# TIME-FREQUENCY BASED CLASSIFICATION†

Berkant Tacer and Patrick Loughlin
Dept. of Electrical Engineering
University of Pittsburgh, Pittsburgh, PA 15261
*E-mail:* tacer@ee.pitt.edu *Tel:* (412) 624-9694 *Fax:* (412) 624-8003

## ABSTRACT

We propose a time-frequency based pattern classification method which utilizes the joint moments of time-frequency distributions (TFDs) for features. The method is applied to a biomedical data set, and compared to a template matching scheme and to methods utilizing only temporal moments or spectral moments. Our results show that a classification algorithm which utilizes joint time-frequency information, as quantified by the joint moments of the TFD, can potentially improve performance over time or frequency-based methods alone, for classification of nonstationary time series.

**Keywords:** time-frequency distributions, classification, joint moments, Fisher's linear discriminant, principal components analysis

## 1. INTRODUCTION

Many natural and man-made signals exhibit time-varying, or nonstationary, characteristics. Examples include speech, music, biological signals, and AM and FM waves. Time-frequency distributions (TFDs) have been successfully applied to such signals to study their changing spectral properties. [1, 2, 3, 4, 5, 6] It is natural to expect that TFDs should offer the potential for enhanced classification performance when dealing with nonstationary signals. However, there is a dimensionality problem: For every $N$-point discrete-time signal, the TFD is $O(N^2)$, and hence it would appear that far more data are needed for time-frequency based classification than otherwise.

A potential solution to the dimensionality problem can be found in image processing, where joint moments have been extensively used for recognition purposes. [7] In this paper, we propose a feature set based on the joint moments of TFDs for classification of time-varying signals. Using only a few joint moments of the TFD, significant reduction in the dimensionality of the problem is achieved, while at the same time salient time-frequency information is captured, enabling accurate and efficient classification.

In our approach, several joint moments of a TFD for each time series in a training set are computed, and then normalized. The dimensionality of this set is then reduced using one of two different approaches: principal components analysis, or exhaustive search. This reduced feature vector is used to train a linear discriminant function, which is then used to perform classification on features from time-series outside the training set. The proposed method is tested using a Jack-knife method on a (limited) biomedical data set, with encouraging results.

## 2. BACKGROUND

### 2.1 Temporal, Spectral and Joint Time-Frequency Moments

The temporal and spectral moments of a signal $s(t)$ are given by

$$\langle t^n \rangle = \int_{-\infty}^{\infty} t^n |s(t)|^2 dt \tag{1}$$

$$\langle \omega^m \rangle = \int_{-\infty}^{\infty} \omega^m |S(\omega)|^2 d\omega \tag{2}$$

$(n, m = 1, 2, \dots)$ respectively, where $S(\omega) = \int_{-\infty}^{\infty} s(t)\, e^{-j\omega t} dt$ is the Fourier transform of the signal. Similarly, the joint time-frequency moments are given by

† This research was supported in part by grants from Boeing Commercial Airplane Group, the Office of Naval Research and the Whitaker Foundation.



0-8194-2234-7/96/$6.00

$$\langle t^n \omega^m \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t^n \omega^m P(t, \omega)\, dt d\omega \text{ for } n, m = 1, 2, \ldots \tag{3}$$

where $P(t, \omega)$ is the TFD of the signal.

While there are currently many possibilities for the TFD of a signal, we utilize the Cohen-Posch TFDs, [8] implemented via the method of Loughlin, Pitton and Atlas. [9] These TFDs are non-negative and yield the correct marginal densities when integrated over time or frequency:

$$P(t, \omega) \geq 0 \tag{4}$$

$$\int_{-\infty}^{\infty} P(t, \omega)\, dt = |S(\omega)|^2 \tag{5}$$

$$\int_{-\infty}^{\infty} P(t, \omega)\, d\omega = |s(t)|^2. \tag{6}$$

Accordingly, these TFDs are interpretable as joint energy density functions of the signal. The marginals ensure that temporal and spectral moments are correctly given. The marginals, together with non-negativity, further ensure that the joint (and conditional) moments are sensible and consistent with the notion of a joint density function. The TFDs are unit-energy normalized for further consistency with distribution theory.

### 2.2 Fisher's Linear Discriminant Function

The decision surfaces which separate different classes are estimated from training samples. If the classes in a feature space are linearly separable, one can find linear discriminant functions that correctly classify the training samples. For the 2-class problem, Fisher's method provides an analytic solution for the classification (decision) function by mapping an $m$-dimensional feature space onto a line.

In Fisher's algorithm, a linear function $w^T x_i$ which maps a set of $n$ feature vectors $x_1, x_2, \ldots, x_n$ from $R^m$ to $R^1$ is found by maximizing the criterion function [11]

$$C(w) = \frac{|\mu_1 - \mu_2|^2}{s_1^2 + s_2^2} \tag{7}$$

where $\mu_1$ and $\mu_2$ are the sample means and $s_1^2$ and $s_2^2$ are the scatters of the two classes in one dimensional space ($R^1$). They are defined as

$$\mu_i = \frac{1}{n_i} \sum_{x \in X_i} w^T x \text{ and } s_i^2 = \sum_{x \in X_i} \left( w^T x - \mu_i \right)^2 \text{ for } i = 1, 2 \tag{8}$$

where $n_i$ is the number of feature vectors in each class $X_i$. The solution of the maximization of the above criterion function is

$$w_s = (S_1 + S_2)^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \tag{9}$$

where $\underline{\mu}_1$ and $\underline{\mu}_2$ are the sample mean vectors, and $S_1$ and $S_2$ are the scatter matrices of the two classes $X_1$ and $X_2$ respectively.

Then, the decision for an unlabeled sample $x_u$ is made in favor of class one if

$$|w_s^T x_u - \mu_1| < |w_s^T x_u - \mu_2|, \tag{10}$$

or in favor of class two otherwise.

We used this method to classify the signals using their univariate or joint moments as the feature vectors.

### 2.3 Principal Components Analysis

Principal components analysis [12] is a technique for reducing the dimensionality of feature space. Using this technique, the original coordinates of feature vectors are mapped to a smaller number of coordinates (principal components) that account for a large portion of the total variance. The mapping is accomplished by taking linear combinations of the original coordinates.

The principal components of a feature space are determined by calculating the eigenvalues and eigenvectors of the sample covariance matrix of the feature vectors. If the sample covariance matrix $K$ is a non-singular square matrix, then

$$K = PDP^T \tag{11}$$

where $D$ is a diagonal, and $P$ is an orthonormal matrix, i.e., $PP^T = I$. The diagonal elements of $D$ are the eigenvalues of $K$, and the columns of $P$ are the eigenvectors (principal components) of $K$. Since the eigenvectors form a complete basis, the feature vectors can be written in terms of these eigenvectors. The principal components whose contribution to the total variance is negligible (i.e., below a predetermined threshold) are discarded to decrease the dimension of the feature space. To determine how many principal components to keep, a number of methods can be used. [12]

## 3. METHODS

In this section, we describe in greater detail the generation of the joint moment-based features and our classification procedure. The method is applied to a biomedical data set (23 postural sway time series, to be classified as balance impaired (10) or unimpaired (13)), and compared to other (univariate) moment-based methods and a template matching procedure.

### 3.1 Classification Using Joint Moments

We begin by calculating $p \ll N^2$ (normalized) joint moments of the $O(N^2)$ discrete TFD of each $N$ point time series in the training set. These are then log-normalized for variance stabilization and to reduce the dynamic range of the features. The feature vector is thus the set††

$$f_l = \{\log \frac{\langle t^n \omega^m \rangle}{n!m!}\} \text{ for } n, m = 1, 2, \ldots < N. \tag{12}$$

For the postural sway data set, 25 joint moments were computed for each time series. For comparison, the size of the TFD for each series was 180x128. Often, the number of training samples is limited (as is the case here), which limits the number of moments that can be used (for robust classification). In general, the number of samples should be significantly greater than the number of features. [11] Accordingly, while the joint moments represent a reduced feature space with respect to the TFD, further dimensional reduction of the feature space is required, particularly when limited training samples are available. In such cases, methods are available to obtain a minimal set of moment-features for classification. We investigated the use of exhaustive search techniques, and principal components analysis.

For exhaustive search, we computed the criterion (see eqn. (7))

$$\alpha = \frac{|\mu_1 - \mu_2|^2}{s_1^2 + s_2^2} \tag{13}$$

for 5000 randomly chosen (with replacement) sets of five (out of the 25) moment features of the training samples. A histogram of all sets with an $\alpha$ value greater than the threshold $0.8max(\alpha)$ was made (approximately 250 sets), from which the five moments that occurred most frequently were selected to comprise the reduced feature set.

For principal components analysis, the feature set needs to be further normalized to prevent moments with large magnitudes from dominating the total variance. After the feature vectors are normalized such that all of the coordinates have zero mean and unit variance, the sample covariance matrix is calculated. Five of the eigenvectors of this normalized covariance matrix were kept to obtain the reduced feature vectors. To determine the selected eigenvectors, the eigenvalues were plotted in descending order to see an elbow in the curve. [12] Then, the number of eigenvectors kept was found by determining the point around the elbow at which the performance of the classification is the best.

†† We note that the elements $\langle t^n \omega^m \rangle / n!m!$ are the Taylor series coefficients of the characteristic function of the TFD. The factorial normalization of the moments $\langle t^n \omega^m \rangle$ enhances numerical stability in computations for large $m, n$.

A Jack-knife method was used for training and testing both procedures: 22 of the 23 time series were used to define the reduced feature set and train the weight vector. The remaining time series was then classified via Fisher's method using this weight vector. This procedure was repeated for each one of the time series, and performance scores were taken as the number of correct classifications out of the total number classified.

### 3.2 Classification Using Temporal and Spectral Moments

To determine whether or not the joint moments carried any time-frequency information pertinent for classification, the time series were also classified using temporal moments only, and spectral moments only. From the first ten log and variance normalized temporal moments ($n = 1, \ldots, 10$), a reduced feature set of five was obtained via principal components analysis. This set was then used in a Jack-knife method to train and test classification using temporal moment features only. An analogous procedure was carried out for spectral moments.

### 3.3 Classification Using Template Matching

Finally, for a "baseline" comparison, we used a simple template matching algorithm to classify the time series. The distance measure

$$\iint (P(t, \omega) - T_k(t, \omega))^2 dt d\omega \quad \text{for } k = 1, 2 \tag{14}$$

was computed, where $P(t, \omega)$ is an unlabeled TFD from the test set and $T_k(t, \omega)$ is the ensemble TFD of the $k$-th training class (impaired or unimpaired here). The unlabeled TFD was assigned to the class it was closest to. The population ensembles and representative TFDs of subjects are shown in Figures 1 and 2. The performance of this classification method was again assessed using a Jack-knife method.

## 4. RESULTS AND DISCUSSION

Table 1 summarizes the classification performance of each method.

**Table 1: Classification of Postural Sway Time Series**

| Approach Used | Correct Classification |
|---|---|
| Joint Moments (Principal Components) | 87% |
| Joint Moments (Exhaustive Search) | 87% |
| Spectral Moments | 83% |
| Template Matching | 65% |
| Temporal Moments | 61% |

As seen, the performance of the joint moments is superior to the other methods. However, given the small data set used in this experiment, the performance of spectral moments can be considered as good as that of the joint moments. We believe that the lack of significant frequency modulations in the time series (see Figures 1 and 2) is a contributing factor to the similar performance of spectral and joint moments. We expect that for time series with class-dependent frequency modulations, the joint moments would outperform the spectral moments. We are currently investigating this hypothesis. However, we do note that there are indeed time-dependent spectral changes within and between the two populations, as exemplified by the abrupt appearance of the 0.25 Hz frequency component at 30 seconds, and the decay of this component over time in the healthy population but not in the impaired group. In a separate study, these differences were found to be statistically significant ($p<0.05$).[10] Accordingly, we would expect that the small performance advantage of joint moments over spectral moments for this classification task would hold up under a larger population study.

With regard to the particulars of our approach, we make a few general observations. Although we used Fisher's linear discriminator here because of its computational simplicity, there are no doubt tasks for which other linear or nonlinear discriminators may be better suited. Exhaustive search methods, which in theory are optimal for obtaining the best reduced feature set, are clearly not an option in most cases. Even for the limited feature set investigated here, computational requirements precluded a true exhaustive search of all possible feature combinations. Therefore, one has to resort to other methods such as principal components analysis to avoid this computational problem. The main disadvantage of principal components analysis, on the other hand, is that one has to calculate all $p$ joint moments for each new test sample in order to classify it, whereas with exhaustive search, only the previously identified subset of moments need be computed. This is due to the fact that with principal components analysis, the new features are formed by using linear combinations of the old features.

Finally, we make no claims as to the "optimality" of our proposed joint moment-based classification approach. It is very often possible to tailor a particular approach to a specific problem by finding features unique to that process. Our goal, however, was to investigate the potential of generic time-frequency based features for classifying nonstationary time series. As we have shown, information in the time-frequency plane pertinent to classification is readily extracted via joint moments. This, we believe, is an attractive property for a "general" classifier, since it obviates the time and effort required to find unique, specific features for different types of signals (e.g., the presence or absence of a specific frequency or modulation pattern, which would not normally generalize to other processes). Other general time-frequency features may certainly be good candidates, as well, provided they capture salient time-frequency differences and are sparse.

## 5. CONCLUSION

In this paper, we presented a method for classifying nonstationary time series utilizing a feature set based on the joint moments of a time-frequency distribution of the time series. The method was applied to biomedical time series, with comparison to template matching and univariate moment-based methods. The results suggest that a classification algorithm which utilizes a sparse feature set that captures joint time-frequency information can improve performance over time or frequency-based methods alone, for classification of nonstationary time series.

Although the joint moments of a time-frequency distribution were utilized in this paper, our approach can be applied to other joint distributions, such as time-scale [13], scale-frequency, etc. In addition, it may be possible to compute the joint moments directly from the signal, without the need to compute the joint distribution. [14] Such an option is clearly attractive from a computational viewpoint, and we believe merits further investigation, as does the general topic of joint distribution-based classification of nonstationary time series.

## 6. REFERENCES


1. *Proceedings of the IEEE*, special issue on "Time-Frequency Analysis: Biomedical, Acoustical and Industrial Applications," P. Loughlin (ed.), vol. 84, no. 9, Sep. 1996.
2. E. Velez and R.Absher, "Segmentation and classification of phonemes in continuous speech based on their time-frequency representations," *Proc. SPIE*, vol. 1348, pp. 188-196, 1990.
3. H. Oehlmann et al., "Examination of gear box cracks using time-frequency distributions," *IEEE Proc. ICASSP'95*, pp. 925-928, 1995.
4. B. Boushash and P. O'Shea, "A methodology for Detection and Classification of Some Underwater Acoustic Signals Using Time-Frequency Analysis techniques," *IEEE Trans. Acoust., Speech and Sig. Proc.*, vol. 38, no. 11, pp.1829-1841, 1990.
5. M. Brown, W. Williams and A. Hero, "Non-orthogonal Gabor representation of biological signals," *Proc. IEEE ICASSP-94*, vol.4, pp. 305-308, 1994.
6. L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, 1995.
7. R. Prokop and A. Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition," *CVGIP Graphical Models and Image Processing*, Vol. 54, No. 5, September, pp. 438-460, 1992.
8. L. Cohen and T. E. Posch, "Positive time-frequency distribution functions," *IEEE Transactions on ASSP*, vol. 33, no. 1, pp. 31-37, 1985.
9. P. Loughlin, J. Pitton and L. Atlas, "Construction of positive time-frequency distributions," *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2697-2705, 1994.
10. P. Loughlin, M. Redfern and B. Tacer, "Time-varying characteristics of postural sway are different for patients and healthy controls," *IEEE Proc. ICASSP'96.*, vol. 3, pp. 1439-1442, 1996.
11. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons,1973.
12. W. Dillon and M. Goldstein, *Multivariate Analysis*, John Wiley & Sons, 1984.
13. B. Tacer and P. Loughlin, "Time-scale energy density functions," *IEEE Trans. on Signal Processing*, Vol. 44, no. 5, pp.1310-1314, 1996.
14. B. Tacer and P. Loughlin, "What are the joint time-frequency moments of a signal?," *IEEE Proc. ICASSP'96*, vol. 3, pp. 1427-1430, 1996.

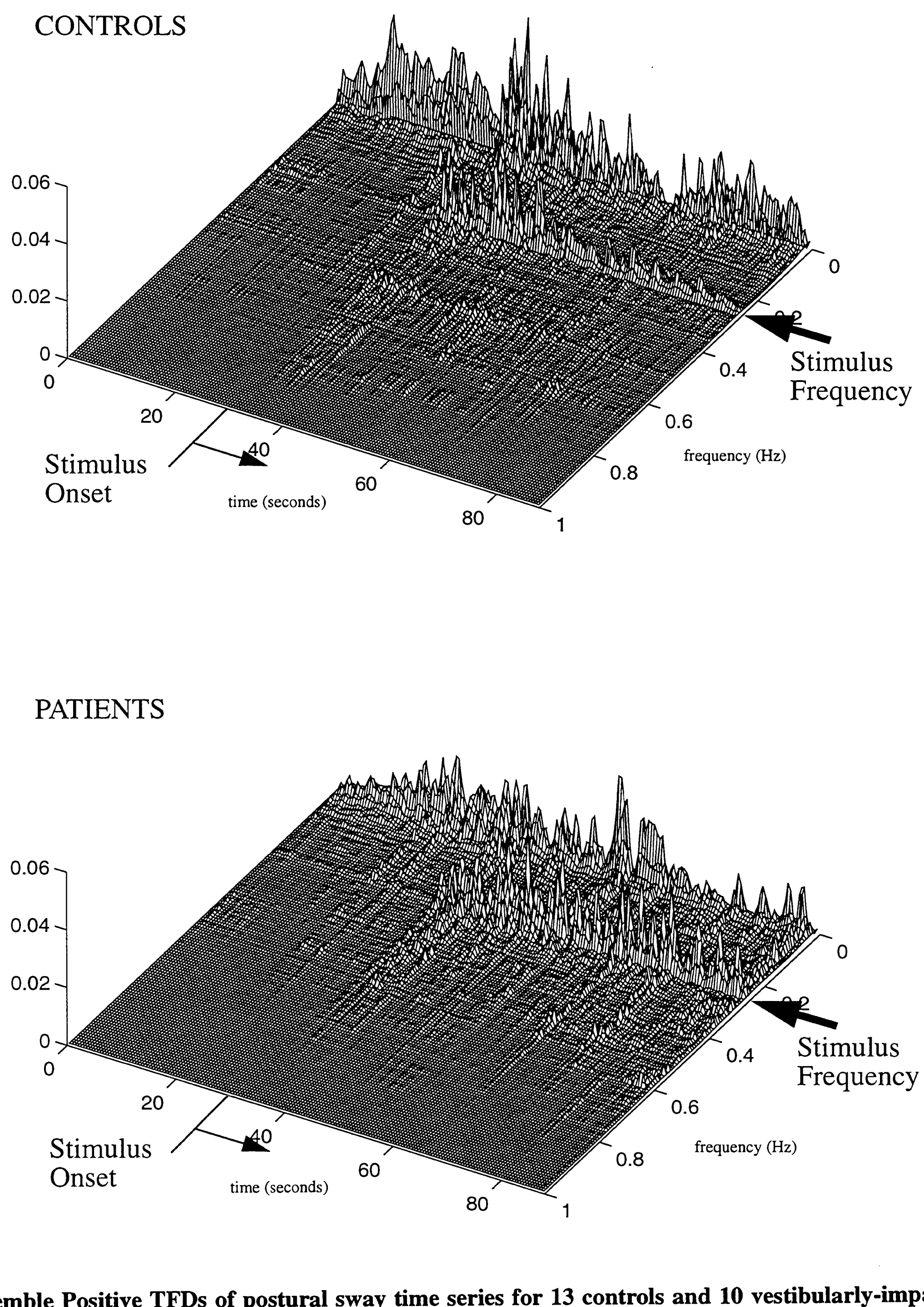


**Figure 1: Ensemble Positive TFDs of postural sway time series for 13 controls and 10 vestibularly-impaired subjects. Sinusoidal visual perturbation of 0.25 Hz began at 30 secs and continued for 60 secs. Note the decrease in energy at 0.25 Hz over time in controls, but not in the vestibularly impaired.**

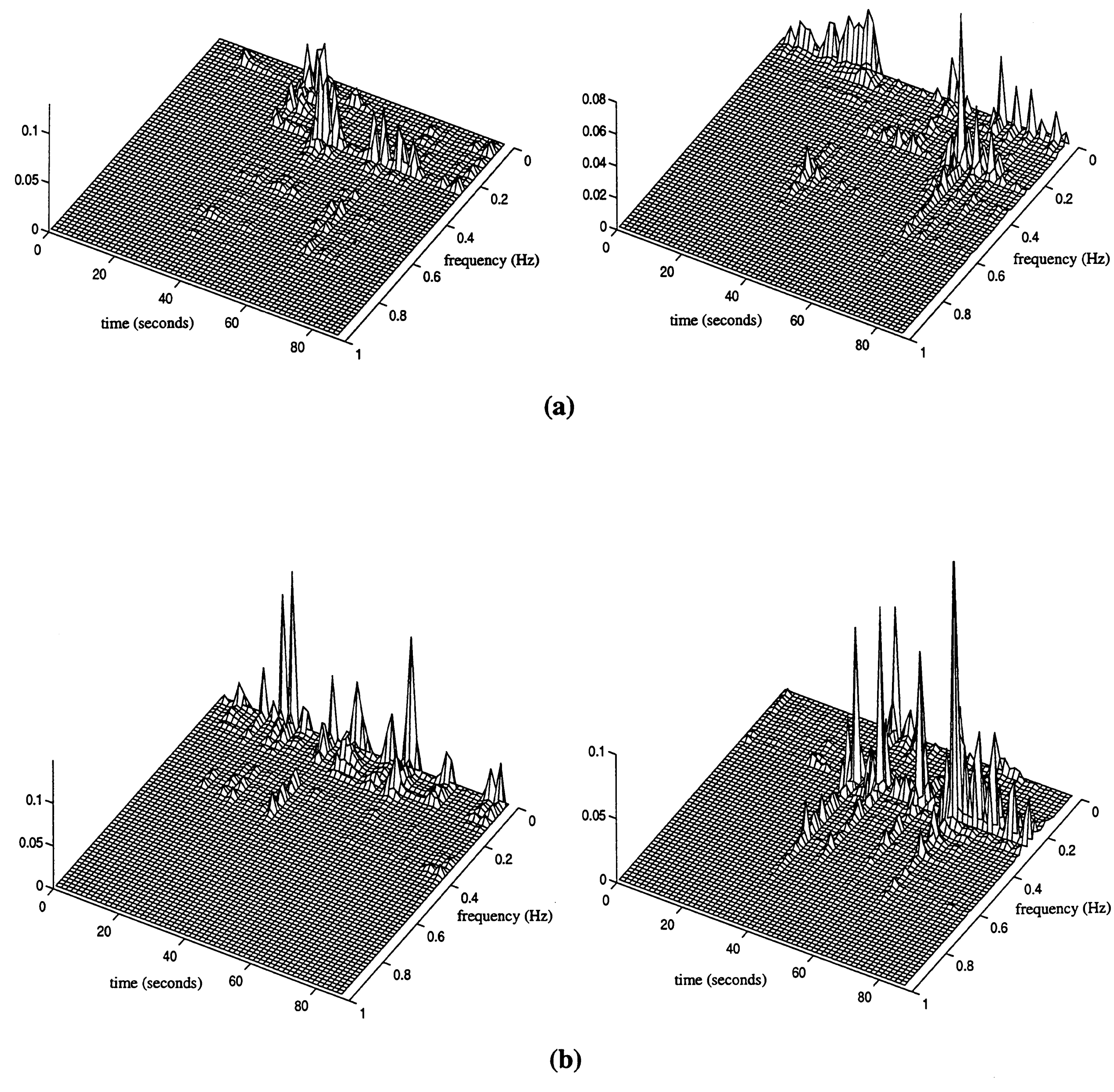


**Figure 2: The positive TFDs of two representative samples of (a) controls (b) patients.**