COMPARATIVE STUDY OF THREE TIME-FREQUENCY REPRESENTATIONS WITH APPLICATIONS TO A NOVEL CORRELATION METHOD

Ervin Sejdić and Jin Jiang

Department of Electrical and Computer Engineering
The University of Western Ontario
London, Ontario, N6A 5B9, Canada
E-mail: {esejdic, jjiang}@uwo.ca

ABSTRACT

In this paper, the effect of three time-frequency representations on a novel correlation algorithm is studied. By representing a signal in the time-frequency domain, a redundant representation of the signal is obtained. The algorithm presented herein relies on such redundancies to extrapolate some significant features of the signal. The developed scheme has been applied to heart sound analysis using real recordings from patients, where the opening snap (OS) is distinguished from the third heart sound (S3). The results for the three time-frequency transforms are compared and very encouraging results have been obtained with S-transform.

1. INTRODUCTION

Automatic recognition and pattern matching for signals with particular characteristics buried in other signals or noise can be a difficult task. If a correlation-type scheme is used, the corrupting signals/noise will affect the accuracy of the pattern matching and may subsequently lead to errors in the decision [1]. This is because corrupting signals may also bear some resemblance to the template being matched. This is particularly true if the pattern of interest is a transient signal embedded within a non-stationary environment. A particular application considered in this paper is the analysis of heart sounds. Even though the correlation is a very powerful tool, which has been used extensively in many applications [1], the potential limitations with traditional time-domain correlation-based pattern recognition methods lie in the fact that they did not fully utilize the frequency characteristics of the template and the signal being analyzed. It is important to point out that, if the template has bandlimited characteristics, significant improvement in the performance of the pattern recognition scheme can be readily made by a relatively simple pre-processing of the signal and the template in time-frequency domain. Such pre-processing can separate the intertwined time and frequency domain features of the signals effectively and allow important features to be exposed in the time-frequency domain.

A novel scheme for improving the performance of pattern recognition methods of bandlimited non-stationary signals has been developed in this paper, based on time-frequency analysis tools. The pre-processing is carried out by converting a one-dimensional (1D) time domain signal into a two-dimensional (2D) time-frequency representation. By doing so, the true time-frequency compositions of the signal can be revealed clearly. Hence, it allows the pattern matching to be conducted only in selected regions in the time-frequency domain. For this reason, this newly developed technique is referred to as *Selective Regional Correlation*. To emphasize the novelty of this new method and its performance with different time-frequency transforms, the S-Transform [2], Short-Time Fourier Transform (STFT) [3], and Wavelet analysis [4] have been employed in the analysis and numerical studies.

This paper is organized as follows: The detailed time-frequency domain pre-processing and the theoretical development of the selective regional correlation are presented in Section 2. Section 3 illustrates the performance of the proposed scheme through an application in heart sounds analysis. With this example, effects of the three time-frequency representations on the resolution of the pattern classifier are examined. The conclusions are drawn in Section 4 followed by a list of references.

Authors would like to thank Natural Sciences and Engineering Research Council of Canada for financial support for this project.

2. SELECTIVE REGIONAL CORRELATION

2.1. Problem Statement

Based on the above introduction, the problem investigated in this paper can be stated as follows:

For a given bandlimited template $p(t)$ and the signal $s(t)$, design a pre-processing scheme so that when $s(t)$ contains the known template $p(t)$, the pre-processing increases the value of correlation coefficient, otherwise decreases the value of correlation coefficient.

2.2. A Time-Frequency Decomposition of the Signal and the Template

The essence of the proposed scheme is to represent a 1D time domain signal in 2D time-frequency domain in order to reveal its true characteristics for more accurate pattern matching. In principle, any time-frequency representation method can be utilized for such a purpose; however, due to the fact that signal decompositions are involved, the bilinear class of time-frequency distributions, such as Wigner distribution or Cohen's class [5], may be more involved for their cross products terms.

An idea behind the linear time-frequency transform is to correlate or convolve a function with a waveform (sometimes called time-frequency atoms [4]) which is well concentrated in time and frequency.

Definition 2.1 *For a waveform $\phi_{\tau,\gamma} \in \boldsymbol{L}^2(R)$, the corresponding time-frequency transform of function $f \in \boldsymbol{L}^2(R)$ is defined as [4]:*

$$Tf(\tau,\gamma) = \int_{-\infty}^{\infty} f(t)\phi_{\tau,\gamma}(t)dt = \langle f, \phi_{\tau,\gamma}\rangle \quad (1)$$

Similarly, the inverse time-frequency transform can be represented as:

$$f(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Tf(\tau,\gamma)K(\tau,\gamma)d\tau d\gamma \quad (2)$$

where $K(\tau,\gamma)$ is a kernel used for the inversion.

A short-time Fourier transform can be constructed with a finite length window g which is translated by τ and modulated by ξ [3] [4]:

$$\phi_\gamma(t) = g_{\xi,\tau}(t) = e^{j\xi t} g(t-\tau) \tag{3}$$

A wavelet transform is constructed from dilation by s and translation by τ of a mother wavelet ψ [4]:

$$\phi_\gamma(t) = \psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi(\frac{t-\tau}{s}) \tag{4}$$

An S-Transform is constructed using a Gaussian window which is translated by τ, dilated by f and modulated by ξ [2] [9] [10]:

$$\phi_\gamma(t) = g_{\xi,\tau,f}(t) = e^{j\xi t} g(\frac{t-\tau}{f}) \tag{5}$$

Clearly, the S-Transform can be viewed from two different perspectives. It can be seen as a short-time Fourier transform with a variable window length [2], or as a special type of continuous wavelet transform (CWT) with a Gaussian mother wavelet multiplied by a phase factor [2]:

$$STf(\tau,\gamma) = e^{j2\pi f\tau} CWT_f(\tau,\gamma) \tag{6}$$

2.3. Selective Regional Correlation

To present the concept of selective regional correlation, let's assume that a bandlimited template and the signal being analyzed are represented by $p(t)$ and $s(t)$, respectively.

Lemma 2.1 *If the template is represented by $p(t)$, its time-frequency transform can be represented as:*

$$Tp(\tau,\gamma) = \int_{-\infty}^{\infty} p(t)\phi_{\tau,\gamma}(t)dt \tag{7}$$

and

$$Tp(\tau,\gamma) \equiv 0 \quad \forall\gamma \notin \{\gamma : \gamma_1 \le \gamma \le \gamma_2\} \tag{8}$$

where γ_1 and γ_2 are the lower and upper limits of frequency band of the template.

Lemma 2.2 *If a finite duration signal being processed and its time-frequency transform are represented by $s(t) \subseteq [t_1, t_2]$ and $Ts(\tau,\gamma)$, respectively, the following signal decomposition is in order:*

$$Ts(\tau,\gamma) = Ts_1(\tau,\gamma) \cup Ts_2(\tau,\gamma) \tag{9}$$

In other words, $Ts_1(\tau,\gamma)$ represents the portion of the signal in the time and frequency range of the template, and $Ts_2(\tau,\gamma)$ is its complement, representing the remaining signal elements.

Different 2D windows can be used to effectively extract $Ts_1(\tau,\gamma)$ from $Ts(\tau,\gamma)$. However, care should be taken when extracting $Ts_1(\tau,\gamma)$ from $Ts(\tau,\gamma)$, since window functions could introduce additional transients when returning to the time-domain.

Lemma 2.3 *If a 2D window is represented in the time-frequency domain as: $W(\tau,\gamma) \quad \exists\,\forall \quad \tau \in [\tau_1,\tau_2] \quad \gamma \in [\gamma_1,\gamma_2]$, then*

$$Ts_1(\tau,\gamma) = Ts(\tau,\gamma) \cdot W(\tau,\gamma) \quad \tau \in [\tau_1,\tau_2]\, \gamma \in [\gamma_1,\gamma_2] \tag{10}$$

In this paper, for the interest of reducing the edge effect, elliptic type of window is used. A Gaussian elliptic window is a window whose boundary is defined by an ellipse [6], while weights of the points inside the ellipse are assigned by a Gaussian distribution. The definition of a Gaussian elliptic window is

$$W_g(\tau,\gamma) = \left\{ \begin{array}{ccc} \frac{1}{2\pi\sigma_\tau\sigma_\gamma} e^{-\frac{\tau^2}{2\sigma_\tau^2} - \frac{\gamma^2}{2\sigma_\gamma^2}} & \exists\,\forall & \tau,\,\gamma \\ 0 & otherwise & \end{array} \right\} \tag{11}$$

where the Gaussian elliptic supports the region $R = \{(\tau,\gamma) : \tau \in [\tau_1,\tau_2], \quad \gamma \in [\gamma_1,\gamma_2], (\tau/a)^2 + (\gamma/b)^2 = 1\}$, and constants a and b represent half of the major and minor axes, respectively. The variances in time and frequency are represented by σ_τ^2 and σ_γ^2 respectively. They control the shape of this 2D window.

Corollary 2.1 *Based on Lemma 2.2 and Lemma 2.3, the corresponding time-domain signal has the following properties:*

$$s(t) = s_1(t) + s_2(t) \tag{12}$$

$$\int_{-\infty}^{\infty} s_1(t)s_2(t)dt = 0 \tag{13}$$

where $s_1(t)$ and $s_2(t)$ are the inverse time-frequency transforms of $Ts_1(\tau,\gamma)$ and $Ts_2(\tau,\gamma)$, respectively.

Theorem 2.1 *If the pattern similar to that of the template $p(t)$ is present in the signal $s(t)$, then the following is true:*

$$\max[|corr(s_1(t),p(t))|] > \max[|corr(s(t),p(t))|] \tag{14}$$

$$\max[|corr(s_2(t),p(t))|] < \max[|corr(s(t),p(t))|] \tag{15}$$

where $\max[|corr(x(t),y(t))|]$ is defined as:

$$\max[|corr(x(t),y(t))|] = \max\left[\left|\frac{\int_{-\infty}^{\infty} x(t)y(t+\tau)dt}{\sqrt{\int_{-\infty}^{\infty} x(t)^2dt}\sqrt{\int_{-\infty}^{\infty} y(t)^2dt}}\right|\right] \tag{16}$$

Proof: Eqn. (12) and Eqn. (13) represent the signal $s(t)$ and its decomposition. Assuming that $s_1(t)$ contains the pattern $p(t)$, and $s_2(t)$ lies in the frequency and the time bands outside those of the pattern $p(t)$, this means that

$$\left|\int_{-\infty}^{\infty} s_1(t)p(t)dt\right| \gg 0 \tag{17}$$

$$\left|\int_{-\infty}^{\infty} s_2(t)p(t)dt\right| \approx 0 \tag{18}$$

Using Eqn. (18) and Eqn.(13), the left hand sides of Eqn. (14) and Eqn. (15) are equal to

$$\max\left[\left|\frac{\int_{-\infty}^{\infty} s_1(t)p(t+\tau)dt}{\sqrt{\int_{-\infty}^{\infty} s_1(t)^2dt}\sqrt{\int_{-\infty}^{\infty} p(t)^2dt}}\right|\right] \tag{19}$$

$$\max[|corr(s_2(t),p(t))|] \approx 0 \tag{20}$$

and the right hand side of Eqn. (14) and Eqn. (15) is equal to

$$\max\left[\left|\frac{\int_{-\infty}^{\infty} s_1(t)p(t+\tau)dt}{\sqrt{\int_{-\infty}^{\infty} \{s_1(t)^2 + s_2(t)^2\}dt}\sqrt{\int_{-\infty}^{\infty} p(t)^2dt}}\right|\right] \tag{21}$$

Since the denominator of Eqn. (21) is greater than the denominator of Eqn. (19), and $s_1(t)$ and $s_2(t)$ are complements to each other, it follows that

$$\max[|corr(s_1(t),p(t))|] > \max[|corr(s(t),p(t))|] \tag{22}$$

$$\max[|corr(s_2(t),p(t))|] < \max[|corr(s(t),p(t))|] \tag{23}$$

So far, only a binary case is dealt with. The concept of the selective regional correlation is also valid for M-ary case, but certain precautions must be taken when selecting the templates. The major point is that the templates must be mutually exclusive.

Theorem 2.2 *Let's assume that $p_1(t)...p_m(t)$ are the templates with the time-frequency representation $Tp_1(\tau,\gamma)...Tp_m(\tau,\gamma)$ respectively. The selective regional correlation can then be applicable in two ways:*

1. *Case I*
   If the time-frequency representations of the templates, $Tp_1(\tau,\gamma), Tp_2(\tau,\gamma)...Tp_m(\tau,\gamma)$ are disjoint sets, that is

   $$Tp_k(\tau,\gamma) \cap Tp_l(\tau,\gamma) = \emptyset \quad (24)$$
   $$\forall\, k,l \in \{k,l\,|k \neq l \ \wedge\ 1 \leq k,l \leq m\}$$

   then a template $p_i(t)$ is simply found by multiplying a function $\varsigma_i(t)$ (where $i = 1,2...m$ in each case) that contains the desired template by a 2D window and inverting back to the time domain:

   $$p_i(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T\varsigma_i(\tau,\gamma)W_i(\tau,\gamma)K(\tau,\gamma)d\tau d\gamma \quad (25)$$

   where $T\varsigma_i(\tau,\gamma)$ is the time-frequency representation of $\varsigma_i(t)$ and $W_i(\tau,\gamma)$ is 2D window.
2. *Case II*
   When the time-frequency representations of the templates, $Tp_1(\tau,\gamma), Tp_2(\tau,\gamma)...Tp_m(\tau,\gamma)$ are not disjoint, that is

   $$Tp_k(\tau,\gamma) \cap Tp_l(\tau,\gamma) \neq \emptyset \quad (26)$$
   $$\forall\, k,l \in \{k,l\,|k \neq l \ \wedge\ 1 \leq k,l \leq m\}$$

   it is necessary to introduce a mutually exclusive template in order to reduce the highest correlation coefficient when the template is absent. This exclusivity is introduced in the time-frequency domain as follows:

   $$p_k(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_k(\tau,\gamma)K(\tau,\gamma)d\tau d\gamma \quad (27)$$
   $$p_l(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_l(\tau,\gamma)K(\tau,\gamma)d\tau d\gamma \quad (28)$$

   where

   $$P_k(\tau,\gamma) = Tp_k(\tau,\gamma)w_k(\tau,\gamma) \setminus Tp_l(\tau,\gamma)w_k(\tau,\gamma) \quad (29)$$
   $$P_l(\tau,\gamma) = Tp_l(\tau,\gamma)w_k(\tau,\gamma) \setminus Tp_k(\tau,\gamma)w_l(\tau,\gamma) \quad (30)$$
   $$\forall\, k,l \in \{k,l\,|k \neq l \ \wedge\ 1 \leq k,l \leq m\}$$

In both cases, the selective regional correlation is defined as:

$$SRC(\zeta) = \frac{\int_{-\infty}^{\infty} \tilde{s}(t)p_i(t+\zeta)dt}{\sqrt{\int_{-\infty}^{\infty}\tilde{s}(t)^2dt}\sqrt{\int_{-\infty}^{\infty}p_i(t)^2dt}} \quad (31)$$

where

$$\tilde{s}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Ts(\tau,\gamma)W(\tau,\gamma)K(\tau,\gamma)d\tau d\gamma \quad (32)$$

and where $Ts(\tau,\gamma)$ is the time-frequency transform of the signal $s(t)$, and $p_i(t)$ is ith template.

3. HEART SOUND CLASSIFICATION BY THE SELECTIVE REGIONAL CORRELATION

3.1. Heart Sounds

Heart problems associated with heart valves, known as mitral stenosis, are caused by rheumatic heart disease which leads to narrowing of mitral valve. Clinical experience has shown that heart auscultation can be an effective tool to diagnose the above condition, since it allows the detection of abnormal behavior of the heart before it can be detected with the other techniques such as ECG [7] [8]. Heart sounds are result of sudden closure of the heart valves during different phases of the cardiac contraction. They are non-stationary, non-deterministic signals that carry information about the anatomical and physiological state of the heart. Each heart beat consists of at least the first heart sound (S1) and the second heart sound (S2). Further, the opening snap (OS) and the third heart sound (S3) represent two different pathological states of the heart. Thus, they do not occur simultaneously. The OS occurs in a patient with potential mitral stenosis and the S3 often occurs in patients with impaired myocardial reserve [7]. However, the difficulty lies in the fact that there are significant similarities between OS and S3 as shown in top graphs of Fig. 1 and Fig. 2, respectively. It is generally difficult for physicians to distinguish them without specific training [7]. It will be shown in this section that the selective regional correlation based pattern recognition techniques can improve the detection of the presence or absence of these conditions significantly.

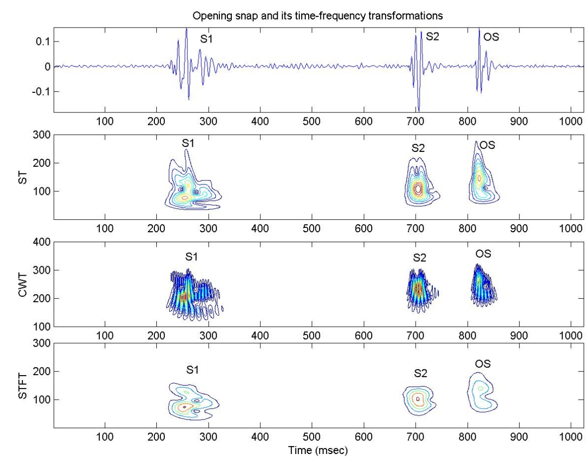

Fig. 1. Opening snap and its time-frequency representation

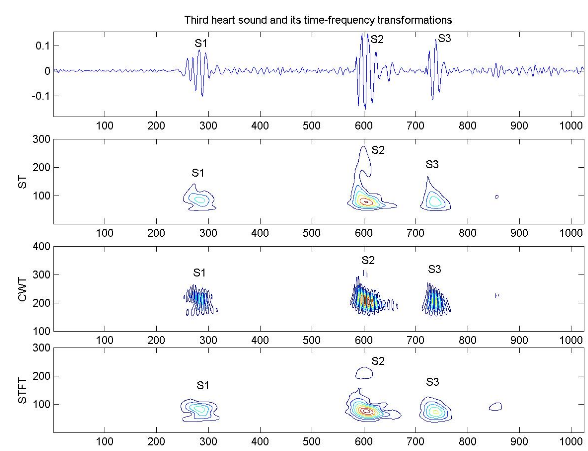

Fig. 2. Third heart sound and its time-frequency representation

Table 1. Comparison of highest correlation coefficient for selective regional correlation and general correlation

Method		MATCH	NO MATCH	RESOLUTION	RATIO
Selective Regional Correlation	S-Transform	0.5248	0.1759	0.3489	2.9835
	Short Time Fourier Transform	0.5972	0.2404	0.3568	2.4842
	Continuous Wavelet Transform	0.7292	0.5735	0.1557	1.2715
General Correlation		0.3885	0.3575	0.0310	1.087

3.2. Numerical Analysis

Phonocardiograph recordings of actual heart sounds sampled at 4000 Hz, are obtained from patients at the St. Joseph's Hospital in Toronto, Canada. In order to take advantage of the fast algorithms each recording is 1.024 seconds long [9]. The sampling rate is high enough since the maximum frequency content of the heart sounds is usually below 600 Hz [7]. These recordings have been carefully studied by the chief cardiologist to classify them as either containing the opening snap or the third heart sound. There are total of nineteen signals used in our analysis. Twelve of the signals contain the opening snap and seven contain the third heart sound. Two of the signals were set as the templates.

In the time-frequency analysis, three different transforms are used as outlined in the previous sections. A Gaussian window with a length of 260 points is used for STFT. A Gaussian mother wavelet is used for the CWT. In the calculations of the S-transform, an original Gaussian window was used [2]. The time-frequency representations of the OS and S3 in Fig. 1 and Fig. 2 depict that there exist minor differences between the two signals in the time domain. These differences can be clearly seen in the time-frequency domain. Very often, the S1 is used as the onset of the heart beat, and in order to detect the beginning of the S1 one would need to use the ECG signal which has been recorded simultaneously [7]. The simultaneous recording of ECG and the heart sounds can be easily performed with SimulScope III from Cardionics Inc. Webster, Texas, USA. If the onset of S1 is used as the time reference, at the same rate of heart beat, the OS will be usually ahead of S3 slightly [9]. Naturally, the selective regional correlation can be applied to assist physicians for correct diagnosis.

The goals of the numerical study are to show that the selective regional correlation improves the resolution of the classifier significantly, that is, it is sensitive to different patterns, and to show that it is a robust technique independent of the type of time-frequency transforms and slight variations in patterns. These objectives have been fulfilled by comparing the results of the selective regional correlation with general correlation. In order to do so, two states are used: MATCH or NO MATCH. MATCH represents the situation that the signal contains the patterns as specified. NO MATCH represents the situation where the signal does not contain the pattern of the template. The performance can be seen by calculating the differences and ratios between correlation coefficients in these two situations.

3.3. Performance of the Selective Regional Correlation

The sensitivity of the proposed scheme has been evaluated by comparing it with general correlation, and the results are shown in Table 1. From the results, it is clear that the selective regional correlation performs significantly better than the general correlation.

The results clearly indicate that the selective regional correlation is superior method to the general correlation. It is rather important to examine how different time-frequency transforms will affect the resolution of the classifier when applied to the heart sounds. The results in Table 1 confirm earlier work of [9] [10] which have speculated that the S-transform is a better method for the time-frequency analysis of the heart sounds than the STFT and CWT. From the Table 1 it is clear that when the S-transform is used the highest correlation coefficient when there is no match present is on average one third of the highest correlation coefficient when there is a match present. For other two time-frequency representations, this ratio is higher. However, authors would like to emphasize that the superior performance of the S-transform is only confirmed for the heart sounds.

4. CONCLUSION

This paper has examined effects of the three time-frequency transforms on the performance of a newly developed technique known as the selective regional correlation, which relies on redundant representation of a 1D signal in a 2D time-frequency domain for both the template and the signals being processed. The resolution of the classifier is investigated when the proposed method is applied to heart sound analysis to distinguish the opening snaps (OS) from the third heart sound (S3). All three time-frequency methods proved to be reliable for the selective regional correlation with the S-transform being the best.

5. REFERENCES

[1] D.H. Kil, F.B. Shin, *Pattern Recognition and Prediction with Applications to Signal Characterization*, Woodbury, N.Y., AIP Press, 1996.

[2] R.G. Stockwell, L. Mansinha and R.P. Lowe, "Localization of the Complex Spectrum: The S-Transform", *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 998-1001, April 1996.

[3] K. Grochenig, *Foundations of Time-Frequency Analysis*, Boston, Birkhauser, 2001.

[4] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd edition, San Diego, Academic Press, 1999.

[5] L. Cohen, *Time-Frequency Analysis*, Englewood Cliffs, N.J, Prentice Hall PTR, 1995.

[6] T.S. Huang (ed.), *Two-dimensional Digital Signal Processing I: Linear Filters*, Berlin New York, Springer-Verlag, 1981.

[7] A. Ravin, *Auscultation of the Heart*, 3rd edition, Chicago, Year Book Medical Publishers, 1977.

[8] L.G. Durand, P. Pibarot, "Digital Signal Processing of the Phonocardiogram: Review of the Most Recent Advancements", *Critical Reviews in Biomedical Engineering*, vol. 23, no. 3-4, pp. 169-219, 1995.

[9] G. Livanos, N. Ranganathan, J. Jiang, "Heart Sound Analysis Using the S-Transform", *Computers in Cardiology Proceedings*, pp. 587-590, 2000.

[10] M. Varanini, G. De Paolis *et al.*, "Spectral Analysis of Cardiovascular Time Series by the S-Transform", *Computers in Cardiology Proceedings*, vol. 24, pp. 383-386, 1997.