말소리와 음성과학 제2권 제3호 (2010)
pp. 165~170

Acoustic Analysis with Moving Window in Normal and Pathologic Voices

Choi, Seong Hee[1)] · Lee, JiYeoun Jiang · J. Jack

ABSTRACT

In this study, the most stable portion was identified using 5% moving window during /a/ sustained phonation in normal and pathologic voice signals and the perturbation values were compared between normal and pathologic voices at the mid-point and at the most stable portion using moving window, respectively.

The results revealed that some severe pathologic voice signals can be eligible for perturbation analysis by identifying the most stable portion with Err less than 10. In addition, the perturbation acoustic parameters did not differentiate the pathologic voice signals from the normal voice signals when the mid-point was selected to measure the perturbation analysis($p>0.05$). However, significantly higher %shimmer and lower SNR values were observed in pathologic voices ($p<0.05$) when the most stable portion was selected by moving window. In conclusion, moving window could identify the most stable portion objectively which can allow toget the minimum perturbation values (%jitter, %shimmer) and maximum SNR values. Thus, moving window technique can be applicable for more reliable and accurate perturbation acoustic analysis.

Keywords: acoustic analysis, perturbation, normal, pathologic voices, moving window

1. Introduction

Acoustic analysis has been commonly used for an objective voice assessment method of voice quality [1]. Generally, perturbation analysis such as jitter, shimmer and SNR requires the accurate determination of fundamental frequency (F0). Jitter is a measure of the cycle-to-cycle variation in the fundamental frequency and shimmer measures the cycle-to-cycle variation in amplitude. SNR (Signal-to-Noise Ratio) is a measure of the ratio between the harmonic signal and turbulent noise [2-3].

General standard statements of acoustic analysis supposed by Titze [4] give the guidelines to determine the reliabilityfor perturbation analysis with signal typing [4]. In terms of signal typing suggested by Titze, the voice signals can be classified into three types of signals. Titze suggested that type 1 signals define a nearly periodic, which could be suitable for perturbation analysis. Type 2 signals present strong modulations or subharmonics and type 3 signals are irregular or aperiodic characteristics and thereby, spectrographic analysis or nonlinear dynamic analysis as correlation dimension D2 estimation has been applied for type 2 and type 3 signals [5]. Recently, Sprecher et al. [6] updated signal typing in the voice by adding the type 4 signals which are characterized by stochastic broadband white noise characteristics [6]. It is suggested that perturbation analysis can be only reliable in nearly periodic voice signals [4].

To get the more accurate and reliable acoustic perturbation measures, some available commercialized voice analyzers provide cut off reference for estimating thereliability of perturbation measures. For instance, trk or error values which indicates the failure to estimate the pitch period in the voice signals are generated in *TF32*software [10] and *CSpeech* [11] and a signal with a greater than 10 error values indicates a higher aperiodic voice sample that cannot be appropriate for perturbation analysis[7-8]. Titze [10] also suggested less than 5 % perturbation

1) Department of Communicative Disorder, Goodnight Hall, 1975 Willow Dr., Madison, WI, 53706 & Department of Surgery, Division of Otolaryngology-Head and Neck Srugery, University of Wisconsin Medical School, 5745 a Medical Sciences Center, 1300 University Avenue, Madison, WI, 53706 USA.
choi @ surgery.wisc.edu

Received: August 11, 2010
Revision: September 14, 2010
Accepted: September 29, 2010

values of both %jitter and % shimmer might be reliable level of confidence [4].

Additionally, the alternative option to increase the reliability of objective acoustic analysis is selection of the most stable portion of the voice signals. A common clinical and laboratory procedure is to avoid the negative effects of onset and offset corresponding to the beginning and end of the signal. Such sections of a waveform are usually correlated with rapidly changing fundamental frequency and amplitude, which results in unreliable jitter and shimmer measurements. These waveform segments are extremely complex due to the many changes in aerodynamic and biomechanical parameters present during voicing onset and offset [9]. Overall, these currently practiced procedures of sample selection are very subjective. Therefore, it can be expected that the currently practiced subjective selection ofa sample will produce different perturbation measurements depending upon the exact location of the sample taken from the waveform. With respect to this concern, the segment of minimum perturbation (the most stable portion with the smallest variations) should be identified objectively in the given voice signals for measuring the aperiodicity. The aim of study is to apply the moving window as means of identifying the most stable portion and to compare the perturbation measures at the mid‐portion in the waveform with those at the most stable portion found using moving window.

2. Methods

2.1 Voice samples

The normal voice samples were obtained from five healthy males ranged in age from 21 to 23 years (mean of 21.8) and five healthy females ranged in age 20 to 22 years with a mean age of 20.8 years in the study. Subject participation was approved by the Institutional Review Board (IRB) of the University of Wisconsin, Madison. All participants were nonsmoking native speakers of American English. They reported normal hearing ability, no laryngeal and airway infection, and good health condition. Also, they were judged to present normal language skills determined by a certified speech-language pathologist. The pathological voice samples used were selected from the Disordered Voice Database, model 4337, Version 1.03, developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. 32 pathologic voice samples were chosen including 11 males and 21 females from the data base 9 vocal fold edemas, 4 vocal fold polyps, 3 vocal nodules, 1 vocal fold scarring, 5 vocal fold paralyses, 1 papilloma, 1 spasmodic dysphonia, 1 abnormal vocal process, 4 leukoplakias. 1 Parkinson's disease, 1 laryngitis, and 1 sulcus vocalis.

2.2 Recording

Recordings were made in a standardized manner [4],[9]. Participants wereasked to produce sustained /a/ vowel phonation with comfortable pitch and loudness for 3 sec at three times using a head-set microphone (AKG-c410, Kay Elemetrics) positioned at 10cm from the mouth at a 45 degree angle. This was done to evaluate how much variation would appear over the entire cycles during phonation A DAT-recorder(SONY TCD-8, Japan)was used with software of the Computerized Speech Lab, Model 4300(Kay Elemetrics, Lincoln Park, NJ) at a sampling rate of 44.1kHz.

2.3 Moving Window

The perturbation measures were repeated for each new voice segment as the window was moved forward by 25ms increments from the onset of phonation through its offset as shown in <Figure 1>. A moving window length of 500ms was selected based on previous data that suggested signal lengths of 500 to 825ms to keep variance in the parameters below 5% of the minimum. Then, all perturbation values corresponding to the each frame were obtained from onset to offset and finally, lowest %jitter, %shimmer and highest SNR were selected automatically. Thus, minimum perturbation values and maximum SNR values do not imply to average all perturbation values from onset to offset using moving window.

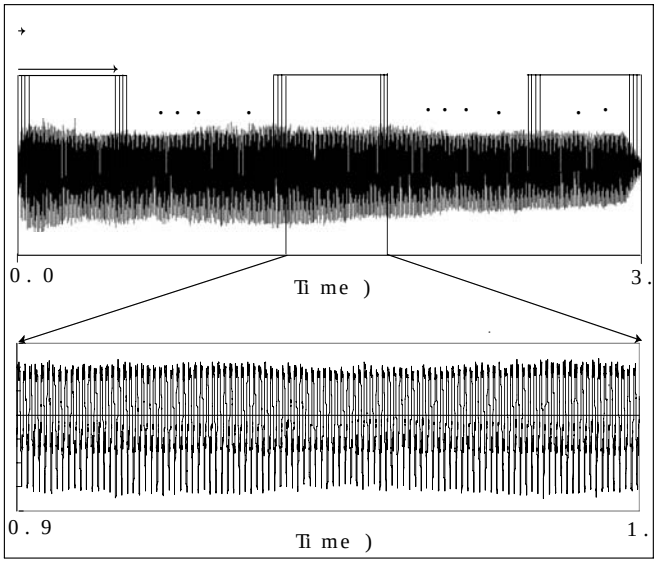

Figure 1. A 500ms moving window was shifted by 25 ms from voice onset through offset during a sustained /a/ phonation. The top window shows the whole 3 second sample while the bottom window shows one 0.5 second window used for analysis.

2.4 Perturbation analysis

Percent jitter, shimmer, and SNR were measured with TF32 software [6]. Least mean square measures were used for voice perturbation measures [2],[6]. This algorithm determines a pitch period during autocorrelation method. The reliability of these measures was estimated using TF32 values of "trk," which quantifies the number of dramatic fluctuations in pitch and "error," which indicates discrepancies in the calculated fundamental frequency likely due to voice breaks present in the sample[7]. Only voice signals which generated Error <10were chosen to compare the perturbation measures between normal and pathologic voice signals.

The perturbation values of normal voice signals at the mid-point were compared with pathologic voice signals at the mid-point for 500msec. Similarly, the perturbation values at the most stable portion which indicates the minimum perturbation (% jitter, %shimmer) and maximum SNR values, in normal voice signals identified by 5% moving window were compared with those of pathologic voice signals.

2.5 Statistical analysis

Statistical analysis was conducted using Sigma Stat 3.0 (Jandel Scientific, SanRafael, CA). Mann-Witney Rank Sum Tests were performed to test differences between normal and pathologic voice signals at the mid-point selection and at the most stable portion identified in the moving window. A level of significance was 95% for all measures.

3. Results

3.1 Perturbation analysis

Perturbation values of 10 normal voice signals at the mid-point in the waveform and minimum perturbation (minimum %jitter, shimmer) and maximum SNR values with Error values were summarized in <Table 1>, <Table 2> respectively. All Error values for normal voice signals were 0, which indicate there is no pitch loss tracking in given voice signals.

Perturbation values of 32 pathological voice signals at the mid-point in the waveform and minimum perturbation (minimum %jitter, shimmer) and maximum SNR values with Error values were summarized in <Table 3>, <Table 4> respectively. Error values ranged 0 ~ 283 in the perturbation analysis at the mid-point selection whereas 0 ~ 60 in perturbation analysis at the most stable segments obtained by moving window.

Table 1. Perturbation values of Normal voice signals at the mid-point in the waveform.

N	%jitter	%shimmer	SNR	Err
1	0.58	2.40	22.6	0
2	0.47	4.07	19.2	0
3	0.58	3.60	17.9	0
4	0.58	3.64	16.9	0
5	0.20	1.27	27.5	0
6	0.21	1.57	25.5	0
7	0.38	5.25	13.6	0
8	0.35	1.67	24.6	0
9	0.40	1.18	27.5	0
10	0.34	2.35	23.7	0

Table 2. Minimum perturbation and maximum SNR values of normal voice signals obtained by moving window.

N	%jitter	%shimmer	SNR	Err
1	0.30	1.50	25.4	0
2	0.41	3.09	19.9	0
3	0.42	2.67	20.2	0
4	0.50	2.97	17.3	0
5	0.18	1.06	28.0	0
6	0.17	1.34	26.2	0
7	0.26	2.95	18.4	0
8	0.28	1.50	25.5	0
9	0.26	1.05	28.6	0
10	0.21	1.86	25.0	0

3.2 Comparisons of perturbation measures at the mid-point selection vs. at the most stable portion selection

When we chose the voice signals with Err less than 10 for reliable perturbation analysis, 10 voice signals of normal, and 19 voice signals of pathologic voices at the mid-point selection were eligible, showing as gray column given in <Table 1>and <Table 3>.

As shown in <Figure 1>, Mann-Whitney Rank Sum test results showed that there is no significant difference in %jitter between normal and pathologic voices at the mid-point selection ($p=0.891$) and at the most stable portion selection ($p=0.235$).

Additionally, % shimmer and SNR values at the mid-point were not significantly different between normal and pathologic voices ($p=0.207$), ($p=0.215$) in <Figure 2>. However, significantly higher % shimmer and lower SNR values were observed in pathologic voices group when perturbation measures estimated at the most stable portion with moving window ($p<0.01$).

Table3. Perturbation values of pathologic voice signals at the mid-point in the waveform. The gray columnsrepresent perturbation values with less than 10 Error values.

P	%jitter	%shimmer	SNR	Err
1	0.17	1.74	23	0
2	0.32	3.39	20.9	0
3	0.25	1.32	24.5	0
4	0.26	1.47	27.5	0
5	0.17	1.93	26.6	0
6	0.42	2.32	20.3	0
7	0.26	3.33	21.6	0
8	0.23	2.48	23.5	0
9	0.42	2.49	18.5	0
10	0.53	2.33	17.3	0
11	0.94	5.87	13.4	2
12	0.36	2.59	17.9	0
13	0.63	7.44	15.3	2
14	0.38	5.05	19.3	0
15	0.31	7.59	17.6	0
16	0.26	1.72	26	0
17	4.7	28.19	5.9	61
18	1.69	14.27	12.4	11
19	1.17	13.79	6.8	1
20	3.54	31.24	5.2	30
21	0.61	5.41	12.4	0
22	3.35	23.83	6.3	129
23	1.13	16.71	6.5	19
24	3.2	13.77	9	8
25	3.29	41.17	3.2	77
26	10.63	46.2	3.6	85
27	7.63	43.16	2.7	46
28	9.5	42.02	3.7	78
29	5.26	11.58	5	35
30	2.32	43.38	2.3	38
31	6.53	41.79	2.6	40
32	5.56	25.99	1.7	283

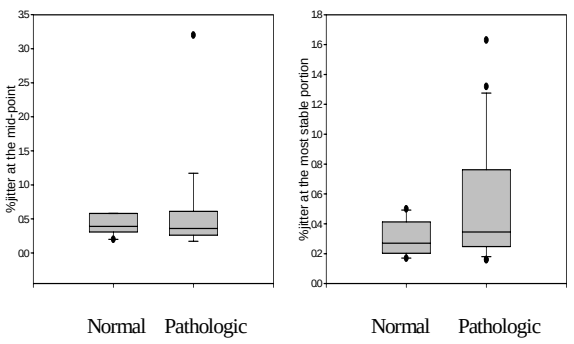

Figure 1. Box plots for % jitter values at the mid-point selection (Left) and at the most stable portion (Right). The midline represents the median, with the lower and upper boundaries of the box indicating the first and third quartile, respectively.

Table4. Minimum Perturbation values and maximum SNR values in the pathologic voice signals obtained using 5% moving window. The gray columns represent perturbation values with less than 10 Error values.

	%jitter	%shimmer	SNR	Err
1	0.16	1.56	23.8	0
2	0.27	3.24	21.2	0
3	0.24	1.25	24.6	0
4	0.25	1.19	27.8	0
5	0.16	1.81	26.7	0
6	0.34	2.24	20.7	0
7	0.24	2.54	23.4	0
8	0.23	2.29	23.7	0
9	0.40	2.18	18.9	0
10	0.43	2.16	18.1	0
11	0.83	4.96	13.6	1
12	0.35	2.51	18.5	0
13	0.52	5.83	15.8	0
14	0.26	4.05	20.3	0
15	0.30	6.92	18.0	0
16	0.25	1.65	26.3	0
17	1.32	7.28	8.50	7
18	1.17	7.35	16.7	1
19	0.74	12.83	8.00	0
20	2.93	28.95	7.40	60
21	0.48	5.23	12.5	0
22	1.63	13.0	9.00	10
23	1.04	10.33	7.80	5
24	2.51	13.1	10.4	30
25	2.63	24.71	4.80	30
26	3.98	41.82	6.20	32
7	4.38	37.47	4.00	50
28	4.45	39.17	5.80	108
29	4.87	9.95	5.6	19
30	3.63	40.42	6.2	162
31	3.19	40.46	3.9	44
32	4.49	23.31	4.5	59

4. Discussion and conclusion

Acoustic perturbation analyses have been used in determining phonatory stability characteristics [12]. In this study, vowel segment selection is of particular interest to obtain a reliable perturbation measure. Moving window was used to identify the most stable portion in the waveform objectively. When moving window was applied to estimate the perturbation analysis, the minimum perturbation (minimum %jitter, minimum %shimmer)

and maximum SNR values were obtainedby going through from voice onset to offset, showing that those values were always lowest %jitter, % shimmer , and highest SNR value in each given voice signal. Some researchers usedthe mid-portion of vowels in their studies as a stable sound source in a vowel for the perturbation analysis [11-13]. Apparently, in the present study, however, the mid-point selection from the waveform didn't provide any evidence of the most stable portion in a given voice signal. Although only periodic or nearly periodic voice signals can be suitable for reliable perturbation analysis [4], the moving window could allow getting reliable perturbation measures for some severe aperiodic pathologic voices by identifying the most stable portion which generated err less than 10.

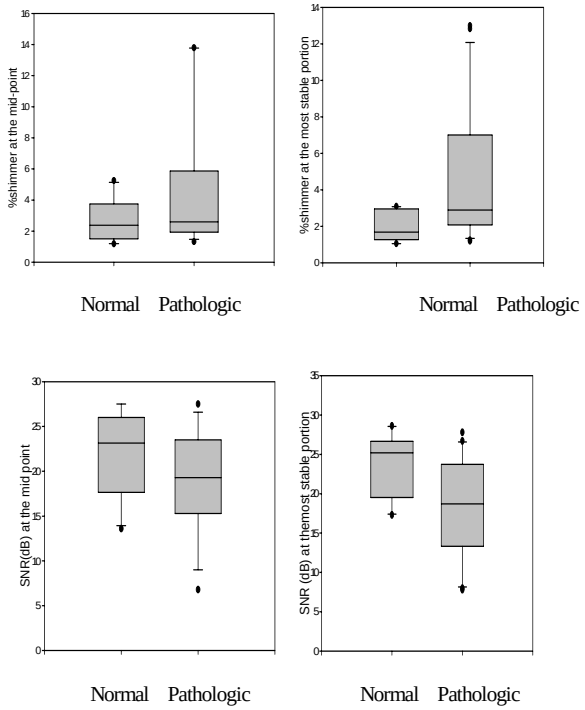

Figure 2. Box plots for % shimmer and SNR values at the mid-point selection (Left) and at the most stable portion selection (Right). The midline represents the median, with the lower and upper boundaries of the box indicating the first and third quartile, respectively.

In this study, the voice signals which demonstrated Errormore than 10 were not included in comparing the normal and pathologic voices to get the more reliable and accurate perturbationacoustic measures. Our results revealed that when we chose the voice signals which have Error less than 10 in the perturbation measures, %jitter, %shimmer and SNR values at mid-point segment in the vowel signal did not distinguish pathologic voice signals from the normal voice signals. In contrast, based on the moving window technique, % shimmer in the perturbation measures, and SNR values at the most stable portion presented significant differences between normal and pathologic voice signals when voice signals with Error less than 10 values were selected, which couldallow us for differential diagnosis between groups.

In conclusion, moving window can be potentially applicable for clinical practice for reliable and accurate perturbation acoustic analysis as an objective method of sample selectionby identifying the most stable portion in a given voice signal. The program generates a folder of the cut samples which can easily be run as a batch in *CSpeech* or *TF32* [7]. To clarify the clinical evidence supporting this technique, in the future research, larger data sets should be used to compare the different vowel segment selection methods with the most stable portions in the perturbation analysis. Developing effective technique for the practical implementation of acoustic perturbation analysis is highly essential for the advancement of acoustic voice analysis. Thus, moving window technique can be useful to increase the reliability of acoustic perturbation measures with an expansive scope including both clinicians and researchers.

References

[1] Mehta, D.D., & Hillman, R.E. (2008). "Voice assessment : updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Current Opinion in Otolaryngology & Head and Neck Surgery*, Vol. 16, No. 3, pp. 211-215.

[2] Milenkovic, P. (1987). "Least mean square measures of voice perturbation," *Journal of Speech and Hearing Research*, Vol. 30, pp. 529-538.

[3] Feijoo, S., & Hernandez, C. (1990). "Short-term stability measures for the evaluation of vocal quality," *Journal of Speech and Hearing Research*, Vol. 33, pp. 324-334.

[4] Titze, I. (1995). "Workshop on acoustic voice analysis : Summary statement," National Center for Voice and Speech, Denver, CO.

[5] Zhang, Y., Jiang, J.J., Wallace, S, M., & Zhou, L. (2005). "Comparison of nonlinear dynamic methods and perturbation methods for voice analysis," *Journal of Acoustic Society of America*, Vol. 118, pp. 2551-2560.

[6] Sprecher, A., Olszewski, A., Jiang, J.J., & Zhang Y. (2010). "Updating signal typing in voice: addition of type 4 signals," *Journal of Acoustical Society of America*, Vol. 127, No. 6, pp.

3710-3716.

[7] Milenkovic, P. (2001). *TF 32 User's Manual*, Madison, WI.

[8] Milenkovic, P., & Read, C. (1992). *CSpeech Version 4 User's Manual*. Madison, WI: University of Wisconsin.

[9] Regner MF, Tao, C., Zhuang, P, & Jiang JJ. (2008). "Onset and offset phonation threshold flow in excised canine larynges, "*Laryngoscope*, Vol.118, No. 7, pp. 1313-1317.

[10]Dejonckere, P.H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van, De Heyning, P., Remacle, M., Woisard, V., & Committee on Phoniatricsof the European laryngological Society (ELS). (2001). "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatricsof the European Laryngological Society (ELS)," *European Archieves of oto-rhino-laryngology*, Vol. 258, No. 2, pp. 77-82.

[11] Glaze, L.E., Bless, D.M., & Susser, R.D. (1990). "Acoustic analyisis of vowel and loudness differences in children's voice," *Journal of Voice*, Vol. 4, No. 1, pp. 37-44.

[12] Gelfer, M.P. (1995). "Fundamental frequency, intensity, and vowel selection : Effects on measures of phonatory stability," *Journal of Speech and Hearing Research*, Vol. 38, No. 6, pp. 1189-1198.

[13] Bielamowicz, S., Kreiman, J., Gerratt, B.R., Daier, M.S., 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.

• **Choi, Seong Hee,** Corresponding author
Address: Department of Communicative Disorder, Goodnight Hall, 1975 Willow Dr., Madison, WI, 53706 & Dept of Surgery, Division of Otolaryngology, University of Wisconsin Medical School, 5745a Medical Sciences Center, laryngeal physiology lab, 1300 University Avenue, Madison, WI, 53706
Telephone: 1- 714-309-6012 Email: choi@surgery.wisc.edu
Research Interests: voice disorder, swallowing disorder, tissue engineering, etc.
2007~ present Postdoctoral fellow
2009~ present Graduate School, Dept of Communicative disorders
Ph.D. Dept of Speech and Language Pathology, Yonsei Univ., 2006

• **Lee, JiYeoun**
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 Email: leeji@surgery.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, KIST, 2008.

• **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 Larygeal Physiology Lab.
Telephone : +1- 608-265-7888 E-mail : jjjiang@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, Univ. of Iowa, 1991.
M.D., Shanghai Medical Univ., 1983.