33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Investigating the Statistical Properties of the Swallowing Sounds

S. Sarraf Shirazi, Student *member, IEEE,* and Z. Moussavi, Senior *member, IEEE*

Abstract— **In this paper the statistical properties of the swallowing sound is discussed. This knowledge is required for the acoustical modeling of the swallowing mechanism as it is important to select an appropriate type of the system (i.e. linear vs. nonlinear) for modeling. The tests of linearity and gaussianity were performed. The results of the statistical test of gaussianity showed a nongaussian distribution of the swallowing sound signals. Also, the test of linearity exhibited the nonlinear characteristics of the model that represents the swallowing sound generation.**

I. INTRODUCTION

SWALLOWING is one of the most complex mechanisms in human body. Any impairment such as lack of coordination of swallowing events or delays in initiating swallows may result in swallowing disorders (dysphagia). In recent years, swallowing sound analysis has received considerable attention as a non-invasive tool to study the characteristics of swallowing mechanism in healthy and dysphagic individuals [2]-[4]. This study attempts to derive the statistical properties of the swallowing sound. This will help in modeling the swallowing sound generation by providing a priori information about the acoustical signature of the swallowing.

There are two techniques currently used to evaluate the swallowing mechanism: videofluoroscopy (VFS), fiber-optic endoscopy (FEES); each of them provides information of the swallowing process related to the anatomic structure it visualizes. VFS study is considered the gold standard technique for swallowing evaluation. Also, FEES has been reported to be a valid tool for detecting swallowing disorders such as aspiration, penetration and pharyngeal residue [5]. However, both methods are invasive.

Acoustical analysis of swallowing comes into play as a need for a non-invasive method of swallowing evaluation with minimal interference on the normal eating procedure. Swallowing sound analysis has received considerable attention as a convenient and accurate method for swallowing assessment [2], [6]-[9]. Swallowing sound is described by two distinct segments [10]: initial discrete sound (IDS) and bolus transit sound (BTS). IDS is associated with the opening of the upper esophageal sphincter that occurs during the pharyngeal phase. BTS is the gurgle sound generated as the bolus passes through the esophagus with peristaltic contraction during the esophageal phase.

Recently, the acoustical studies of swallowing mechanism mainly focused on the classification of the control and dysphagic groups [6], [11], [12]. There have been few attempts toward modeling the swallowing sound generation and transmission [13], [14], assuming that the swallowing sound is produced by exciting the pharyngeal wall tissue with train of impulses coming through the pharynx. Much of what is known about physiological systems such as swallowing mechanism has been learned using linear and time-invariant (LTI) system theory. The main advantage of linear system analysis is the availability of analytical tools to deal with the modeling. However, many real systems have complex properties that cannot be studied by restricting them to linear techniques.

The input of the system is an important issue in modeling as it is neither known nor accessible for most physiological systems, such as swallowing. In such cases, it is convenient to make some assumption of the input signal, i.e. a random Gaussian noise signal. However, if the model is considered as an LTI, then it is not correct to assume a white Gaussian noise input while the output is not Gaussian. Hence, the statistical characteristics of the signal (the output of the system) should be studied before making any assumption of the input distribution and the type of the system. This paper investigates the gaussianity and linearity of the swallowing sound signal, and whether it is different between the two groups of control and dysphagic individuals.

II. METHOD

A. Data

Data in this study included swallowing sound recordings of 10 dysphagic (stroke and/or head trauma patients) and 10 agematched individuals without any swallowing disorder as the control subjects. The swallowing sounds were recorded by a Sony (ECM-88B) microphone placed over the suprasternal notch of the trachea and digitized at 44 kHz. The swallowing sounds of the dysphagic group were recorded simultaneously with the VFS or the FEES assessment at Health Sciences Centre, and Riverview Health Centre, Winnipeg, Canada. The experimental protocol was the same for all data recordings: Subjects were fed 5-8 boluses of a thin liquid texture (i.e. juice) with a 5ml spoon. The study was approved by the Biomedical Ethics Board of the University of Manitoba, and participants signed a written consent prior to experiments.

All signals were normalized to their maximum amplitude. IDS and BTS segments of the swallowing sound signals were separated manually by an expert by auditory and visual

S. Sarraf Shirazi is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada (email:umsarraf@ee.umanitoba.ca).

Z. Moussavi is with the Department Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada (e-mail: mousavi@ee.umanitoba.ca).

inspection of signals in the time and frequency domain. The IDS part of the swallowing sound was considered in this study since it was assumed to be a stationary process.

B. Statistical Analysis

According to Wold decomposition theory, any weak or wide-sense stationary process {x(t)} with innovations {ε(t)} has a moving average (MA) representation as (1) [15]:

$$x(t) = \sum_{k=1}^{\infty} h(k)\varepsilon(t-k), \quad (1)$$

where $\varepsilon(t)$ are independent identically distributed random variables with $E\{\varepsilon(t)\} = 0$ and $h(0) = 1$, $\sum_{k=1}^{\infty}|h(k)|^2 < \infty$.

In other words, a linear model can approximate $\{x(t)\}$ because the innovations are independent. In this case, any input with the Gaussian distribution will result in a Gaussian output. Thus, the 2nd order statistics, i.e. correlation and spectral analyses, can readily determine the statistical properties of the process and be used for system identification. However, if the innovations are not normal and $E\{\varepsilon^3(t)\} \neq 0$ or the signal comes from a nonlinear system, then higher order statistics become important. The 3rd order cumulant of a zero mean stationary process is defined as:

$$c_{xxx}(m,n) = E\{x^*(t)x(t+m)x(t+n)\} \quad (2)$$

which is nonzero since $E\{\varepsilon^3(t)\} \neq 0$. The cumulants of a Gaussian signal are zero at the orders higher than 2. Therefore, the Gaussianity of the process should be determined before making any assumption of the type of the system (i.e. linearity). It should be noted that the signal is assumed to be zero-mean in all the analyses.

Similar to the definition of the power spectrum as the Fourier transform of the second cumulant (the autocorrelation), the Bispectrum is defined as the two-dimensional Fourier transform of the 3rd order comulant (3). This is the indirect method for calculating the bispectrum:

$$B(\omega_1,\omega_2) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} c_{xxx}(m,n)e^{-j(\omega_1 m+\omega_2 n)} \quad (3)$$

This function is periodic in both ω_1 and ω_2 with period 2π. The symmetry properties of the bispectrum of a real signal can be expressed as:

$$B(\omega_1,\omega_2) = B(\omega_2,\omega_1) = B(-\omega_1-\omega_2,\omega_2) = B(\omega_1,-\omega_1-\omega_2) \quad (4)$$

Therefore, $B(\omega_1,\omega_2)$ can be determined in terms of the values inside a triangle whose vertices are located at (0,0), (1/2,0), (1/3,1/3).

The indirect method is not a consistent estimate of the bispectrum [15]. Also, this method is time consuming and has computational burden. This motivates the use of frequency domain analysis known as the direct estimation of the bispectrum assuming the linear representation shown in (1). The direct estimate of the bispectrum is calculated based on the Fourier transform of the signal ($X(\omega_j)$) is:

$$\hat{B}(j,k) = 1/N X(\omega_j)X(\omega_k)X^*(\omega_{j+k}), \quad (5)$$

where N is the length of the signal, and $\omega_j = 2\pi j/N$ for $j = 0,1,\ldots,N$. Since $X(\omega_j)$ is periodic with N, the values of $\hat{B}(j,k)$ will be computed over the region inside a triangle formed by three lines as: $k = 0$, $j - k = 0$, and $2j + k = N$. It was shown that this estimation is not consistent [16]. Thus, the concept of averaging such as what is done in the power spectrum estimation methods is applied. The simplest approach is to average the values of $\hat{B}(j,k)$ over a 2D window of size $M \times M$. All the points must be located in the triangular region introduced above. It is shown that to have an asymptotically unbiased estimate, M should be an increasing function of N and satisfy the criteria in (6) [15].

$$lim_{N\to+\infty}\frac{M(N)}{N} = 0, \text{ and } lim_{N\to+\infty}\frac{M^2(N)}{N} = +\infty \quad (6)$$

Thus, M(N) is proposed to be equal to N^c , where $0.5 < c < 1$. The area inside each window, over which the averaging is performed, consists of M^2 points centered at $\frac{(2j-1)M}{2}, (2k-1)M/2$, where $j = 1,\ldots,k$ and $k \leq \frac{N}{2M} - \frac{N}{2}$. Therefore, the new estimator can be obtained as:

$$\hat{B}_{av}(j,k) = 1/M^2 \sum_{p=(j-1)M}^{jM-1}\sum_{q=(k-1)M}^{kM-1}\hat{B}(p,q) \quad (7)$$

$\hat{B}_{av}(j,k)$ is an asymptotically unbiased, consistent estimate [16].

The approximate asymptotic distribution of the 2nd order spectra estimates was shown to be independent complex normal variables [17]. Therefore, a complex normal distribution with unit variance can represent the distribution of $Z(j,k) = \hat{B}_{av}(j,k)/\sqrt{var\{\hat{B}_{av}(j,k)\}}$.

The two random variables $Re(\boldsymbol{Z})$ and $Im(\boldsymbol{Z})$ are independent and normally distributed ($N(\mu_i,1)$). Thus, the variable $|\boldsymbol{Z}(j,k)|^2 = Re(\boldsymbol{Z})^2 + Im(\boldsymbol{Z})^2$ is distributed according to the chi-square distribution denoted by $\chi_2^2(\lambda)$ with two degrees of freedom [18] and the non-centrality parameter $\lambda(j,k) = \mu_1^2 + \mu_2^2$ [18] that can be written as [19]:

$$\lambda(j,k) = |B(j,k)|^2 \, (NM^{-4}Q)^{-1} S_{xx}^{-1}\left(\frac{(2j-1)f}{2}\right) S_{xx}^{-1}\left(\frac{(2k-1)f}{2}\right) S_{xx}^{-1}\left(\frac{(2j-1)f}{2} + \frac{(2k-1)f}{2}\right) \quad (8)$$

Therefore, the statistic of $Y(j,k) = \sum_{j,k} 2|\boldsymbol{Z}(j,k)|^2$ has a chi-square distribution with 2n degrees of freedom with the non-centrality parameter as the summation of (j,k).

1) Test of Gaussianity

The test for gaussianity is designed based on the statistics introduced above. In case of a Gaussian signal, $B(j,k)$ is equal to zero, which results in the central chi-square distribution as the non-centrality parameter becomes zero. To investigate the guassianity of a signal statistically, we need two null and alternative hypotheses. The null hypothesis (H_0) is considered such that Y has approximately chi-square distribution $\chi_{2n}^2(0)$, and the alternative hypothesis (H_1) would be that Y has the noncentral chi-square distribution. In all instances a p value less than 0.05 is considered as the significance level. Rejecting the null hypothesis is equivalent to the rejection of the gaussianity assumption. If the signal is nongaussian, then the linearity test can be performed. Otherwise, the test of gaussianity does not convey any information about the linearity of the system.

2) Test of linearity

Given that a stationary signal has a linear representative as mentioned in Eq. 1, then the the spectrum and bispectrum are obtained according to (10), (11), respectively:

$$S_{xx}(\omega) = \sigma_\varepsilon^2 |H(\omega)|^2 \quad (10)$$

$$B(\omega_i, \omega_j) = \mu_3 H(\omega_i) H(\omega_j) H^*(\omega_i + \omega_j) \quad (11)$$

where $E\{\varepsilon(t)\} = 0, \sigma_\varepsilon^2 = E\{\varepsilon^2(t)\},\ \mu_\varepsilon = E\{\varepsilon(t)\}$ and $H(\omega)$ is the Fourier transform of the filter coefficients. If $\mu_\varepsilon \neq 0$, then the linear process is non-Gaussian.

The linearity is investigated by analyzing the characteristics of the noncentrality parameter of $\chi_2^2(\lambda)$ denoted by $\lambda(j,k)$ which equals to λ_0 (12) if the signal is generated by a linear process.

$$\lambda(j,k) = \lambda_0 = C\mu_3^2/\sigma_\varepsilon^6, \quad (12)$$

where $C = NM^{-4}Q$.

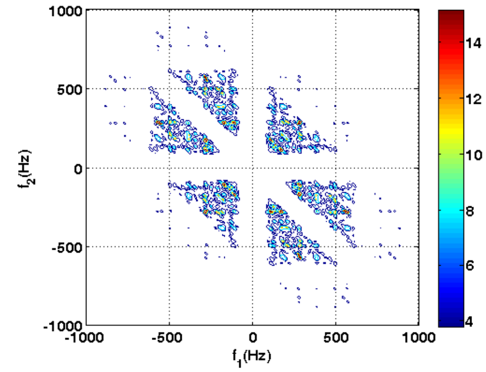
As mentioned before, the statistic $2|\boldsymbol{Z}(j,k)|^2$ has a chi-square distribution ($\chi_2^2(\lambda)$). It is assumed that $\lambda(j,k)$s are generated by a random variable that has a degenerate distribution equivalent to $\chi_2^2(\lambda_0)$ when $\lambda(j,k) = \lambda_0$. The statistical test is performed based on the distribution of the random variable that generates λ. The null hypothesis assumes a linear process generates the signal, which results in a chi-square distribution with a constant noncentrality parameter $\lambda_0 = C\mu_3^2/\sigma_\varepsilon^6$. The alternative hypothesis assumes that a nonlinear system generates the signal. Hence, the random variable generating λ has a non-degenerate chi-square distribution.

The noncentral chi-square distribution of $\chi_P^2(\lambda)$ can be expressed as the Poisson-weighted mixture of central chi-square distributions [18] as indicated in (13). Therefore, $\chi_P^2(\lambda)$ has a thicker tail under the null hypothesis than under the alternative hypothesis. This can be measured by calculating the interquartile range of the distribution. The result of statistical test is obtained based on the comparison between the interquartile ranges of $\chi_P^2(\lambda)$ under each of the two hypotheses. The null hypothesis is rejected if the interquartile value of $\chi_P^2(\lambda)$ is greater than the interquartile value of $\chi_P^2(\lambda_0)$.

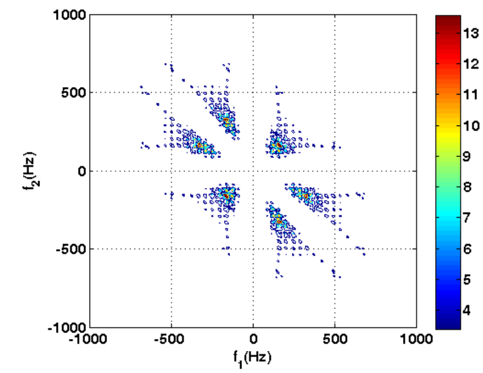
C. Results

The bispectra of typical swallowing sounds of a control and a dysphagic subject are depicted in Fig. 1. The amplitude of the counter plot of these samples of the bispectrum confirms the nongaussian characteristics of the swallowing sound signal. The swallowing sound signals were tested for the nonlinearity. Interestingly, all the statistical significance *p* values, were zero. Thus, the null hypothesis of gaussianity was rejected.

Next, we performed the linearity test since the signals were categorized as nongaussains. The results show significant differences between the theoretically calculated value of the interquartile range and the estimated one for every swallow of each individual. The mean and standard error, averaged among the swallows in each group of data, are shown in Table I. Thus, it can be concluded that swallowing sound is generated by a nonlinear system, and should be analyzed by the nonlinear techniques. Moreover, no trend was found to be characterizing each group (Fig. 2). Also, the results of both tests didn't exhibit any obvious difference between the two groups of data.



(a)



(b)

Fig. 1. The bispectrum of the swallowing sound of (a) a control and (b) a dysphagic subject.

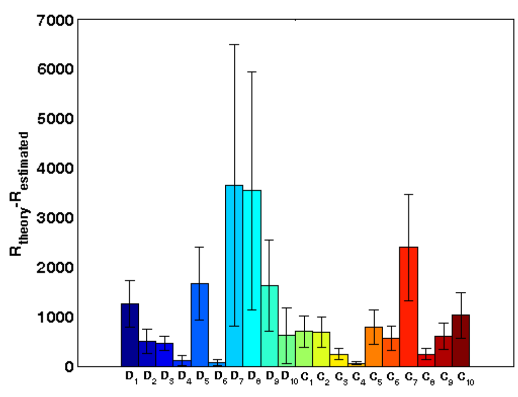


Fig. 2. The test of linearity results: the difference between the estimated and the theoretical interquartile range for all the 10 control and 10 dysphagic data (averaged over all the swallows of each individual).

D. Discussion

The statistical analysis of the swallowing sound reveals the nonlinear properties of this signal. Recently, the nonlinear techniques were used to automatically detect the swallowing sound segments from the breath sounds, as well as classifying normal and dysphagic swallowing sounds

[20]. For example, nonlinear dynamic analysis, recurrence quantification analysis (RQA), (HMM), and multiresolution wavelet analysis were among the nonlinear techniques applied to detect characteristic features of swallowing sounds [20]. However, none of the few studies done on the swallowing sound modeling assumed nonlinear models for the swallowing sound generation.

Table I. The difference between the theoretical and estimated values of the interquartile range averaged for each group.

Data	R_{theory}-$R_{estimated}$
Control Group	730.88 ± 21.59
Dysphagic Group	1408 ± 66.24

In the model that was suggested in [13] and further studied in [14], swallowing sound was assumed to be produced by exciting the pharyngeal wall structure and tissue with an impulse train coming from the pharynx. The swallowing sound was thought to be the output of an LTI system representing the pharyngeal muscle and tissue responses to the neural activities that trigger the swallow, and are represented by an impulse train. That model benefits from the simple known relationship between the generated sounds during the swallow, and the physiological events occurred as the result of the neural excitation. These were the first attempts toward modeling the swallowing mechanism.

Although the linear model can shed light on some aspect of the swallowing sound generation model, it may lead to an oversimplification of the actual system dynamics. This study investigated the validity of the linear assumption. Based on the statistical properties defined for the signals generated by nonlinear systems, two statistical tests were designed: test of gaussianity and test of nonlinearity. Following the method introduced in [16], the test of nonlinearity can be performed if the test of gaussianity doesn't hold. The results confirmed that all the swallowing sounds (in either group of control or dysphagic) demonstrated the nongaussian behavior. Hence, the sounds were investigated for nonlinearity properties.

$$P(\chi_P^2(\lambda) < x) = \sum_{m=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^m}{m!} P(\chi_{P+2m}^2(0) < x) \quad (13)$$

The statistic test based on which the linearity is determined was originated from the test statistics defined for the gaussianity. Both tests deal with the properties of the chi-square distribution of the estimated bispectrum of the signal. The criterion for the gaussianity test uses the level of confidence, which is the error in accepting the central chi-square distribution for an observed data while the true distribution is a noncentral one. The linearity hypothesis is approved if there is not a significant difference between the theoretical and the estimated values of the interquartile range of the distribution. The results showed the nongaussain and nonlinear characteristics of the swallowing sound of both groups of subjects.

The bispectrum plots of both groups showed more and less the same pattern. However, further studies may find some characteristic differences between the bispectra of the two groups as this was beyond the scope of this study. The outcome of this study would be considered in future studies on modeling the swallowing mechanism.

ACKNOLEDGMENT

This study was supported financially by Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors would like to acknowledge A. Meakin, C. Buchel, R. Daun, and L. Lentile at Riverview Health Center and Dr. B. Hassouneh and I. Mike at Health Science Center, for their help in recruiting patients and collecting data.

REFERENCES

[1] B. Jones, Normal and Abnormal Swallowing*: Imaging in Diagnosis and Therapy.* Springer, New York, 2002.
[2] G. Rempel and Z. Moussavi, "The Effect of Viscosity on the Breath Swallow Pattern of Young People with Cerebral Palsy," *Dysphagia*, vol. 20, no. 2, pp. 108-112, 2005.
[3] S. L. Hamlet, R. L. Patterson, S. M. Fleming, and L. A. Jones, "Sounds of Swallowing Following Total Laryngectomy," *Dysphagia*, vol. 7, no. 3, pp. 160-165, 2006.
[4] J. A. Logeman, *Evaluation and Treatment of Swallowing Disorders*, Austin, Tx, Pro-Ed, 1998.
[5] S. E. Langmore, *Endoscopic Evaluation and Treatment of Swallowing Disorders*. Thieme, New York, 2000.
[6] L. Lazareck and Z. Moussavi, "Classification of Normal and Dysphagic Swallowing Sounds by Acoustical Means", *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2013-2112, 2004.
[7] M. Aboofazeli and Z. Moussavi, "Analysis of Temporal Pattern of Swallowing Mechanism," *Proc. IEEE EMBS*, pp. 5592-5595, Sept. 2006.
[8] R. C. Mackowiak, H. S. Brenman, and M. H. F. Friedman, "Acoustic Profile of Deglutition*," Proc. Soc Exp Biol Med.*, vol. 125, pp. 1149-52, 1967.
[9] J. A. Y. Cichero, B. E. Murdoch, "Acoustic Signature of the Normal Swallow: Charecterization by Age, Gender, and Bolus Volume," *Ann. Otol. Rhinol. Laryngol*, vol. 111, pp. 623-632, 2002.
[10] F. L. Vice, J. M. Heinz, G. Giuriati, M. Hood, J. F. Bosma, "Cervical Auscultation of Suckle Feeding in Newborn Infants," *Dev Med Child Neurol.*, vol. 32, no. 9, pp. 760-768, 1990.
[11] M. Aboofazeli and Z. Moussavi, "Analysis and Classification of Swallowing Sounds Using Reconstructed Phase Space Features," *Proc. IEEE ICASSP*, pp. 421-424, March 2005.
[12] A. Yadollahi and Z. Moussavi, "Feature Selection for Swallowing Sounds Classification," *29th IEEE EMBS Conf.*, pp. 3172 – 3175, Aug. 2007.
[13] A. Yadollahi and Z. Moussavi "A Model for Normal Swallowing Sounds Generation Based on Wavelet Analysis," *Canadian Conference on Electrical and Computer Engineering, CCECE*, pp. 827–830, 2008.
[14] S. Sarraf and Z. Moussavi "Acoustical Modeling of Swallowing Mechanism," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 81-87, 2011.
[15] J. M. Le Caillec, and R. Garello, "Asymptotic Bias and Variance of Conventional Bispectrum Estimates for 2-D Signals," *Multidimensional Systems and Signal Processing*, vol. 16, pp. 49–84, 2005.
[16] M. J. Hinich, "Testing for Gaussianity and Linearity of a Stationary Time Series," *Journal of Time Series Analysis*, vol. 3, no. 3, pp. 169-176, 1982.
[17] J. K. Patel, and C. B. Read, *Handbook of the Normal Distribution, Marcel Dekker Inc*., New York and Basel, 1982.
[18] N. L. Johnson, and S. Kotz, *Distributions in Statistics: Continuous Univariate Distribution*, vol. 2, New York: Wiley, 1970.
[19] D. R. Brillinger, and P. R. Krishnaiah, *Handbook of Statistics 3: Time Series in the Frequency Domain*, North-Holland, 1983.
[20] M. Aboofazeli, "Analysis and Modeling of Swallowing Sounds," Ph.D. dissertation, Dept. Elect. and Comput. Eng. Univ. of Manitoba, Winnipeg, MB, 2006.