

# Variance-wise Segmentation for a Temporal-Adaptive SAX

Chao Sun<sup>1</sup>David Stirling<sup>1</sup>Christian Ritz<sup>1</sup>Claude Sammut<sup>2</sup>

<sup>1</sup> School of Electrical, Computer and Telecommunications Engineering  
University of Wollongong  
Wollongong NSW 2522, Australia,  
Email: {chaos,david\_stirling,critz}@uow.edu.au

<sup>2</sup> School of Computer Science and Engineering  
The University of New South Wales  
Sydney 2052, Australia  
Email: claude@cse.unsw.edu.au

## Abstract

The Symbolic Aggregate approXimation algorithm (SAX) is a very popular symbolic mapping technique for time series data, and it is widely employed in pattern identification, sequence classification, abnormality detection and other data mining research. Although SAX is a general approach which is adaptable to most data, it utilises a fixed-size sliding window in order to generate motifs (temporal shapes). When certain target phenomena (activities of interest) are manifested over differing time scales, SAX-motifs are unable to correctly account for all such targets. This paper proposes a new method named the variance-wise segmentation method which can adaptively change the size of the sliding window in a generalised SAX approach. By generating motifs with differing durations, patterns can be found for activities with similar shape but occurring over a significant altered time base. This method is tested on both artificially modified ECG data, as well as, variable tactile vibration data, with improved results compared to the original SAX formulation.

*Keywords:* Time Series Data, Adaptive Segmentation, SAX

## 1 Introduction

The time series is an important type of data due to its frequent appearance in various fields, such as finance, weather forecast, medicine and industry. Analysis and data mining on time series are distinct and often difficult compared to other common data types, mainly due to the natural temporal ordering characteristic.

The Symbolic Aggregate approXimation (SAX) developed by Lin et al. (Lin et al. 2003) is a technique designed for time series data mining, and it is widely employed for analysis on time series from many kinds of sources. SAX has been proven to be efficient and reliable for abnormality detection (Keogh & Lin 2005), sequence classification (Lin et al. 2003), similar patterns locating (Mueen & Keogh 2010) and many other tasks. However, the SAX method uses a fixed-size sliding window on the full range temporal data for the generation of symbolic motifs, which are

further used as the basic elements to decompose non-stationary time series, the same type of events with a temporal distortion may not be disassembled into sets containing similar motifs.

The assumption is that if a certain type of event can be recorded as a series of sequences, the shapes (temporal change patterns) of the recorded sequences will be similar to each other. However time distortions on records often happen in the real world when some activities are happening more quickly or slowly, and in this case, the shape patterns are stretched or compressed along the time axis. If the frequency changes happen as the time series data are being recorded, the traditional SAX technique with a fixed-size sliding window will have difficulty in correctly identifying the actual type of events.

This work proposes a dynamic segmentation method that is designed to improve the SAX technique. The new segmentation method is aimed at making the SAX transformation adaptive to time distortions in non-stationary time series by changing the size of sliding window dynamically. The remainder of this paper is structured as follows: Section 2 reviews related research. Section 3 describes the concept of variance-wise segmentation. Section 4 details experimental trials, and a comparison between the new and original SAX methods. Section 5 summarises the advantages and disadvantages of this method, and discusses possible improvements for the future.

## 2 Related work

A key problem in processing time series data efficiently and effectively is linked to its sequential representation. Transform or approximate representations of time series data can be categorised as either frequency-domain or time-domain (Keogh et al. 1993). Most transform methods, such as the Fourier and Wavelet Transform, belong to the frequency-domain, whilst in the time-domain, approximations such as, Symbolic Mappings and Piecewise Linear Representation (PLR) are widely used for time series representation.

### 2.1 Symbolic Representation – SAX

SAX is a typical symbolic mapping approximation for time series representation. Originally, SAX (Lin et al. 2003) used a sliding window to extract a section of sequential data, where the section is divided into several sub-sequences of equal length based on the specified SAX word length. Values in each subsequence are averaged and the whole section is converted into a

Copyright ©2012, Australian Computer Society, Inc. This paper appeared at the 10th Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, December 2012. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 134, Yanchang Zhao, Jiuyong Li, Paul Kennedy and Peter Christen, Ed. Reproduction for academic, not-for-profit purposes permitted provided this text is included.

Piecewise Aggregate Approximation (PAA) representation (Chakrabarti et al. 2002). The PAA is then discretised into a list of symbols or values based on the quantising region that every PAA value lies within. The SAX mapping approach uses a Gaussian distribution to derive the regional breakpoints resulting in the generation of an equiprobable set of symbols, as illustrated in Figure 1, adapted from one of Keogh's paper (Keogh & Lin 2005) as an illustration.

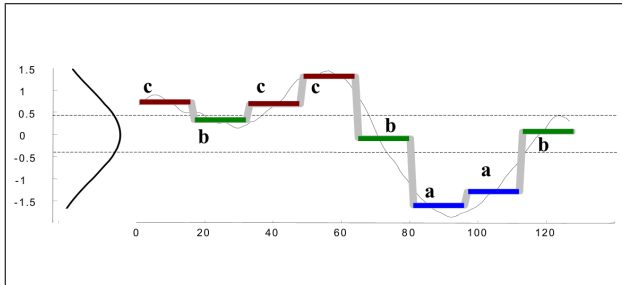


Figure 1: Concept of Original SAX

The SAX representation satisfies a Lower Bounding (LB) distance measurement property such that if two subsequences are similar, then the distance measurement (typically Euclidean) between their SAX motifs will be small, and generally less than the distance between the two original sequences. Lower Bounding is an important characteristic of SAX for pattern searching and indexing, and it promotes the SAX motifs to be used in conjunction with other data mining techniques.

The SAX motif represents the rough shape of the original time series data, which significantly reduces the amount of data to be processed. The technique is designed for quickly locating sections roughly matching a given shape pattern. However, with very long time series, the number of SAX motifs may also grow huge and makes the searching within the motif set difficult. In order to obtain improved index-ability for SAX representations, a multi-resolution extension of SAX, iSAX, was proposed by Shieh and Keogh (Shieh & Keogh 2008). The iSAX representation allows mixed cardinalities in the SAX words, and the symbols are represented as binary strings. This enables resultant SAX words to become hierarchically index-able, and this leads to fast approximate search with various resolutions. However, iSAX still relies on a Gaussian determination of breakpoints for the PAA discretisation, together with a fixed size sliding window as in the original SAX approach.

Because both SAX and iSAX presentation rely on a fixed Gaussian distribution threshold table, they are not adaptive to magnitude distribution changes within time series. Pham et al. (Pham et al. 2010) combined K-Means clustering algorithm to adaptively set the break points for each subsequence in the time series data. With Pham's adaptive SAX/iSAX method, the ranges for SAX symbols change accordingly to the sequential data and generates adaptive SAX representations which are claimed to outperform the original SAX and iSAX approaches. The adaptive (i)SAX can change the breakpoints based on the data in every sliding window, which solves the skew problem over different symbols when the values in datasets have a non-Gaussian distribution. However, for the patterns encased within varying temporal modes, adaptability does not exist for the temporal-modal, or frequency changes within the time series because of the fixed window size. If the SAX words are required to be adaptive in time-domain, a differ-

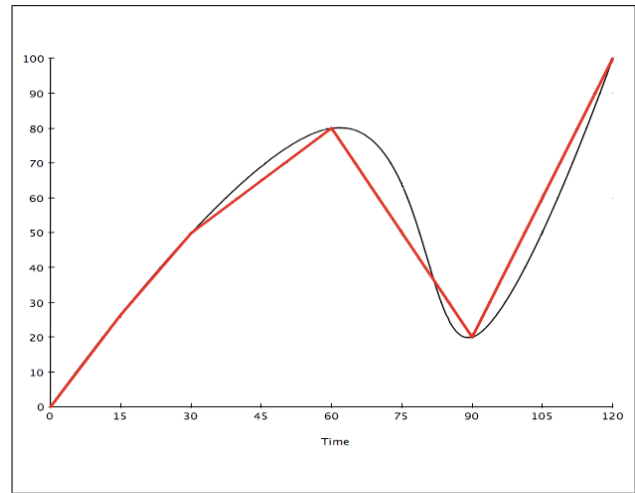


Figure 2: Example of Piecewise Linear Representation

ent segmenting approach has to be employed to adaptively change the size of PAA segments when the time series data alters frequently in nature.

## 2.2 Techniques for Dynamic Segmentation

Appropriate segmentation methods are always an important factor for time series analysis in the time-domain. Unlike the symbolic mapping approach of SAX, reviewed in the previous section, time series data are often compressed into short sequences which contain fewer samples but preserve the most important information of the original data. Segmentation is usually considered to be an optimisation problem which seeks the best fitting approximation with the lowest distance from the original series data.

The Piecewise Linear Representation (PLR) is the most frequently used representation for time series approximation. The PLR approximates a time series with a number of straight lines, and yields the best representation such that the maximum error for any segment is less than a certain threshold, or that the total error of all segments does not exceed the given threshold.

This error can be measured by the Euclidean distance, furthest points or other metrics. Figure 2 illustrates a simple PLR (red line) of the original data (black curve). Various algorithms can be employed to improve the outcome of PLR, such as Neural Networks (Chang et al. 2009), Hidden Markov Models (Ge & Smyth 2001), or Genetic Algorithms (Ghosh et al. 2011).

In financial time series, such as stock price or market index analysis, methods such as, Perceptually Important Points (PIP) are popular for sequence segmentation. The PIP approach was first proposed by Fu et al. (Fu et al. 2001) for financial time series representation. For each subsequence, the PIP segmentation approach initialises the first and last data points as PIPs, then the new PIP is selected by measuring the distance between the data points and the nearby PIPs. The data point with the maximum distance to the closest existing PIPs is subsequently selected as a new PIP, and these are further generated recursively until a stopping criteria is met.

The PIP is a typical Top-down segmentation approach which partitions the whole given sequence into smaller segments, and therefore is not suited for on-line segmentation of streaming data series. The PIP approach is designed to partition temporal sequences

into meaningful patterns similarly to the qualitative perceptions that humans may form, where the fitting error in each segment is not overly restricted in any way. Since there are many empirical financial patterns summarised by technical stock market analysts, the patterns segmented with PIP can be effectively used to match with the known patterns (Chung et al. 2004, Yu et al. 2010).

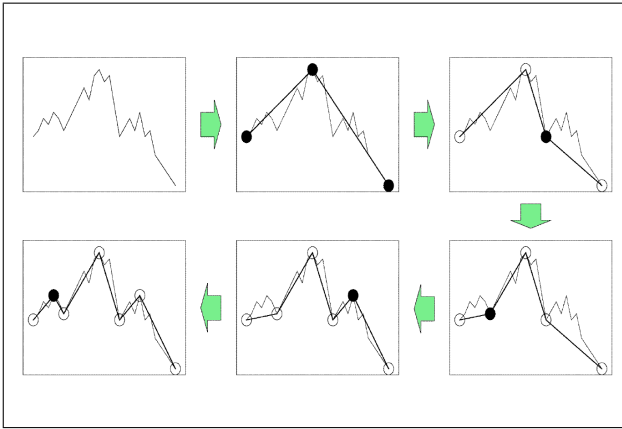


Figure 3: Example of PIP for discovering shoulder-head pattern in financial data

### 3 Temporal Adaptive SAX – using Variance-Wise Segmentation

One advantage of SAX motifs is that they can represent all possible patterns in a sequence as long as its resolution is high enough, and that these motifs can be further analysed by other data mining techniques. PLR produces adaptive segments and detects the change points, however the PLR segments are very likely to be monotonic because they are fitted with straight lines. Using the PLR segments for SAX motif generation is ineffective as a large number of interesting patterns would be ignored.

The other reviewed technique, the PIP, seeks to find important patterns that contain the most distinct points in a given period, however this top-down segmentation method needs the period and a set number of points to be identified prior to the actual segmentation process. Thus the PIP approach does not appear to be useful for dynamic and adaptive segmentation for an online procedure.

Although neither PLR nor PIP can be readily utilised for this purpose, a new segmentation approach has been inspired by both. The essential idea of this new adaptive segmentation approach is that every segment of data requires a sufficient amount of variance to ensure it partially covers some, or the whole, of an important pattern (similar to the concept of maximising distances in PIP), however the variance in any segment of data still needs to be limited to a reasonable extent to ensure motifs are comparable (similar to the maximum error in each segment in PLR).

The assumption here is that if a certain phenomena is recorded as a sequential pattern (shape) in time series data, then a related or similar natured phenomena generated in different time frame will be recorded as a related pattern with time distortion. Most real world temporal processes change in both time and amplitude, therefore distortions can be found on both dimensions. However, in this work we simplify the

**Data:** Sequence, Thresh, WordLen, AlphaSize

**Result:** SAX Motifs, Locations

[saxWords,locs]=vwSax(Seq,Thresh,wLen,alpha)

initialise segStart, segEnd, currSegVar

```

while segEnd < length(Seq) do
  while currSegVar < Thresh do
    | segEnd += 1
  end
  while currSegVar Var >= Thresh do
    newSax = SAXFunc(segment, wLen,
    alpha)
    if newSax ≠ lastWord then
      | update(saxWords, locs, lastWord)
      | segStart += 1
    else
      | continue shrinking
    end
  end
end
return saxWords, locs
    
```

Algorithm 1: Variance-wise Segmenting SAX

problem and limit the distortion within the time domain only, and these situations do readily exist in the real world. For example, ECG data is a typical time series that records activities of the human heart, and the shape of each heartbeat cycle is regular for normal people. The frequency of ECG signals may vary significantly during exercise or rest, however the amplitudes of the ECG waves hardly change (Battler et al. 1979, Dori & Bitterman 2008).

Our method can be described as follows, function  $F()$  is defined to calculate the total variance within a section of data. Threshold  $Th$  is given as the criterion to make segments. For time series data  $T$ , any segment  $[T_i : T_j]$  must satisfy two conditions:  $F(T[i : j]) \geq Th$  and  $F(T[i : j - 1]) < Th$ . ( $i$  and  $j$  are the start and end points of the section, and  $T[i : j]$  are the sequential data for SAX motif).

The function  $F()$  can be defined by different methods, such as the total absolute differences among points, the total square error from the average value or any other variance measurements with similar meaning. The threshold  $Th$  is a fixed value or a value adaptively derived from the data. Unlike the PIP method, this segmentation method does not expect that the extracted patterns are physically meaningful. As long as the SAX motifs are generated for further modelling, the meaning and importance of the motifs can be determined later by other data mining techniques.

The segmentation method uses a sliding window with variable size, therefore it suits online SAX motif generation. When the current variance of the segment is less than the threshold, the window is expanded by moving the end point forward. And when the current variance of the segment is greater than the threshold, the window is reduced by advancing the starting point. Similar to original SAX, every segment with enough variance will be converted to a SAX motif, and the new motif is compared with the previous motif. If the new motif differs from the previous one, both segments and the motif will be saved for future use. Figure 4 illustrates how this method works under different situations. Algorithm 1 is the pseudo code for the new segmentation approach (variance function is not included).

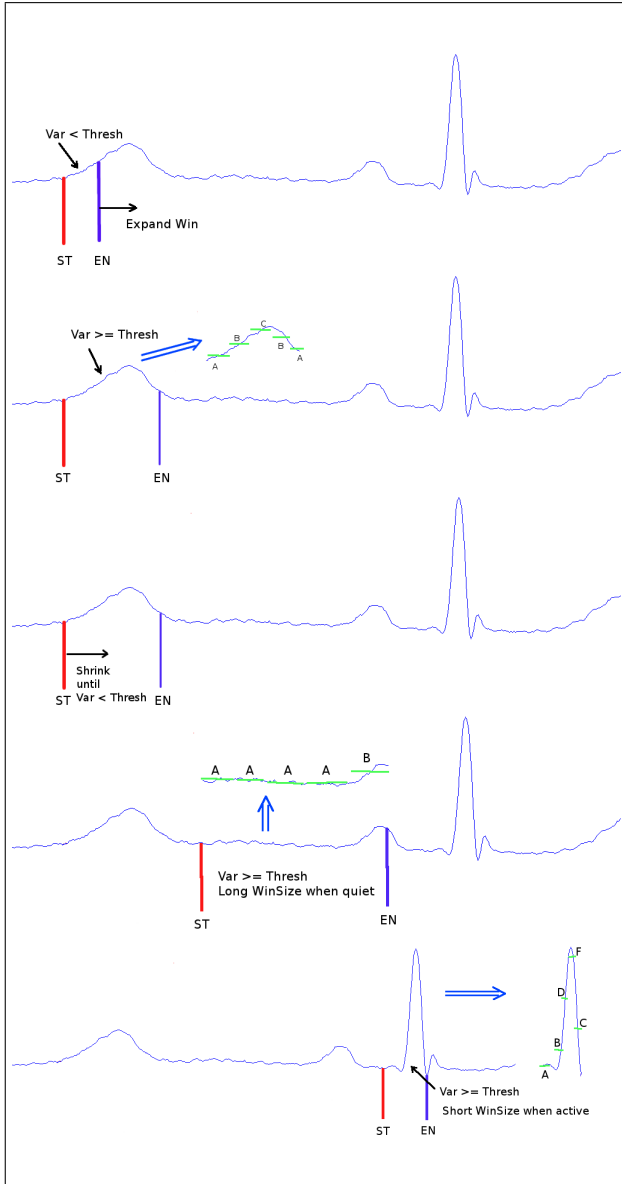


Figure 4: Temporal Adaptive SAX using Variance-wise Segmentation Method

## 4 Experiment

The results of preliminary experiments are presented in this section. Real ECG data obtained from the MIT-BIH database (Moody & Mark 2001) have been artificially modified for testing purposes. The variance function is set to accumulate the absolute differences of nearby records in a specific segment of data, and the threshold for variance is selected to be certain multiples of the standard deviation of the same data.

$$F(T[i : j]) = \Sigma(|diff(T[i : j])|)$$

$$Thresh = N \times std(T)$$

Besides the ECG data, a set of real data from a tactile sensor used by Jamali and Sammut (Jamali & Sammut 2012) are also tested using the same code.

### 4.1 MIT-BIH ECG Data and Setups

Original SAX is used in the first experiment, and the word length and alphabet size are both set to 5. Figure 5 is one of the artificially modified ECG sequences

Table 1: Var-wise SAX Event Stats (ECG)

Evt	SampNo	MotNo	AvgWin	MinWin	MaxWin
N1	327	29	67	22	173
N1*	932	39	200	60	440
N2	309	30	62	20	173
N2*	878	35	170	50	399
PVC	380	18	107	40	304
PVC*	1080	23	283	108	799

tested, captured from the No. 119 ECG record in the MIT-BIH database. The whole sequence in Figure 5 contains 7700 samples, where the first 2000 samples are from the original ECG signal (stage 1) including four normal and a premature ventricular contraction (PVC) heartbeats sampled at 360 Hz. The rest of the data (stage 2) are interpolated from the stage 1 data using the “resample” function in the Matlab time-series toolbox (Matlab 2010b). The signal in stage 2 is increased to 5700 samples and accordingly its sampling rate is 1026 Hz.<sup>1</sup>

In Figure 5, the normal and PVC heartbeats in stage 1 have distinct temporal shape patterns, and these shape patterns are well maintained in stage 2 after the interpolation. The second and third normal heartbeats and the PVC in stage 1 are highlighted and labelled as N1, N2 and PVC, and accordingly their up-sampled versions are N1\*, N2\* and PVC\*. Using the original SAX approach, a sliding window covering the whole ECG event in stage 1 can only cover less than half the event in stage 2, thus obviously the motif set representing N1 will be different from the motif set representing N1\*. In both SAX approaches, a window (with either fixed or varying size) is used to slide through the whole sequence and generates SAX words continuously. Therefore, for a pre-defined event, some SAX motifs are fully included and some others are overlapped with the event. The motif set for an event is constructed with all fully included SAX motifs within the period of such event. Table 1 lists the sample number, number of motifs, average window size and min/max window size in the motifs sets for all the 6 events highlighted in Figure 5.

A few facts can be observed from the statistics summary in Table 1:

1. For the variance-wise SAX, the sliding window does adaptively increase its size during stage 2, and the change ratio is close to the actual up-sample ratio. (2.6–2.9:1 on the average window sizes against the real up-sample ratio at 2.85:1).
2. The sliding window may vary more than double the size according to the activity on time series data. Increasing the threshold increases the possible coverage of each segment, and once the segment is able to cover a whole normal event, the window size will vary less during the normal periods. Some later experiments show that if the threshold is increased to 15 times the standard deviation, the window size for N1 event vary less (185–217 samples). However, for irregular events with significant changes, such as PVC and PVC\*, the window size still varies significantly (100–306 samples) during the abnormal period.

<sup>1</sup>Please note that the frequency difference on the ECG signal is artificially manipulated to an extent that will never happen in real. It is common to see the heartbeat rate on the same subject to increase from 60 bpm to 100 bpm, however for subjects in the MIT-BIH database (mostly patients), there is no record containing such rate variance. Therefore the data used in this work is generated for validation purposes only, and the frequency change of the data can also be variable.

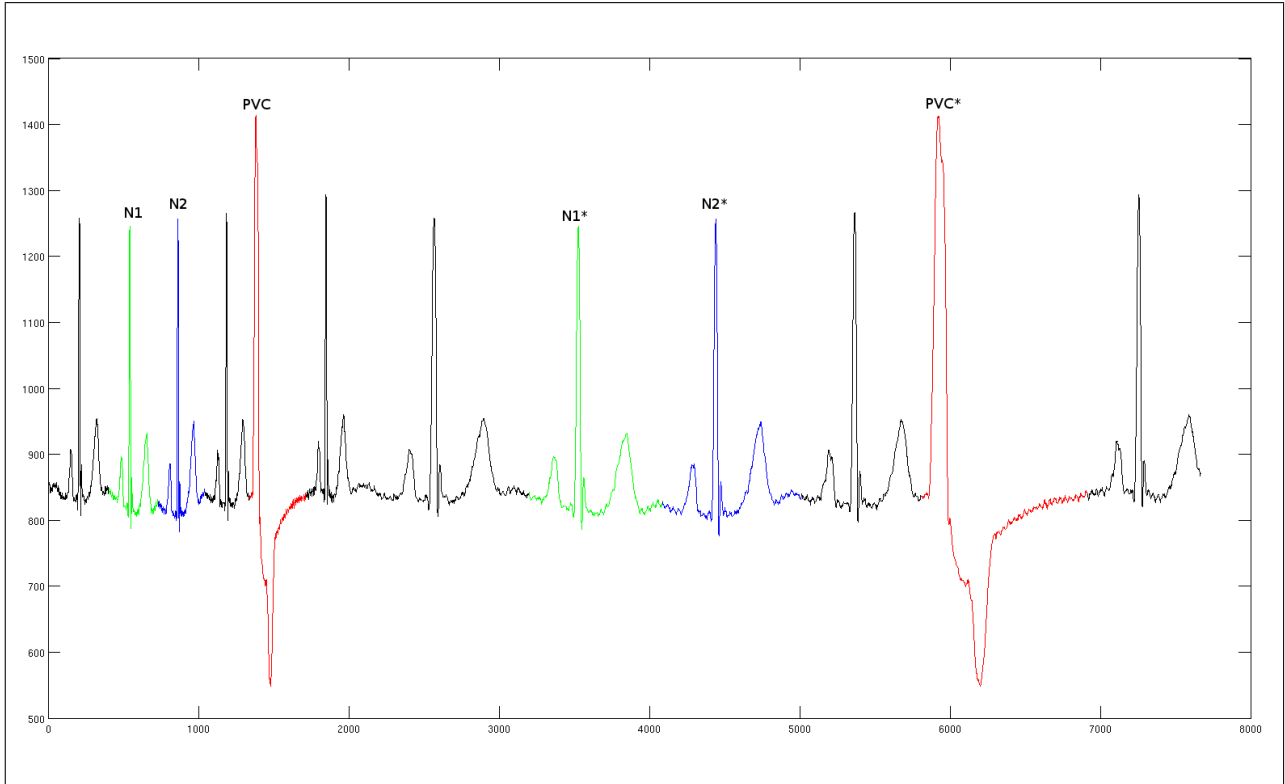


Figure 5: Artificially Changed ECG Signal

- With the adaptive sliding window, the sizes of the motif set for the same event in both stage 1 and 2 are close, and similar motif patterns can be expected from both stages regardless the frequency difference.

#### 4.2 ECG outcomes

As the number of motifs for each event varies from 18 to 39, it is not necessary to list all of the actual SAX motifs. In order to compare the motif sets efficiently, we design a table to show the similarity correlations between different motif sets. Every row of the table represents an event, and the columns indicate the event to be compared. Every element is constructed by two values (C/D), Common(C) is the number of motifs which find a match in the comparison motif set, and Different(D) is the number of motifs that miss a match in the comparison set. When the total absolute difference of two motifs is less or equal to 1, we define it as a similar match. For example, the total difference between [2,2,3,4,5] and [2,2,2,4,5] is 1, so they are considered a match. However the differences from [2,2,3,4,5] to [2,3,4,4,5] and [4,2,3,4,5] are both 2, so they are not similar matches. A high (C/D) ratio means that the SAX representation set for the row event is similar to the motif set or part of the set as comparison.

The threshold for the variance-wise segmentation method is set to 10 times the standard deviation, and for the normal SAX method, a sliding window size of 160 is used, which is close to half the period of a normal heartbeat event in stage 1. Results for both variance-wise SAX and normal SAX are listed in Table 2 and 3 respectively.

In both Table 2 and 3, the comparisons between a normal ECG event and the same event in artificially stretched ECG signal are highlighted with bold fonts. For example, the first row in both tables compares

Table 2: Motif Set Similarity – Var-Wise SAX (ECG)

C/D	N1	N2	PVC	N1*	N2*	PVC*
N1	29/0	28/1	0/29	<b>29/0</b>	25/4	0/29
N2	28/2	30/0	0/30	29/1	<b>27/3</b>	0/30
PVC	0/18	0/18	18/0	1/17	2/16	<b>18/0</b>
N1*	<b>38/1</b>	36/3	1/38	39/0	36/3	1/38
N2*	26/9	<b>29/6</b>	3/32	32/3	35/0	3/32
PVC*	0/23	0/23	<b>23/0</b>	1/22	2/21	23/0

Table 3: Motif Set Similarity – Normal SAX (ECG)

C/D	N1	N2	PVC	N1*	N2*	PVC*
N1	31/0	27/4	0/31	<b>16/15</b>	13/18	2/29
N2	27/4	31/0	1/30	14/17	<b>15/16</b>	7/24
PVC	0/18	1/17	18/0	4/14	4/14	<b>17/1</b>
N1*	<b>15/41</b>	15/41	5/51	56/0	52/4	8/48
N2*	17/56	<b>17/56</b>	6/67	65/8	73/0	10/63
PVC*	1/32	4/29	<b>18/15</b>	6/27	6/27	33/0

the motif set from event N1 with motif sets from all other events. According to Table 2, the variance-wise SAX generates 29 unique motifs during the period of N1, and in which 28 motifs found similar matches in the motif set for event N2, therefore N1 is very similar to N2 based on their SAX motif components. N1 is dissimilar with PVC as no motif finds a match in the PVC's motif set. Comparison between N1 and the stretched events N1\* and N2\* also show good similarities using the variance-wise SAX. All motifs from N1 can find matches in the motif set for N1\*, and 25 out of 29 motifs find matches in the motif set for N2\*. This means the stretched normal heartbeats (N1\*, N2\*) produce motif sets similar to the motif set of original heartbeat (N1). A comparison of the motifs between N1 and PVC\* shows they are very dissimilar as no motif can be matched between these two. Comparisons among all other events are similarly listed in



Table 4: Motif Set Similarity – Tactile Data

C/D	VW SAX			Orig SAX		
	Ev1	Ev2	Ev3	Ev1	Ev2	Ev3
Ev1	24/0	19/5	16/8	41/0	26/15	7/34
Ev2	22/4	26/0	22/4	32/10	42/0	23/19
Ev3	14/13	23/4	27/0	11/17	22/6	28/0

the rest part of Table 2.

In Table 3, 31 motifs are generated for N1 with the normal SAX approach, and the comparison results are close to the variance-wise SAX when they are done within the same frequency range. However, when N1 is compared with stretched normal heartbeats N1\* and N2\*, the number of similar matches drop to about half of the total motif number (16 with N1\* and 13 with N2\*). Comparison between N1 and PVC\* shows that 29 out of 31 motifs are dissimilar, leaving 2 motifs with matches in the PVC\* motif set.

Analysis of the tables above shows the advantage of variance-wise SAX as its C/D ratios are not affected significantly by changes in frequency. The distinctions of motif sets mostly correlate with the difference of event types. However when using the traditional SAX approach for the same task, the C/D ratio between motif sets of an event and its stretched version drops significantly from 100:0 to approximately 50:50.

The work conducted on the artificial ECG data illustrates how the variance-wise segmentation method helps on generating temporal-adaptive SAX motifs. In the next section some data captured from a real application will be used for further evaluation.

### 4.3 Tactile Data

These sequential data were captured during texture recognition experiments with an artificial finger. Vibrations are detected by several polyvinylidene fluoride (PVDF) sensors embedded in the finger (Jamali et al. 2009).

The density of a moving set of mechanical ridges that are drawn across the surface of the artificial finger changes from 10 units to 30 units and then 50 units, these are referred to as three different stages, S1, S2 and S3. We randomly select data from each stage as follows: 1450 samples in S1, including 4 ridges; 1000 samples in S2, including 10 ridges; 1000 samples in S3, including 17 ridges. The resultant data are seen in Figure 6, and one example event is randomly selected from each stage as highlighted with differing colours, similar to previous work with ECG. The sampling rate for PVDF signal is 500 Hz, thus Figure 6 represents about 7 seconds worth of data.

Unlike the artificially stretched ECG data used in previous section, the tactile data contain a significant amount of noise, therefore the performance can be expected to be lower. Both the variance-wise SAX and normal SAX are tested to evaluate the performance with frequency-varying data.

Because the noise and temporal distortion are significant in the tactile data, we allow higher tolerance when comparing motifs. In this case, if the total difference between two motifs is less than or equal to 2, they are considered to match. The threshold for tactile data is set to 7 times the standard deviation. For the normal SAX, the size of sliding window is 30 (Approximately half period of the stage 3 data). The motif set similarity matrices for both methods are listed in Table 4.

Table 5: Match Rate among All Tactile Events

	VW SAX (avg(std)%)		
	Stage1	Stage2	Stage3
Stage1	96.7(6.4)	74.4(16.3)	46.8(17.1)
Stage2	78.2(13.7)	91.2(11.8)	64.6(19.3)
Stage3	40.7(13.1)	59.1(17.2)	66.2(20.9)
	Orig SAX (avg(std)%)		
Stage1	99.5(1.2)	65.2(17.6)	23.9(12.3)
Stage2	70.6(16.8)	89.9(10.3)	43.3(14.4)
Stage3	42.7(16.9)	62.2(17.1)	81.9(16.8)

The temporal adaptive SAX with the variance-wise segmentation method generates similar motif sets across all three stages with different frequency patterns, and the C/D ratios are generally higher than results with traditional SAX approach in Table 4. Table 5 presents the comparison results among all events from Stage 1, 2 and 3. Because the sizes of motif sets are not the same, in this table we use the average percentage of all matching rates instead of the number of matches. The standard deviation of the matching rate is also listed in the brackets.

The high noise level, especially in Stage 3, affects the performance of variance-wise SAX. The noise not only affects the overall shape similarity between stages, but also introduces a lot of variances that affect segmentation and give rise to new shape patterns that do not exist under normal condition. It is believed that improved techniques for calculating the variance function and adding a noise filter to remove random events will provide improved performance for the tactile data.

## 5 Conclusion and Future Work

The paper proposes a new method to dynamically change the size of the sliding window in the SAX algorithm. To generate SAX motifs which are adaptable to time distortions, the total variance within a period is used as the main criterion for segmentation. The method is tested on artificially modified ECG signal and vibration data from a recent real application. The results show the time-adaptive SAX outperforms traditional SAX on sequences with different temporal-modal patterns.

The variance-wise segmenting method is designed to improve the SAX technique, and it can be employed by both the SAX and iSAX. Because the key idea of SAX is not changed, the new time-adaptive SAX motifs are still guaranteed to be lower bounding, and can be utilised for online processing. Further analysis is needed for studying how the new segmenting method changes the lower bound tightness of the SAX motifs.

This work concentrates on frequency varying time series data only, whilst in practice distortions commonly happen on both the time and magnitude axes. Future research will focus on the follows:

1. Appropriate variance function for the type of time series.
2. Evaluation of the method on more real data, such as the tactile signal.
3. Combination of this method and magnitude adaptive (i)SAX for a fully adaptive SAX technique.

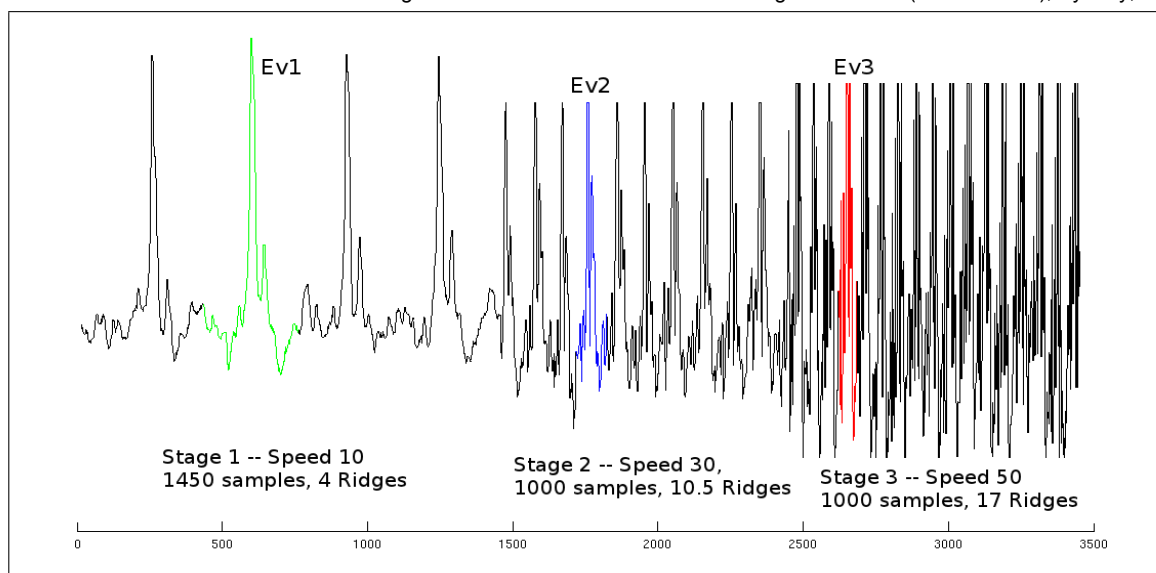


Figure 6: Tactile Data

## References

- Battler, A., Froelicher, V., Slutsky, R. & Ashburn, W. (1979), 'Relationship of qrs amplitude changes during exercise to left ventricular function and volumes and the diagnosis of coronary artery disease.', *Circulation* **60**(5), 1004–13.
- Chakrabarti, K., Keogh, E., Mehrotra, S. & Pazzani, M. (2002), 'Locally adaptive dimensionality reduction for indexing large time series databases', *ACM Trans. Database Syst.* **27**(2), 188–228.
- Chang, P.-C., Fan, C.-Y. & Liu, C.-H. (2009), 'Integrating a piecewise linear representation method and a neural network model for stock trading points prediction', *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* **39**(1), 80–92.
- Chung, F.-L., Fu, T.-C., Ng, V. & Luk, R. (2004), 'An evolutionary approach to pattern-based time series segmentation', *Evolutionary Computation, IEEE Transactions on* **8**(5), 471–489.
- Dori, G. & Bitterman, H. (2008), 'Ecg variability contour-a reference for evaluating the significance of amplitude ecg changes in two states.', *Physiol Meas* **29**(8), 989–997.
- Fu, T.-C., lai Chung, F., Ng, V. & Luk, R. (2001), 'Evolutionary segmentation of financial time series into subsequences', in 'Evolutionary Computation, 2001. Proceedings of the 2001 Congress on', Vol. 1, pp. 426–430 vol. 1.
- Ge, X. & Smyth, P. (2001), 'Segmental semi-markov models for endpoint detection in plasma etching', *IEEE Transactions on Semiconductor Engineering*.
- Ghosh, S., Ray, A., Yadav, D. & Karan, B. (2011), 'A genetic algorithm based clustering approach for piecewise linearization of nonlinear functions', in 'Devices and Communications (ICDeCom), 2011 International Conference on', pp. 1–4.
- Jamali, N. & Sammut, C. (2012), 'Slip prediction using hidden markov models: Multidimensional sensor data to symbolic temporal pattern learning', in 'Robotics and Automation (ICRA), 2012 IEEE International Conference on', pp. 215–222.
- Jamali, N., Byrnes-Preston, P., Salleh, R. & Sammut, C. (2009), 'Texture recognition by tactile sensing', in 'Australasian Conference on Robotics and Automation (ACRA), December 2-4, 2009, Sydney, Australia'.
- Keogh, E. & Lin, J. (2005), 'Hot sax: Efficiently finding the most unusual time series subsequence', pp. 226–233.
- Keogh, E., Chu, S., Hart, D. & Pazzani, M. (1993), 'Segmenting time series: A survey and novel approach', in 'In an Edited Volume, Data mining in Time Series Databases. Published by World Scientific', Publishing Company, pp. 1–22.
- Lin, J., Keogh, E. J., Lonardi, S. & chi Chiu, B. Y. (2003), 'A symbolic representation of time series, with implications for streaming algorithms', in 'DMKD', pp. 2–11.
- Moody, G. & Mark, R. (2001), 'The impact of the mit-bih arrhythmia database', *Engineering in Medicine and Biology Magazine, IEEE* **20**(3), 45–50.
- Mueen, A. & Keogh, E. (2010), 'Online discovery and maintenance of time series motifs', in 'Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining', KDD '10, ACM, New York, NY, USA, pp. 1089–1098.
- Pham, N. D., Le, Q. L. & Dang, T. K. (2010), 'Two novel adaptive symbolic representations for similarity search in time series databases', in 'Proceedings of the 2010 12th International Asia-Pacific Web Conference', APWEB '10, IEEE Computer Society, Washington, DC, USA, pp. 181–187.
- Shieh, J. & Keogh, E. J. (2008), 'isax: indexing and mining terabyte sized time series', in 'KDD', pp. 623–631.
- Yu, H.-H., Tseng, V., Chen, C.-H. & Hong, T.-P. (2010), 'A pip-based evolutionary approach for time series segmentation and pattern discovery', in 'Computer Symposium (ICS), 2010 International', pp. 705–710.

