A MEASURE OF MUTUAL INFORMATION ON THE TIME-FREQUENCY PLANE

Selin Aviyente

Department of Electrical and Computer Engineering,
Michigan State University, East Lansing, MI, 48824

ABSTRACT

Information-theoretic characterization of time-frequency distributions have been successful at quantifying the complexity of non-stationary signals. Information measures such as entropy and divergence have been adapted to the time-frequency domain for counting the number of signal components, evaluating the performance of different kernels and discriminating between signals based on their information content. Inspired by the success of these measures and in order to develop a more comprehensive information processing theory on the time-frequency plane, we introduce a mutual information measure for time-frequency distributions. The properties of this measure are derived and its application to signal classification problems is illustrated with examples.

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. In order to address the issue of information representation on the time-frequency plane, a comprehensive theory of information processing needs to be developed.

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 have 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 an information-theoretic measure that quantifies the dependencies between signals on the time-frequency plane. Mutual information is one such measure that has been used effectively in various statistical signal processing applications including classification and source separation [4].

The mutual information is defined as a measure of independence between random variables. In the case of time-frequency distributions, where the underlying signals are not necessarily random, a modification of the definition of mutual information is necessary. Instead of measuring the statistical dependence between signals, mutual information on the time-frequency plane should measure the dependence between the time-frequency distributions of the individual signals and the joint distribution of the signals. In this paper, we introduce a time-frequency based mutual information measure and present its properties.

In Section 2, a new measure based on mutual information is derived to quantify the dependence/disjointness of signals on the time-frequency plane and its major properties are discussed. Section 3 illustrates the application of this new measure to signal classification problems, and shows the reduction in classification error compared to conventional similarity measures such as correlation. Finally, Section 4 gives the conclusions and discusses some possible future work.

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] [5]:

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

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 $\phi(\theta,0)=\phi(0,\tau)=1\ \ \forall\tau,\theta$.

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

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

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. Definition of Mutual Information on the Time-Frequency Plane

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

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{3}$$

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, where we do not necessarily have random signals and TFDs are not actual probability density functions, 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 (3). Using the same definition, the joint density function in equation (3) will be replaced by the joint energy distribution of $x(t)$ and $y(t)$ defined by the cross-spectrogram of the two signals:

$$C_{xy}(t,f) = STFT_x(t,f)STFT_y^*(t,f), \tag{4}$$

where $STFT_x(t,f) = \int h(\tau - t)x(\tau)e^{-j2\pi f\tau}\,d\tau$ with $h(t)$ being the data window. $C_{xy}(t,f)$ is the joint energy distribution of $x(t)$ and $y(t)$ since its time marginal yields the cross energy of the two signals:

$$\int C_{xy}(t,f)\,df = x(t)y^*(t). \tag{5}$$

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 defined as:

$$I(C_x,C_y) = \int\int |C_{xy}(t,f)| \log \frac{|C_{xy}(t,f)|}{C_x(t,f)C_y(t,f)} dt\,df. \tag{6}$$

2.3. Properties of Mutual Information on the Time-Frequency plane

In this section some important properties of the mutual information measure will be derived.

- $I(C_x,C_y)$ is a symmetric measure. In order to prove this property, one needs to show that $|C_{xy}| = |C_{yx}|$. Since

$$\begin{aligned}C_{yx}(t,f) &= STFT_y(t,f)STFT_x^*(t,f),\\ &= C_{xy}^*(t,f),\end{aligned} \tag{7}$$

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

- If 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:

$$\begin{aligned}I(C_x,C_x) &= \int\int C_x(t,f) \log \frac{C_x(t,f)}{C_x(t,f)^2} dt\,df,\\ &= -\int\int C_x(t,f)\log C_x(t,f) dt\,df,\\ &= H(C_x).\end{aligned} \tag{8}$$

  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. Note that we would end up with a similar result for signals that are equal to each other except for an amplitude scale factor, since time-frequency distributions are normalized before computing any information-theoretic quantity.

- 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 well-separated on the time-frequency plane, $C_x(t,f)C_y(t,f) = 0, \forall t,f$. Therefore,

$$\begin{aligned}|C_{xy}(t,f)| &= |STFT_x(t,f)STFT_y^*(t,f)|,\\ &= |STFT_x(t,f)||STFT_y^*(t,f)|,\\ &= \sqrt{C_x(t,f)}\sqrt{C_y(t,f)},\\ &= 0,\end{aligned} \tag{9}$$

which implies that the mutual information $I(C_x, C_y) = 0$. This is analogous to independent random variables having zero mutual information.

- The two properties derived above can be combined to define another information-theoretic measure: conditional entropy. The difference between the entropy of the signal $x(t)$ and the mutual information is defined as the conditional entropy of signal $x(t)$ given $y(t)$. This quantity is equal to zero when $x(t) = y(t)$ by the second property, and is equal to $H(C_x)$ when $x(t)$ and $y(t)$ are well-separated on the time-frequency plane by the third property. The conditional entropy of signal $x(t)$ given $y(t)$ can be defined as:

$$H(C_x|C_y) = H(C_x) - I(C_x, C_y). \qquad (10)$$

The properties derived above can be easily illustrated for a sample signal. Consider two gabor logons separated by Δt from each other, i.e. $x(t) = g(t) + g(t - \Delta t)$, where $g(t) = \exp(-\frac{(t-t_0)^2}{2\sigma^2})\exp(-j2\pi f_0 t)$. By the second and third properties, the mutual information between $x(t)$ and $g(t)$ equals to the entropy of $g(t)$ when $\Delta t = 0$, and goes to zero as $\Delta t \to \infty$. This phenomenon is illustrated in Fig. 1, where the maximum of the mutual information is equal to 5.98 which is equal to the entropy of a single gabor logon.

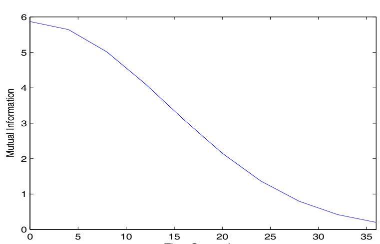

Fig. 1. Mutual information between two gabor logons with respect to the time separation, Δt

3. RESULTS

In this section, the application of the proposed mutual information measure for signal classification on the time-frequency plane will be illustrated through examples. Signal classification on the time-frequency plane has been considered in previous research using different metrics such as correlation and mean square error [6]. The performance of the proposed mutual information measure will be compared to these conventional metrics.

Example 1: The two signal classes considered in this example are:

$$C_1 : y(t) = \exp(-j\omega_0 t) + \exp(-(\beta_1 t^2 + \omega_1 t)) + v(t),$$
$$C_2 : y(t) = \exp(-j\omega_0 t) + \exp(-(\beta_2 t^2 + \omega_2 t)) + v(t), \qquad (11)$$

where $\omega_0 = 0.3$, β_1 is uniformly distributed between [0.01,0.4], ω_1 is uniformly distributed between [0.01, 1.01], β_2 is uniformly distributed between [0.2,0.9], ω_2 is uniformly distributed between [0.2,0.3], and $v(t)$ is white Gaussian noise. A training set consisting of 100 samples of each class is constructed, and the average TFDs representing each class, $\overline{C}_1(t, f)$ and $\overline{C}_2(t, f)$, are computed. A test set consisting of 100 samples of each class is generated and the time-frequency distribution of each sample is compared to the class averages using mutual information, i.e. $I(C_x, \overline{C_1})$ and $I(C_x, \overline{C_2})$ are computed. The classification decision is made to choose the class that has the higher mutual information with the given sample signal as follows:

$$\hat{C} = argmax_{\{i=1,2\}} I(C_x, \overline{C_i}), \qquad (12)$$

where $\hat{C}$ is the class assigned to signal x. The probability of error can then be computed as:

$$P_e = P(\hat{C} \neq C_1|C_1)P(C_1) + P(\hat{C} \neq C_2|C_2)P(C_2). \qquad (13)$$

The performance of the mutual information measure for this classification problem is quantified for different SNRs by computing the probability of error. Fig. 2 illustrates the probability of error in classification versus SNR. It is observed that even at high noise levels the probability of error is low.

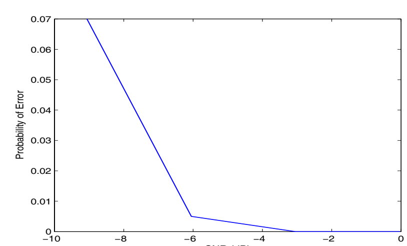

Fig. 2. Probability of error versus SNR for mutual information based classification

The proposed mutual information measure is also compared to conventional measures such as correlation. It is observed that even though the probability of errors are comparable in the case of Gaussian noise, the mutual information measure is superior to correlation-based measures when the noise distribution is non-Gaussian. To illustrate this point, the classification problem formulated by equation (11) is considered with non-Gaussian noise. The probability of error in classification versus SNR for both the mutual information and correlation measures is illustrated in Fig. 3. The results indicate that as the SNR decreases the deviation of the TFD from a Gaussian distribution increases, thus mutual information becomes a better discrimination measure compared to correlation.

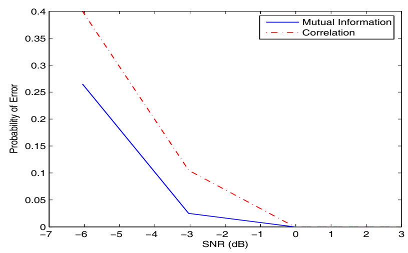

Fig. 3. Probability of error versus SNR for mutual information and correlation based classifiers

Example 2: In this example, the proposed measure is applied to a fault detection problem in fuel pumps. The current through a faulty electric motor is a non-stationary signal with short transients which are indicative of the faults [7] [2]. The proposed mutual information measure is applied to the problem of classifying between 3 different types of faults:

1. One of the coils is poorly fused causing its resistance to be increased.
2. The coil is cut entirely.
3. The commutator face was scored during assembly.

Six motors belonging to each class are available. The proposed mutual information measure is applied to this classification problem, measuring the dependencies between each class:

$$\hat{C} = argmax_{\{i=1,2,3\}} I(C_x, \overline{C_i}), \qquad (14)$$

where $\overline{C_i}$ is the average TFD of class i. One sample is taken out at a time and the classifier is trained on the remaining five samples and is tested on the sixth one. The classification performance of mutual information and correlation measures are compared. Using the mutual information measure 16 out 18 motors are correctly classified whereas using the correlation measure 14 out of 18 motors are correctly classified.

4. CONCLUSIONS

In this paper, we introduced a new information-theoretic measure, mutual information, on the time-frequency plane. The major properties of this measure, such as its maximum and minimum points, are derived. It is shown that mutual information can quantify the disjointness of two signals on the time-frequency plane similar to determining statistical independence in information theory. The application of this measure in signal classification is illustrated through both simulations and real-life signals. It is observed that mutual information is a better measure in classifying non-Gaussian signals compared to the conventional discrimination measures such as correlation.

The results presented in this paper can be extended to non-positive distributions by modifying the definition of mutual information using generalized entropy measures such as Rényi's entropies [8]. The definition of mutual information can also be generalized to include more than two signals at a time.

[2] The author wishes to thank Elias Strangas and Wes Zanardelli of Michigan State University for the electric motor data and for the permission to use it in this paper.

5. REFERENCES

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

[2] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Trans. on Info. Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.

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

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

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

[6] M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels, "Improved optimization of time-frequency based signal classifiers," *IEEE Signal Processing Letters*, vol. 8, no. 2, pp. 52–57, Feb. 2001.

[7] W. G. Zanardelli, E. G. Strangas, and H. K. Khalil, "Comparison of wavelet-based methods for the prognosis of failures in electric motors," in *IEEE Workshop on Power Electronics in Transportation*, 2002, pp. 61–67.

[8] A. Rényi, "On measures of entropy and information," in *Proceedings 4th Berkeley Symp. Math. Stat. and Prob.*, 1961, vol. 1, pp. 547–561.