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

Nonlinear dynamics of voices in esophageal phonation

Nan Yan, Manwa L. Ng, Dongning Wang, Victor Chan and Lan Zhang

Abstract - The present study investigated the difference in voice perturbation measures and parameters obtained from nonlinear dynamic analysis between normal laryngeal phonation and standard esophageal (SE) phonation. Jitter, shimmer, correlation dimension and Kolmogorov entropy were measured from 10 SE and 10 normal male speakers of Cantonese. Jitter and shimmer values were significantly higher for SE than laryngeal voice. But jitter values were found to be significantly different when length of sound samples was altered. In addition, both correlation dimension and Kolmogorov entropy values were significantly higher for SE than laryngeal voice and sample length did not appear to affect the result. These results suggest that SE voices are more chaotic than laryngeal voice. It follows that the technique of nonlinear dynamic analysis may be more reliable and stable for evaluating the acoustic characteristics of SE voices.

I. Introduction

Esophageal [1] speech is an alternative phonation method adopted by laryngeal cancer survivors after total laryngectomy, which is a surgical procedure of removing a pathological larynx (voice box). With the loss laryngeal structures after the surgery, a new voice source, the pharyngoesophageal [2] segment, is adapted by SE speakers for phonation. However, due to the complexity of the PE segment, and the partial loss of control over the new vibratory structure, SE voice appears to be awkward, atypical and rough, compared to other types of pathological voices.

Evidenced from previous studies using biomechanical simulation or acoustical characterization, laryngeal voices exhibit intrinsic characteristics of nonlinear dynamic systems. Titze firstly introduced the technique of using nonlinear dynamics in studying various disordered voices [2]. According to Titze, there are three types of voice signals: (1) nearly-periodic signals, (2) signals with strong subharmonics or modulations, and (3) aperiodic signals. It has been found that both normal and pathological laryngeal voices usually fall into the first two categories, respectively. However, there have been very few nonlinear dynamics studies on SE voices, and it is not known which category SE voice signal falls within.

Compared to biomechanical modeling, acoustic analyses using nonlinear dynamics techniques are simpler and have already been successfully applied to other pathological voices. Zhang et al. carried out a series of studies on various vocal pathologies such as vocal polyps and unilateral laryngeal paralysis [1, 3]. The authors suggested that traditional perturbation measures should be combined with nonlinear dynamics techniques to provide more efficient descriptions of pathological voices for clinical diagnosis. Jiang compared the normal and irregular phonations produced using an excised larynx, and found that nonlinear dynamics techniques yielded significant differences between normal and irregular phonations, while the traditional perturbation methods did not[4]. Robb studied the chaotic phenomena of cries of full-term and preterm infants and confirmed the existence of nonlinear phenomena such as bifurcation in the infant cries [5].There is a study on analyses of nonlinear characteristics of esophageal phonation[6], but this study focused on the Mandarin language and didn't consider reflection on the length of analytic data. The current study focused on Cantonese alaryngeal speech, which is different from Mandarin such as the level of lexical tones. It will help us better understand the nonlinear characteristics of alaryngeal speech of many lexical tone levels.

In the present study, several analytic techniques that are based on nonlinear dynamics will be applied to SE voice signals produced by alaryngeal patients. The techniques included correlation dimension, and Kolmogrov entropy. Perturbation measures were also obtained from these voice signals and used to compare with nonlinear dynamic measures. The aim of the study is to verify the existence of nonlinear phenomena in SE voice comparing to the laryngeal voice, and to estimate the stability and reliability of nonlinear dynamic measures in extremely aperiodic voice signals. Cantonese vowels produced by SE and laryngeal speakers were statistically compared. In addition, the applicability of the nonlinear dynamics techniques was compared to traditional perturbation measures such as jitter and shimmer.

II. METHODS

A. Speakers

Ten standard esophageal (SE) and 10 normal Cantonese male speakers participated in the present study. The SE participants were all superior speakers carefully selected from the New Voice Club of Hong Kong by two practicing speech therapists who had extensive experience in working with laryngectomees (with over 15 years of experience). Since a standardized

Nan Yan is with Divisions of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong (phone: 852-28571504; e-mail: nyan@hku.hk).

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

assessment battery for Cantonese alaryngeal speech is not available, a tailor-made screening test was used to assess the different aspects of speech including consonant, vowel, and tone production in different forms of alaryngeal phonation [7]. Upon listening to the continuous speech samples, the two speech therapists rated the SE speech performance based on the following five aspects of speech using a rating scale of 1-7, with a "7" representing the best production, and a "1" referring to the worst production: (1) voice quality, (2) articulation proficiency, (3) quietness of speech, (4) pitch variation, and (5) overall speech intelligibility [7]. In the study, only those who consistently received a total score of 25 or higher (maximum = 35) were considered as superior speakers.

Upon completion of screening, 10 superior SE speakers and 10 age-matched normal Cantonese speakers were recruited for the experiment. The SE and laryngeal speakers were of ages ranging from 60 to 73 years (mean = 66.3 years) and from 55 to 83 years (mean = 66.0 years). All speakers were physically healthy, who had no reported history of respiratory, speech, language, and/or hearing problems, except that associated with laryngectomy for SE speakers.

B. Speech tasks and Recording procedure

The speech task included vowel prolongation. During the experiment, the participants were instructed to sustain the vowel /i/ at high level tone three times for as long as they could. With no attempt to control for loudness, the laryngectomees were instructed to produce the speech samples at a comfortable level of loudness.

In order to familiarize themselves with the speech tasks and the recording environment, the speakers were instructed to practice the speech tasks several times prior to the actual recording. A brief instruction of the recording procedure was given to each speaker before the recording. Recordings were made in a soundproof room with the microphone (SM58A, Shure) positioned at approximately 8 cm from the speaker's mouth. Audio signals were recorded at a sampling rate of *fs* = 20 kHz and quantization rate of 16 bits/sample using Praat. A steady-state medial segment (window size = 4500 points) was selected for analysis from each participant's recording. Voice onset and offset were excluded to avoid effects of interaction between the larynx and vocal tract on analysis. Perturbation and nonlinear dynamic analysis were then carried out with these signals.

C. Perturbation analysis

Perturbation measures are often used to non-invasively and objectively assess laryngeal function and voice quality. Jitter and shimmer have been commonly used to analyze the perturbation of normal and SE voices. Jitter is a measure of short-term (cycle-to-cycle) variation in the fundamental frequency of a signal, whereas shimmer measures the amplitude variation of a signal. In this study, to assess the reliability of perturbation measurements, jitter and shimmer were measured from the normal and SE voice signals segmented into two lengths (2500 and 4500 points) by using the Praat software [8].

D. Nonlinear dynamic analysis

The dynamics of each voice segment was reconstructed in a phase space, which was then used to calculate correlation dimension and Kolmogorov entropy. The reconstructed phase space can be used to describe the dynamic behavior of a signal: a periodic signal produces a closed trajectory, while an aperiodic signal appears to be irregular and chaotic [9]. Correlation dimension (D_2) and Kolmogorov entropy (K_2) are useful in describing irregular phenomena. D_2 specifies the number of degrees of freedom needed to describe a system; a more complex system has a higher dimension, meaning more degrees of freedom are needed to describe its dynamic state [10]. D_2 allows us to distinguish between deterministic chaos and random noise. And K_2 quantifies the rate of loss of information regarding the state of a dynamic system as it evolves, and $K_2 > 0$ is a sufficient condition for a chaotic system [11]. Detailed descriptions concerning applications of phase space reconstruction, correlation dimension, and Kolmogorov entropy can be widely found in the literature [1, 4, 11]. In the present study, an m-dimensional delay coordinate phase space Xi $\{x(t_i), x(t_i-\tau), ..., x(t_i-(m-1)\tau)\}$ was reconstructed using the time delay technique, where *m* is the embedding dimension and τ the time delay [9]. *m* was determined using the False Nearest Neighbors method[12] and the proper time delay τ was estimated using the C-C methods [13]. Correlation integral *C(r)* measures the number of distances between points in the reconstructed phase space are smaller than the radius *r*. Based on *C(r)*, D_2 and K_2 were manually estimated in the scaling region of the radius *r* with the embedding dimension *m*.

E. Statistical analysis

Since the nonlinear parameters have Non-Gaussian population, nonparametric Mann-Whitney rank sum test was used to assess if jitter, shimmer, D_2 and K_2 (dependent variables) were significantly different among the voice types (independent variable). In addition, Mann-Whitney rank sum test was also used to examine if data length may significantly affect the dependent variables. Statistical significance level was set at the 0.05 level for all tests.

III. RESULTS

The corresponding phase spaces associated with normal and SE voices are shown in Figures 1A and 1B. According to Figure 1A, the reconstructed phase space of the normal voice has a regular structure, where the proper time delay τ was determined to be 8 using the C-C methods and the embedding dimension *m* was estimated as 6 using the False Nearest Neighbors method.

However, the reconstructed phase space of SE voice expressed an irregular structure, as shown in Figure 1B.

The average and standard deviation values of jitter and shimmer of normal and SE voice signals are shown in Table I.

TABLE I
PERTURBATION ANALYSIS FOR NORMAL VOICE AND ESOPHAGEAL VOICE

	Normal Voice	Esophageal Voice	Mann-Whitney Test results
Jitter	$M = 0.0027$	$M = 0.0161$	$U = 35$
	$SD = 0.0016$	$SD = 0.0091$	$P < 0.001$**
Shimmer	$M = 0.0602$	$M = 0.1143$	$U = 142$
	$SD = 0.0325$	$SD = 0.0434$	$P < 0.001$**

** Statistically significance at $P = 0.001$
Abbreviations: M, means; SD, standard deviation

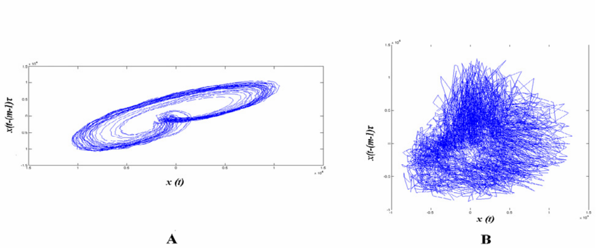

Fig. 1. Reconstructed phase spaces associated with a typical (A) normal voice, and (B) esophageal voice.

Mean jitter and shimmer of normal voices with data length of 4500 points were 0.0027 and 0.0602, respectively. In contrast, mean jitter and shimmer values of SE voices with data length of 4500 points were 0.0161 and 0.1143. Results of Mann-Whitney rank sum test on perturbation results revealed a significant difference between normal voice and SE voice ($P < 0.001$). Perturbation of SE voice was much higher than the normal voice, and the periodicity of SE voice was lower than normal voice.

To describe the stability and reliability of perturbation measures, jitter and shimmer of normal voice and SE voice signals with length of 2500 points were also calculated. Distribution of these results is shown in Figure 2. For data length of 2500 points, mean jitter and shimmer values of normal voice were 0.0023 and 0.0561, respectively. Mann-Whitney rank sum test indicated no significant difference in jitter and shimmer between data lengths of 4500 points and 2500 points ($P = 0.45 > 0.05$ for jitter; $P = 0.46 > 0.05$ for shimmer). This indicated that the perturbation measures were reliably calculated for the normal voice, which resembled nearly-periodic signals. However, mean jitter and shimmer measures in the SE voice signals with length of 2500 points were 0.0308 and 0.1195, respectively. Mann-Whitney rank sum test showed no significant difference in shimmer measures of SE voice using different data length ($P = 0.5857 > 0.05$), but the jitter measure of SE voice was significantly different between different data lengths ($P < 0.001$). This indicates that perturbation measures for aperiodic voice may be questionable and unreliable.

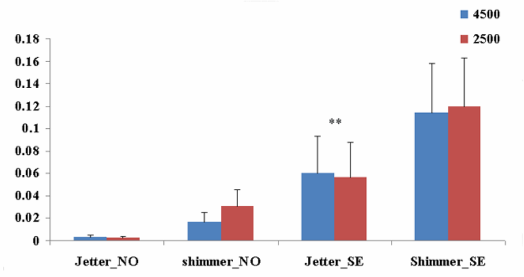

Fig.2. Distribution of perturbation analysis of normal voice and esophageal voice signals with different data lengths.

Table II summarizes the results of nonlinear dynamic analysis of normal and SE voice with data length of 4500 points. A higher mean D2 value was found for SE voice (M = 2.405) than normal voice (M = 1.447). Similarly, a higher K2 value was found for SE voice (M = 0.023) than normal voice (M = 0.014). Results of the Mann-Whitney rank sum tests revealed that, for both D2 and K2, normal voice was significantly different from SE voice (P < 0.001). This indicated that SE voice was more chaotic than normal voice.

TABLE II
NONLINEAR DYNAMIC ANSLYSIS FOR NORMAL VOICE AND OESOPHAGEAL VOICE

	Normal Voice	Esophageal Voice	Mann-Whitney Test results
Correlation	$M = 1.447$	$M = 2.405$	$U = 24$
dimension (C_2)	$SD = 0.0016$	$SD = 0.0091$	$P < 0.001$**
Kolmogorov	$M = 0.014$	$M = 0.023$	$U = 39$
entropy (K_2)	$SD = 0.0325$	$SD = 0.0434$	$P < 0.001$**

**Statistically significance at P= 0.001
Abbreviations: M, means; SD, standard deviation

The D_2 and K_2 values of SE voice using 2500 points were also calculated. Figure 3 shows the distribution of these results compared to values using 4500 points. Mean D_2 and K_2 of SE voice using 2500 points were 2.207 and 0.023, respectively. Mann-Whitney rank sum test indicated no significant difference in D_2 and K_2 of SE voice between different data lengths ($P = 0.73 > 0.05$ for D_2; $P = 0.74 > 0.05$ for K_2). This implied that correlation dimension and Kolmogorov entropy were reliable and stable measures in evaluating the voice quality of SE phonation.

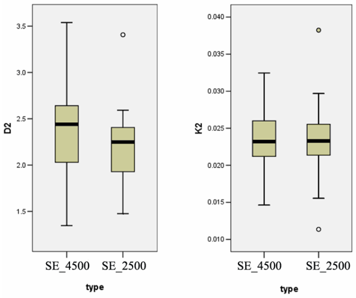

Fig.3. Distribution of nonlinear dynamic analysis of esophageal (SE) voice signals with different data length.

IV. Discussion

The present study examined the perturbation and nonlinear dynamic characteristics associated with normal and SE voice voicing source (neoglottis), SE speech has been found to have reduced intelligibility, fundamental frequency, duration, and intensity, increased formants, and altered aerodynamic characteristics when compared with laryngeal speech [14-18]. Since the PE segment is a composite of both skeletal and smooth muscle fibers [19], the control over PE segment vibration is diminished. Combined with the relatively asymmetric and irregular anatomy of PE segment, SE voice exhibits greater aperiodicity and instability than laryngeal voice. The imprecise, irregular, and relatively slower PE segment vibration contributes to the marked acoustic differences between SE and laryngeal voices, and the poor voice quality associated with SE voices.

Despite perturbation measures can be used to analyze the periodicity of SE voice signals, it appears to be questionable and lack reliability. As discussed previously, voice signals can be qualitatively classified into three types: nearly-periodic signals, signals with strong subharmonics or modulations, and aperiodic signals. Recent studies suggested that perturbation analysis may only be suitable for nearly periodic signals [20]. It may not be ideal for analyzing aperiodic voice signals due to the ill-defined fundamental frequency and peak amplitude. Zhang also demonstrated that perturbation analyses were not suitable for signals with short data length, low sampling rate, and high noise level [21]. Based on our results, it can also be seen that perturbation measures may be inadequate for evaluating SE voice. Results of perturbation measures were affected by the data length of voice signals being analyzed. Nonlinear dynamic analysis, however, appeared to be insensitive to data length. This indicates that the use of perturbation analysis in assessing the voice quality of aperiodic signals was unreliable and questionable.

It has been reported that nonlinear dynamic analysis was more reliable and stable in analyzing aperiodic *laryngeal* voice signals, such as voices of unilateral laryngeal paralysis and hyperfunctional voice disorders [3, 21]. SE speech represents a totally different voice production mechanism, with the PE segment being used in place of the vocal folds. As the PE segment is usually composed of the inferior pharyngeal constrictor muscle, the cricopharyngeus muscle, and the upper portion of the SE sphincter[19], its vibration exhibits nonlinear stress-strain characteristics. To further complicate the scenario, the upper part of the esophagus is also used as a new air reservoir in SE phonation. This may lead to a highly nonlinear relationship between sub-PE segment pressure and the neoglottal area, explaining the nonlinear phenomena.

The present study demonstrates that SE voices were significantly more chaotic than normal laryngeal voices, based on comparisons using correlation dimension (D_2) and Kolmogorov entropy (K_2) measures. These results are very consistent with the study of MacCallum et al. [6]. Moreover, traditional perturbation measures including jitter and shimmer were found to be unreliable and inadequate as they failed to reveal any significant difference in using different data lengths. Nonlinear dynamic analysis appeared to be more superior in evaluating aperiodic signals such as SE voices. It can be used to describe the geometric scaling property of signals in the phase space, which is independent of the periodicity characteristics of the signals.

References

[1] Y. Zhang*, et al.*, "Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis," *Journal of Voice,* vol. 19, pp. 519-528, 2005.

[2] I. R. Titze*, et al.*, *Workshop on acoustic voice analysis: Summary statement*: National Center for Voice and Speech, 1995.

[3] Y. Zhang*, et al.*, "Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps," *The Journal of the Acoustical Society of America,* vol. 115, p. 2270, 2004.

[4] J. J. Jiang*, et al.*, "Nonlinear dynamics of phonations in excised larynx experiments," *The Journal of the Acoustical Society of America,* vol. 114, p. 2198, 2003.

[5] M. P. Robb, "Bifurcations and chaos in the cries of full-term and preterm infants," *Folia phoniatrica et logopaedica,* vol. 55, pp. 233-240, 2003.

[6] J. K. MacCallum*, et al.*, "Acoustic Analysis of Aperiodic Voice: Perturbation and Nonlinear Dynamic Properties in Esophageal Phonation," *Journal of Voice,* vol. 23, pp. 283-290, 2009.

[7] M. L. Ng*, et al.*, "Speech performance of adult cantonese-speaking laryngectomees using different types of alaryngeal phonation," *Journal of Voice,* vol. 11, pp. 338-344, 1997.

[8] P. B. D. Weenink. (2009, *Praat: doing phonetics by computer (Version 5.1.05)*. Available: http://www.praat.org/

[9] N. H. Packard*, et al.*, "Geometry from a time-series," *Physical Review Letters,* vol. 45, pp. 712-716, 1980.

[10] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena,* vol. 9, pp. 189-208, 1983.

[11] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov entropy from a chaotic signal," *Physical Review A,* vol. 28, pp. 2591-2593, 1983.

[12] M. B. Kennel*, et al.*, "Determining embedding dimension for phase-space reconstruction using a geometrical construction.," *Physical Review A,* vol. 45, pp. 3403-3411, 1992.

[13] H. S. Kim*, et al.*, "Nonlinear dynamics, delay times, and embedding windows," *Physica D,* vol. 127, pp. 48-60, Mar 1999.

[14] J. Gandour and B. Weinberg, "Production of intonation and contrastive stress in esophageal and trachoesophageal speech," *Journal of Phonetics,* vol. 13, pp. 83-95, 1985.

[15] M. L. Ng and R. Chu, "An Acoustical and Perceptual Study of Vowels Produced by Alaryngeal Speakers of Cantonese," *Folia phoniatrica et logopaedica,* vol. 61, pp. 97-104, 2009.

[16] M. L. Ng*, et al.*, "Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech," *Folia phoniatrica et logopaedica,* vol. 53, pp. 36-47, 2001.

[17] M. L. Ng*, et al.*, "Perceptions of tonal changes in normal laryngeal, esophageal, and artificial laryngeal male Cantonese speakers," *Folia phoniatrica et logopaedica,* vol. 50, pp. 64-70, 1998.

[18] M. L. Ng and J. Wong, "Voice Onset Time Characteristics of Esophageal, Tracheoesophageal, and Laryngeal Speech of Cantonese," *Journal of Speech Language and Hearing Research,* vol. 52, pp. 780-789, 2009.

[19] Y. Edels, *Pseudo-voice: Its theory and practice*. Sydney,Australia: Croom Helm, 1983.

[20] H. Herzel*, et al.*, "Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling," *Chaos,* vol. 5, pp. 30-34, 1995.

[21] Y. Zhang*, et al.*, "Comparison of nonlinear dynamic methods and perturbation methods for voice analysis," *Journal of the Acoustical Society of America,* vol. 118, pp. 2551-2560, 2005.