ANALYSIS AND CLASSIFICATION OF SWALLOWING SOUNDS USING RECONSTRUCTED PHASE SPACE FEATURES

Mohammad Aboofazeli, Zahra Moussavi
Department of Electrical and Computer Engineering,
University of Manitoba, Winnipeg, Manitoba, Canada
{mohammad, mousavi}@ee.umanitoba.ca

ABSTRACT

This paper presents quantative analysis of swallowing sounds in normal and dysphagic subjects based on nonlinear dynamic metric tools. In addition, an automated method is proposed to identify patients at risk of dysphagia. Multidimensional phase space representation of the swallowing sound was reconstructed using Takens method of delays. Rosenstein and False Nearest Neighbor (FNN) methods were employed to evaluate the optimum time delay and proper embedding dimension, respectively. Grassberger-Procaccia algorithm was utilized to calculate the correlation dimension as a measure of the complexity of reconstructed attractor. The analysis demonstrated the low-dimensional dynamic characteristics of normal and dysphagic swallowing sounds. The optimum time delay and correlation dimension of opening and transmission phases of swallowing sounds were used as features for 3-nearest neighbor classifier to identify individuals at risk of dysphagia. The method was applied to tracheal sound recordings of 15 healthy subjects and 11 patients with some degrees of dysphagia. The algorithm was able to classify 83% of swallows correctly. Finally, a screening algorithm was used which correctly classified 24 subjects of 26 subjects. This study suggests the nonlinear analysis as a promising tool for quantative analysis of swallowing sounds and swallowing disorders.

1. INTRODUCTION

Swallowing disorder (dysphagia) occurs in individuals with different congenital abnormalities, structural damage, and/or medical conditions [1]. A controlled and coordinated swallow does not consistently occur in individuals with severe neurological impairments [2]. Dysphagic patients are at risk of choking, malnutrition, dehydration and breathing difficulty [1].

Currently, the gold standard technique in dysphagia assessment is videofluoroscopy (VFS) test in which the patients are fed barium-mixed boluses and their bolus movement pattern is monitored on the screen. VFS allows examination of swallowing mechanism and detection of aspiration (entering the bolus into airway instead of esophagus) when it occurs. However, VFS is based on X-ray and since the maximum amount of patients' exposure to X-ray must be limited, it can be run only for a short period of time. On the other hand, even in dysphagic patients aspiration may occur only 10% of the time of a feeding assessment. Therefore, it is quite possible to miss aspiration or other disorders during VFS assessment. Hence there is a need to develop non-invasive techniques to assess swallowing mechanism and its abnormalities.

Cervical auscultation (listening at the throat with a stethoscope) is a common routine for clinicians as a non-invasive component of clinical evaluation of swallowing. Cervical auscultation is applicable in a wide variety of feeding circumstances and reveals additional criteria for evaluation of the pharyngeal stage of feeding [3]. As in any observational method, the quality of information obtained by auscultation depends on the perceptual skills of the examiner [4].

In recent years, acoustical analysis of swallowing mechanism has received considerable attention [2-8] as a complement for cervical auscultation. Respiratory and swallowing sounds are recorded by microphones and/or accelerometers and analyzed by digital signal processing techniques.

Recent developments in the theory of nonlinear dynamics have developed some methods for quantative analysis of experimental time series representing signals measured from nonlinear systems. Nonlinear techniques are able to describe more details of the process generated by nonlinear biological systems.

The production of swallowing sound is a highly nonlinear process involving biomechanical and aerodynamic effects. The methods employed in this work are based on the analysis of data obtained from swallowing sound which is a single variable time series generated by swallowing mechanism. Characteristics of swallowing sound using nonlinear dynamic metric tools in normal subjects were recently studied in our lab [5]. It was shown that swallowing sound is well characterized by a small number of dimensions. In addition, the largest Lyapunov exponent was estimated to evaluate the presence of chaos. As the largest Lyapunov exponent for some cases was negative, it was concluded that swallowing sound is not necessarily a chaotic process [5].

An acoustical method for classification of normal and dysphagic swallows was also proposed in [6]. Waveform dimension trajectories (WFD) was used to segment the swallowing sound into characteristics segments and discriminant analysis was used to identify patients at risk of dysphagia. The feature set included WFD, duration of swallows, average power and magnitude of the swallowing sounds. The result of that study was encouraging and therefore we aimed to continue that line of research applying nonlinear analysis.

The goal of this work was to study the characteristics of swallowing sound using nonlinear dynamic metric tools for both normal and dysphagic individuals for dynamic assessment of swallowing mechanism. In addition, these metric tools were used as features for classification of normal and dysphagic swallows by acoustical means.

2. METHOD

2.1. Data

Data were adopted from a previous study [2]. Tracheal sound recordings of two groups of subjects were used in this study. The first group consisted of 12 healthy children (3-16 years) and three healthy adults (ages 35, 38, and 54 years). All subjects of first group were in good health without any history of swallowing disorder, eating or nutrition problems, or lower respiratory tract infection. The second group consisted of 11 young adult patients (ages 16-25 years) with swallowing disorder. It should be noted that all the swallows of this group were in the category of normal swallows as we excluded the swallows in which aspiration had occurred. However, since they belonged to a dysphagic patient, we considered them as marginally normal.

During the test, participants were fed three textures: pre-packaged pudding (semisolid texture), diluted pudding (thick liquid texture) and fruit juice (thin liquid) in bolus size of 5 ml throughout the experiment. For acoustical monitoring, two Siemens accelerometers (EMT25C) were placed (by double-sided adhesive tape rings) over suprasternal notch to record tracheal breath and swallowing sounds and the left or right second intracostal space, in midclavicular line, to record lung sounds. In this study, however, only tracheal sounds were used. The sound signals were amplified, bandpass filtered (30-2500 Hz) and digitized at 10240 Hz.

2.2. Nonlinear Analysis and Feature Extraction

The first step in the analysis of nonlinear dynamical systems is the reconstruction of the attractor [9-13]. It is common to make a reconstruction of the attractor using a single variable time series using the Takens method of delays [9,10]. In this study, swallowing sound is available time series. These data must be transformed into multi-dimensional phase-space plots. Swallowing sound is described by the one-dimensional time series $x(i)$ where $i = 1,2,...,N$ is the time index.

The method of delays reconstructs the attractor by using delay coordinates to form state-space vectors $X(k)$ in a multi-dimensional phase space. The E-dimensional vectors $X(k)$ can be constructed as:

$$X(k) = [x(k), x(k+\tau), x(k+2\tau),..., x(k+(E-1)\tau)] , \quad (1)$$

where $X(k)$ is one point of the trajectory in the phase space at time $k = 1,2,..., N-(E-1)\tau$, τ is an appropriate time lag (an integer multiple of the sampling period) and E is called the embedding dimension and determines the smallest number of independent variables that uniquely describe the character of the system.

The quality of the reconstruction using method of delays depends on the delay parameter τ [10, 11]. Rosenstein *et al.* [11] proposed a geometry based method for choosing best time delay. This method measures the average displacement $< S_E >$ of the embedding vectors from their original locations on the line of identity. $< S_E >$ is evaluated as a function of τ such that:

$$< S_E(\tau) >= \frac{1}{M}\sum_{k=1}^{M} \| X^{\tau}(k) - X^{0}(k) \| , \quad (2)$$

where the superscripts denote the time delay between successive embedding components and M is the number of vectors for the corresponding dimension E . As the lag increases, the average displacement increases accordingly. With larger values of E , reconstruction expansion reaches a plateau at smaller values of τ . We choose the best time lag as the point where the slope of the curve decreases to less than 25% of its initial value.

The purpose of time-delay embedding is to unfold the attractor in a Euclidean space large enough such that all self –crossings of the orbit can be eliminated. The attractor will be unfolded if we use the minimum embedding dimension $E_{\min}$, or any $E > E_{\min}$. In an embedding dimension that is too small to unfold the attractor, points that lie close to each other will not be neighbors because of the dynamics. Some will be far from each other but appear as neighbors because the geometric structure of the attractor has been projected onto a smaller space. On the other hand, working in any dimension larger than the minimum required by the data leads to excessive computation when we evaluate any metric parameters. In this study, the method discussed in [12] was used to find a value for the minimum embedding dimension which is based on increasing dimension, $E = 1,2,3,...$ until no false neighbors remained.

In dimension E , each vector, $X(k)$, has a nearest neighbor, $X^{NN}(k)$. Let $R^2{}_E(k)$ be the distance between the vectors $X(k)$ and $X^{NN}(k)$:

$$R^2{}_E(k) = \| X(k) - X^{NN}(k) \|^2 = \sum_{j=0}^{E-1} [x(k+j\tau) - x^{NN}(k+j\tau)]^2 . \quad (3)$$

In dimension $E+1$, the distance between the vectors $X(k)$ and $X^{NN}(k)$ will be:

$$R^2{}_{E+1}(k) = R^2{}_E(k) + [x(k+E\tau) - x^{NN}(k+E\tau)]^2 . \quad (4)$$

If $R^2{}_{E+1}(k) >> R^2{}_E(k)$, the closeness $X(k)$ and $X^{NN}(k)$ is due to the projection from some higher dimensional attractor, down to dimension E . When we increase from dimension E to dimension $E+1$, we have put away these two points. A threshold T is required to decide when neighbors are false. The neighbors that fulfill:

$$\frac{| x(k+E\tau) - x^{NN}(k+E\tau) |}{R_E(k)} > T , \quad (5)$$

at time point k are considered false.

Several kinds of dimensions have been proposed in order to give a precise description of the complexity of the system or an equivalent way to give lower bound for number of variables required to model the system under study. The correlation dimension is the most widely used measure in literature [10,13].

The Correlation Dimension D_c given in [13] by Grassberger and Procaccia based on determining the relative number of pairs of points in the phase space set that is separated by a distance of less than r. It is computed from:

$$D_c = \lim_{r \to 0} \frac{\log C(r)}{\log r} , \quad (6)$$

where the correlation sum $C(r)$ is:

$$C(r) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Theta(r - | X_i - X_j |) , \quad (7)$$

where X_i, X_j are the points of the trajectory in the phase space, N is the number of data points in the phase space , the distance r is a radius around each reference point X_i, and Θ is the Heavyside function which excludes values outside of radius.

2.3. Classification

A 3-nearest neighbor classifier and a two-layer feed-forward neural network were created to classify normal and dysphagic swallows. The optimum time delay and correlation dimension of the opening and transmission sounds (4 features in total) were used as features of these classifiers.

We began by removing the swallows of one subject from a set of labeled swallows of all subjects. 3-nearest neighbor rule classified each swallow of the subject whose swallows had been removed by assigning it the label most frequently represented among 3 nearest samples of other subjects' swallows. In other words a decision is made by examining the labels on the 3 nearest neighbors and taking a vote. 3-nearest neighbor classifier was used 26 (i.e. the number of subjects) separate times, and each time all swallows of one subject was removed.

Since the number of subjects was not large enough for randomly dividing the recordings to training and test data sets, jackknife approach was utilized for training and testing of the neural network classifier, in which the accuracy is estimated by training the classifier 26 (i.e. the number of subjects) separate times, and each time testing it on the left out data. The jackknife estimates of the classification were averaged between the subjects.

3. RESULTS AND DISCUSSION

In this study, swallowing sound was decomposed into 2 major sections and each section was individually studied. These two sections were called Opening and Transmission. The Opening section was from the beginning of swallow to the end of Initial Discrete Sound (IDS) as defined in [3]. The Transmission section was from the sample point after IDS termination to swallow termination.

A typical curve of percentage of false nearest neighbors for different values of embedding dimension for thin liquid texture and opening part of swallow of a dysphagic subject is shown in Fig. 1. For embedding dimension $E = 8$, the false nearest neighbor percentage is less than 1%. Therefore, we may conclude the optimum value for the minimum embedding dimension, $E_{\min}$, based on false nearest neighbor method is equal to 8.

The average displacement of the embedding vectors from their original locations on the line of identity as a function of time delay $< S_E(\tau) >$, for embedding dimensions $E = 8$ was calculated for each texture and for each of the opening and transmission sections. The best time lag was chosen as the point, where the slope of the curve decreased to less than 25% of its initial value. The optimum values of τ depended on subject, bolus textures and swallowing sound section and are shown in table 1.

Using Grassberger-Procaccia algorithm, correlation dimension of swallowing sounds were evaluated for different subjects, textures and sections. Correlation dimension do not appear to change appreciably among the subjects. Correlation dimension were calculated for embedding dimension 8 and the corresponding optimum time delay. The results are shown in table 1.

Classification of different swallows was performed as explained in section 2.3. The 3 nearest neighbor classifier was found to be superior to the feed-forward neural network. Their accuracy was 83% and 71%, respectively. As a second stage classification, a screening algorithm was used in which if more than 50% of the swallows of a subject were classified as being normal, the subject was considered as normal. Otherwise, the subject was considered as at risk of dysphagia. This algorithm classified 24 subjects of 26 subjects, correctly.

Classification accuracy obtained in this study is comparable to the results of [6]. Among 24 different features used in [6], two features were stated to be more important than the others: waveform fractal dimension and time duration of the opening and transmission sections of swallow. In this study, we added time durations to our features. But classification accuracy became less for most subjects mainly due to the high variability of this feature within the swallows of a subject. However, the waveform fractal dimension, if added to features used in this study will likely increase the accuracy.

Overall, the results of this study are very encouraging as the proposed method is able to classify normal and marginally normal swallows from each other with a high accuracy. Although physical interpretations of nonlinear metric tools are usually difficult, however based on the results of this study, they seem to reveal some hidden characteristics of a very complex mechanism, i.e. swallowing process. This study suggests the nonlinear analysis as a promising tool for quantitative assessment of swallowing sounds and confirms discriminatory capability of the reconstructed phase space features for swallowing disorders.

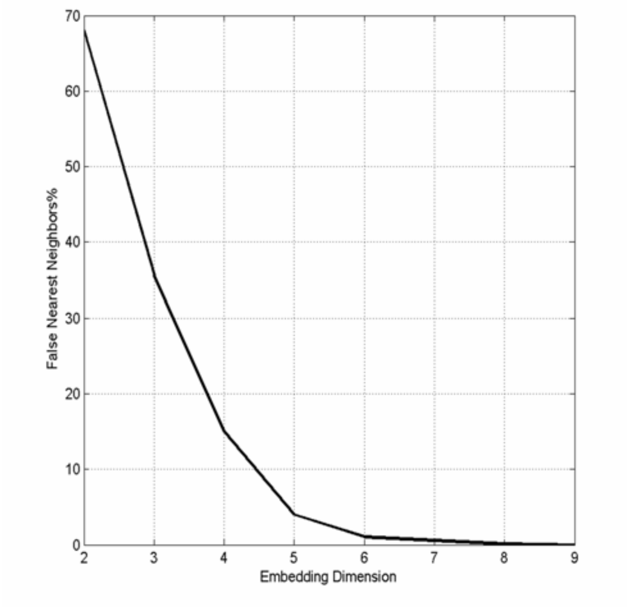

Fig. 1. False nearest neighbors as a function of embedding dimension

4. ACKNOWLEDGEMENT

This research was supported by Telecommunication Research Laboratories (TRLabs) of Canada.

5. REFERENCES

[1] J.A. Logemann, *Evaluation and Treatment of Swallowing Disorders*. Austin, Texas: proed, 1998.

[2] G. Rempel, H. Pasterkamp, and Z. Moussavi, "The effect of altering the consistency of food on the breath and swallow pattern of young people with cerebral palsy," presented at the 53rd Annual congress of American Academy for cerebral palsy and developmental Medicine, Sept. 1999.

[3] F. L. Vice, J. M. Heinz, G. Giuriati, M. Hood, and J.F. Bosma, "Cervical auscultation of suckle feeding in newborn infants," *Developmental Medicine and Child Neurology,* vol. 32, pp.760-768, 1990.

[4] S. Hamlet, D. Penney, and J. Formolo, "Stethoscope acoustics and cervical auscultation of swallowing," *Dysphagia*, vol. 9, pp. 63-68, 1994

[5] M. Aboofazeli, and Z. Moussavi; "Analysis of Normal Swallowing Sounds Using Nonlinear Dynamic Metric Tools," in *Proc. of 26th Annual International Conference of the IEEE Eng. in Med. and Biol. Society (EMBS 2004)*, pp. 3812-3815, September 2004

[6] L. Lazareck and Z. Moussavi, "Classification of Normal and Dysphagic Swallows by Acoustical Means," *IEEE Trans. Biomed. Eng.,* in press

[7] I.H. Gewolb, J.F. Bosma, V.L. Taciak, F.L. Vice, "Abnormal Developmental Patterns of Suck and Swallow Rhythms During Feeding in Preterm Infants with Bronchopulmonary Dysplasia," *Developmental Medicine and Child Neurology,* vol. 43, no. 7, pp. 454-459, July 2001.

[8] W.G. Selley, F.C. Flack, and R.E. Ellis, "The Exeter dysphagia assessment technique," *Dysphagia,* vol. 4, pp. 227-235, 1990.

[9] F. Takens; Detecting strange attractors in turbulence, in: D.A. Rand and L.S. Young, (Eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Springer, vol. 898, pp. 366 - 381, 1981

[10] R. C. Hilborn, Chaos and nonlinear dynamics, 2nd edition, oxford university press, 2000

[11] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D*, vol. 73, pp. 82-98, 1994

[12] H.D.I. Ababarnel, R. Brown, J. L. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, Vol. 65, No. 4, pp. 1331-1392, 1993

[13] P. Grassberger, and I. Procaccia, "Measuring the Strangeness of Strange Attractors," *Physica D*, vol. 9, pp. 189-208, 1983

[14] M. T. Rosenestein, J.J. Collins, and C.J. De Luca, "A practical method for calculating largest Lyapunov exponents," *Physica D*, vol. 65, pp. 117-134, 1993

Table 1. Optimum time delay range and Correlation dimension (mean ± std) for opening and transmission phases in normal and dysphagic subjects.

		Normal			Dysphagic		
		Thin Liquid	Thick liquid	Semisolid	Thin Liquid	Thick liquid	Semisolid
Optimum Time Delay	Opening Phase	3-9	3-9	3-9	6-9	5-8	6-8
	Transmission Phase	4-10	3-10	4-10	7-11	6-10	6-9
Correlation Dimension	Opening Phase	3.09±0.58	2.74±0.52	3.05±0.52	2.22±0.24	2.19±0.24	2.37±0.32
	Transmission Phase	3.20±0.51	2.90±0.54	3.13±0.47	2.30±0.18	2.21±0.26	2.25±0.35