

Perturbation and Nonlinear Dynamic Analysis of Sustained Vowels in Normal and Pathological Voices

Lee, JiYeoun¹⁾ · Choi, Seong Hee · Jiang, Jack J. · Hahn, MinSoo · Choi, Hong-Shik

ABSTRACT

In this paper, we investigate the acoustic characteristics of sustained voices from normal subjects and patients with laryngeal pathologies. Perturbation methods (including jitter and shimmer), signal-to-noise ratio (SNR), and nonlinear dynamic methods (such as correlation dimension) are used to analyze normal and pathological voices. We find that jitter does not statistically discriminate between normal and pathological voices, but a significant difference is found for shimmer, SNR, and correlation dimension. The results suggest that nonlinear dynamic analysis may be valuable for the analysis of normal and pathological voices but perturbation analysis should be applied with caution for pathological voice analysis.

Keywords: Nonlinear dynamic analysis, perturbation analysis, laryngeal pathologies, normal voices

1. Introduction

Over the past few years a considerable number of studies have been applied on the acoustic analyses, including perturbation methods including jitter and shimmer for the laryngeal pathologies [1]. However, since these parameters are based on the fundamental frequency, a very reliable pitch detection algorithm is essential to measure voicing irregularities [2][3]. In a severely chaotic voice signal which exhibits an irregular and aperiodic waveform, it tends to show extreme and unstable perturbation values [4]. In addition, perturbation analysis has been found to be sensitive to variations in pitch extraction algorithm and analysis tools such as multi-dimensional voice profile (MDVP) and CSpeech. It is also sensitive to aperiodicity as well as to error that can be created by environmental noise and measurement noise from recording

and sampling [2-6]. Although signal-to-noise ratio (SNR) is greatly influenced by various noises, the variance of the perturbation methods allows for the utilization of SNR parameter.

Nonlinear dynamic methods have recently received interest in the field of speech signal processing and enable us to quantitatively describe aperiodic and chaotic phenomena [7][8]. They have shown the potential ability to reliably quantify both periodic and aperiodic signals, to describe disordered voices, to classify pathological voices from normal ones, finally, and to quantify the degree of aperiodicity and irregularity [9-14].

The objective of this paper is to introduce nonlinear dynamic methods which have recently discussed in the United States of America (USA). We will compare acoustic characteristics of normal and pathological voices by using perturbation analysis (including jitter and shimmer), SNR, and nonlinear dynamic analysis (such as correlation dimension). We will then examine the ability of nonlinear dynamic and perturbation analyses to distinguish between normal and pathological voices by comparing the results before and after adding voice samples which have unreliable error estimates (error > 10).

1) Department of Surgery, Division of Otolaryngology—Head and Neck Surgery, University of Wisconsin Medical School, 5745a Medical Sciences Center, 1300 University Avenue, Madison, WI 53706. E-mail: leeji@surgery.wisc.edu

2. Database

The voice samples utilized in this study were selected from the Disordered Voice Database, model 4337, Version 1.03 (Kay Elemetrics Corporation, Lincoln Park, NJ), developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. We used 20 normal subjects (including 10 males and 10 females, ranging in age between 26 and 55 years) and 20 patients (including 7 males and 13 females, ranging in age between 18 and 75 years) from this database. The subject information is shown in <Table 1>, and more detailed information has been given in the Disordered Voice Database. The voice samples with the sampling rate of 44.1 kHz were selected by subjects sustaining the vowel /a/.

3. Data analysis

3.1 Perturbation analysis

The acoustic perturbation measures (percent jitter and percent shimmer) and SNR were obtained from CSpeech software, version 4.0 (Milenkovic and Read, Madison, WI). Jitter is a measure of cycle-to-cycle fluctuations in the fundamental period. Shimmer is a measure of cycle-to-cycle variation in waveform amplitude. SNR indicates the amount of noise present in the speech waveform.

analyzed [2-6]. Error in perturbation measurement is calculated by the CSpeech program to determine reliability. The error indicates the number of times the analysis algorithm overlooks a pitch period consistent with the peak of the autocorrelation function used to calculate jitter, shimmer, and SNR values [15]. Therefore, the error value acts as a reliability measurement for all three parameters. In accordance with the CSpeech user manual, an error count greater than 10 indicates that perturbation analysis methods is unreliable [15].

3.2 Nonlinear dynamic analysis

Detailed descriptions of nonlinear dynamic analysis methods, such as phase space reconstruction, correlation dimension, and second-order entropy to human voice production are widely found in the literature [4][7-8][11-14]. We brought almost all background and theory of nonlinear dynamic analysis described in this paper from previous papers.

The dynamics of each voice segment can be reconstructed in a phase space [8]. A reconstructed phase space is created by plotting a voice signal against itself at some time delay. The reconstructed phase space shows the dynamic behavior of a signal: a periodic signal produces a closed trajectory, while an aperiodic signal produces a chaotic trajectory as shown in <Figure 1. (a) and (b)>. Correlation dimension, D_2 , quantifies the complexity of a reconstructed phase space: $D_2 = 0$

Table 1. Subject information.

Subject No.	Sex	Age (y)	Diagnosis	Subject No.	Sex	Age (y)	Diagnosis
1	M	34	Normal	21	F	50	Chronic laryngitis
2	F	32	Normal	22	F	63	Bowing
3	F	34	Normal	23	M	69	Paralysis
4	M	39	Normal	24	F	49	Vocal fold edema
5	M	32	Normal	25	M	29	Papilloma
6	M	28	Normal	26	F	61	Polyp
7	F	40	Normal	27	F	25	Vocal fold edema
8	F	26	Normal	28	F	31	Polypoid degeneration (Reinke's)
9	F	43	Normal	29	M	40	Scarring
10	M	55	Normal	30	F	75	Parkinson's Disease
11	F	43	Normal	31	F	38	Keratosis / leukoplakia
12	F	45	Normal	32	F	57	Vocal fold edema
13	M	26	Normal	33	F	43	Polypoid degeneration (Reinke's)
14	F	28	Normal	34	M	38	Bowing
15	M	29	Normal	35	M	42	Keratosis / leukoplakia
16	M	59	Normal	36	M	29	Polyp
17	M	40	Normal	37	F	42	Vocal fold edema
18	M	46	Normal	38	F	18	Vocal nodule s
19	F	33	Normal	39	M	18	Paralysis
20	F	32	Normal	40	F	53	Paralysis

Previous studies have shown that perturbation analysis of aperiodic voices is unreliable. However, nearly periodic voices with jitter and shimmer values less than 5% can be reliably

corresponds to a static state; $D_2 = 1$ corresponds to a periodic oscillation; $D_2 = 2$ describes a quasi-periodic signal, and fractal D_2 describes an aperiodic or chaotic oscillation. Therefore, a

more complex system has a higher dimension, which means that more degrees of freedom may be needed to describe its behavior[16]. Kolmogorov entropy quantifies the rate of loss of information about the state of a dynamic system as it evolves. Second-order entropy (K_2) is the lower bound of Kolmogorov entropy, and a positive, K_2 , provides a sufficient condition for chaos. For periodic behavior, this entropy is equal to zero. A chaotic system with a finite degree of freedom has a finite K_2 value, whereas the K_2 value of true random behavior approaches infinity [17].

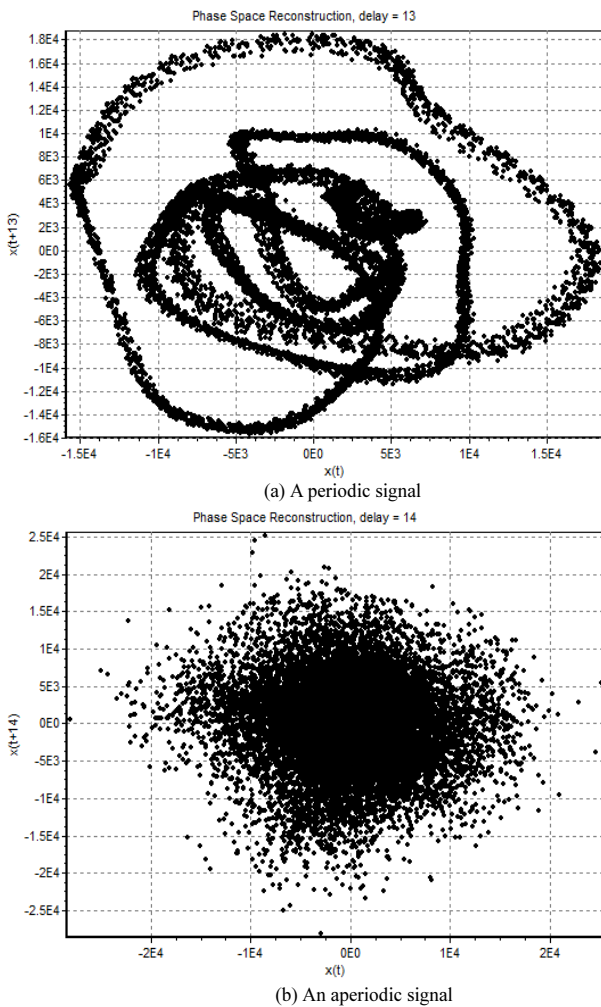


Figure 1. Reconstructed phase space of a periodic and an aperiodic signal.

In this study, correlation dimension is performed using nonlinear dynamic analysis software developed by the Laryngeal Physiology Laboratory at the University of Wisconsin. Calculations made by the software are based on the numerical algorithms described for studies analyzing excised larynx phonations and pathological human voices [11-12][14]. Briefly, an m -dimensional delay-coordinate phase space, $X_i = \{x(t_i),$

$x(t_i - \tau), \dots, x(t_i - (m - 1)\tau)\}$, is reconstructed using the time delay technique, where m is the embedding dimension and τ is the time delay [18]. Dimension, m , is determined according to the embedding theorem [19].

The proper time delay, τ , is estimated using the mutual information method proposed by Fraser and Swinney[20]. The improved algorithm proposed by Theiler is used to calculate the correlation integral $C(r)$, where r is the radius around X_i [21]. Correlation integral $C(r)$ measures the number of distances between points in the reconstructed phase space that are smaller than the radius r . $C(r)$ has a power law behavior $C(r) \propto r^{D_2} e^{-m\tau K_2}$, which reveals the geometrical scaling property of the attractor [18]. Based on $C(r)$, D_2 is estimated in the scaling region of the radius r with the increase of the embedding dimension m as shown in <Figure 2>. For sufficiently large m , the correlation dimension and its standard deviation are derived using a curve fit to the curve of $\log_2 C(r)$ versus $\log_2 r$ in the scaling region. <Figure 2> gives the estimated correlation dimension versus the radius r , where the curves from bottom to top correspond to $m=1, 2, \dots, 12$, respectively. When r is within the scaling range ($2^{7.2} = r_1 < r < r_2 = 2^{8.1}$), with the increase of m , the estimated correlation dimension approaches 1.106 ± 0.003 . This result is shown more clearly in <Figure 3>, where D_2 is plotted as a function of m . When m is sufficiently large, the voice is estimated as 1.106 ± 0.003 , which differs from the nonconvergent estimate of the random noise, as shown in <Figure 3>.

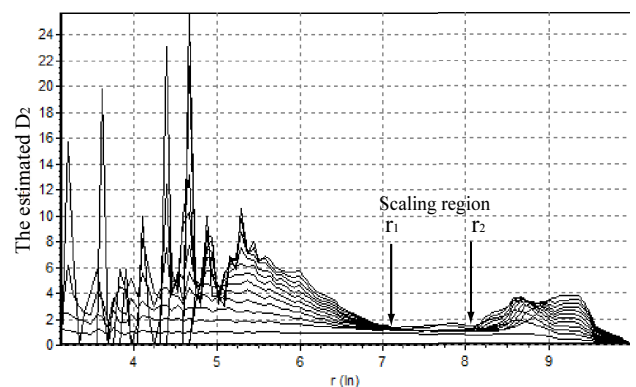


Figure 2. The estimated dimension versus r , where the curves from bottom to top correspond to $m=1, 2, \dots, 12$, respectively.

The reliability of nonlinear dynamic analysis calculations is determined for each voice signal using the standard deviations of the estimated D_2 values which refers to as SDD_2 . For reliable estimation of dimension in a particular signal, SDD_2 values

should be less than 5%. SDD_2 values greater than 5% indicate that the nonlinear dynamic analysis method was unreliable for a signal.

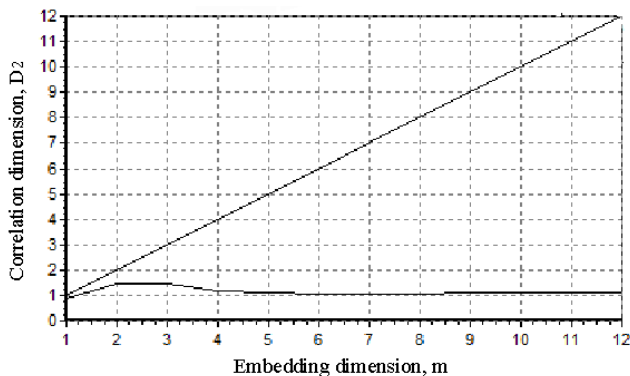


Figure 3. D_2 is plotted as a function of m , where the estimate of D_2 converges to 1.106 ± 0.003 with the increase of m

3.2 Statistical analysis

Percent jitter, percent shimmer, SNR, and correlation dimensions are compared for the two groups (20 normal and 20 pathological sustained vowel samples). The Mann-Whitney rank sum test is employed using jitter, shimmer, SNR, and correlation dimension as dependent variables and the subject groups (normal and pathological) as independent variables. Statistical significance level is set at the level of 0.05. SPSS 12.0 software is used for statistical analysis.

4. Results and discussion

<Figure 4 and 5> show the distributions of jitter (%) and shimmer (%), respectively. The box plots represent better visualization between normal and pathological voices. It is made by min, first quartile, median, third quartile, max values, and outliers. As mentioned in chapter 3.1, many samples analyzed via perturbation methods have unreliable error estimates (error > 10). In this paper, 25% of analyzed pathological data showed unreliably high errors. So, to compare the reliability of perturbation method, statistical analyses before and after adding the pathological data which have high errors are discussed in <Tables 2 and 3>. When pathological data with error > 10 were removed, the median jitters of normal and pathological groups

were 0.29 and 0.48, respectively. Then, no significant difference was seen in jitter ($P = 0.053$). When all data samples were included, regardless of error, they were 0.29 and 0.57, respectively. Mann-Whitney rank sum tests showed that jitter reveals a statistically significant difference between pathological and normal voices ($P = 0.005$). And the median shimmers of normal and pathological voices were 1.93 and 2.71, respectively, when pathological data with error > 10 were removed. When all data samples were included, they were 1.93 and 5.48, respectively. Mann-Whitney rank sum tests showed that shimmer reveals a statistically significant difference between pathological and normal voices in two cases ($P = 0.011$ and $P = 0.001$).

<Figure 6> shows the distributions of SNR (dB). In <Table 2>, when unreliable error estimates were removed, the median SNR values of normal and pathological voices were 24.22 and 17.80, respectively. In <Table 3>, when unreliable error estimates were included, the median SNR values of normal and pathological voices were 24.22 and 16.20, respectively. Mann-Whitney rank sum tests showed that SNRs shown in <Table 2 and 3> reveal a statistically significant difference between pathological and normal voices ($P = 0.004$ and $P < 0.001$). <Figure 7> shows the distributions of D_2 . Results of D_2 analysis for normal and pathological voices are given in <Tables 2 and 3>. In case of removing pathological data with error > 10, the median D_2 of normal and pathological voices were 1.16 and 2.62, respectively. On the other hands, in case of including pathological data with error > 10, the median D_2 of normal and pathological voices were 1.16 and 2.68, respectively. Mann-Whitney rank sum tests showed that both demonstrated a statistically significant difference between pathological and normal voices ($P < 0.001$ and $P < 0.001$). As shown in <Figure 4 to 6>, the distributions of the acoustic parameters such as jitter, shimmer, and SNR showed a similar characteristic between normal and pathological voices as mentioned in other researches[4-6][11]. Specifically, compared to other distributions, D_2 had a definite threshold to classify normal and pathological voices as easily found in <Figure 7>. Also, we can confirm that the D_2 values of pathological voices have a broad one with ratings distributed from 1.82 to 3.50 and are higher than that of normal voices.

Table 2. Comparisons of normal and pathological voices for differences in jitter, shimmer, SNR, and D_2 (removing pathological data which show unreliable error estimates).

	Normal voices (N=20)		Pathological voices (N=15)		Mann-Whitney (P value)
	Median	25%-75% range	Median	25%-75% range	
Jitter (%)	0.29	0.25-0.48	0.48	0.33-0.62	0.053
Shimmer (%)	1.93	1.34-2.96	2.71	2.09-6.09	0.011*
SNR (dB)	24.22	19.72-24.74	17.80	13.50-18.30	0.004*
D2	1.16	1.13-1.44	2.62	1.78-2.93	< 0.001*
Error count	mean = 0, std. = 0		mean = 0.47, std. = 0.99		

* Significant

Table 3. Comparisons of normal and pathological voices for differences in jitter, shimmer, SNR, and D_2 (including pathological data which show unreliable error estimates).

	Normal voices (N=20)		Pathological voices (N=20)		Mann-Whitney (P value)
	Median	25%-75% range	Median	25%-75% range	
Jitter (%)	0.29	0.25-0.48	0.57	0.38-1.42	0.005*
Shimmer (%)	1.93	1.34-2.96	5.48	2.32-9.25	0.001*
SNR (dB)	24.22	19.72-24.74	16.20	11.20-18.18	< 0.001*
D2	1.16	1.13-1.44	2.68	1.82-3.05	< 0.001*
Error count	mean = 0, std. = 0		mean = 28.50, std. = 67.93		

* Significant

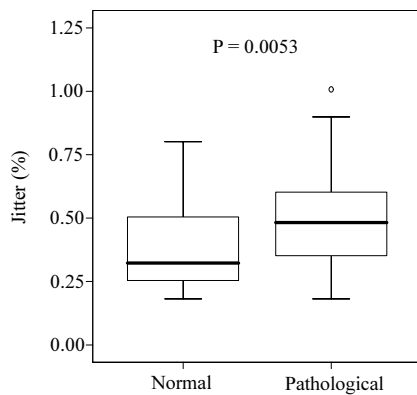


Figure 4. Distributions of jitter between normal and pathological voices.

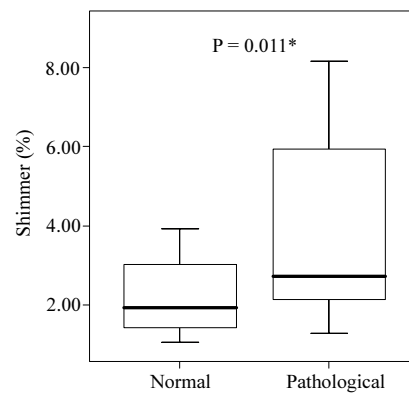


Figure 5. Distributions of shimmer between normal and pathological voices

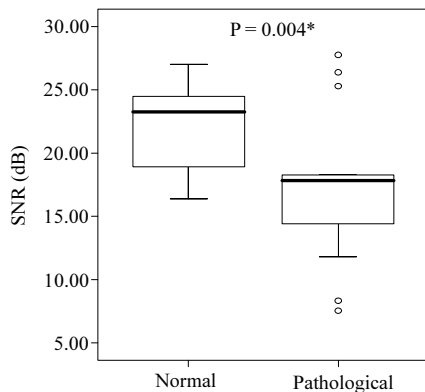


Figure 6. Distributions of SNR between normal and pathological voices.

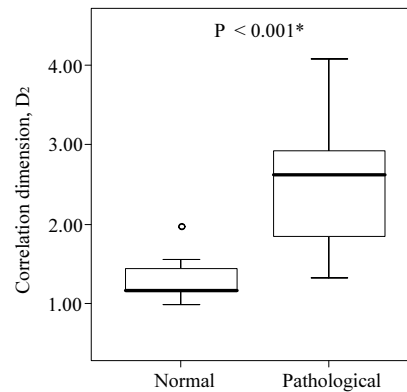


Figure 7. Distributions of correlation dimension between normal and pathological voices.

In this study, the traditional acoustic analysis methods (jitter, shimmer, and SNR) and nonlinear dynamic analysis method (correlation dimension, D_2) have been applied for the analysis of sustained vowels. Jitter and shimmer describe the temporal properties of a voice by measuring variations in the fundamental frequency and peak amplitude. Differing from jitter, shimmer, and SNR, the correlation dimension describes the properties of complexity and predictability of a voice in a state space and does not require the definition of cycle period. Thus, perturbation analysis and nonlinear dynamic analysis provide different but complementary information on the analysis of voice[10-12]. As shown in <Table 3>, when all data samples are included, regardless of error, all these acoustic parameters except jitter show a significant difference between normal and pathological voices. Also, although Mann-Whitney rank sum tests show that shimmer and SNR reveal a statistically significant difference between pathological and normal voices in <Table 2 and 3>, the their P values present the high variation degree compared to before removing pathological data with error > 10. On the other hand, P values of D_2 show same value as shown in <Table 2 and 3>. These facts suggest that it can be unreliable to use the perturbation methods to estimate jitter, shimmer, and SNR in pathological voices due to a failed pitch extraction. In contrast, error count for all normal voices is 0 as shown in <Table 2 and 3>, indicating the perturbation methods are reliably calculated for these nearly periodic signals. Therefore, jitter, shimmer, and SNR should be cautiously applied to the analysis of pathological voices, whereas correlation dimension may represent valuable methods for normal and pathological voice analysis.

5. Conclusion

In this paper, we have applied acoustic measures of nonlinear dynamics and perturbation to normal and pathological voices. Although the distributions of perturbation measures such as jitter, shimmer, and SNR significantly differentiated between pathological and normal voice, error count demonstrated the insufficient reliability of these measures in quantifying the pathological voice signal. No significant difference was also seen in jitter ($P = 0.053$). On the other hand, nonlinear dynamic measure like correlation dimension obviously differentiated between pathological and normal voices. As evidenced by P value, the measure was able to reliably quantify normal and pathological signals. In conclusion, shimmer, SNR, and correlation dimension each successfully discriminated between normal and pathological

voices, where shimmer and SNR values were calculated for periodic voices and correlation dimension values were calculated for both periodic and aperiodic voices. Therefore, nonlinear dynamic analysis may provide more information and present a valuable procedure to objectively classify normal and pathological voices; but perturbation analysis should be applied with caution for pathological voice analysis.

Future research should focus on clinical application of nonlinear dynamic analysis as a valuable, reliable tool for measurement of extremely aperiodic voice, such as noise like voice.

Acknowledgements

Gratitude is expressed to Dr. Yu Zhang for work in nonlinear dynamic analysis expertise.

References

- [1] Lieberman, P. (1961). "Perturbations in vocal pitch", *Journal of the Acoustical Society of America*, Vol. 33, No. 5, pp. 597-603.
- [2] Titze, I. R. (1981). "Workshop on acoustic voice analysis: summary statement", National Center for Voice and Speech, pp. 1-36.
- [3] Kay Elemetrics Corp. (1993). "Multi-dimensional voice program: software instruction manual", Pine Brook, NJ, Kay Elemetrics Corp.
- [4] Zhang, Y., Wallace, S. M., and Jiang, J.J. (2005). "Comparison of nonlinear dynamic methods and perturbation methods for voice analysis", *Journal of the Acoustical Society of America*, Vol. 118, No. 4, pp. 2551-2560.
- [5] Titze, I. R. & Liang, H. (1993). "Comparison of F0 extraction method for high-precision voice perturbation measurements", *Journal of Speech and Hearing Research*, Vol. 36, No. 6, pp. 1120-1133.
- [6] Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., and Berke, G. S. (1996). "Comparison of voice analysis systems for perturbation measurement", *Journal of Speech and Hearing Research*, Vol. 39, No. 1, pp. 126-134.
- [7] Titze, I. R., Baken, R., and Herzel, H. (1993). "Evidence of chaos in vocal fold vibration", In: Titze, I. R. ed. *Vocal Fold Physiology: New Frontiers in Basic Science*, San Diego, CA, Singular Publishing Group, pp. 143-188.
- [8] Jiang, J. J., Zhang, Y., and McGilligan, C. (2006). "Chaos in

- voice, from modeling to measurement”, *Journal of Voice*, Vol. 20, No. 1, pp. 2-17.
- [9] Hertrich, I., Lutzenberger, W., Spieker, S., and Ackermann, H. (1997). “Fractal dimension of sustained vowel productions in neurological dysphonias: an acoustic and electroglottographic analysis”, *Journal of the Acoustical Society of America*, Vol. 102, No. 1, pp. 652-654.
- [10] Giovanni, A., Ouaknine, M., and Triglia, J. M. (1998). “Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis”, *Journal of Voice*, Vol. 13, No. 3, pp. 341-354.
- [11] Zhang, Y., Jiang, J. J., Biazzo, L., Jorgensen, M., and Berman, M. (2005). “Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis”, *Journal of Voice*, Vol. 19, No. 4, pp. 519-528.
- [12] Zhang, Y., McGilligan, C., Zhou, L., Vig, M., and Jiang, J. J. (2004). “Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps”, *Journal of the Acoustical Society of America*, Vol. 115, No. 5, pp. 2270-2277.
- [13] Herzel, H. & Berry, D. (1994). “Analysis of vocal disorders with methods from nonlinear dynamics”, *Journal of Speech and Hearing Research*, Vol. 37, No. 5, pp. 1008-1019.
- [14] Jiang, J. J., Zhang, Y., and Ford, C. N. (2003). “Nonlinear dynamics of phonations in excised larynx experiments”, *Journal of the Acoustical Society of America*, Vol. 114, No. 4, pp. 2198-2205.
- [15] Milenkovic, P. & Read, C. (1992). “CSpeech Version 4 User’s Manual”, Madison, WI.
- [16] Grassberger, P. & Procaccia, I. (1983). “Measuring the strangeness of strange attractors”, *Physica D: Nonlinear Phenomena*, Vol. 9, No. 1-2, pp. 189-208.
- [17] Grassberger, P. & Procaccia, I. (1983). “Estimation of the Kolmogorov entropy from a chaotic signal”, *Physical Review A - General Physics*, Vol. 28, No. 4, pp. 2591-2593.
- [18] Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S. (1980). “Geometry from a time series”, *Physical Review Letters*, Vol. 45, No. 9, pp. 712.
- [19] Takens, F. (1981). “Detecting strange attractors in turbulence”, In: Rand DA, Young LS, eds. *Lecture Notes in Mathematics*. Berlin, Springer-Verlag, pp. 366-381.
- [20] Fraser, A. M. & Swinney, H. L. (1986). “Independent coordinates for strange attractors from mutual information”, *Physical Review A - General Physics*, Vol. 33, No. 2, pp. 1134-1140.
- [21] Theiler, J. (1986). “Spurious dimension from correlation algorithms applied to limited time series data”, *Physical Review A - General Physics*, Vol. 34, No. 3, pp. 2427-2432.
- **Lee, JiYeoun:** Corresponding author
Address: Department of Surgery, Division of Otolaryngology-Head and Neck Surgery, University of Wisconsin Medical School, 5745a Medical Sciences Center, 1300 University Avenue, Madison, WI 53706.
Affiliation: UW Laryngeal Physiology Lab.
Telephone: +1-213-598-4410
E-mail: leej@surgey.wisc.edu
Research Interests: speech signal processing - voice measurement in patients with laryngeal pathology, etc.
2008~present Postdoctoral Fellow.
Ph.D., Dept. of Information & Communications Engineering, KAIST, 2008.
- **Choi, Seong Hee**
Address: Department of Surgery, Division of Otolaryngology-Head and Neck Surgery, University of Wisconsin Medical School, 5745a Medical Sciences Center, 1300 University Avenue, Madison, WI 53706.
Affiliation: UW Laryngeal Physiology Lab.
Telephone: 1-714-309-6012
E-mail: Choi@surgey.wisc.edu
Research Interests: voice disorder, dysphagia, tissue engineering, etc.
2007~present Researcher
Ph.D., Dept. of Speech Pathology, Yonsei Uni., 2007
- **Jiang, Jack J**
Address: Department of Surgery, Division of Otolaryngology—Head and Neck Surgery, University of Wisconsin Medical School, 5745a Medical Sciences Center, 1300 University Avenue, Madison, WI 53706.
Affiliation: UW Laryngeal Physiology Lab.
Telephone: +1- 608-265-7888
E-mail: jiang@surgey.wisc.edu
Research Interests: the vibratory properties of the vocal folds via studies of excised larynges, biomechanical modeling, aerodynamics, and analysis of laryngeal microstructure, speech signal processing, etc.
1998~present Professor, Division of Otolaryngology—Head and Neck Surgery
Ph.D., Speech Pathology and Audiology, Uni. of Iowa, 1991.
M.D., Shanghai Medical Uni., 1983.
- **Hahn, MinSoo**
Address: KOREA Advanced Institute of Science and Technology, 335 Gwahak-ro, Yuseong-gu, Daejeon, 305-701, Korea.
Affiliation: Speech and Audio Information Lab.
Telephone: +82-42-866-6123
E-mail: mshahn@kaist.ac.kr

• Choi, Hong-Shik

Address: Kangnam Severance Hospital, Yonsei University
College of Medicine. Institute of Logopedics and
Phoniatrics, 612 Enjuro Kangnamgu, Seoul, Korea.

Affiliation: Department of Otorhinolaryngology, Institute of
Logopedics and Phoniatrics.

Telephone: +82-2-2019-3460

E-mail: hschoi@yumc.yonsei.ac.kr