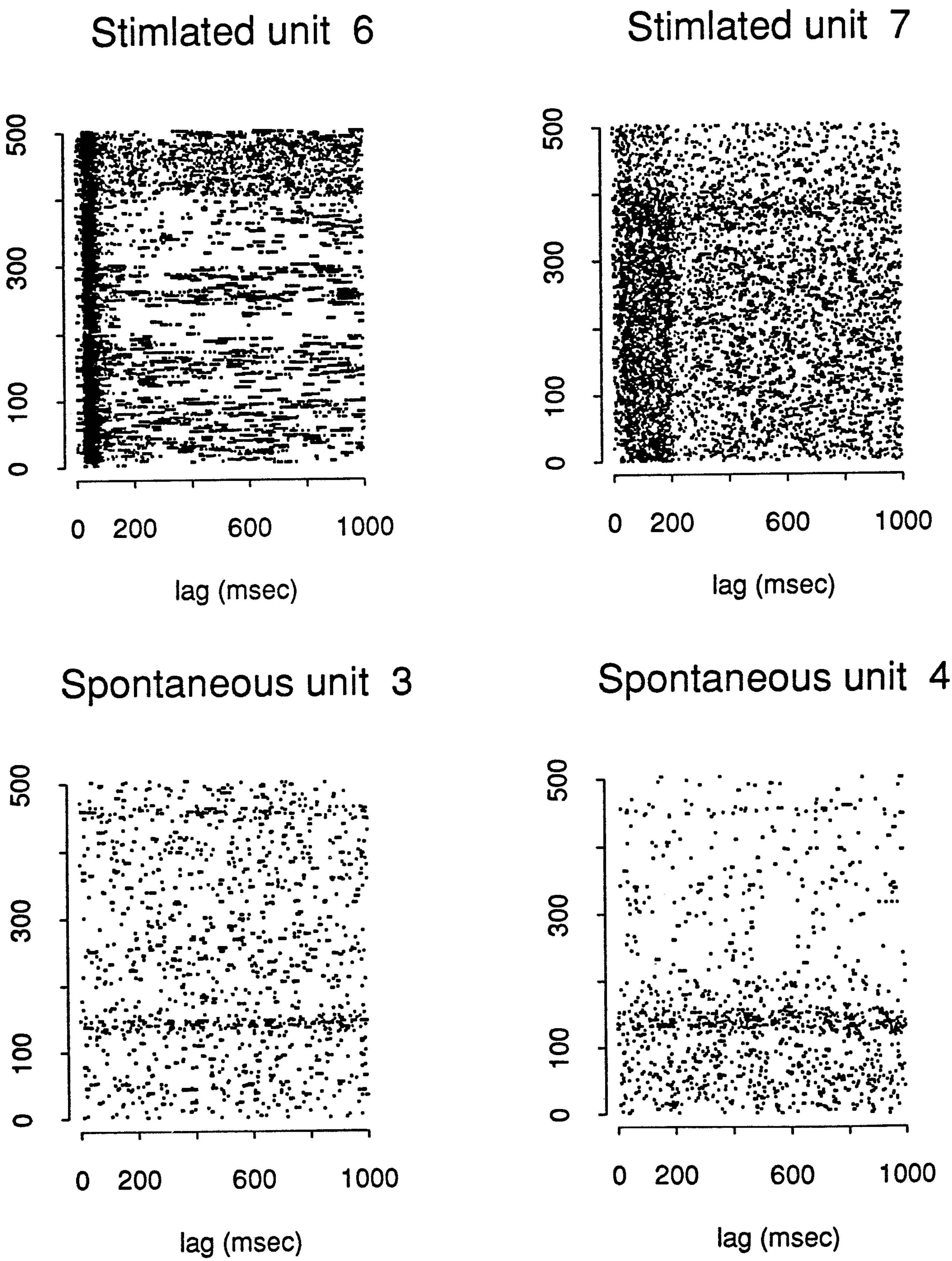
Figure 2. The square root of a multiple of the quantity (6). Were the processes independent and stationary then about 5% of the values should lie outside the band defined by the two horizontal dashed lines.

Figure 3. Estimated coherences of cells 6 and 7 in the stimulated case and 3 and 4 when the firing is spontaneous.

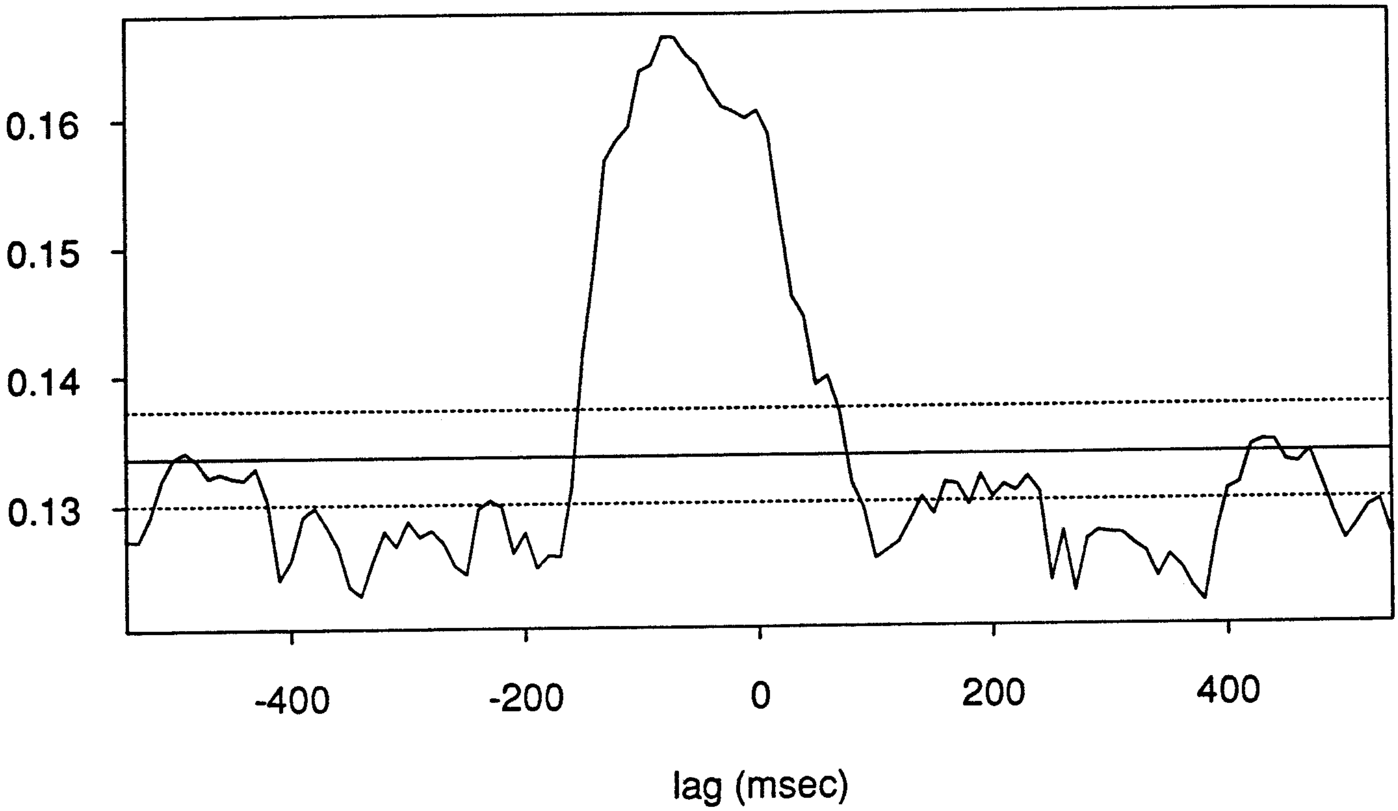
Figure 4. Estimated partial coherence of cells 6 and 7 "removing" the effect of the stimulus.

Figure 5. The estimated coherence of cells 3 and 4 in the case of stimulation.

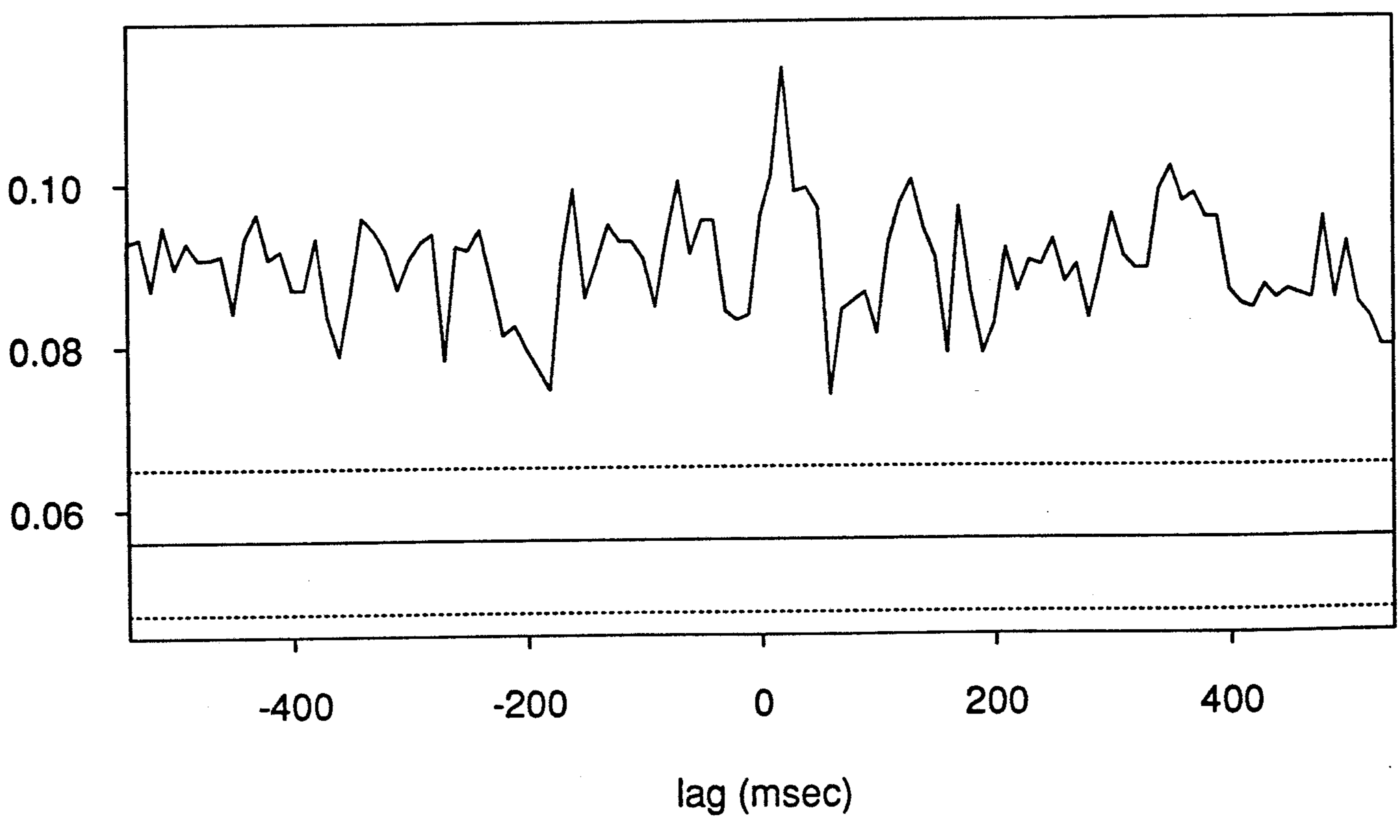
Spike times following stimulus



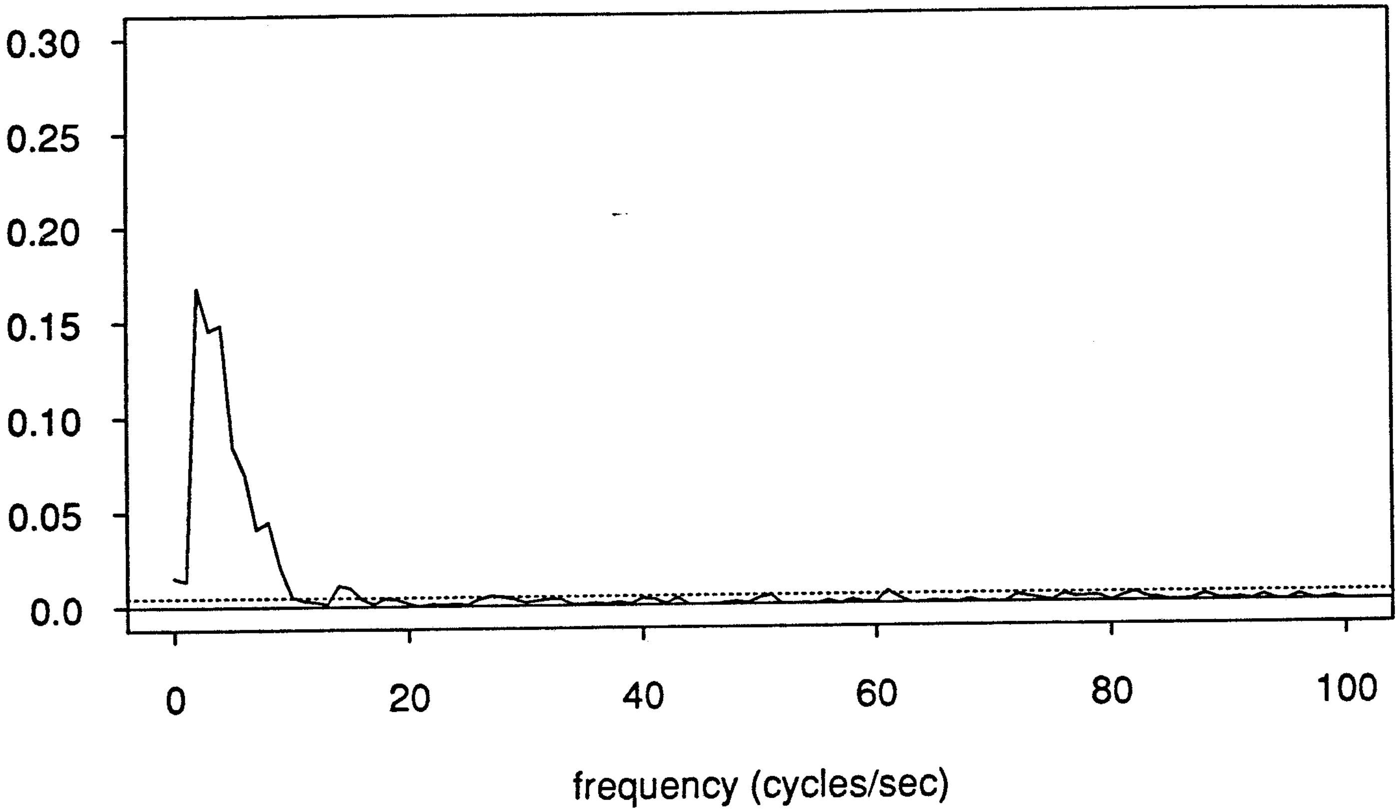
Stimulated units 6&7, sqrt(crossint)



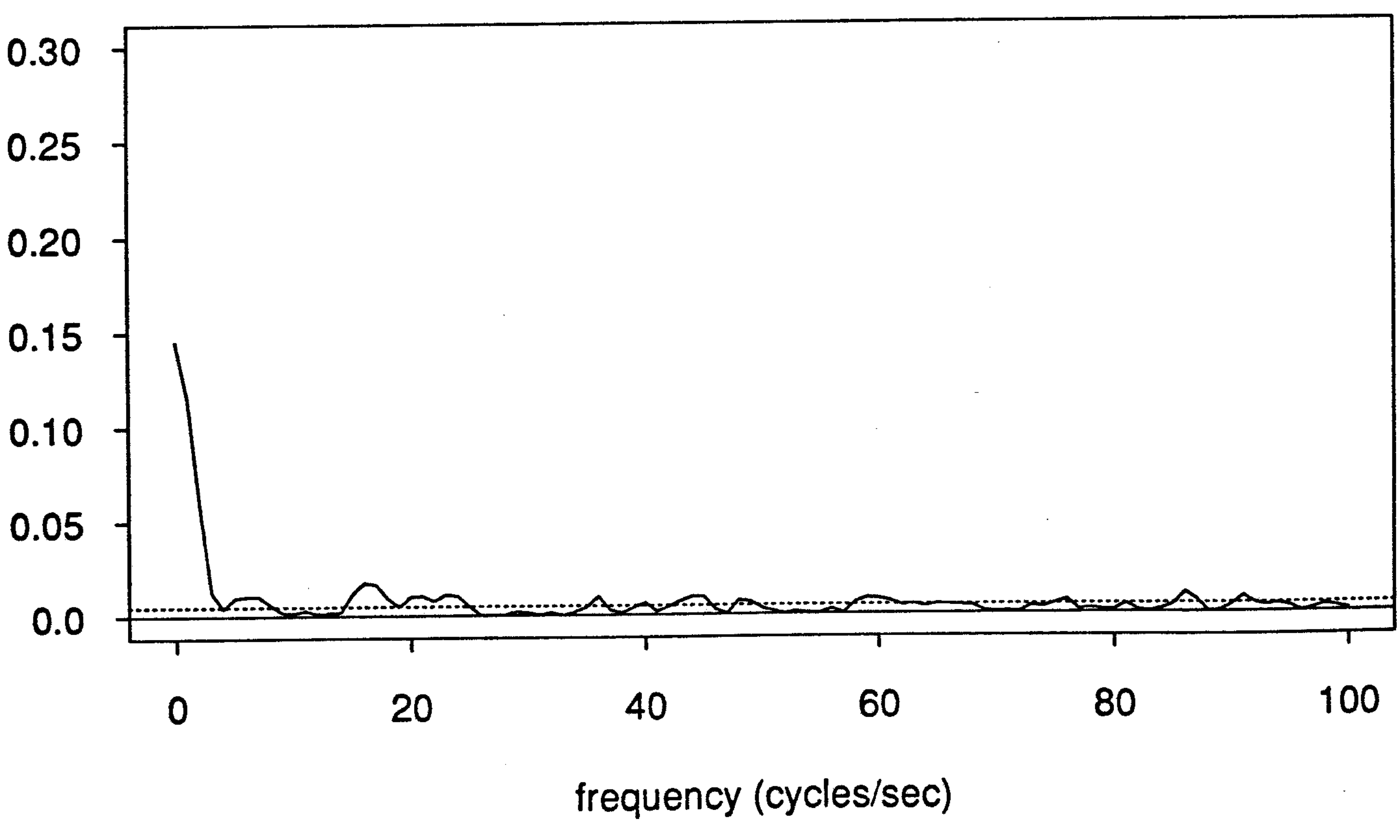
Spontaneous units 3&4, sqrt(crossint)



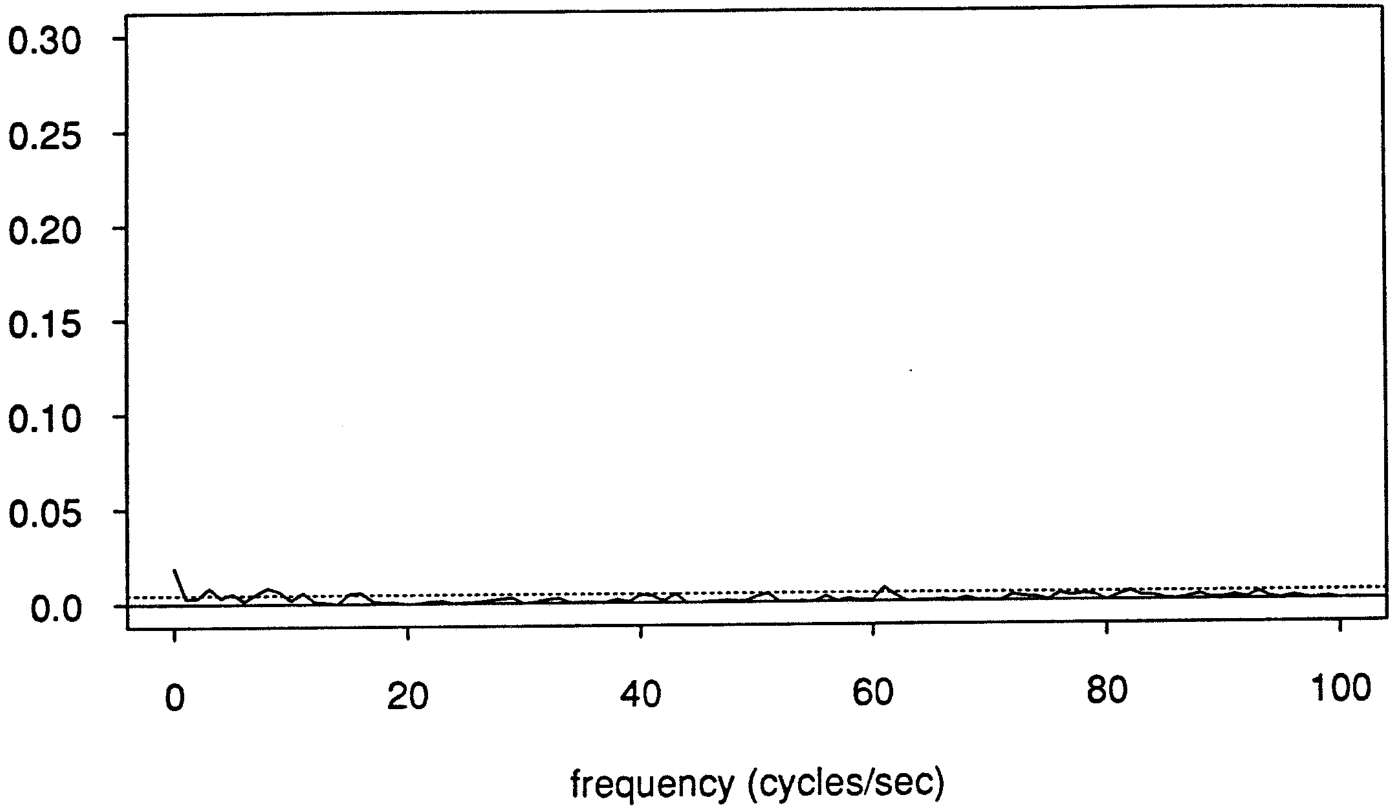
Stimulated units 6&7, coherence

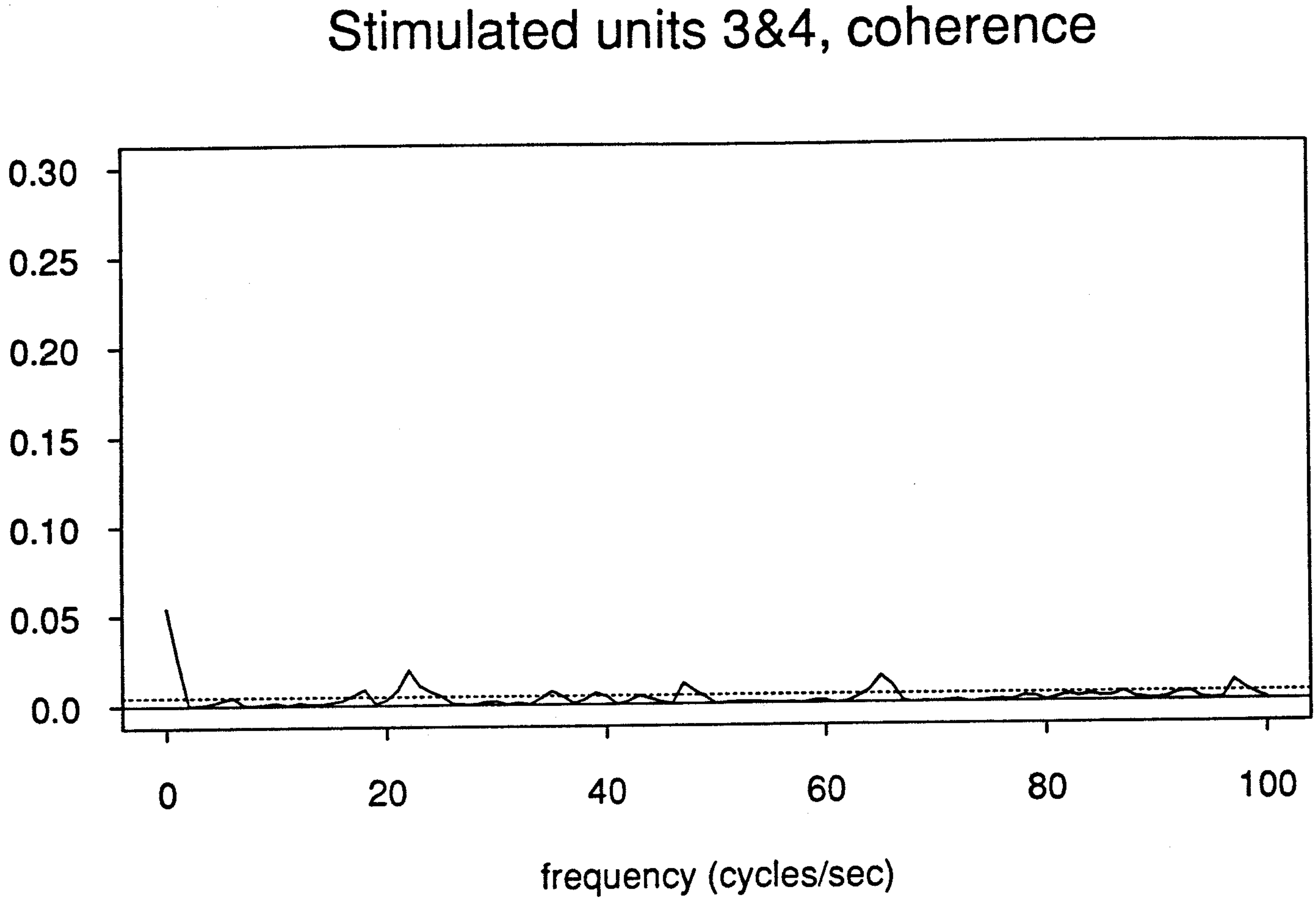


Spontaneous units 3&4, coherence



Units 6&7, partial coherence




Stimulated units 3&4, coherence
0.30
0.25
0.20
0.15
0.10
0.05
0.0
0
20
40
60
80
100
frequency (cycles/sec)

NPS-OR-92-012

NAVAL POSTGRADUATE SCHOOL

Monterey, California

MOMENTS AND SIGNAL PROCESSING

Proceedings of the Conference held at the Naval Postgraduate School March 30-31 1992, Monterey, CA

Edited by:

P. Purdue, Naval Postgraduate School, Monterey, CA
H. Solomon, Stanford University, Stanford, CA

26 August 1992

Prepared for:
National Security Agency
Office of Naval Research