말소리와 음성과학 제2권 제2호 (2010)
pp. 109~115

Perturbation and Perceptual Analysis of Pathological Sustained Vowels according to Signal Typing

이 지 연[1] · 최 성 희 · Jiang, Jack J. · 한 민 수 · 최 홍 식
Lee, JiYeoun · Choi, Seong Hee · Jiang, Jack J. · Hahn, MinSoo · Choi, Hong-Shik

ABSTRACT

In this paper, we investigate a signal typing on the basis of visual impression of distinctive spectrogram. Pathological voices are classified into signal type 1, 2, 3, or 4 to estimate perturbation parameters and to mark perceptual rating based on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). The results suggest that perturbation analysis can be applied to only type 1 and 2 signals and the perceptual ratings of overall grade increase with each signal type, overall. A good inter-rater reliability is showed among three raters. We recommend that pathological voices should be marked the signal typing and CAPE-V, together, to definitely describe the characteristics of pathological voices.

Keywords: Signal typing, perturbation analysis, CAPE-V, perceptual rating

1. Introduction

In a 1995 summary statement from workshop on acoustic voice analysis, Titze proposed that signals should be assessed and categorized as type 1, 2, or 3 to determine whether a particular signal is appropriate for perturbation analysis [1]. In his system, type 1 signals are nearly periodic and therefore suitable for perturbation analysis. Type 2 signals contain strong modulations or sub-harmonics and type 3 signals are irregular and aperiodic. Such signals might not be appropriate for perturbation analysis [1-2]. Today, Titze's recommendations continue to be employed to determine the suitability of voice signals for perturbation analysis [3-6]. Recently, the addition of signal type 4 to Titze's voice classification scheme is proposed [7]. This signal type 4 is primarily stochastic in behavior and is therefore unsuitable for both perturbation and nonlinear dynamic analysis [7-8].

Recently, the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) has discussed and standardized as a tool for clinical auditory-perceptual assessment of voice. The CAPE-V has developed from a consensus meeting sponsored by the American Speech-Language-Hearing Association's (ASHA) Division 3: Voice and Voice Disorders, and the Department of Communication Science and Disorders, University of Pittsburgh, held in Pittsburgh on June 10-11, 2002. More detailed information has been given in [9-11].

The objective of this paper is to introduce signal typing and CAPE-V which have recently discussed in the United States of America (USA) and to look at their relation. We will classify pathological voices into type 1, 2,

3, and 4 in the basis of characteristic spectrogram patterns. Then, we will compare acoustic characteristics of pathological voices by using perturbation analysis (including jitter and shimmer), and signal-to-noise ratio (SNR) according to signal types. Finally, we will examine the relation between signal typing and perceptual rating using CAPE-V.

[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: leej@surgery.wisc.edu

Received May 13, 2010.
Revision June 17, 2010.
Accepted June 23, 2010.

2. Methods

2.1 Material

The pathological voice samples utilized in this study were selected from the Disordered Voice Database, model 4337, Version 1.03, developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. [12]. We selected 32 pathological subjects including 11 males and 21 females, ranging in age between 18 and 76 years from this database. The subject information is shown in <Table 1>, and more detailed information has been given in the Disordered Voice Database. Sustained vowel /a/ phonations (0.8-1.3 seconds in length) were used and all voice data were sampled at 44.1 kHz.

2.2 Spectrogram analysis

The signal typing was conducted during voice team meetings consisting of 3 trained speech-language pathologists. Complete agreement on the signal typing for each patient was required. For instances in which there was disagreement, discussion was held and data were reviewed until complete agreement was achieved. Signal typing was chosen to achieve a visual impression of the acoustic content of the voice samples. Narrow band spectrograms were generated using the Praat software version 5.1.02 (Website: www.praat.org, P. Boersma and D. Weenink, Amsterdam, Netherlands). Narrow band spectrograms were created with a window length of 50 millisecond, a time step of 0.002 seconds, a frequency step of 5Hz, and a dynamic range of 40dB. A hamming window shape was used to generate the spectrogram.

2.3 Perturbation analysis

The acoustic perturbation measures (percent jitter and percent shimmer) and SNR were obtained from the TF32 software [13]. 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. They were extracted from the whole recording of average 1 second in length excluding the onset and offset. The reliability of jitter, shimmer, and SNR was assessed using the TF32 generated values of “Trk” and “Err”. “Trk” provides an indication of the number of dramatic fluctuations in pitch, while “Err” quantifies large variations in the calculated fundamental frequency indicative of voice breaks [13].

Table 1. Subject information

Subject	Sex	Age (y)	Diagnosis	Signal typing
1	F	18	Vocal fold edema	Type 1
2	F	39	Abnormal vocal process	Type 1
3	M	29	Vocal fold polyp	Type 1
4	F	18	Vocal nodules	Type 1
5	F	25	Vocal nodules	Type 1
6	F	24	Vocal fold edema	Type 1
7	F	21	Nodular swelling	Type 1
8	F	34	Vocal nodules	Type 1
9	F	49	Vocal fold edema	Type 2
10	F	25	Vocal fold edema	Type 2
11	F	31	Polypoid degeneration	Type 2
12	M	40	Scarring	Type 2
13	F	38	Keratosis / leukoplakia	Type 2
14	M	38	Bowing / sulcus vocalis	Type 2
15	M	42	Keratosis / leukoplakia	Type 2
16	F	42	Vocal fold edema	Type 2
17	F	50	Chronic laryngitis	Type 3
18	F	61	Vocal fold polyp	Type 3
19	F	75	Parkinson's disease	Type 3
20	F	43	Polypoid degeneration	Type 3
21	M	76	Vocal fold polyp	Type 3
22	F	65	Vocal fold polyp	Type 3
23	M	39	Keratosis / leukoplakia	Type 3
24	F	32	Paralysis	Type 3
25	M	69	Paralysis	Type 4
26	M	49	Paralysis	Type 4
27	M	53	Paralysis	Type 4
28	M	52	Paralysis	Type 4
29	F	38	Spasmodic dysphonia	Type 4
30	F	40	Generalized edema of larynx	Type 4
31	F	47	Keratosis / leukoplakia	Type 4
32	M	29	Papilloma	Type 4

2.4 Perceptual analysis

All of the voices were subjectively rated. Each voice sample was independently rated and required to complete the ratings only once by three certified raters. Raters were blinded to the voice type and pathology. Each rater could replay a sample as many times as necessary to determine a rating. The CAPE-V was used to evaluate overall grade.

After the clinician has completed all ratings, he or she should measure ratings from each scale. To do so, he or she should physically measure the distance in mm from the left end of the scale. The mm score should be written in the blank space to the far right of the scale, thereby relating the results in a proportion to the total 100 mm length of the line. In CAPE-V, the results can be reported in two possible ways [9-11]. First, results can indicate distance in mm to describe the degree of deviancy, for example “73/100” on “strain.” Second, results can be reported using descriptive labels that are typically employed clinically to indicate the general amount of deviancy, for example “moderate-to-severe” on “strain.”

In here, pathological voices were described by the first rating method according to overall dysphonia: Moderate to severe degree of overall dysphonia (78/100), moderate roughness (56/100), moderate to severe breathiness (74/100) and strain (62/100). The three raters' scores were averaged and an overall mean score for each voice type was determined.

3. Results

3.1 Signal typing

<Figure 1> shows waveforms and spectrogram of type 1 signal. Waveforms of this signal were considered nearly periodic as shown in <Figure 1.(a) and (b)>. In <Figure 1.(c)>, the spectrogram for type 1 signal showed clearly defined, nearly straight harmonics of a variable number and spacing. Noise between harmonics was minimal in type 1 voice.

<Figure 2> shows waveforms and spectrogram of type 2 signal. In type 2 signal, noise between harmonics formed clearly defined subharmonics. In some cases, modulations caused the harmonics to appear wavy. Areas of subharmonics or modulations were often transient. Signals were rated type 2 if they contained one or more segments with a substantial lack of periodic structure in the waveform and modulation existed as shown in spectrogram of <Figure 2.(c)>.

<Figure 3> shows waveforms and spectrogram of type 3 signal. Type 3 signal showed a smearing of energy across multiple harmonics. Although the fundamental frequency was often apparent, higher harmonics could not be distinguished in <Figure 3. (b)>. Most of the harmonics were obscured by low frequency noise. Signals were rated type 3 if they contained segments of strong subharmonics, modulations, or other bifurcations (a sudden qualitative change in the pattern of the signal).

<Figure 4> shows waveforms and spectrogram of type 4 signal. Type 4 signal was characterized by complete absence of harmonics. It also showed a destroyed spectrogram so that we couldn't see any evidence showing subharmonics, modulations, and bifurcations. Finally, the type 4 signal was characterized by diffuse energy spanning the range of frequencies displayed.

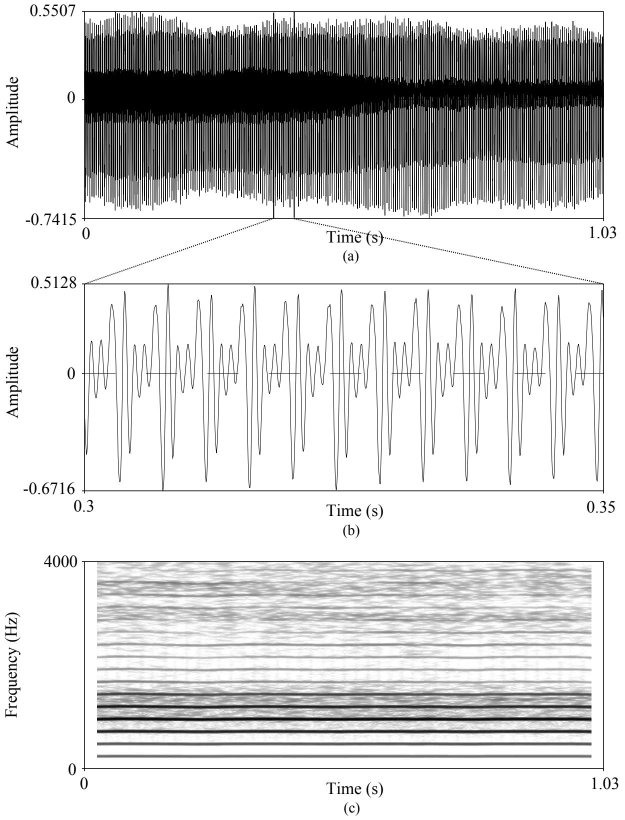

Figure 1. Signal type 1. (a) Whole waveform (b) an enlarged waveform of specific frame (0.3s – 0.35s) (c) spectrogram

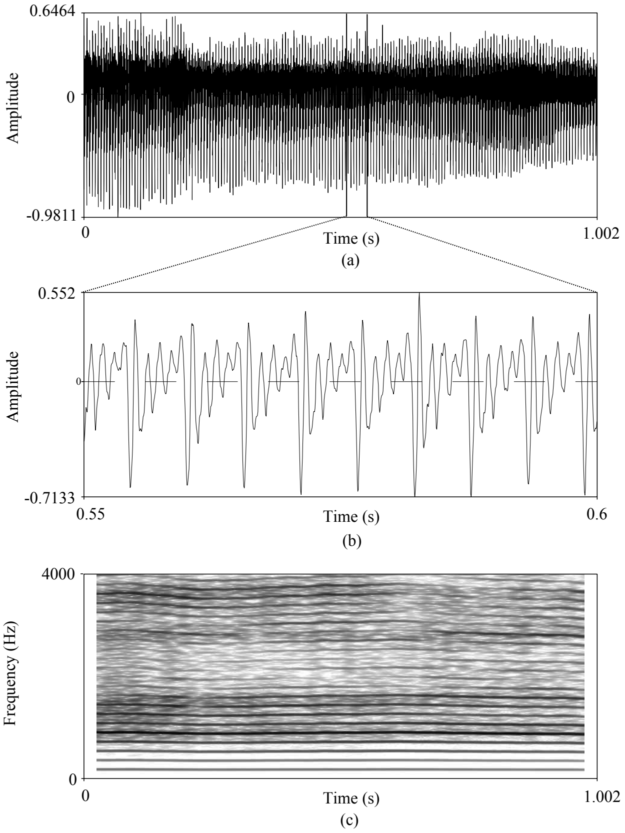

Figure 2. Signal type 2. (a) Whole waveform (b) an enlarged waveform of specific frame (0.55s – 0.6s) (c) spectrogram

3.2 Perturbation analysis

<Figure 5, 6, 7, and 8> show the "Trk", "Err", jitter (%), shimmer (%), and SNR (dB) estimated in signal type 1, 2, 3, and 4 signals, respectively. Both "Trk" and "Err" values increased significantly from type 1 to type 4 signals. Using our cutoff of "Err" less than 10, only type 1 and type 2 voices were appropriate for acoustic analysis. Both jitter (%) and shimmer (%)

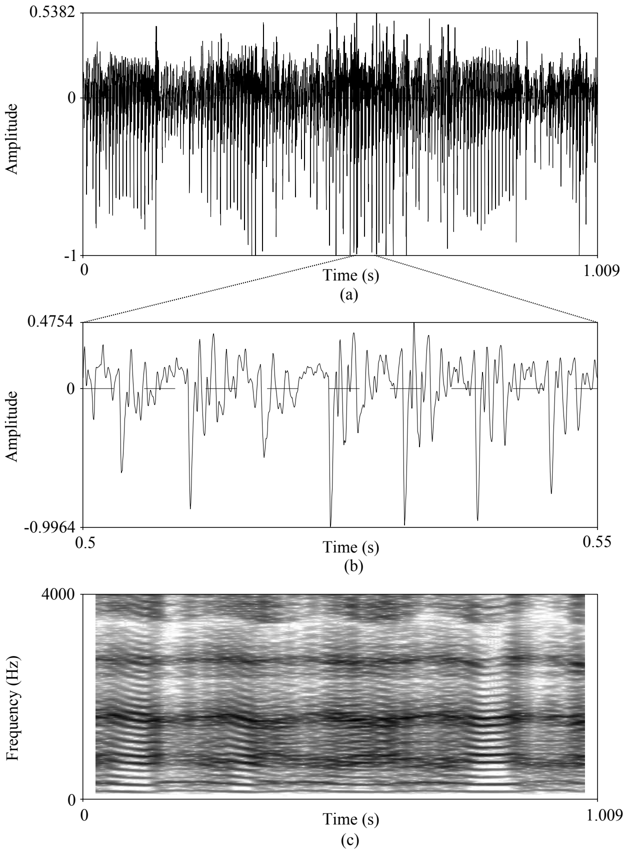

Figure 3. Signal type 3. (a) Whole waveform (b) an enlarged waveform of specific frame (0.5s – 0.55s) (c) spectrogram

increased with each voice type. Similarly, SNR decreased from type 1 through type 4 voices, indicating that the evidence of harmonics decreased as signal type increased.

3.3 Perceptual analysis

Results of perceptual analysis are given in <Figure 9> and <Table 2>. <Figure 9> shows that perceptual ratings of overall grade increased with each signal type. In <Table 2>, although some signals like subject 11, 13, 14, and 15 were determined to type 2, their perceptual ratings were similar to those of type 1 signals. The perceptual rating using CAPE-V of each signal type tended to be a little overlapped to neighboring signal type.

Inter-rater reliability analysis was conducted for the overall severity rating of voice impairment between the three certified raters who had completed CAPE-V ratings using the statistics package SPSS 12.0. A Pearson product moment correlation (r) was used to examine inter-rater reliability on the overall severity score on the CAPE-V. There was a statistically significant (P > 0.01) relationship among all raters' responses, with r ranging from 0.885 to 0.893 as shown in <Table 3>.

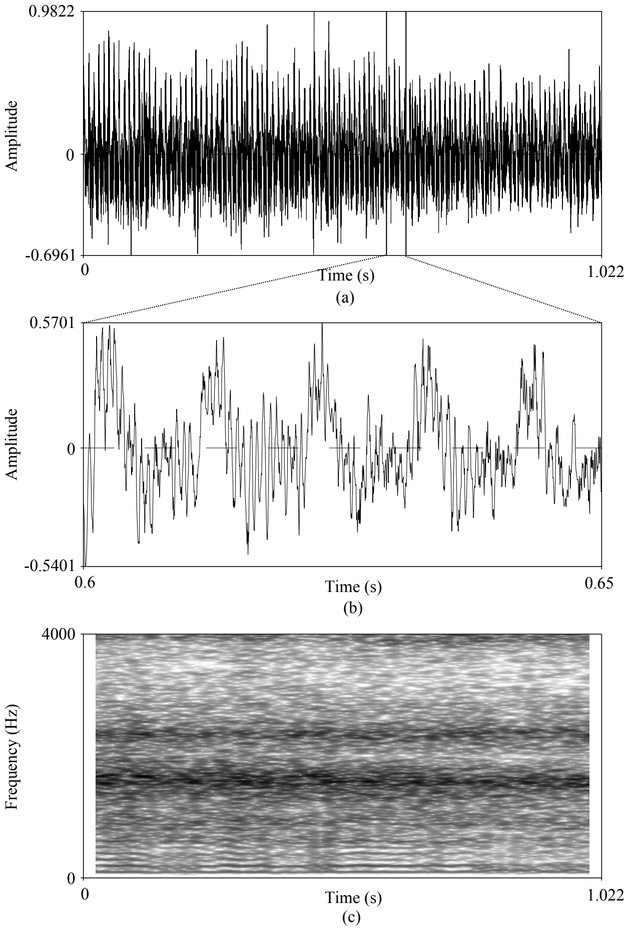

Figure 4. Signal type 4. (a) Whole waveform (b) an enlarged waveform of specific frame (0.6s – 0.65s) (c) spectrogram

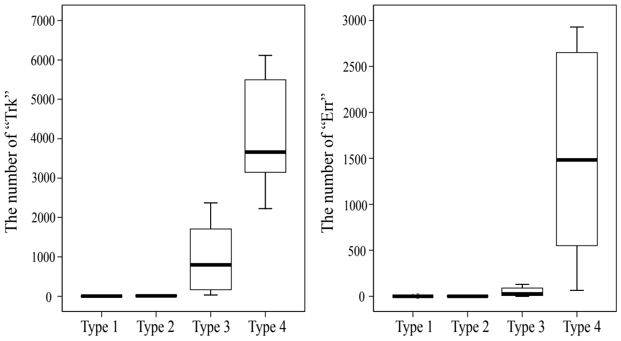

Figure 5. The "Trk" and "Err" distributions estimated in type 1, 2, 3, and 4 signals. The midline represents the median, with the lower and upper boundaries of the box indicating the first and third quartile, respectively.

4. Discussion

In this study we introduced a signal typing and CAPE-V as perceptual rating method compared their relation. The signal typing was chosen to achieve a visual impression of the acoustic content of the voice samples. Voice samples were classified into 1, 2, 3, or 4 on the basis of distinctive spectrogram patterns.

Next we applied perturbation and perceptual analysis to all voice samples. All measures indicated increasing disorder from the type 1 voices through the type 4 voices.

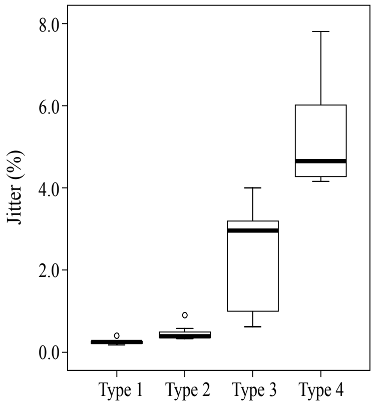

Figure 6. The distributions of jitter (%) estimated in type 1, 2, 3, and 4 signals. The midline represents the median, with the lower and upper boundaries of the box indicating the first and third quartile, respectively. Circle indicates the maximum value.

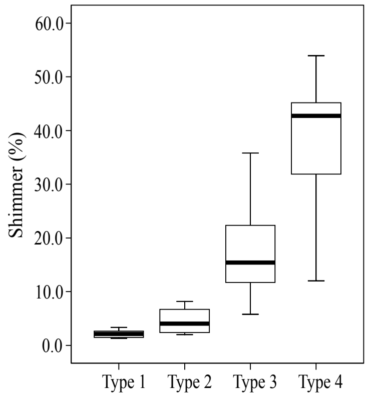

Figure 7. The distributions of shimmer (%) estimated in type 1, 2, 3, and 4 signals. The midline represents the median, with the lower and upper boundaries of the box indicating the first and third quartile, respectively.

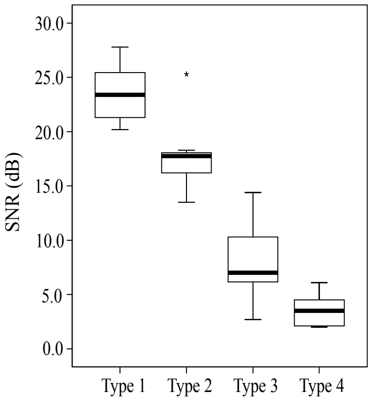

Figure 8. The distributions of SNR (dB) estimated in type 1, 2, 3, and 4 signals. The midline represents the median, with the lower and upper boundaries of the box indicating the first and third quartile, respectively. Whisker indicates the outlier.

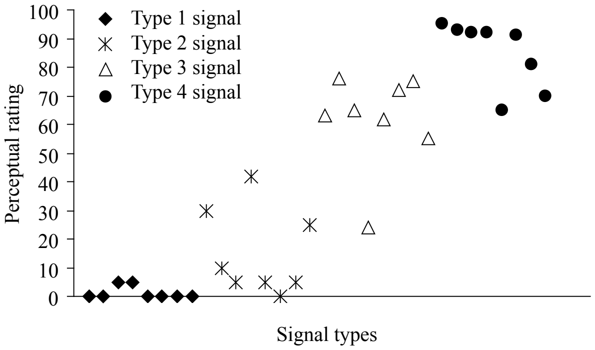

Figure 9. Distributions of perceptual ratings for each voice type.

Table 2. Mean scores of the perceptual ratings in each signal type.

Subject	Signal typing	Perceptual rating (mm)
1	Type 1	0 / 100
2	Type 1	0 / 100
3	Type 1	5 / 100
4	Type 1	5 / 100
5	Type 1	0 / 100
6	Type 1	0 / 100
7	Type 1	0 / 100
8	Type 1	0 / 100
9	Type 2	30 / 100
10	Type 2	10 / 100
11	Type 2	5 / 100
12	Type 2	42 / 100
13	Type 2	5 / 100
14	Type 2	0 / 100
15	Type 2	5 / 100
16	Type 2	25 / 100
17	Type 3	63 / 100
18	Type 3	76 / 100
19	Type 3	65 / 100
20	Type 3	24 / 100
21	Type 3	62 / 100
22	Type 3	72 / 100
23	Type 3	75 / 100
24	Type 3	55 / 100
25	Type 4	95 / 100
26	Type 4	93 / 100
27	Type 4	92 / 100
28	Type 4	92 / 100
29	Type 4	65 / 100
30	Type 4	91 / 100
31	Type 4	81 / 100
32	Type 4	70 / 100

In this study, we classified 32 voice samples into signal type 1, 2, 3, or 4 based on distinctive spectrogram patterns. Narrow band spectrograms were created with a window length of 50 millisecond, a time step of 0.002 seconds, a frequency step of 5Hz, and a dynamic range of 40dB. A hamming window shape was used to generate the spectrogram. The spectrogram of the type 1 signal showed clearly defined harmonics. In contrast to the type two sample, there was no evidence of subharmonics. A fundamental frequency was visible in the type 3 signal; however,

Table 3. Inter-rater reliability of the CAPE-V overall scores for each rater.

	Rater 1	Rater2	Rater3
Rater 1			
Pearson correlation	1	0.893†	0.892†
Sig. (2-tailed)		0	0
N	32	32	32
Rater 2			
Pearson correlation	0.893†	1	0.885†
Sig. (2-tailed)	0		0
N	32	32	32
Rater 3			
Pearson correlation	0.892†	0.885†	1
Sig. (2-tailed)	0	0	
N	32	32	32

† Correlation is significant at the 0.01 level (2-tailed).

most of the harmonics were obscured by low frequency noise. Finally, the type 4 signal was characterized by diffuse energy spanning the range of frequencies displayed.

The perturbation analysis suggests that type 1 and 2 signals only were suitable for acoustic analysis. Type 1 and 2 signal produced “Err” values below the cutoff of 10 and maintained similarly low values for “Trk”. However, the potential subharmonics in type 2 signals as shown in <Figure 2> suggest a need of the careful consideration for an application of perturbation analysis. Both “Trk” and “Err” increased significantly and showed large variations in type 3 and 4 signals because perturbation analysis is based on fundamental frequency in time domain. It is well known that nonlinear dynamic analysis (correlation dimension, D_2) was able to generate results for all type 3 signals [7-8]. The calculation of D_2 does not require a determination of fundamental frequency; therefore, it is unaffected by modulations in pitch or tracking errors [8]. Notably, the type 4 voices could not be quantified using correlation dimension. There is a present no objective method for evaluating these voices.

CAPE-V was used to evaluate overall grade of pathological voices [9-11]. After the clinician has completed all ratings, he or she should measure ratings from each scale in a proportion to the total 100 mm length of the line. In this investigation we found the perceptual ratings of overall grade increased with each signal type: however, the perceptual ratings of each signal type tend to be a little overlapped to those of neighboring signal types. In inter-rater reliability analysis, a Pearson correlation showed a higher value among three raters’ opinions, with ranging from 0.885 to 0.893. To clearly describe the characteristics of pathological voices, we recommend that the pathological voices should be marked along with signal typing and perceptual rating using CAPE-V as the following description: signal type 1 (5/100) or signal type 4 (95/100).

5. Conclusion

We introduce a signal typing and perceptual rating method. Pathological voices are classified into 1, 2, 3, or 4 to present perturbation analysis and perceptual rating based on CAPE-V. Using “Err” value of TF32 software, we determine that perturbation analysis can be applied to type 1 and 2 signals. However, the potential for period doubling in type 2 signals suggests a need for caution in the calculations of perturbation measures. Type 3 and 4 signals cannot be quantified using perturbation analysis: however nonlinear dynamic analysis can be applied to type 3 signals. Although the perceptual ratings of overall grade using CAPE-V increased with each signal type, we suggest that pathological voices should be marked the signal typing and CAPE-V, together. A good inter-rater reliability is showed among three raters, which presents the relationship between CAPE-V and signal typing. However, intra-rater reliability testing of three raters’ rating is not done because the raters are required to complete the ratings only once per voice sample.

References

[1] Titze, I. R. (1981). "Workshop on acoustic voice analysis: summary statement", National Center for Voice and Speech, pp. 1-36.

[2] Behrman, A., Agresti, C. J., Blumstein E., & Lee, N. (1998). " Microphone and Electroglottographic Data from Dysphonic Patients: Type 1, 2 and 3 Signals", *Journal of Voice, Vol.* 12, No. 2, pp. 249-260.

[3] Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, 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.

[4] Vieira, M. N., McInnes, F. R., & Jack, M. A. (2002). "On the influence of laryngeal pathologies on acoustic and electrographic jitter measures", *Journal of the Acoustical Society of America, Vol.* 111, No. 2, pp. 1045 – 1055.

[5] Shaw, H. S., & Deliyski, D. D. (2008). "Mucosal wave: a

normophonic study across visualization techniques", *Journal of Voice, Vol.* 22, No. 1, pp. 23-33.

[6] Zhang, Y., & Jiang, J. J. (2008). "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies", *Journal of Voice, Vol.* 22, No. 1, pp. 1-9.

[7] Sprecher, A., Zhang, Y., & Olszewski, A. (2010). "Updating signal typing in voice: addition of type 4 signals", *Journal of the Acoustical Society of America* (article in press).

[8] Zhang, Y., & Jiang, J. J., (2003). "Nonlinear dynamic analysis of signal typing of pathological human voices", *Electronics Letters, Vol.* 39, No. 13, pp. 1021-1023.

[9] American Speech-Language-Hearing Association: consensus auditory-perceptual evaluation of voice (CAPE-V). Available at: www.asha.org/uploadedFiles/members/divs/D3CAPEVprocedures.pdf.

[10] Kempster, G. B. Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). "Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol", *American Journal of Speech-Language Pathology, Vol.* 18, No. 2, pp. 124-132.

[11] Braden M. N., Johns, M. M. 3rd, Klein, A. M., Delgaudio, J. M., Gilman, M., & Hapner, E. R. (2010). "Assessing the effectiveness of botulinum toxin injections for adductor spasmodic dysphonia: Clinician and patient perception", *Journal of Voice, Vol.* 24, No. 2, pp. 242-249.

[12] Kay Elemetrics Corp. (1993). "Multi-dimensional voice program: software instruction manual", Pine Brook, NJ, Kay Elemetrics Corp.

[13] Milenkovic P. (2001). *TF32 user's manual.* Madison, WI.

• 이지연 **(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 Larygeal Physiology Lab.
Telephone : +1-213-598-4410
E-mail : 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, KAIST, 2008.

• 최성희 **(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 Larygeal Physiology Lab.
Telephone : 1-714-309-6012
E-mail : Choi@surgery.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 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, Uni. of Iowa, 1991.
M.D., Shanghai Medical Uni., 1983.

• 한민수 **(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

• 최홍식 **(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