Article

Comparative Study between Healthy Young and Elderly Subjects: Higher-Order Statistical Parameters as Indices of Vocal Aging and Sex

Hee-Jin Choi ¹ and Ji-Yeoun Lee ^{2,*}
¹ Department of Electrical Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea; change@kaist.ac.kr

² Department of Biomedical Engineering, Jungwon University, 85 Munmu-ro, Goesan-eup, Goesan-gun 28024, Korea

* Correspondence: jylee@jwu.ac.kr; Tel.: +82-43-830-8607

Abstract: The objective of this study was to test higher-order statistical (HOS) parameters for the classification of young and elderly voice signals and identify gender- and age-related differences through HOS analysis. This study was based on data from 116 subjects (58 females and 58 males) extracted from the Saarbruecken voice database. In the gender analysis, the same number of voice samples were analyzed for each sex. Further, we conducted experiments on the voices of elderly people using gender analysis. Finally, we reviewed the standards and reference models to reduce sex and gender bias. The acoustic parameters were extracted from young and elderly voice signals using Praat and a time–frequency analysis program (TF32). Additionally, we investigated the gender- and age-related differences in HOS parameters. Young and elderly voice signals significantly differed in normalized skewness ($p = 0.005$) in women and normalized kurtosis ($p = 0.011$) in men. Therefore, normalized skewness is a useful parameter for distinguishing between young and elderly female voices, and normalized kurtosis is essential for distinguishing between young and elderly male voices. We will continue to investigate parameters that represent important information in elderly voice signals.

Keywords: vocal aging; Praat; time–frequency analysis program; higher-order statistics; elderly voice analysis

Citation: Choi, H.-J.; Lee, J.-Y. Comparative Study between Healthy Young and Elderly Subjects: Higher-Order Statistical Parameters as Indices of Vocal Aging and Sex. *Appl. Sci.* **2021**, *11*, 6966. <https://doi.org/10.3390/app11156966>

Academic Editor: Michael Döllinger

Received: 29 April 2021

Accepted: 22 July 2021

Published: 28 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to an analysis conducted by the National Statistics Office, 15.7% of Korea's population was aged 65 or older in 2020. This percentage is expected to continue to increase in the future, reaching 20.3% in the year 2025 [1]. At that time, Korea is expected to be a super-aged society. The aging of laryngeal tissue changes the movement of vocal cords, their vibration, and their opening and closing processes. Therefore, the recognition of elderly voices requires an understanding of the characteristics caused by changes in vocal cord tissue owing to anatomical or physiological aging. Voice characteristics are measured by the frequency of vocal cord oscillations per second, that is, the fundamental frequency (F0), jitter, shimmer, etc. [2,3]. As the voices of children, adolescents, seniors, and middle-aged people have distinct features, voice characteristics should be measured across different age groups.

Gender analysis aims to provide new knowledge in industries and markets by determining the effects of biological, social, and behavioral differences between men and women. In 2005, the European Medicines Agency announced that clinical trials should be performed and reported in consideration of gender. In 2011, the United States (U.S.) National Institutes of Health (NIH) published guidelines for the research and evaluation of gender differences in clinical trials of medical devices. Since 2013, *The Lancet* and the

Journal of the American Heart Association have been calling for the inclusion of gender-based differences in the publication of scientific articles. In Europe, Horizon 2020 presented a framework for gender equality in research and innovation. Additionally, the U.S. NIH has recommended that both sexes be included in its financially supported clinical trials when using animals, cells, and tissues and stated that the sex of animals and cell origins must be included as important variables in submitted research plans. From basic research to clinical research and applications, gender analysis should be actively utilized, and its effectiveness in research should be enhanced by verifying and correcting the analysis [4–7].

To respond to changes in a super-aged society, voice recognition developed based on gender analysis should be used in clinical practice to interpret voices in terms of public support services and to enable the more active use of various information technology (IT) devices by older people. Given that the interfaces of current voice recognition systems analyze voice patterns of all ages, the voice recognition performance tends to deteriorate if it deviates slightly from the average pattern [8,9]. Therefore, a voice recognition system should account for elderly voices [10]. Research on voice signal processing for the elderly is necessary, and the analysis of elderly voices can be regarded as an extension of gender analysis.

Previous studies [11,12] have found that the fundamental frequency (F0) can be affected by different factors, such as age, vocal fold length, and language or ethnological background. It was also observed that the most commonly used acoustic parameters depend on F0. Aging has been known to influence the classification performance results of acoustic parameters. Natalie et al. conducted a study to provide preliminary acoustic standards for the voices of elderly (60–80 years old) and young populations (20–30 years old). In older individuals, they found direct relationships between the tone of voice and the degradation of acoustic parameters, such as jitter and shimmer [13–25].

However, acoustic studies on new parameters for the classification of elderly voices are insufficient, as most have focused on parameters such as F0, mel-frequency cepstral coefficients (MFCCs), and linear prediction cepstrum coefficients (LPCCs). Therefore, it is necessary to establish new parameters for this population and to identify their effects in young and elderly voices for different genders [13–25]. Accordingly, the aim of the present study is to provide higher-order statistical (HOS) parameters for the classification of young and elderly voice signals to identify gender- and age-related differences in HOS parameters and to associate the HOS parameters with extensively used acoustic parameters extracted from Praat and a time–frequency analysis program (TF32). Therefore, the originality of this work is the proposal of a new parameter based on gender analysis to differentiate the voice signals of young and elderly people.

2. Materials and Methods

2.1. Database

This work used the Saarbruecken voice database (SVD) recorded by the Phonetics Research Institute at Saarland University, Germany [26]. We used the sustained vowel /a/ sound recorded from 116 normal speakers (58 females and 58 males) at neutral pitch, including 36 voice samples of “vox senilis”. Inclusion criteria were the absence of physiological and organic anomalies on the SVD datasheet. Exclusion criteria were neurological disorders affecting laryngeal functions, chronic degenerative diseases, vocal cord lesions, etc., that can cause voice disorders. We classified the voice signals into two different groups, namely, those from young and elderly subjects, based on a recent publication [13]. Although there may be a large gap in the voice signals of young subjects, they were divided into two groups in a large frame according to the purpose of our study. We also analyzed voice differences by comparing young and elderly subjects according to sex. Thus, we subdivided the voice signals into four subcategories. Group 1 consisted of 21 young men between the ages of 22 and 59 years (mean age = 39.12). Group 2 consisted of 21 elderly men between the ages of 60 and 89 years (mean age = 71.2). Group 3 included 37 young

women between the ages of 20 and 58 years (mean age = 39.17). Finally, Group 4 consisted of 37 elderly women between the ages of 60 and 87 years (mean age = 70.7).

2.2. Gender Analysis Methods

The Gender Innovation Website presents the 12 most common methods of gender analysis for utilization in scientific technology [4]. The first involves rethinking research priorities and outcomes. This entails reviewing how gender can affect the priorities of future studies. The second involves rethinking concepts and theories. This is a means of considering (a) how the concept and theory of a study can be formed from a gender perspective, (b) which gender hypotheses are implicit in these concepts and theories, and (c) how the concept and theory of gender affects the selection of research subjects and methods and the review and interpretation of the data. The third involves the formulation of research questions. Similar to other research and development processes, re-examining existing research practices with the importance of gender in mind can lead to creative and innovative development. The fourth method involves the analysis of sex. Sex, which distinguishes the biological differences between men and women, plays an important role in prioritizing research, establishing hypotheses, and designing experiments. The fifth method involves the analysis of gender. The ideas that researchers have about gender affect the prioritization of research, the development of research problems, and the selection of research methods. This can lead to stereotypes and prejudices in scientific and engineering research. The sixth method aims to analyze how sex and gender interact. In reality, sex and gender interact to form individual bodies, cognitive abilities, and disease patterns. In turn, the seventh method aims to analyze factors intersecting sex and gender. Factors or variables such as genetic characteristics, age, sex hormones, reproductive state, body composition, physical size, disability status, ethnicity, nationality, geographical location, socioeconomic status, educational background, religion, lifestyle, language, etc. reflect the biological, social, cultural, and psychological aspects of the user and customer. The eighth method relates to engineering innovation processes. By incorporating gender analysis into engineering innovation and technologies, we can develop new products, processes, infrastructure, and services that can promote gender equality and wellbeing and discover new markets and business opportunities. The ninth method relates to health and biomedical research studies. When conducting various types of studies, such as surveys, experimental studies, clinical trials, case studies, etc., sex and gender analysis should be incorporated into many stages of the research design process. The 10th method is for participatory research and design. This approach analyzes experiences specified by sex and gender. Additionally, the 11th method requires rethinking standards and reference models. Standards and reference models developed on the basis of research results for specific groups of men and women may lead to erroneous results in their future applications. Finally, the 12th method involves rethinking language and visual representations. Consideration should be given to whether unconscious assumptions about gender are implicit in metaphors and visual presentations of data, and inclusive language should be used.

In this study, the fourth method of sex analysis was applied using the same number of voice samples for each sex. The fifth gender analysis was conducted by only setting the voice of the elderly as a target without using the voices of all ages. The characteristics of gender-specific parameters extracted from the voices of elderly female and male groups were analyzed. Finally, to reduce sex and gender bias, we reviewed the standards and reference models by applying the 11th analysis. Therefore, it is not necessary to utilize all 12 gender analysis methods, but it is important to select and apply them according to the purpose of the research.

2.3. Praat and TF32 Software: Setting

Praat is a representative tool for speech evaluation. The advantage of Praat is that the scripts allow researchers to simultaneously process large amounts of data quickly. TF32 is

a 32-bit window-based time–frequency analysis program that can analyze speech sounds or audible frequency waveforms, and it has recently become increasingly used by voice scientists and voice clinicians. When the same voice data are analyzed by different analysis programs, there may be differences in data values depending on the analysis program. Analytical programs detect F0 using various methods, such as the autocorrelation function, zero-crossing rate, etc. Therefore, in this work, we selected the Praat and TF32 programs and analyzed the similarities or differences in the acoustic measures produced by these two programs, since TF32 and Praat both detect F0 using the cross-correlation method [27].

In the Praat software, it is crucial to appropriately set the parameters for computing the spectrogram, such as “view range”, “window length”, and “dynamic range”. For the analysis, we set two different pitch ranges: one specifically for female voices (100–500 Hz) and the other specifically for male voices (75–300 Hz). The pitch range was chosen to be the same as that used in the previous study [11] so that the results of the two studies could be compared. In TF32, the standard range and values recommended by Milenkovic were used [28].

2.4. HOS Analysis

HOS analysis in the time domain has shown massive potential as a classification index for pathological signals. The primary advantage is that it does not require periodic or quasiperiodic signals to enable reliable analysis [29–32]. Among the various HOSs, the 3rd- and 4th-order cumulants were used as characteristic parameters in this study. These parameters are called normalized skewness γ_3 and normalized kurtosis γ_4 , and they are defined as shown in (1).

$$\gamma_3 = \frac{\sum_{n=1}^N (x_n - \mu)^3}{(N-1)\sigma^3}, \quad \gamma_4 = \frac{\sum_{n=1}^N (x_n - \mu)^4}{(N-1)\sigma^4}. \quad (1)$$

where, x_n is the n th sample value, N is the number of samples, and μ and σ represent the mean and standard deviation, respectively.

Normalized skewness is a measure of the symmetry in a distribution. A normal distribution has a skewness of 0. If the skewness is between -0.5 and 0.5 , the data are fairly symmetrical. If the skewness is between -1 and -0.5 or between 0.5 and 1 , the data are moderately skewed. If the skewness is less than -1 or greater than 1 , the data are highly skewed. Normalized kurtosis is a measure of the combined sizes of the two tails. It measures the amount of probability in the tails. The value is often compared to the kurtosis of the normal distribution, which is equal to 3. If the kurtosis is greater than 3, then the dataset has heavier tails than a normal distribution. If the kurtosis is less than 3, then the dataset has lighter tails than a normal distribution.

2.5. Statistical Analysis

Statistical Package for the Social Sciences, version 24.0 (SPSS, Inc., Chicago, IL, USA) (IBM Corp., Armonk, NY, USA) was used for the statistical analysis. The normality of the distribution of the data was investigated by the Kolmogorov–Smirnov test. The two-sample t -test assumes normality and the Mann–Whitney U-test does not assume normality [33]. In this study, if the data satisfied normality, the distribution of the two groups were compared using a two-sample t -test using means and standard deviation. However, if normality was not satisfied, the Mann–Whitney U-test was used. The significance level was set a priori at $p < 0.05$.

3. Results

With the exception of the “Mean H/N and fundamental frequency (F0),” which follows a normal distribution in Praat, as shown in Table 1, all parameters were tested using the Mann–Whitney U-test. The analysis aimed to determine whether the mean ranks were significantly different between the two groups (young and elderly voices in females and males) to identify systematic differences in various parameters between the

two vocal classes. The medians of all groups were analyzed, except for two Gaussian-distributed parameters with arithmetic (mean and standard deviation) values and 95% confidence intervals (Table 1). Acoustic parameters were analyzed to compare the voices of younger and older people. The following parameters were statistically significant in Praat: fundamental frequency (F0, $p = 0.002$), jitter local ($p = 0.002$), jitter local abs ($p = 0.04$), jitter rap ($p = 0.01$), jitter ppq5 ($p = 0.008$), shimmer local ($p = 0.002$), shimmer local (dB) ($p = 0.001$), shimmer apq3 ($p = 0.016$), shimmer apq5 ($p = 0.001$), mean N/H ($p = 0.008$), and mean H/N ($p = 0.002$) in men; and fundamental frequency (F0, $p = 0.038$), jitter local ($p = 0.028$), and jitter local abs ($p = 0.017$) in women. The values of the fraction of locally unvoiced frames in all men and women were all zero. Therefore, the standard deviations were all zero, which means that all values were the same.

Table 1. Means, 95% confidence intervals (CIs) for the mean and median, and standard deviations (SDs) for all parameters analyzed with Praat.

		Men		Women	
		Young	Elderly	Young	Elderly
Fundamental frequency (Hz)	Arithmetic mean	121.64	142.02	216.93	201.06
	95% CI for the mean	112.91–130.09	133.73–150.31	207.19–226.02	191.88–210.40
	SD	4.55	4.28	4.85	4.76
	p value	0.002 *		0.038 *	
Jitter local (%)	Arithmetic median	0.36	0.51	0.32	0.44
	95% CI for the median	0.28–0.40	0.41–0.75	0.30–0.41	0.34–0.56
	SD	0.04	0.10	0.03	0.05
	p value	0.002 *		0.028 *	
Jitter local abs (us)	Arithmetic median	29.60	39.61	15.80	22.10
	95% CI for the median	24.61–35.72	28.77–56.78	14.06–20.16	16.26–28.58
	SD	2.95	8.20	1.36	3.17
	p value	0.04 *		0.017 *	
Jitter rap (%)	Arithmetic median	0.16	0.27	0.18	0.22
	95% CI for the median	0.12–0.23	0.19–0.31	0.17–0.23	0.17–0.28
	SD	0.03	0.04	0.02	0.02
	p value	0.01 *		0.279	
Jitter ppq5 (%)	Arithmetic median	0.20	0.30	0.19	0.24
	95% CI for the median	0.16–0.24	0.21–0.42	0.16–0.24	0.17–0.29
	SD	0.02	0.05	0.02	0.02
	p value	0.008 *		0.124	
Shimmer local (%)	Arithmetic median	2.69	4.91	2.69	2.65
	95% CI for the median	2.13–2.98	3.23–6.96	2.39–3.26	2.09–3.00
	SD	0.24	1.03	0.25	0.23
	p value	0.002 *		0.969	
Shimmer local (dB)	Arithmetic median	0.23	0.43	0.23	0.25
	95% CI for the median	0.18–0.27	0.28–0.63	0.20–0.28	0.18–0.28
	SD	0.02	0.09	0.02	0.02
	p value	0.001 *		0.617	
Shimmer apq3 (%)	Arithmetic median	1.29	2.64	1.42	1.32
	95% CI for the median	1.00–1.57	1.43–3.71	1.28–1.72	1.06–1.48
	SD	0.16	0.60	0.11	0.12
	p value	0.016 *		0.596	
Shimmer apq5 (%)	Arithmetic median	1.48	3.02	1.63	1.56
	95% CI for the median	1.29–1.77	2.07–4.03	1.50–2.05	1.34–1.74
	SD	0.13	0.58	0.15	0.09
	p value	0.001 *		0.961	
Mean N/H	Arithmetic median	0.008	0.018	0.005	0.01
	95% CI for the median	0.006–0.0111	0.008–0.041	0.004–0.007	0.006–0.011
	SD	0.001	0.009	0.001	0.001
	p value	0.008 *		0.106	
Mean H/N (dB)	Arithmetic mean	22.80	19.50	23.65	23.26
	95% CI for the mean	21.73–23.77	17.68–21.19	22.49–24.79	22.05–24.39
	SD	0.51	0.88	0.58	0.59
	p value	0.002 *		0.102	

Note: * $p < 0.05$.

Table 2 shows the results of the Mann–Whitney’s U-test, which was applied to all parameters except for “SNR and fundamental frequency (F0)”, which followed a normal distribution (analysis of variance) in TF32. Acoustic parameters were analyzed to compare the voices of male and female young people and male and female seniors. The following parameters were statistically significant: fundamental frequency (F0, $p = 0.004$), ppd ($p = 0.011$), jitter ($p = 0.001$), shimmer ($p = 0.001$), signal to noise ratio (SNR; $p = 0.001$), Trk ($p = 0.044$), and Err ($p = 0.000$) in men; and fundamental frequency (F0, $p = 0.008$), ppd ($p = 0.048$), jitter ($p = 0.018$), Trk ($p = 0.000$), and Err ($p = 0.004$) in women.

Table 2. Means, 95% CIs for the mean and median, and SDs for all parameters analyzed with TF32.

		Men		Women	
		Young	Elderly	Young	Elderly
Fundamental frequency (Hz)	Arithmetic mean	125.32	142.23	213.75	199.06
	95% CI for the mean	116.24–135.11	134.41–150.68	202.65–224.42	188.23–209.14
	SD	4.74	4.17	5.42	5.22
	p value	0.004 *		0.008 *	
Ppd (%)	Arithmetic median	8.06	6.87	4.63	4.88
	95% CI for the median	7.39–8.61	6.58–7.64	4.53–5.01	4.60–5.24
	SD	0.32	0.28	0.17	0.16
	p value	0.011 *		0.048 *	
Jitter (%)	Arithmetic median	0.34	0.53	0.31	0.42
	95% CI for the median	0.28–0.43	0.37–0.71	0.28–0.41	0.32–0.56
	SD	0.03	0.09	0.03	0.05
	p value	0.001 *		0.018 *	
Shimmer (%)	Arithmetic median	2.17	3.54	1.91	2.03
	95% CI for the median	1.96–2.46	2.54–5.07	1.61–2.31	1.72–2.50
	SD	0.14	0.79	0.15	0.20
	p value	0.001 *		0.765	
SNR (dB)	Arithmetic mean	23.89	20.05	24.13	23.42
	95% CI for the mean	22.31–25.33	18.31–21.81	22.76–25.40	22.29–24.71
	SD	0.76	0.91	0.65	0.62
	p value	0.001 *		0.233	
Trk	Arithmetic median	90.00	146.00	34.00	65.00
	95% CI for the median	60.00–135.00	121.00–200.00	21.00–43.00	35.00–144.00
	SD	23.62	17.20	7.01	28.22
	p value	0.044 *		0 *	
Err	Arithmetic median	0.00	5.00	0.00	1.00
	95% CI for the median	0.00	0.00–0.16	0.00	0.00–6.00
	SD	0.03	3.93	0.09	1.65
	p value	0 *		0.004 *	

Note: * $p < 0.05$.

The similarity in the acoustic parameters of the two programs was evident in parameters such as shimmer (%) and SNR (dB) in female voice samples, as these did not statistically differ. Although there were slight differences in the shimmer (%) values extracted from female and male voice signals, it can be concluded that the two programs extract almost the same values because they both extract acoustic parameters based on the cross-correlation method.

The statistical analysis between young and elderly voice signals in Table 3 was performed using Mann–Whitney’s U-tests for non-Gaussian distributions and two-sample t -tests for independent samples. The significance level was set to $p < 0.05$. Young and elderly voice signals significantly differed in normalized skewness in women ($p = 0.005$) and normalized kurtosis in men ($p = 0.011$). Thus, they can be viewed as age-related parameters that differ according to gender. Since the mean of normalized skewness estimated

in young female voices is close to zero, the data are normally distributed. However, the values extracted from elderly male and female voices range from -0.5 to 0.5 , so the data are fairly symmetrical. As the means of the normalized kurtosis are all less than 3, the dataset is considered to have a light-tailed distribution.

Table 3. Means, 95% CIs for the mean and median, and SDs for all parameters analyzed with higher-order statistics (HOS).

		Men		Women	
		Young	Elderly	Young	Elderly
The normalized skewness	Arithmetic mean	−0.133	−0.141	−0.080	0.182
	95% CI for the mean	−0.238–−0.026	−0.205–0.143	−0.340–0.182	−0.046–0.384
	SD	0.054	0.092	0.135	0.110
	<i>p</i> value	0.65		0.005 *	
The normalized kurtosis	Arithmetic median	2.976	2.483	2.722	2.539
	95% CI for the median	2.846–3.333	2.275–2.601	2.461–2.989	2.311–2.751
	SD	0.156	0.092	0.121	0.147
	<i>p</i> value	0.011 *		0.494	

Note: * $p < 0.05$.

Figure 1 shows the results obtained in Praat. The figure presents various parameters extracted from Praat in the form of box plots, which provide better visualization of young and elderly voice signals for men and women. As shown in Figure 1a, F0 tends to increase in older men, whereas it tends to significantly decrease in older women [13]. This may be attributed to a change in the posture of elderly people who bend forward, which lowers the vocal cords. Notably, the results are the same as those obtained by other authors [34,35]. As shown in Figure 1b, jitter was significantly higher in both men and women of all ages. The jitter represents the regularity of the oscillation cycle and the perturbation of the F0 mean, and it is related to the degree of roughness. Therefore, the results can justify the use of speech sounds as clinical cues in presbyphonia. We also obtained similarly meaningful results by analyzing other parameters, such as jitter local abs, rap, and ppq5, for the same properties (Figure 1c–e). For women, there were no significant differences in certain jitter parameters, such as jitter local rap and jitter ppq5. The schematic in Figure 1f shows the perturbation of the glottic vibration, indicating the amplitude of the sound wave. This is related to changes in the degree of voice breathiness and intensity variations. Similar results in Figure 1g–i are from the study of acoustic parameters, such as shimmer local (dB), apq3, and apq5, which describe similar speech characteristics. According to the results, the shimmer was significantly lower in young men than in elderly men. However, in females, no significant difference was observed in the shimmer, which tended to be constant in elderly female voices and mostly increased in young female voices. The shimmer did not change considerably with respect to age in women. These results agree with indications that age-related changes in the larynx are greater in men than in women and may begin earlier in males [36].

The mean N/H in Figure 1j shows the amount of noise associated with the harmonics of the waveform: the higher the value, the lower the overall sound quality level. The results show that the parameter values were much higher in older people of both sexes (women and men over the age of 60). Noise is caused by changes in frequency and amplitude (jitter and shimmer), subharmonic components, and instantaneous voice interruption; thus, the reliability of the above results is supported. The mean N/H in Figure 1j also confirms that voice aging leads to a general deterioration of speech quality that can be objectively measured. The mean H/N dB shown in Figure 1k defines the relationship between the intensity of harmonic and nonharmonic components in the overall spectrum of the measured voice signal. The greater the value, the clearer the speech quality. In this study, older male subjects scored much lower than younger males, indicating that the vocal performance was worse for men than women over the age of 60. The reduction in harmonic components can be explained by changes in the resonant structure of the vocal tract.

Figure 2 shows the results of the analysis with TF32. F0 (Figure 2a) tended to increase in elderly male voices and decreased in elderly female voices, yielding significant differences from values extracted from Praat. Ppd (Figure 2b) presents the pitch period in milliseconds. As the frequency and period are reciprocal, the graphs of F0 and pitch period have opposite trends. Therefore, F0 tended to decrease in elderly male voices and increase in elderly female voices, and the difference was significant. Jitter (Figure 2c) is the cycle-to-cycle variation in the pitch period during voicing. Shimmer (Figure 2d) is the variation in amplitude between cycles. The SNR (Figure 2e) compares the magnitude of a voice signal to the magnitude of the aperiodic component as defined by the TF32 manual. They tend to be similar to the values extracted through Praat. For men, all acoustic parameters extracted from Praat and TF32 yielded significant results ($p < 0.05$). For women, there was no significant difference in the shimmer and SNR based on the acoustical parameters extracted from Praat and TF32. The numbers trk and err are the reliability measures for the pitch tracker. A high trk count reflects large swings in F0, while a high err count indicates voice breaks that disrupt the pitch track. The trk (Figure 2f) and err (Figure 2g) tended to increase in elderly male and female voices. In elderly voices, there are many large swings in pitch and breaks that may exaggerate the jitter and shimmer values and diminish the SNR in a voice that is already unsteady.

Figure 3 shows the results of the HOS analysis. In Figure 3a, the mean of the normalized skewness estimated in female voices tends to be larger than zero. However, the value extracted from male voices tends to be less than zero. In particular, there is a difference between the distributions of the normalized skewness of young and elderly female signals ($p = 0.005$). In elderly female signals, the distribution of the normalized skewness tends toward higher values and is slightly wider than that of young female signals. In this case, the distribution refers to the range of the normalized skewness parameter. Overall, it is evident that the normalized skewness is sufficiently distinct and can be used to analyze young and elderly female voice signals in terms of statistics such as the p value and the mean of normalized skewness, although there is an overlap in the normalized skewness parameter between young and elderly female voices. In addition, it can be used as a basis for the automatic classification of young and elderly female voice signals. For normalized kurtosis in Figure 3b, the estimated values of the young male voice signals tend to be higher and wider than the estimated values of the elderly male voice signals. As the normalized kurtosis estimated for young male voices tends to be larger than or equal to three, the distribution has heavier tails and is called a leptokurtic distribution. However, because the normalized kurtosis for elderly female voices tends to be less than 3, the distribution has light tails and is called a platykurtic distribution. In particular, there is a clear difference between the distributions of the normalized kurtosis of young and elderly male signals. In female voice signals, the normalized kurtosis tends to be less than three, so the distribution has light tails and is called a platykurtic distribution. Therefore, the normalized kurtosis sufficiently differentiates the young and elderly male voice signals and can be used for automatic classification between young and elderly male voice signals.

Figure 1. Cont.

Figure 1. Distributions of acoustic parameters analyzed with Praat.

Figure 2. Distributions of acoustic parameters analyzed with TF32.

Figure 3. Distributions of higher-order statistical parameters.

4. Discussion

The proportion of the elderly in the total population will increase significantly over the next few decades. Anatomical and physiological changes in the larynx owing to aging may change the pitch of the voice. These changes can be distinguished from normal signals [5–7]. However, in most smart devices, elderly voice signals have been neglected because the interface does not consider age as a factor [7]. As speech interfaces currently use an optimized method based on the average speech pattern of people of all ages, the performance of voice analysis and recognition may degrade when an elderly voice is input into the voice recognition system [3,4,13,14,16,17].

To respond to changes in a super-aged society, voice recognition developed based on gender analysis should be used in clinical practice. Therefore, research on voice signal processing for the elderly is needed, and the analysis of elderly voices can constitute an extension of gender analysis. In this study, the fourth method of sex analysis was applied using the same number of voice samples for each sex. The fifth gender analysis method was used by setting the voice of the elderly as a target without using voices of all ages. Finally, to reduce sex and gender bias, we reviewed the standards and reference models by applying the 11th analysis method. Standard voices are often categorized as male. This study divided voice signals into four categories and used them as standard models. In the future, we will use a similar reference model to apply deep learning algorithms. Therefore, it is not necessary to utilize all 12 methods of gender analysis, but it is important to select and apply them according to the purpose of the research.

The purpose of this study was to classify the voices of the elderly and to quantify age-related changes in voices using gender analysis, provide HOS parameters for the classification of young and elderly voice signals, identify gender- and age-related differences regarding HOS analysis, and associate HOS parameters with extensively used acoustic parameters extracted from Pratt and TF32. Our analyses highlighted statistically significant differences that can be regarded as useful parameters for classification between young and elderly voice signals in terms of gender. The following parameters were statistically significant using Praat and TF32 in all studied men and women: fundamental frequency, jitter local, jitter local abs, ppd, jitter, Trk, and Err.

The most important discovery of this study is that normalized skewness is a useful parameter for distinguishing between young and elderly female voice signals, and normalized kurtosis can differentiate between young and elderly male voices. There is merit in combining acoustic measures with age- and gender-related differences because they

contain important information and can thus improve the characterization of voices. We will continue to study parameters that reflect important information about older voice signals to achieve high classification performance between younger and older voice signals in real-world environments. We will also strive to spread awareness of gender analysis in the field of elderly voice signal processing.

5. Conclusions

In summary, the results of this study imply that normalized skewness is a useful parameter for distinguishing between young and elderly female voices, and normalized kurtosis is essential for distinguishing between young and elderly male voices. In future work, parameters to reflect important information about elderly voice signals will be studied for high-classification performance and with the use of various deep learning methods. It will also be necessary to develop the combination of various parameters and deep learning methods to predict elderly voice signals more sensitively.

Author Contributions: Conceptualization, J.-Y.L.; methodology, J.-Y.L.; software, J.-Y.L. and H.-J.C.; validation, J.-Y.L. and H.-J.C.; writing—original draft preparation, J.-Y.L. and H.-J.C.; writing—review and editing, J.-Y.L.; visualization, J.-Y.L.; funding acquisition, J.-Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2017R1A2B4011373). The sponsor had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Statistics Korea. Available online: http://kostat.go.kr/portal/korea/kor_nw/1/1/index.board?bmode=read&aSeq=385322 (accessed on 6 June 2021).
2. Kahane, J.C. Anatomic and physiologic changes in the aging peripheral speech mechanism. In *Aging: Communications Processes and Disorders*; Beasley, D.S., Davis, A., Eds.; Grune and Stratton: New York, NY, USA, 1981; pp. 21–45.
3. Lee, S.Y. The Overall Speaking Rate and Articulation Rate of Normal Elderly People. Graduate Program in Speech and Language Pathology. Master's Thesis, Yonsei University, Seoul, Korea, 2011.
4. KOFWST. Gendered Innovations. Available online: <http://gister.re.kr/#!/main> (accessed on 2 March 2017).
5. Lee, J.Y. Gender analysis in elderly speech signal processing. *J. Digital Conver.* **2018**, *16*, 351–356. [[CrossRef](#)]
6. Lee, J.Y. Elderly speech signal processing: A systematic review for analysis of gender innovation. *J. Conver. Inf. Technol.* **2019**, *9*, 148–154. [[CrossRef](#)]
7. Lee, J.Y.; Lee, H.S. Gendered innovation for algorithm through case studies. *J. Digital Conver.* **2018**, *16*, 459–466. [[CrossRef](#)]
8. Song, Y.-K. Prevalence of Voice Disorders and Characteristics of Korean Voice Handicap Index in the Elderly. *Phon. Speech Sci.* **2012**, *4*, 151–159. [[CrossRef](#)]
9. Lee, S.J.; Kwon, S.I. Elderly speech analysis for improving elderly speech recognition. *J. Korean Inst. Inf. Sci. Eng.* **2014**, *32*, 16–20.
10. Jeong, J.H.; Jang, J.H.; Moon, M. Development of AI Speaker with Active Interaction Customized for the Elderly. *J. Korean Inst. Electron. Commun. Sci.* **2020**, *15*, 1223–1230.
11. Braun, A. Fundamental Frequency—How Speaker-specific It Is? In *Studies in Forensic Phonetics*; Braun, A., Koster, J.P., Eds.; BEIPHOL: Berlin, Germany, 1995; pp. 9–23.
12. Mennen, I.; Schaeffler, F.; Docherty, G. Cross-language differences in fundamental frequency range: A comparison of English and German. *J. Acoust. Soc. Am.* **2012**, *131*, 2249–2260. [[CrossRef](#)]
13. Mezzedimi, C.; Di Francesco, M.; Livi, W.; Spinosi, M.C.; De Felice, C. Objective Evaluation of Presbyphonia: Spectroacoustic Study on 142 Patients with Praat. *J. Voice* **2017**, *31*, 257.e25–257.e32. [[CrossRef](#)] [[PubMed](#)]
14. Yamauchi, A.; Yokonishi, H.; Imagawa, H.; Sakakibara, K.-I.; Nito, T.; Tayama, N.; Yamasoba, T. Age- and Gender-Related Difference of Vocal Fold Vibration and Glottal Configuration in Normal Speakers: Analysis With Glottal Area Waveform. *J. Voice* **2014**, *28*, 525–531. [[CrossRef](#)]

15. Yamauchi, A.; Imagawa, H.; Yokonishi, H.; Nito, T.; Yamasoba, T.; Goto, T.; Takano, S.; Sakakibara, K.-I.; Tayama, N. Evaluation of Vocal Fold Vibration With an Assessment Form for High-Speed Digital Imaging: Comparative Study between Healthy Young and Elderly Subjects. *J. Voice* **2012**, *26*, 742–750. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Silva, M.; Vellasco, M.M.; Cataldo, E. Evolving Spiking Neural Networks for Recognition of Aged Voices. *J. Voice* **2017**, *31*, 24–33. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Ferrand, C.T. Harmonics-to-Noise Ratio. *J. Voice* **2002**, *16*, 480–487. [\[CrossRef\]](#)
18. Ambreen, S.; Bashir, N.; Tarar, S.A.; Kausar, R. Acoustic Analysis of Normal Voice Patterns in Pakistani Adults. *J. Voice* **2019**, *33*, 124.e49–124.e58. [\[CrossRef\]](#)
19. Ahmadi, A.; Hosseini, S.; Faham, M.; Shahramnia, M.M.; Ebadi, A.; Etter, N.M.; Shiani, A.; Dehghan, M. Translation, Validity, and Reliability of the Persian Version of the Aging Voice Index. *J. Voice* **2021**, *35*, 327.e13–327.e21. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Da Silva, P.T.; Master, S.; Andreoni, S.; Pontes, P.A.D.L.; Ramos, L.R. Acoustic and Long-Term Average Spectrum Measures to Detect Vocal Aging in Women. *J. Voice* **2011**, *25*, 411–419. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Maslan, J.; Leng, X.; Rees, C.; Blalock, D.; Butler, S.G. Maximum Phonation Time in Healthy Older Adults. *J. Voice* **2011**, *25*, 709–713. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Schaeffer, N.; Knudsen, M.; Small, A. Multidimensional Voice Data on Participants With Perceptually Normal Voices From Ages 60 to 80: A Preliminary Acoustic Reference for the Elderly Population. *J. Voice* **2015**, *29*, 631–637. [\[CrossRef\]](#)
23. Linville, S.E. Source Characteristics of Aged Voice Assessed from Long-Term Average Spectra. *J. Voice* **2002**, *16*, 472–479. [\[CrossRef\]](#)
24. De Machado, F.C.M.; Lessa, M.M.; Cielo, C.A.; Barbosa, L.H.F. Spectrographic Acoustic Vocal Characteristics of Elderly Women Engaged in Aerobics. *J. Voice* **2016**, *30*, 579–586. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Linville, S.E.; Rens, J. Vocal Tract Resonance Analysis of Aging Voice Using Long-Term Average Spectra. *J. Voice* **2001**, *15*, 323–330. [\[CrossRef\]](#)
26. William, J.B.; Manfred, P. *Saarbrücken Voice Database*; Institute of Phonetics, University of Saarland: Saarbrücken, Germany, 2007; Available online: <http://www.stimmdatenbank.coli.uni-saarland.de/> (accessed on 13 May 2018).
27. Ko, H.-J.; Woo, M.-R.; Choi, Y. Comparisons of voice quality parameter values measured with MDVP, Praat, and TF32. *Phon. Speech Sci.* **2020**, *12*, 73–83. [\[CrossRef\]](#)
28. Milenkovic, P.H. TF32 and Cspeech Home Page. Available online: <http://userpages.chorus.net/cspeech/> (accessed on 2 March 2001).
29. Nemer, E.; Goubran, R.; Mahmoud, S. Robust voice activity detection using higher-order statistics in the LPC residual domain. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 217–231. [\[CrossRef\]](#)
30. Lee, J.-Y.; Jeong, S.; Choi, H.-S.; Hahn, M. Objective Pathological Voice Quality Assessment Based on HOS Features. *IEICE Trans. Inf. Syst.* **2008**, *E91.D*, 2888–2891. [\[CrossRef\]](#)
31. Lee, J.-Y.; Jeong, S.; Hahn, M. Pathological Voice Detection Using Efficient Combination of Heterogeneous Features. *IEICE Trans. Inf. Syst.* **2008**, *E91.D*, 367–370. [\[CrossRef\]](#)
32. Lee, J.Y.; Hahn, M. Automatic Assessment of Pathological Voice Quality Using Higher-Order Statistics in the LPC Residual Domain. *EURASIP J. Adv. Signal Process.* **2010**, *2009*, 748207. [\[CrossRef\]](#)
33. Kwak, S.G.; Park, S.-H. Normality Test in Clinical Research. *J. Rheum. Dis.* **2019**, *26*, 5–11. [\[CrossRef\]](#)
34. Harnsberger, J.D.; Shrivastav, R.; Brown, W.; Rothman, H.; Hollien, H. Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age. *J. Voice* **2008**, *22*, 58–69. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Hollien, H.; Shipp, T. Speaking Fundamental Frequency and Chronologic Age in Males. *J. Speech Hear. Res.* **1972**, *15*, 155–159. [\[CrossRef\]](#)
36. Goy, H.; Fernandes, D.N.; Pichora-Fuller, M.K.; van Lieshout, P. Normative Voice Data for Younger and Older Adults. *J. Voice* **2013**, *27*, 545–555. [\[CrossRef\]](#)