# A FACIAL PATTERN RECOGNITION APPROACH FOR DETECTION OF TEMPOROMANDIBULAR DISORDER

*M. Ghodsi*[1]*, S. Sanei*[1]*, Y. Hicks*[1]*, T. Lee*[2]*, and S. Dunne*[3]

[1] Centre of Digital Signal Processing, Cardiff University, UK
[2] School of EEE, Singapore Polytechnic, Singapore
[3] Dental Institute, King's College London , UK

## ABSTRACT

*The aim of this study is to automatically classify individuals with temporomandibular disorder and healthy subjects. The process of automated classification requires measurement of features that can be used to distinguish between different classes. We used maximum Lyapunov exponents to measure the changes in the dynamics of the chewing pattern, the number of peaks in the normalized highpass filtered data to find the abnormalities in both opening and closing of mouth, normalized skewness and kurtosis to measure the distribution profile of the data samples, likelihood information to quantify the probability of the click events in either opening or closing process, and peak amplitude to measure how severe the abnormality is. Finally, using the above features together with Support vector machine to classify all subjects as belonging to individuals with TMD or not. The early experiments show encouraging results.*

**Keywords:** *Temporomandibular disorder (TMD), maximum Lyapunov exponents, support vector machine (SVM).*

## 1. INTRODUCTION

Temporomandibular disorders (TMDs) occur as a result of problems with the jaw, temporomandibular joint, and surrounding facial muscles that control chewing and moving the jaw[1][2].Temporomandibular joint (TMJ) is the hinge joint that connects the lower jaw (mandible) to the temporal bone of the skull, which is immediately in front of the ear on each side of your head. The joints are flexible, allowing the jaw to move smoothly up and down and side to side. Symptoms of TMD include headaches, tenderness of the chewing muscles, and clicking or locking of the joints [3].

Current methods for TMD detection involve a physical examination by an expert in the area [4]. A dentist or clinician almost always diagnoses a TMD based solely on a person's medical history and on a physical examination. Recent research attempts to detect TMD on the basis of audio analysis [5]. Electronic recording of sounds emitted from the jaw during jaw opening and closing. TMJ sound has been suggested as a potential tool to characterize the TMDs.

We attempt to develop a method for detection of TMD based on visual analysis of facial movement. In this study, we look at changing facial features in the video frames recorded using one video camera through frontal-lateral direction. Afterwards, we analyse the motion of colour markers placed to the locations of interest on subjects' faces. We use image processing methods to extract the positions of markers in video frames.

Emails: {ghodsim, saneis, hicksya}@cf.ac.uk, tlee@sp.edu.sg, and stephen.dunne@kcl.ac.uk

The important locations with significant changes during mouth movement are around the TM bone. The patients used in our experiments were examined by our clinical expert collaborator. The marker designations are: left side of TM bone (lt). Two additional markers are attached to tip of the chin and tip of the nose. We use the latter markers to suppress the head movement in various directions. In our experiments, we used video sequences of eight subjects, two of which had TMD (class 1) and six healthy people (class -1).

## 2. METHODS

In this study, we analyse the information related to the dynamics of movement of the colour markers placed on the face around the mandible. Next, we explore how these dynamics can be measured and used here.

### 2.1 Chaotic Biological Movement

Van Emmerik et al. [6] in a tutorial overview, discuss how various seemingly simple human actions are the result of the interaction of complex systems. West and Scafetta [7] analyse the stride length of humans which have been shown to be slightly multifractal and can be modelled using nonlinear oscillators. Tracey et. al used chaotic measures to identify humans by their gait, which also set a significant precedent [8]. We expect that the process of chewing has a chaotic behavior and it is more chaotic for the individual with TMD rather than for healthy subjects. To test this hypothesis, we use the maximum Lyapunov exponent to evaluates the dynamics of the system.

The concept of phase-space analysis of chaotic systems is extended here to enable joint analysis of a number of motion trajectories at the same time. The trajectories specify motion of a number of colour markers during the process of opening and closing mouth. We can thus characterize their behavior.

### 2.2 Dynamical System

Consider the real-valued nonzero time series $(x(1),\ldots,x(N))$ of sufficient length $N$. We create the time delay vectors as follows:

$$\mathbf{V} = (\mathbf{x}(\mathbf{n}), \mathbf{x}(\mathbf{n}+\tau), \ldots, \mathbf{x}(\mathbf{n}+(\mathbf{d}-\mathbf{1})\tau)) \qquad (1)$$

where $\tau$ is the time delay, $d$ is the embedding dimension, and $n = N-(d-1)\tau$.

According to the embedding theorems [9] [10], if the time series is generated from a deterministic system, there generally exists a function $\mathbf{F} : \mathscr{R}^{\mathbf{d}} \mapsto \mathscr{R}^{\mathbf{d}}$ such that:

$$\mathbf{V}_{n+1} = \mathbf{F}(\mathbf{V}_{\mathbf{n}}) \qquad (2)$$

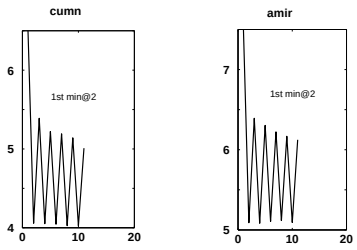


Figure 1: Mutual information of the left TM marker (lt) for two samples.

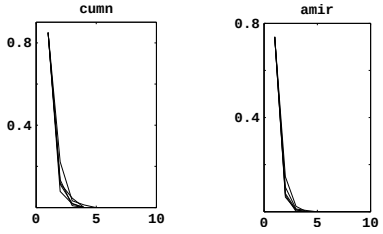


Figure 2: False nearest neighbors of the left TM marker (lt) for two samples.

when the embedding dimension $d$ is sufficiently large this mapping has the same dynamic behavior as that of the underlying unknown system in the sense of topological equivalence.

The remaining problem is how to choose a $\tau$ and $d$, *i.e*, a time-delay and embedding dimension, such that the above mapping exists. From Takens' embedding theorem [9] it does not matter what time delay is selected in a 'generic' sense. In practice, however, because we have only a finite number of data points available with limited measurement precision, a good choice of $\tau$ is deemed to be important in the phase space reconstruction. In addition, determining a good embedding dimension $d$ depends on the judicious choice of $\tau$. For more discussion on the topic, see the related articles such as [11].

To ensure selection of a minimum embedding dimension, we select the time-delay by the *mutual information* [12]. A sample plot is shown in Fig. 1. Each plot shows the mutual information of colour marker against the time delay. The point at which the first minimum of the plot is taken is the best value for $\tau$ which $\tau = 2$ in this case and for all samples.

The next challenging problem is selection of the embedding dimension. The method of *false nearest neighbors* [13] is employed to choose the embedding dimension. A sample plot is shown in Fig. 2. Each plot shows the false nearest neighbors of the colour markers against the embedding dimension. The best value for all samples is $d = 5$.

### 2.3 Maximum Lyapunov Exponent

Lyapunov exponents measure the average exponential separation between the nearby phase space trajectories. It is the generic mechanism for deterministic randomness and unpredictability. There are several measures of chaotic behaviour, the maximum Lyapunov exponent $\lambda_1$ being the most useful and commonly used. Positive Lyapunov exponents, for almost all initial conditions in a bounded dynamical system, are widely used as the definition for deterministic chaos. One of the more recent methods to calculate $\lambda_1$ is by Rosenstein [14] and independently, by Kantz [15]. This method is suitable for small and noisy data sets. The maximum Lyapunov exponent can be defined using the following equation where $d(t)$ is the average divergence at time $t$ and $C$ is a constant that normalizes the initial separation:

$$d(t) = C e^{\lambda_1 t}. \tag{3}$$

The maximum Lyapunov exponent is estimated as the mean rate of separation of the nearest neighbors:

$$\lambda_1 = \frac{1}{i\Delta t}\frac{1}{M-i}\sum_{j=1}^{M-i} \ln \frac{d_j(i)}{d_j(0)} \tag{4}$$

where $\Delta t$ is the sampling period of the time series and $d_j(i)$ is the distance between the $j$-th pare of nearest neighbors after $i$ discrete time steps, i.e., $i.\Delta t$ seconds. Here $M = N-(d-1)\tau$ is the number of reconstructed points, $d$ is the embedding dimension and can be selected using false nearest neighbors [13] and $\tau$ is time delay which can be selected using mutual information [12]. From the definition of $\lambda_1$ given in Eq (3) we assume the $j$-th pair of nearest neighbors diverge approximately at rate given by the maximum Lyapunov exponent.

$$d_j(i) = C_j e^{\lambda_1 (i\Delta t)} \tag{5}$$

where $C_j$ is the initial separation. By taking the logarithm of both side of Eq (5), we obtain

$$\ln d_j(i) = \ln C_j + \lambda_1 (i\Delta t). \tag{6}$$

where ln denotes natural logarithm. The above equation represents a set of approximately parallel lines (for $j = 1, 2, ..., M$), each with a slope roughly proportional to $\lambda_1$. The maximum Lyapunov exponent is easily and accurately calculated using a least-squares fit to the average line defined by

$$y(i) = \frac{1}{\Delta t}\langle \ln d_j(i) \rangle. \tag{7}$$

where $\langle \cdot \rangle$ denotes the statistical average over all values of $j$. We calculate the values of maximum Lyapunov exponent, $\lambda_1$, to identify chaotic behavior of the colour marker signals.

Next, in order to use both dynamic and static features in classification of TMD, support vector machine (SVM) is briefly explained.

## 3. CLASSIFICATION USING SVM

SVMs have been proven to be a successful machine learning tool for variety of classification problems since they were introduced by Vapnik [16]. SVMs have demonstrated good generalization performance in face recognition [17], and they also bring some hopes for the biomedical data classification [18]. The goal of an SVM is to find an optimal separating hyperplane (OSH) for a given feature set. The OSH is found by solving the following constrained optimisation,

$$\min_{\mathbf{Z},\mathbf{b},\gamma_i} \left( \frac{1}{2} \| \mathbf{Z} \|^2 + C \sum_{i=1}^{l} \gamma_i \right) \tag{8}$$

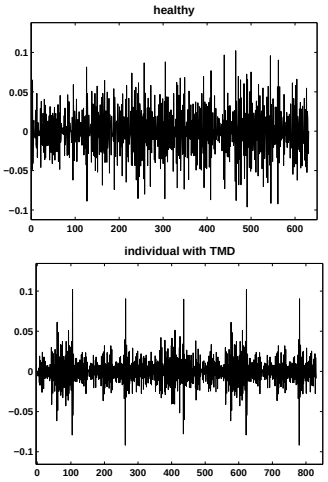


Figure 3: Normalized highpass filtered data of TM marker for healthy (top) and an individual with TMD (bottom).

$$s.t. \quad q_i(\mathbf{Z}.\mathbf{g}_i - b) + \gamma_i \geq 0 \quad (i = 1, \cdots, l)$$

where, $l$ is the number of training vectors and $q_i \in \{\pm 1\}$ are the output targets, $\| \mathbf{Z} \|^2 = \mathbf{Z}^T\mathbf{Z}$ is the squared Euclidean norm and $(.)$ is the dot product. The parameter $\mathbf{Z}$ determines the orientation of the separating hyperplane, $\gamma_i$ is the $i$-th positive slack parameter, and $\mathbf{g}_i$ is a vector containing the features $\mathbf{g}_i = [f_1(i), \ldots, f_n(i)]$, where $f$ is the feature and $n$ is the number of features. The non negative parameter $C$ is the (misclassification) penalty term, and can be considered as the regularization parameter selected by the user.

A larger $C$ is equivalent to assigning a higher penalty to the training errors. The parameter $C$ is set to a value which yields the lowest cross-validation (CV) test error. SVs are the points from the dataset that fall closest to the separating hyperplane. Any vector $\mathbf{g}_i$ that corresponds to a non-zero $\alpha_i$ is a support vector (SV) of the optimal hyperplane. It is desirable to have the number of SVs small to have a more compact and parsimonious classifier. The OSH (generally nonlinear) is then computed as a decision surface of the form

$$f(g) = \text{sgn}\left(\sum_{i=1}^{L_s} q_i \alpha_i K(\mathbf{g}_i^s, \mathbf{g})\right). \tag{9}$$

where $\text{sgn}(.) \in \{\pm 1\}$, $\mathbf{g}_i^s$ are SVs, $K(\mathbf{g}_i^s, \mathbf{g})$ is the nonlinear kernel function (if $K(\mathbf{g}_i^s, \mathbf{g}) = \mathbf{g}_i^s.\mathbf{g}$ the SVM is linear), and $L_s$ is the number of support vectors. A Kernel for a nonlinear SVM projects the samples to a feature space of higher dimension via a nonlinear mapping function.

## 4. FEATURES

The most effective features are measured in this section. 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. Fig. 3 shows the time sequence of a healthy (top figure) and that of an individual with TMD (bottom figure) in the normalized highpass filtered data.

### Feature 1

Detecting the presence of chaos in a dynamical system is an important problem that is solved by measuring the maximum Lyapunov exponents. Here, we use the value of $\lambda_1$ as a feature. This is measured from the normalized data and it is observed that the chewing signal for subjects with TMD is more chaotic than healthy individuals' signals. We denote $f_1 = \lambda_1$. We used the normalized data to calculate the value of $\lambda_1$.

### Feature 2

In many applications, it is important to detect the outliers i.e., unusual abnormal values. In medicine, unusual values may indicate the diseases (see, e.g., [19]). One approach to outlier detection is that we start with $N$ normal values $x_1, \cdots, x_N$, compute the sample average $\bar{x}$ , the sample standard deviation $\sigma$ , and then mark a value $x$ as an outlier if $x$ is outside the interval $(\bar{x} - a\sigma, \bar{x} + a\sigma)$ (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. [20]). Here, we selected $a = 3$ and used the normalized highpass filtered data.
$f_2 = the\ number\ of\ observations > 3\sigma$
or $the\ number\ of\ observations < -3\sigma$.

### Feature 3

This feature is a measure of the likelihood of a peak subject to the gradient of the smoothed waveform. Let $u(t)$ be the lowpass filtered data. Denote $\nabla_t u(t)$ (approximated as $\nabla_t u(t) = u(t) - u(t-1)$ ) and define

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

We denote $f_3 = I_{\nabla_t u(t)}$.

### Feature 4

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

$$f_4 = \frac{\max\{|\boldsymbol{x}(t)|\}}{\sigma_x^2} \quad t = 1, \cdots, N \tag{11}$$

where $\boldsymbol{x}(t)$ is normalized highpass filtered data , max(.) is a scalar valued function that returns the maximum element in a vector, $\sigma$ is the standard deviation of $\boldsymbol{x}(t)$ and $|.|$ is the absolute value applied element-wise.

### Feature 5

This feature corresponds to a third order statistic of the data. The normalized skewness for each signal is given by

$$f_5 = \left| \frac{E\{\boldsymbol{x}^3(t)\}}{\sigma_x^3} \right| \quad t = 1, \cdots, N \tag{12}$$

Hence we take the absolute value of the skewness.

**Feature 6**

Kurtosis is a measure of how sharp a symmetric distribution is when compared to a normal distribution of the same variance. It is defined as:

$$f_6 = \frac{E\{x^4(t)\}}{\sigma_x^4} \quad t = 1, \cdots, N \tag{13}$$

Kurtosis is actually more influenced by scores in the tails of the distribution than scores in the centre of a distribution. Accordingly, it is often appropriate to describe a leptokurtic distribution as "fat in the tails" and a platykurtic distribution as "thin in the tails".

## 5. RESULTS

In our experiments, we used video sequences of eight subjects, two of which had TMD (class 1) and six healthy people (class -1). Three blue markers were placed around the TM bone on the face of each subject. Each subject was filmed performing five cycles of chewing motion using a high resolution (640×480 pixels) video camera at 30fps. On average, 400 video frames were obtained per subject.

The positions of the markers were automatically extracted using image processing techniques and used to obtain the above six features, which were subsequently used in SVM classification. In addition, we used cross-validation to test the accuracy of SVM performance.

Table 1 represents a summary of the results which were obtained. Column 2 of this table shows the individuals in this study. For each individual we divided the data into two trials. The first four rows show individuals with TMD and the rest show healthy individuals. Column 3–8 represents the value of $f_1$–$f_6$.

As it appears from column 3 of Table 1, the value of $\lambda_1$ (is rounded) for each colour marker is positive indicating that they have chaotic behavior. It must also be mentioned that the value of $\lambda_1$ for the individuals with TMD, rows 1–4, is larger than those for healthy subjects, rows 5–16 as we expected.

The results of feature 2 are presented in column 4. As it is shown in this column, the number of peaks for the individuals with TMD is greater than those for healthy subjects, which confirms the significance of this feature for the classification. We can see a similar pattern for the rest of the features. For all features there are significant discrepancies between the values for individuals with TMD, and healthy subjects. Feature 3 is useful to distinguish between the peaks in the signals during the chewing process related to TMD and other non-relevant peaks.

After applying SVM to the above features extracted from the video sequences for eight subjects, two sets per subject, we were able to classify correctly all sixteen video sequences as belonging to patients with TMD or not.

## 6. CONCLUSIONS

An effective method for classification of TMD using video has been proposed here. In our approach, both static and dynamic features are measured from a number of time sequences and classified using SVM. The SVM correctly classified the data for all subjects. A general system for determination of TMD types may be developed by fusing other recording modalities such as sound and electromyogram (EMG) with the above procedure.

Table 1: list of feature value.

| N | Subject | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ |
|---|---|---|---|---|---|---|---|
| 1 | A1 | 0.099 | 8 | 1 | 254.14 | 0.16 | 6.79 |
| 2 | A2 | 0.099 | 8 | 1 | 253.88 | 0.16 | 6.76 |
| 3 | C1 | 0.098 | 6 | 1 | 177.66 | 0.16 | 5.64 |
| 4 | C2 | 0.098 | 6 | 1 | 186.24 | 0.01 | 6.25 |
| 5 | K1 | 0.061 | 0 | 0 | 92.11 | 0.01 | 3.15 |
| 6 | K2 | 0.061 | 0 | 0 | 78.81 | -0.03 | 2.73 |
| 7 | L1 | 0.075 | 0 | 0 | 56.75 | -0.03 | 2.78 |
| 8 | L2 | 0.075 | 0 | 0 | 55.59 | -0.05 | 2.78 |
| 9 | M1 | 0.040 | 0 | 0 | 61.91 | -0.01 | 2.96 |
| 10 | M2 | 0.040 | 0 | 0 | 61.02 | 0.03 | 2.95 |
| 11 | Q1 | 0.073 | 1 | -1 | 96.24 | -0.08 | 3.15 |
| 12 | Q2 | 0.073 | 2 | 1 | 96.31 | 0.14 | 3.31 |
| 13 | S1 | 0.069 | 2 | 1 | 88.46 | 0.07 | 3.12 |
| 14 | S2 | 0.069 | 0 | 0 | 72.36 | 0.02 | 2.89 |
| 15 | Y1 | 0.069 | 0 | 0 | 72.83 | -0.08 | 2.99 |
| 16 | Y2 | 0.069 | 1 | 1 | 80.41 | -0.06 | 3.06 |

## REFERENCES


[1] C. McNeill, "Temporomandibular Disorders; Guidelines for Classification," *Assessment and Management*, Quintessence, Chicago, pp. 1-38, 1993.

[2] J. P. Okeson, "Temporomandibular disorders in the medical practice," *Journal of Family Practice*, **43**, 347, 1996.

[3] S. F. Dworkin, and L. LeResche, "Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, criteria," *J Craniomand Disord Facial Oral Pain*, vol. 6, pp. 301-355, 1992.

[4] R. Ohrbach, and C. G. Widmer, "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, 1992.

[5] C. C. Took, S. Sanei, J. Chambers, and S. Dunne, " Underdetermined Blind Source Separation of Temporomandibular Joint Sounds," *IEEE Trans. on Biomed. Eng*, vol. 53, pp. 2123- 2126, 2006.

[6] R. E. A. Van Emmerik, M. T. Rosenstein, W. J. McDermott and J. Hamill, "Nonlinear Dynamical Approaches to Human Movement," *Journal of Applied Biomechanics*, vol. **20**, 2004.

[7] B. J. West and N. Scafetta, "Nonlinear dynamical model of human gait," it Physical Review E, vol. **67**, no. 051917, 2003.

[8] T. K. M. Lee, R. Surendra and S. Sanei, "Frontal view-based identifiction using the largest lyapunov exponent," to be published in Pattern Recognatation Letters, 2007.

[9] F. Takens, "Detecting strange attractors in turbulence," *Lecture notes in Math*, **898**, 336, 1981.

[10] Y. Sauer, J. A. Yorke and M, Casadgli "Embedology", *Journal of Statistical Physics*, **65**, pp. 579-616, 1991.

[11] E. Ott, T. Sauer, and J. A. Yorke, *Copying with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems*, John Wiley & Sons, Inc., New York (1994).

[12] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, **33**, pp. 1134-1140, 1986.

[13] K. B. Kennel, R. Brown, H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruc-

tion using a geometrical construction," *Phys. Rev.* A, vol. **45**, no. 3403, 1992.

[14] M. T. Rosenstein, J. J. Collins and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D*, vol. **65**, pp.117, 1993.

[15] H. Kantz, "A robust method to estimate the maximal Lyapunov exponent of a time series," *Physics. Letters A*, vol **185**, pp.77, 1994.

[16] V. Vapnik, *The nature of statistical learning theory*, Springer, New York, 1995.

[17] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," *in Proc. Computer Vision and Pattern Recognition* **97**, Puerto Rico, pp. 130136.

[18] V. N. Vapnik, *Statistical Learning Theory*, john Wily, 1998.

[19] O. Kosheleva, S. Cabrera, R. Osegueda, S. Nazarian, D. L. George, M. J. George, V. Kreinovich and K. Worden, "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*, vol. **3459**, pp. 128-135, San Diego, CA , 1998.

[20] H. M. Wadsworth Jr. (ed.): *Handbook of statistical methods for engineers and scientists*, McGraw-Hill, NY. 1990.