Fundamental Frequency Characteristics using Moving Window Method for Korean Elderly Voices

JiYeoun Lee
Department of Biomedical Engineering, Jungwon University
Donbu-ri #5, Goesan-eup, Goesan-gun, Chungcheongbuk-do 367-805, South Korea
jylee@jwu.ac.kr

Abstract— **It is necessary to provide the individual services with elderly voice in health medical tools for the elderly who aren't accustomed to manage IT equipment. This study focuses on acoustic analysis by means of moving window method in an entire portion of the given elderly signal. Acoustic parameters based on fundamental frequency such as Jitter (%), Shimmer (%), and Signal-to-noise ratio (SNR) are calculated in each window frame using the TF32 software. The distributions estimated from female and male elderly signals overall exists abnormal ranges. Also variation between frames tends to be greater than that of normal voices. It can guess that the acoustic results vary with depending on how to set the analysis region. Therefore the decision method by the minimum perturbation with moving window method for elderly signals is recommended in this research. Moving window analysis only detects the areas where "Err" value dropped below 10 or 20. If the acoustical parameters are estimated in the same frame, the time point is decided as the analysis section. If they have a different minimum perturbation location, then various situations must be considered for the selection of analysis section. Future investigations will develop analysis methods to improve the performance for elderly voices.**

Keyword- Elderly voice, Disordered voice, Fundamental frequency, Acoustic analysis, Minimum perturbation, Moving window method.

I. INTRODUCTION

Nowadays, voice recognition technology has emerged as a key sector which can be applied for medical field. It is possible to provide the individual services in health medical field, because many people can apprehend a psychological and physical condition, linguistic ability, and emotion with speaker's voice [1]. Therefore, researchers should be aware of the necessity of products which can handle the medical instruments with an elderly voice in the view of the elderly who aren't accustomed to manage IT utility. Therefore, they must develop the products to deliver accurate information with elderly voices as well as normal voices.

Although we can analyse the elderly voices using various voice analysis software such as Dr.Speech, TF32, Praat, and etc, they have disadvantage of performance degradation caused from different unsolved problems such as non-stationary noise, inaccurate fundamental frequency extraction, and etc [2]-[7]. In the level of clinical circumstance, it can be needed to apply for the user's viewpoints to actively utilize the software using elderly voice.

Generally, many studies have applied acoustic analysis including perturbation measures of jitter (%), shimmer (%), and signal to noise ratio (SNR), to Korean elderly voices [8]-[11]. Most elderly voices exhibit irregular or aperiodic waveforms leading to the elevated and unstable perturbation values. As a result, variability in the sentence makes the method of selecting an analysis segment vitally important for the acoustic analysis [4][7]. Although the general consensus is to avoid the negative effects of onset, offset, and silence of phonation, the location and length of the selected segment varies and results in different perturbation values among studies [7]. Many researchers select a portion of the voice signal visually deemed to be most stable [12]-[13]. Visual selection often focuses on amplitude variability; however, frequency may vary independent of amplitude and the samples that selected using this method may not represent the most stable section of the sample with regard to frequency. Recently, two papers related with moving window are already published and the methods are substantially referred in this paper [4][7].

This study proposes the moving window method as an objective and reliable method of sample selection for Korean elderly voices. It utilizes minimum perturbation value of fundamental frequency based measures such as jitter (%), shimmer (%), and SNR (dB) to extend perturbation analysis. It also investigates the impact of the moving window on perturbation measures generated from these voice sample and compare minimum perturbation with average values of the acoustic parameters.

II. RESEARCH METHODS

A. Database

The elderly voice samples used in this research were collected in The Speech Information Technology&Industry Promotion Center (SiTEC). The database includes the elderly voices of 20 Korean subjects (10 female and 10 male) ranging in age from 70 to 80 years. Two sentences were used and all voice data were sampled at 22.5 kHz. The information is detailed in Table Ⅰ. Silence sections are manually deleted using information such as structure of waveform and spectrum.

TABLE Ⅰ. Elderly database. Displayed are number, age range, and the used two sentences.

Elderly voices	Number	Age	Korean Sentences
Female	10	70–79	1. 그때 누가 그녀의 책상 앞으로 다가왔다. 2. 그때 웬 낯선 사람이 다가와 물었다.
Male	10	70–78	1. 그때 누가 그녀의 책상 앞으로 다가왔다. 2. 그때 웬 낯선 사람이 다가와 물었다.

B. Waveform and spectrogram

Narrowband spectrograms were generated by using the Praat software. They were created with a window length of 0.05 s, time step of 0.002 s, frequency step of 5 Hz, and dynamic range of 40 dB. A hamming window shape was used to generate the spectrograms. Above method to generate narrowband spectrogram is frequency used in many researches and is quoted in [4][7]. The typical waveform and spectrogram characteristics for elderly signals are shown in Figs. 1 and 2. Figure 1 shows the typical waveforms of female and male elderlies. The waveforms become increasing complex and are non-periodic. In the waveform of Fig. 1 (a) and (b), the periodic aspect of the waveform is different from the left to right patterns, respectively.

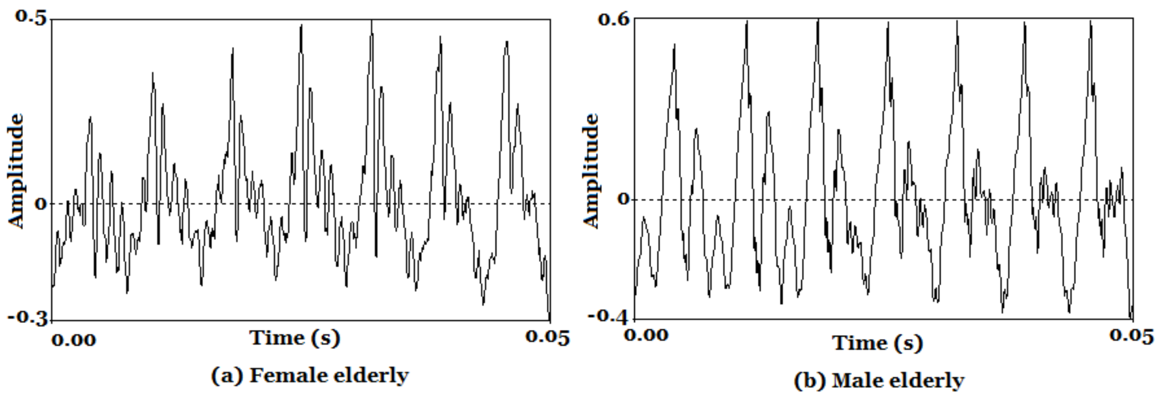

Fig. 1. The typical waveforms of female and male elderlies.

In Fig. 2(a) and (b), the spectrograms of female and male elderly signals show subharmonic or modulating frequencies in the overall sections. Although the fundamental frequency is often apparent, it also shows a smearing of energy across multiple harmonics.

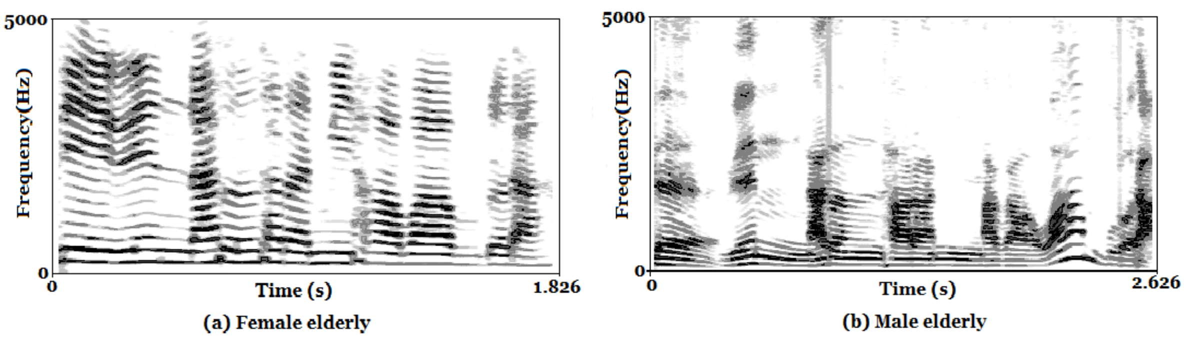

Fig. 2. The typical spectrograms generated from female and male elderlies.

C. Moving window method

A window of 0.5s length was used and was shifted forward in 0.025s increments across the entire signal as shown in Fig. 3[4][7]. Perturbation parameters such as jitter (%), shimmer (%), and Signal-to-noise ratio (SNR) were calculated in each window frame using the TF32 software [14]. Generally, the reliability of jitter (%), shimmer (%), and SNR(dB) is assessed using the values of "Trk" and "Err" that generated through TF32. The "Trk" count means the number of sharp fluctuations in pitch and "Err" quantifies the number of voice breaks between the samples. A large number of "Err" are an indication that a sample is ill-suited for acoustic analysis. Above method to move windows is quoted in [4][7].

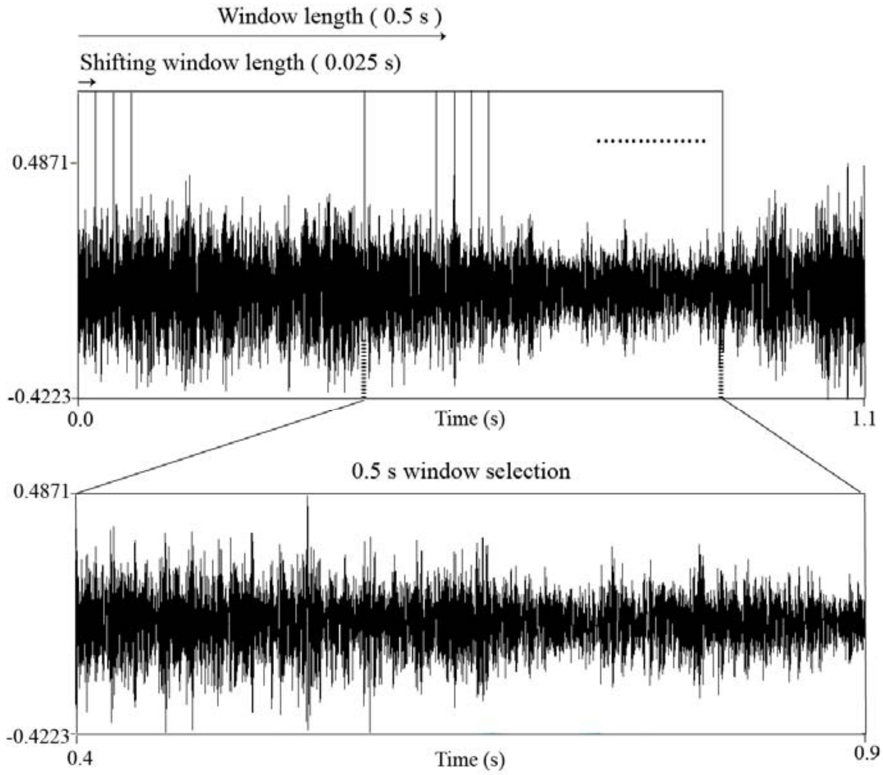

Fig. 3. Selection procedure of analysis segment through windows[4][7]

D. Minimum perturbation

Perturbation measures such as jitter (%) and shimmer (%) mean variation of the frequency and amplitude, respectively [15]. Therefore, the smallest value of the measures in entire speech sample indicates the most stable segment. In this study, the section which is ideally in the minimum value of jitter (%), shimmer (%), and SNR (dB) in the same time is selected as section to analyse the elderly voices. When the values are in some speech section simultaneously, the interval will be determined for the acoustic analysis of the elderly voices.

III. RESULTS

Figure 4 shows "Trk" and "Err" estimated from the whole phonation. These figures were obtained by calculating the minimum, first quartile, median, third quartile, maximum values, and outliers. Both "Trk" and "Err" values increased dramatically in elderly signals. By the pre-defined cutoff of an "Err" less than 10, elderly signals were not appropriate for acoustic analysis using TF32.

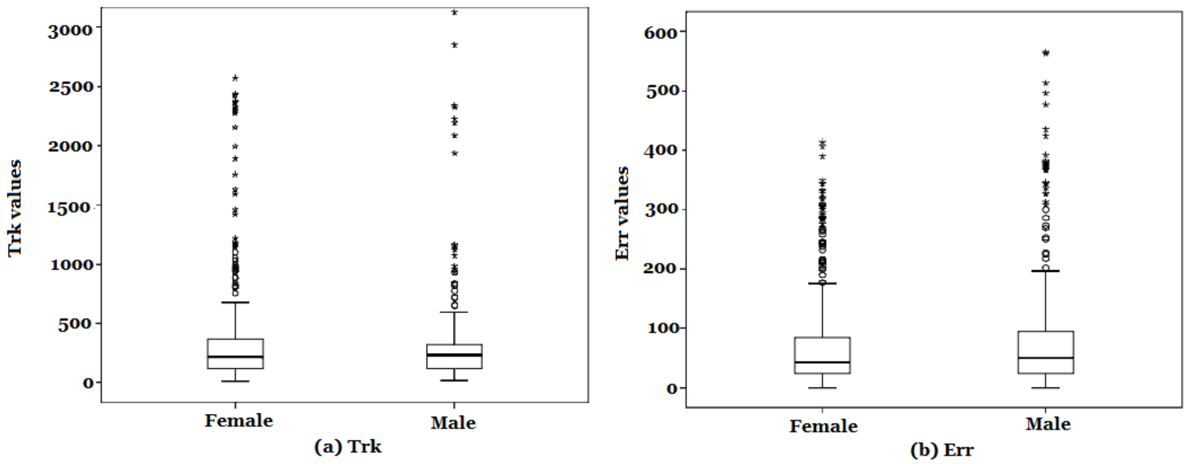

Fig. 4. "Trk" and "Err" estimated from elderly signals.

Figure 5 (a), (b), (c), and (d) represent "Err", jitter (%), shimmer (%), and SNR (dB) distributions estimated from female elderly signals(subject 1~5) using the moving window method, respectively. When the normal range of jitter (%) is 0.0~0.6, the overall values exists abnormal ranges in female elderly signals (subject 1~5). When the normal range of shimmer (%) is 0.0~6.5, the overall values also exists abnormal ranges in female elderly signals (subject 1~5). Also variation between frames tends to be high in jitter (%) and shimmer (%) distributions. Considering lower SNR (dB) distribution, noises are included in voice signals.

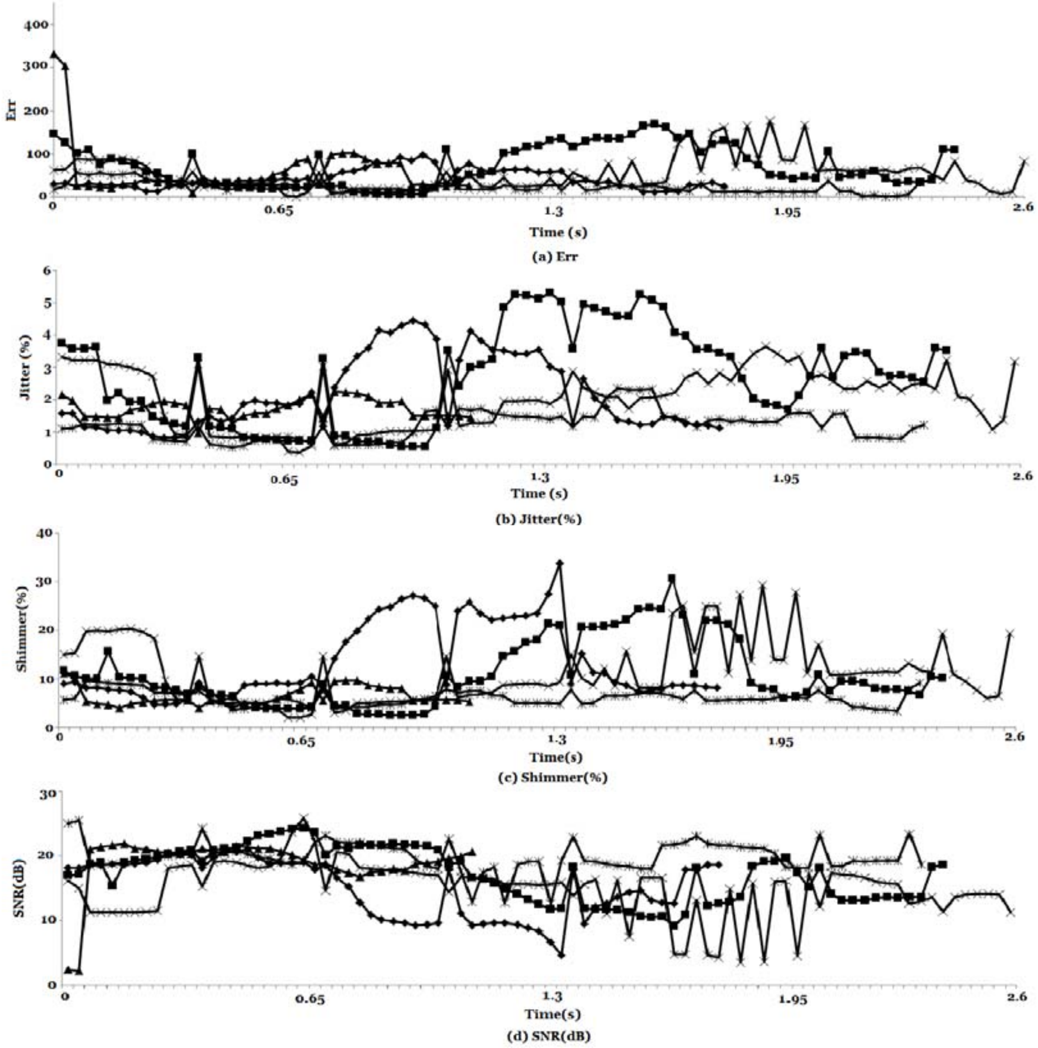

Fig. 5. Distributions estimated from female elderly signals using the moving window method (—◆— : subject 1, —■— subject 2, —▲— : subject 3, —×— : subject 4, —✳— : subject 5)

Figure 6 (a), (b), (c), and (d) show “Err”, jitter (%), shimmer (%), and SNR (dB) distributions estimated from female elderly signals(subject 6~10) using the moving window method, respectively. The range of jitter (%) and shimmer (%) overall exists abnormal ranges in female elderly signals (subject 6~10). Also variation between frames tends to be greater than distributions of normal voices. Considering lower SNR (dB) distribution, noises are blended in voice signals. Therefore author thinks that characteristics of the elderly voice are similar to those of the pathological voice or disordered voice. Deep valleys are generally shown in Fig.5. The points are boundaries between silence and voice section. So the information can be used for classification between silence and voicing.

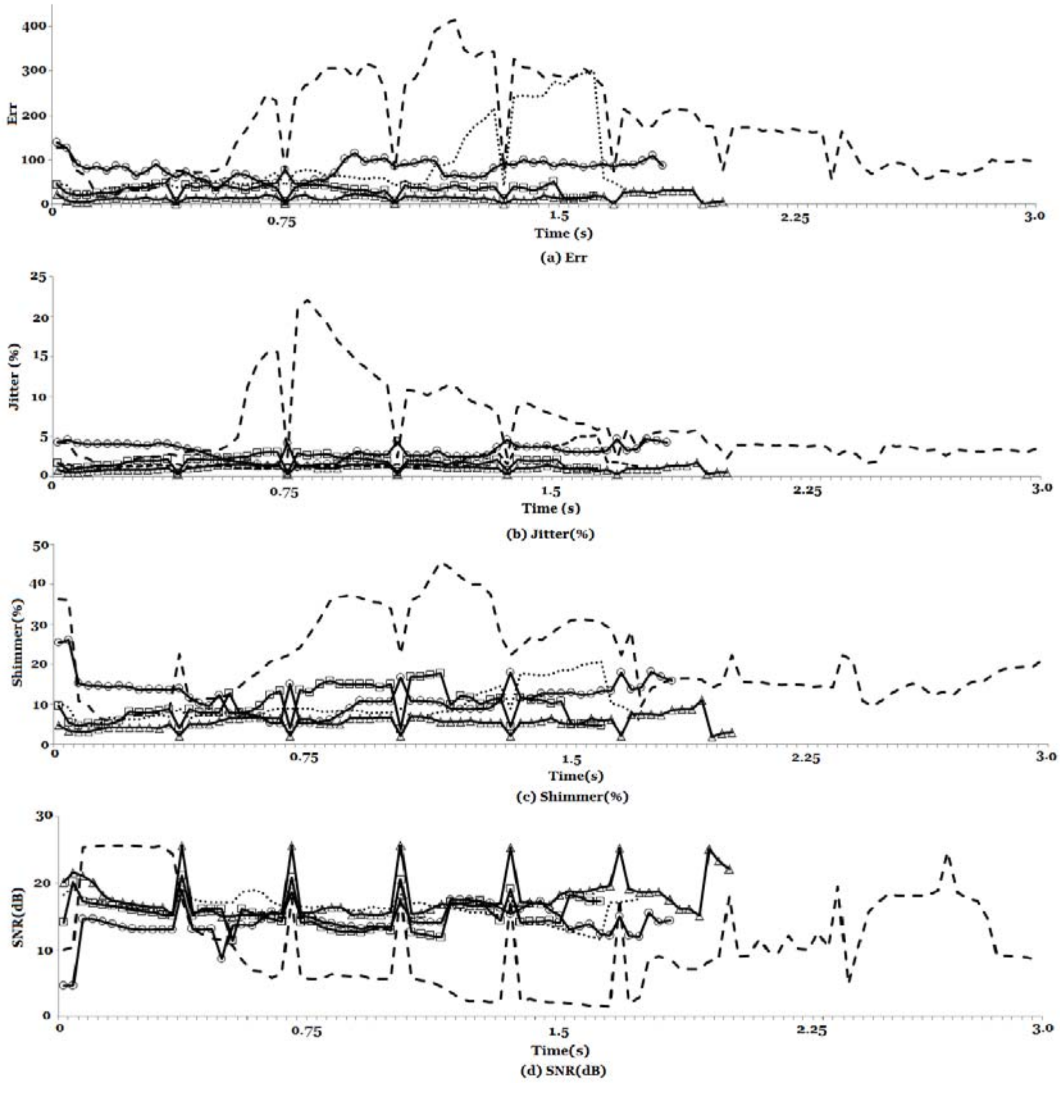

Fig. 6. Distributions estimated from female elderly signals using the moving window method (.... subject 6, —○— : subject 7, —△— : subject 8, – – : subject 9, —□— : subject 10)

Figure 7 (a), (b), (c), and (d) show “Err”, jitter (%), shimmer (%), and SNR (dB) distributions estimated from male elderly signals(subject 1~5) using the moving window approach, respectively. The range of jitter (%) and shimmer (%) overall are abnormal in male elderly signals (subject 1~5). Considering lower SNR (dB), noises are mixed in voice signals. Most values of acoustic parameters extracted in female elderly tend to be higher than ones estimated in male elderly. Also rate of change between frames tends to be high in jitter (%), shimmer (%), and SNR (dB).

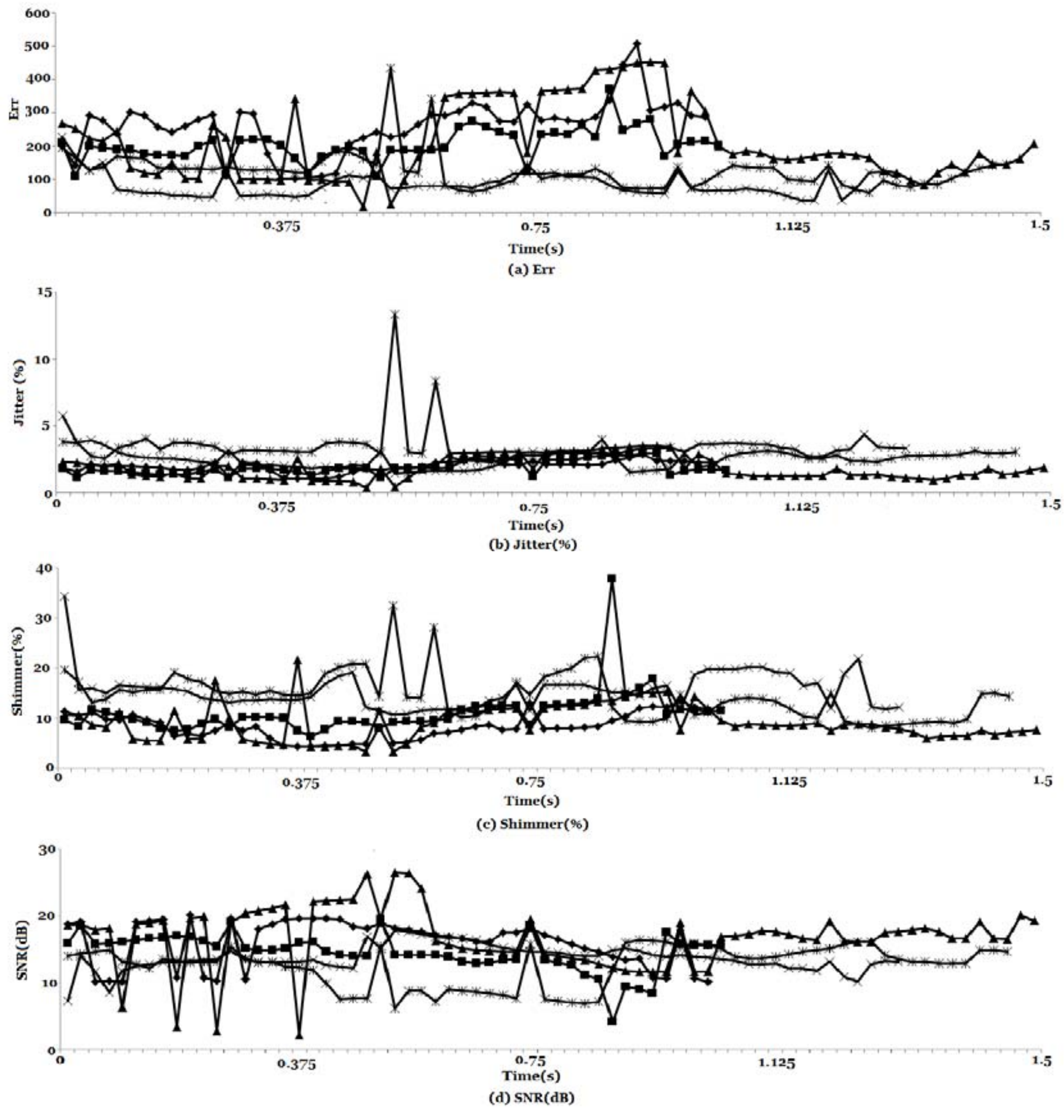

Fig. 7. Distributions estimated from male elderly signals using the moving window method (—◆— : subject 1, —■— subject 2, —▲— : subject 3, —✕— : subject 4, —✱— : subject 5)

Figure 8 (a), (b), (c), and (d) show “Err”, jitter (%), shimmer (%), and SNR (dB) distributions estimated from male elderly signals(subject 6~10), respectively. The values of jitter (%) and shimmer (%) overall are abnormal in male elderly signals (subject 6~10). Also variation between frames tends to be high in jitter (%), shimmer (%), and SNR (dB) distributions. Noises are included in voice signals by observing lower SNR (dB) distribution. Sometimes, the extreme values are shown in jitter (%) and shimmer (%) distribution.

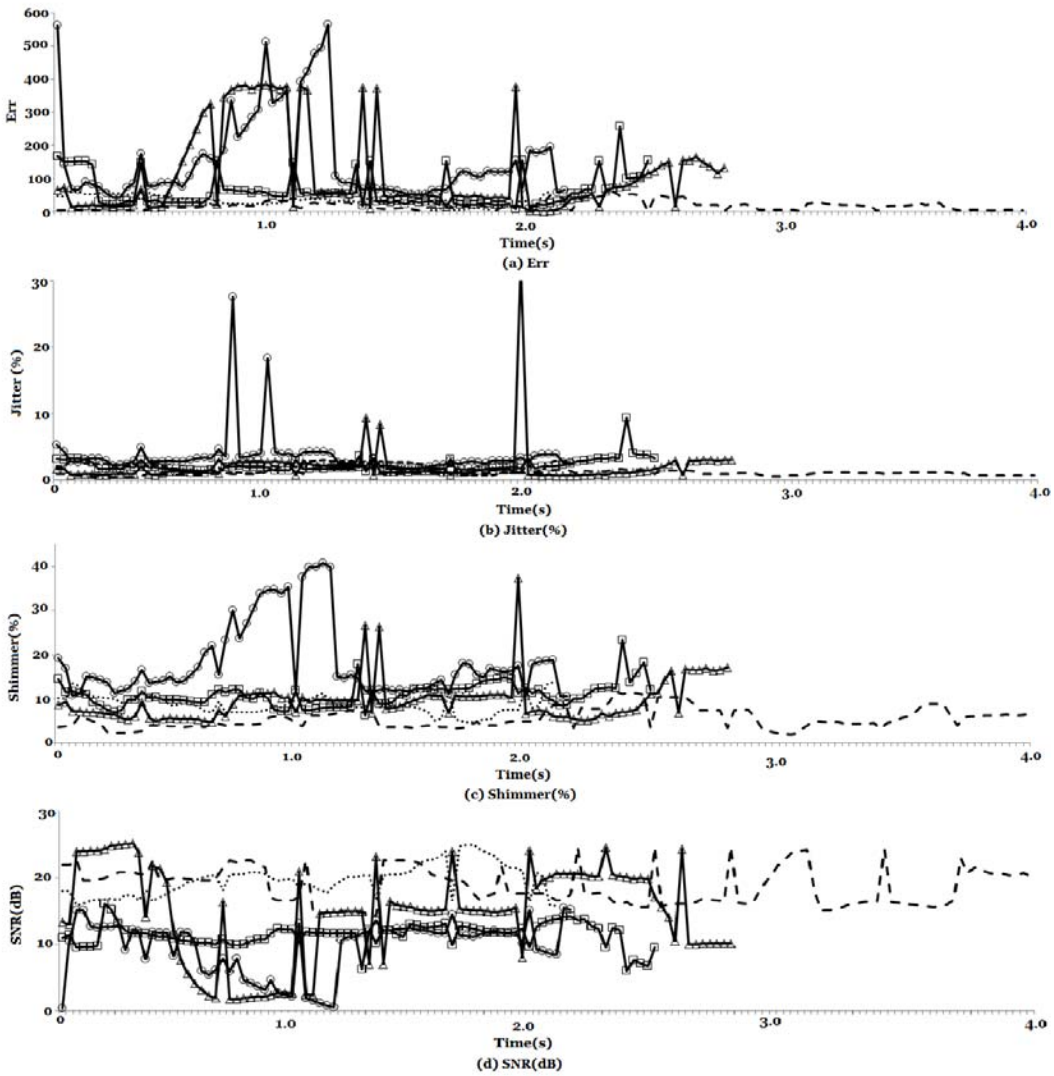

Fig. 8. Distributions estimated from male elderly signals using the moving window method (.... subject 6, : subject 7, : subject 8, - - : subject 9, : subject 10)

Table Ⅱ shows the minimum perturbation of acoustic parameters estimated in each elderly signal and the selected analysis sections. Moving window analysis firstly detects the areas where "Err" value dropped below 10 and the segment could be used for decision of the analysis section in this study. If there aren't the areas where "Err" value dropped below 10, then the areas where "Err" value dropped below 20 are secondly chosen as analysis section. Finally if all frames are where "Err" value is more than 20, then no section is detected. For female elderly voices such as subject 7 and male elderly voices such as subject 4, 5, and 7, there are no frames below 20 of "Err" values. So it is indicated as 'x'.

For female elderly voices such as subject 3~6 and 8~10 and male elderly voices such as subject 2, 3, and 6, all samples achieved minimum values for the three parameters (jitter (%), shimmer (%), and SNR (dB)) in same places. For female elderly voices such as subject 1, 2 and male elderly voices such as subject 1, 3, 8, 9, and 10, at least two parameters achieved minimum values among the three parameters in same places. Although acoustic parameters estimated in "Err" less than 10 are appropriate for acoustic analysis, the elderly signals tend to be higher "Err" than normal signals.

Recommendation of the minimum perturbation decision for moving window method is as follows:

1. If the acoustical parameters such as jitter (%), shimmer (%), and SNR (dB) are estimated in the same location, the time point is the analysis section.
2. If the acoustical parameters have similar minimum perturbation location, the location corresponding to more than two acoustical parameters is the time point for the sample selection.
3. If the acoustical parameters have a different minimum perturbation location, then various situations must be considered for selection of analysis section.

TABLE Ⅱ. Minimum perturbation extracted in each elderly signal and the selected analysis sections

Subject	Female/Male	Jitter (%)	Shimmer (%)	SNR (dB)	Selected section
1	F	0.82 (0.225s)	4.62 (0.2s)	20.4 (0.225s)	0.2s ~ 0.225s
2	F	0.55 (0.775s)	2.46 (0.75s)	21.8 (0.75s)	0.75s ~ 0.775s
3	F	0.97 (0.3s)	4.09 (0.3s)	21.2 (0.3s)	0.3s
4	F	0.38 (0.525s)	1.94 (0.525s)	25.9 (0.525s)	0.525s
5	F	0.53 (0.375s)	3.64 (0.375s)	20.8 (0.375s)	0.375s
6	F	1.15 (0.05s)	4.55 (0.05s)	17.8 (0.05s)	0.05s
7	F	x	x	x	x
8	F	0.32 (0.3s)	2.01 (0.3s)	25.7 (0.3s)	0.3s
9	F	1.29 (0.125s)	6.17 (0.125s)	25.6 (0.125s)	0.125s
10	F	0.46 (0.3s)	4.07 (0.3s)	21.2 (0.3s)	0.3s
1	M	1.07 (0.4s)	4.27 (0.425s)	19.7 (0.475s)	0.4s ~ 0.475s
2	M	1.11(0.575s)	8.08 (0.575s)	19.5 (0.575s)	0.575s
3	M	0.39 (0.3s)	3.28 (0.3s)	26.3 (0.3s)	0.3s
4	M	x	x	x	x
5	M	x	x	x	x
6	M	0.59 (1.475s)	4.5 (1.475s)	25 (1.475s)	1.475s
7	M	x	x	x	x
8	M	0.6 (1.8s)	5.78 (1.825s)	20.7 (1.825s)	1.8s ~ 1.825s
9	M	0.43 (0.025s)	1.97 (2.625s)	24.2 (2.625s)	2.625s
10	M	1.02 (1.375s)	6.12 (1.1s)	12.3 (1.1s)	1.1s

Table Ⅲ shows the values of acoustic parameters estimated in an entire phonation except for onset, offset, and silence sections. Since it doesn’t consider the number of voice breaks between the samples, I think that the values have no credibility. “Err” values are greater than ones shown as in Table Ⅱ. Also the values of jitter (%) and shimmer (%) tend to be abnormal and SNRs(dB) is lower than normal condition.

TABLE Ⅲ. Jitter (%), Shimmer (%), and SNR (dB) estimated in an entire phonation

Female	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Jitter (%)	2.12	2.64	2.09	2.26	1.14	2.4	3.38	1.46	4.12	2.17
Shimmer (%)	12.38	11.77	8.6	15.49	5.69	11.35	13.87	7.18	41	17.78
SNR (dB)	14.2	15.7	6.7	10.9	20.6	17.4	8	16.6	4.6	11.1
Err	117	302	393	191	64	363	298	74	913	132
Male	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Jitter (%)	2.21	2.24	1.88	3.64	3.58	1.97	3.44	1.82	1.07	2.71
Shimmer (%)	9.94	12.81	8.93	24.38	17.54	9.15	20.01	9.67	5.56	19.96
SNR (dB)	17.9	12.5	17.1	10.5	10.5	19.2	3	9.1	18.4	7.5
Err	200	109	180	376	486	127	1130	583	113	378

Table Ⅳ presents average jitter (%), shimmer (%), and SNR (dB) calculated in frames less than “Err” 10 or 20. The values of the acoustic parameters are also estimated in an entire phonation except for onset, offset, and silence sections. Since the number of voice breaks between the samples is considered in this method, I think that the values have more believable than ones of Table Ⅲ. Minimum perturbations and average values of acoustic parameters are compared with Fig. 9. Most average values of acoustic parameters are greater than the values estimated in minimum perturbation and some are found in the same location. I think that I cannot determine at this time what method is more reliable. Importantly, a reference method is determined and it is used to analyze the characteristics of elderly voices.

TABLE Ⅳ. Average Jitter (%), Shimmer (%), and SNR (dB) calculated in frames less than “Err” 10 or 20

Female	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Jitter (%)	1.18	0.61	0.97	0.62	0.82	1.15	x	0.54	1.37	0.48
Shimmer (%)	6.83	2.61	4.09	3.25	3.8	4.55	x	2.96	6.33	3.94
SNR (dB)	15.51	21.7	21.2	21.2	19.28	17.8	x	22.3	25.6	20.87
Err	13	7	9	4.25	2.6	19	x	4.07	17.5	5.67
Male	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10
Jitter (%)	1.12	1.18	0.66	x	x	0.7	x	0.67	0.72	1.61
Shimmer (%)	4.39	7.83	3.81	x	x	5.07	x	6.68	4.06	9.61
SNR (dB)	19.6	18.35	26.4	x	x	24.18	x	19.96	21.70	12.43
Err	14	17.5	3.67	x	x	6.5	x	2.6	3.8	12

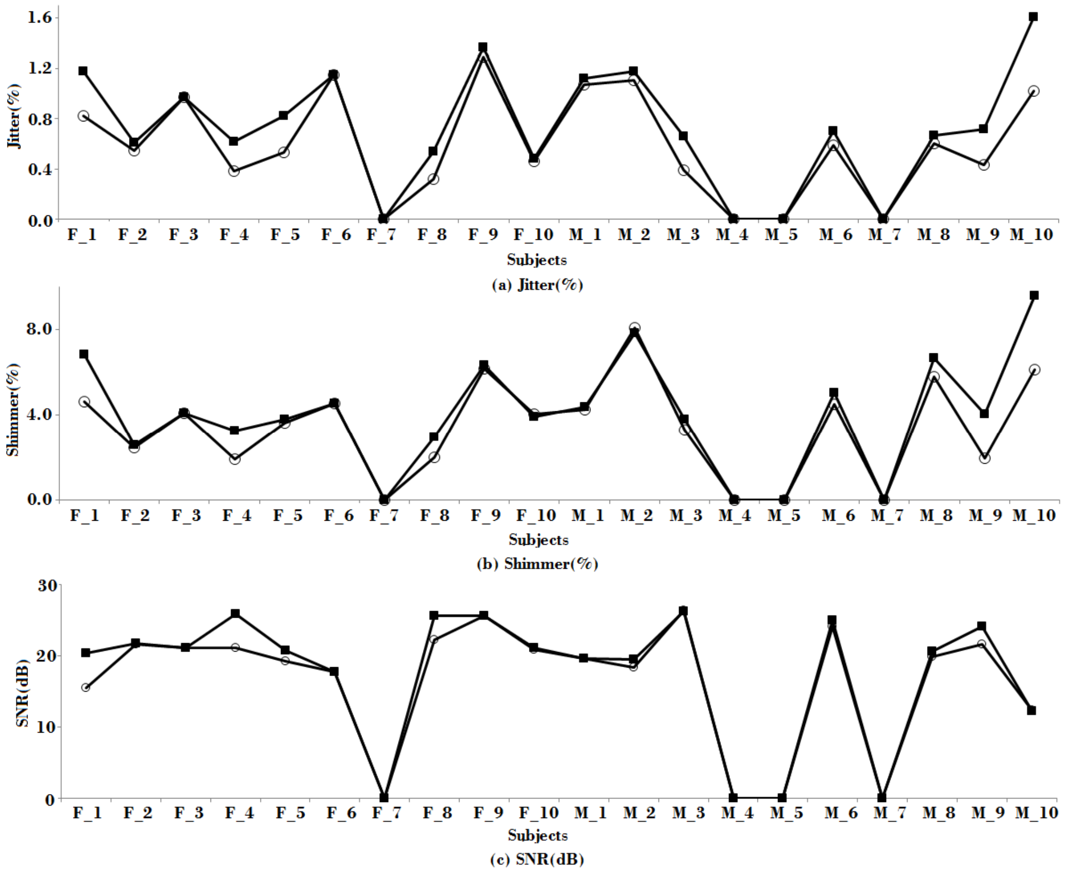

Fig. 9. Comparison between minimum perturbations and average values of acoustic parameters. ‘F’and ‘M’ mean female and male, respectively. (—■— : minimum perturbation, —○— : average value)

IV. CONCLUSION

It is important to provide the individual services with voice in health medical equipment. Especially, it is necessary to handle the medical instruments with an elderly voice for the elderly who aren't accustomed to manage IT utility. Furthermore researchers must develop the systems to deliver accurate information with elderly voices as well as normal voices.

A number of researches have been applied on the acoustic analysis with acoustic parameters such as jitter (%), shimmer (%), and SNR (dB) for the pathological and normal voices. Especially, since waveform and spectrum of elderly voices are unstable and irregular, it needs a reliable method to analyse correct information. This study focuses on acoustic analysis by means of moving window method using an entire portion of the given elderly signal. A window of 0.5s length is shifted forward in 0.025s increments across the entire signal except for onset, offset, and silence parts. Perturbation parameters such as jitter (%), shimmer (%), and Signal-to-noise ratio (SNR) are calculated in each window frame using the TF32 software[14]. "Err", jitter (%), shimmer (%), and SNR (dB) distributions estimated from female (subject 1~10) and male (subject 1~10) elderly signals using the moving window method overall exists abnormal ranges. Also variation between frames tends to be greater than distributions of normal voices. Considering lower SNR (dB), noises tend to be blended in voice signals. Therefore author thinks that characteristics of the elderly voice are similar to those of the disordered voice. The observed deep valleys seem to boundaries between silence and voice section. So the information can be used for classification between silence and voicing. As a result, the acoustic results very vary depending on how to set the analysis region.

Recommendation of the minimum perturbation decision with moving window method for elderly signals is as follows:

1. Moving window analysis firstly detects the areas where "Err" value dropped below 10
2. If minimum perturbation values of the acoustical parameters such as jitter (%), shimmer (%), and SNR (dB) are estimated in the same location, the time point is the analysis section.
3. Otherwise, the location corresponding to more than two acoustical parameters is the time point for the sample selection.
4. If it does not correspond to the above cases, then various situations must be considered for selection of analysis section and it is the problem to be solved later.
5. If there aren't the areas where "Err" value dropped below 10, then the areas where "Err" value dropped below 20 are secondly chosen as analysis section and step 2~4 are repeated.
6. Finally if all frames are where "Err" value is more than 20, then no section is detected.

As an experiment, average jitter (%), shimmer (%), and SNR (dB) calculated in frames less than "Err" 10 or 20 are compared with minimum perturbations of the above parameters. Most average values of acoustic parameters are greater than the values estimated in minimum perturbation and some are found in the same location. Although it is determined at this time what method is more reliable, it is firstly very important to decide a reference method when it is used to analyze the characteristics of elderly voices. To acoustically analyse elderly voices is of great interest. This method is attractive for its potential ability to detect illnesses early and for its applicability in telemedicine. Future investigations will develop analysis methods to improve performance and implement more reliable detectors.

ACKNOWLEDGMENT

This work was supported by the Jungwon University Research Grants.

REFERENCES

[1] J. Lee, KHIDI Brief, Korea Health Industry Development Institute, vol. 140, pp1-2, 2014.
[2] K. C. Chang, S. H. Lee, J. N. Choi, H. S. Choi, C. H. Nam, D. W. Kim, "Comparison of vowel pitch results among several commercial voice analysis programs", ICS'05, pp.54-55, 2005.
[3] J. Y. Lee and S. H. Choi, "Perturbation analysis using a moving window for disordered voices", International Journal of Engineering, Science, and Innovative Technology, vol. 3(1), pp. 1-10, 2012.
[4] J. Y. Lee, "Moving window approach with minimum perturbation for acoustic anaylsis of pathological voices", Jokull journal, http://www.jokulljournal.com/coredoux/index.php/jTracker/index/7kfNH, 2014.
[5] J. Y. Lee, S. H. Choi, J. J. Jiang, M. S. Hahn, and H. S, Choi, "Perturbation and Perceptual Analysis of Pathological Sustained Vowels According to Signal Typing", Malsori and Speech Science, vol. 2(2), pp. 109-115, 2010.
[6] J. Y. Lee, S. H. Choi, J. J. Jiang, M. S. Hahn, and H. S. Choi, "Perturbation and Nonlinear Dynamin Analysis of Sustained Vowels in Normal and Pathological voices with Signal Typing", Malsori and Speech Science, vol. 2(1), pp. 113-120, 2010.
[7] J. Y. Lee, "Sample selection approach using moving window for acoustic analysis of pathological sustained vowels according to signal typing", Journal of the Korean society of speech sciences, vol.3(3), 2011.
[8] J. Y Lim, "Current Researches on Vocal Fold Aging", 대한후두음성언어의학회지, vol. 25(1), pp.24-26, 2014.
[9] K. H. Pae, J. H. Wang, S. H. Choi, S. Y. Kim, and S. Y. Nam, "Glottic Characteristics and Voice Complaint in the Elderly", Journal of Special Education & Rehabilitation Science, 대한음성언어의학회지, vol.16(2), pp. 135-139, 2014.
[10] S. I. Kim and B. C. Kim, "Dialog System based on Speech Recognition for the Elderly with Dementia", 한국해양정보통신학회논문지, vol. 6(6), pp. 923-930, 2002.

[11] Y. Song, “Prevalence of Voice Disorders and Characteristics of Korean Voice Handicap Index in the Elderly”, 말소리와 음성과학, vol. 4(3), pp. 151-159, 2012.
[12] M. Petrović-Lazić, S. Babac, M. Vuković, R. Kosanović, and A. Ivanković, “Acoustic Voice Analysis of Patients with Vocal Fold Polyp,” J. Voice, vol. 26, pp. 92-97, 2010.
[13] Y. Zhang and J. J. Jiang, “Acoustic Analyses of Sustained and Running Voices From Patients With Laryngeal Pathologies”, J. Voice, vol. 22, pp.1-9, 2008.
[14] P. Milenkovic, “TF32 User’s Manual”. Madison, WI, 2001.
[15] D. K. Raymond and J. B. Martin, “Voice Quality Measurement”, Thomson Learning, pp. 134-159, 2007.

AUTHOR PROFILE

JiYeoun Lee obtained B.S. in division of Electronics and Computer Engineering from Hanyang University in the year of 2001, M.S. and Ph.D in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST) in the year of 2003 and 2008. Dr. JiYeoun Lee is now an assistant professor at the faculty of Biomedical Engineering, Jungwon University. Her research interest fields are Biomedical Signal Processing, Noise Reduction, Sound Localization, and Audio Signal Processing.