A SIGNAL CLASSIFICATION APPROACH USING TIME-WIDTH VS FREQUENCY BAND SUB-ENERGY DISTRIBUTIONS

Karthikeyan Umapathy

Dept. of Electrical and Computer Engg.,
The University of Western Ontario,
London, ON N6A 5B8, Canada
Email: kumapath@uwo.ca

Sridhar Krishnan

Dept. of Electrical and Computer Engg.,
Ryerson University,
Toronto, ON M5B 2K3, Canada
Email: krishnan@ee.ryerson.ca

ABSTRACT

Time-frequency (TF) signal decompositions provide us with ample information and extreme flexibility for signal analysis. By applying suitable processing on the TF decomposition parameters, even subtle signal characteristics can be revealed. In many real world applications, identification of these subtle differences make a significant impact in signal analysis. Particularly in classification applications using TF approaches, there may be situations where a localized high discriminative signal structure is diluted due to the presence of other overlapping signal structures. To address this problem we propose a novel approach to construct multiple time-width vs frequency band mappings based on the energy decomposition pattern of the signal. These mapping are then analyzed to locate the highly discriminative features for classification. Initial results with two real world biomedical signal databases (1) Vibroarthrographic (VAG) signals and (2) Pathological speech signals, indicate high potential for the proposed technique.

1. INTRODUCTION

Time-frequency (TF) transformations have significantly contributed towards complex signal analysis and automatic classification. In classification applications using TF approach, it is often a small area or pockets of areas in the TF plane that actually exhibit the difference between the classes of signals. Within these small areas, there may be overlapping multiple signal components with varying discriminative characteristics. The overall discriminative power of the area is normally decided by the high energy signal components which dilute the discriminative characteristics of less energy signal components. It may so happen that a high discriminative but less energy component is masked by a less discriminative but high energy component. Typical biomedical signals contain a mixture of coherent and non-coherent signal structures with varying localized overlaps. Using some criteria, if we can separate these localized overlapping structures, it may lead to a better understanding of the signal thereby to extract high discriminative features for classification applications.

In general, all real world signals contain both coherent and non-coherent structures. Coherent structure have definite TF localization unlike the non-coherent structures. Any iterative decomposition algorithm such as matching pursuits with TF dictionaries model the coherent structures during the initial iterations as they correlate well with the dictionary elements. The non-coherent structures on the other hand are broken into finer and finer structures till the information is diluted across the whole dictionary [1]. The contribution of coherent and non-coherent structures in a signal decide the energy capture pattern of the decomposition algorithms.

The previous work [2] of the authors, introduced a novel time-width vs frequency band mapping (constructed from the decomposition parameters) to identify the high discriminative TF tilings between different classes of signal using Local Discriminant Bases (LDB) algorithm. The proposed work uses a similar mapping, however splitting it into multiple mappings for identifying better discriminatory features.

The paper is organized as follows: Section II covers methodology consisting of adaptive time-frequency transform, multiple TFD slices, multiple s_n vs f_n mappings, databases, feature extraction and pattern classification. Results and discussion are given in Section III and conclusions in Section IV.

Thanks to NSERC for funding this research work.

2. METHODOLOGY

2.1. Adaptive Time-frequency Transform (ATFT)

The signal decomposition technique used in this work is based on the matching pursuit (MP) [1] algorithm. MP is a general framework for signal decomposition. The nature of the decomposition varies according to the dictionary of basis functions used. When a dictionary of TF functions is used, MP yields an adaptive time-frequency transformation [1]. In MP any signal $x(t)$ is decomposed into a linear combination of K TF functions $g(t)$ selected from a redundant dictionary of TF functions as given by:

$$x(t) = \sum_{n=0}^{K-1} \frac{a_n}{\sqrt{s_n}}\, g\left(\frac{t-p_n}{s_n}\right) \exp\left\{j(2\pi f_n t + \phi_n)\right\}, \quad (1)$$

where a_n is the expansion coefficient, the scale factor s_n also called as octave or time-width parameter is used to control the width of the window function, and the parameter p_n controls the temporal placement. The parameters f_n and ϕ_n are the frequency and phase of the exponential function respectively. The signal $x(t)$ is projected over a redundant dictionary of TF functions with all possible combinations of scaling, translations and modulations. The dictionary of TF functions can either suitably be modified or selected based on the application in hand. In our technique, we are using the Gabor dictionary (Gaussian functions) which has the best TF localization properties. At each iteration, the best correlated TF functions to the local signal structures are selected from

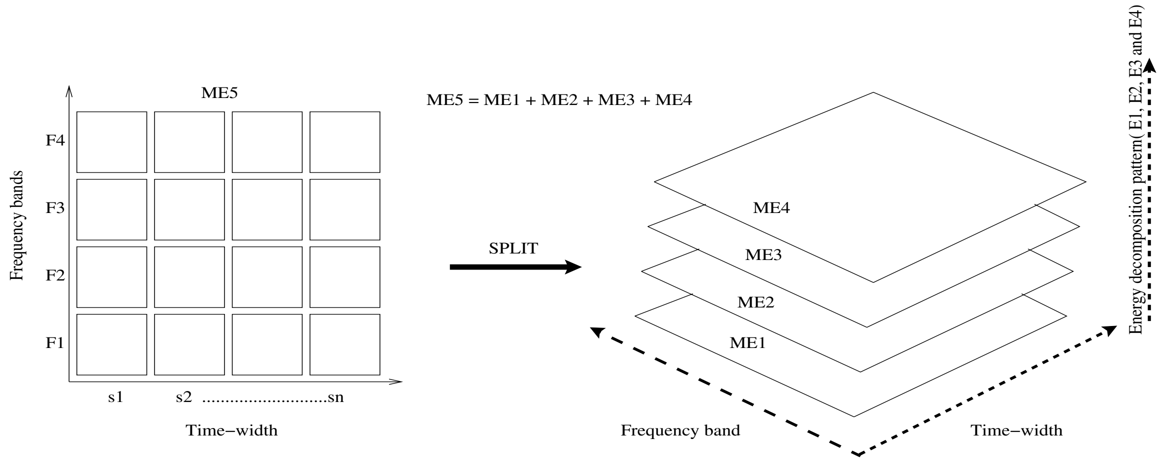

Fig. 1. Time-width vs frequency band mapping

the dictionary. The remaining signal called the residue is further decomposed in the same way at each iteration subdividing them into TF functions.

2.2. Multiple TFD slices

As explained in Section 1, in the initial iterations, the ATFT algorithm captures the coherent signal structures which have correlated TF dictionary elements and then as the number of iterations grows, it tries model the non-coherent structures by breaking them finer and finer till the information is diluted across the whole dictionary. The energy capture pattern can be extracted from the normalized decomposition parameter a_n. In order to explain how this energy capture pattern can be utilized to extract overlapping signal structures, let us take an example of a synthetic signal $y(t)$ which is composed of a sinusoid, two chirps and random noise. The signal $y(t)$ is given by:

$$y(t) = w_1 s(t) + w_2 c_1(t) + w_3 c_2(t) + w_4 r(t) \quad (2)$$

where $s(t)$ represent the sinusoid at approximately Fs/4, $c_1(t)$ is a linear chirp with increasing frequency cutting the sinusoid, $c_2(t)$ is another linear chirp with decreasing frequency again cutting both the sinusoid and $c_1(t)$. $r(t)$ represents the random noise. The weight factors $w_{1,2,3,4}$ are (1, .1, .01, .001) respectively. We performed the ATFT decomposition (1000 iterations) of $y(t)$ using a Gabor dictionary. Figures 3(a) and 3(b) show $y(t)$ in time domain and TF domain (spectrogram is used inorder to show all the three components at the same time). Here we deliberately introduced energy differences between the components so as to demonstrate the significance of energy capture pattern. Most of the times, the first few iterations capture significant amount of signal energy (coherent structures). Thereafter with the increase in the number of iterations we move from modeling coherent structures to non-coherent structures. The energy capture pattern of the ATFT decomposition for $y(t)$ is shown in Fig. 2 (the first 50 iterations). The curve represents the normalized energy captured per iteration. We can see the energy captured per iteration drops as we move along the iterations. In this work as an example we split the energy capture pattern into 4 parts i.e. (E1) the number of iterations at which the energy captured per iteration drops to 10% of its initial value (initial value= 1), (E2) the number of iterations between 10% of initial value and 1% of initial value, (E3) the number of iterations between 1% of initial value and 0.1% of initial value, and (E4) the number of iterations between 0.1% of initial value to the end of decomposition.

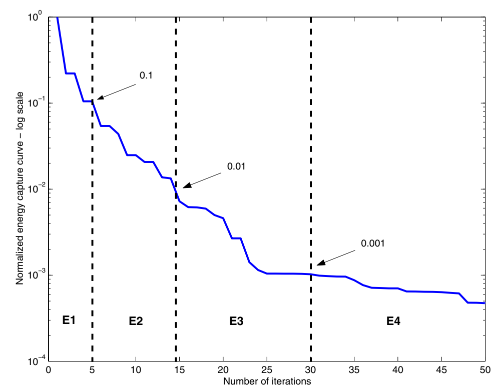

Fig. 2. Energy capture pattern of the sample signal $y(t)$.

Following the energy capture pattern we accumulate the TF functions into the above explained four parts (E1-4). For this example, we had 5 TF functions for E1, 9 TF functions for E2, 16 TF functions for E3 and 970 TF functions for E4. The number of TF functions will give an idea that almost 99% of the signal energy needs only 30 (1 to E3) TF fucntions, whereas the remaining 1% signal energy (mostly noise like strutcures) needs 970 TF functions or more. Using these 4 sets of TF fucntions we construct 4 different TFDs. i.e. splitting the original TFD of $y(t)$ into 4 TFDs based on the energy capture pattern. The corresponding 4 TFDs are shown in Figs. 3(c), 3(d), 3(e) and 3(f). If we closely look at the TFDs, we can see the TFD in Fig. 3(c) showing the sinusoid $s(t)$ alone, the TFD in Fig. 3(d) shows the disappearing sinusoid, the TFD in Fig. 3(e) shows the evolving chirp $c_1(t)$ signal from the sinusoid background and the TFD in Fig. 3(f) showing a stronger but noisy chirp $c_1(t)$, a faint evolving chirp $c_2(t)$ and the random noise. It is obvious to see that TFDs 3(c) to 3(f) are better individual representations of the signal components than the combined TFD 3(b). In this example if it so happens that one of the components that was masked by the overlapping strong component is

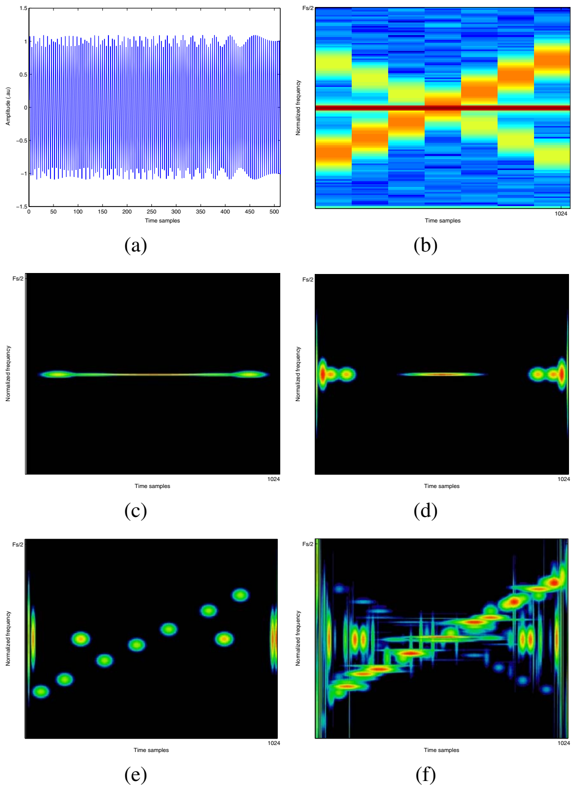

Fig. 3. (a) sample signal $y(t)$, (b) TFD of the sample signal, (c) TFD of sample signal with TF functions of E1, (d) TFD of sample signal with TF function of E2 (e) TFD of sample signal with TF functions of E3 and (f) TFD of the residue signal

the discriminator that we are looking for, then the proposed technique of generating multiple TF mapping using the energy capture pattern will be of immense help. Here it should be noted that the energy split shown in this example is not the best to show all the components individually and separately. This is just to give an idea about the possiblity of using the energy capture patttern for removing overlapping structures in complex situations. Also this approach may not work in all situations unless there are hidden signal structures either with (a) different energy contribution or (b) different contributions from coherent and non-coherent structures or both (a) and (b). Extending this same concept of multiple TF mappings, we now apply it on a novel time-width vs frequency band mapping as will be explained in the next Section 2.3.

2.3. Multiple s_n vs f_n mappings

In order to effectively analyze for classification applications, the ATFT signal decomposition parameters need to be rearranged in a pseudo dictionary format. There are five parameters as explained in Section 2.1 viz. a_n, s_n, f_n, p_n and ϕ_n that represent the index of each of the dictionary element. After a signal is decomposed into TF functions, we group the TF functions with the time-width parameter s_n in X axis and the the f_n parameter in the Y axis. In order to reduce the computational complexity instead of using all the possible values of the f_n parameter we break the frequency range into M bands only. whereas s_n takes all the possible values ($2^{1..14}$) depending on the length of the signal. Each combination of s_n with one of the M frequency bands form a cell which contains the cumulative normalized energy of all the TF functions falling in that particular combination of s_n and frequency band. The left side of the Fig. 1 shows a sample time-width vs frequency band mapping. In the proposed work we used 4 frequency bands, which means we transform the decomposition parameters of a signal into 14 time-widths (s_n) vs 4 frequency band mapping.

From this time-width vs frequency band mapping we can readily obtain the energy distribution of the signal in terms of the time-width and frequency band (center frequency) decomposition parameters. Depending upon the application one can choose say K number of cells that covers an area corresponding to certain amount of signal energy. This area will provide the s_n and f_n ranges which are significant for that particular application. This area can be arrived by averaging the time-width vs frequency band of N sample signals. For classification applications this can be done using LDB as demonstrated in authors previous work [2]. Considering the benefits of multiple TFD slices in signal analysis as explained in Section 2.2, instead of using one time-width vs frequency band mapping that covers all the signal energy, we slice it into L time-width vs frequency band mappings as shown in Fig. 1 ($L = 4$). This L sliced time-width vs frequency band mappings are expected to separate out the overlapping energy distribution of the TF functions based on the energy capture pattern and thereby enhance the discriminatory power of the cells. We performed classification on two biomedical signal databases to verify the effectiveness of the proposed technique of splitting the time-width vs frequency band mapping.

2.4. Databases

(1) Vibroarthographic (VAG) signals: These are the vibration signals emitted from the human knee joints during an active movement of the leg and can be used to detect the early joint degeneration or defects. Extensive work [3] has been done using time-frequency approach in analyzing these signals. Few important characteristics of the VAG signals which make them difficult to analyze are as follows: (i) Highly non-stationary, (ii) Varying frequency dynamics, and (iii) Multi-component signal. The database consists of 36 signals with 19 normal and 17 abnormal signals.
(2) Pathological speech signals: These are speech signals recorded from the pathological and normal talkers in a sound-attenuating booth at the Massachusetts Eye and Ear Infirmary. All signals were sampled at 25 kHz. The signals were the first sentence of the rainbow passage, 'when the sunlight strikes rain drops in the air, they act like a prism and form a rainbow'. More details on the classification of this database can be found in author's previous work [4]. The database used in this study consists of 30 signals with 15 normal and 15 pathological signals.

2.5. Feature Extraction and Pattern Classification

Signals from both the databases were decomposed using the ATFT algorithm (5000 iterations) as explained in Section 2.1. For each signal, 4 time-width vs frequency band mappings were created using the decomposition parameters. The energy split used for generating these 4 mappings were same as the one used in the example of synthetic signals (E1-4). In these 4 mappings, each row of the mapping represents the signal energy distribution over all time-widths for a particular band of frequencies. Let us name the mappings as ME1, ME2, ME3 and ME4 and the frequency bands as F1, F2, F3 and F4 as shown in Fig. 1. Now for each combination of MEx and Fx we extract $P \times 14$ energy values from the cells as feature matrix, where P is the number of signals in the database. From the 16 combinations of MEx and Fx, only non-zero feature matrices are used for classification. In order to compare the results with the original non-split time-width vs frequency mappings (let it be ME5), another set of 4 feature matrices were generated using the same procedure. When tested with the Ho-Kashyap [5] algorithm, most of these 20 combinations (MEx and Fx) for both the databases favored non-linear separability to achieve maximum classification accuracies. However, as the main focus of the proposed technique is to demonstrate the relative improvement in discrimination between the split and non-split time-width vs frequency mappings, we restrict our analysis to a linear classifier. The extracted features were fed to a Linear Discriminant Analysis (LDA) based classifier using SPSS [6]. The classification accuracy was validated using the leave-one-out method which is known to provide a least bias estimate.

3. RESULTS AND DISCUSSION

A two stage classification was performed for the VAG database. In the first stage, we performed a two group classification classifying the VAG signals into normal and abnormal. Table 1 shows the highest classification accuracy achieved out of the 20 combinations of MEx and Fx. We observed an overall classification accuracy of 88.9% using leave-one-out (Cross. V) based LDA for the combination of ME4 and F3. This is higher than the classification accuracies reported by existing works for this database. There is no difference in the classification accuracy comparing it with the combination of non-split ME5F3. This is because F3 is non zero only for ME4 which means, there is no overlap in F3. So eventually ME4F3 and ME5F3 were the same. The results also gave a clue that the discriminatory information between normal and abnormal lies in F3.

Table 1. Table showing 2 group classification accuracy achieved for the VAG database. Cross.V = Leave-one-out method LDA, % = Percentage of classification

Method	Groups	Normal	Abnormal	Total
Cross.V	Normal	15	4	19
	Abnormal	0	17	17
%	Normal	**78.9**	21.1	100
	Abnormal	0	**100**	100

We now performed the second stage of classification on the 17 abnormal VAG signals. The abnormal VAG signals in the database are from different kinds of knee pathologies. Chondromalcia patella (CMP) [3] is one of the pathologies which has four categories (I, II, III and IV) of grading based on the severity. It is a difficult task to classify them by their gradings using the VAG signals. Out of the 17 abnormal signals, 10 were CMP signals. We performed a three groups classification on this 10 signals viz. grade(I and II), grade (II and III) and grade (III and IV). We observed a perfect classification of 100% using leave-one-out based LDA for the combination of ME2 and F1. None of the other combinations including the non-split ME5F1 could achieve 100% classification. This result explains the fact that splitting the time-width vs frequency band mappings does enhance the discriminatory power and also indicates the discriminatory features for CMP lies in the ME2 and F1 mapping.

Similarly we performed a 2 group classification (normal and pathological) for the pathological speech database. Table 2 shows the highest classification accuracy achieved out of the 20 combinations of MEx and Fx. An overall classification accuracy of 93.3% was achieved using the leave-one-out based LDA. The reported classification is for the combination of ME1F1 and non-split ME5F1. In which case we observe the classification accuracy to remain same with or without splitting the time-width vs frequency mapping. However the results give a clue that the discriminatory information lies in ME1 and F1.

Table 2. Table showing the 2 group classification accuracy achieved for the pathological speech database. Cross.V - Leave-one-out method LDA, % = Percentage of classification

Method	Groups	Normal	Pathological	Total
Cross.V	Normal	13	2	15
	Pathological	0	15	15
%	Normal	**86.7**	13.3	100
	Pathological	0	**100**	100

4. CONCLUSIONS

Enhancing the discriminatory power of the TF representations using a TFD splitting approach was proposed. The technique was explained using a synthetic signal and two real world signal databases. Using the technique on the VAG database showed a significant improvement in the sub classification of abnormal signals. Although the results are inconclusive for the real world databases, this approach may better suit for identifying finer discriminatory features inside global classifications. Adaptively choosing the energy split might improve the significance of the proposed technique. Future work involves in arriving at a suitable energy split ratio based on the nature of the signal, increase the number of frequency bands and extract visual feature treating the time-width vs frequency mapping as an image.

5. REFERENCES

[1] S. G. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[2] K. Umapathy, S. Krishnan, and A. Das, "Sub-dictionary selection using local discriminant bases algorithm for signal classification," in *Proceeding of IEEE Canadian Conference on Electrical and Computer Engineering*, Niagara falls, Canada, May 2004, pp. 2001–2004.

[3] S. Krishnan, "Adaptive signal processing techniques for analysis of knee joint vibroarthrographic signals," in *Ph.D dissertation*, University of Calgary, June 1999.

[4] K. Umapathy, S. Krishnan, V. Parsa, and D. Jamieson, "Time-frequency modeling and classification of pathological voices," in *Proceedings of IEEE Engineering in Medicine and Biology Society (EMBS) 2002 Conference*, Houston, Texas, USA, Oct 2002, pp. 116–117.

[5] M. H. Hassoun and J. Song, "Adaptive Ho-Kashyap rules for perceptron training," *IEEE Trans. on Neural Networks*, vol. 3, no. 1, pp. 51–61, 1992.

[6] SPSS Inc., "SPSS Advanced statistics user's guide," in *User manual, SPSS Inc., Chicago, IL*, 1990.