# ADAPTIVE MINIMUM ENTROPY DECOMPOSITION ON THE TIME-FREQUENCY PLANE

*Zeyong Shan and Selin Aviyente*

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

## ABSTRACT

In many applications, such as array processing and sensor networks, it is desirable to extract the source signals that generate the observed output signals. Some common approaches include principal component analysis, which assumes uncorrelated source signals, and independent component analysis, which assumes the independence of the underlying sources. In recent years, there has been efforts to perform source separation in the time-frequency domain since most real life signals of interest are non-stationary [1]. In this paper, we introduce one such component extraction approach on the time-frequency plane. The proposed approach extracts components that are well-concentrated on the time-frequency plane. In order to quantify the compactness or the concentration of the extracted components, we use the entropy measure as adapted to the time-frequency distributions. It has been shown that signals which achieve minimum entropy on the time-frequency plane are gabor logons. Based on this idea, we propose an adaptive gabor logon extraction method from a given set of observed signals. The proposed method extracts the most significant gabor logons as the components using an adaptive filtering approach. The method is applied on an example data set to show the effectiveness of the component extraction algorithm.

## 1. INTRODUCTION

In many signal processing applications, it is desirable to extract the sources underlying the observed signals. Some examples include source separation in array processing and sensor network applications. The most common methods for source or component extraction include principal component analysis (PCA) and independent component analysis [2]. These methods are effective at extracting orthogonal components and assume the stationarity of the underlying signals. Most real life signals are not stationary and thus do not obey this underlying assumption. For this reason, we are introducing a component extraction method on the time-frequency plane. This method relies on extracting components that are well-concentrated on the time-frequency plane. The concentration of the components are quantified through an entropy measure on the time-frequency plane. Since it has been shown in the literature that signals that achieve a small entropy value on the time-frequency plane are gabor logons, our component extraction algorithm reduces to extracting the gabor logons that best describe the given data set in a minimum mean square sense. Unlike the traditional gabor decomposition [3], where the signal is expressed as an infinite sum of time and frequency shifted gabor logons, the components extracted by this algorithm have time and frequency centers determined by the signal. Moreover, the components extracted in this approach have chirp rates and local spread adapted to the given set of signals. The goal is to represent the given data set with a few number of chirped gabor logons.

This paper is organized as follows. Section 2 reviews the background on entropy measures on the time-frequency plane, formulates the decomposition problem, and introduces the adaptive time-frequency decomposition algorithm. Section 3 presents an example and illustrates the results. Finally, Section 4 discusses the major contributions of this paper as well as some future work.

## 2. METHOD

### 2.1. Entropy Measure on the Time-Frequency Plane

A time-frequency distribution (TFD), $C(t,f)$, from Cohen's class can be expressed as [1] [4]:

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

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 $\infty$ unless otherwise stated.

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

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 well-known Shannon entropy for TFDs is defined as:

$$H(C) = -\iint C(t,\omega)\log_2 C(t,\omega)dt\,d\omega. \tag{3}$$

This measure is only defined when $C(t,\omega) > 0$, $\forall t,\omega$. Therefore, it is valid for positive distributions such as the spectrogram. In [5], Rényi entropy was introduced as an alternative way of measuring the complexity of TFDs and the properties of this measure were proved extensively in [6]:

$$H_\alpha(C) = \frac{1}{1-\alpha}\log_2 \iint \left(\frac{C(t,\omega)}{\iint C(u,v)du\,dv}\right)^\alpha dt\,d\omega. \tag{4}$$

It has been shown that the minimum value of entropy on the time-frequency plane is achieved for a gabor logon [6]. This is also consistent with the fact that the gabor logon is the signal that achieves the lower bound on the uncertainty on the time-frequency plane [4]. For this reason, our signal decomposition algorithm is based on extracting a set of well-concentrated components, that best described the given data set.

### 2.2. Problem Statement

Given $M$ measurements of a signal, $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$, we want to extract the first $L$ components, $L < M$, that minimize entropy on the time-frequency plane. Each signal, $\mathbf{x}_i$, is transformed to the time-frequency plane as:

$$C_i(n,\omega;\psi) = \sum_m \sum_l \psi(n-l,m) x_i\left(l+\frac{m}{2}\right) x_i^*\left(l-\frac{m}{2}\right) e^{-j\omega m} \tag{5}$$

The time-frequency distribution corresponding to each trial is vectorized and a matrix of time-frequency distributions is formed:

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_M \end{bmatrix}, \tag{6}$$

where $C_i$ is a vector of length $N \times K$ points, $N$ and $K$ being the number of time and frequency points, respectively. The components on the time-frequency plane are found based on this time-frequency data matrix. Each component $TF_i$ is a linear combination of the rows of this matrix, i.e.

$$TF_i(n,k) = \sum_{j=1}^{M} a_j C_j(n,k), \tag{7}$$

where $\sqrt{\sum_j {a_j}^2} = 1$ and $a_j$s are chosen such that $H_\alpha(TF_i)$ is minimized on the time-frequency plane.

### 2.3. The Proposed Approach

Since gabor logon signals have minimum entropy on the time-frequency domain, the cost function is chosen as $e = H(TF_i) - H^*$, where $H(TF_i)$ is the entropy of $i$th component, and $H^*$ the entropy of the corresponding desired logon signal. The weight vector $\mathbf{a} = [a_1, a_2, \cdots, a_M]^T$ is updated using the method of Steepest Descent [7], which is

$$\hat{\mathbf{a}} = \mathbf{a} - \mu\frac{\partial e}{\partial \mathbf{a}} \tag{8}$$

where $\mu$ is the step size parameter. In the discrete case, Rényi entropy of the component $TF_i$ is

$$H(TF_i) = H(\mathbf{a}^T C) = \frac{1}{1-\alpha}\log_2 \sum_n \sum_k \left(\sum_{j=1}^{N} a_j C_j(n,k)\right)^\alpha \tag{9}$$

where $TF_i$ and $C_j$ are normalized. The gradient of the cost function $e$ with respect to the $l$th weight coefficient $a_l$ is derived as:

$$\frac{\partial e}{\partial a_l} = \frac{\alpha}{1-\alpha}\frac{\sum_n \sum_k (TF_i(n,k))^{\alpha-1} C_l(n,k)}{\sum_n \sum_k (TF_i(n,k))^\alpha} \tag{10}$$

where $l = 1, \cdots, M$. For the special case of $\alpha = 2$,

$$\frac{\partial e}{\partial a_l} = -2\frac{\sum_n \sum_k TF_i(n,k) C_l(n,k)}{\sum_n \sum_k (TF_i(n,k))^2} \tag{11}$$

Substituting the results in equation (11) into equation (8) yields the update equation for **a** as:

$$\hat{a}_l = a_l + 2\mu\frac{\sum_n \sum_k TF_i(n,k) C_l(n,k)}{\sum_n \sum_k (TF_i(n,k))^2} \tag{12}$$

The block diagram of the proposed approach is shown in Fig. 1.

The algorithm can be summarized as follows:

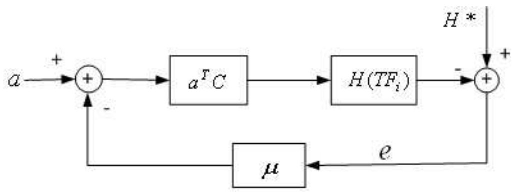


**Fig. 1**. The block diagram of the system

1. Find the gabor logon that best describes the average of all trials, $C_{av} = \frac{1}{M}\sum_j C_j$. This first gabor logon is found by finding the average time duration, average frequency, the spread, and the chirp rate of $C_{av}$. A logon with these estimated parameters is constructed and chosen as the first desired signal,$G(n, k; n_0, k_0, \sigma, \beta)$.

2. Set the initial value for $a_j = \frac{1}{\sqrt{N}}$, and use the adaptive filtering algorithm to update the weights until the error converges. The first component is then determined as, $TF_1 = \mathbf{a_{opt}}^T \mathbf{C}$, where $\mathbf{a}_{opt}$ is the optimal weighting vector.

3. Project all the trials on $TF_1$ and compute the residue.

$$\hat{C}_i = C_i - < TF_1, C_i > C_i \; i = 1, 2, \cdots, M \quad (13)$$

4. Repeat the same algorithm on this residue matrix $\hat{\mathbf{C}}$, and extract the next component.

5. Stop when the average energy of the residues drops below a pre-determined threshold value.

## 3. EXPERIMENTAL RESULTS AND ANALYSIS

In order to evaluate the effectiveness of our method, we consider the following example. The set of observed signals are linear combinations of two gabor logons and a chirp signal, i.e. $x_i = w_{i1}s_1 + w_{i2}s_2 + w_{i3}s_3$, where $w_{i1}, w_{i2}, w_{i3}$ are the weights for each signal and are distributed as $N(0, 1)$. The first gabor logon is centered at the time sample point 50 and normalized frequency of 0.7, the second gabor logon is centered at time sample point 150 and normalized frequency of -0.7. The linear chirp signal has an initial normalized frequency of -0.2 and its instantaneous frequency increases to a normalized frequency of 0.2. Rényi entropy with $\alpha = 2$ is used as the cost function to ensure that entropy is well-defined. The data set consists of $M = 128$ linear combinations of these three signals. Each signal is transformed to the time-frequency domain with $N = 50$ time samples and $K = 64$ frequency samples. Each TFD is then vectorized to form a TFD matrix of size $128 \times 3200$.

First, the average of $M$ TFDs corresponding to each trial is computed. Then, the time-frequency location of the peak energy on the time-frequency plane is found as $n_0$ and $k_0$. A window centered at $(n_0, k_0)$ is constructed to determine a local region around this peak. The size of the window is determined based on the energy distribution of the signal, i.e. the window is expanded until the energy value drops below $10\%$ of the peak value. This windowing approach around the peak helps us extract local features. The same window is applied to all trials to extract the corresponding regions in each trial. The standard deviation $\sigma$ and gradient (the chirp rate), $\beta$, of this local TFD are estimated. Based on the parameters $(n_0, k_0, \sigma, \beta)$, a gabor logon is constructed and chosen as the first desired signal. Using the steepest descent algorithm, the weight coefficients $a_j$s are updated to minimize the difference of Rényi entropy between the linear combination of the $M$ local TFDs and the TFD of the first desired logon to obtain the first time-frequency component, $TF_1$. This first component is projected onto all of the $M$ trials and the residue is found. This same algorithm is repeated for the residue on the time-frequency plane, i.e. pick the peak, construct a window, determine the desired gabor logon, and adaptively filter the signals to get close to the desired gabor logon. This process is repeated until the energy of the residue is below a certain threshold. In this example, five components were enough to represent $99\%$ of the total energy of the signal.

Table 1 gives the entropy values for these five components, the corresponding desired logon signals, and the first five components obtained using PCA. It is shown in Table 1 that the entropy of the gabor logons are close to the entropy of the extracted components. Since the entropy differences between these five components and desired logon signals are small, we can infer that the extracted signals are quite close to actual logons. It is also seen that the entropy of components extracted by our method is less than the entropy of PCA components. This indicates that we obtain components that are more compact than the ones obtained by PCA.

The time-frequency surfaces in Fig. 2 indicate that the five components extracted include both the logon signals and three chirped logons that represent the linear chirp signal. The topographical plots of the extracted components make it clear that each component was appropriately isolated in terms of the topographical region of origin.

The results of this preliminary example show that the decomposition of time-frequency energy using our approach can be used to extract meaningful time-frequency components for analysis of large sets of data. This decomposition algorithm achieves several goals. First, time-frequency data reduction is accomplished by producing a few meaningful components on the time-frequency plane that explain most of the signal's energy. A second benefit of this time-frequency domain decomposition is that it can extract activity that overlaps in time, but not in frequency, which is

**Table 1**. Entropy Comparison

| Entropy | Our Comps | PCA Comps | Desired Logon |
|---|---|---|---|
| 1 | 2.8119 | 2.8123 | 2.7609 |
| 2 | 2.8144 | 2.9341 | 2.7613 |
| 3 | 2.8117 | 2.8118 | 2.2248 |
| 4 | 2.8002 | 2.8065 | 2.0377 |
| 5 | 2.8002 | 2.8017 | 2.0352 |

not possible using time domain decomposition approaches. Finally, another benefit of our method is the ability to separate and extract parts of chirped signals, which cannot be achieved using the conventional gabor expansion.

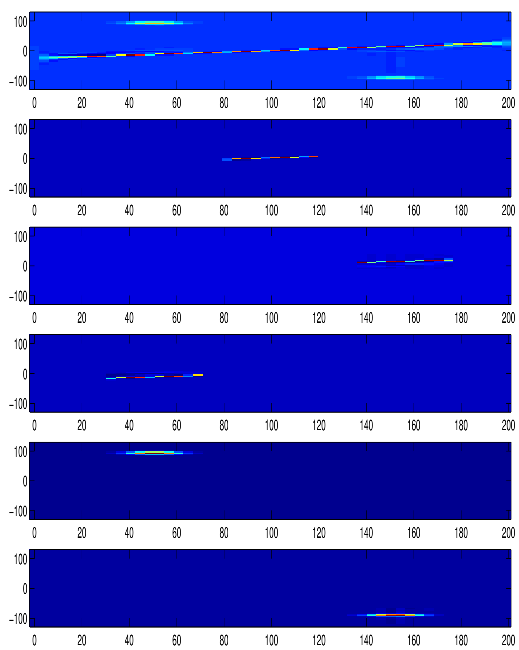

**Fig. 2**. The average time-frequency distribution of 128 trials and the first 5 components

## 4. CONCLUSIONS

In this paper, a new signal decomposition method on the time-frequency plane is proposed based on the minimum entropy criterion. The major difference of the proposed approach from conventional component extraction or decomposition methods is the cost function. The cost function that is minimized is entropy on the time-frequency plane, thus producing compact components that are similar to gabor logons. Using entropy as the cost function, and adopting an adaptive filtering method to update the weights corresponding to each trial we extract 'minimum' entropy components orthogonal to each other. Experimental results show that our approach is effective in determining a few number of components that can be used to represent a large set of data.

Future work includes determining the convergence rate of the algorithm and applying the proposed method to real life data sets. The proposed approach may be applied to different areas such as array processing, sensor data fusion and source extraction in biomedical signal processing.

## 5. REFERENCES


[1] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. on Signal Processing*, vol. 46, pp. 2888 – 2897, 1998.

[2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley and Sons, 2001.

[3] S. Qian and D. Chen, "Discrete gabor transform," *IEEE Trans. on Signal Processing*, vol. 41, pp. 2429–2438, 1993.

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

[5] 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.

[6] 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.

[7] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985.