Kulback-Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients

R. Quian Quiroga,[1] J. Arnhold,[1,2] K. Lehnertz,[2] and P. Grassberger[1]
[1]*John von Neumann Institute for Computing, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany*
[2]*Clinic of Epileptology, University of Bonn, Sigmund-Freud Straße 25, D-53105 Bonn, Germany*
(Received 20 September 1999; revised manuscript received 7 April 2000)

Recently, ''renormalized entropy'' was proposed as a novel measure of relative entropy [P. Saparin *et al.*, Chaos, Solitons and Fractals **4**, 1907 (1994)] and applied to several physiological time sequences, including electroencephalograms (EEGs) of patients with epilepsy. We show here that this measure is just a modified Kullback-Leibler (KL) relative entropy, and it gives similar numerical results to the standard KL entropy. The latter better distinguishes frequency contents of, e.g., seizure and background EEGs than renormalized entropy. We thus propose that renormalized entropy might not be as useful as claimed by its proponents. In passing, we also make some critical remarks about the implementation of these methods.

PACS number(s): 87.90.+y, 05.45.Tp, 87.19.Nn

I. INTRODUCTION

Since Shannon's classical works, information theoretic concepts have found many applications in practically all fields of science. In particular, tools derived from information theory have been used to characterize the degree of randomness of time sequences, and to quantify the difference between two probability distributions. Indeed there are a number of constructs which qualify as distances between two distributions. Although the Kullback-Leibler (KL) relative entropy [1,2] is not a distance in the mathematical sense (it is not symmetric), it plays a central role as it has numerous applications and numerous physical interpretations. Another, seemingly independent, observable measuring a dissimilarity between two distributions was recently introduced in [3]. This ''renormalized entropy'' was subsequently applied to various physiological time sequences, including heart beats [4,5] and electroencephalograms (EEGs) recorded in patients with epilepsy [6]. The relation between KL and renormalized entropy and their application to EEGs recorded in patients with epilepsy are the subject of the present paper.

Ever since the first recordings in the late 1920s, the EEG has been one of the most powerful tools in neurophysiology [7]. An important application of EEGs in clinical practice is the diagnosis of epilepsy. Characteristic abnormal patterns help to classify epilepsies, to localize the epileptogenic focus, and eventually to predict seizures [8]. About 20% of patients suffering from focal epilepsies do not improve with antiepileptic medication and are therefore assumed candidates for a surgical resection of the seizure-generating area. Successful surgical treatment of focal epilepsies requires exact localization of the seizure-generating area and its delineation from functionally relevant areas. Recording the patient's spontaneous habitual seizures by means of long-term (several days), and in some cases intracranial, EEGs (i.e., with electrodes implanted within the skull) is currently assumed most reliable.

Although EEG recordings have been in clinical use for more than half a century, conventional EEG analysis relies mostly on visual inspection or on linear methods such as the Fourier transform (see, e.g., [9] for a comprehensive description of Fourier analysis in EEGs). Particularly for the diagnosis of epilepsy, quantitative methods of analysis are in need to give additional information (for a review of quantitative methods in EEG analysis, see, e.g., [7]). It is precisely in this context that the authors of [6] found renormalized entropy to be much more significant than any of the other methods they looked at.

In the following, we argue that renormalized entropy is very closely related to KL entropy. Indeed, it *is precisely* a KL entropy, although not between the two distributions one started out to compare. Nevertheless, we can relate renormalized entropy to the KL entropy between these two distribution. Moreover, when extracting these measures from EEGs, we find both to be very similar. It seems indeed from these analyses that standard KL entropy is more useful than renormalized entropy.

In the next section we recall Shannon and KL entropies, and show how renormalized entropy is related to KL entropy. In Sec. III we present applications to seizure EEG data. In this section we also address several technical points concerning the implementation in the case of EEG data, and we discuss the importance of the results from a neurophysiological point of view. Finally, in Sec. IV we draw our conclusions.

II. ENTROPY MEASURES

We consider a discrete random variable having n possible outcomes $x_k (k=1,\ldots,n)$ with respective probabilities p_k, satisfying $p_k \geqslant 0$ and $\Sigma_{k=1}^{n} p_k = 1$. The Shannon entropy of p is defined as [10]

$$H[p] = -\sum_k p_k \ln p_k . \tag{1}$$

In the following we shall take k as a frequency index and p_k as a normalized spectral density,

$$p_k = \frac{S(\omega_k)}{\sum_k S(\omega_k)} . \tag{2}$$

Moreover, the spectrum will be estimated from gliding windows over a scalar (''univariate'') time sequence x_n,

$$S(\omega_k)=S_t(\omega_k)=[|X_t(\omega_k)|^2]_{\text{smooth}}, \tag{3}$$

where $X_t(\omega_k)$ is the discrete Fourier transform of x_n taken over a window of length T centered at time t (see Sec. III for details), and the bracket $[\ ]_{\text{smooth}}$ indicates a local averaging over nearby frequencies. We should stress, however, that all the results of the present section apply to any probability distribution.

Shannon entropy is equal to 0 in the case of δ distributions, and positive otherwise. It can be interpreted as the average amount of code length [measured in bits, if the logarithm in Eq. (1) is taken with base 2] needed to encode a randomly chosen value of k (randomly with respect to p). The essential point here is that the minimal (average) code length is obtained by codes which are optimal for a specific probability distribution—see, e.g., the Morse code which uses shorter codes for the more frequent letters.

Let us now suppose we have two different probability distributions $p=\{p_k\}$ and $q=\{q_k\}$. We can then define the KL (relative) entropy as [1,2]

$$K(p|q)=\sum_k p_k \ln\frac{p_k}{q_k}. \tag{4}$$

It is also positive and vanishes only if $p_k\equiv q_k$, thus measuring the degree of similarity between both probability distributions. Notice, however, that it is in general not symmetric, $K(p|q)\neq K(q|p)$, therefore it is not a distance in the usual mathematical sense. Its most important interpretation is the following: assume that p is the correct distribution, but the encoding is made using a code which would have been optimal (i.e., would have produced the shortest average code length) if the distribution were q instead. Then, $K(p|q)$ measures the average excess of the code length (again measured in bits, if the logarithm is base 2) over the shortest code (which would have been based on p). But there are also several different interpretations in different contexts. For instance, mutual information [1] can be considered as KL entropy with p the true joint distribution and q the product of the marginal distributions. Also, Boltzmann's H theorem is most easily derived using KL entropies [2].

A supposedly different and independent distance measure between two distributions was introduced in [3]. These authors called q the ''reference distribution.'' They defined a ''renormalized'' reference distribution $\tilde{q}$ as

$$\tilde{q}_k=C[q_k]^\beta, \tag{5}$$

where C and β are uniquely fixed by demanding

$$\sum_k \tilde{q}_k \ln q_k=\sum_k p_k \ln q_k \tag{6}$$

and

$$\sum_k \tilde{q}_k=1. \tag{7}$$

Then they define ''renormalized entropy'' as

$$\Delta H=H[p]-H[\tilde{q}] \tag{8}$$

and show that it is negative definite, except when $p\equiv q$. When applying it to time-resolved spectra of several physiological time series, it is claimed in [3–6] that ΔH gives more significant results (e.g., shows more clearly the onset of an epileptic seizure [6]) than any other observable studied by these authors. We want to show now that (i) the renormalized entropy is just the negative of the KL entropy between p and $\tilde{q}$,

$$\Delta H=-K(p|\tilde{q}); \tag{9}$$

(ii) the absolute value $|\Delta H|$ is less than the KL entropy between p and q, since the difference between both is also a KL entropy,

$$|\Delta H|=K(p|q)-K(\tilde{q}|q)\leqslant K(p|q). \tag{10}$$

This strongly suggests that renormalized entropy cannot be more useful than the standard KL relative entropy between the unrenormalized distributions.

To prove our claims, we notice that we can rewrite Eq. (6), using Eqs. (5) and (7), as

$$\sum_k \tilde{q}_k \ln\tilde{q}_k=\sum_k p_k \ln\tilde{q}_k. \tag{11}$$

Therefore,

$$\begin{aligned}\Delta H&=\sum_k \tilde{q}_k \ln\tilde{q}_k-\sum_k p_k \ln p_k\\ &=\sum_k p_k \ln\tilde{q}_k-\sum_k p_k \ln p_k=-\sum_k p_k \ln\frac{p_k}{\tilde{q}_k},\end{aligned} \tag{12}$$

which proves our first claim. Furthermore, we can write

$$\begin{aligned}\Delta H+K(p|q)&=\sum_k p_k \ln\tilde{q}_k-\sum_k p_k \ln q_k\\ &=\sum_k \tilde{q}_k \ln\tilde{q}_k-\sum_k \tilde{q}_k \ln q_k=\sum_k \tilde{q}_k \ln\frac{\tilde{q}_k}{q_k},\end{aligned} \tag{13}$$

which proves the second claim.

III. APPLICATION TO EEG DATA

A. Details of the data

We will illustrate the result of the preceding section by reanalyzing some of the same data used in [6]. The data correspond to an intracranial multichannel EEG recording of a patient with medial temporal lobe epilepsy; it was sampled with 173 Hz and band pass filtered in the range 0.53 $-$85 Hz. In Fig. 1 we show EEG time sequences (500 000 data points, approximately 48 min of continuous recording) from three different recording sites prior to, during, and after an epileptic seizure. Seizure starts at about point 270 000 (minute 26) and lasts for 2 min. The recording sites are located within the seizure-generating area (upper trace), adja-

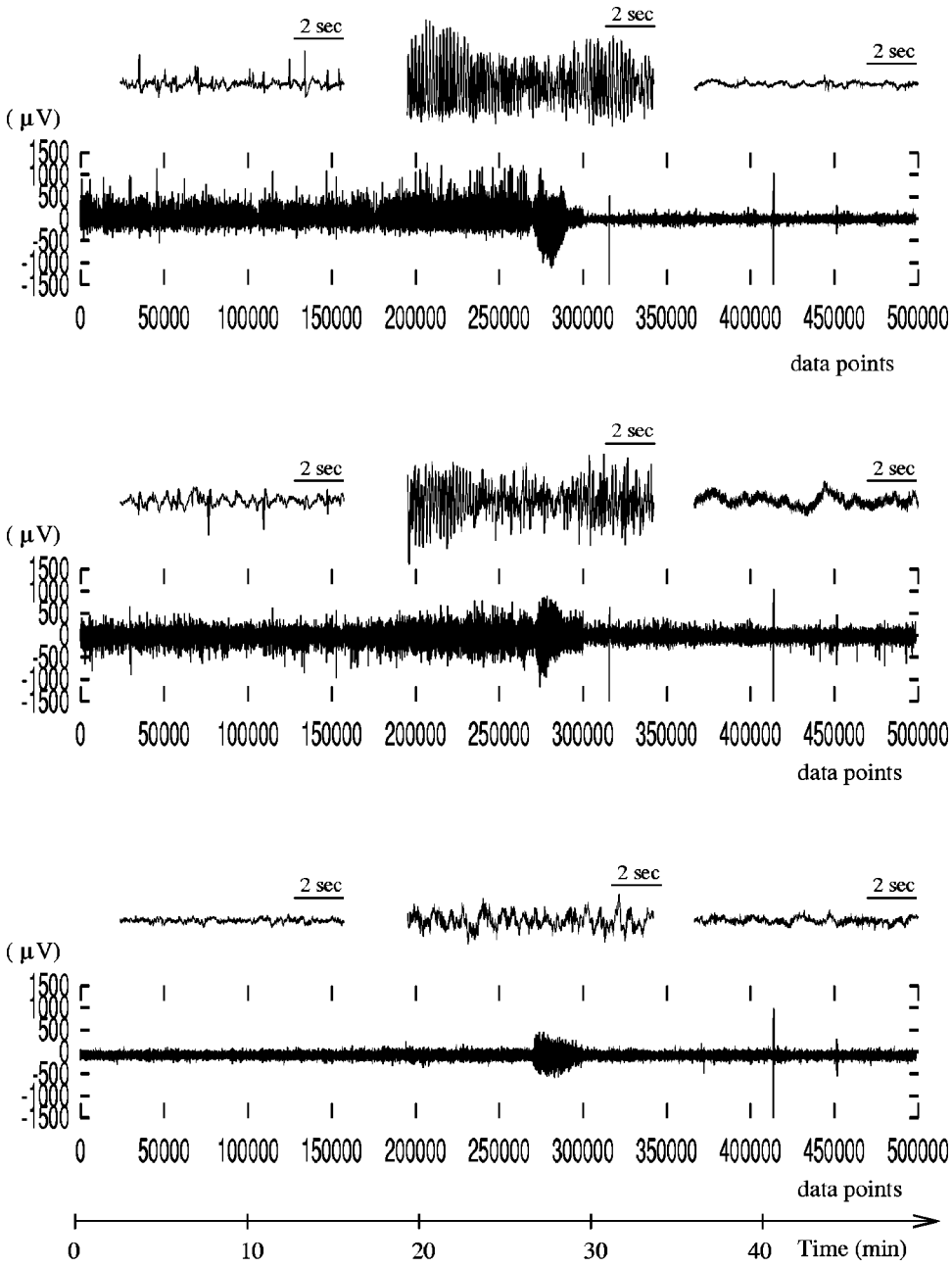

FIG. 1. Intracranial EEG recordings prior, during, and after an epileptic seizure of right medial temporal origin. Recordings were taken from within (upper plot) and adjacent (middle plot) to the seizure-generating area as well as from the nonaffected brain hemisphere (lower plot). For each state and electrode, a 10 sec zoom of the signal is shown. See text for further details. The vertical lines at about 316 000, 415 000, and 451 000 data points are due to artifacts in the recording. The data corresponding to these artifacts were not considered for further analysis.

cent to it (middle trace), and on the nonaffected brain hemisphere (lower trace) To better visualize the dynamics, insets drawn on top of each signal show typical EEG sequences of 10 sec duration during the preseizure (left), seizure (middle), and the postseizure stage (right).

B. Power spectrum

For a finite data set x_n sampled at discrete times $t_n = n\,\Delta t, n=1,\ldots,N, T=N\,\Delta t$, we denote by $X(\omega_k)$ its discrete Fourier transform at $\omega_k = 2\,\pi\, k/T$, with $k=1,\ldots,N$. We estimate the power spectrum as

$$S(\omega_k) = C \sum_{n=-b}^{b} w(n) |X(\omega_{k+n})|^2, \tag{14}$$

where $w(n)$ is a smoothing function of window size $B = 2b+1$, and C is a normalization factor. As in Ref. [6], a Bartlett-Priestley smoothing function was used,

$$w(n) \propto \begin{cases} [1-(n/b)^2], & |n| \leqslant b \\ 0, & |n| > b. \end{cases} \tag{15}$$

As in [6] and for comparison purposes, we subdivide the data in (half overlapping) epochs of $T \simeq 24$ s ($N=4096$ data points), and choose the window size of the Bartlett-Priestley function as $B=33$. This window length corresponds to a frequency resolution of 0.042 Hz. In the following, we consider the spectrum in the region $\omega < 30$ Hz. Moreover, since we are not interested in the absolute power, the normalization factor C is adjusted such that the sum over all frequencies below 30 Hz gives unity.

C. Shannon entropy

Parts (a)–(c) of Figs. 2–4 show the EEG signals recorded at the three sites, contour plots of the corresponding normalized power spectra, and time-dependent estimates of the

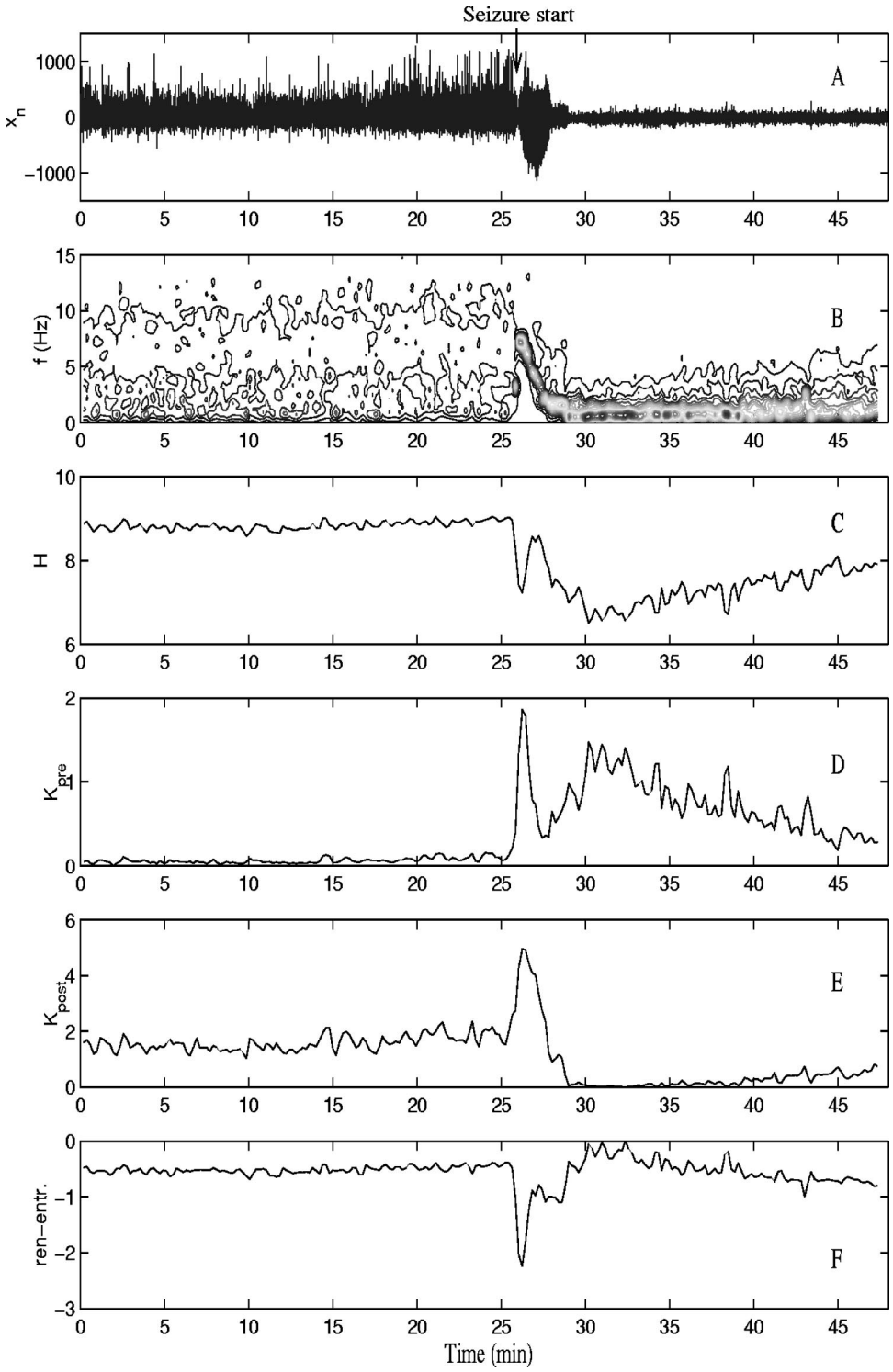

FIG. 2. (A) EEG recording from the electrode contact within the seizure-generating area, (B) its corresponding power spectrum, (C) Shannon entropy, and Kullback-Liebler entropy taking a preseizure (D) and a postseizure (E) reference window, and (F) renormalized entropy (same postseizure reference window).

Shannon entropy *H*. Prior to the seizure, power spectra exhibit an almost stable but spread frequency composition which is reflected in high values of *H*.

When the seizure starts, the spectra in Figs. 2 and 3 are dominated by a single frequency component (~ 7 Hz). This is reflected in Fig. 2 by an abrupt decrease of *H* by about 20%. Actually, the decrease is even more pronounced for smaller time windows, since the period of strong coherence is much shorter than 24 sec. As the seizure evolves, the dominant frequency decreases rapidly. This dynamics is characteristic of seizures originating from the medial temporal lobe (see, e.g., [11]) but it is not the only possible one [12]. The rise of *H* in both Figs. 2 and 3 immediately before the final drop can partially be attributed to this fast change of dynamics. The estimated entropy is high during this phase because of several subsequently appearing frequencies in the same window. The following concentration of activity at lower frequencies finally leads to a decrease of *H*. To a lesser degree this is also seen in Fig. 4. Within or close to the seizure-generating area, *H* remains small throughout the entire recorded postseizure stage. Finally, it slowly increases towards values that compare to those obtained during the

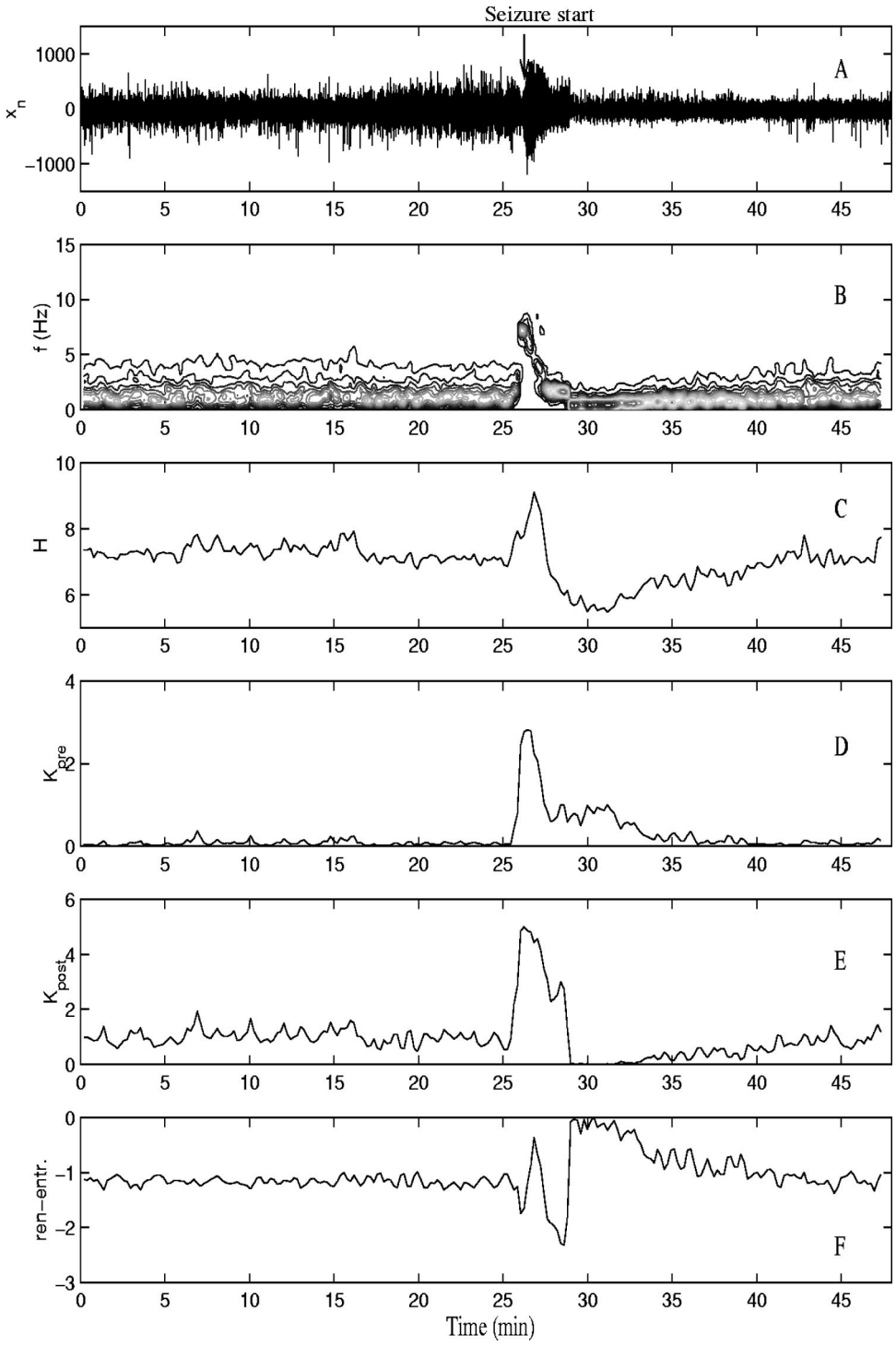

FIG. 3. Same as Fig. 2 but for the electrode adjacent to the seizure-generating area.

preseizure stage. Using a Shannon entropy defined from the wavelet transform, similar results were obtained in Ref. [13] from an analysis of a scalp recorded seizure.

D. Kullback-Leibler entropy

The time courses of the KL entropy $K(p|q)$ are shown in parts (d) of Figs. 2–4. As reference segments we used the signals from the preseizure stage consisting of 4096 data points and starting at $n=20\,480$. The sensitivity [i.e., increase of $K(p|q)$ during the seizure relative to the background level] is notably improved when compared to that of the Shannon entropy. Background fluctuations during the preseizure stage only slightly affected $K(p|q)$ since preseizure power spectra from different windows are almost similar. Also, $K(p|q)$ proved nearly independent of the choice of the reference segment, as long as it was chosen from the preseizure stage.

As with the Shannon entropy, we see in Figs. 2 and 3 a marked change at seizure onset due to a concentration of spectral power at frequencies ~ 7 Hz. $K(p|q)$ clearly detects this difference. It also detects the spectral difference when lower frequencies dominate in the postseizure stage. But again the rapid frequency change after seizure onset is hard to distinguish from a broadband spectrum due to our somewhat large window size T.

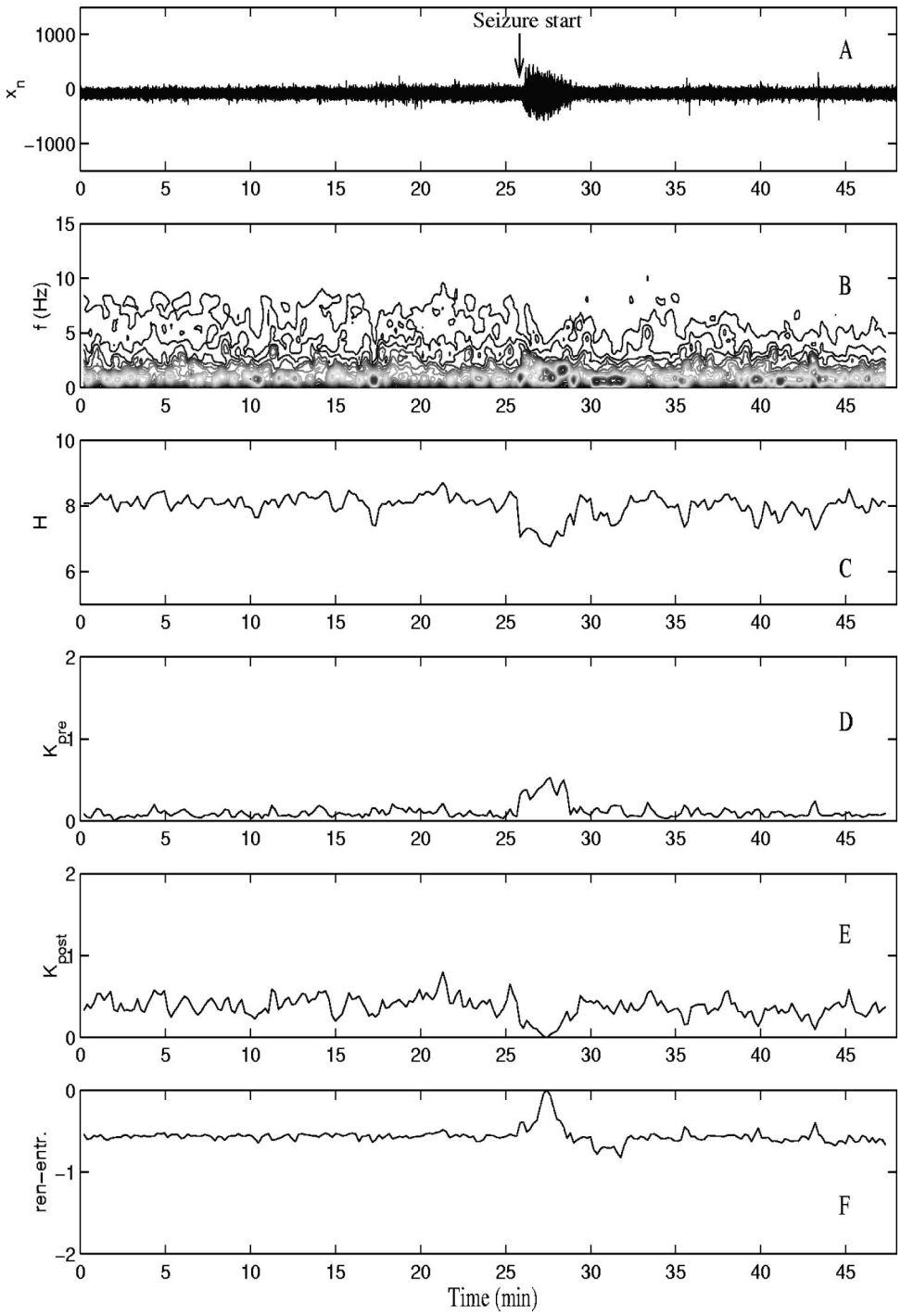

FIG. 4. Same as Fig. 2 but for the electrode located in the nonaffected brain hemisphere.

The last two parts of Figs. 2–4 show time courses of the KL entropy and the renormalized entropy calculated using a reference segment with lowest Shannon entropy as was done by the authors of [6]. For Figs. 2 and 3 this was after the seizure (4096 data points starting at $n=335\,872$ and $n=315\,392$, respectively), while it was during the seizure for data shown in Fig. 4 (4096 data points starting at $n=284\,672$).

Here KL and renormalized entropies give similar results. This illustrates the similarity between renormalized and KL entropies as already pointed out in Sec. II. Differences with results in [6] can be attributed partly to differences in the exact choice of the reference segment. We see that peak values of $K(p|q)$ are larger than those based on calculations using a preseizure reference window. However, the relative increases over preseizure values are much less pronounced. Therefore, we consider postseizure reference segments as not very useful for seizure detection. Moreover, postseizure reference segments obviously cannot be used in real-time applications. In addition, a postseizure reference segment is not very reasonable physiologically. Immediately after a seizure, the state of the patient and, accordingly, the EEG are highly abnormal. Typically, the postseizure EEG exhibits slow fluctuations of high amplitude, sometimes superposed with high-frequency activity (see Fig. 1). This is obviously not a typical background EEG. Moreover, the postseizure stage is often contaminated by artifacts, some of which are not as easily recognizable as those shown in Fig. 1.

We therefore disagree with the procedure proposed in Ref. [6] of automatically choosing a reference as the segment with lowest entropy for each recording channel. Instead, we propose to choose a reference segment recorded during a state as ''normal'' as possible, i.e., far from a seizure (we should note, however, that there is still a lot of controversy in neurophysiology over what is considered to be ''far''), free of artifacts and, if possible, free of abnormal alterations (admittedly, this is not always possible). Moreover, the reference segment should be exactly the same time interval for all channels. Otherwise comparisons between different recording sites are not reliable. Also, one might consider taking shorter time segments. This would of course enhance statistical fluctuations, but would allow better time resolution.

Even then it would be difficult to detect the recording site showing the very first sign of the seizure, which is necessary for an exact focus localization. We verified this for windows down to 1.5 sec (data not shown). This is in agreement with clinical experience, which shows that the time scales relevant for this detection can be less than 1 sec. Because of these problems, the suggestions of [6] concerning clinical applications such as seizure detection or localization of epileptic foci seem too optimistic.

Finally, we remark that none of the entropy measures appeared to show information prior to the onset of the seizure exceeding naked eye visualization or spectral analysis of the EEG. An unequivocal definition of a long-lasting preseizure state, however, is of great importance. Apart from an early warning for the patient, this definition would allow pharmacological or electrotherapeutic interventions as well as basic research about seizure-generating mechanisms in humans in the preseizure period. Since all the entropies described in this study were defined from the Fourier power spectrum, our findings support the view that with traditional linear methods relevant information of an impending seizure is restricted to at most a few seconds prior to its onset [8]. In contrast, measures defined within the analysis framework of the theory of nonlinear dynamical systems have recently shown more promising results [8,14–16].

IV. CONCLUSION

The aim of the present paper was twofold. First, we showed that ''renormalized entropy,'' a novel entropy measure for differences in probability distributions, is closely related to Kullback-Leibler entropy. We also argued that it is very unlikely that more information is obtained from the former than from the latter. Second, we checked recent claims that renormalized entropy (and thus also KL entropy) is very useful in applications to intracranial EEGs from epilepsy patients. We found some of these claims to be unjustified. Nevertheless, the fact remains that KL entropy applied to spectral distributions is a very promising tool which has not yet been studied much in this context. In fact, ''abnormal'' frequency patterns corresponding to epileptic seizures were better identified with KL than with the Shannon entropy. While the present study was performed on a limited amount of data, we suggest KL entropy to be an interesting tool for a more systematic study.

Finally, we point out that the KL entropy can also be defined from other time-frequency distributions rather than the windowed Fourier transform. In particular, we consider wavelets as good candidates, since they have optimal resolution both in the time and the frequency range (see [17,18] for theoretical background and [19,20] for application to EEGs).

ACKNOWLEDGMENTS

K.L. acknowledges support from the Deutsche Forschungsgemeinschaft.

[1] R. Gray, *Entropy and Information Theory* (Springer-Verlag, New York, 1990).
[2] S. Guiasu, *Information Theory with Applications* (McGraw-Hill, New York, 1977).
[3] P. Saparin, A. Witt, J. Kurths, and V. Anishenko, Chaos, Solitons and Fractals **4**, 1907 (1994).
[4] J. Kurths *et al.*, Chaos **5**, 88 (1995).
[5] A. Voss *et al.*, Cardiovasc. Res. **31**, 419 (1996).
[6] K. Kopitzki, P.C. Warnke, and J. Timmer, Phys. Rev. E **58**, 4859 (1998).
[7] F.H. Lopes da Silva, in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, 3rd ed., edited by E. Niedermeyer and F.H. Lopes da Silva (Williams and Wilkins, Baltimore, 1993), p. 1097.
[8] K. Lehnertz and C.E. Elger, Phys. Rev. Lett. **80**, 5019 (1998).
[9] G. Dumermuth and L. Molinari, in *Handbook of Electroencephalography and Clinical Neurophysiology, Vol. I: Methods of Analysis of Brain Electrical and Magnetic Signals*, edited by A. Gevins and A. Rémond (Elsevier, Amsterdam, 1987), p. 85.
[10] C.E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[11] R. Quian Quiroga *et al.*, Electroencephalogr. Clin. Neurophysiol. **103**, 434 (1997).
[12] D. King and S. Spencer, J. Clin. Neurophysiol. **12**, 32 (1995).
[13] S. Blanco *et al.*, Phys. Rev. E **57**, 932 (1998).
[14] J. Martinerie *et al.*, Nat. Med. **4**, 1173 (1998).
[15] H.R. Moser, B. Weber, H.G. Wieser, and P.F. Meier, Physica D **130**, 291 (1999).
[16] J. Arnhold, P. Grassberger, K. Lehnertz, and C.E. Elger, Physica D **134**, 419 (1999).
[17] A. Grossmann and J. Morlet, SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **15**, 723 (1984).
[18] C. Chui, *An Introduction to Wavelets* (Academic Press, San Diego, 1992).
[19] R. Quian Quiroga and M. Schürmann, Clin. Neurophysiol. **110**, 643 (1999).
[20] S.J. Schiff, A. Aldrouby, M. Unser, and S. Sato, Electroencephalogr. Clin. Neurophysiol. **91**, 442 (1994).