# MULTIVARIATE WALSH–FOURIER ANALYSIS

BY DAVID S. STOFFER

*University of Pittsburgh*

*First version received December 1988*

**Abstract.** In this paper we establish a statistical methodology for the spectral analysis of stationary multivariate time series via the Walsh–Fourier transform. Theoretical results pertaining to the definition and estimation of the Walsh–Fourier spectral matrix and functions of that matrix including cross-spectra, coherency and phase are given. An example of the statistical techniques developed in this paper is given; in particular, the methodologies are applied to neonatal sleep data collected from a study of the effect of maternal substance use during pregnancy.

**Keywords.** Multivariate time series; Walsh spectral analysis; Walsh–Fourier cross-spectra; Walsh–Fourier coherency; sequency domain; neonatal EEG sleep patterns

## 1. INTRODUCTION

Recently, attention has been focused on the statistical applications of Walsh–Fourier analysis to real-time stationary time series. Kohn (1980a,b) laid the groundwork by showing that many of the results concerning the decomposition of stationary time series using trigonometric functions have their Walsh function analogs, although Morettin (1974) had obtained limit theorems for the Walsh–Fourier transform of stationary time series earlier. Other papers establishing Walsh–Fourier theory for real-time stationary processes are those of Morettin (1981, 1983) and Stoffer (1985, 1987). Statistical data analysis via the Walsh–Fourier transform can be found in Ott and Kronmal (1976), where the transform is used in classification problems for strictly stationary binary data, and in Stoffer *et al.* (1988), where an analysis of variance based on the Walsh–Fourier transform is used to assess the effect of maternal alcohol consumption on neonatal sleep-state cycling. Further applications of Walsh spectral analysis can be found in the *Proceedings of the Symposium on the Applications of Walsh Functions*, Ahmed and Rao (1975); and Beauchamp (1975, 1984), to mention a few. Beauchamp (1975, Section VF; 1984, Section 3.3.4) empirically demonstrated that the roles of Walsh and Fourier spectral analysis for discontinuous and smooth-varying signals respectively are clear. He concluded that where the signal is derived from a sinusoidally based waveform, Fourier analysis is relevant, and where the signal contains sharp discontinuities and a limited number of levels, Walsh analysis is appropriate. The aforementioned works demonstrate that the Walsh–Fourier transform can be a powerful tool in the statistical analysis of spectra. Hence it is of considerable importance that the existing Walsh–Fourier theory for the

0143-9782/90/01 057–17 $02.50/0
JOURNAL OF TIME SERIES ANALYSIS Vol. 11, No. 1

statistical analysis of time series data (both univariate and multivariate), and in particular processes with a limited number of levels such as discrete-valued and categorical processes (i.e. square waveforms), be extended at least to the point of development of the statistical theory of Fourier (trigonometric) analysis for sinusoidal waveforms.

The Walsh functions form a complete orthonormal sequence on $[0, 1)$ and take on only two values, $+1$ and $-1$ (or 'on' and 'off'). They are ordered by the number of zero-crossings (or switches) which is called *sequency* (although other orderings exist, sequency ordering is the easiest to interpret since it is comparable with the frequency ordering of sines and cosines). Let $W(n, \lambda)$, $n = 0, 1, 2, \ldots, 0 \leq \lambda < 1$, denote the $n$th sequency-ordered Walsh function; then $W(n, .)$ makes $n$ zero-crossings in $[0, 1)$. The first eight sequency-ordered Walsh functions $W(n, m/N)$, $m, n = 0, 1, \ldots, 7$, corresponding to a sample of length $N = 2^3$ are shown below as the rows (columns) of a symmetric matrix called the Walsh-ordered Hadamard matrix $\boldsymbol{H}_W(3)$:

$$\boldsymbol{H}_W(3) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}.$$

Let $\boldsymbol{X}(0), \boldsymbol{X}(1), \ldots, \boldsymbol{X}(N-1)$ be a sample of length $N = 2^p$, $p > 0$ integer, from an $r \times 1$ vector-valued stationary zero-mean time series $\{\boldsymbol{X}(t), t = 0, \pm 1, \pm 2, \ldots\}$ with an $r \times r$ autocovariance function matrix $\boldsymbol{\Gamma}(h) = \operatorname{cov}\{\boldsymbol{X}(t+h), \boldsymbol{X}(t)\}$, $h = 0, \pm 1, \pm 2, \ldots$. Then

$$\boldsymbol{d}_N(\lambda) = N^{-1/2} \sum_{t=0}^{N-1} \boldsymbol{X}(t) W(t, \lambda) \qquad (0 \leq \lambda < 1)$$

is the $r \times 1$ finite (or discrete) Walsh–Fourier transform of the data. The $r \times r$ covariance matrix of $\boldsymbol{d}_N(\lambda)$ is

$$\operatorname{var}\{\boldsymbol{d}_N(\lambda)\} = \sum_{j=0}^{N-1} \boldsymbol{\tau}(j) W(j, \lambda)$$

where $\boldsymbol{\tau}(j)$ is the $r \times r$ logical covariance matrix of $\boldsymbol{X}(t)$ and is given by

$$\boldsymbol{\tau}(j) = 2^{-(q+1)} \sum_{k=0}^{2^q-1} \{\boldsymbol{\Gamma}(j \oplus k - k) + \boldsymbol{\Gamma}'(j \oplus k - k)\} \qquad (2^q \leq j < 2^{q+1});$$

$j \oplus k$ denotes the dyadic addition of $j$ and $k$ (the concept of the logical covariance function was introduced by Robinson (1972) and formalized by Kohn (1980a)). Denote the elements of $\boldsymbol{\Gamma}(h)$ by $c_{ab}(h)$, $a, b = 1, \ldots, r$. If the $c_{ab}(h)$ are absolutely summable (i.e. $\Sigma |c_{ab}(h)| < \infty$, although slightly weaker conditions will suffice), then var $\{\boldsymbol{d}_N(\lambda)\} \to \boldsymbol{f}(\lambda)$ as $N \to \infty$ where

$$\boldsymbol{f}(\lambda) = \sum_{j=0}^{\infty} \boldsymbol{\tau}(j) W(j, \lambda)$$

is the $r \times r$ Walsh–Fourier spectral density matrix of $\boldsymbol{X}(t)$.

In the univariate case $r = 1$, various results relating the convergence of $d_N(\lambda)$ to $f(\lambda)$ exist. For example, Kohn (1980a) shows that if $\lambda_N$ is dyadically rational (i.e. its binary representation is finite) and $\lambda_N \oplus \lambda \to 0$ as $N \to \infty$, then $E\{d_N^2(\lambda_N)\} \to f(\lambda)$ as $N \to \infty$. Although the asymptotic covariance of the Walsh–Fourier transform at two distinct sequences is not in general zero, Kohn (1980a) also shows that, if $\lambda_{1,N}$ and $\lambda_{2,N}$ are dyadically rational, $|\lambda_{1,N} - \lambda_{2,N}| \geqslant N^{-1}$ and $\lambda_{i,N} \oplus \lambda \to 0$ $(i = 1, 2)$ as $N \to \infty$, then $E\{d_N(\lambda_{1,N}) d_N(\lambda_{2,N})\} \to 0$. These results are useful in obtaining consistent estimators of $f(\lambda)$. Morettin (1974, 1983), Kohn (1980a,b), and Stoffer (1985, 1987) have established central limit theorems for the Walsh–Fourier transform under a wide range of conditions; the basic result is that under appropriate conditions (which typically include some type of mixing condition) $d_N(\lambda) \xrightarrow{\mathscr{D}} N\{0, f(\lambda)\}$. Estimation of the Walsh–Fourier spectrum is discussed by Kohn (1980b) and Stoffer (1987). With the exception of the work of Stoffer (1987) and Stoffer *et al.* (1988), little has been done to extend the statistical theory of Walsh–Fourier analysis beyond the basic kinds of results just mentioned.

Similarly, little is known beyond the basics about the multivariate case $r > 1$. Kohn (1980b, section 3) addresses the multivariate case in a brief note claiming that, with a few exceptions, the results obtained by Kohn (1980a,b) for the scalar case carry over, in an obvious way, to the vector case; while this is true, it ignores a most important aspect of the analysis of multivariate time series, namely cross-spectral analysis.

In this paper we establish the analysis of multivariate stationary time series via the Walsh–Fourier transform with emphasis on the analysis of cross-spectra, coherency and phase. Since Walsh analysis is appropriate when the signal contains sharp discontinuities and a limited number of levels, our assumptions will be fairly general. Thus, for example, we do not assume that the process of interest is generated by a linear process. Theoretical results pertaining to the Walsh–Fourier spectral matrix, cross-spectra, coherency, gain and phase spectra and their estimation are given in Section 2. An example of the application of the techniques given in Section 2 is presented in Section 3. Here, a bivariate Walsh–Fourier analysis of neonatal sleep data collected from a study of the effects of maternal substance use during pregnancy is used to aid the detection of alterations or disruptions in the ultradian rhythms of sleep as a result of exposure to alcohol.

One final comment should be made. As will be seen, Walsh–Fourier analysis is not a replacement for Fourier (trigonometric) analysis, nor is it simply a replicate of Fourier analysis. Thus it is not necessary to choose between the two approaches—both may be simultaneously useful in some problems.

## 2. ESTIMATION OF THE WALSH–FOURIER SPECTRAL MATRIX

In this section we present theoretical results pertaining to the definition and estimation of the Walsh–Fourier spectral matrix and functions of the matrix. As previously stated, in order to remain fairly general we follow the approach of Brillinger (1975) by assuming that the process of interest has higher moments (which is reasonable for processes with a limited number of levels) and that cumulants of all orders exist (see Brillinger, 1975, p. 25).

Given an $r \times 1$ vector time series $\boldsymbol{X}(t)$, $t = 0, \pm 1, \pm 2, \ldots$, with components $X_a(t)$ such that $E|X_a(t)|^k < \infty$, $a = 1, \ldots, r$, let

$$c_{a_1,\ldots,a_k}(t_1, \ldots, t_k) = \operatorname{cum}\{X_{a_1}(t_1), \ldots, X_{a_k}(t_k)\} \tag{2.1}$$

for $a_1, \ldots, a_k = 1, \ldots, r$ and $t_1, \ldots, t_k = 0, \pm 1, \pm 2, \ldots$, be the joint cumulant of order $k$ of the series $\boldsymbol{X}(t)$. For $\boldsymbol{X}(t)$ strictly stationary we use the asymmetric notation (Brillinger, 1975, p. 23)

$$c_{a_1,\ldots,a_k}(t_1, \ldots, t_{k-1}) = c_{a_1,\ldots,a_k}(t_1, \ldots, t_{k-1}, 0).$$

The following assumption is made throughout.

ASSUMPTION 1. $\{\boldsymbol{X}(t),\ t = 0, \pm 1, \pm 2, \ldots\}$ is a strictly stationary zero-mean $r \times 1$ vector series with components $X_a(t)$, $a = 1, \ldots, r$, all of whose moments exist with

$$\sum_{j_1,\ldots,j_{k-1}=-\infty}^{\infty} |c_{a_1,\ldots,a_k}(j_1, \ldots, j_{k-1})| < \infty$$

for $a_1, \ldots, a_k = 1, \ldots, r$ and $k = 2, 3, \ldots$.

Note that, by Assumption 1,

$$\sum_{h=-\infty}^{\infty} |c_{ab}(h)| < \infty \qquad (a, b = 1, \ldots, r)$$

and hence the Walsh–Fourier spectral matrix is well defined.

We now discuss the asymptotic properties of the transform $\boldsymbol{d}_N(\lambda)$. Theorem 1 is the multivariate extension of the central limit theorem given by Morettin (1983, Theorem 1). Let $N_r(\boldsymbol{0}, \boldsymbol{V})$ denote the $r$-dimensional normal distribution with zero mean and covariance matrix $\boldsymbol{V}$.

THEOREM 1. *Let $\boldsymbol{X}(t)$ satisfy Assumption 1; then as $N \to \infty$*

$$\boldsymbol{d}_N(\lambda) \xrightarrow{\mathcal{D}} \mathrm{N}_r\{\boldsymbol{0}, \boldsymbol{f}(\lambda)\}.$$

PROOF. Since $E\{\boldsymbol{d}_N(\lambda)\} = \boldsymbol{0}$ and $\operatorname{var}\{\boldsymbol{d}_N(\lambda)\} \to \boldsymbol{f}(\lambda)$ as $N \to \infty$, the first- and second-order cumulants behave as required in the theorem. Next, we show that all cumulants of order $k > 2$ tend to zero as $N \to \infty$ and the theorem will follow from Brillinger (1975, p. 403, Lemma P4.5). Using the notation (2.1), we obtain

$$\operatorname{cum}\{d_{a_1}(\lambda_1), \ldots, d_{a_k}(\lambda_k)\} = N^{-k/2} \sum_{j_1,\ldots,j_k=0}^{N-1} W(j_1, \lambda_1) \ldots W(j_k, \lambda_k) c_{a_1,\ldots,a_k}(j_1, \ldots, j_k).$$

Thus

$$\operatorname{cum}\{d_{a_1}(\lambda_1), \ldots, d_{a_k}(\lambda_k)\} \leqslant N^{-k/2+1} \sum_{j_1,\ldots,j_{k-1}=-\infty}^{\infty} |c_{a_1,\ldots,a_k}(j_1, \ldots, j_{k-1})|$$

from which it follows that all cumulants of order $k > 2$ tend to zero as $N \to \infty$. ∎

In order to estimate the spectral matrix $\boldsymbol{f}(\lambda)$ consistently, the following lemma is needed. This lemma follows from Theorem 1 and from Kohn (1980a, Corollary 3); the result of the lemma was discussed in Section 1.

LEMMA 1. *Let $\lambda_{j,N} = j/N$ for $1 \leqslant j \leqslant N - 1$ and suppose that the collection $\{\lambda_{j(m),N};\ m = 1, \ldots, M\}$ is close to $\lambda$ such that $\lambda \oplus \lambda_{j(m),N} \to 0$ as $N \to \infty$ and $|\lambda_{j(m),N} - \lambda_{j(k),N}| \geqslant N^{-1}$ for $m \neq k = 1, \ldots, M$. Then, under Assumption 1, the $\boldsymbol{d}_N(\lambda_{j(m),N})$, $m = 1, \ldots, M$ are asymptotically independent $\mathrm{N}_r\{\boldsymbol{0}, \boldsymbol{f}(\lambda)\}$ vector variables.*

As previously stated, the asymptotic covariance of the Walsh–Fourier transform at two distinct sequencies is not necessarily zero. Hence, Lemma 1 is not true for a general collection of sequencies $\{\lambda_{j,N}\}$. With this being considered, we also note that, while some of the results that follow may look similar to their trigonometric counterparts, they are not trivial consequences of the analogous Fourier results.

Let $\boldsymbol{I}_N(\lambda) = \boldsymbol{d}_N(\lambda)\boldsymbol{d}'_N(\lambda)$ be the $r \times r$ periodogram matrix with components $I_{ab}(\lambda)$, $a, b = 1, \ldots, r$. The next theorem establishes asymptotic properties of the periodogram matrix.

THEOREM 2. *Let $\boldsymbol{X}(t)$ satisfy Assumption 1 and let the collection $\{\lambda_{j(m),N}, m = 1, \ldots, M\}$ be as in Lemma 1. Then*

$$\operatorname{cov}\{I_{a_1a_2}(\lambda_{j(m),N}), I_{b_1b_2}(\lambda_{j(k),N})\} = \begin{cases} \mathrm{O}(N^{-1}) & m \neq k \\ f_{a_1b_1}(\lambda)f_{a_2b_2}(\lambda) + f_{a_1b_2}(\lambda)f_{a_2b_1}(\lambda) + \mathrm{o}(1) & m = k \end{cases}$$

for $m, k = 1, \ldots, M$.

PROOF. To simplify the notation, put $\lambda_l = \lambda_{j(l),N}$, $l = 1, \ldots, M$. Then using Hannan (1970, p. 23)

$$\begin{aligned}E\{I_{a_1a_2}(\lambda_m)I_{b_1b_2}(\lambda_k)\} &= N^{-2}\sum_{t_1..}^{N-1}\ldots\sum_{..t_4=0}^{N-1}\{c_{a_1a_2}(t_2-t_1)c_{b_1b_2}(t_4-t_3)\\ &\quad + c_{a_1b_1}(t_3-t_1)c_{a_2b_2}(t_4-t_2) + c_{a_1b_2}(t_4-t_1)c_{a_2b_1}(t_3-t_2)\\ &\quad + c_{a_1a_2b_1b_2}(t_1,t_2,t_3,t_4)\} \times W(t_1,\lambda_m)W(t_2,\lambda_m)W(t_3,\lambda_k)W(t_4,\lambda_k)\\ &\equiv S_1+S_2+S_3+S_4.\end{aligned}$$

It follows from the above that $S_2 + S_3 + S_4$ is the desired covariance. By Lemma 1, as $N \to \infty$,

$$S_2 = E\{d_{a_1}(\lambda_m)d_{b_1}(\lambda_k)\}E\{d_{a_2}(\lambda_m)d_{b_2}(\lambda_k)\} \to \begin{cases} 0 & m \neq k \\ f_{a_1b_1}(\lambda)f_{a_2b_2}(\lambda) & m = k. \end{cases}$$

Similarly $S_3 \to 0$ when $m \neq k$ and $S_3 \to f_{a_1b_2}(\lambda)f_{a_2b_1}(\lambda)$ when $m = k$. Also, by Assumption 1 we have that $S_4 \to 0$ as $N \to \infty$. To establish the bound when $m \neq k$ note that, following Kohn (1980a, Corollary 2), it can be shown that there exists an integer $N_{ab}(m, k)$ such that $E\{d_a(\lambda_m)d_b(\lambda_k)\} = 0$ for $N > N_{ab}(m, k)$ and $m \neq k$; $a, b = 1, \ldots, r$. This implies that there is an integer $N(m, k)$ such that $S_2 = S_3 = 0$ for $N > N(m, k)$, $m \neq k$. Finally, by Assumption 1 and following the proof of Theorem 1,

$$\begin{aligned}|S_4| &\leq N^{-2}\sum_{t_1..}^{N-1}\ldots\sum_{..t_4=0}^{N-1}|c_{a_1a_2b_1b_2}(t_1,\ldots,t_4)|\\ &\leq N^{-1}\sum_{t_1..}^{\infty}\ldots\sum_{..t_3=-\infty}^{\infty}|c_{a_1a_2b_1b_2}(t_1,\ldots,t_3)|\end{aligned}$$

which completes the proof. ■

The result of Theorem 2 suggests the following consistent estimate of the Walsh–Fourier spectral density matrix $\boldsymbol{f}(\lambda)$. For the collection $\{\lambda_{j(m),N};\ m = 1, \ldots, M\}$, as defined in Lemma 1, define the smoothed periodogram estimate of $\boldsymbol{f}(\lambda)$ to be the $r \times r$ matrix

$$\hat{\boldsymbol{f}}_{M,N}(\lambda) = \sum_{m=1}^{M} H_N(m)\boldsymbol{I}_N(\lambda_{j(m),N}) \tag{2.2}$$

such that $M \to \infty$ and $M/N \to 0$ as $N \to \infty$, and such that $H_N(m) \geq 0$, $m = 1, \ldots, M$, $\Sigma_{m=1}^{M} H_N(m) = 1$, but with $H_N(m) = O(M^{-1})$ so that $\Sigma_{m=1}^{M} H_N^2(m) \to 0$ as $N \to \infty$ (for notational convenience the dependence of $M$ on $N$ will be understood). Note that the estimator given in (2.2) is in general different from the estimator given by Kohn (1980b) for the univariate case. This estimator also has some advantages over that described by Kohn: one advantage is that the estimator is computationally simpler, but more important is that the estimator will be easy to interpret while interpretation of Kohn's estimator may be difficult owing to its dependence on dyadic addition. These claims will become evident in the application. The consistency of $\hat{\boldsymbol{f}}_{M,N}(\lambda)$ is now established.

THEOREM 3. *Let $X(t)$ satisfy Assumption 1 and let $\hat{f}_{M,N}(\lambda) = \{\hat{f}_{ab}(\lambda)\}$ be defined in (2.2). Put $h_{M,N}^2 = \Sigma_{m=1}^M H_N^2(m)$; then as $N \to \infty$*

(a) $E\{\hat{f}_{M,N}(\lambda)\} \to f(\lambda)$,

(b) $h_{M,N}^{-2} \operatorname{cov}\{\hat{f}_{a_1a_2}(\lambda), \hat{f}_{b_1b_2}(\lambda)\} \to f_{a_1b_1}(\lambda)f_{a_2b_2}(\lambda) + f_{a_1b_2}(\lambda)f_{a_2b_1}(\lambda)$.

PROOF. To establish (a) note that

$$E\{\hat{f}_{M,N}(\lambda)\} = \sum_{m=1}^{M} H_N(m)E\{I_N(\lambda_{j(m),N})\}$$
$$= \sum_{m=1}^{M} H_N(m)\{f(\lambda) + \mathbf{o}(1)\} = f(\lambda) + \mathbf{o}(1).$$

To establish (b) put $\lambda_{j(l),N} = \lambda_l$, $l = 1, \ldots, M$; then

$$\operatorname{cov}\{\hat{f}_{a_1a_2}(\lambda), \hat{f}_{b_1b_2}(\lambda)\} = \sum_{m=1}^{M} H_N^2(m) \operatorname{cov}\{I_{a_1a_2}(\lambda_m), I_{b_1b_2}(\lambda_m)\}$$
$$+ \sum\sum_{m \neq k}^{M} H_N(m)H_N(k) \operatorname{cov}\{I_{a_1a_2}(\lambda_m), I_{b_1b_2}(\lambda_k)\}.$$

From Theorem 2 when $m \neq k$, $|\operatorname{cov}\{I_{a_1a_2}(\lambda_m), I_{b_1b_2}(\lambda_k)\}| \leq cN^{-1}$, so that

$$\left|\sum\sum_{m \neq k}^{M} H_N(m)H_N(k) \operatorname{cov}\{I_{a_1a_2}(\lambda_m), I_{b_1b_2}(\lambda_k)\}\right| \leq cN^{-1}\left\{\sum_{m=1}^{M} H_N(m)\right\}^2$$
$$\leq cN^{-1}\left\{\sum_{m=1}^{M} H_N^2(m)\right\}M$$
$$= o(h_{M,N}^2).$$

From Theorem 2 when $m = k$, $\operatorname{cov}\{I_{a_1a_2}(\lambda_m), I_{b_1b_2}(\lambda_m)\} = f_{a_1a_2b_1b_2}(\lambda) + o(1)$, where $f_{a_1a_2b_1b_2}(\lambda)$ denotes the right-hand side of (b). Hence

$$\operatorname{cov}\{\hat{f}_{a_1a_2}(\lambda), \hat{f}_{b_1b_2}(\lambda)\} = h_{M,N}^2 f_{a_1a_2b_1b_2}(\lambda) + o(h_{M,N}^2) + o(h_{M,N}^2)$$

and the theorem follows. ∎

It is possible to improve on the result of Theorem 3(a) which pertains to the asymptotic bias of the spectral estimate (2.2) and to calculate the rate at which the bias tends to zero. First the following assumption is needed.

ASSUMPTION 2. $\Sigma_{j=0}^{\infty} j|c_{ab}(j)| < \infty$, $a,b = 1, \ldots, r$.

THEOREM 4. *Let Assumptions 1 and 2 hold. Then*

$$(h_{M,N}^{-2})^{1/2}|E\{\hat{f}_{ab}(\lambda)\} - f_{ab}(\lambda)| \to 0 \qquad (a,b = 1, \ldots, r)$$

*as $N \to \infty$, where $h_{M,N}^2$ is defined in Theorem 3.*

PROOF. First note that, since $W(k, \lambda_1)W(k, \lambda_2) = W(k, \lambda_1 \oplus \lambda_2)$,

$$E\{\hat{f}_{ab}(\lambda)\} = \sum_{m=1}^{M} H_N(m) \sum_{k=0}^{N-1} \tau_{ab}(k) W(k, \lambda_{j(m),N})$$
$$= \sum_{k=0}^{N-1} \tau_{ab}(k) W(k, \lambda) \sum_{m=1}^{M} H_N(m) W(k, \lambda \oplus \lambda_{j(m),N}).$$

Using the fact that $1 = \{\Sigma_{m=1}^{M} H_N(m)\}^2 \leqslant h_{M,N}^2 M$, we obtain

$$(h_{M,N}^{-2})^{1/2} |E\{\hat{f}_{ab}(\lambda)\} - f_{ab}(\lambda)|$$
$$\leqslant M^{1/2} \sum_{k=0}^{N-1} |\tau_{ab}(k)| \sum_{m=1}^{M} H_N(m)\{1 - W(k, \lambda \oplus \lambda_{j(m),N})\} + M^{1/2} \sum_{k=N}^{\infty} |\tau_{ab}(k)|$$
$$\equiv S_1 + S_2.$$

By Assumption 2, $\Sigma_{k=N}^{\infty} |\tau_{ab}(k)| \leqslant cN^{-1}$ for some positive constant $c$ (see Kohn, 1980b, Lemma 3) so that $S_2 \leqslant cM^{1/2} N^{-1} \to 0$ as $N \to \infty$. Also, since $H_N(m) = O(M^{-1})$ there is a positive constant $c$ such that

$$S_1 \leqslant cM^{-1/2} \sum_{k=0}^{\infty} |\tau_{ab}(k)| \sum_{m=1}^{M} \{1 - W(k, \lambda \oplus \lambda_{j(m),N})\}.$$

However, by the definition of $\{\lambda_{j(m),N}\}$,

$$\lim_{N \to \infty} \sum_{m=1}^{M} \{1 - W(k, \lambda \oplus \lambda_{(j(m)),N})\} < \infty$$

independent of $k$ (by the root test for example), and hence $S_1 \to 0$ as $N \to \infty$. ■

Next define the Walsh–Fourier coherency between the component series $X_a(t)$ and $X_b(t)$ as

$$\mathcal{K}_{ab}(\lambda) = \frac{f_{ab}(\lambda)}{\{f_{aa}(\lambda) f_{bb}(\lambda)\}^{1/2}}$$

provided that the denominator is not zero; note that $-1 \leqslant \mathcal{K}_{ab}(\lambda) \leqslant 1$. A consistent estimate of $\mathcal{K}_{ab}(\lambda)$ based on the smoothed periodogram defined in (2.2) is now established.

THEOREM 5. *Let Assumptions 1 and 2 be satisfied and let $\hat{f}_{M,N}(\lambda) = \{\hat{f}_{ab}(\lambda)\}$ be the estimate given in (2.2). Then for $a, b = 1, \ldots, r$, such that $a \neq b$, the random vector $(\hat{f}_{aa}(\lambda), \hat{f}_{bb}(\lambda), \hat{f}_{ab}(\lambda))'$ is asymptotically multivariate normal with mean $(f_{aa}(\lambda), f_{bb}(\lambda), f_{ab}(\lambda))'$ and covariance matrix $h_{M,N}^2 \mathbf{\Delta}$ where $h_{M,N}^2$ is defined in Theorem 3 and*

$$\mathbf{\Delta} = \begin{bmatrix} 2f_{aa}^2(\lambda) & 2f_{ab}^2(\lambda) & 2f_{aa}(\lambda)f_{ab}(\lambda) \\ & 2f_{bb}^2(\lambda) & 2f_{bb}(\lambda)f_{ab}(\lambda) \\ \text{symmetric} & & f_{aa}(\lambda)f_{bb}(\lambda) + f_{ab}^2(\lambda) \end{bmatrix}.$$

The proof of Theorem 5 is based on the results of Theorems 3 and 4 and is similar to the proof given by Brillinger (1975, Theroem 7.4.4)—the details are omitted. The asymptotic covariance matrix $\mathbf{\Delta}$ is evaluated via Theorem 3. The reader is also referred to Kohn (1980b, Theorem 3).

An estimate of the Walsh–Fourier coherency $\mathcal{K}_{ab}(\lambda)$ is given by

$$\hat{\mathcal{K}}_{ab}(\lambda) = \frac{\hat{f}_{ab}(\lambda)}{\{\hat{f}_{aa}(\lambda)\hat{f}_{bb}(\lambda)\}^{1/2}} \tag{2.3}$$

provided that the denominator is not zero, where $\hat{\boldsymbol{f}}_{M,N}(\lambda) = \{\hat{f}_{ab}(\lambda)\}$ is the estimate given in (2.2). The next theorem establishes the asymptotic distribution of (2.3).

THEOREM 6. *Let the conditions of Theorem 5 hold. Then for* $\mathcal{K}^2_{ab}(\lambda) > 0$, $\hat{\mathcal{K}}_{ab}(\lambda)$ *is asymptotically normal with mean* $\mathcal{K}_{ab}(\lambda)$ *and variance* $h^2_{M,N}\{1 - \mathcal{K}^2_{ab}(\lambda)\}^2$, *where* $h^2_{M,N}$ *is given in Theorem 3.*

PROOF. Let $g(y_1, y_2, y_3) = y_3/(y_1 y_2)^{1/2}$ for $y_1,\ y_2 > 0$. Fix $0 < \lambda < 1$ and henceforth omit it from the notation. For $a \neq b = 1, \ldots, r$, put $\hat{\boldsymbol{f}} = (\hat{f}_{aa}, \hat{f}_{bb}, \hat{f}_{ab})$ and $\boldsymbol{f} = (f_{aa}, f_{bb}, f_{ab})$ so that $\hat{\mathcal{K}}_{ab} = g(\hat{\boldsymbol{f}})$ and $\mathcal{K}_{ab} = g(\boldsymbol{f})$. Define the row vector $\boldsymbol{D}$ as

$$\boldsymbol{D} = \left[\frac{\partial g(\boldsymbol{f})}{\partial f_{aa}}, \frac{\partial g(\boldsymbol{f})}{\partial f_{bb}}, \frac{\partial g(\boldsymbol{f})}{\partial f_{ab}}\right].$$

Then by Theorem 5 and Brockwell and Davis (1987, Proposition 6.4.3), $g(\hat{\boldsymbol{f}})$ is asymptotically normal with mean $g(\boldsymbol{f})$ and variance $h^2_{M,N}\boldsymbol{D}\mathbf{\Delta}\boldsymbol{D}'$ where $\mathbf{\Delta}$ is given in Theorem 3. The results of the theorem now follow. ∎

Theorem 6 admits an approximate confidence interval estimate for $\mathcal{K}_{ab}(\lambda)$; however, the width of the interval will vary with $\lambda$. This problem can be alleviated by applying a variance-stabilizing transformation. As in the trigonometric case, the appropriate transformation is $\tanh^{-1}$, in which case $\tanh^{-1}\{\hat{\mathcal{K}}_{ab}(\lambda)\}$ is asymptotically normal with mean $\tanh^{-1}\{\mathcal{K}_{ab}(\lambda)\}$ and variance $h^2_{M,N}$.

Although it is possible to define a Walsh–Fourier amplitude spectrum and phase spectrum, there is no natural interpretation of such parameters. For example, a Walsh–Fourier phase spectrum can be defined using the notion of 'cosine' and 'sine' Walsh functions. Note that the even-numbered Walsh functions in sequency order $\{W(j, \lambda),\ j = 2k,\ k = 0, 1, 2, \ldots\}$ are even functions about $\lambda = \frac{1}{2}$ and the odd-numbered Walsh functions in sequency order $\{W(j, \lambda),\ j = 2k - 1,\ k = 1, 2, \ldots\}$ are odd functions about $\lambda = \frac{1}{2}$. For example, using the discrete Walsh functions displayed in Section 1 in the Hadamard matrix $\boldsymbol{H}_{\mathrm{W}}(3)$ with $N = 8$, we can see that $W(n, m/8)$ is an even function about the midpoint ($\lambda = 4/8$) when $n = 0, 2, 4, 6$ and is an odd function about the midpoint when $n = 1, 3, 5, 7$, as $\lambda$ varies, $0 \leqslant \lambda < 1$. Hence

$$\mathrm{cal}(k, \lambda) = W(2k, \lambda) \qquad (k = 0, 1, 2, \ldots),$$
$$\mathrm{sal}(k, \lambda) = W(2k-1, \lambda) \qquad (k = 1, 2, \ldots)$$

are called the cal (cosine Walsh) and sal (sine Walsh) functions respectively by analogy with the even cosine functions and the odd sine functions (for further discussions of the cal and sal functions, refer to Ahmed and Rao (1975) or Beauchamp (1975, 1984)). The Walsh–Fourier spectral matrix can then be decomposed as $f(\lambda) = f_c(\lambda) + f_s(\lambda)$ where $f_c(\lambda) = \Sigma_{k=0}^{\infty}\tau(k)\,\mathrm{cal}(k, \lambda)$ and $f_s(\lambda) = \Sigma_{k=1}^{\infty}\tau(k)\,\mathrm{sal}(k, \lambda)$. However, unlike the sine and cosine functions, the sal and cal functions are not shifted versions of each other (again refer to the Hadamard matrix in Section 1 for an example). Thus, while it is possible to give an analogy of the trigonometric case and define a Walsh–Fourier phase spectrum, say $\varphi_{ab}(\lambda) = \tan^{-1}\{f_{s,ab}(\lambda)/f_{c,ab}(\lambda)\}$, between the component series $X_a(t)$ and $X_b(t)$, $\varphi_{ab}(\lambda)$ will not have a constant slope of $\delta$ if $X_b(t) = X_a(t-\delta)$ for example.

This type of problem is also inherited by coherency $\mathcal{K}_{ab}(\lambda)$. That is, since the Walsh–Fourier transform is not invariant to cyclic shifts of the data (see the discussion in Ahmed and Rao (1975, p. 115) for example), it will not be the case that $\mathcal{K}_{ab}^2(\lambda) \equiv 1$ if $X_b(t) = \Sigma_{j=0}^{\infty}\psi(j)X_a(t-j)$, where $\Sigma|\psi(j)| < \infty$, unless $\psi(j) = 0$ for $j \geq 1$. It would, however, be the case that $\mathcal{K}_{ab}^2(\lambda) \equiv 1$ if $X_b(t) = \Sigma_{j=0}^{\infty}\psi(j)X_a(t \oplus j)$, although such filters have no realistic interpretation; see Morettin (1981) for a discussion of dyadic filters.

## 3. EXAMPLE

In this section the techniques of Section 2 are applied to neonatal sleep–state data collected in a study of the effects of moderate maternal alcohol consumption on neonatal electroencephalographic (EEG) sleep patterns; this study is part of a larger study of the effects of maternal substance use during pregnancy. A detailed description of the study design, the methods used for measuring alcohol use and the scoring of the neonatal EEG sleep records can be found in Day *et al.* (1985) and Scher *et al.* (1988). Briefly, an EEG sleep recording of duration approximately 2 h is obtained on a full-term infant 24–36 h after birth, and recordings are scored (by an electroencephalographer who is not aware of the prenatal substance exposure of the infant) for EEG sleep state, rapid eye movements (REMs), arousals and body movements using scoring epochs of 1 min. Sleep state is categorized (per minute) into one of six possible states: (1) quiet sleep—trace alternant; (2) quiet sleep—high voltage; (3) indeterminate sleep; (4) active sleep—low voltage; (5) active sleep—mixed; (6) awake.

First we compare the bivariate Walsh–Fourier spectra of the EEG sleep state and total number of body movements (discounting sucking) per minute for an infant whose mother abstained from using alcohol during pregnancy (ID 465) and for an infant whose mother used alcohol regularly at a rate of

0.5 drinks per day throughout pregnancy (ID 223). Finally, the results of this analysis are used to examine the coherency between sleep state and body movements for unexposed versus exposed groups of infants.

Figures 1 and 2 show the sleep states (using the preceding state labels 1–6) and total number of body movements for the unexposed infant (ID 465) and for the exposed infant (ID 223) respectively for 120 min of sleep. Spectral analysis was accomplished using the fast Walsh–Fourier transform given in Ahmed and Rao (1975) by padding each record to 128 time points. Spectral estimation was based on the estimate given in (2.2) with $M = 5$, centering on the sequency of interest and using the following weights: 0.1 {2, 1, 4, 1, 2}. The significance of this type of filter reveals a major difference between the Walsh–Fourier approach and the Fourier approach as well as the flexibility of an estimator of the form of (2.2). It is clear that power associated with odd (even) sequencies correspond to sleep cycles that are odd (even) functions about the middle of the sleep record. Thus the aforementioned filter yields an estimate that will emphasize the cal and sal components of the spectrum rather than simply smooth the periodogram. It is important to the analysis to be able to distinguish between the odd and even cycles around the 1 h time point since normal sleep is expected to cycle according to odd cycles of about 60 min (i.e. the middle of the sleep record).

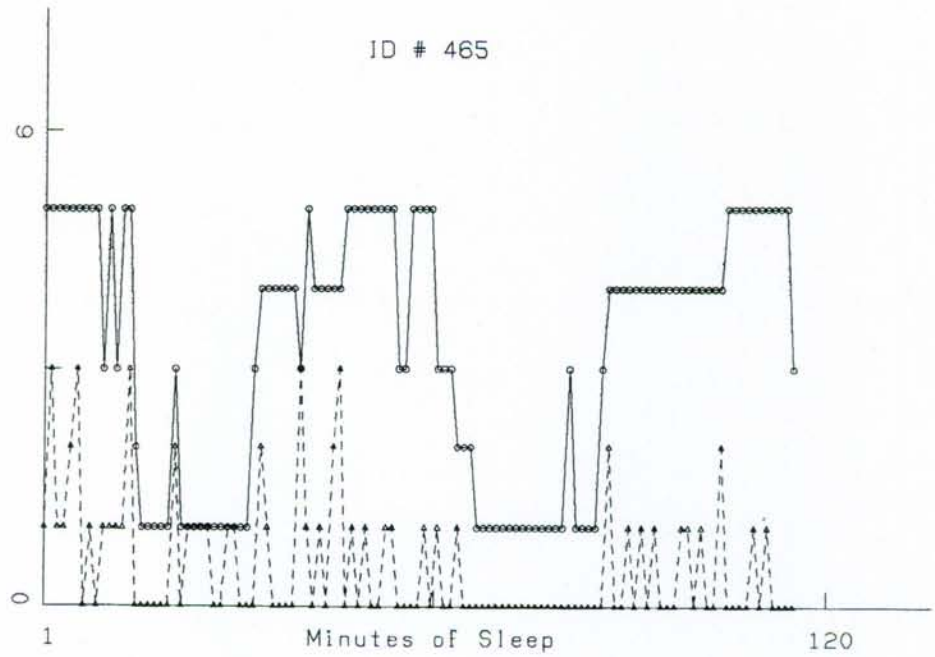


FIGURE 1. Sleep state (——) and total number of body movements (– – –) of the unexposed infant.

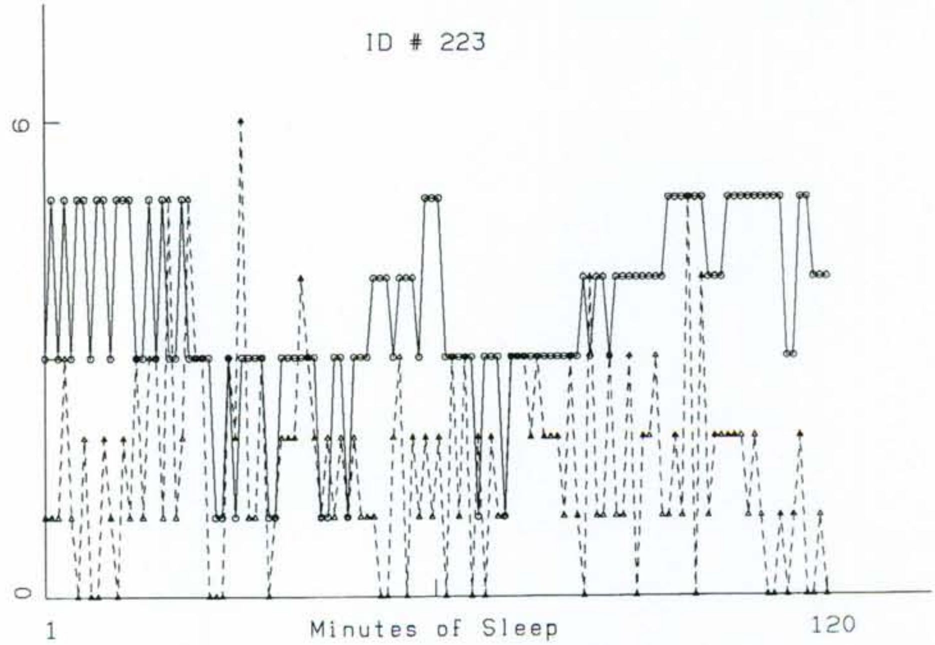


FIGURE 2. Sleep state (——) and total number of body movements (---) of the exposed infant.

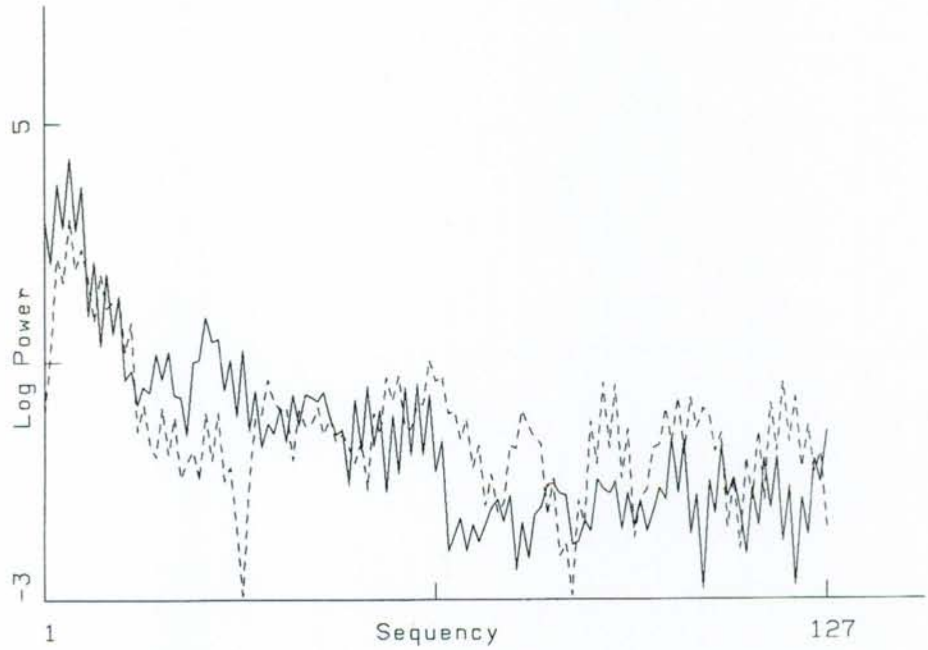


FIGURE 3. Estimated sleep-state spectrum for the unexposed infant (——) and for the exposed infant (---).

The estimated log spectra of sleep state for the unexposed and the exposed infant are compared in Figure 3. There are many similarities between the two spectra; however, there is typically more power at the slower sequencies for the unexposed infant and more power at the higher sequencies for the exposed infant. The peak power for both infants is at the odd sequencies of 3, 5 and 7 (per 128 min) which correspond to periods of roughly 45, 25 and 18 mins respectively. Note that in the Walsh–Fourier domain a peak period of $p$ min means 'one switch every $p$ min' as opposed to the usual trigonometric definition of 'one cycle every $p$ min'.

The log spectra of body movements for each infant are compared in Figure 4, and there are considerable differences here. The spectrum of the unexposed infant has its peak power at the lower sequencies with some additional power at the middle sequencies. There is considerable more power throughout for the exposed infant than for the unexposed infant, and these peak periods are spread over a wide range of sequencies, These results are expected from the sleep records shown in Figures 1 and 2 where it is noted that the number of body movements is much more variable throughout the record of the exposed infant than throughout that of the unexposed infant.

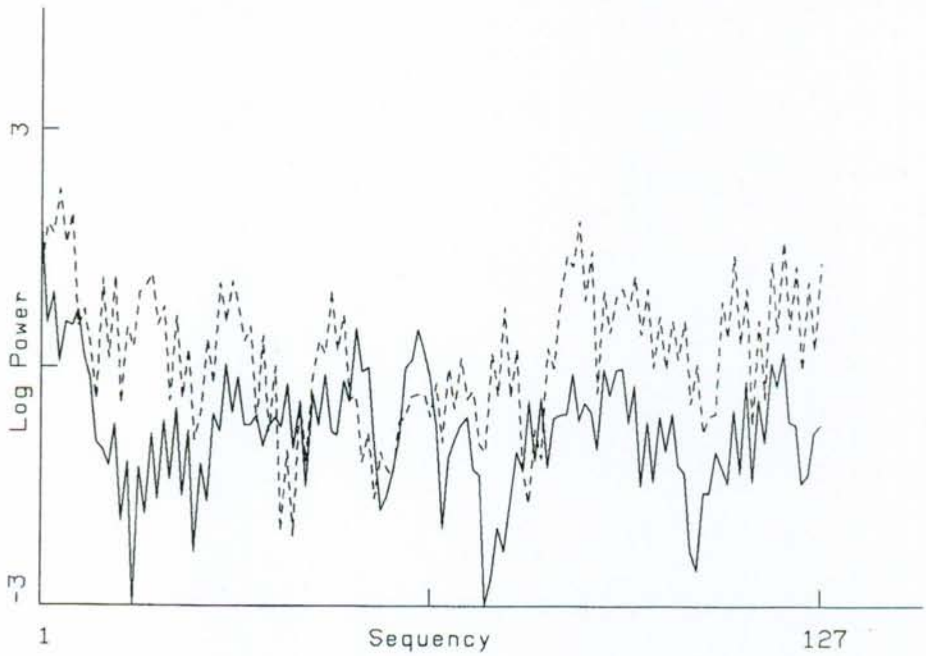


FIGURE 4. Estimated spectrum of the total number of body movements for the unexposed infant (——) and the exposed infant (– – –).

Figure 5 shows the estimated cross-spectrum between sleep state and body movements for each infant. This sort of plot will be new to investigators of spectra, and it reveals another major difference between this approach and the trigonometric approach, i.e. the Walsh–Fourier cross-spectrum is real whereas the Fourier cross-spectrum is complex. In the Fourier domain we typically decompose the complex-valued cross-spectrum into polar coordinates which results in an amplitude (or gain) spectrum and a phase spectrum, both real. In the sequency domain, the cross-spectrum is analyzed directly. Comparison of the unexposed and the exposed infant via figure 5 shows that there is typically more power associated with the unexposed infant than with the exposed infant. The peak periods for the unexposed infant correspond to the sequencies 1, 3, 5 and 7 (per 128 min), the very slowest odd Walsh signals, whereas the peak periods for the exposed infant corespond to the sequencies 4, 6, 8 and 13 (per 128 min). It is of interest to note the differences between the two cross-spectra at the very fastest periods; this difference becomes more noticeable when the cross-spectra are normalized and the coherence is computed.

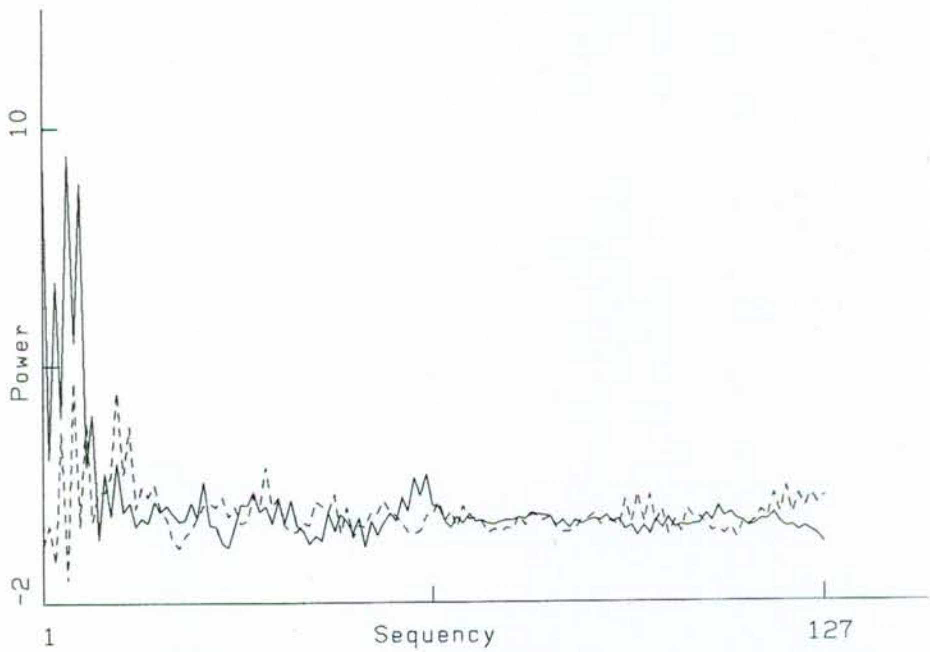


FIGURE 5. Estimated cross-spectrum between sleep state and body movements for the unexposed infant (——) and the exposed infant (---).

The coherencies of the unexposed infant and the exposed infant are compared in Figure 6. Since the Walsh–Fourier cross-spectrum is real, unlike the Fourier coherency (which is absolute correlation relative to frequency), the Walsh–Fourier coherency can be negative; this is a considerable advantage over the trigonometric case, especially in this example as will be seen. It is evident from Figure 6 that there are small isolated ranges of sequencies at which the coherency between the sleep-state signal and the body-movement signal for the unexposed infant and the exposed infant differ markedly in sign, most notably at the very fast end of the sequency range (120–127). Here, the coherency between the two signals is negative for the unexposed infant and positive for the exposed infant. It is believed that this difference will be an aid in the identification of a disturbance in the sleep cycle. This distinction might have been missed in a trignometric analysis.

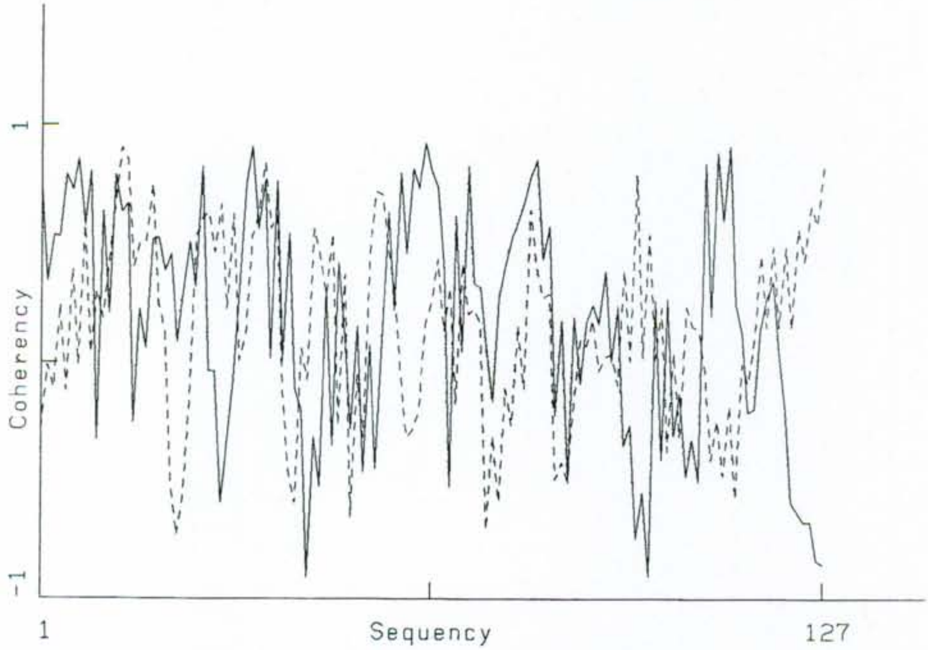


FIGURE 6. Coherency between sleep state and body movements for the unexposed infant (——) and the exposed infant (– – –).

Finally, the average coherency of a group of 12 infants whose mothers abstained from using alcohol during pregnancy is compared with another group of 12 infants whose mothers used alcohol on a regular basis throughout pregnancy. These are the same infants as those analyzed by Stoffer *et al.* (1988). While there were differences between the average coherency of the two groups at various isolated ranges of sequencies (that were consistent with the differences between ID 465 and ID 223 previously mentioned), differences in the sequency range of 120–127 (per 128 min) will be illustrated. The average coherency with its standard error for each group of 12 infants and the normalized difference between the groups at sequencies 120–127 are listed in Table I. While the difference between the exposed group and the unexposed group do not remain as marked as the individual analysis, a trend prevails. It appears that, for the unexposed group, the sleep-state signal and the body-movement signal have zero coherency on the average at the fast sequencies (although there is a tendency to be on the negative side of zero as in the case of ID 465), while the average coherency between the signals for the exposed group is positive (which is consistent with the individual analysis of ID 223).

TABLE I

COMPARISON OF THE AVERAGE COHERENCY BETWEEN SLEEP STATE AND NUMBER OF BODY MOVEMENTS

| *Sequency* | *Exposed group*[a] (mean (se)) | *Unexposed group*[b] (mean (se)) | *Normalized difference* |
|---|---|---|---|
| 120 | 0.07 (0.14) | −0.10 (0.13) | 0.91 |
| 121 | 0.21 (0.13) | −0.04 (0.11) | 1.46 |
| 122 | 0.17 (0.13) | −0.08 (0.17) | 1.16 |
| 123 | 0.20 (0.14) | 0.01 (0.12) | 1.08 |
| 124 | 0.30 (0.10) | −0.03 (0.15) | 1.94[c] |
| 125 | 0.29 (0.10) | −0.00 (0.12) | 1.91[c] |
| 126 | 0.23 (0.14) | 0.00 (0.17) | 0.99 |
| 127 | 0.29 (0.12) | −0.06 (0.16) | 1.77[c] |

[a] Group of 12 infants whose mothers used alcohol regularly during pregnancy.
[b] Group of 12 infants whose mothers abstained from alcohol during pregnancy.
[c] Exceeds 0.05-level critical value for a one-sided $t$-test.

ACKNOWLEDGEMENTS

The author would like to thank Professor Nancy L. Day and Professor Mark S. Scher, School of Medicine, University of Pittsburgh, for providing the data set used in the example. The data were collected as part of a study on the effect of maternal substance use during pregnancy and was funded in part by National Institute on Alcohol Abuse and Alcoholism Grant AA06390 and National Institute on Drug Abuse Grant DA03209. The author is also grateful to a referee for comment and suggestions which improved the presentation.

REFERENCES

AHMED, N and RAO, K. R. (1975) *Orthogonal Transforms for Digital Signal Processing*. New York: Springer-Verlag.

BEAUCHAMP, K. G. (1975) *Walsh-Functions and Their Applications*. London: Academic Press.

—— (1984) *Applications of Walsh and Related Functions*. London: Academic Press.

BRILLINGER, D. R. (1975) *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

BROCKWELL, P. J. and DAVIS, R. A. (1987) *Time Series: Theory and Methods*, New York: Springer-Verlag.

DAY, N. L., WAGENER, D. K. and TAYLOR, P. M. (1985) Measurement of substance use during pregnancy: methodologic issues. In *Consequences of Maternal Drug Abuse* (ed. T. M. Pinkert), NIDA Research Monograph 59.

HANNAN, E. J. (1970) *Multiple Time Series*. New York: Wiley.

KOHN, R. (1980a) On the spectral decomposition of stationary time series using Walsh functions I. *Adv. App. Prob.* 12, 183–99,

—— (1980b) On the spectral decomposition of stationary time series using Walsh functions II. *Adv. Appl. Prob.* 12, 462–74.

MORETTIN, P. A. (1974) Limit theorems for stationary and dyadic-stationary processes. *Bull. Braz. Math. Soc.* 5, 97–104.

—— (1981) Walsh spectral analysis. *SIAM Rev.* 23, 279–91.

—— (1983) A note on a central limit theorem for stationary processes. *J. Time Ser. Anal.* 4, 49–52.

OTT. J. and KRONMAL, R. A. (1976) Some classification procedures for multivariate binary data using orthogonal functions. *J. Am. Statis. Assoc.* 71, 391–99.

ROBINSON, G. S. (1972) Discrete Walsh and Fourier power spectra. *Proceedings of the Symposium on the Applications of Walsh Functions, Washington, DC*, pp. 298–309.

SCHER, M. S., RICHARDSON, G. A., COBLE, P. A., DAY, N. L. and STOFFER, D. S. (1988) The effects of prenatal alcohol and marijuana exposure: disturbances in neonatal sleep cycling and arousal. *Pediat. Res.* 24, 101–5.

STOFFER, D. S. (1985) Central limit thoerems for finite Walsh–Fourier transforms of weakly stationary time series. *J. Time Series Anal.* 6, 261–67.

—— (1987) Walsh–Fourier analysis of discrete-valued time series. *J. Time Series Anal.* 8, 449–67.

——, SCHER, M. S., RICHARDSON, G. A., DAY, N. L. and COBLE, P. A. (1988) A Walsh–Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *J. Am. Statist. Assoc.* 83, 954–63.