

## Support Vector Machines for Classification of Temporomandibular Disorders from Facial Pattern Values

Mansoureh Ghodsi<sup>1</sup> and Saeid Sanei<sup>2</sup>

<sup>1</sup>Cardiff University and <sup>2</sup>University of Surrey

*Abstract:* The aim of this study is to develop a method for detection of temporomandibular disorder (TMD) based on visual analysis of facial movements. We analyse the motion of colour markers placed on the locations of interest on subjects faces in the video frames. We measured several features from motion patterns of the markers that can be used to distinguish between different classes. In our approach, both static and dynamic features are measured from a number of time sequences for classification of the subjects. A measure of nonlinear dynamics of the variations in the movement of colour markers positioned on the subjects faces was obtained via estimating the maximum Lyapunov exponent. Static features such as the number of outliers and kurtosis have also been evaluated. Then, Support Vector Machines (SVMs) are used to automatically classify all the subjects as belonging to individuals with TMD and healthy subjects.

*Key words:* Temporomandibular disorder, maximum Lyapunov exponents, support vector machine.

### 1. Introduction

One of the most dynamic biomechanical junctures in the human body is the temporomandibular joint (TMJ). The TMJs are joints located on either side of the face that connect the lower jaw to the skull (McNeill, 1993) (illustrated in Figure 1).

Temporomandibular disorders (TMDs) often results when the chewing muscles and the TMJ do not work together correctly. When this occurs, the muscles often cramp. This spasm can then become part of a cycle that results in tissue damage, pain and muscle tenderness (Okeson, 1996a; Ohrbach and Widmer, 1992).

TMD is a collective term used to describe a number of related disorders affecting the temporomandibular joints, masticatory muscles, and associated structures, all of which have common symptoms such as pain in or around the ear,

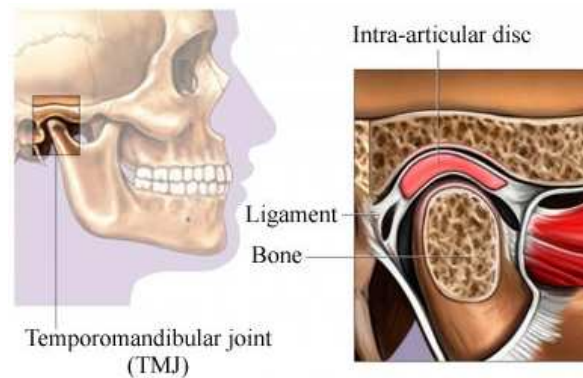


Figure 1: An illustration of the temporomandibular joint and its location adopted from (<http://www.stevenschnolldds.com/tmjdisorders.htm>)

limited mouth opening, tenderness of the jaw muscles, Clicking noises when one opens or closes the mouth and difficulty in opening and closing mouth (McNeill, 1993; Okeson, 1996b).

The clinical diagnosis of TMD has been based on determine the cause of these symptoms by conducting a series of diagnostic tests. These may include complete medical history and clinical examination which is to consider the possibility of temporomandibular joint pain and dysfunction, particularly if the pain is accompanied by clicking jaw joints and limited mouth opening (Carlsson, 1984). Therefore, poor detection of these signs and symptoms can lead to misdiagnosis of TMD.

Currently, there are three methods for measurement of the functional features for diagnosis of TMD; (a) the computerized mandibular scan (CNS) that records the delicate functioning movements of the jaw, (b) Electromyography (EMG) to measure masticatory muscle, and (c) The electrosonograph (ESG) to measure and graphically display or represent sounds made by TMJ components (Deng *et al.*, 2006). These methods, however, are very rarely used in only a few places around the world. Different types of imaging systems may also be utilized for diagnosis of TMD. Arthrography and magnetic resonance imaging (MRI) are the most popular ones. Plain X-ray and computerized tomography (CT) are valuable for determining the presence of osseous changes and traumatic injury to the osseous components of the joint. MRI is costly and unable to visualize perforations of the posterior attachment or the disc. CT is too hazardous to be used frequently and not comfortable for the patients.

The above mentioned reasons indicate that there is a clear need for additional research on alternative methods to be used to supplement clinical examinations with automated classification and detection of TMD. Signal processing methods allow for enhanced clinical utility and an automated approach to the diagnosis

of TMD.

Automatic motion analysis in human recognition is easier and most effective if the subject movement is cyclostationary. There are two obvious cyclic motions: walking and chewing. Analysis of motion vectors in walking subjects has been widely reported in gait recognition (Lee *et al.*, 2006). During normal chewing, the lower jaw and connecting joints on both sides are synchronized; the joints on each side slide and rotate right in front of each ear. Both the cyclic and the chaotic features for these cases can be quantified and exploited in detection of any abnormality (Ghodsí, 2008). In the cases when the jaw twists during one of these motions, it causes pain and click (Ghodsí *et al.*, 2008). The most common symptom of TMD is clicking of the TMJ. In some research the clicking sound has been suggested as a potential data to characterize TMD (Took *et al.*, 2006; Took *et al.*, 2008a; Took *et al.*, 2008b). Since the clinicians hear the mixtures of TMJ sources recorded from inside each auditory canal, it makes the task difficult for them to diagnose TMD. It is difficult to know which type of TMJ source come from the right/left TMJ.

In this paper we develop a method for detection of TMD based on visual analysis of facial movement which is comfortable and safe. The combine of the sound and visual data can help audio analysis to characterize TMD. Here, we consider visual analysis of facial movement. For this purpose we attached a number of markers to the points of interest on the individuals' faces and tracked their positions over a large number of frames in the video sequences. We used image processing methods to extract the positions of the markers in the video frames. The important locations with significant changes during mouth movement are around the TMJ. We analyse the information related to the dynamics of movement of the colour markers placed on the face around the TMJ.

The paper is organized as follows. In Section 2, we provide the necessary mathematical background of the support vector machine (SVM), then we discuss the experimental data in Section 3. Section 4 presents the Bootstrap method and in section 5 the results are derived. Section 6 concludes the paper.

## 2. Support Vector Machines

Support vector machine (SVM) is an effective non-parametric classifier suitable for highdimensional datasets and has been found competitive with the best machine learning algorithms (Vapnik, 1995; Taylor and Cristianini, 2000). Unlike many classification algorithms SVM performs efficiently when

- It is optimum for two-class classification.
- The number of features is high.

- There is a limited time for performing the classification.
- There is a non-uniform weighting among the features.
- There is a nonlinear map between the inputs and the outputs.
- The distribution of the data is not known.
- A convex (monotonic) is required so it does not fall into a local minima.

In recent years, SVM has found a wide range of real-world applications, including face detection from images (Osuna *et al.*, 1997; Shih and Liu, 1996), object recognition (Blanz *et al.*, 1996; Hayasaka *et al.*, 2006), speaker identification (Schmidt, 1996; Moreno and Ho, 2003), text categorization (Joachims, 1997), and biomedical data classification (Shoker *et al.*, 2005).

The formulation of SVM learning is based on the principle of structural risk minimization. Instead of minimizing an objective function based on the training samples (such as mean square error (MSE)), the SVM attempts to minimize a bound on the generalization error (i.e., the error made by the learning machine on test data not used during training) (Vapnik, 1995).

Consider a data set and their corresponding labels as:  $(X_1, y_1), \dots, (X_N, y_N)$ , where  $X_i \in \mathbf{R}^p$  is a feature vector and  $y_i \in \{-1, +1\}$  are the class labels. Then it is possible to partition the  $p$ -dimensional pattern space into two half-spaces with a separating hyper-plane of equation

$$\{X | X \cdot W + b = 0\}. \quad (1)$$

Of all the boundaries determined by  $W$  and  $b$ , the one that maximizes the margin (i.e. maximizes the distance between the hyperplane and the nearest data point of each class) generalizes better than other possible separating hyperplane. Mathematically, this hyperplane can be found by minimizing the following cost function:

$$\vartheta(W, \xi) = \frac{1}{2} \|W\|^2, \quad (2)$$

subject to the separability constraints:

$$y_i [\langle W, X_i \rangle + b] \geq 1, \quad i = 1, \dots, N, \quad (3)$$

where  $\langle \cdot, \cdot \rangle$  means cross-product. This specific problem formulation may not be useful in practice because the training data may not be completely separable by a hyperplane. In this case, slack variables, denoted by  $\xi_i$ , can be introduced to relax the separability constraints in Eq. (3) as follows:

$$y_i[\langle W, X_i \rangle + b] \geq 1 - \xi_i, \quad \xi_i \geq 0; i = 1, \dots, N. \quad (4)$$

Accordingly, the cost function in Eq. (2) can be modified as follows:

$$\vartheta(W) = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^N \xi_i, \quad (5)$$

where  $\|\cdot\|$  refers to the vector geometrical norm. Solving the above equation determines the Lagrangian multipliers and a classifier by implementing the optimal separating hyperplane in the feature space given by

$$f(x) = \text{sign}\left(\sum_{SV} y_i \alpha_i K(X_i, X) + b\right), \quad (6)$$

where  $sv$  refers to number of support vectors. Consequently, everything that has been derived concerning the linear case is also applicable for a non-linear case by using a suitable kernel instead of the dot product. The choice of kernel to fit non-linear data into a linear feature space depends on the structure of the data.

In this project we exploit these distinguished properties of SVMs in order to build up one classifier and apply that to a number of carefully selected and estimated features.

### 3. Data Acquisition

In this study; we used seven subjects. We have two individuals with TMD in the left side of their faces and the rest are healthy subjects. The patients used in our experiments were examined by our clinical expert collaborator <sup>1</sup>. We captured the video of all subjects' faces (healthy and individual with TMD) from the left and right sides in frontal-lateral direction by two cameras (illustrated in Figure 2).

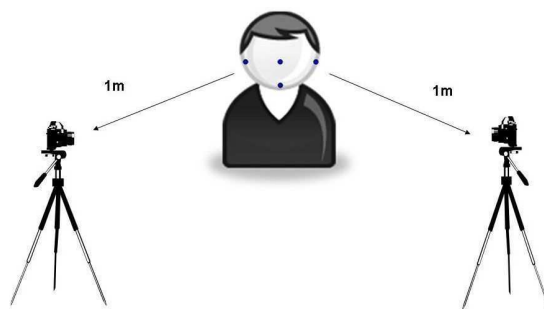


Figure 2: The diagram of cameras and subject position

<sup>1</sup>We are grateful to Prof. S. Dunne of the Dental Institute, Kings College London.

Each subject was captured performing three cycles of chewing motion using a high resolution ( $640 \times 480$  pixels) colour video cameras at 30fps. On average, 400 video frames were obtained per subject. We placed four blue round markers at the locations of interest on each subject's face. The size of each marker is 6mm. We attached two markers on the TMJ at the left and right sides of the face. We also attached two additional markers of the same colour on nose and chin level. The distance of subject from camera can vary thus we used the distance of latter markers as a scaling measure for the images of two sides. We then found the coordinates of the center of each marker in each frame, which we used to find the correspondences between the markers detected in different frames. Therefore, for each marker we obtained a time sequence representing its movement, in direction  $x$ , in the video sequences. We then analyze the motion patterns of the TMJ markers during cycles of chewing. Figure 3 shows the time sequence of a TMJ marker for individual with TMD (left side) and healthy one (right side) in the original scale. As it appears from Figure 3 it is not possible to detect which series is related to individuals with TMD through visual inspection of the signals.

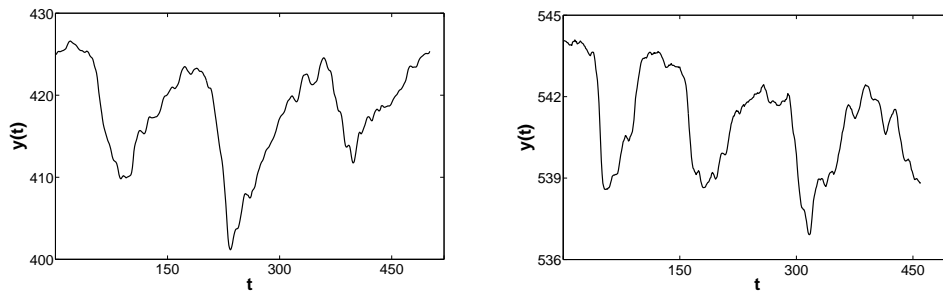


Figure 3: Original time series of the TMJ marker from an individual with TMD (in the left) and that of a healthy individual (in the right)

Some features are measured from chewing motion patterns as described in the following section, that can be used to distinguish between different classes and used SVM as our classification method. It is believed that opening and closing patterns with click is more frequent in patients with TMD than normal subjects and patients with TMD demonstrated a restricted range of motion and reduced velocity than normal subjects.

The static features are measured from the normalized highpass filtered data. The highpass filter is used to remove the effect of mouth movement and enhance the changes in the TMJ within the time sequence. The signals are normalized to suppress the changes in picture/video size, here we define the new signal  $x_{new} = (x_i - \bar{x})/s$  where  $\bar{x}$  is the sample mean and  $s$  is the sample standard deviation. It is believed that the visual features presented to the SVM are distinct enough

to be separated using either a linear SVM or an SVM with a kernel as discussed above.

### 3.1 Features

Although any various order statistics of the data can be considered as features for this purpose, the following features have been empirically found as the most effective and justifiable features for this classification application (Ghodsi, 2008). These features provide both the statistical and dynamical information regarding the effects on the TMD during normal chewing.

**Feature 1.** Any abnormalities in the chewing process produce chaotic behaviour of the motion. Estimating chaos in a dynamical system is an important problem. Measuring maximum Lyapunov exponents (MLE) is a way to solve this problem (Kantz, 1994). As a dynamic feature, we used MLE  $\lambda_1$  to measure the changes in the dynamics of the chewing pattern (Rosenstein *et al.*, 1993). This is measured from the normalized data and it is observed that the chewing signal for the subjects with TMD is more chaotic than for healthy individuals (Ghodsi *et al.*, 2007). We denote  $f_1 = \lambda_1$ . Although in general the estimates of Lyapunov exponents for short data sequences are not very accurate, but the method adopted here (Rosenstein *et al.*, 1993) provides sufficiently distinct values of  $\lambda$  for TMD and normal subjects.

**Feature 2.** Outliers are observations which are presumed to come from a different distribution than those of majority of the data (Han and Kamber, 2001). They can have a profound influence on the data analysis, often leading to erroneous conclusions because of their powerful influence on most parametric tests. Outliers (unusual abnormal values) are often the special points of interest in many practical situations and their identification is the main purpose of the investigation. In medicine, unusual values may indicate the diseases (see, e.g., (Kosheleva *et al.*, 1998)). Accurate identification of outliers plays an important role in data analysis.

One approach to outlier detection is to start with  $N$  normal values  $x_1, \dots, x_N$ , compute the sample mean  $\bar{x}$ , the sample standard deviation  $s$ , and then mark a value  $x$  as an outlier if  $x$  is outside the interval  $(\bar{x} - a s, \bar{x} + a s)$  (for some preselected number  $a$ ). We can therefore identify the outliers as those values that are outside the  $a\sigma$  intervals (for an application of this method in engineering, see, e.g. (Wadsworth, 1990)). Here, we selected  $a = 3$  and used the normalized highpass filtered data.  $f_2 =$  the number of observations  $> 3|\sigma|$ .

The skewness and the kurtosis tests are useful, among the most powerful

tools available for testing the presence of outliers in an otherwise normal sample, especially when the number of outliers is unknown (Hawkins, 1980).

**Feature 3.** A large ratio between the peak (outlier) amplitude and the variance of a signal suggests that there is an unexpected value in the data. The equation describing this feature is given by

$$f_4 = \frac{\max\{|X|\}}{s}, \quad (7)$$

where,  $X = (x_1, \dots, x_N)$  is normalized highpass filtered data,  $\max(\cdot)$  is a scalar valued function that returns the maximum element in a vector,  $s$  is the sample standard deviation of  $X$  and  $|\cdot|$  is the absolute value applied element-wise.

**Feature 4.** Kurtosis can be formally defined as the standardized fourth order moment. Kurtosis is a measure of how sharp a symmetric distribution is when compared to a normal distribution of the same variance.

Note that the kurtosis of a normal distribution is 3. If a distribution has a large central region which is flatter than a normal distribution with the same mean and variance, it has a kurtosis of less than 3 (i.e. sub Gaussian). If the distribution has a central maximum more peaked and with longer tail than the equivalent normal distribution, its kurtosis is higher than 3 (i.e. super Gaussian) (Brooks and Carruthers, 1953).

As noted above, kurtosis largely reflects tail behavior, and so its use for detecting outliers has been considered. Discussions of approaches to detecting outliers using kurtosis can be found in (Barnett and Lewis, 1996; Jobson, 1991). Kurtosis is defined as:

$$f_4 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{s^4}. \quad (8)$$

**Feature 5.** Skewness is a measure of the asymmetry of a distribution and is zero for a normal distribution. If the longer tail of a distribution occurs for values of  $x$  higher than the mean, that distribution is said to have positive skewness. If the longer tail occurs for values of  $x$  lower than the mean, the distribution is said to have negative skewness (Kerbaol and Chapron, 1999). Great skewness may motivate the researcher to investigate outliers. The normalized skewness for each signal is given by

$$f_5 = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{s^3}. \quad (9)$$



Figure 4 confirms that why we used features 2-5.

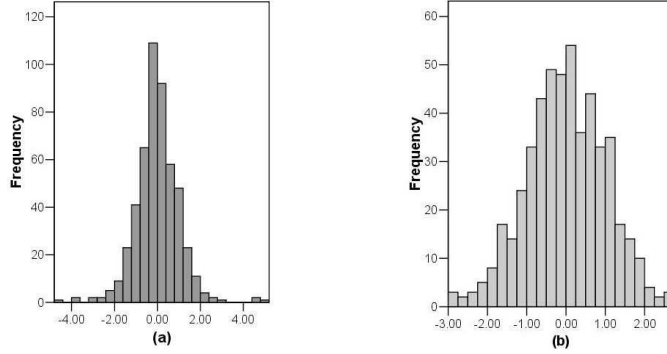


Figure 4: Histogram of the signal of TM marker from (a) the individual with TMD and (b) a healthy individual

**Feature 6.** This feature is a measure of likelihood of a peak subject to the gradient of the smoothed waveform. This feature identifies whether the peak appears in opening or closing the mouth. Let  $u(t)$  be the lowpass filtered data. We denote  $\nabla_t u(t)$  (approximated as  $\nabla_t u(t) = u(t) - u(t-1)$ ) and define

$$I_{\nabla_t u(t)} = \begin{cases} 1, & \text{peak at } \nabla_t u(t) \geq 0, \\ 0, & \text{no peak,} \\ -1, & \text{peak at } \nabla_t u(t) < 0. \end{cases} \quad (10)$$

We denote  $f_6 = I_{\nabla_t u(t)}$ . To acquire a better understanding of  $f_6$ , Figure 5 shows the highpass filtered data (thin line) and  $\nabla_t u(t)$  (thick line) together. As it appears from Figure 5 all peaks occur when  $\nabla_t u(t) \geq 0$ .

#### 4. Data Sample Generation Using Bootstrap

Although a reasonably large number of data samples can be provided by multiple recordings from the same subjects, in those cases where the statistical features are mainly with respect to the data distribution, a number of data samples can be produced. Also, in places where provision of sufficient data with reasonable lengths is difficult a number of data samples can be generated according to the actual data distribution. One of the methods for doing that is bootstrap (Efron and Tibshirani, 1993). This is explained next. We used bootstrap average signal

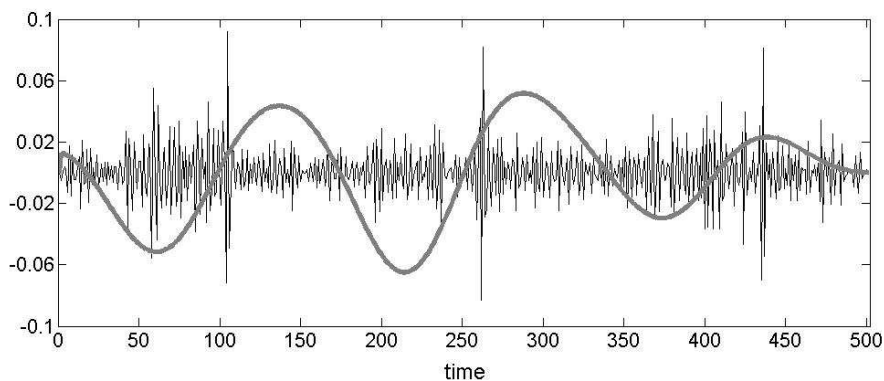


Figure 5: Highpass filtered (thin line) and lowpass filtered data (thick line)

to test the reliability and accuracy of the results obtained from the original signal. Let us consider the method of constructing bootstrap average signal for the signal  $X_t$  (for more information see, e.g., (Efron and Tibshirani, 1993; Golyandina *et al.*, 2001)). Under a suitable choice of embedding dimension  $m$  (we select this parameter using false nearest neighbor) and the corresponding eigentriples in the singular value decomposition (SVD), we have the representation  $X_t = \tilde{S}_t + \tilde{N}_t$ , where  $\tilde{S}_t$  (the reconstructed signal) approximates  $X_t$ , and  $\tilde{N}_t$  is the noise series. Suppose now that we have a (stochastic) model for the noise  $\tilde{N}_t$  (for instance, pure noise). Then, simulating  $n$  independent copies  $N_{t,i}$  of the noise series  $\tilde{N}_t$ , we obtain  $n$  signals  $X_{t,i} = \tilde{S}_t + \tilde{N}_{t,i}$  and produce  $n$  reconstruction results  $\tilde{X}_{t,i}$  (Hassani and Zhigljavsky, 2009).

When the sample  $\tilde{X}_{t,i}$  ( $1 \leq i \leq n$ ) of the reconstruction results is obtained, we can calculate its bootstrap average signal by averaging the bootstrap results. The simplest model for  $N_t$  is the Gaussian white noise model. The corresponding hypotheses can be checked with the help of a standard test for randomness and normality (Golyandina *et al.*, 2001; Hassani, 2007; Hassani *et al.*, 2009).

## 5. Results

As we mentioned above, we used both static and dynamic features extracted from movement of the markers positioned on subjects' faces to detect individuals with TMD. Table 1 represents a summary of the obtained results. The second column represents the left (L) and right (R) sides of face for all subjects; two individuals with TMD in left side (L1, L2) and the rest are healthy people. The values of features 1–6 for all samples are calculated. We followed the same procedure for the observations obtained by bootstrap method. The symbol '\*' in Table 1 indicates that the results obtained by bootstrap method. Columns 3–14 represent the values of  $f_1$ – $f_6$  for both original and bootstrap averaged signals.

Table 1: Values of the features for all subjects/trials

N	S	$f_1$	$f_1^*$	$f_2$	$f_2^*$	$f_3$	$f_3^*$	$f_4$	$f_4^*$	$f_5$	$f_5^*$	$f_6$	$f_6^*$
1	L1	0.091	0.101	6	5	5.24	4.87	6.77	6.10	0.16	0.15	1	1
2	R1	0.074	0.075	0	0	2.87	3.13	2.92	3.00	-0.11	-0.10	0	0
3	L2	0.094	0.102	4	5	3.99	3.98	5.93	5.57	0.14	0.13	1	1
4	R2	0.063	0.071	0	0	2.82	2.85	3.15	3.09	0.08	0.08	0	0
5	L3	0.066	0.691	0	0	2.21	2.36	2.21	2.22	0.10	0.08	0	0
6	R3	0.068	0.081	2	1	2.95	3.06	2.90	2.96	-0.15	-0.13	1	1
7	L4	0.074	0.084	0	0	2.68	2.82	2.96	2.89	-0.02	-0.03	0	0
8	R4	0.077	0.080	1	0	3.12	3.05	3.23	3.28	0.06	0.08	-1	0
9	L5	0.078	0.084	0	0	2.65	2.84	2.84	2.84	-0.03	-0.03	0	0
10	R5	0.073	0.081	1	2	3.2	3.31	3.22	3.31	-0.12	-0.08	1	-1
11	L6	0.065	0.077	0	0	2.69	2.78	3.17	3.10	-0.08	-0.09	0	0
12	R6	0.060	0.062	1	0	3.02	2.89	2.85	2.61	-0.02	0.01	-1	0
13	L7	0.074	0.083	0	0	2.81	2.74	2.88	2.85	0.06	0.06	0	0
14	R7	0.074	0.081	1	0	2.87	2.80	3.00	2.86	0.11	0.09	0	0

As appears from Table 1, the results of bootstrap averaged signals are close to the original signal indicating that our results are reliable. Columns 3 and 4 of Table 1 show the values of  $\lambda_1$  (is rounded) for each colour marker. The values  $\lambda_1$  are positive for all samples (individuals with TMD and healthy subjects) indicating that they have chaotic behavior. However, this value for the individuals with TMD, rows 1 and 3, is larger than those for healthy subjects.

Feature 2 represents the number of outliers for all subjects. It should be noted that outliers are often the special points of interest and their identification is the main purpose of the investigation for instance in medicine may indicate the diseases. Here, the outliers represent the click events in either opening or closing process. The number of outliers for individuals with TMD is greater than those for healthy ones indicating that click is more frequent in the group of individuals with TMD.

A large ratio between the peak amplitude and the variance of a signal is a typical identifier for the click which is shown in feature 3. The normal chewing process is usually distributed about its mean value. Therefore, a low ratio is expected for a health subject, whilst the chewing process for an individuals with TMD has a high value. As the results show, the same pattern obtained here.

Features 4 and 5 indicate how clicks change the distribution of chewing process. Feature 6 is useful to distinguish between the peaks in the signals during

the chewing process related to TMD and other non-relevant peaks. The results of feature 2 are presented in columns 5 and 6. As it is shown in these columns, the number of extreme values for the individuals with TMD is greater than those for healthy subjects which confirms the significance of this feature for classification.

SVM is used to separate TMD and non-TMD classes based on the above features. 14 data segments from 7 subjects were used for training and the data from another 7 subjects were used for testing. A linear kernel was used. In order to test the classification we used cross-validation to test the accuracy of the SVM performance. In the cross-validation procedure we used 70% of the data as training examples and 30% for testing with no overlapping. The cross-validation was performed 10 times, each time the data were randomly rearranged, in order to yield a better estimate of the error. To test the classification results, we compared the classification results with the impressions by our expert clinician. As the result, we were able to classify correctly all the subjects into TMD or non-TMD subjects.

## 6. Conclusions

In this paper, we attempted to use an efficient classification system followed by measurement of a set of carefully selected features to classify the subjects suffering from the most common type of TMD, namely click, from visual data. This provides a simple non invasive and non intrusive procedure for TMD diagnosis. The TMD classifier works based on visual analysis of facial movement. We used a number of carefully selected features extracted from movement of the markers positioned on subjects' faces. In our approach the features are related to both static and dynamic visual variables measured from a number of time sequences corresponding to different subjects within different time intervals, and classified using SVM. The SVM correctly classified the two-class data for all subjects. SVM has been verified as a computationally cost effective method capable of classifying separable and nonseparable data through application of linear and nonlinear kernels. Furthermore, using bootstrap technique the designed classifier was tested and the results obtained from the original data were confirmed. It is achieved that even in the cases of mild TMD the classifier can obtain close to 100% correct classification.

## References

- Barnett, V. and Lewis, T. (1996). *Outliers in Statistical Data*, 3rd ed. Wiley, New York.
- Blanz, V., Schölkopf, B., Bühlhoff, H., Burges, C., Vapnik, V. and Vetter, T.

- (1996). Comparison of view based object recognition algorithms using realistic 3D models, in *Proceedings of International Conference on Artificial Neural Networks*, 251-256.
- Brooks, C. E. P. and Carruthers, N. (1953). *Handbook of Statistical Method in Meteorology*. Her Majesty's Stationery Office.
- Carlsson, G. E. (1984). Epidemiological studies of signs and symptoms of temporomandibular joint-pain-dysfunction a literature review. *Australian Society of Prosthodontists Bulletin* **14**, 7-12.
- Deng, M., Long, X., Dong, H., Chen, Y. and Li, X. (2006). Electrosonographic characteristics of sounds from temporomandibular joint disc replacement. *Journal of Oral and Maxillofacial Surgery* **35**, 456-460.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Ghods, M., Sanei, S., Hicks, Y., Lee, T. and Dunne, S. (2007). Detection of Temporomandibular Disorder from Facial Pattern, in *Proceedings of 15th International Conference on Digital Signal Processing*. Cardiff, UK.
- Ghods, M. (2008). *Non-Linear Dynamical System and Support Vector Machine for Detection of Temporomandibular Disorder*. Master thesis, Cardiff University.
- Ghods, M., Hassani, H., Sanei, S. and Hicks, Y. (2009). The use of noise information for detection of temporomandibular disorder. *Journal of Biomedical Signal Processing and Control* **4**(2), 79-85.
- Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and related techniques*. Chapman & Hall/CRC, New York - London.
- Hassani, H. and Zhigljavsky, A. (2009). Singular spectrum analysis: methodology and application to economics data. *Journal of System Science and Complexity* **22**(3), 372-394.
- Hassani, H., Heravi, S. and Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting* **25**(1), 103-118.
- Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of Data Science* **5**(2), 239-257.

- Han, J. and Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, Massachusetts.
- Hawkins, D. M. (1980). *Identification of Outliers*. Chapman and Hall, London, 19-20.
- Hayasaka, T., Nakauchi, S. and Usui, S. (2006). An Analysis of viewpoint dependency in three dimensional object recognition using support vector machines. *Systems and Computers in Japan* **37**(1), 105-115.
- <http://www.stevenschnoldds.com/tmjdisorders.htm>.
- Joachims, T. (1997). Text categorisation with support vector machines. *Technical Report, LS VIII*, no. 23. University of Dortmund.
- Jobson, J. D. (1991). *Applied Multivariate Data Analysis: Vol. 1: Regression and Experimental Design*. Springer-Vedag, New York.
- Kantz, H. (1994). A robust method to estimate the maximal Lyapunov exponent of a time series. *Physics Letter A* **185**, 77.
- Kerbaol, V., and Chapron, B. (1999). Analysis of the asymmetric distribution of altimeter return signals. *Rapport Techynique IFREMER*.
- Kosheleva, O., Cabrera, S., Osegueda, R., Nazarian, S., George, D. L., George, M. J., Kreinovich, V. and Worden, K. (1998) "Case study of non-linear inverse problems: mammography and non-destructive evaluation", in: Mohamad-Djafari, A. (ed.): *Bayesian Inference for Inverse Problems, Proceedings of the SPIE/International Society for Optical Engineering* **3459**, 128-135. San Diego, CA.
- Lee, T. K. M., Ranganath, S., and Sanei, S. (2006). Frontal view-based gait identification using largest lyapunov exponents, in *Proceedings of IEEE Conference on Signals, Acoustics and Speech Processing*. Toulouse, France.
- McNeill, C. (1993). *Temporomandibular Disorders Guidelines for Classification, Assessment and Management*, 2nd ed. Quintessence Publishing Co., Chicago.
- Moreno, P. J. and Ho, P. P. (2003). A new SVM approach to speaker identification and verification using probabilistic distance kernels, in *Proceedings of Eurospeech*. Geneva, Switzerland.

- Ohrbach, R. and Widmer, C. G. (1992). Review of the literature. Research diagnostic criteria for temporomandibular disorders review, criteria, examinations and specifications, critique. *Journal of Craniomandibular Disorders: Facial and Oral Pain* **6**, 307.
- Okeson, J. P. (1996a). Temporomandibular disorders in the medical practice. *Journal of Family Practice* **43**, 347.
- Okeson, J. P. (1996b). *Orofacial Pain: Guidelines for Assessment, Diagnosis and Management*. Quintessence Publishing Co., Chigaco.
- Osuna, E., Freund, R. and Girosi, F. (1997). Training support vector machines: an application to face detection, in *Proceedings of Computer Vision and Pattern Recognition*, 130-136.
- Rosenstein, M. T., Collins, J. J. and De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D* **65**, 117.
- Shih, P. and Liu, C. (2006). Face detection using discriminating feature analysis and support vector machine. *Pattern Recognition* **39**, 260-276.
- Schmidt, M. (1996). Identifying speaker with support vector networks, in *Proceedings of Interface*. Sydney.
- Shoker, L., Sanei, S. and Chambers, J. (2005). Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm. *IEEE Signal Processing Letters* **12**(10), 721-724.
- Taylor, J. S. and Cristianini, N. (2000). *Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge.
- Took, C. C., Sanei, S., Rickard, S., Chambers, J. and Dunne, S. (2006). Fractional delay estimation for blind source separation and localization of temporomandibular joint sounds. *IEEE Transactions on Biomedical Engineering* **53**, 2123-2126.
- Took, C. C., Sanei, S., Chambers, J. and Dunne, S. (2008). Underdetermined blind Source Separation of Temporomandibular Joint Sounds. *IEEE Transactions on Biomedical Engineering* **55**, 949-956.
- Took, C. C. and Sanei, S. (2008). Exploiting sparsity, sparseness and super-Gaussianity in underdetermined blind identification of temporomandibular joint sounds. *Journal of Computers* **2**(6), 65-71.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.

Wadsworth, H. M. (1990). *Handbook of Statistical Methods for Engineers and Scientists*. McGraw-Hill, New York.

Received November 20, 2009; accepted May 19, 2010.

Mansoureh Ghodsi  
School of Mathematics  
Cardiff University  
Senghennydd Road, Cardiff, CF24 4AG, UK  
GhodsiM@cardiff.ac.uk

Saeid Sanei  
Nature Inspired Computational Engineering (NICE) Group  
Faculty of Engineering and Physical Sciences University of Surrey  
Guildford GU2 7XH, Surrey, UK  
Saneis@cardiff.ac.uk