

## Article

# Benefits of Zero-Phase or Linear Phase Filters to Design Multiscale Entropy: Theory and Application

Eric Grivel <sup>1,\*</sup> , Bastien Berthelot <sup>2</sup>, Gaetan Colin <sup>3</sup>, Pierrick Legrand <sup>4</sup> and Vincent Ibanez <sup>2</sup>

<sup>1</sup> IMS Laboratory, Bordeaux INP, Bordeaux University, UMR CNRS 5218, 33400 Talence, France

<sup>2</sup> Thales AVS France, Campus Merignac, 75-77 Av. Marcel Dassault, 33700 Mérignac, France; bastien.berthelot@fr.thalesgroup.com (B.B.); vincent.ibanez@fr.thalesgroup.com (V.I.)

<sup>3</sup> ENSEIRB-MATMECA, Bordeaux INP, 33400 Talence, France

<sup>4</sup> IMB Laboratory, Bordeaux University, UMR CNRS 5251, ASTRAL Team, INRIA, 33400 Talence, France; pierrick.legrand@u-bordeaux.fr

\* Correspondence: eric.grivel@ims-bordeaux.fr; Tel.: +33-556844466

**Abstract:** In various applications, multiscale entropy (MSE) is often used as a feature to characterize the complexity of the signals in order to classify them. It consists of estimating the sample entropies (SEs) of the signal under study and its coarse-grained (CG) versions, where the CG process amounts to (1) filtering the signal with an average filter whose order is the scale and (2) decimating the filter output by a factor equal to the scale. In this paper, we propose to derive a new variant of the MSE. Its novelty stands in the way to get the sequences at different scales by avoiding distortions during the decimation step. To this end, a linear-phase or null-phase low-pass filter whose cutoff frequency is well suited to the scale is used. Interpretations on how the MSE behaves and illustrations with a sum of sinusoids, as well as white and pink noises, are given. Then, an application to detect attentional tunneling is presented. It shows the benefit of the new approach in terms of  $p$  value when one aims at differentiating the set of MSEs obtained in the attentional tunneling state from the set of MSEs obtained in the nominal state. It should be noted that CG versions can be replaced not only for the MSE but also for other variants.

**Keywords:** MSE; linear-phase filter; coarse-grained; entropy rate



**Citation:** Grivel, E.; Berthelot, B.; Colin, G.; Legrand, P.; Ibanez, V. Benefits of Zero-Phase or Linear Phase Filters to Design Multiscale Entropy: Theory and Application. *Entropy* **2024**, *26*, 332. <https://doi.org/10.3390/e26040332>

Academic Editor: José María Amigó

Received: 25 December 2023

Revised: 16 March 2024

Accepted: 6 April 2024

Published: 14 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In statistical signal processing, a standard processing chain consists of extracting some markers from the data initially collected with sensors. These markers make it possible to characterize the signal samples, which sometimes describe a physical phenomenon. For instance, in biomedical applications, the power in some frequency bands can be representative of the activity of the sympathetic or parasympathetic nervous system or the activity of the brain. As an alternative, the practitioner can search for an a priori model to represent the data using a small set of parameters. Thus, according to the Wold decomposition, moving average (MA), autoregressive (AR), ARMA, fractionally integrated (FI), and also ARFIMA processes can be considered. They can be seen as the filtering of a white noise. The first ones allow the short-memory processes to be modeled, while the last ones are relevant for long-memory processes. Discrete fractional Gaussian noise (dfGn) and discrete fractional Brownian motion (dfBm) can be also of interest. Sums of complex exponentials can be also well suited to some signals whose spectrum is discrete. Whatever the model, the model parameters are related to the power spectral density (PSD) and consequently to the correlation function. Other markers are used. For instance, they are related to the regularity and the self-similarity property of the signal under study. A time-continuous process  $y(t)$  is said to be self-similar with the Hurst parameter  $H$  if and only if:

$$x(Nt) \stackrel{d}{=} N^H x(t) \quad (1)$$

where  $x(Nt)$  is the timescaled signal by a scale factor  $N$ , and  $\stackrel{d}{=}$  means an equivalency in terms of the distribution. In that case, the Hurst exponent can be estimated by using different approaches such as the fluctuation analysis, the detrended fluctuation analysis, or some variants. See for instance [1–5].

In this paper, we propose to focus our attention on another marker, namely the sample entropy (SE or *Sampen*). In the first appendix, we recall some successive steps that led to the SE. For a vector  $\underline{x}_1$  storing the samples  $\{x_n\}_{n=1,\dots,N}$  of the signal  $x$ , the SE can be expressed as

$$SE(\underline{x}_1, \tau, m, r) = -\log \frac{P_r^{m+1} \left( \begin{bmatrix} x_1 & x_{1+\tau} \dots & x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau} \end{bmatrix} \right)}{P_r^m \left( \begin{bmatrix} x_1 & x_{1+\tau} \dots & x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau} \end{bmatrix} \right)} \quad (2)$$

where  $\lfloor \cdot \rfloor$  is the floor function. In addition,  $P_r^m \left( \begin{bmatrix} x_1 & x_{1+\tau} \dots & x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau} \end{bmatrix} \right)$  is the estimated probability that the distance between two different vectors storing  $m$  consecutive values of the sample set is smaller than the tolerance level  $r$ .

Among the results that were established, the SE is known to be less sensitive to the noise than the approximate entropy (ApEn) [6]. In addition, the SEs of white noise and  $1/f$  noise were studied for different scales  $\tau$ . An analytical expression of the SE for unit variance zero-mean white noise was obtained [7,8]:

$$SE(\underline{x}_1, \tau, m, r) = -\log \left( \int_{-\infty}^{+\infty} \sqrt{\frac{\tau}{8\pi}} \left[ \operatorname{erf} \left( \frac{x+r}{\sqrt{\frac{2}{\tau}}} \right) - \operatorname{erf} \left( \frac{x-r}{\sqrt{\frac{2}{\tau}}} \right) \right] e^{-\frac{1}{2}x^2} dx \right) \quad (3)$$

where  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$  refers to the error function. Given (3), the SE of white noise monotonically decreases with  $\tau$ , whereas the SE of CG  $1/f$  noise is almost constant. In addition, the SE is sensitive to short-duration signals. Indeed, Richman mentioned that more than  $10^m$  samples are required to get a “good” estimation of  $SE(\underline{x}_1, \tau, m, r)$ .

However, distinguishing the interbeat interval time series of different diseased and healthy states is not necessarily an easy task if only a single scale of the signal is considered. In addition, there is no clear relationship between the entropy-based regularity and the “complexity” (it should be noted that different “complexity” measures of random sequences were proposed in the literature: the linear complexity, the maximum-order complexity, the nonlinear complexity [9], the 2-adic complexity [10], the Lempel–Ziv (LZ) complexity [11,12] in which the signal is converted to a binary sequence by comparing the signal with threshold(s) determined by CG methods (the mean, the median, the midpoint, or the  $k$  means), some variants of the LZ complexity such as its extension to the multiscale [13], the permutation LZ complexity [14] and its extensions to the multiscale [15,16], the dispersion LZ complexity and its extension to the multiscale analysis [17], the eigen complexity [18,19], the statistical complexity based on the Rényi entropy [20], and the T complexity. Moreover, some studies were conducted to define the relationship between complexity measures and Shannon entropy. See for instance [18]). Indeed, when the above quantity was directly used with physiologic signals and more particularly heartbeat interval series, larger values were obtained from pathological subjects than healthier ones. However, the contrary was rather expected by the experts, as the complexity is related to the ability of living systems to adapt themselves to the situation. Therefore, different scales of the signal were considered [7]. This led to the multiscale entropy (MSE) [7]. It has been used in a wide range of applications and more particularly in biomedical applications, traffic time series, real vibration data, diagnosing rotating machinery faults, and financial markets. Different families of MSE have been proposed for the last decades. The reader can refer to the second appendix to get more details.

The standard version of the MSE consists of estimating the SE of the process at various “scales”. These sequences are obtained from a coarse-graining (CG) process. Initially, a

CG process is known to be a way to obtain robust information of a dynamical system by mapping original signals into symbol sequences. The mapping can be based on the rounding, min, max, midpoint, and averaging operators. When dealing with the latter and from a signal processing point of view, the signal is filtered by an averaging finite impulse response (FIR) causal filter whose order defines the scale. The frequency response of the filter is low-pass, and so all the more as the scale is high. Then, the filter output is decimated by a factor equal to the scale (in [21], the time-shift MSE consists of computing the SE of the original data but also the mean of the SEs of the  $\tau$  sequences deduced from the data decimated by a factor  $\tau$ , with  $\tau$  chosen up to a maximum value. Although the author motivates the design of this variant of the MSE by drawing inspiration from Higuchi's fractal dimension, it corresponds to the composite MSE (cMSE) [22] without the averaging step (See (A3)). Once again, the consequence of the decimation is not taken into account). Using this CG process could be a priori relevant for multifractal analysis or when dealing with processes such as dfGn whose normalized covariance function remains the same after CG. However, the signals processed in practice are unlikely to have these properties. In [23], Zhang combines CG and SE in order to introduce the concept of the complexity measure, which is between regular and completely disordered data. His aim was to show that the sum of the sample entropies evaluated on the signal and their CG versions is maximized when the signal that is studied is a  $1/f$  noise signal. As mentioned in [24–26], the decimation may be problematic if the design of the filter is not done properly. Indeed, the Shannon theorem should be satisfied for each scale. Although aliasing is a priori avoided in [25–27] by using a Butterworth filter with a well-suited cutoff frequency, the phase distortions induced by the filter in the pass band can be a source of problems. In this paper, we propose to consider two types of solutions: the first one consists of designing a linear-phase FIR filter. The window method or the Remez algorithm can be used, and we will see which method is the most relevant. The second one is based on an infinite-impulse filter (IIR)-based structure, which leads to a null-phase equivalent filter. Three studies are then conducted: one based on a sum of sinusoids, a second one on white and pink noises, and the last one on real data in order to detect the tunneling of aircraft pilots, which corresponds to a cognitive state during which the pilots are unable to detect visual alarms because they are focused on the task to be done. We will see that our approach can be more relevant than the standard CG-based MSE.

The remainder of this paper is organized as follows. In Section 2, the MSE based on the linear-phase or null-phase filter is first introduced. In Section 3, illustrations are given with white noise and  $1/f$  noise. More particularly, we present how the measure evolves with the scale based on a certain number of trials. We also show the relevance of our approach with a sum of sinusoids. As the SE is defined from the ratio of two probabilities, we propose to better understand the link between both probabilities. Finally, the application on real data is proposed. It should be noted that this paper is an extension of a conference paper [28].

## 2. MSE Based on a Linear-Phase or Null-Phase Low-Pass Filter Well Suited for a Post Decimation Step

In this section, after introducing the MSE from the signal processing point of view, the alternative to the CG procedure is presented.

### 2.1. Multiscale Entropy: A Measure of Complexity

Twenty years ago, the MSE was proposed by Costa et al. [7] to evaluate the complexity of a signal. It consists of summing the SE of the signal itself but also the SEs of the time series derived from a CG procedure applied on the signal at different scales. The  $\left\lfloor \frac{N}{\tau} \right\rfloor$  samples  $y_{n,k}^{(\tau)}$  are deduced by computing the following arithmetic mean of the samples of the signal  $x$ :

$$y_{n,k}^{(\tau)} = \frac{1}{\tau} \sum_{i=(n-1)\tau+k}^{n\tau+k-1} x_i \quad \text{for } k = 1, \dots, \tau \quad (4)$$

It should be noted that  $y_{n,1}^{(1)} = x_n$ .

Applying the CG on the signal  $x$  with a factor  $\tau$  amounts to constructing the vector  $\underline{y}_k^{(\tau)}$  and stacking the values  $\{y_{n,k}^{(\tau)}\}$  with  $n \in \{1, \dots, \lfloor \frac{N}{\tau} \rfloor\}$ .

From a signal processing point of view, the following two steps define the CG:

**Step 1: Applying on the signal  $x$  an averaging filter** defined by a causal FIR  $h_n^{(\tau)} = \frac{1}{\tau}$  for  $n = 0, \dots, \tau - 1$  (with  $\tau \leq \tau_{max}$  being the maximum scale) and zero elsewhere. The resulting transfer function  $H(z)$ , which is the  $z$  transform of  $h_n^{(\tau)}$ , satisfies

$$H^{(\tau)}(z) = \begin{cases} \frac{1}{\tau} \sum_{n=0}^{\tau-1} z^{-n} = \frac{1}{\tau} \frac{1-z^{-\tau}}{1-z^{-1}} & \text{if } z \neq 1 \\ 1 & \text{if } z = 1 \end{cases} \quad (5)$$

Consequently, the frequency response of this causal filter for the normalized frequency  $f \in [-\frac{1}{2}, \frac{1}{2}]$  is given by

$$|H^{(\tau)}(f)| = \begin{cases} \frac{1}{\tau} \left| \frac{\sin(\pi f \tau)}{\sin(\pi f)} \right| & \text{if } f \neq 0 \\ 1 & \text{if } f = 0 \end{cases} \quad (6)$$

It is a low-pass filter. The zeros of  $H^{(\tau)}(z)$  are equal to  $z = e^{j2\pi \frac{k}{\tau}}$ , with  $k \in \{0, \dots, \tau - 1\}$ , and are located on the unit circle in the  $z$  plane. This means that the frequency response of the filter totally rejects the normalized frequencies  $\frac{k}{\tau}$  for  $k \in \{0, \dots, \tau - 1\}$ . Finally, due to the symmetry of the impulse filter, the filtering has the advantage of having a linear phase, thereby leading to a constant group delay and no phase distortion of the signal in the pass band. Only the steady state is considered. In other words, the first  $\tau - 1$  samples of the output filter are not considered.

**Step 2: Decimating by a factor  $\tau$  the filter output:** The samples whose indices are multiples of  $\tau$  are kept for  $n \in \{0, \dots, \lfloor \frac{N}{\tau} \rfloor - 1\}$  and  $k \in \{0, \dots, \tau - 1\}$ :

$$y_{n,k}^{(\tau)} = y_{n\tau+k} \quad (7)$$

In the following,  $\underline{y}_k^{(\tau)}$  is the vector storing the values  $y_{n,k}^{(\tau)}$ , with  $n \in \{0, \dots, \lfloor \frac{N}{\tau} \rfloor - 1\}$ . It should be noted that in the frequency domain, due to the decimation, the normalized frequencies above  $\frac{1}{2\tau}$  will be a source of aliasing. As the frequency responses of the filters defined above have their main lobes between  $-\frac{1}{\tau}$  and  $\frac{1}{\tau}$  and the side lobes are not necessarily much attenuated (See (6)), the filters are not well suited to be antialiasing filters. Hence, this can be a source of problems, as reported by Valencia [25], Humeau [24], or more recently Zhao [26].

**Step 3: Computing the SE for each scale and summing them:** the SE is computed for each scale, and the MSE is defined by the sum  $\sum_{\tau=1}^{\tau_{max}} SE(\underline{y}_1^{(\tau)}, 1, m, r)$ . This can be rewritten as

$$MSE(\underline{x}_1, \tau_{max}, m, r) = - \sum_{\tau=1}^{\tau_{max}} \log \frac{P_r^{m+1}(\underline{y}_1^{(\tau)})}{P_r^m(\underline{y}_1^{(\tau)})} \quad (8)$$

Note that only one decimated sequence, namely  $\underline{y}_k^{(\tau)}$  with  $k = 1$ , is used in the standard MSE. In addition, one of the problems of the MSE algorithm is the use of the same tolerance level  $r$  for all scales [29].

## 2.2. Our Theoretical Contribution

We propose to derive an MSE by taking care of the filtering step. The first step of the MSE described above is hence replaced by the following step, while the others remain unchanged:

**New step 1: Applying on the signal  $x$  a linear-phase or null-phase low pass filter.** For each scale  $\tau \in [2, \dots, \tau_{max}]$ , the goal is to select a low-pass causal FIR filter with a linear phase or a null phase and a normalized cutoff frequency that is smaller than  $\frac{1}{2\tau}$ .

This can be done by using the window method for a causal linear-phase FIR filter design or Remez algorithm.

**About the window method:** It operates with the following steps: defining the specifications of the low-pass digital filter  $H(f)$ , taking the inverse Fourier transform, and windowing the resulting impulse response to obtain the FIR of the low-pass filter.

**About the Remez algorithm:** Also known as the Parks–McClellan optimal equiripple FIR filter design, the Remez algorithm is a curve fitting method proposed in the 1970s minimizing the error between the actual frequency response and the frequency response of the designed filter. The starting point of the approach is that the frequency response of a filter whose impulse response is symmetric and whose order is odd can be expressed as a linear combination of  $\cos(2\pi n f)$  or equivalently as a linear combination of Chebyshev polynomials of the first kind in  $\cos(2\pi f)$ . Therefore, this frequency response can be modeled as a polynomial in  $\cos(\theta)$ . The next step is to search the coefficients of the polynomial that best approximate this frequency response so that the error between both frequency responses in the pass and stop bands are minimized. Using the alternation theorem (a polynomial fit of degree  $n$  to a set of bounded points is said to be minimax if and only if it attains its maximal error at  $n + 2$  points with alternating signs), the normalized frequencies leading to the maximal errors and the polynomial coefficients are estimated in an iterative way. Usually less than 15 iterations are required.

An alternative is to design an IIR filter using the bilinear transform and then to design a zero-phase linear filter. Two types of processing chains can be considered:

**First strategy:** The signal  $x_n$  is filtered by a filter whose real impulse response is  $h_n$ , thus leading to the signal  $g_n$ . Then, the time-reversed version  $g_{-n}$  is filtered by a filter whose impulse response is still equal to  $h_n$ . The output is denoted as  $r_n$ . Finally,  $r_n$  is time reversed to obtain the filtered signal of interest, which is denoted as  $y_n$ . Let  $H(f) = |H(f)| \exp(j\Phi(f)) = H^*(-f)$ ; one has

$$\begin{cases} G(f) = H(f)X(f) \\ R(f) = H(f)G(-f) \\ Y(f) = R(-f) = |H(f)|^2 X(f) \end{cases} \quad (9)$$

The Fourier transform of the impulse response of the equivalent filter is given by

$$H_{equ}(f) = |H_{equ}(f)| \exp(j\Phi_{equ}(f)) = |H(f)|^2 \quad (10)$$

It is hence a null-phase filter ( $\Phi_{equ}(f) = 0$ ) characterized by  $|H_{equ}(f)| = |H(f)|^2$ .

**Second strategy:** The sequences  $x_n$  and  $x_{-n}$  are filtered by the same filter whose impulse response is  $h_n$ . This leads to two sequences respectively denoted as  $g_n$  and  $r_n$ . The final output is equal to  $g_n + r_{-n}$ . In that case, the Fourier transform of the output is  $Y(f) = 2|H(f)|^2 \cos(\Phi(f))$ . It is a zero-phase filter if  $\Phi(f) \in [0, \frac{\pi}{2}]$  in the pass band.

As there is no guarantee this is always the case, only the first strategy is considered. Note that the filter can be either FIR or IIR. For IIR, and without being exhaustive, different filter families exist such as the Butterworth filters, whose order can be large due to the constraint of a maximally flat magnitude in the pass band, and the Chebyshev filters, which are known to exhibit equiripple either in the pass band (type I) or in the stop band (type II).

The specifications of the low-pass filter frequency response have to be defined. It consists of choosing the transition bandwidth  $\Delta f = f_s - f_p$ , where  $f_s$  and  $f_p$  respectively

denote the normalized stop-band corner frequency and the normalized pass-band corner frequency. For the window method and Remez approach,  $f_p$  and  $f_s$  are defined as  $\frac{1}{2\tau} - \Delta f/2$  and  $\frac{1}{2\tau} + \Delta f/2$ , respectively. For the null-phase filter,  $f_p$  is defined as  $\frac{1}{2\tau} + 10^{-3}$  and  $f_s$  as  $f_p + \Delta f$ . Moreover, the maximum permissible pass-band loss  $R_p$ , as well as the stop-band attenuation  $R_s$ , have to be defined.

In the Section 3, three applications of the MSE and our proposed variants are presented.

### 3. Applications

#### 3.1. Application on the Sum of Sinusoids: Interpretation and Comments

This section aims to analyze how the SE evolves with the scale  $\tau$  to better understand the results one obtains when applying the MSE on basic signals. As recalled in (2),  $SE(x_1, \tau, m, r)$  depends on the ratio between  $P_r^{m+1}([x_1 \ x_{1+\tau} \dots \ x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau}])$  and  $P_r^m([x_1 \ x_{1+\tau} \dots \ x_{1+\lfloor \frac{N-1}{\tau} \rfloor \tau}])$

In the following and more particularly in the Figures 1a,b, 7 and 8, we propose to address the computation of these probabilities and consequently the SE through a geometric approach by considering a 2D representation: The pixel located at the  $i$ th column and the  $j$ th row corresponds to the Chebyshev distance between the  $i$ th and  $j$ th vectors storing  $m$  consecutive samples of the data. As an example, when  $\tau = 1$  and  $m = 2$ , the pixel located at the 2nd column and the 5th row corresponds to the Chebyshev distance between the vector storing the samples  $x_2$  and  $x_3$  and the vector storing the samples  $x_5$  and  $x_6$ . When the distance is smaller than  $r$ , the pixel is yellow, whereas it is turquoise when it is greater. The blue pixels correspond to the case when one vector is compared with itself (hence, this is the antidiagonal in the image. This case is not taken into account when the SE is computed. This is the reason why the corresponding pixels have another color). Computing the probabilities amounts to counting the number of yellow pixels. In that case, if  $d_{i,j}^{(m)}$  denotes the Chebyshev distance between the  $i$ th and  $j$ th vectors of dimension  $m$ , one has

$$d_{i,j}^{(m+1)} = \max(d_{i,j}^{(m)}, d_{i+1,j+1}^{(m)}) \quad (11)$$

This amounts to saying that

$$d_{i,j}^{(m+1)} < r \iff d_{i,j}^{(m)} < r \text{ and } d_{i+1,j+1}^{(m)} < r \quad (12)$$

Therefore, due to (12), a pixel whose coordinates are  $(i, j)$  is yellow when the size of the vectors is equal to  $m + 1$  if and only if both pixels at coordinates  $(i, j)$  and  $(i + 1, j + 1)$  are yellow when the size of the vector is equal to  $m$ .

In the rest of this subsection, three cases are studied. Moreover, three approaches are used: the standard CG-based MSE, the variant based on a Chebyshev filter, and the variant we propose based on the null-phase filter.

**First case:** Let us first look at the case of  $N = 400$  samples of one sinusoid defined as follows:

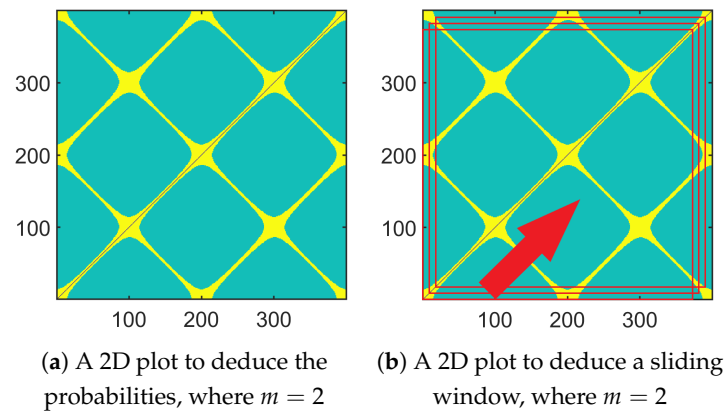
$$x_k = 2 \cos(2\pi f_1 k) \quad (13)$$

with the normalized frequency  $f_1 = 0.005$ . Due to (13), the minimum and maximum values of the signal under study are obtained when  $k$  is a multiple of 100.

Let us now look at the corresponding 2D representation of the Chebyshev distances in Figure 1a, where  $m = 2$  and  $\tau = 1$ . In that case,  $N - m + 1$  vectors of size  $m$  have to be compared. Moreover, the abscissa and the ordinate in the 2D plot can vary between 1 and  $N - m + 1 = 399$ . The antidiagonal necessarily corresponds to blue pixels. The pixels located along the main antidiagonal ( $i = j$ ) are yellow because the sets of vectors to be compared can have samples in common. This phenomenon is retrieved along the sub-antidiagonals, whose index is a multiple of  $\frac{1}{f_1}$  (i.e.,  $j = \pm 200k + i$  with  $k = 0, 1, \dots$ ). Due to the symmetry of a sinusoid around its minima and maxima, one can also observe yellow straight lines that are orthogonal to these antidiagonals (i.e., when  $j = 200k - i$ ) in

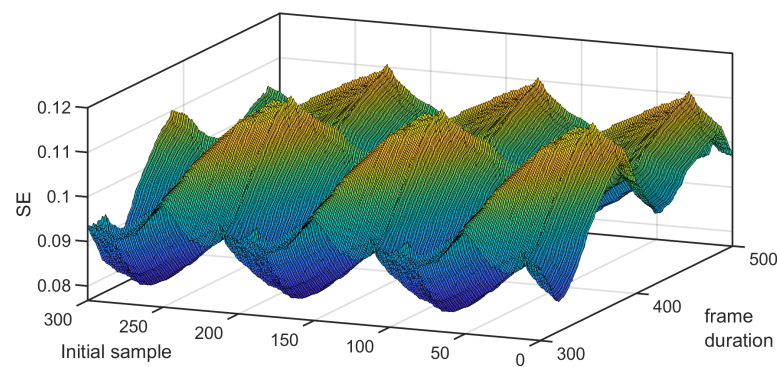


the 2D plot. The intersections correspond to the locations of the minima and maxima of the sinusoid. The rest of the pixels in Figure 1a are turquoise, since the Chebyshev distance is smaller than the threshold  $r$ .

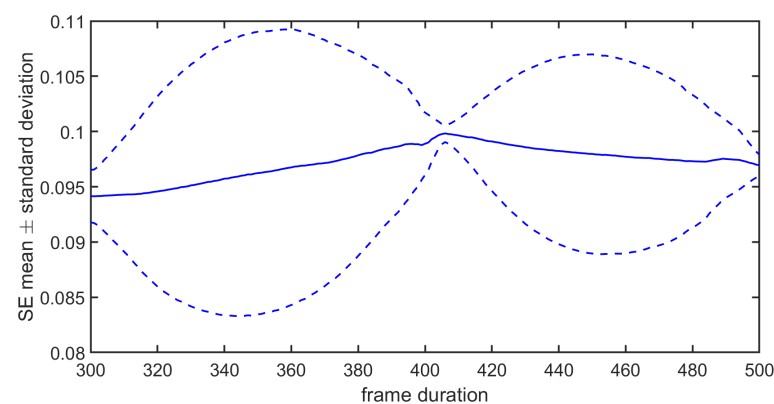


**Figure 1.** Way to compute the probability based on a geometric interpretation when using the standard SE (a) and when using the SE with a sliding window (b). The abscissa and ordinate correspond to the indices of the vectors of size  $m$  to be compared.

*Remark on the sensitivity of the SE to the set of samples that are used:* The value of the SE was sensitive to the number of samples available and/or to the initial sample. This is due to the fact the probabilities  $P_{\tau}^m$  and  $P_{\tau}^{m+1}$  (or similarly the numbers of yellow pixels for  $m = 2$  and  $m = 3$ ) can change when the number of samples is modified and/or when the initial sample changes. In Figure 2, one can see that the SE varied with the number of samples available (from 300 to 500) and the first sample that was considered (from 1 to 300). We have the mean and the standard deviation of the SE computed on the frames for a given length  $N$  but whose initial sample changes are shown in Figure 3. One can notice that both statistics can vary considerably. Therefore, to make the computation of the SE less sensitive to the first sample, we suggest using a sliding window, computing the SE for each frame, and averaging them. An illustration is given in Figure 1b, where the window length was equal to  $N_w = 380$ . As  $N_w < N$ , the resulting 2D plot is depicted by a red square. When the analysis window was shifted, the red square moved. In that case, the computational cost can be optimized. Indeed, by taking into account the computation of the probability using the 2D plot, one can just consider the upper triangular part (without the main antidiagonal). This provides the cardinal of the sample space, i.e.,  $\frac{(N_w - m + 1)(N_w - m)}{2}$ . When the analysis window was shifted by one sample, this amounted to shifting the upper triangular part corresponding to the last  $N_w - m - 1$  columns and the first  $N_w - m - 1$  rows to the left and down. Only  $N_w - m$  new distances must be computed to deduce the event space. As the number  $N$  of samples available is assumed to remain the same, there may be  $N - N_w + 1$  SEs that can be computed and averaged. The resulting computation cost is equal to  $\frac{N_w - m}{2}(2N - N_w - m + 3)$ . The number of SEs to be computed for the averaging, as well as the time shift, are chosen by the practitioner. It should be noted that this averaging is different from the cMSE. See (A3). Indeed, for the scale  $\tau = 1$ , the SE computed in the cMSE is the SE, whereas it is an average in this variant. In the cMSE, if  $\tau = 2$  and  $N$  is even, there are  $\frac{N}{2}$  samples in the first decimated filter output and  $\frac{N}{2} - 1$  samples in the second one to compute the SE and average them. In this variant, the number of samples per decimated sequence is always the same. The way to compute the probabilities for a given time shift (that is not equal to 1 but that is larger) is given in Figure 1b.



**Figure 2.** SE computed for the signal under study for different numbers of samples available and different initial values.



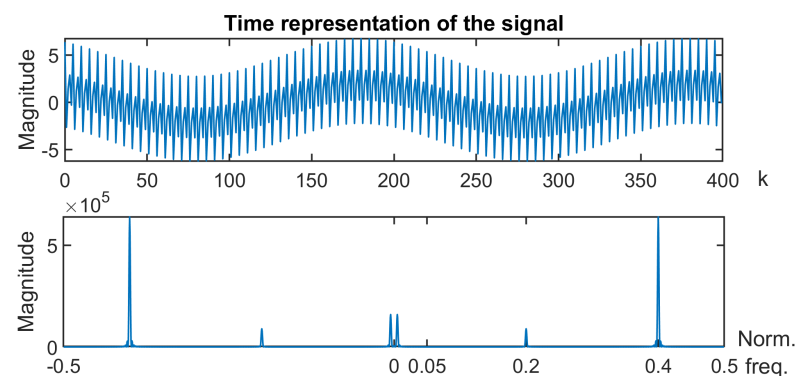
**Figure 3.** Mean  $\pm$  its standard deviation of SE for different frame lengths.

The above suggestion is not applied in the rest of this subsection in order to analyze each phenomenon separately.

**Second case:** Let us now address the case of  $N = 400$  samples of a sum of three sinusoids defined as follows:

$$x_k = 2 \cos(2\pi f_1 k + \frac{\pi}{5}) + 1.5 \cos(2\pi f_2 k + \frac{\pi}{3}) + 4 \cos(2\pi f_3 k) \quad (14)$$

with normalized frequencies  $f_2 = 0.2$  and  $f_3 = 0.4$ . The time representation of the signal and its spectrum are given in Figure 4.

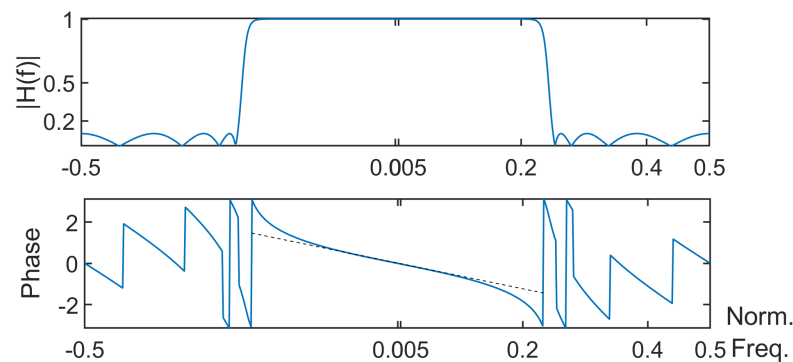


**Figure 4.** Time representation and spectrum of the signal under study.

Moreover, the scale  $\tau$  was set at 2, and the frequency response of the type-II Chebyshev filter is given in Figure 5. One can notice that the filter does not have a linear phase. Indeed,

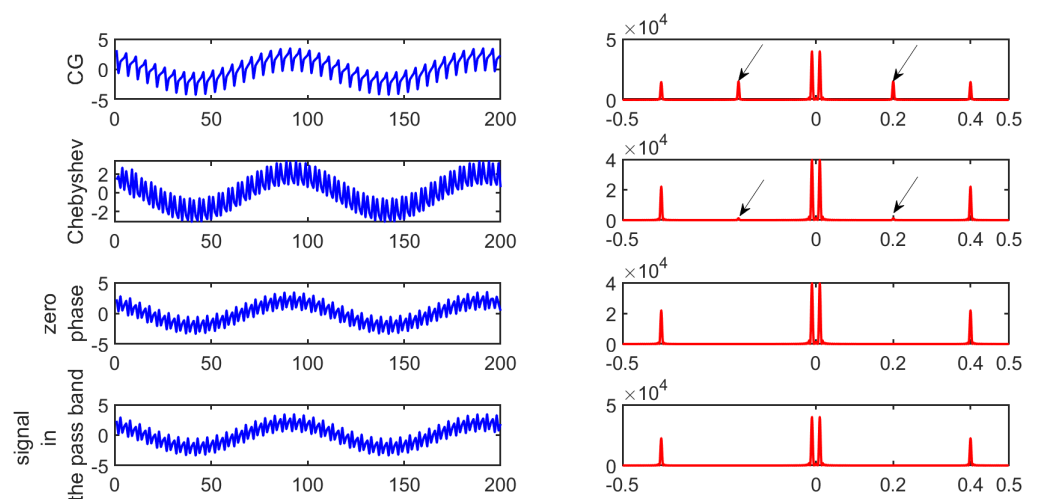


the phase curve should follow the black dashed line when the normalized frequency is in the interval  $[-0.22, 0.22]$ .



**Figure 5.** Frequency response of the Chebyshev filter: modulus and argument of  $H(f)$ .

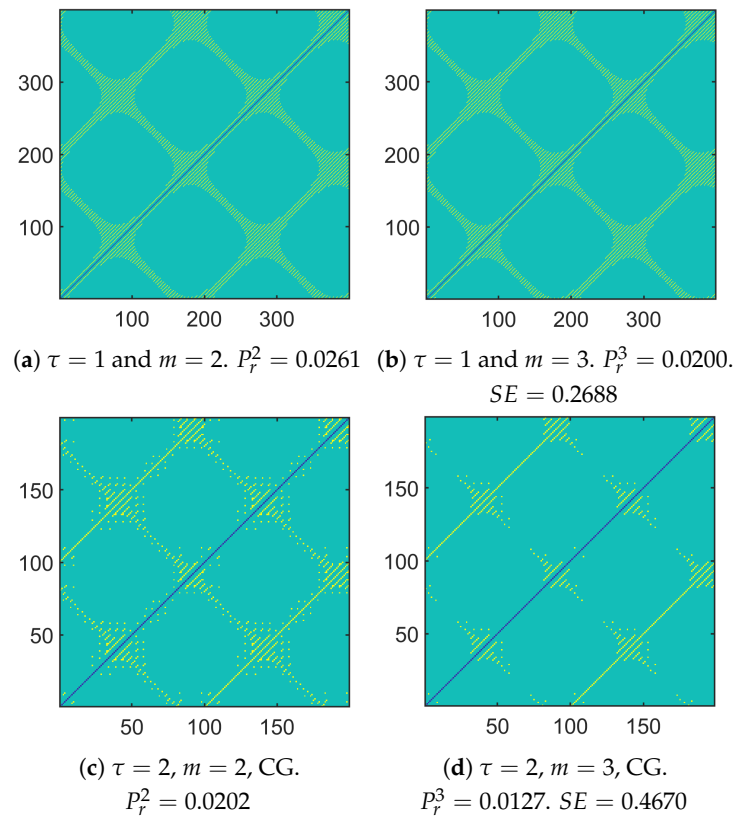
When applying the CG case, aliasing occurred. When considering a low-pass filter whose cutoff frequency is well suited, this should be no longer the case in theory. However, a filtering- phase distortion occurred. Moreover, depending on the level of attenuation in the stop band, there may be a small amount of aliasing due to the decimation that follows. When using a zero-phase filter based approach, the above phenomena were greatly alleviated, as shown in Figure 6.



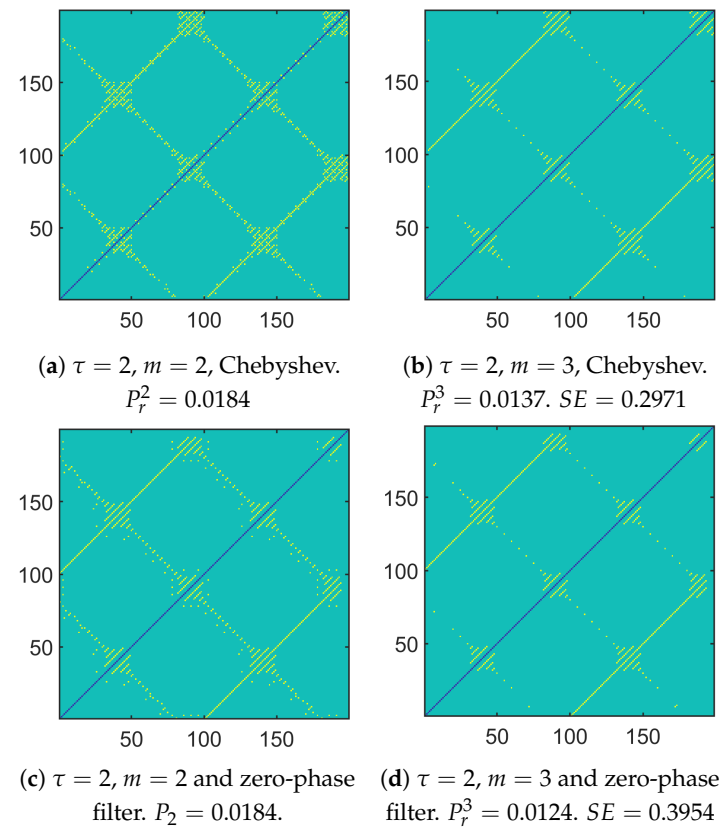
**Figure 6.** Time representation and spectrum of the signal at the scale  $\tau$  equal to 2 when using CG, a Chebyshev filter, or the zero-phase filter based approach with their corresponding spectrum. The resulting signals are compared with the part of the signal located in the pass band.

When looking at Figure 7a with  $\tau = 1$ , one can notice the influence of the phase at the origin of the component whose normalized frequency is  $f_1$ . There was a global shift down and to the left when comparing Figure 7a with Figure 1a. Moreover, the two components whose normalized frequencies are  $f_2$  and  $f_3$  had an impact on the 2D plot by modifying the initial pattern presented in Figure 1a. Depending on  $f_2$  and  $f_3$  and the magnitude of both components, only segments of yellow straight lines remained in the 2D plot. The gap between them can be related to  $\frac{1}{f_2}$  and/or  $\frac{1}{f_3}$ .

Let us now compare Figures 7c,d and 8, which provide the 2D plots of the vector distances when applying a CG, the Chebyshev filter, and our approach, respectively. One can see how the number of yellow pixels changed. Some yellow pixels disappeared when applying our approach, especially in the upper right corner. As a consequence, depending on the approach that is considered, there may be some differences in the probabilities and consequently in the SEs that are obtained.



**Figure 7.** 2D plot to deduce the probabilities and then computing the SE: CG-based approach.



**Figure 8.** A 2D plot to deduce the probabilities and then computing the SE: Chebyshev-based approach and zero-phase filter-based approach.

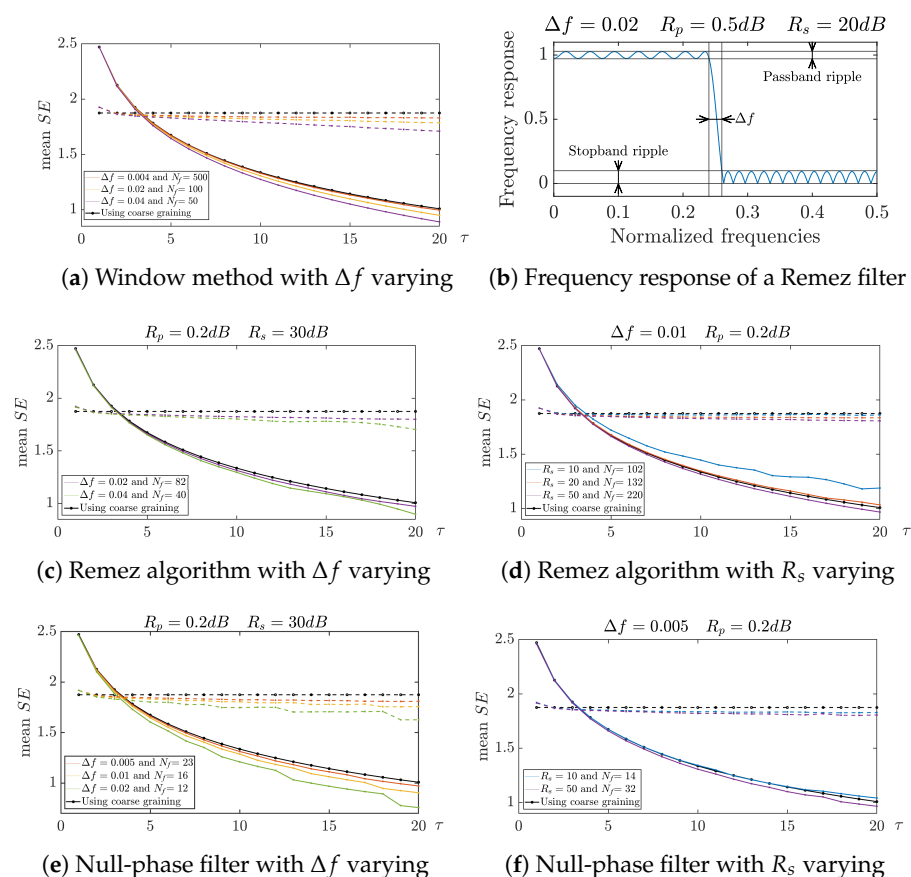
### 3.2. Application on White Noises and $1/f$ Process

In this section, we propose to follow the same methodology as the one used in [7,25] by providing the way the entropy measure evolves with the scale for white noises and  $1/f$  noise based on a certain number of trials.

**Remark 1.** Some authors study how entropies evolve with respect to the scale. Thus, the multiscale PE was applied to dfGn [30]. According to the analysis made by the authors, the PE value of a dfGn was shown to be invariant to the time scale. In [31], the authors analyzed how the  $K_2$  entropy evolves with the scale for dfGn. They showed that it can be approximated by an affine function of the logarithm of the scale, whose slope is equal to  $H - 1$ , with  $H$  being the Hurst exponent measuring the long-term memory of the time series.

As the approaches we propose depend on the specifications of the low-pass filter, we looked at the way the curves varied with one of the specification parameters. They were compared with the curves obtained through the standard MSE [7].

Figure 9,  $2N_f + 1$  denotes the length of the impulse response for the window method (Figure 9a) and the Remez algorithm (Figure 9c,d), whereas it denotes the order of the type-II Chebyshev filter for the null-phase filter structure (Figure 9e,f). The approaches were performed on 50 signals of length 30,000. The results were then averaged to obtain the curves.



**Figure 9.** Evolution of the SE according to  $\tau$  for different choices of  $\Delta f$  and  $R_s$ . White noise: —. Pink noise: -- [28].

According to our simulations, when  $R_p$  was modified in the design of the IIR filter, it did not change the shape of the curve much. For this reason, we do not present it in Figure 9. When  $R_s$  was taken too small, the linear-phase or null-phase low-pass filter did not filter the frequencies in the stop band enough. Therefore, like the standard MSE, spectrum overlapping may occur. The resulting evolution of the entropy measure tended to

the one obtained with the standard MSE. When the transition bandwidth became too wide, the stop band became smaller. Some frequencies in the transition band were not attenuated enough. Therefore, the resulting curve did not differ much from the curve obtained with the standard CG-based approach.

The reader should not conclude that obtaining curves that are different from the one obtained with CG is bad. Rather, our purpose is to show how the curve is modified when using our approach.

### 3.3. Application on Real Data to Detect Attentional Tunneling

In this section, the CG-based MSE and the new variant based on antialiasing linear-phase or null-phase filtering has been applied on real data. The goal is to detect a degraded physiological and/or cognitive state of an aircraft pilot or crew, such as attentional tunneling. It is defined as “the allocation of attention to a particular channel of information, diagnostic hypothesis or task goal, for a duration that is longer than optimal, given the expected cost of neglecting events on other channels, failing to consider other hypotheses, or failing to perform other tasks” [32]. The risk associated with this state is that the pilot may neglect new crucial information.

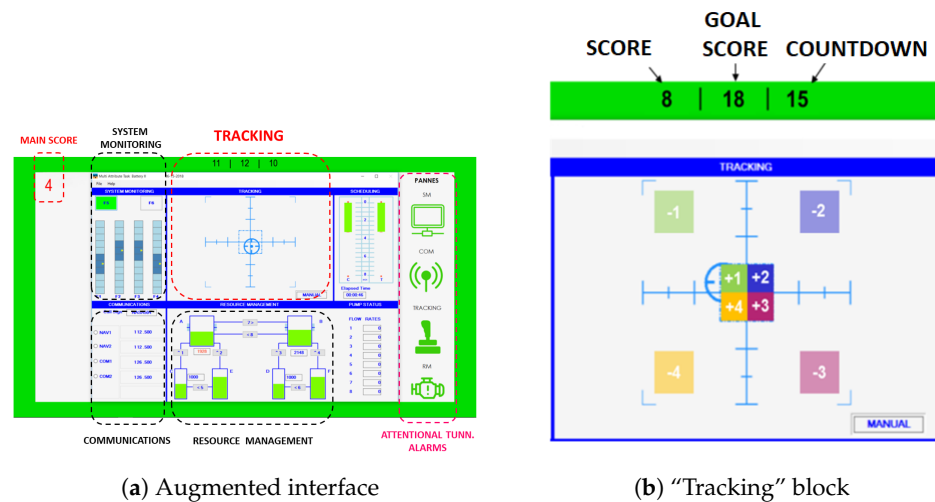
This attentional tunneling, which is responsible for almost all fatal controlled flight into terrain (CFIT) accidents [33], can be distinguished by two main variants: inattentive blindness and inattentive deafness [34,35]. Readers can refer to [36] for an in-depth study of attentional tunneling in the aeronautical domain. Inattentive blindness can be defined as an individual’s tendency to not perceive changes in a visual scene [37]. This state can occur in situations of high cognitive demand [38] and for a subject heavily engaged in a task [39]. Various physiological measures have been proposed to characterize this state using electroencephalogram (EEG) [40] and electrocardiogram (ECG) [41] signals. In this paper, the study is about ocular signals for two main reasons:

- Ocular data have proven to be relevant to characterize tunneling [42]. It has been shown that this state is correlated with a reduction in the number of eye movements (saccades) and an increase in fixation duration.
- Ocular data are collected from cameras or eyetrackers, which are sensors that can be accepted by a user in ecological conditions, unlike electrodes for example. This facilitates the integration of a visual tunneling monitoring solution.

The rest of this section is organized as follows: The protocol to collect the data is first presented. Then, the different methods are used for classification. We will see if the new variant of the MSE is more relevant or not.

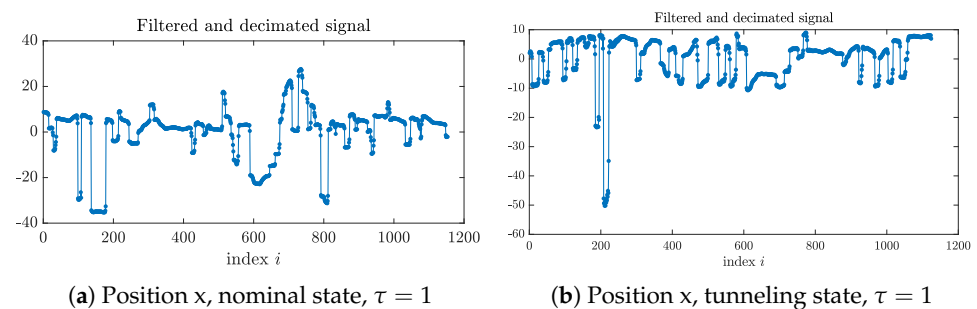
#### 3.3.1. Information about the Protocol

Let us briefly describe the protocol. Due to its similarities with the activities of a pilot in a cockpit, the NASA MATB-II software was considered as a multitask simulator [43]. As no previous knowledge of piloting is required to use the MATB-II software, any subject with no previous piloting experience can participate in the experiment. It consists of different blocks: system monitoring, communications, resource management, and tracking. Among the various tasks that can be done, the “tracking” task is considered as the primary one, because continuous involvement is required compared with the others. To facilitate the appearance of tunneling, a scoring interface was coded and added above the “tracking” task, thus leading to an augmented interface. See Figure 10a. The user must obtain the “goal score” by placing the cursor in a specific area of the task block and clicking the joystick trigger. Four areas near the center increase the score (from one to four points), while four others (in the periphery) decrease it (from one to four points). The user must reach the goal in fifteen seconds to increment the main score (located at the top left of the interface in Figure 10b) by one point. When the user achieves it, a new goal score is proposed. Otherwise, the main score is decremented by one point and the timer is reset. See Figure 10.



**Figure 10.** Presentation of the augmented interface used for the protocol.

An eye tracker (VT2 model from the company EyeTech, with a sampling rate at 40 Hz) was used to collect ocular data, namely the horizontal and vertical position of the subject gaze on the screen denoted as "position x" and the "position y", respectively. See Figure 11. Our purpose is to study if an ocular metric, namely the SE of the gaze direction at different scales, can be used to distinguish the nominal states from the tunneling ones.



**Figure 11.** Examples of a "position x" signal.

The protocol is composed of three phases.

1. During the 30-min training phase, the subject becomes familiar with each task that can be done with the simulator. Then, the subject faces a situation combining several tasks during 4 min.
2. During the reference phase, subjects are first asked to relax and do nothing during two minutes. Then, a two-minute scenario leading to a "Low" mental workload is launched.
3. During the experimental phase, there are two parts with a short break in between. Each part is composed of a task of "Low" mental workload lasting four minutes, a four-minute scenario corresponding to "Medium" mental workload, and a final task of "High" mental workload. The order of the three tasks is randomly set.

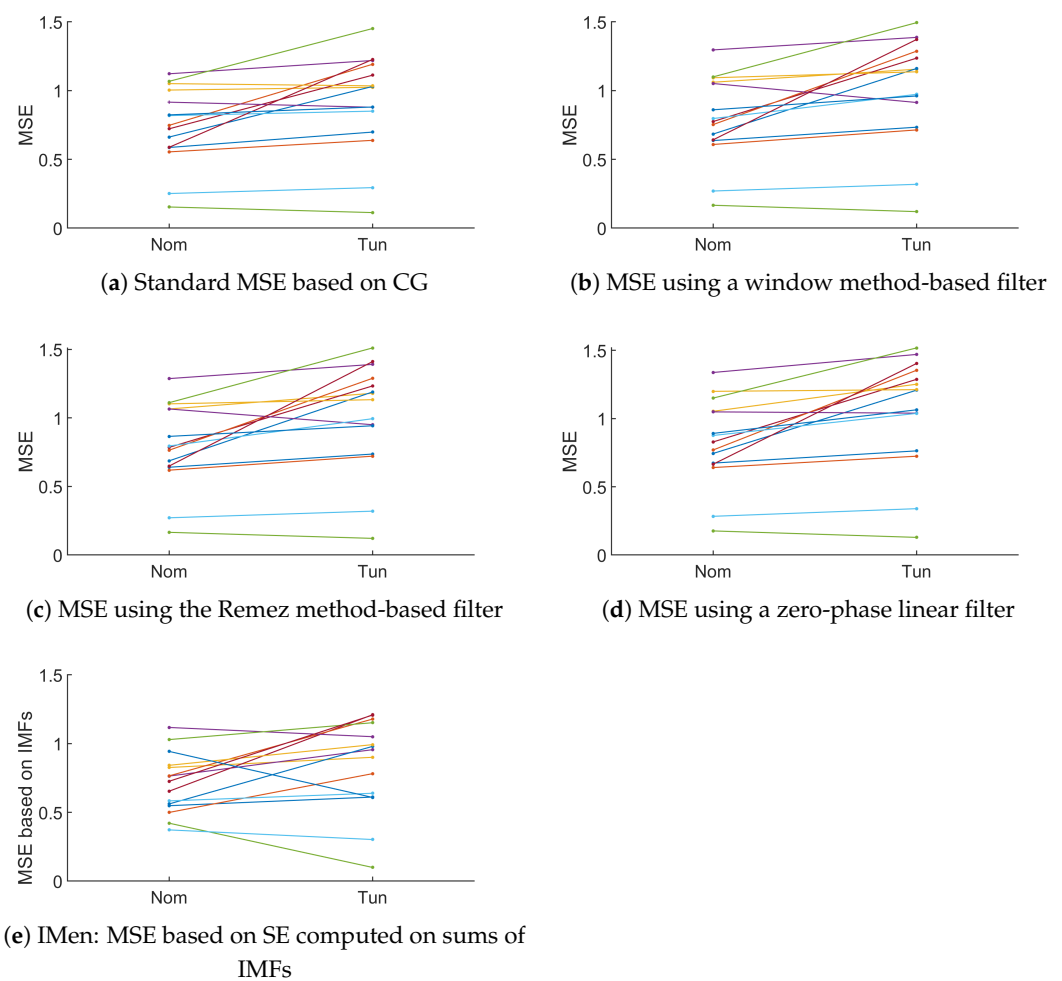
Twenty subjects filled in a NASA-TLX form [44] to subjectively evaluate their mental workload during the scenario that was proposed.

In the context of this experimentation, an occurrence of visual tunneling is detected when an alarm from the "system monitoring" task goes unnoticed, and no interaction with secondary tasks is performed for a duration exceeding twenty seconds. The data used for this analysis consist of an aggregation of performance metrics across various tasks within the NASA MatB-II software.

Among the twenty participants, five had to be removed due to technical issues. Half of the participants were observed to undergo a state of visual tunneling at least once, thus resulting in the identification of fifteen occurrences of this state. Notably, some subjects experienced visual tunneling multiple times.

### 3.3.2. Using CG-Based MSE or Its New Variant

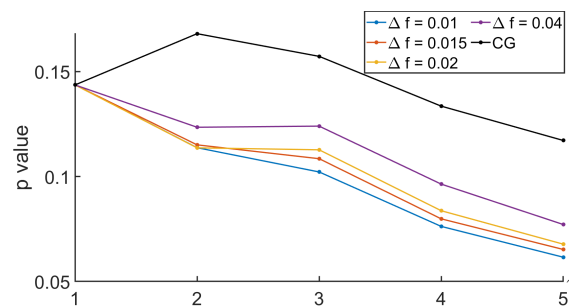
Based on the data collected for the fifteen subjects and during the different phases of the protocol, the next step was to use the CG-based MSE and three variants that we have proposed based either on the window method, Remez algorithm and the null-phase filter, or applied on the position  $x$  (i.e., the horizontal gaze position). In every case, instead of considering the standard MSE, the averaging used in the cMSE was considered. The purpose is to see whether the two classes, namely the nominal states and the tunneling ones, could be distinguished. According to Figure 12, most of the MSEs increased from the nominal state to the tunneling one (four decreased when using the standard CG, whereas two to three decreased when using the variants we proposed). Depending on the variant of the MSE, the values of the marker varied. In addition, the intrinsic mode entropy (IMen), i.e., the MSE based on the decomposition of the signal using EMD, is provided in Figure 12e. However, even when the IMen increased for some participants, it also decreased for four of them quite significantly. At first sight, it is less easy to identify a trend when IMen was used.



**Figure 12.** Evolution of the MSE from the nominal to the tunneling states using the standard MSE based on CG and the variants using different low-pass filter design with linear-phase or null-phase ( $\Delta f = 0.02$  and  $R_s = 30$  dB). Comparison with the MSE based on the IMen.  $\tau$  or number of IMS equal to 3. Each color corresponds to one participant.



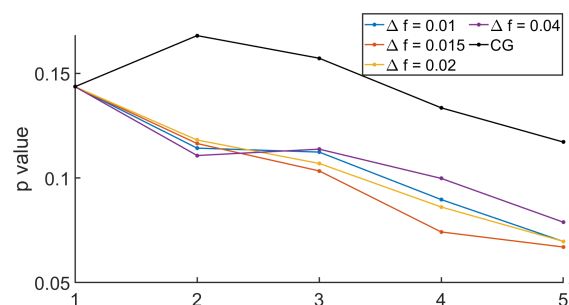
To analyze the relevance of the approaches, the comparison between the MSE computed during the nominal and the tunneling states was based on the Behrens–Fisher problem. It is a hypothesis testing dealing with the difference between the means of two normally distributed populations when the variances of the two populations are not assumed to be equal based on two independent samples. The indicator we are interested in is the  $p$  value. It is a score between 0 and 1 indicating whether or not two sets come from distributions having the same mean. Thus, a low  $p$  value indicates that there is low probability that the two sets come from equal mean distributions. Thus, a low  $p$  value indicates that the two sets can easily be differentiated. Regarding our approaches based on the null-phase filter, the Remez approach, or the window method, we proposed to see how the  $p$  value evolved with respect to some specifications of the filter such as  $\Delta f$  and  $R_s$ . See Figures 13–16.



**Figure 13.**  $p$  value when comparing MSEs between nominal and tunneling states with different values of  $\Delta f$ , with the window method filter, and with  $R_s = 30$  dB.

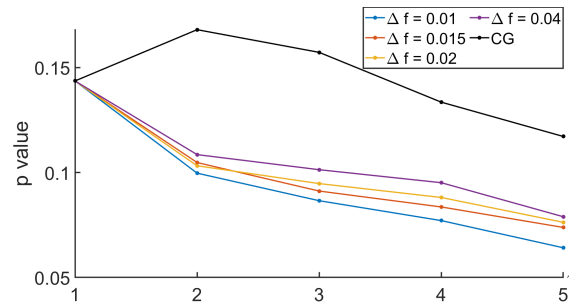
In Figure 13, one can see that the  $p$  values decreased with the scale  $\tau$ . Using the MSE up to a scale  $\tau$  equal to 5 should be better than using the MSE up to a smaller scale. In addition, when studying the influence of the specification parameters used to design the linear-phase low-pass filter with the window method, the smaller  $\Delta f$  was, the more distinct the pass band and the rejected band were and the smaller the  $p$  value was. Finally, the  $p$  values were much smaller than those obtained on this test when comparing the MSEs computed with the CG-based MSE.

In Figure 14, similar comments can be made when analyzing the way the  $p$  value evolved with the scale and  $\Delta f$ , which is one of the parameters defining the low-pass filter designed with the Remez algorithm.

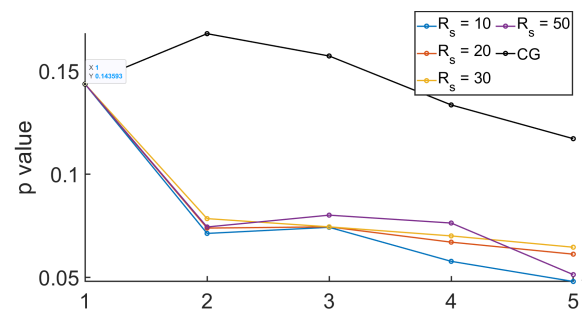


**Figure 14.**  $p$  value when comparing MSEs between nominal and tunneling states with different values of  $\Delta f$ , with Remez filter, and with  $R_s = 30$  dB.

In Figures 15 and 16, we propose to look at the sensitivity of the  $p$  value when the parameters of the filter used to get the null-phase low-pass filter were modified. Firstly,  $R_s$  was set at 30 dB, and  $\Delta f$  varied. Then,  $\Delta f$  was equal to 0.005, and different values of  $R_s$  were considered. Once again, the  $p$  value decreased with the scale for  $\tau \leq 2$ . The smaller the  $\Delta f$ , the smaller the  $p$  value. Therefore, our variant was more reliable than the MSE based on a CG.



**Figure 15.**  $p$  value when comparing MSEs between nominal and tunneling states with different values of  $\Delta f$ , with a null-phase filter, and with  $R_s = 30$  dB.



**Figure 16.**  $p$  value when comparing MSEs between nominal and tunneling states with different values of  $R_s$ , with a null-phase filter, and with  $\Delta f = 0.005$ .

#### 4. Conclusions and Perspectives

To replace the standard coarse-graining step of the MSE that induces aliasing, a linear-phase FIR filter and a null-phase structure based on an IIR filter have been used in our first variant. Following the same type of analysis as in [7,25], the sample entropies of white noise and  $1/f$  noise at different scales were studied. As the parameter of the filters could be tuned by the practitioner, our simulations confirmed that if the stop-band ripple is too small or if the transition bandwidth of the filter is too wide, the antialiasing properties of the filter become lesser. As the FIR filter order may be large, it may be less relevant when the number of samples available is low, and we suggest using the null-phase filter structure. Our proposal, based on a linear- or null-phase filter, seems moderate from the signal processing point of view, but the resulting MSE could serve a large community of users. It shows its relevance in the application dedicated to attentional tunneling.

**Author Contributions:** Conceptualization, E.G., B.B. and P.L.; methodology, E.G., B.B. and P.L.; software, E.G., G.C., B.B. and P.L.; validation, E.G., B.B. and P.L.; formal analysis, E.G., B.B. and P.L.; investigation, E.G., B.B. and P.L.; resources, E.G., B.B., P.L. and V.I.; data curation, B.B.; writing—original draft preparation, E.G., B.B. and P.L.; writing—review and editing, E.G., B.B. and P.L.; visualization, E.G., B.B., G.C. and P.L.; supervision, E.G. and P.L.; project administration, E.G., B.B., P.L. and V.I.; funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The authors would like to thank GIS Albatros for their support.

**Conflicts of Interest:** Authors Bastien Berthelot and Vincent Ibanez were employed by Thales AVS France. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AME	adaptive MSE
ApEN	approximate entropy
AR	autoregressive
ARFIMA	autoregressive fractionnaly integrated moving average
ARMA	autoregressive with moving average
CFIT	controlled flight into terrain
CG	coarse-graining
CMSE	composite MSE
COSEn	coefficient of the sample entropy
dfBm	discrete fractional Brownian motion
dfGn	discrete fractional Gaussian noise
DE	dispersion entropy
EMD	empirical mode decomposition
ECG	electrocardiogram
EEG	electroencephalogram
ER	Eckmann–Ruelle
FI	fractionally integrated
FIR	finite impulse response
FuzEn	fuzzy entropy
HE	hierarchical Entropy
IIR	infinite impulse response
IMEn	intrinsic mode entropy
IMFs	intrinsic mode functions
IMPE	improved multiscale PE
KS	Kolmogorov–Sinai
LZ	Lempel–Ziv
MEMD	multivariate EMD
MFE	multiscale fuzzy sample entropy
MFPE	multiscale fractional-order PE
MGMSE	multivariate generalized MSE
MGrcMSE	multivariate generalized rcMSE
MMSE	modified MSE
MrcMSE	multivariate
MSampEn	multivariate sample entropy
MSE	multiscale entropy
MSSE	multiscale symbolic entropy analysis
mv	multivariate
mvMPE	mv multiscale PE
PE	permutation entropy
PSD	power spectral density
rcMFE	refined composite multiscale fuzzy entropy
rcMPE	refined composite multiscale permutation entropy
rcMSE	refined composite MSE
rMSE	refined MSE
SE	sample entropy
SV	singular value
SVD	singular value decomposition
TSMBE	time shift multiscale bubble entropy
WPE	weighted permutation entropy
WMFPE	weighted multiscale fractional-order PE

## Appendix A

### Appendix A.1. About Entropies

Entropy has been used in a wide range of applications, from thermodynamic to information theory via statistical mechanics and the structure of graphs. This is a measure of different properties such as the amount of disorder, uncertainty, lack of predictability, and a measure of information.

Shannon entropy has been one of the most popular entropies for more than 70 years. Then, a great deal of interest has been paid to the definitions of alternative entropies, especially in the 1960s and the 1970s with many contributors such as Arimoto and Picard when dealing with discrete random variables. Some generalizations of Shannon entropy were proposed such as Sharma–Mittal entropy where the practitioner can tune two real parameters:  $\alpha \neq 1$ , which is called the order, and  $\beta \neq 1$ , which is known as the degree. When  $\beta = \alpha$ , Tsallis entropy is obtained. When  $\alpha$  tends to 0 (resp. infinity), Hartley entropy (resp. min entropy) is deduced. When  $\alpha = 2$ , one obtains the collision entropy. In addition, the Kapur entropy can be defined, up to a multiplicative factor, as the difference between the Rényi entropies for two different orders. A generalization of the entropy is given through the  $(h, \Phi)$  entropy, whose simplified definition is given by using the functions  $h$  and  $\Phi$ , which are increasing and concave, respectively (or decreasing and convex):

$$H_{\Phi}^h = h\left[\int \Psi(p(x))dx\right] \quad (A1)$$

If  $h(t) = t$  and  $\Psi(t) = -t \log(t)$ , one obtains Shannon entropy. Other functions can be defined to obtain Taneja entropy, Arimoto entropy, Varma entropy of the first and second kind, Havrda, and Charvat entropy. The reader may refer to [45] for more details.

The notion of metric entropy of a dynamical system with  $F$  degrees of freedom, whose state measured at time instant multiples of  $\delta$  that can be in the phase space partitioned into a collection of boxes of size  $r$ , was introduced at the end of the 1950s. This led to the Kolmogorov–Sinai (KS) entropy, whose value makes it possible to distinguish an ordered system from a chaotic one. This was the starting point of other research activities whose purpose was to obtain an approximation of the KS entropy in practice. This led to quantities such as  $K_2$  entropy proposed by Grassberger and Procaccia, the Eckmann–Ruelle (ER) entropy, and finally the approximate entropy (ApEn) proposed by Pincus in 1991 [6]. In that case, given a set of samples  $\{x_n\}_{n=1,\dots,N}$ , one first computes the probability that the distance between one reference vector storing  $m$  consecutive values of the sample set starting by  $x_i$  and another vector of length  $m$  starting by  $x_j$  is smaller than a threshold, which is also called tolerance level  $r$ . Note that the vectors can start by the same sample and, hence, can be the same. For a given value of  $m$ , the logarithms of the  $N - m + 1$  probabilities are summed. The same procedure is done for  $m + 1$ . At that stage, the ApEn is the difference between both quantities when  $N$  tends to infinity. It is hence defined for a given length  $m$  and tolerance level  $r$ . The ApEn has been generalized through the kernel-based entropies. See for instance [46]. The SE is another extension of the ApEn [47]. As it depends on the tolerance level  $r$ , the quadratic differential entropy rate can be considered where  $\log(2r)$  is added. To address the problem of atrial fibrillation, one can also use the coefficient of the SE (COSEn). The latter corresponds to the SE from which one subtracts the sum of two terms: the  $\log(2r)$  and the logarithm of the RR interval mean [48].

In the intrinsic mode entropy (IMEn), the authors propose to use the empirical mode decomposition (EMD) in order to extract the intrinsic mode functions (IMFs) of the signal. Then, the SEs of the cumulative sums of the IMFs are computed [49].

As an alternative to the SE, other approaches can be used. Thus, some authors like Chen et al. [50,51] suggested using the concept of Zadeh’s fuzzy sets in the classification procedure in the ApEn and the SE. This led to the fuzzy entropy (FuzEn). However, its computational cost of the FE is larger than that of the SE. In addition, when short time series are studied, this may lead to poor results.

In the range entropy [52], instead of considering the Chebyshev distance, the authors suggested using the difference between the maximum and the minimum divided by the sum of the maximum and the minimum. Note that this is no longer a distance.

Another type of approaches is based on a bubble sorting algorithm that consists of repeatedly swapping two adjacent values when they are in the wrong order to obtain values ordered in the ascending order. The rank-based entropy was also proposed in [53].

In [54], the data vectors of length  $m$  were sorted in an ascending order by using the bubble sorting algorithm. In each case, the number of swaps were saved to obtain an histogram. It was used to compute the second-order Rényi entropy. The standard bubble entropy (bEn) is given as the difference between the entropies computed for the vector of length  $m + 1$  and  $m$ . It is finally normalized by  $\log(\frac{m+1}{m-1})$ .

As an alternative to these types of entropies, starting from the definition of Shannon entropy, some authors have proposed to define a distribution that is different from the pdf of the data. Thus, singular value decomposition (SVD)-based entropy has been presented in [55]: in that case, the SVD of the Hankel data matrix can be computed; each singular value (SV) is then divided by the sum of the SVs, and the Shannon entropy of the resulting distribution is computed. When dealing with the EMD energy entropy, the pdf is built from the the normalized energy of each IMF [56]. In [57], the signal was first filtered using a high-pass difference filter. Then, the normalized difference between two consecutive samples was computed and expressed in percentage. The probability that the resulting index belongs to the interval  $[i, i + 1[$  with  $i$  an integer was computed. In [58], Fourier entropy was used, where the principle is to decompose the signal of interest in a Fourier basis. The modules of the Fourier coefficients are normalized so that the sum is equal to 1. Shannon entropy of the resulting distribution is computed. Note that a time evolving entropy based on a sliding window can be also considered. As an alternative, a wavelet-based entropy can be defined. See [59].

Regarding permutation entropy (PE) [60] based on the order relations among values of a signal, the principle is the following: when considering  $m$  consecutive samples whose values are assumed to be all different,  $m!$  patterns can be obtained. This number corresponds to the number of permutations that can be done on a set whose cardinal is equal to  $m$ . The probability that each pattern appears can then be calculated on the signal and its delayed version. This consists of counting the number of times the pattern appears and then normalizing it. Finally, Shannon entropy is evaluated. However, it is sensitive to noise. In addition, this approach reflects how adjacent values are organized, but the data magnitudes are omitted. For this reason, the phase PE was developed [61]. It is based on the analytical signal deduced from the signal under study. In addition, the weighted PE (WPE) has been proposed. In that case, the estimation of the probability no longer consists of only counting the number of times each pattern appears. When a pattern is identified, the variance of the  $m$  data samples is considered instead. The variances are summed and then normalized. These approaches have been also extended to the Rényi entropy of order  $\alpha$  [62] and to Havrda–Charvat entropy [63]. The dispersion entropy (DE) is based on a linear or a nonlinear mapping technique, the definition of the possible dispersion patterns, the probability they appear, and the computation of the corresponding Shannon’s entropy [64]. Note that the computational cost of DE is much smaller than that of SE. However, DE is sensitive to the number of classes. Some variants were recently developed: one called fluctuation-based DispEn (FDispEn) considers the differences between adjacent elements of the dispersion patterns [65], whereas the fuzzy DE combines DE and fuzzy membership functions for signal quantization and was proposed in [66].

Finally, in neuroscience and biomedical applications for instance, various signals can be recorded at the same time (electroencephalogram (EEG) for instance). For this reason, multivariate (mv) entropy can be considered. The reader may refer to [67].

### Appendix A.2. Some Variants of the MSE

In 2009, the refined MSE (rMSE) was proposed by Valencia et al. [25], where the FIR filter was replaced by an IIR low-pass Butterworth filter whose cutoff frequency is chosen to downsample the signal. Note that in 2021, the authors in [26] suggested using the same kind of filter (the resulting method was called parallel MSE by the authors). However, this filter does not have a linear phase. Therefore, the design of this filtering step must be improved.

In 2013, the composite MSE (cMSE) consisted of using the  $\tau$  sequences that can be defined after the decimation step [22]:

$$cMSE(\underline{x}_1, \tau_{max}, m, r) = \sum_{\tau=1}^{\tau_{max}} cSE(\underline{x}_1, \tau, m, r) \quad (A2)$$

where  $cSE(\underline{x}_1, \tau, m, r) = \frac{1}{\tau} \sum_{k=1}^{\tau} SE(y_k^{(\tau)}, 1, m, r)$  corresponds to the average of the sample entropies computed on the  $\tau$  sequences obtained from the filtering and the decimation. Averaging has the advantage of reducing the variance of the estimated SE values. The above equation can be rewritten as follows:

$$\begin{aligned} cMSE(\underline{x}_1, \tau_{max}, m, r) &= - \sum_{\tau=1}^{\tau_{max}} \frac{1}{\tau} \sum_{k=1}^{\tau} \log \frac{P_r^{m+1}(y_k^{(\tau)})}{P_r^m(y_k^{(\tau)})} \\ &= - \sum_{\tau=1}^{\tau_{max}} \log \frac{(\prod_{k=1}^{\tau} P_r^{m+1}(y_k^{(\tau)}))^{\frac{1}{\tau}}}{(\prod_{k=1}^{\tau} P_r^m(y_k^{(\tau)}))^{\frac{1}{\tau}}} \end{aligned} \quad (A3)$$

What is called “short-term MSE” and proposed by [68] is described by the same process.

As pointed out in (A3), the cMSE amounts to computing the mean of logarithms or equivalently computing the log of geometric means. In contrast, the refined composite MSE (rcMSE) consists of computing the log of the arithmetic means [69]:

$$rcMSE(\underline{x}_1, \tau_{max}, m, r) = - \sum_{\tau=1}^{\tau_{max}} \log \frac{\sum_{k=1}^{\tau} P_r^{m+1}(y_k^{(\tau)})}{\sum_{k=1}^{\tau} P_r^m(y_k^{(\tau)})} \quad (A4)$$

The geometric mean is less sensitive than the arithmetic mean to the highest values.

In the modified MSE (MMSE) [70], the authors proposed to apply the averaging filter. Instead of decimating the filter output  $y$ , the sample entropies were directly computed on  $y$ . According to the authors, this has the advantage of increasing the number of vector pairs, but the computational cost becomes larger.

$$MMSE(\underline{x}_1, \tau_{max}, m, r) = \sum_{\tau=1}^{\tau_{max}} SE(y_1, \tau, m, r) \quad (A5)$$

In [21], Pham et al. suggested introducing different time shifts on various intervals of time series. In some applications, the MSE was applied on data that had been preprocessed. See for instance [71].

The concept of multiscale can be extended to other entropy metrics such as those recalled in Section 2.1. Among them, the multiscale PE is detailed in [72]. Some variants also exist such as the refined composite multiscale PE (rcMPE) [73], the improved multiscale PE (IMPE) [74], the multiscale fractional-order PE (MFPE), and its weighted version—whose acronym is WMFPE in [75] or the fractional multiscale phase PE [76]. When fractional approaches are considered, the  $\alpha$  derivative of the information element  $-\log(p)$  with  $-1 \leq \alpha \leq 1$  and  $p$  probability is considered in the definition of the entropy instead of  $-\log(p)$ .



Multiscale fuzzy SE (MFE) has been derived, as well as composite and refined composite multiscale fuzzy entropy (rcMFE) [77].

In 2015, the generalized MSE was proposed by Costa et al. [78]. Instead of computing the mean in the first step of the MSE approach, a higher-order moment (for instance the variance) was considered. When dealing with the variance, one obtains the  $MSE\sigma^2$ . As the length of the CG series decreases when the scale increases, the estimation of the SE may not be accurate. For this reason, the refined  $MSE\sigma^2$  was proposed in [79], thereby following the same methodology as in the rcMSE. The MSE based on skewness and the kurtosis have been also developed in [80].

The solution proposed by Hu et al. [81] is to decompose the signal into IMFs by using the EMD and its variant known as multivariate EMD (MEMD), which has the advantage of fixing some problems such as the mode misalignment. As the EMD and MEMD essentially act as dyadic filter banks, the authors suggest summing a certain number of IMFs in order to obtain different sets of data, whose spectrums tend to exhibit lower and lower frequencies. The number of IMFs that are considered corresponds to a scale. Then, the SE is computed on each signal. This leads to the adaptive MSE (AME). However, this method has some limitations such as the sensitivity to the noise.

Finally, in [82], the authors suggested applying the CG procedure, where the sample average is replaced by the median value. From a signal processing point of view, the low-pass linear filter is replaced by a nonlinear filter. Then, the authors suggested deducing the sign time series. When the difference between two consecutive samples is positive, the value is equal to 1. Otherwise, it is equal to 0. As a consequence, the number of possible patterns was found to be finite, and the authors used both the SE and this second entropy. This method is called multiscale symbolic entropy analysis (MSSE).

In hierarchical entropy (HE) [8], the signal is decomposed into two sequences: the low-pass filtered one obtained by the averaging filter and the high-pass filtered one obtained by the difference filter. Then, each resulting sequence is decomposed into two sequences: a low-pass and a high-pass one. This leads to a “hierarchical” decomposition of the signal. Then, the SE is computed for each component.

Extension of other entropies to the multiscale has been done. Thus, in [83], the bubble entropy led to time shift multiscale bubble entropy (TSMBE).

In [84], the authors suggested using the refined composite multiscale DE and applying it to biomedical signals. More recently, they considered the refined composite generalized MDE based on the mean, variance, and skewness [85], and they proposed the multiscale fuzzy dispersion entropy (MFDE) and refined composite MFDE (RCMFDE) [86].

Instead of processing the signals independently, mv MSE, as well as multivariate rcMSE (MrcMSE), have been proposed [67]. They are based on multivariate SE (MSampEn). Humeau proposed to combine the above concepts to obtain what is called multivariate generalized MSE (MGMSE) and multivariate generalized rcMSE (MGrcMSE) [87]. Weighted versions were recently studied in [88]. The mv multiscale PE (mvMPE) has been also developed, as well as the mv multiscale DE [89] and the multiscale multivariate fractional dispersion entropy (MMFDE) [90]. The reader could finally notice that multiscale approaches have been also derived for crossentropy to analyze the dynamical characteristics of the coupling behavior between two sequences on multiple scales [91].

The reader can refer to Figures A1 and A2 to see the links between the entropies and the variants of the MSE.

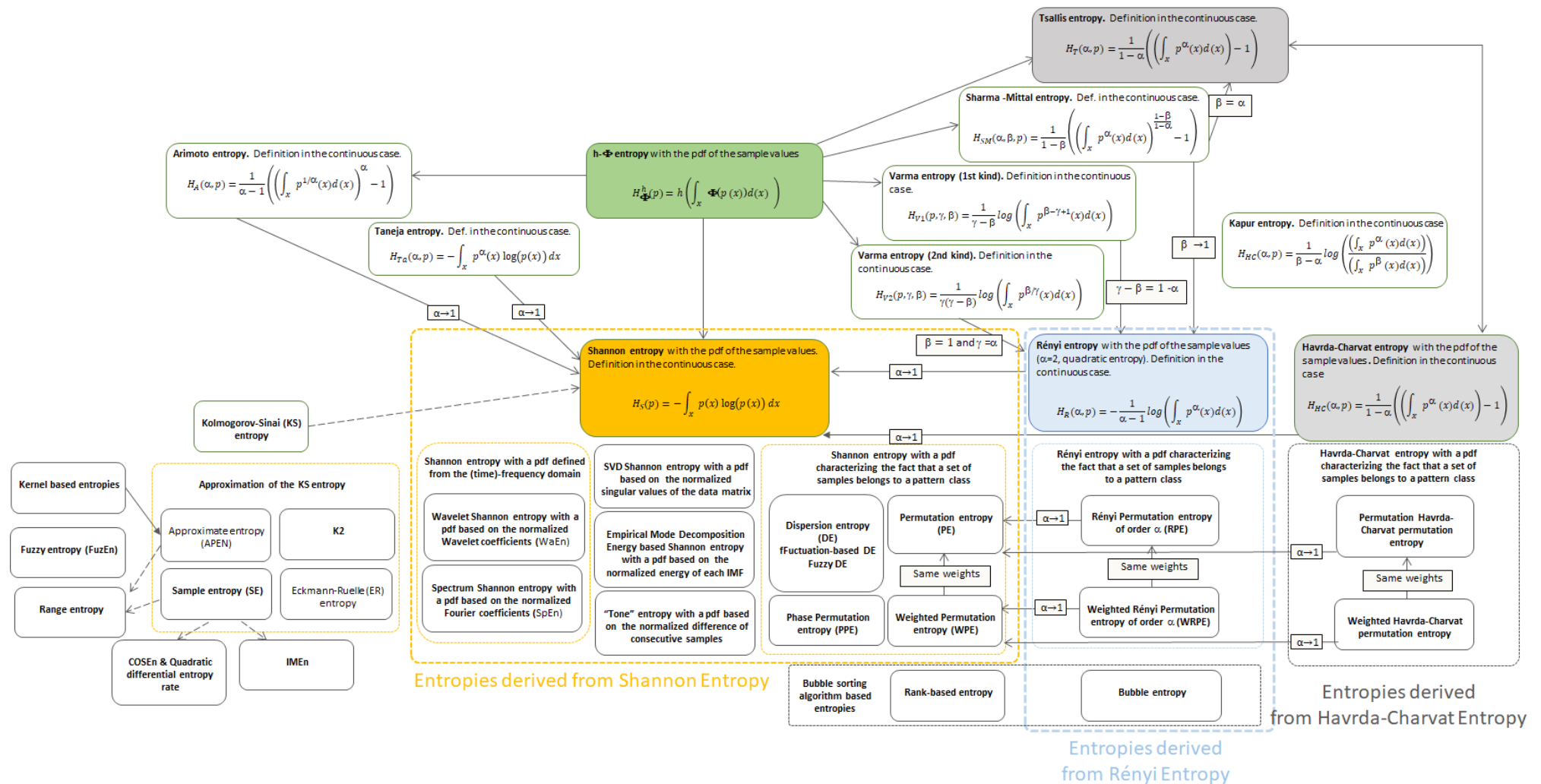


Figure A1. Links between the various entropies.

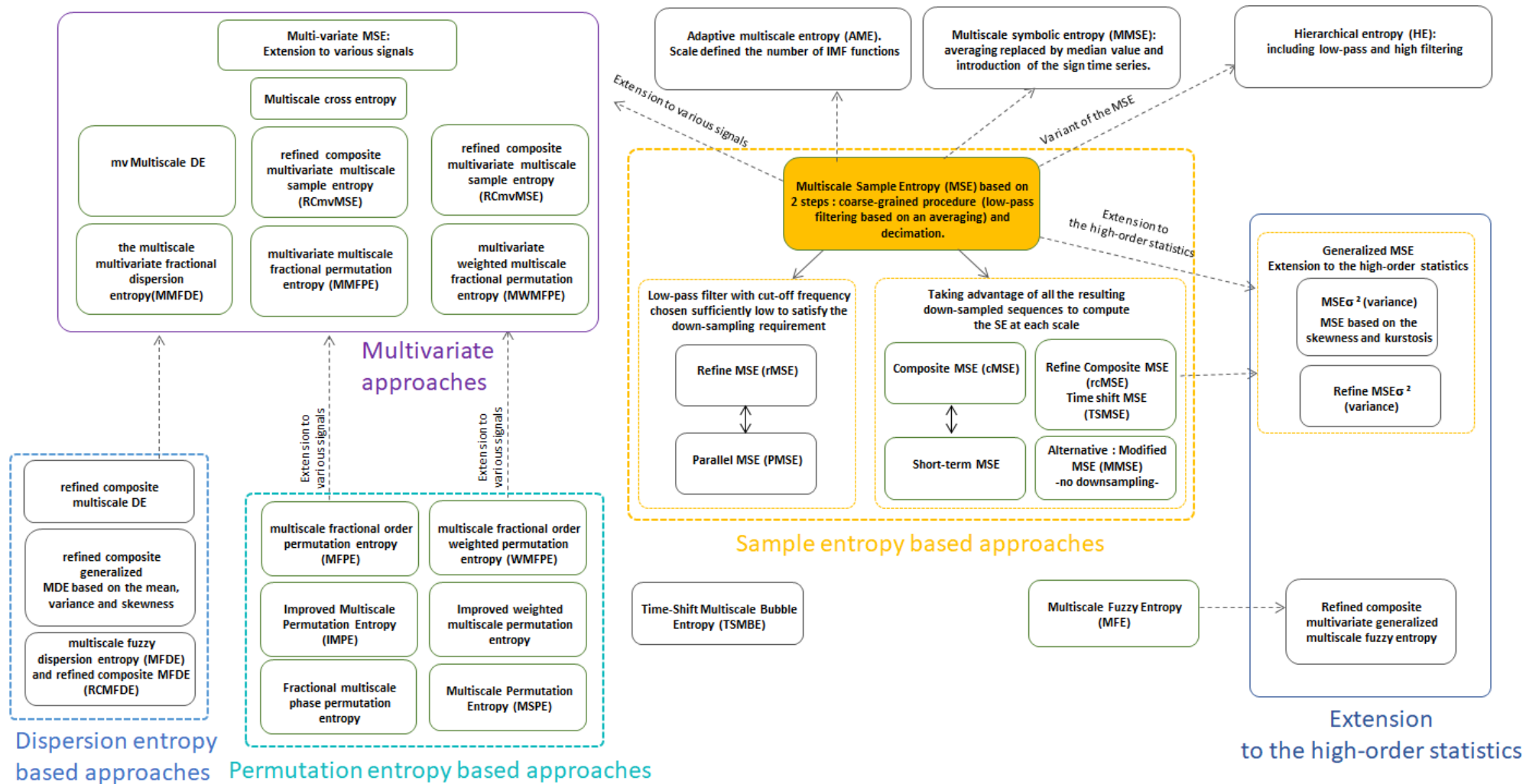


Figure A2. Multiscale entropies and its variants.

## References

- Berthelot, B.; Grivel, E.; Legrand, P.; André, J.-M.; Mazoyer, P. Alternative ways to compare the detrended fluctuation analysis and its variants. application to visual tunneling detection. *Digit. Signal Process.* **2020**, *108*, 102865.
- Berthelot, B.; Grivel, E.; Legrand, P. New Variants of DFA based on LOESS and LOWESS methods: Generalization of the detrended moving average. In Proceedings of the ICASSP 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021.
- Grivel, E.; Berthelot, B.; Legrand, P.; Giremus, A. Dfa-based abacuses providing the hurst exponent estimate for short-memory processes. *Digit. Signal Process.* **2021**, *116*, 103102.
- Peng, C.K.; Buldyrev, S.V.; Goldberger, A.L.; Havlin, S.; Sciortino, F.; Simons, M.; Stanley, H.E. Long-range correlations in nucleotide sequences. *Nature* **1992**, *356*, 168–170.
- Peng, C.K.; Havlin, S.; Hausdorff, J.M.; Mietus, J.E.; Stanley, H.E.; Goldberger, A.L. Fractal mechanisms and heart rate dynamics. *J. Electrocardiol* **1996**, *28*, 59–64.
- Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301.
- Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102.
- Jiang, Y.; Peng, C.K.; Xu, Y. Hierarchical entropy analysis for biological signals. *J. Comput. Appl. Math.* **2011**, *236*, 728–742.
- Chen, Z.; Gómez, A.I.; Gómez-Pérez, D.; Tirkeld, A. Correlation measure, linear complexity and maximum order complexity for families of binary sequences. *Finite Fields Their Appl.* **2022**, *78*, 101977.
- Xiong, H.; Qu, L.; Li, C. A New Method to Compute the 2-adic Complexity of Binary Sequences. *IEEE Trans. Inf. Theory* **2014**, *60*, 2399–2406.
- Lempel, A.; Ziv, J. On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81.
- Ziv, J.; Lempel, A. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory* **1977**, *23*, 337–343.
- Yan, X.; She, D.; Xu, Y.; Jia, M. Application of Generalized Composite Multiscale Lempel–Ziv Complexity in Identifying Wind Turbine Gearbox Faults. *Entropy* **2021**, *23*, 1372.
- Bai, Y.; Liang, Z.; Li, X. A permutation Lempel–Ziv complexity measure for EEG analysis. *Biomed. Signal Process. Control* **2015**, *19*, 102–114.
- Borowska, M. Multiscale Permutation Lempel–Ziv Complexity Measure for Biomedical Signal Analysis: Interpretation and Application to Focal EEG Signals. *Entropy* **2021**, *23*, 832.
- Ibáñez-Molina, A.J.; Iglesias-Parro, S.; Soriano, M.F.; Aznarte, J.I. Multiscale Lempel–Ziv complexity for EEG measures. *Clin. Neurophysiol.* **2015**, *126*, 541–548.
- Mao, X.; Shang, P.; Xu, M.; Peng, C.-K. Measuring time series based on multiscale dispersion Lempel–Ziv complexity and dispersion entropy plane. *Chaos Solitons Fractals* **2020**, *37*, 109868.
- Liu, L.; Miao, S.; Hu, H.; Deng, Y. On the Eigenvalue and Shannon’s Entropy of Finite Length Random Sequences. *Complexity* **2015**, *21*, 154–161.
- Liu, L.; Xiang, H.; Li, R.; Hu, H. The Eigenvalue Complexity of Sequences in the Real Domain. *Entropy* **2019**, *21*, 1194.
- Jauregui, M.; Zunino, L.; Lenzi, E.K.; Mendes, R.S.; Ribeiro, H.V. Characterization of time series via Rényi complexity–entropy curves. *Physica A* **2018**, *498*, 74–85.
- Pham, T.D. Time-Shift Multiscale Entropy Analysis of Physiological Signals. *Entropy* **2017**, *19*, 257.
- Wu, S.D.; Wu, C.W.; Lin, S.G.; Wang, C.C.; Lee, K.Y. Time series analysis using composite multiscale entropy. *Entropy* **2013**, *15*, 1069–1084.
- Zhang, Y.-C. Complexity and  $1/f$  noise. A phase space approach. *J. Phys. I* **1991**, *1*, 971–977.
- Humeau-Heurtier, A. The multiscale entropy algorithm and its variants: A review. *Entropy* **2015**, *17*, 3110–3123.
- Valencia, J.F.; Porta, A.; Vallverdu, M.; Claria, F.; Baranowski, R.; Orłowska-Baranowska, E.; Caminal, P. Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 2202–2213.
- Zhao, D.; Liu, S.; Cheng, S.; Sun, X.; Wang, L.; Wei, Y.; Zhang, H. Parallel multi-scale entropy and its application in rolling bearing fault diagnosis. *Measurement* **2021**, *168*, 108333.
- Azami, H.; Escudero, J. Coarse-Graining Approaches in Univariate Multiscale Sample and Dispersion Entropy. *Entropy* **2018**, *20*, 138.
- Grivel, E.; Berthelot, B.; Legrand, P.; Colin, G. Null or linear-phase filters for the derivation of a new variant of the MSE. In Proceedings of the 2023 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 4–8 September 2023.
- Nikulin, V.V.; Brismar, T. Comment on multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2004**, *92*, 089803.
- D’avalos, A.; Jabloun, M.; Ravier, P.; Buttelli, O. Theoretical study of multiscale permutation entropy on finite-length fractional gaussian noise. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018.
- Gao, J.; Hu, J.; Liu, F.; Cao, Y. Multiscale entropy analysis of biological signals: a fundamental bi-scaling law. *Front. Comput. Neurosci.* **2015**, *9*, 1–9.

32. Wickens, C. *Attentional Tunneling and Task Management*; International Symposium on Aviation Psychology: Dayton, OH, USA, 2005; pp. 812–817.
33. Shappell, S.A.; Wiegman, D.A. *Definition and Computation of Oculomotor Measures in the Study of Cognitive Processes*; Elsevier Science Ltd.: Amsterdam, The Netherlands, 2003.
34. Dehais, F.; Roy, R.; Scannella, S. Inattentional deafness to auditory alarms: inter-individual differences, electrophysiological signature and single trial classification. *Behav. Brain Res.* **2018**, *360*, 51–59.
35. Rock, I.; Linnett, C.; Grant, P.; Mack, A. Perception without attention: results of a new method. *Cogn. Psychol.* **1992**, *24*, 502–534.
36. McAtee, A.; Feltman, K.; Swanberg, D.; Russell, D.; Statz, J.; Harding, T.; Ramiccio, J. *Pilot Cueing Synergies for Degraded Visual Environments*; Society of Photo-Optical Instrumentation Engineers, SPIE Defense + Security: Anaheim, CA, USA, 2017.
37. Simons, D.; Chabris, C. Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception* **1999**, *28*, 1059–1074.
38. Wilson, G.; Eggemeier, F. Psychophysiological assessment of workload in multi-task environments. In *Multiple-Task Performance*; CRC Press: Boca Raton, FL, USA, 1991; pp. 329–360.
39. Posner, M.I.; Dehaene, S. Attentional networks. *Trends Neurosci.* **1994**, *17*, 75–79.
40. Gaume, A.; Dreyfus, G.; Vialatte, F.B. A cognitive brain–computer interface monitoring sustained attentional variations during a continuous task. *Cogn. Neurodynamics* **2019**, *13*, 257–269.
41. Chanel, P.C.C.; Wilson, M.D.; Scannella, S. Online ECG-based features for cognitive load assessment. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 3710–3717.
42. Regis, N.; Dehais, F.; Rachelson, E.; Thooris, C.; Pizzoli, S.; Causse, M.; Tessier, C. Formal detection of attentional tunneling in human operator– automation interactions. *IEEE Trans. Hum. Mach. Syst.* **2014**, *44*, 326–336.
43. Santiago-Espada, Y.; Myer, R.R.; Latorella, K.A.; Comstock, J.R., Jr. The Multi-Attribute Task Battery II (Matb-II) Software for Human Performance and Workload Research: A User’s Guide; NASA/TM–2011-217164; 2011. Available online: <https://ntrs.nasa.gov/api/citations/20110014456/downloads/20110014456.pdf> (accessed on 5 April 2024).
44. Hart, S.; Staveland, L. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Adv. Psychol.* **1988**, *52*, 139–183.
45. Menendez, M.L.; Morales, D.; Pardo, L.; Salicrú, M. (h;φ)-entropy differential metric. *Appl. Math.* **1997**, *42*, 81–98.
46. Xu, L.S.; Wang, K.Q.; Wang, L. Gaussian kernel approximate entropy algorithm for analyzing irregularity of time-series. *Int. Conf. Mach. Learn. Cybern.* **2005**, *9*, 5605–5608.
47. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049.
48. Lake, D.E.; Moorman, J.R. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am. J. Physiol. Heart Circ. Physiol.* **2011**, *300*, H319–H325.
49. Amoud, H.; Snoussi, H.; Hewson, D.; Doussot, M.; Duchêne, J. Intrinsic mode entropy for nonlinear discriminant analysis. *IEEE Signal Process. Lett.* **2014**, *14*, 297–300.
50. Chen, W.; Wang, Z.; Xie, H.; Yu, W. Characterization of surface emg signal based on fuzzy entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15*, 266–72.
51. Chen, W.; Zhuang, J.; Yu, W.; Wang, Z. Measuring complexity using fuzzyen, apen, and sampen. *Med Eng. Phys.* **2009**, *15*, 61–68.
52. Omidvarnia, A.; Mesbah, M.; Pedersen, M.; Jackson, G. Range entropy: A bridge between signal complexity and self-similarity. *Entropy* **2018**, *20*, 962.
53. Citi, L.; Guffanti, G.; Mainardi, L. Rank-based multi-scale entropy analysis of heart rate variability. *Comput. Cardiol.* **2014**, *41*, 597–600.
54. Manis, G.; Bodini, M.W.; Rivolta, M.; Sassi, R. A two-steps-ahead estimator for bubble entropy. *Entropy* **2021**, *23*, 761.
55. Roberts, S.J.; Penny, W.; Rezek, I. Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. *Med Biol. Eng. Comput.* **1999**, *37*, 93–98.
56. Yu, Y.; Junsheng, C. A roller bearing fault diagnosis method based on emd energy entropy and ANN. *J. Sound Vib.* **2006**, *294*, 269–277.
57. Oida, E.; Moritani, T.; Yamori, Y. Tone-entropy analysis on cardiac recovery after dynamic exercise. *J. Appl. Physiol.* **1997**, *82*, 1794–1801.
58. Powell, G.E.; Percival, I.C. A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems. *J. Phys. A: Math. Gen.* **1979**, *12*, 2053.
59. Rosso, O.A.; Blanco, S.; Yordanova, J.; Kolev, V.; Figliola, A.; Schürmann, M.; Başar, E. Wavelet entropy: A new tool for analysis of short duration brain electrical signals. *J. Neurosci. Methods* **2001**, *105*, 65–75.
60. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102.
61. Kang, H.; Zhang, X.; Zhang, G. Phase permutation entropy: A complexity measure for nonlinear time series incorporating phase information. *Physica A* **2021**, *568*, 9125686.
62. Zhao, X.; Shang, P.; Huang, J. Permutation complexity and dependence measures of time series. *Europhys. Lett.* **2013**, *102*, 40005.
63. Shi, Y.; Shang, Y.; Wu, P. Research on weighted havrda–charvat’s entropy in financial time series. *Physica A* **2021**, *572*, 125914.
64. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time series analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614.
65. Azami, H.; Escudero, J. Amplitude and Fluctuation-Based Dispersion Entropy. *Entropy* **2018**, *20*, 210.

66. Rostaghi, R.M.; Khatibi, M.M.; Ashory, M.R.; Azami, H. Fuzzy Dispersion Entropy: A Nonlinear Measure for Signal Analysis. *IEEE Trans. Fuzzy Syst.* **2021**, *30*, 3785–3796.
67. Ahmed, M.U.; Mandic, D.P. Multivariate multiscale entropy: A tool for complexity analysis of multichannel data. *Phys. Rev. E* **2011**, *84*, 061918.
68. Chang, Y.C.; Wu, H.T.; Chen, H.R.; Liu, A.B.; Yeh, J.J.; Lo, M.T.; Tsao, J.H.; Tang, C.J.; Tsai, I.T.; Sun, C.K. Application of a modified entropy computational method in assessing the complexity of pulse wave velocity signals in healthy and diabetic subjects. *Entropy* **2014**, *16*, 4032–4043.
69. Wu, S.D.; Wu, P.H.; Wu, C.W.; Ding, J.J.; Wang, C.C. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **2012**, *14*, 1343–1356.
70. Wu, S.D.; Wu, C.W.; Lee, K.Y.; Lin, S.G. Modified multiscale entropy for short-term time series analysis. *Physica A* **2013**, *392*, 5865–5873.
71. Shi, W.; Feng, H.; Zhang, X.; Yeh, C.-H. Amplitude modulation multiscale entropy characterizes complexity and brain states. *Chaos Solitons Fractals* **2023**, *173*, 113646.
72. Vakharia, V.; Gupta, V.K.; Kankar, P.K. A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings. *J. Vib. Control*. **2015**, *21*, 3123–3131.
73. Humeau-Heurtier, A.; Wu, C.-W.; Wu, S.-D. Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence. *IEEE Signal Process. Lett.* **2015**, *22*, 2364–2367.
74. Azami, H.; Escudero, J. Improved multiscale permutation entropy for biomedical signal analysis: interpretation and application to electroencephalogram recordings. *Biomed. Signal Process. Control* **2016**, *23*, 28–41.
75. Chen, S.; Shang, P.; Wu, Y. Multivariate multiscale fractional order weighted permutation entropy of nonlinear time series. *Physica A* **2019**, *515*, 217–231.
76. Wan, L.; Ling, G.; Guan, Z.-H.; Fan, Q.; Tong, Y.-H. Fractional multiscale phase permutation entropy for quantifying the complexity of nonlinear time series. *Physica A* **2022**, *600*, 127506.
77. Zheng, J.; Pan, H.; Cheng, J. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **2017**, *85*, 746–759.
78. Costa, M.D.; Goldberger, A.L. Generalized multiscale entropy analysis: Application to quantifying the complex volatility of human heartbeat time series. *Entropy* **2015**, *17*, 1197–1203.
79. Liu, Y.; Lin, Y.; Wang, J.; Shang, P. Refined generalized multiscale entropy analysis for physiological signals. *Physica A* **2018**, *490*, 975–985.
80. Xu, M.; Shang, P. Analysis of financial time series using multiscale entropy based on skewness and kurtosis. *Physica A* **2018**, *490*, 1543–1550.
81. Hu, M.; Liang, H. Adaptive multiscale entropy analysis of multivariate neural data. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 12–15.
82. Lo, M.T.; Chang, Y.C.; Lin, C.; Young, H.W.V.; Lin, Y.H.; Ho, Y.L.; Peng, C.K.; Hu, K. Outlier-resilient complexity analysis of heartbeat dynamics. *Sci. Rep.* **2015**, *5*, 8836.
83. Chen, F.; Tian, W.; Zhang, L.; Li, J.; Ding, C.; Chen, D.; Wang, W.; Wu, F.; Wang, B. Fault diagnosis of power transformer based on time-shift multiscale bubble entropy and stochastic configuration network. *Entropy* **2022**, *24*, 1135.
84. Azami, H.; Rostaghi, M.; Abasolo, D.; Escudero, J. Refined Composite Multiscale Dispersion Entropy and its Application to Biomedical Signals. *IEEE Trans. Biomed. Engineering* **2017**, *64*, 2872–2879.
85. Rostaghi, M.; Khatibi, M.M.; Ashory, M.R.; Azami, H. Bearing Fault Diagnosis Using Refined Composite Generalized Multiscale Dispersion Entropy-Based Skewness and Variance and Multiclass FCM-ANFIS. *Entropy* **2021**, *23*, 1510.
86. Rostaghi, R.M.; Khatibi, M.M.; Ashory, M.R.; Azami, H. Refined Composite Multiscale Fuzzy Dispersion Entropy and Its Applications to Bearing Fault Diagnosis. *Entropy* **2023**, *25*, 1494.
87. Humeau-Heurtier, A. Multivariate generalized multiscale entropy analysis. *Entropy* **2016**, *18*, 411.
88. Minhasa, A.S.; Kankar, P.K.; Kumarc, N.; Singh, S. Bearing fault detection and recognition methodology based on weighted multiscale entropy approach. *Mech. Syst. Signal Process.* **2021**, *147*, 107073.
89. Azami, H.; Fernández, A.; Escudero, J. Multivariate Multiscale Dispersion Entropy of Biomedical Times Series. *Entropy* **2019**, *21*, 913.
90. Zhang, B.; Shang, P.; Zhou, Q. The identification of fractional order systems by multiscale multivariate analysis. *Chaos Solitons Fractals Nonlinear Sci. Nonequilibrium Complex Phenom.* **2021**, *144*, 110735.
91. Jamin, A.; Humeau-Heurtier, A. (Multiscale) Cross-Entropy Methods: A Review. *Entropy* **2020**, *22*, 45.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.