JOINT TIME AND TIME - FREQUENCY OPTIMAL DETECTION OF

K-COMPLEXES IN SLEEP EEG

Cédric RICHARD and Régis LENGELLE

Laboratoire LM2S, Université de Technologie de Troyes

12 rue Marie Curie - BP 2060

F 10010 TROYES Cedex

Correspondence should be sent to Cédric RICHARD   tel: +33.3.25.71.76.00 (57.82)

fax: +33.3.25.71.56.99

email: cedric.richard@univ-troyes.fr

Abstract

Automated detection of waveforms such as delta and K-complexes in the EEG is an important component of sleep stage monitoring. The K-complex is a key feature that contributes to sleep stages assessment. However, its automated detection is still difficult due to the stochastic nature of the EEG. In this paper, we propose a detection structure which can be interpreted as joint linear filtering operations in the time and time-frequency domains. We also introduce a method of obtaining the optimum detector from training data, and we show that the resulting receiver offers better performances than the one obtained via the Fisher criterion maximization. The efficiency of this approach for K-complexes detector design is explored. It results from this study that the obtained receiver is potentially the best one which can be found in the literature. Finally, it is emphasized that this methodology can advantageously be used to solve many other detection problems.

I. INTRODUCTION

In the last 50 years, polygraphic sleep techniques have permitted to greatly improve human sleep understanding (1). Unfortunately, the visual scoring of whole nights EEG, EOG and EMG records is still time-consuming and expensive, playing the role of a limiting factor in sleep studies development. This justifies the growing interest for automated sleep data analyzers. In the past two decades, significant advances have been made in this domain, starting with hybrid systems as in (2). Recently, in addition to the latter heuristic approaches, a large number of algorithms based on statistical pattern recognition techniques (3) and some expert systems have been proposed (4).

Automated detection of waveforms such as alpha, delta and K-complex waves in the EEG is an important component of sleep stage monitoring. The K-complex is one of the key features that contributes to sleep stages assessment. It can occur both spontaneously or as an evoked response to an auditory stimuli, and seems to be associated with arousal during sleep. This transient EEG pattern has a total duration of between 500 and 1500 ms, and is characterized by a sharp upward wave followed by a downward one, orientating the y-axis from top to bottom as in electrophysiologists' convention. Its amplitude is 3 times background activity, and is generally larger than 75 mV (5). However, due to the highly non-stationary nature of the superimposed EEG background activity, the K-complex morphology can vary drastically, and no complete description is available (6). The automated detection of K-complexes is a challenging proposition which has been the purpose of numerous published efforts. In (7) for example, Da Rosa *et al*. propose a detector based on a K-complex model. The resulting good detection and false alarm rates are respectively 89% and 49%. A multi-layer neural net is involved in (8): the integrated responses of two band-pass filters are used as inputs. This approach yields a detection rate ranging from 55 to 67%. In (6), Bankman *et al*. also use a neural net with 14 features extracted from EEG as inputs. Around 90% true positives are obtained for a 8% false positives rate. Finally, Tang and Ishii successfully introduce a new approach based on the discrete wavelet transform in (9), obtaining respectively 87% and 10% true and false positives. However, they consider that the K-complex is always overridden by a spindle whereas Jansen et *al*. (5) stated that there was no fixed relationship between these

two events. This emphasized that detector performances proposed in the literature strongly depends on the training and test sets composition.

Time-frequency (TF) representations have been extensively used for detection in applications ranging from radar to machine fault diagnostics, due to the need for dealing with non-stationary signals. These representations are parameterised in terms of time and frequency, and describe the non-stationary signal characteristics via their time-varying spectral content. Most of the TF-based detection schemes which have been proposed are based on Moyal's relationship (10), and implement classical optimum receptors equivalently in the TF domain (11). Recently, Sayeed and Jones identified several scenarios in which TF-based detectors are both optimum and fully exploit the many degrees of freedom available in TF representations (12). It is stated that invertible bilinear TF representations are only promising for situations in which the optimum detector is a quadratic function of the observations. This justifies for example the TF matched filter sub-optimality for known deterministic signal detection in the presence of white Gaussian noise. Finally, it is noteworthy that all these approaches require substantial knowledge on signals whereas phenomena are complex and poorly understood in many applications. However, the collection of substantial amounts of labeled data is often feasible. In (13), Jones and Sayeed adopt this approach, and derive near-optimum TF-based detectors directly from training data by maximizing the Fisher linear criterion. It should be noticed that Fisher discriminants can be arbitrarily bad: there are distributions such that even though the two classes are linearly separable, the Fisher linear discriminant has a probability of error close to one (14). In (15), Richard and Lengellé develop an original method of obtaining optimum TF-based detectors (probability of error minimum) from training data sets, and show the sub-optimality of the approach proposed by Jones and Sayeed (13).

In this paper, we introduce a new detection structure which can be interpreted as joint linear filtering operations in the time and TF domains. We also identify a scenario for which this statistic naturally appears. Then, we extend the method of obtaining optimum TF-based detectors from training data introduced in (15) to the studied detection structure. This original approach is used to define a new detector of K-complexes. The paper is organized as follows. In Section II, we start with a brief review of Cohen's class TF distributions, followed by a description of the usual time and TF-based approaches to

decision problems. Then, we introduce the new concept of joint time and time-frequency (T-TF) based detection, and we identify a scenario for which the considered statistic is optimal. In Section III, we expose our method of obtaining optimum linear detector operating in the T-TF domain, and we show the sub-optimality of any receiver derived from the Fisher criterion maximization. A simple example illustrates the efficiency of our approach. In Section IV, our procedure is applied to a set of EEG events (K-complexes and delta waves), and the performances of the obtained T-TF-based receiver are evaluated. Finally, we present some conclusions regarding our approach in Section V.

II. TIME-FREQUENCY-BASED APPROACH TO DECISION PROBLEMS

1. Cohen's class time-frequency representations review

i) Definition: The Wigner-Ville distribution, which has been extensively studied in recent years, is defined as follows ([10]):

$$W_x(t,f) = \int_{-\infty}^{+\infty} R_x(t,\tau) \exp(-2j\pi f\tau)\, d\tau,$$ [1]

where $R_x(t,t) = x(t+t/2)\, x^*(t-t/2)$ is the instantaneous autocorrelation function of the signal $x(t)$.

This distribution is well-known for its high resolution in the TF domain and the large number of properties it satisfies ([10]). Unfortunately, its use in practical applications is limited due to the numerous non-negligible cross-components generated by its bilinear structure. This difficulty can be removed by filtering the instantaneous autocorrelation function R_x. This operation leads to the following definition of Cohen's class TF distributions (CTFD) ([10])([16]):

$$C_x(t,f) = \iint_{-\infty} F(t,u) R_x(u,$$ [2]

where F is called the autocorrelation-domain kernel ([17]).

One can also show that members of Cohen's class corresponds to filtered versions of the Wigner-Ville distribution:

$$C_x(t,f) = \iint_{-\infty} \Pi(t-u,$$ [3]

where $\Pi(t,f) = \int_{-\infty}^{+\infty} F(t,\tau) \exp(2j\pi f\tau)\, d\tau$. [4]

Finally, it is noteworthy that Cohen's class includes numerous well-known TF distributions as special cases: as an example, the classical Spectrogram with short-time window w(t) can be obtained from Eq. {3} if the kernel P is chosen as the Wigner-Ville distribution of w(t) (10).

ii) Discrete definition: In practice, sampled data are usually processed, and F has a finite support S_F defined as:

$$S_F = \left\{ (p,m) \in Z : |p| \le P, |m| \le M - 1 \right\},$$ [5]

in which case the following discrete equivalent {6} of the definition {2} can be used. The sampling frequency f_s is normalized to unity.

$$C_x(k, f_i) = 2 \sum_{m=1-M} \sum_{p=-P} F(p,m)$$ [6]

In the expression introduced above, the bidimensional function $R_x(n, m)$ denotes the discrete time instantaneous autocorrelation function of the signal x(k):

$$R_x(n,m) = x(n+m) x^*(n-m).$$ [7]

The discrete CTFDs properties are similar to the continuous time case except for the periodicity in the frequency variable, which period is equal to one-half the sampling frequency. This implies that the sampling frequency f_s must be at least twice the well-known Nyquist rate to avoid aliasing, when the signal x(k) is real (18). This constraint is equivalent to:

$f_s \ge 4B$, where B denotes the bandwidth of the signal x(k). [8]

To remove this difficulty, one can replace the real signal x(k) by its corresponding analytic signal z_x (k), which is defined as follows in the frequency domain (18):

$$Z_x(f) = \begin{cases} 2 X(f), & 0 < f < 1/2 \\ X(f), & f = 0 \\ 0, & -1/2 < f < 0. \end{cases}$$ [9]

Consequently, even if the periodicity is unchanged, the absence of a negative frequency spectrum eliminates the problem of aliasing which otherwise occurs when data is sampled at the Nyquist rate (18). Then, the previous sampling frequency constraint {8} obviously becomes:

$f_s \ge 2B$, where B denotes the bandwidth of the signal x(k).

6

Finally, the use of analytic signals is also important in avoiding artifacts around the frequency origin, due to cross products between negative and positive frequencies. A detailed discussion can be found in (18). For the two reasons mentioned in this section, each real signal studied in this paper will be first transformed into an analytic one using Eq. {9}.

iii) Fast implementation: Evaluating CTFDs by the direct application of Eq. {6} requires the calculation of the smoothed autocorrelation function, and of a 2M-sample FFT at each time-instant. This algorithm is computationally expensive, and consequently cannot easily be used to compute the TF representations of long biomedical signals. In order to reduce computation time, several ideas have been proposed (17-21). The recursive approach presented below, introduced by Amin (20) and adopted by Richard and Lengellé (21) in their fast algorithm dedicated to CTFDs modified by the reassignment method (22), seems to be the most efficient one.

In the case when the distribution kernel F belongs to the family K_{rec} defined as follows:

$$K_{rec} = \left\{ F \, / \, F(p,m) = h(m) \sum_{t=1}^{T} g_t \right\}, \tag{10}$$

it is straightforward to show that the corresponding TF distribution satisfies the following recursive formulation:

$$C_x^F(k+1,f_i) = \sum_{t=1}^{T} C_x^{F_t}(k+1,f_i), \tag{11}$$

where $F_t(p,m) = h(m)\, g_t(p)$, $\beta_t = r_t \exp(j\theta_t)$,

and $C_x^{F_t}(k+1,f_i) = \beta_t C_x^{F_t}(k,f_i) +$. $\tag{12}$

In the above expression, $j_t(k,m)$ is given by:

$$\psi_t(k,m) = \lfloor g_t(t) \rfloor K_x(k+1-t,m) \tag{13}$$

We can notice that K_{rec} is composed of functions with separable (p,m) variables. Then, this algorithm only authorizes the evaluation of smoothed pseudo Wigner-Ville distributions (SPWVD). This constraint on the distribution choice is not restrictive. With its separable kernel, the SPWVD allows the time and frequency smoothing to be adjusted independently, making this representation one of the most versatile of

7

Cohen's class. However, it should be mentioned that SPWVDs do not allow a directional time-frequency smoothing, which is sometimes useful (when analyzing multi-component chirp signals for example). Finally, it is noteworthy that the following truncating windows can advantageously be used since they belong to the class K_{rec}: rectangular (T=1), half-sine (T=2), Hamming (T=3), Hanning (T=3) and Blackman (T=5) windows, considered as functions of the variable p, and post multiplied by any function h of the variable m.

2. Time and time-frequency-based detection frameworks

i) Linear detection in the time domain: The detection problem we consider is as follows. Given a discrete-time (complex) signal $\underline{x}$ received over the interval $\{0, ..., K\}$, where $\underline{x} = [x(0) ... x(K)]^T$, one must decide between the two competing hypotheses H_0 and H_1:

$$\begin{cases} H_0: & x(k) = n(k) \\ H_1: & x(k) = s(k) + n(k) \end{cases}, k \text{ Î } \{0, ..., K\}, \tag{14}$$

where $\underline{s}$ is the underlying discrete-time (complex) signal to be detected, and $\underline{n}$ some additive (complex) noise.

The decision between H_0 and H_1 is often made by comparing a statistic $l(\underline{x})$, computed from the observation, to some preset threshold (23). As an example, when $\underline{s}$ is a known deterministic signal, and $\underline{n}$ some white Gaussian noise with a known variance s^2, the following statistic obtained via the Neyman-Pearson criterion has been shown to be the optimal solution of {14}:

$$\lambda_L(\underline{x}) = \sum_{k=0}^{K} s(k) x(k) = \underline{s}^T \underline{x}. \tag{15}$$

This detection structure, which is a *linear transformation* of the samples $x(k)$, is called a matched filter. Note that {15} is often used when the statistical properties of the observations differ from those considered above, and can be obtained by maximizing the output signal to noise ratio of an imposed linear detector.

ii) Linear detection in the time-frequency domain: In the perspective of a TF-based detection scheme in the Cohen's class, the binary hypothesis testing problem described above can be rewritten in time-frequency terms (24).

$$\begin{cases} H_0: & C_x^F(k,f_i) = C_n^F(k,f_i) \\ H_1: & C_x^F(k,f_i) = C_{s+n}^F(k,f_i) \end{cases}, \ k \ \hat{I} \ \{0, ..., K\}, \ i \ \hat{I} \ \{0, ..., 2(M-1)\}. \tag{16}$$

The signal $\underline{x}$ is supposed to be known over the interval $\{0, ..., K\}$, and equal to zero elsewhere. The bidimensional function C_x^F denotes a discrete Cohen's class TF distribution.

By analogy with the classical matched filter theory, one can consider the general class of linear TF-based detectors ([24]):

$$\lambda_Q(\underline{x}) = \sum_{k=0}^{K} \sum_{i=0}^{2(M-1)} g(k,f_i) \, C_x^\Gamma(k,f_i), \tag{17}$$

where $\underline{g}$ is a TF-based filter response to be determined using the *a priori* known characteristics of the signal $\underline{s}$. Such a determination can be achieved by maximizing the Fisher criterion ([12]), or the signal to noise ratio between the two competing hypotheses ([25]), when *a priori* probabilities and probability densities of C_x^F (conditionally to H_0 and H_1) are unknown. In ([15]), an optimal approach is also proposed.

The statistic {17} quantifies the similarity between a TF structure of reference $\underline{g}$ and the TF representation C_x^F of the received signal $\underline{x}$, which can be considered from Eq. {3} as a filtered version of the Wigner-Ville distribution W_x. In an equivalent way, one can smooth the reference $\underline{g}$ rather than the observation W_x. This operation leads to the following dual definition of the statistic $l_Q(\underline{x})$:

$$\lambda_Q(\underline{x}) = \sum_{k=0}^{K} \sum_{i=0}^{2(M-1)} v(k,f_i) \, W_x(k,f_i), \tag{18}$$

where $\quad v(k, f_i) = \sum_{k'=0} \sum_{i'=0} g(k \tag{19}$

and $\quad H(k, f_i) = \dfrac{}{2M-1} \sum_{m=0}. \tag{20}$

From these definitions, Flandrin has described two roles played by the distribution kernel F in the statistic $l_Q(\underline{x})$ ([11]). On the one hand, from Eq. {18}-{20}, the smoothing function F authorizes to handle *a priori* uncertainties on the reference $\underline{g}$ by broadening it on the TF plane. On the other hand, from Eq. {17}, it also plays the role of an *a posteriori* smoothing of the observation W_x, which is necessary if one wants to jointly estimate the TF structure of the signal $\underline{x}$ ([11]). This interpretation of the role of the

function F emphasizes one of the main interests of the TF-based approach to decision problems: by using the same tool, it permits to combine the analysis of signals, conveniently performed in the TF domain, and the decision stage (11).

Finally, the TF-based detection structure defined above allows to generate very different well-known receivers from suitable choices of the reference $\underline{v}$. Two examples are mentioned in Table I. This additional property over standard statistics justifies our interest for the unique formulation {18}. Nevertheless, it should be noticed that this test statistic necessarily yields to a sub-optimum solution of the detection problem {14}, when $\underline{s}$ is a known deterministic signal and $\underline{n}$ some white Gaussian noise of known variance s^2, since W_x is a *"pure" quadratic function* of the observations (without linear terms). Note that an optimum statistic based on the cross Wigner-Ville distribution W_{xs} has been already proposed (26): it is obviously a *linear function* of $\underline{x}$.

3. Joint time and time-frequency based approach to decision problems

i) Problem: The optimum solution of some detection problems combines linear and quadratic statistics l_L and l_Q. As an example, consider the following discrete-time binary hypothesis testing problem:

$$\begin{cases} H_0: & x(k) = s_0(k) + n_0(k) \\ H_1: & x(k) = s_1(k) + n_1(k) \end{cases}, k \in \{0, ..., K\}, \qquad [21]$$

where $\underline{n}_0$ (res. $\underline{n}_1$) is a zero-mean white Gaussian noise, with known covariance matrix S_0 (res. S_1). The deterministic signals $\underline{s}_0$ and $\underline{s}_1$ are supposed to be known.

ii) Resolution and interpretation: It is straightforward to show that the optimal statistic corresponding to {21}, in the sense of the Neyman-Pearson criterion, is given by the following log-likelihood ratio (23):

$$= \Lambda_Q(\underline{X}) + \Lambda_L(\underline{X} \qquad [22]$$

which must be compared to a threshold n_0.

The statistic $l(\underline{x})$ consists in a linear combination of a *"pure" quadratic function* $l_Q(\underline{x})$ of the observations, and a *linear* one $l_L(\underline{x})$. Consequently, as it has been introduced in § (II.2), the detection problem {21} can be solved by joint linear filtering operations in the time and time-frequency domains:

10

$$\lambda_{T-TF}(\underline{\Delta}) = \sum_{k=0} \sum_{i=0} g(k, t_i) \, W_x(k, t_i) \qquad [23]$$

As a conclusion, the approach proposed above permits to combine the analysis of signals, conveniently performed in the T-TF domain, and the decision stage of the problem {21}. In Section III, we propose a method to design the optimum linear T-TF-based detector {23}, i.e. which minimizes the error probability, via the optimization of the references g and $\underline{h}$ from training data.

III. OPTIMUM DETECTOR DESIGN FROM TRAINING DATA

1. Procedure for linear detectors design

i) Problem formulation and resolution: Linear receivers are optimum, in the sense of the Neyman-Pearson criterion, for Gaussian distributions with equal covariance matrices, conditionally to the hypotheses H_0 and H_1. However, even if these assumptions on probability density functions are not reasonable in many applications, the simplicity and robustness of this approach often compensate the loss in performance. In this way, we discuss now how to design optimum linear detectors from training data, regardless of the signals distributions. This method has been introduced by Fukunaga to design linear discriminants in the context of pattern recognition (27), and used by Richard and Lengellé to automatically design optimum TF-based detectors from training data (15).

Using the expression of the statistic introduced in Eq. {23}, the detection problem {21} can be rewritten as follows:

$$\begin{cases} \text{if } \lambda_{T-TF}(\underline{Y}) = \underline{V}^T \underline{Y} \geq v_0, \text{ then } H_1 \\ \text{else } H_0 \end{cases}, \qquad [24]$$

where

$$\underline{V} = \lfloor g(0, t_0) \cdots g(0, t_{2(M-1)}) \, g(1, t_0) \cdots g($$

and n_0 a threshold to be determined.

Our design work consists in finding the optimum vector $\underline{V}$ and the optimum threshold value n_0 in the sense of a pre-selected criterion, and for a given data set. Using a minimal *a priori* knowledge, the statistic l_{T-TF} can be characterized by the following expected values and variances:

11

$$\eta_i = \ldots \{ \lambda_{T-TF}(\underline{1})|H_i \} = \underline{V} \cdot \underline{M}_i + \nu_0,$$

where $\underline{M}_i = \ldots \{\underline{1}|H_i\}, \quad \Sigma_i = \ldots \{$

Let $f(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2)$ be any measure (depending only on these parameters) of H_0 and H_1 separability in the l_{T-TF} - space. This criterion must be maximized so that the following derivatives of f, with respect to V and n_0, are equated to zero:

$$\left\{ \frac{\partial f}{\partial \nu_0} = \frac{\partial f}{\partial \sigma_1^2} \frac{\partial \sigma_1^2}{\partial \nu_0} + \frac{\partial f}{\partial \sigma_2^2} \frac{\partial \sigma_2^2}{\partial \nu_0} \right.$$

[25]

The resolution of {25} provides a particularly interesting analytical solution for the detector {23} design since $\underline{V}_{opt}$ satisfies:

$$\left[\alpha \Sigma_0 + (1 - \alpha) \Sigma_1 \right] \underline{V}_{opt} = (\underline{M}_1 - \underline{M}_0$$

[26]

.

Thus, the optimum $\underline{V}_{opt}$ has the form {26} regardless of the selection of f: the criterion effect only appears in the parameter a ($0 \pounds a \pounds 1$). In our case, we choose the a value which minimizes an estimation of the error probability. This operation, which is described in § (III.2), can be carried out by using an iterative procedure.

As a conclusion, this method allows to determine the optimal T-TF-based detector {23} (probability of error minimum) in the sense of the best criterion $f(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2)$, without setting it up. Note that this procedure can also be used to design linear time or time-frequency based detectors directly from training data, as it is shown in (15).

ii) Fisher linear discriminant sub-optimality: The Fisher linear discriminant is obtained by maximizing the following criterion (14):

$$l_{Fisher}(\eta_0, \eta_1, \sigma_0, \sigma_1) =$$

[27]

where P_i denotes the *a priori* probability of the hypothesis H_i.

The derivatives of $f_{Fisher}(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2)$ with respect to σ_0^2 and σ_1^2 are:

$$\frac{}{\partial \sigma_i^2} \qquad , i \in \{0, 1\}.$$

Therefore, $\alpha = P_0$ and $\underline{V}_{Fisher}$ satisfies:

$$\left[P_0 \Sigma_0 + (1 - P_0) \Sigma_1 \right]. \tag{28}$$

In this way, we show that the Fisher linear discriminant is a particular case of {26}, where a is equal to P_0. Consequently, this criterion is not necessarily the one which minimizes the probability of error, among those which belong to the class $f(\eta_0, \eta_1, \sigma_0^2, \sigma_1^2)$. In (14), it is also stated that Fisher discriminants can be arbitrarily bad: there are distributions such as even though the two classes are linearly separable, the Fisher linear discriminant has a probability of error close to one.

2. Algorithm

Assuming that N_i samples are available from each hypothesis H_i, the iterative algorithm mentioned in § (III.1.i) to design the optimum T-TF-based detector is as follows.

i) Declaration of the variables used in the algorithm:

$\underline{x}_i^j$: j^{th} observation, conditionally to the hypothesis H_i, $j \in [1, ..., N_i]$, $i \in \{0, 1\}$.

$\underline{Y}_i^j$: T-TF representation of $\underline{x}_i^j$ (cf. notation in § (III.1.i)).

$\hat{M}_i$: estimate of $\underline{Y}_i$ expected value conditionally to the hypothesis H_i, $i \in \{0, 1\}$.

$\hat{\Sigma}_i$: estimate of $\underline{Y}_i$ covariance matrix conditionally to the hypothesis H_i, $i \in \{0, 1\}$.

α : variable which belongs to the interval [0, 1] (cf. definition in Eq. {26}).

$\Delta\alpha$: variable α increment.

ii) Iterative procedure to find a:

1) Initializations:

· for each realization $\underline{x}_i^j$, evaluate the vector $\underline{Y}_i^j$.

· compute $\hat{M}_i$ and $\hat{\Sigma}_i$, $i \in \{0, 1\}$.

· set the variable α to 0.

2) While (a £ 1) repeat:

· solve: $\left[\alpha \hat{\Sigma}_0 + (1 - \alpha) \hat{\Sigma}_1 \right] V^{\alpha} = \left(\hat{M}_1 - \hat{M}_0 \right)$.

· find the threshold v_0^{α} which minimizes the estimate of the error probability P_e^{α}.

· change the variable α by: a ¬ a + Da.

3) Select the detector $\left(V^{\alpha}, v_0^{\alpha} \right)$ which minimizes P_e^{α}.

Finally, it should be reminded that in this process, no assumption is made on the distribution of the observations. The procedure is only based on our knowledge that $\underline{V}_{opt}$ must have the form of Eq. {26}, and depends on one parameter a. Then, an estimate of the error probability is used to select this parameter, and to adjust the threshold n_0.

3. Experiment on simulated data

The following application illustrates the proposed method. Consider the detection problem of a deterministic transient signal s(k) embedded in a zero-mean white Gaussian noise n(k):

$$\begin{cases} H_0: & x(k) = \rho.n(k) \\ H_1: & x(k) = s(k) + n(k) \end{cases}, k \ Î \ \{0, ..., 14\}, \tag{29}$$

where $s(k) = e^{-(k-7)^2/32} \cos(0.50 \pi k)$. Its representation is shown in Figure 1. The variance s^2 of the noise n(k) is chosen in order that the signal to noise ratio (SNR) is equal to -3 dB.

From Eq. {22}, it is straightforward to show that the Neyman-Pearson optimal statistic corresponding to the binary hypothesis testing problem {29} is given by the following log-likelihood ratio:

$$\lambda(\underline{x}) = \frac{1 - \rho^2}{2\rho^2} \underline{x}^T \underline{x} + \underline{s}^T \underline{x}, \tag{30}$$

which must be compared to a threshold n_0. By applying Moyal's relationship, one can also show that the previous statistic test is equivalent to the following joint linear filtering operations in the time and TF domains:

$$\Lambda_{T-TF}(\underline{x}) = \overline{\frac{}{60\rho^2} \sum_{k=0} \sum_{l=0}}. \tag{31}$$

14

Using the notations introduced in Eq. {23}, note that the reference $g(k,f_i)$ is constant over the TF domain, and corresponds to the one of an energy detector (see Table 1). The reference $h(k)$ is equal to the signal $s(k)$ to be detected.

In order to illustrate our approach, the following experiment of blind detector design from training data was conducted: the parameter r was set to 3 and 16,000 realizations of the hypotheses H_0 and H_1 were generated. This data set was used to train the detector {23} as described in § (III.2.ii). The resulting references are shown in Figures 2 and 3: $\underline{g}$ is nearly constant and $\underline{h}$ closely resembles the signal $\underline{s}$ to be detected. This result conforms to theory and illustrates the ability of the proposed method to closely approach the optimal receiver configuration. This original approach is used, in the next section, to define a new detector of K-complexes from a training set of EEG events.

IV. JOINT K-COMPLEXES DETECTION IN THE T-TF DOMAIN

1. Problem formulation

Automated detection of waveforms such as alpha, delta and K-complex waves in sleep EEG is an important component of sleep stage monitoring. The K-complex, which is characterized by a sharp upward wave followed by a downward one, is a key feature that contributes to sleep stages assessment. However, its automated detection is still difficult due to the stochastic nature of the EEG: the K-complex can have a large variety of shapes, and is not always distinctly different from the EEG background activity. Some examples of K-complexes and delta waves are respectively shown in Figures 4 and 5.

If we consider that the EEG background activity can be modified when a K-complex occurs, the detection problem can be intuitively posed as follows:

$$\begin{cases} H_0: & x(k) = n_0(k) \\ H_1: & x(k) = s(k) + n_1(k) \end{cases}, k \in \{0, ..., K\},$$ [32]

where n_1 (res. n_0) represents the EEG activity in the presence (res. in the absence) of the K-complex $\underline{s}$ to be detected. By analogy with the problem {21}, the T-TF-based approach proposed in § (II.3.i) is now used to solve {32}.

2. Experiments on a set of EEG events

i) Description of the EEG recordings: The EEG signals used for the K-complexes detector {23} design were stored from three channels (Fz, Cz, and Pz) but data from Cz was only used in the present study. The raw EEG data was digitized with a 8 bit A/D converter at a sampling frequency of 128 Hz, and segmented as follows. Two-second intervals, either containing K-complexes, and paroxysmal delta bursts bearing some resemblance to K-complexes were selected. The K-complexes had been selected by three EEGer at least, among the five experienced neurophysiologists who had individually scored the proposed whole nights records. Then, each segment was cropped by placing the x-axis intersection, present between the upward and downward peaks, in the middle of the interval. Each segment was also decimated by a factor of 4, and transformed into an analytic signal using Eq. {9}: the resulting sampling frequency of 32 Hz was adequate since the main energy of the K-complex remains below 6 Hz, as shown in Figure 4. Thus, a set including 609 K-complexes and 1196 waves bearing some resemblance to genuine K-complexes was used to train and test the detector {23}. As shown in Table 2, this data base was obtained from the EEG of three subjects.

ii) Reduction of the problem dimensionality: In the present study, most of the difficulty in designing the optimal T-TF-based detector came from the estimation of covariance matrices, since we were confronted with high-dimensional data whereas few training samples were available:

4096 variables from the (64´64)-dimensional TF representation,

+ 64 variables from the representation of the signal in the time domain,

= 4160 variables for 609 K-complexes and 1196 other waves.

In order to remove this difficulty, the number of variables coming from the TF representation was reduced as follows. First, only the informations $W_x(k, f_i)$ from the frequency band [0 - 8Hz] were extracted since the K-complex energy remains below 6 Hz. Secondly, these informations were compressed by applying the following singular transformation:

$$W_x(k, f_i)$$

.   [33]

In this way, the resolution of each TF representation was arbitrarily set to $\{Dt = 0.125 \text{ s} ; Df = 1 \text{ Hz}\}$, as shown in Figure 6. It should be noticed that this operation is also equivalent to impose the unknown reference $g(k,f_i)$, introduced in Eq. $\{23\}$, to be constant over $(0.125 \text{ s} \times 1 \text{ Hz})$ regions in the TF domain.

Thus, each 4160-dimensional sample $\underline{Y}^j$ was transformed in a 192-dimensional sample $\underline{Y}^j_{compressed}$, losing some of the discriminant information. This allowed us to accurately approximate the data base covariance matrices. We also noticed that the detector obtained after the data transformation $\{33\}$ was quite insensitive to K-complex shape variations, which justifies our choice. However, this operation should be replaced by a selection of the efficient T-TF regions for detection in a subsequent study.

3. Results and discussion

We performed the T-TF-based detector $\{23\}$ design on a training set containing 450 K-complexes and 850 delta waves. We used a distinct set (159 K-complexes, 346 delta waves) to test the obtained receiver, and to estimate its ROC curve. This experiment was repeated 500 times, using different random constitutions for the training and test sets each time. The mean ROC curve is shown in Figures 7(a) and 7(b) (solid). In Figure 8, the contribution of the time (l_L) and time-frequency (l_Q) statistics to the optimal T-TF-based detector is illustrated by the projection of a random data test set considered above onto the T-TF map. This example explicitly demonstrates the advantage offered by the combination of linear operations in the time and time-frequency domains.

The efficiency of our method was evaluated by comparing the T-TF-based detector performances to those of the matched filter $\{15\}$. The design of this receiver was repeated 500 times, using the same sets as those introduced before. Its mean ROC curve, shown in Figures 7(a) and 7(b) (dashed), clearly indicates the loss in performances due to the use of a linear statistic to solve the detection problem $\{32\}$, and emphasized the validity of our approach.

The best performances obtained by a K-complex detector were reported in (6): a morphological feature-based detector using an artificial neural network provided 90% true positives with 8.1% false positives while our approach yielded 9.2% false positives at the same sensitivity level. However, Bankman *et al.*'s training and test sets were containing non-K-complexes which were not necessarily

similar to K-complexes. In our study, we did not include such waveforms because it advantageously affects the false positives rate. In addition, the authors justifiably stated that their detector was not "the universal solution" for the K-complex detection problem because of two reasons:

a) Their feature-based receiver provided good agreement with the visual recognition of the single EEGer who had scored the EEG records when the morphological criteria differ across experts ([28]). In our study, the K-complexes we used had been selected by three EEGers at least, among the five experienced neurophysiologists who had individually scored the whole nights records.

b) The wide variety of EEG waveforms cannot be represented in a finite data set. In order to reduce the effect of a finite number of samples, we used a larger data base (see Table 2) than the one described in ([6]). The latter was containing 251 K-complexes, and 249 non-K-complexes.

As far as we are concerned, our T-TF-based detector has great potential since it provides satisfactory performances whereas the experimental conditions we chose to design it were intentionally against us. Besides, this receiver can significantly be improved. In the first place, we have supposed that the transformation {33} preserves as much discriminant information as possible. This assumption is not reasonable, and we now need to evaluate the effectiveness of individual T-TF variables or their combinations for detection, and to only select the effective one. Branch and Bound procedures, which have been developed to obtain optimal solution to some combinatorial problems, may advantageously be used. In the second place, informations from the channel Cz were only used in the present study. The decision regarding the detection of K-complexes may need to be corroborated by their presence in other channels.

V. CONCLUSION

In this paper, we have introduced the new concept of joint time and time-frequency based detection. As for classical time-frequency approaches to decision problems, this method allows to combine the analysis of signals, conveniently performed in the time-frequency domain, and the decision stage.

Unfortunately all these methods require substantial knowledge of signals, whereas phenomena are complex and poorly understood in many applications (detection in EEG signals, ...). To improve the

18

detector design phase in these situations, we have developed a method to determine the optimum detector, i.e., which minimizes the error probability, directly from labeled training data. The resulting detector theoretically offers better performances than the one obtained via the maximization of the Fisher criterion (13), as it has been shown in (15).

Finally, our T-TF-based detection structure, designed by the method described in § (III.2), has been validated with the problem of K-complexes detection in sleep EEG. It results from the present study that the obtained receiver is potentially the best one which can be found in the literature. Besides, it can be improved by using for example a Branch and Bound procedure to select the efficient T-TF regions for detection. The decision regarding the detection of K-complexes may also need to be corroborated by their presence in other channels. Finally, it should be emphasized that the methodology we have introduced in this paper can advantageously be used to solve many other transient signals detection problems (sharp vertex waves, ...).

ACKNOWLEDGMENT

REFERENCES

1. REYNOLDS, C., KUPFER, C. F., HOUCK, D. J., STACK, C. C., BERMAN, S. R., AND ZIMER, B. Sleep Research in Affective Illness: State of the Art. *Sleep* **10,** 189 (1987).

2. GAILLARD, J. M., AND TISSOT, R. Principles of Automatic Analysis of Sleep Records with Hybrid System. *Comput. Biomed. Res.* **6,** 1 (1973).

3. SCHALTENBRAND, N., LENGELLE, R., AND MACHER, J.-P. Neural Network Model: Application to Automatic Analysis of Human Sleep. *Comput. Biomed. Res.* **26,** 157 (1993).

4. RAY, S., LEE, W. D., MORGAN, C. D., AND AIRTH-KINDREE, W. Computer Sleep Stage Scoring. An expert System Approach. *Int. J. Bio-Med. Comput.* **19,** 43 (1986).

5. JANSEN, B. H., AND DESAI, P. R. K-complex Detection using Multi-Layer Perceptron and Recurrent Network. *Int. J. Bio-Med. Comput.* **37,** 249 (1994).

6. BANKMAN, I. N., SIGILLITO, V. G., WISE, R. A., AND SMITH, P. L. Feature-Based Detection of K-Complex Wave in the Human Electroencephalogram using Neural Networks. *IEEE Trans. Biomed. Eng.* **39,** 1605 (1992).

7. DA ROSA, A. C., KEMP, B., AND PAIVA, T. A Model-based Detector of Vertex Sharp Waves and K-Complexes in Sleep Electroencephalogram. *Electroencephalogr. Clin. Neurophysio.* **78,** 71 (1991).

8. JANSEN, B. H. Artificial Neural Nets for K-Complex Detection. *IEEE Eng. Med. Biol.* **9,** 50 (1990).

9. TANG, Z., AND ISHII, N. Detection of the K-Complex using a New Method of Recognizing Waveform Based on the Discrete Wavelet Transform. *IEICE Trans. Inf. Syst.* **78,** 77 (1995).

10. FLANDRIN, P. "Temps - Fréquence." HERMES, Paris, France, 1993.

11. FLANDRIN, P. A Time-Frequency Formulation of Optimum Detection. *IEEE Trans. Acoust. Speech Signal Process.* **36,** 1377 (1988).

12. SAYEED, A. M., AND JONES, D. L. Optimal Detection using Bilinear Time-Frequency and Time-Scale Representations. *IEEE Trans. Signal Process.* **43,** 2872 (1995).

13. JONES, D. L., AND SAYEED, A. M. Blind Quadratic and Time-Frequency based Detectors from Training Data. *In* "Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing", 1995.

14. DEVROYE, L., GYORFI, L., AND LUGOSI, G. "A Probabilistic Theory of Pattern Recognition." SPRINGER VERLAG, New York, USA, 1996.

15. RICHARD, C., AND LENGELLE, R. Une nouvelle approche pour la détection linéaire optimale dans le plan temps-fréquence. *In* "Proceedings, Seizième Colloque GRETSI sur le Traitement du Signal et des Images", 1997.

16. MARTIN, W., AND FLANDRIN, P. Wigner-Ville Spectral Analysis of Nonstationary Processes. *IEEE Trans. Acoust. Speech Signal Process.* **33,** 1461 (1985).

17. CUNNINGHAM, G. S., AND WILLIAMS, W. J. Fast Implementations of Generalized Discrete Time-Frequency Distributions. *IEEE Trans. Signal Process.* **42,** 1496 (1994).

18. BOASHASH, B., AND BLACK, P. J. An Efficient Real-Time Implementation of the Wigner-Ville Distribution. *IEEE Trans. Acoust. Speech Signal Process.* **35,** 1611 (1987).

19. CUNNINGHAM, G. S., AND WILLIAMS, W. J. Fast Computation of the Wigner Distribution for Finite Length Signal. *In* "Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing", 1992.

20. AMIN, M. G. Computationally Lag-invariant Recursive Spectrum Estimator. *IEEE Trans. Acoust. Speech Signal Process.* **35,** 1713 (1987).

21. RICHARD, C., AND LENGELLE, R. Joint Recursive Implementation of Time-Frequency Representations and their Modified Version by the Reassignment Method. *Signal Process.* **60,** (1997).

22. AUGER, F., AND FLANDRIN, P. Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. *IEEE Trans. Signal Process.* **43,** 1068 (1995).

23. POOR, H. V. "An Introduction to Signal Detection in Noise." Springer-Verlag, New York, USA, 1994.

24. FLANDRIN, P. Signal Detection in the Time-Frequency Plane. *In* "Proceedings, Processing in Automatic Control", 1989.

25. SAMIMY, B., RIZZONI, G., SAYEED, A. M., AND JONES, D. L. Design of Training Data-Based Quadratic Detectors with Application to Mechanical Systems. *In* "Proceedings, IEEE International Conference on Acoustics, Speech and Signal Processing", 1996.

26. KUMAR, B. V. K. V., AND CARROLL, C. W. Performance of the Wigner Distribution based Detection Methods. *Opti. Eng.* **23,** 732 (1984).

27. FUKUNAGA, K. "Introduction to Statistical Pattern Recognition." Academic Press, London, GB, 1990.

28. BREMER, G., SMITH, J. R., AND KARACAN, I. Automatic Detection of the K-Complex in Sleep Electroencephalograms. *IEEE Trans. Biomed. Eng.* **17,** 314 (1970).

$v(k,f_i)$	$\lambda_{T-TF}(\underline{x})$	Type
$W_s(k,f_i)$	$4(2M-1)\left|\sum_{k=0}^{K} x(k).s^*(k)\right|^2$	Quadrature matched filter
1	$2(2M-1)\sum_{k=0}^{K}|x(k)|^2$	Energy detector

Table 1: Quadratic detector configurations resulting from the choice of suitable TF references.

	K-complexes	Delta waves
Subject 1	263	397
Subject 2	162	399
Subject 3	184	400
Total	609	1196

Table 2: Number of K-complexes and delta waves per subject used to train and test the detector.

FIGURES CAPTIONS

Fig.1:   Representation of the transient signal $\underline{s}$ to be detected from training data.

Fig.2:   Time reference $\underline{h}$ used by the T-TF-based receiver to detect the transient signal $\underline{s}$ represented in Figure 1. This reference has been obtained from training data by means of our algorithm.

Fig.3:   TF reference $\underline{g}$ used by the T-TF-based receiver to detect the transient signal $\underline{s}$ represented in Figure 1. This reference has been obtained from training data by means of our algorithm.

Fig.4:   Examples of K-complexes.

Fig.5:   Examples of delta waves.

Fig.6:   K-complexes detection: reduction of the problem dimensionality.

Fig.7a: ROC curves respectively provided by the T-TF-based (solid) and the linear (dashed) detectors of K-complexes.

Fig.7b: Zoom in on the ROC curves respectively provided by the T-TF-based (solid) and the linear (dashed) detectors of K-complexes.

Fig.8:   Projection onto the T-TF domain of the test set: K-complexes (^{2}), delta waves (l).

Title: figure1.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:23:38

Figure 2:

Title: figure2.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:23:45

Figure 3:

Title: figure3.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/13/98 10:09:27

Title: figure4.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:23:58

Figure 5:

Title: figure5.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:24:14

Figure 6:

(64 by 64) TF representation

transformation {33}

of the frequency band [0-8]

$\Delta_{f=1/4\,Hz}$

$\Delta_{t=1/32\,s}$

16

8

0

frequency (Hz)

0   time(s)   2

(16 by 8) TF representation

$\Delta_{f=1\,Hz}$

$\Delta_{t=1/8\,s}$

16

8

0

frequency (Hz)

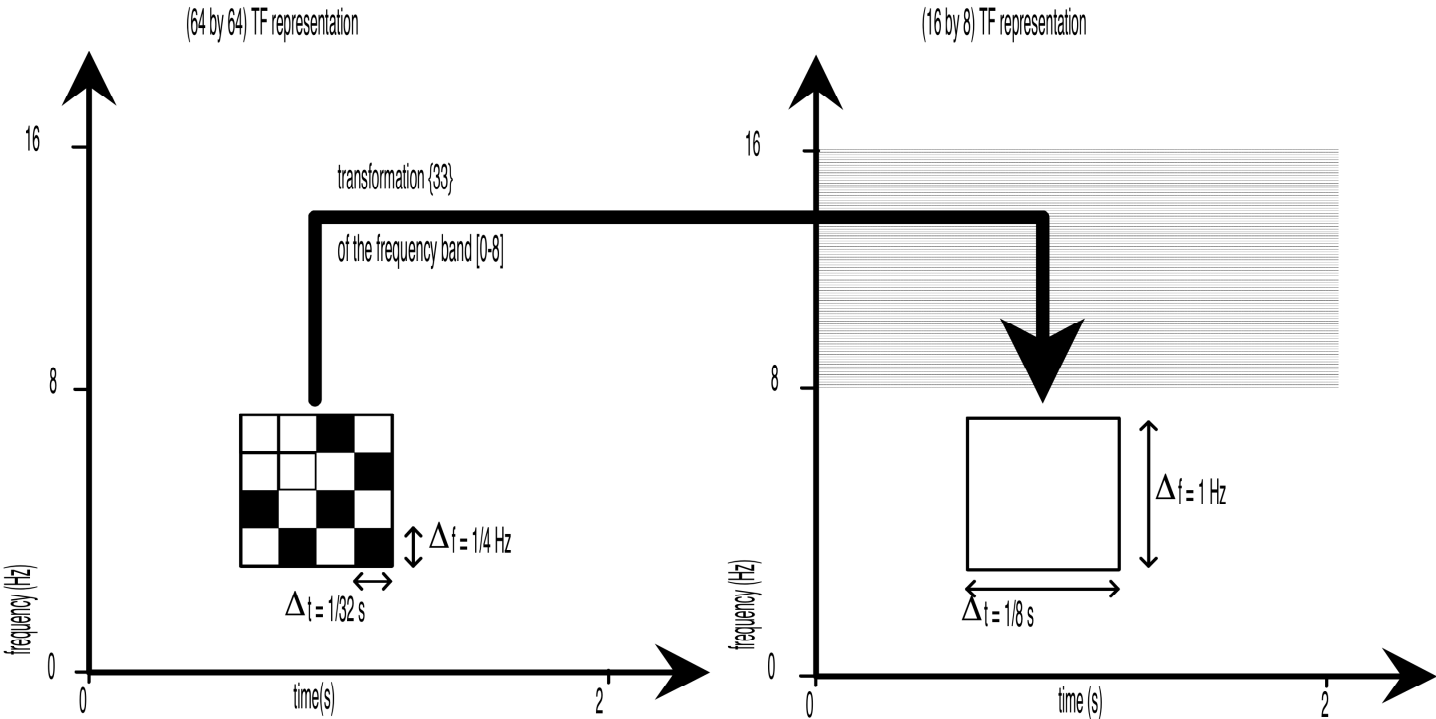0   time (s)   2

Figure 7(a):

Title: figure7a.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:24:22

Figure 7(b):

Title: figure7b.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:24:27

Figure 8:

Title: figure8.eps
Creator: MATLAB, The Mathworks, Inc.
CreationDate: 03/12/98 18:48:32