INFORMATION-THEORETIC SIGNAL PROCESSING ON THE TIME-FREQUENCY PLANE AND APPLICATIONS

Selin Aviyente

Department of Electrical and Computer Engineering, Michigan State University
2120 Engineering Building, East Lansing, MI 48824
phone: + (1) 517 355 7649, fax: + (1) 517 353 1980, email: aviyente@egr.msu.edu

ABSTRACT

Time-frequency analysis is a major tool in representing the energy distribution of time-varying signals. There has been a lot of research on various properties of these representations. However, there is a general lack of quantitative measures in describing the amount of information encoded into a time-frequency distribution. Recently, information-theoretic measures such as entropy and divergence have been adapted to the time-frequency plane to quantify the complexity of individual signals as well as the difference between signals. In this paper, we present a variety of information-theoretic measures and their definitions on the time-frequency plane. The properties of these measures and how they can be applied to signal classification problems are discussed in detail. We then present an application of information-theoretic signal processing to the analysis of event- related brain potentials.

1. INTRODUCTION

Time-frequency distributions (TFDs) are used for representing the energy distribution of time-varying signals simultaneously in time and frequency. Despite their wide use in areas such as detection and classification of signals, their capacity in representing information has not been evaluated quantitatively. This paper aims at addressing this issue by introducing information-theoretic measures such as entropy and divergence measures on the time-frequency plane.

In recent years, there has been an interest in adapting information-theoretic measures to the time-frequency plane in order to quantify signal complexity [1, 2, 3]. The application of information-theoretic measures such as entropy and divergence has made it easier to quantify the complexity of non-stationary signals on the time-frequency plane as well as differentiate between different signals. Despite the success of entropy in characterizing a signal's complexity on the time-frequency plane, it is not sufficient in quantifying the dependencies between signals. In order to have an effective information-theoretic signal characterization and classification system, we need information-theoretic measures that quantify the dependencies between signals on the time-frequency plane. One such measure that can effectively quantify the differences between signals is divergence measures. Distance measures between statistical models have been widely used in signal processing applications. Using entropy based distance functionals is a well-known discrimination method in signal processing. These functionals are known as divergence measures and are applied directly on statistical models describing the signals. Measures of divergence between two probability distributions are used to associate, cluster, classify, compress, and restore signals, images and patterns, in many applications. Many different measures of divergence have been constructed and characterized [5, 6]. Another measure that quantifies dependency is mutual information. Mutual information has been used effectively in various statistical signal processing applications including classification and source separation [4]. In this paper, it will be extended to the time-frequency plane to quantify the 'independence' between signals.

2. TIME-FREQUENCY EQUIVALENT OF MUTUAL INFORMATION

2.1 Background on Information-Theoretic Measures on the Time-Frequency Plane

A time-frequency distribution, $C(t,f)$, from Cohen's class can be expressed as [1] [7]:

$$C(t,f) = \iiint \phi(\theta,\tau) s(u+\frac{\tau}{2}) s^*(u-\frac{\tau}{2}) e^{j(\theta u - \theta t - 2\pi\tau f)} du\, d\theta\, d\tau, \quad (1)$$

where $\phi(\theta,\tau)$ is the kernel function and s is the signal. Some of the most desired properties of TFDs are the energy preservation and the marginals. They are given as follows and are satisfied when $\phi(\theta,0) = \phi(0,\tau) = 1 \;\; \forall \tau, \theta$.

$$\begin{aligned} \iint C(t,f)\, dt\, df &= \int |s(t)|^2 dt = \int |S(f)|^2 df, \\ \int C(t,f)\, df &= |s(t)|^2 \;\; , \int C(t,f)\, dt = |S(f)|^2. \end{aligned} \quad (2)$$

The formulas given above evoke an analogy between a TFD and the probability density function (pdf) of a two-dimensional random variable. This analogy has inspired the adaptation of information-theoretic measures such as entropy to the time-frequency plane. The main difference between TFDs and pdfs is that TFDs are not always positive. Therefore, in this paper the analysis focuses on spectrograms since they are always positive. Another important point is that the distributions have to be normalized by their energy before applying any information-theoretic measure.

2.2 Distance Measures

The most general class of distance measures is known as Csiszar's f-divergence which includes some well-known measures like Hellinger distance, Kullback-Leibler divergence and Rényi divergence [6]. The divergence between

[1] All integrals are from $-\infty$ to ∞ unless otherwise stated.

two probability density functions, p_1 and p_2 for this class of distance measures can be expressed as:

$$d(p_1,p_2) = g\left[E_1\left[f\left(\frac{p_2}{p_1}\right)\right]\right], \quad (3)$$

where f is a continuous convex function, g is an increasing function and E_1 is the expectation operator with respect to p_1. The distance measures and their properties for time-frequency distributions are given below.

1. **Kullback-Leibler divergence**: The most common distance measure used for probability distributions is the Kullback-Leibler divergence measure. This measure can be adapted to the time-frequency distributions as follows:

$$K(C_1,C_2) = \int\int C_1(t,f)\log\frac{C_1(t,f)}{C_2(t,f)}dt\,df. \quad (4)$$

This measure belongs to the class of Csiszar's f-divergence with $f(x) = -\log x$, and $g(x) = x$. $0 \leq K(C_1,C_2) \leq \infty$, the first equality holds if and only if $C_1 = C_2$ and the second equality holds if and only if $Supp\ C_1 \bigcap Supp\ = \emptyset$. This is not a symmetric distance measure but can easily be symmetrized by taking the average of $K(C_1,C_2)$ and $K(C_2,C_1)$.

2. **Rényi Divergence**: Rényi divergence is a generalized formulation of Kullback-Leibler divergence and can be expressed as:

$$D_\alpha(C_1,C_2) = \frac{1}{\alpha-1}\log\int\int C_1^\alpha(t,f)C_2^{1-\alpha}(t,f)dt\,df. \quad (5)$$

where $\alpha \in [0,1]$ is the order of Rényi divergence. This measure converges to Kullback-Leibler distance as $\alpha \to 1$. It is also a member of Csiszar's f-divergence with $f(x) = x^{1-\alpha}$, and $g(x) = \frac{1}{\alpha-1}\log(x)$. $0 \leq D_\alpha(C_1,C_2) \leq \infty$, the first equality holds if and only if $C_1 = C_2$ and the second if and only if $Supp\ C_1 \bigcap Supp\ C_2 = \emptyset$.

3. **Jensen-Shannon Divergence**: One common approach for constructing divergence measures is to apply Jensen inequality on the entropy functional. For time-frequency distributions, Jensen-Shannon divergence can be defined as:

$$J(C_1,C_2) = H\left(\frac{C_1+C_2}{2}\right) - \frac{H(C_1)+H(C_2)}{2}. \quad (6)$$

This distance measure is always positive since

$$H\left(\frac{C_1+C_2}{2}\right) \geq \frac{H(C_1)}{2} + \frac{H(C_2)}{2} \quad (7)$$

by concavity of H. It is equal to zero when $C_1 = C_2$ and is a symmetric divergence measure. Unlike the Kullback-Leibler divergence, Jensen-Shannon distance does not diverge when the two distributions are disjoint.

4. **Jensen-Rényi Divergence**: The Rényi entropy is derived from the same set of axioms as the Shannon entropy, the only difference being the employment of a more general exponential mean instead of the arithmetic mean in the derivation. This realization inspires the modification of Jensen-Shannon divergence from an arithmetic to a geometric mean, and the following quantity is obtained for two positive TFDs C_1 and C_2.

$$J_1(C_1,C_2) = H_\alpha(\sqrt{C_1C_2}) - \frac{H_\alpha(C_1)+H_\alpha(C_2)}{2}, \quad (8)$$

where $(\sqrt{C_1C_2})(t,f) = \sqrt{C_1(t,f)C_2(t,f)}$. This quantity is obviously null when $C_1 = C_2$. The positivity of this quantity can be proven using the Cauchy-Schwartz inequality.

$$\left|\int\int [C_1(t,f)C_2(t,f)]^{\alpha/2}\,dt\,df\right|^2 \leq$$
$$\int\int C_1^\alpha(t,f)dt\,df \int\int C_2^\alpha(t,f)\,dt\,df, \quad (9)$$

and since the log function is monotonically increasing, for $\alpha > 1$

$$\frac{1}{1-\alpha}\quad \log\left|\int\int [C_1(t,f)C_2(t,f)]^{\alpha/2}\,dt\,df\right|^2 \geq$$
$$\frac{1}{1-\alpha}\quad \left[\log\int\int C_1^\alpha(t,f)\,dt\,df + \log\int\int C_2^\alpha(t,f)\,dt\,df\right]. \quad (10)$$

Thus $H_\alpha(\sqrt{C_1C_2}) \geq \frac{H_\alpha(C_1)+H_\alpha(C_2)}{2}$, which proves that the distance measure is always positive.

2.3 Definition of Mutual Information on the Time-Frequency Plane

For two random variables, X and Y, the mutual information is defined as:

$$I(X;Y) = \sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}, \quad (11)$$

where $p(x,y)$, $p(x)$ and $p(y)$ are the joint and marginal probability density functions of X and Y, respectively. $I(X;Y)$ achieves its minimum when X and Y are independent and is equal to 0.

In the case of time-frequency distributions, we will adapt the definition of mutual information using energy density functions instead of probability density functions. Therefore, the individual energy distributions of signals $x(t)$ and $y(t)$ defined as $C_x(t,f)$ and $C_y(t,f)$, respectively, correspond to the marginal densities, $p(x)$ and $p(y)$, given in equation 11. Using the same definition, the joint density function in equation 11 will be replaced by the joint energy distribution of $x(t)$ and $y(t)$ defined as the cross-TFD of the two signals. For example, in the case of the spectrogram this cross-distribution is:

$$C_{xy}(t,f) = STFT_x(t,f)STFT_y^*(t,f), \quad (12)$$

where $STFT_x(t,f) = \int h(\tau-t)x(\tau)e^{-j2\pi f\tau}$ with $h(t)$ being the data window. Since $C_{xy}(t,f)$ can be complex-valued, its absolute value will be used in the definition of mutual information. Therefore, mutual information between two non-stationary signals as measured through their time-frequency distributions is:

$$I(C_x,C_y) = \int\int |C_{xy}(t,f)|\log\frac{|C_{xy}(t,f)|}{C_x(t,f)C_y(t,f)}dtdf. \quad (13)$$

It is important to note that this measure is closely related to Jensen-Rényi divergence. Since $|C_{xy}(t,f)| =$

$\sqrt{|C_x(t,f)|}\sqrt{|C_y(t,f)|}$, the mutual information can be written as:

$$I(C_x,C_y) = \int\int \sqrt{|C_x(t,f)|}\sqrt{|C_y(t,f)|}\log\frac{\sqrt{|C_x(t,f)|}\sqrt{|C_y(t,f)|}}{C_x(t,f)C_y(t,f)}dt\,df,$$
$$= \int\int \sqrt{|C_x(t,f)|}\sqrt{|C_y(t,f)|}\log\frac{1}{\sqrt{|C_x(t,f)|}\sqrt{|C_y(t,f)|}}dt\,df,$$
$$= H(\sqrt{|C_x(t,f)|}\sqrt{|C_y(t,f)|}). \quad (14)$$

This is equivalent to computing the entropy of the overlap between the two distributions and is similar to Jensen-Rényi divergence of order 1 with the difference being that the average entropy of the individual distributions is not subtracted. Therefore, the mutual information quantifies how similar the two distributions are rather than measuring how their 'joint' information is different than their individual information contents.

Some other important properties of this measure on the time-frequency plane are:

- $I(C_x,C_y)$ is a symmetric measure. Since

$$C_{yx}(t,f) = STFT_y(t,f)STFT_x^*(t,f), \quad = C_{xy}^*(t,f), \quad (15)$$

the magnitudes of the joint energy distributions are equal to each other.

- When the two signals, $x(t)$ and $y(t)$, are equal to each other, $I(C_x,C_y)$ equals to the entropy of the individual signals. This can be shown as follows:

$$I(C_x,C_x) = \int\int C_x(t,f)\log\frac{C_x(t,f)}{C_x(t,f)^2}dt\,df,$$
$$= -\int\int C_x(t,f)\log C_x(t,f)dt\,df,$$
$$= H(C_x). \quad (16)$$

For deterministic signals, this constitutes the maximum of mutual information since when the two signals are equal to each other, the dependence between the signals reaches its maximum.

- If $x(t)$ and $y(t)$ are well-separated on the time-frequency plane, i.e. their TFDs do not overlap, then the mutual information between them is equal to zero. When the two signals are separated on the time-frequency plane, $C_x(t,f)C_y(t,f) = 0, \forall t,f$. Therefore,

$$|C_{xy}(t,f)| = |STFT_x(t,f)STFT_y^*(t,f)|$$
$$= |STFT_x(t,f)||STFT_y^*(t,f)|,$$
$$= \sqrt{C_x(t,f)}\sqrt{C_y(t,f)},$$
$$= 0, \quad (17)$$

which implies that the mutual information $I(C_x,C_y) = 0$. This is analogous to the case where independent random variables have zero mutual information. Unlike random variables, the signals have energy distributions that are disjoint on the time-frequency plane, and are not statistically independent.

3. RESULTS

The event-related potentials (ERPs) analyzed in this paper are collected during an experiment that aims at differentiating between the responses of two different groups of subjects: spider phobics and non-phobics [2]. Nine spider phobics and seven non-phobics, serving as controls, were shown subliminally 40 blanks, 20 rectangles and 20 spiders in randomized order using a Harvard-type 3-field tachistoscope. The participants were asked to say if they saw a blank or a picture. Stimulus duration was set at 1ms and luminance at 5ft/lamb. Confirming stimulus subliminality was the finding that d', the probability of detection, was not significantly different from zero for the behaviorial response. The event-related potentials are recorded for 1 second before the stimulus and 1 second after the stimulus. In this study, the event-related potentials in response to the phobic stimulus, i.e. spider, will be analyzed at two electrodes, Cz, the central electrode, and Oz, the occipital electrode.

For each trial, the time-frequency distribution for the pre- and post-stimulus activities are computed. For each subject and each electrode the average of the pre- and post-stimulus activities is computed over 20 trials. The information-theoretic measure is then applied to these averages to determine the difference between pre- and post-stimulus activities. A two-way analysis of variance (ANOVA) is used to explore the interactions between the two factors, i.e. the electrode and the phobic group. A significant interaction was found such that the spider phobics showed a greater difference at Cz than Oz, while the non-phobics showed the opposite. This major finding is sustained for all of the information-theoretic measures, except mutual information, at the 10% significance level.

The results for the different information-theoretic measures can be summarized as follows:

- Kullback-Leibler Distance: This distance measure is only applicable to positive distributions, such as the spectrogram. There was significant interaction between the phobia group and the electrode at the 10% significance level ($p = 0.0776$). The mean and the standard errors for different subject groups and electrodes are summarized in Table 1.

Phobic Group	Electrode	Average Distance	Standard Error
Spider Phobics	Oz	0.028	0.0046
	Cz	0.042	0.0074
Non-Phobics	Oz	0.037	0.0091
	Cz	0.027	0.0035

Table 1: Kullback-Leibler Divergence for the Interaction of Phobic Group and Electrode for the Spider Stimulus

- Rényi Divergence: The Rényi divergence is tested for different values of α in the range $(0,1)$. It is observed that as $\alpha \to 1$, the distance values get closer to the one reported for Kullback-Leibler distance as expected. The interaction between the phobic group and the electrode is found to be significant for all tested values of α, though it is observed that the significance increases as α increases ($p = 0.0752$ for $\alpha = 0.8$.). The mean and the standard errors for $\alpha = 0.8$ are summarized in Table 2.
- Jensen-Shannon Divergence: This measure is based on Shannon entropy, and thus requires positive distributions. There was significant interaction between the phobia group and the electrode at the 10% significance level

[2]The author wishes to acknowledge Dr. Howard Shevrin and his research group at the University of Michigan for sharing this data.

Phobic Group	Electrode	Average Distance	Standard Error
Spider Phobics	Oz	0.0193	0.0036
	Cz	0.0315	0.0057
Non-Phobics	Oz	0.0278	0.0071
	Cz	0.0197	0.0025

Table 2: Rényi Divergence for the Interaction of Phobic Group and Electrode for the Spider Stimulus ($\alpha = 0.8$)

($p = 0.0752$). The mean and the standard errors for different subject groups and electrodes are summarized in Table 3.

Phobic Group	Electrode	Average Distance	Standard Error
Spider Phobics	Oz	0.006	0.0011
	Cz	0.009	0.0017
Non-Phobics	Oz	0.009	0.0021
	Cz	0.006	0.0008

Table 3: Jensen-Shannon Divergence

- Jensen-Rényi Divergence: This measure is based on the Rényi entropy, and can be applied on non-positive TFDs without any modification. The order α is greater than one, since for fractional order α the Rényi entropy will result in complex values. Different values of α are tested to understand the effect of even and odd orders. The interaction between the phobia group and the electrode was significant at the 5% significance level ($p = 0.0338$ for $\alpha = 2$ and $p = 0.023$ for $\alpha = 3$.) The mean and the standard errors for $\alpha = 3$ are summarized in Table 4.

Phobic Group	Electrode	Average Distance	Standard Error
Spider Phobics	Oz	0.04278	0.01150
	Cz	0.08767	0.01843
Non-Phobics	Oz	0.08453	0.02192
	Cz	0.04786	0.01326

Table 4: Jensen-Rényi Divergence for the Interaction of Phobic Group and Electrode for the Spider Stimulus ($\alpha = 3$)

- Mutual Information: The final information-theoretic measure tested was the mutual information. The mutual information was the only measure that did not show any significant interaction between the phobia group and the electrode. The major reason is that this measure quantifies the complexity of the overlap between the two distributions, but not the difference between 'joint' entropy and individual entropies. The mean and the standard errors for different subject groups and electrodes are summarized in Table 5.

Phobic Group	Electrode	Average Distance	Standard Error
Spider Phobics	Oz	12.444	0.0749
	Cz	12.540	0.0840
Non-Phobics	Oz	12.484	0.1239
	Cz	12.409	0.0515

Table 5: Mutual Information for the Interaction of Phobic Group and Electrode for the Spider Stimulus

From the results, it is clear that all of the applied measures give similar findings in terms of the interactions between the electrode and the phobic group. It is important to mention that Jensen-Rényi divergence yielded the most significant interaction between phobic group and electrode. This is mainly due to the robustness of Rényi entropy compared to Shannon entropy [3], as well as the way this measure is constructed. This measure quantifies the difference between the 'joint' entropy or entropy of the overlap and the individual entropies. All the other information-theoretic measures introduced in this paper quantify only the entropy of the overlap between the two distributions.

4. CONCLUSIONS

In this paper, we introduced a collection of information-theoretic measures that can be used to quantify the difference between signals on the time-frequency plane. These measures are based on treating the time-frequency distributions as two-dimensional probability density functions and adapting the well-known information-theoretic measures such as divergence measures and mutual information. We have applied the proposed measures on an ERP data set to discriminate between pre- and post-stimulus brain activities. It has been shown that all of the measures result in significant interactions between phobia group and the electrode. It is also observed that information-theoretic measures, such as Jensen-Rényi divergence, which quantify the difference between the entropy of the 'overlap' of TFDs and the individual entropies, perform better than measures that just quantify the complexity of the 'joint' TFD.

The proposed measures can be modified so that they are applicable to a large class of distributions including non-positive TFDs. The information-theoretic measures can also be optimized for the underlying signal classes by choosing the optimal parameter and the measure.

REFERENCES

[1] W. J. Williams, M. Brown, A. Hero, Uncertainty, information and time-frequency distributions, in: SPIE-Advanced Signal Processing Algorithms, Vol. 1556, 1991, pp. 144–156.

[2] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, O. Michel, Measuring time-frequency information content using the Rényi entropies, IEEE Trans. on Info. Theory 47 (4) (2001) 1391–1409.

[3] S. Aviyente, Information processing on the time-frequency plane, in: Proc. IEEE Int. Conf. Acoustics,Speech and Signal Processing, Vol. 2, 2004, pp. 617–620.

[4] A. J. Bell, T. J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.

[5] M. Basseville, Distance measures for signal processing and pattern recognition, Signal Processing 18 (4) (1989) 349–369.

[6] I. Csiszar, Information-type distance measures and indirect observations, Stud. Sci. Math. Hungar. 2 (1967) 299–318.

[7] L. Cohen, Time-Frequency Analysis, Prentice Hall, New Jersey, 1995.