Open Access

Spectral and Cepstral Based Acoustic Features of Voices with Muscle Tension Dysphonia

Hee-Jeong Shim[1], Hun Jung[2], Rajinder Koul[3], Do-Heung Ko[4]

[1]GU Developmental Center, Gumi University, Korea; [2]Department of Speech and Language Rehabilitation, Gumi University, Korea; [3]School of Health Professions, Texas Tech University Health Sciences Center, USA; [4]Division of Speech Pathology and Audiology, Hallym University, Korea

Purpose: This study aimed to examine the cepstral and spectral acoustic features of patients with muscle tension dysphonia (MTD).

Methods: A total of 30 patients with MTD and 30 healthy women (control group) were enrolled. All participants were asked to vocalize a sustained vowel /a/ for more than 3 seconds, which was recorded and analyzed using the Analysis of Dysphonia in Speech and Voice program.

Results: Compared with the control group, patients with MTD had significantly lower cepstral peak prominence (CPP) and CPP F0, and significantly higher cepstral and spectral index of dysphonia (CSID) and a low- to high-frequency spectral energy ratio. Additionally, in patients with MTD, there was a high correlation between CPP and perceptual parameters such as grade, roughness, breathiness and strain. Receiver operating characteristic analysis found that a threshold of 11.815 for CSID achieved a good classification for MTD, with 73.3% sensitivity and specificity.

Conclusions: By applying cepstral and spectral analysis and identifying the acoustic features of patients with MTD, this study demonstrated the feasibility and usefulness of cepstral parameters in clinical practice.

Keywords: Voice, Muscle tension dysphonia, Acoustic analysis, Cepstral analysis, CSID

CrossMark
click for updates

CACD

INTRODUCTION

Muscle tension dysphonia (MTD) is a voice disorder that occurs as a result of misuse or abuse of the voice [1]. While there is no neurological damage to the larynx nor any structural abnormality present, patients with MTD have limitations in vocal vibrations due to laryngeal muscle tension. Consequently, such patients demonstrate pathological voice problems, such as hoarseness, strained and strangled voice, and voice tremors.

Speech and language pathology has attempted to describe the acoustic features of patients with MTD or other voice disorders based on perceptual assessment, which evaluates dysphonia severity or information regarding voice quality, and instrumental assessment, which evaluates acoustic analysis results. The goal of objective assessment, in particular, is to take a variety of acoustic approaches to distinguish the voice of dysphonic patients from that of normal speakers.

Recent findings on the clinical usefulness of cepstral and spectral parameters have

been reported [2-14,16]. Cepstral and spectral parameters not only analyze the voice of patients with severe voice disorders, such as MTD and neurogenic speech and language disorders, but they also predict the effect of voice treatment and provide the results of voice quality analysis in connected speech. Because they show high correlations with perceptual assessment results and high sensitivity for diagnosis of voice disorders, the clinical value of cepstral and spectral parameters has been highlighted [2-7].

Cepstral and spectral parameters, which could be useful in analyzing the voice quality of patients with voice disorders, could be examined in more detail. For example, cepstral peak prominence (CPP) is a parameter that expresses the degree of harmonics as a peak. The CPP is high in normal speakers who show a well-defined harmonic structure, but is low in dysphonic patients who have severe voice problems due to harmonic formation being restricted by irregular adduction of the vocal folds [8]. More importantly, the CPP is reported to be highly correlated with breathiness, roughness, and harshness. In addition, if such symptoms appear more pronounced, the CPP tends to decrease because of a decline in abnormal cepstral peak [7-9]. Mean CPP F0 is a parameter that expresses the average frequency of the CPP in the range of 60 to 300 Hz when a voiced sound is produced. In patients with a strained and strangled voice, CPP F0 is higher because there is more tension in the vocal folds [10]. The low- to high-frequency spectral energy ratio (L/H ratio) is a parameter that expresses the energy ratio of low to high frequency based on the frequency range with 4 kHz as the cut-off point. In this regard, pathologic voices with high frequency energy, such as breathiness and tension, demonstrate a lower L/H ratio [2,9]. The cepstral and spectral index of dysphonia (CSID) quantifies the level of voice abnormality, and is derived from the CPP, L/H ratio, and sex information through weighted multiple regression. The CSID tends to increase as the severity of voice disorders increases. For this reason, the CSID has very high correlations with the level of perceptual voice abnormality, and has recently gained attention as an influential parameter in determining voice disorders [11,12].

Previous cepstral and spectral analysis studies of dysphonic patients have mainly evaluated those suffering from vocal fold paralysis, in whom breathiness is noticeable, or those with vocal cord polyps and nodules or presbyphonia. The analysis methods mentioned above show high correlations with perceptual assessment results, such as roughness and hoarseness, therefore they are regarded as useful in being able to ob-

jectively explain the severity and acoustic features of patients with MTD, whose vocal cord vibrations are restricted because of excessive laryngeal muscle tension. This study attempted to examine the acoustic features of patients with MTD by applying cepstral and spectral analyses, and to determine the correlations between acoustic analysis results and perceptual assessment results. In addition, this study also attempted to determine the diagnostic predictability of cepstral and spectral parameters when the voice of patients with MTD was identified and diagnosed.

METHODS

Study participants

This study enrolled a total of 60 participants. Thirty of the participants were patients diagnosed with MTD by otolaryngologists and speech therapists through voice assessment after visiting the otorhinolaryngology clinic located in Seoul, Korea from January 2013 to October 2015. The remaining participants were 30 normal age-matched women (control group).

All of the study subjects had no breathing, visual, auditory, or language problems, and were speakers with no neurologic or structural problems in the larynx. Professional voice users were excluded. This study was approved by the Bioethics Committee of Hallym University (No. HIRB-2016-011), and all subjects provided consent prior to participation. Demographic data of the subjects are presented in Table 1.

Test tools

In order to examine the acoustic features of the study subjects, a directional microphone (SM 48, SHURE) was connected to a desktop equipped with the Analysis of Dysphonia in Speech and Voice (ADSV™) program (model 5109, KayPENTAX), which recorded and digitized the voice of participants at a sampling rate of 25,000 Hz. In addition, in order to examine the perceptual features of patients with MTD, the Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS) scale (Hirano, 1981), which utilizes a 4-point scale from 0 (normal) to 3 (severe) was used.

Table 1. Demographic information of the participants (60 women)

	MTD (n = 30)	Control (n = 30)	t	degree of freedom
Age ± SD (range)	35.85 ± 5.45 (21-48)	33.40 ± 2.28 (20-47)	-685	58
Gender	Female	Female		

Test procedure

The participants were recorded in a voice testing room inside the hospital where the noise level was lower than 40 dB. During voice recording, a microphone was placed 7 cm from the mouth of the speaker and fixed at 90 degrees. All participants were asked to vocalize a sustained vowel /a/ for more than 3 seconds with a voice reflecting their usual pitch, strength, and quality in as comfortable a state as possible.

Data analysis

The recorded voice data were edited and analyzed by using more than 2 seconds of the stabilized parts of the sustained vowel vocalization at approximately one-third from the beginning of the vowel. At this point, pitch, strength, and quality remained consistent and flat. These analyzed parts were automatically captured using the Apply Automatic Data selection tool of the ADSV program. The analyzed cepstral and spectral parameters are CPP, CPP F0, L/H ratio and CSID. For perceptual assessment of patients with MTD, 3 speech therapists (30 MTD sustained vowel tasks), who were blinded to the patients' information, assessed the 30 voice samples based on the GRBAS scale. The assessment results were averaged and used in the statistical analysis.

Statistical analysis

The differences in acoustic analysis parameters between the MTD and control groups were analyzed using an independent t test with SPSS version 19.0 (IBM Corporation). Parameters were expressed as the mean ± SD, and a p value less than 0.05 was regarded as statistically significant. The GRBAS results of patients with MTD were examined using descriptive statistics to determine whether the cepstral and spectral parameters were correlated with the perceptual assessment re-

sults. Lastly, the diagnostic predictability of cepstral and spectral parameters were evaluated when the voice of patients with MTD was identified and diagnosed, and a receiver operating characteristic (ROC) analysis was conducted to determine the cut-off point that could distinguish between the 2 groups.

Reliability

To verify reliability, the Pearson product-moment correlation coefficient (PPMCC) was used. After acoustic analysis, 12 samples (6 MTD samples and 6 normal samples, 20% of the total samples) were randomly selected and reanalyzed. Intrarater reliability of cepstral and spectral acoustic analysis was high at 1.000 ($p < 0.001$). To verify the reliability of perceptual assessment, 6 samples (20% of the total MTD samples) were randomly selected and reanalyzed. Inter-rater reliability was high and ranged from 0.853 and 0.964 ($p < 0.001$).

RESULTS

Cepstral and spectral acoustic analysis results

There were significant differences in the CPP, L/H ratio, CPP F0, and CSID between the two groups (Table 2). In particular, all values, except for CPP and CPP F0, tended to be higher in the MTD group compared to the control group.

Perceptual assessment results

The perceptual assessment results of patients with MTD according to the GRBAS scale are shown in Table 3. The most severe parameter of MTD was hoarseness, followed by strain, breathiness, and strangled acoustic features. The overall severity in patients was mild to moderate.

Correlations between acoustic analysis results and perceptual assessment results

The PPMCC analysis revealed both positive and negative correlations between acoustic analysis results and perceptual assessment results. In particular, the CPP decreased as overall grade, roughness, breathiness, and strain increased. Additionally, there were positive correlations in which CSID increased

Table 2. Spectral & cepstral measures between MTD and Control

Measures	Group	N	M ± SD	t	p
CPP (dB)	MTD	30	8.88 ± 1.78	-6.302	0.000***
	Contol	30	11.38 ± 1.25		
L/H ratio (dB)	MTD	30	30.93 ± 4.31	2.894	0.005**
	Contol	30	27.96 ± 3.60		
CPP Fo (Hz)	MTD	30	207.65 ± 20.23	-2.658	0.010*
	Contol	30	219.50 ± 13.69		
CSID	MTD	30	18.51 ± 10.21	4.242	0.000***
	Contol	30	8.96 ± 6.93		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 3. GRBAS score of MTD

Group		G	R	B	A	S
MTD	Mean	1.87	1.60	1.47	0.13	0.80
	SD	0.571	0.621	0.629	0.346	0.847

as grade, roughness, breathiness, and strain increased (Table 4). More specifically, the correlation coefficients between CPP and grade, roughness, breathiness, and strain were $r = -0.657$ ($p < 0.001$), $r = -0.555$ ($p < 0.01$), $r = -0.492$ ($p < 0.01$), and $r = -0.428$ ($p < 0.05$) respectively; demonstrating a high correlation with the grade scale. Meanwhile, the correlation coefficients between CSID and grade, roughness, and breathiness were $r = 0.762$ ($p < 0.001$), $r = 0.635$ ($p < 0.001$), and $r = 0.464$ ($p < 0.05$) respectively; revealing high correlations with the grade and roughness scales.

Diagnostic predictability of cepstral and spectral parameters
A ROC analysis was conducted in order to examine the diagnostic predictability and diagnostic threshold of cepstral and spectral parameters when the voice of patients with MTD was identified and diagnosed. The areas under the ROC curve for the CPP, L/H ratio, CPP F0, and CSID were 0.131, 0.707, 0.304, and 0.774, respectively; indicating that CSID had the highest diagnostic predictability. The optimal threshold for CSID was 11.815, which showed a sensitivity and specificity of 73.3% for distinguishing the voice of MTD patients from that of normal speakers (Table 5, Figure 1).

Discussion & Conclusions
This study evaluated 60 subjects in total, 30 patients with MTD and 30 normal speakers, and, by applying cepstral and spectral acoustic analysis, attempted to determine the most

useful acoustic parameter to predictively diagnose MTD and its threshold value.

Among cepstral and spectral values acquired from sustained vowel /a/ tasks, CPP tended to be higher in normal speakers than in patients with MTD. This finding could indicate that the voice of normal speakers easily formed a harmonic structure, whereas that of dysphonic patients formed limited periodicity because of irregular adduction of the vocal cords, resulting in lower CP*P* values. This finding is consistent
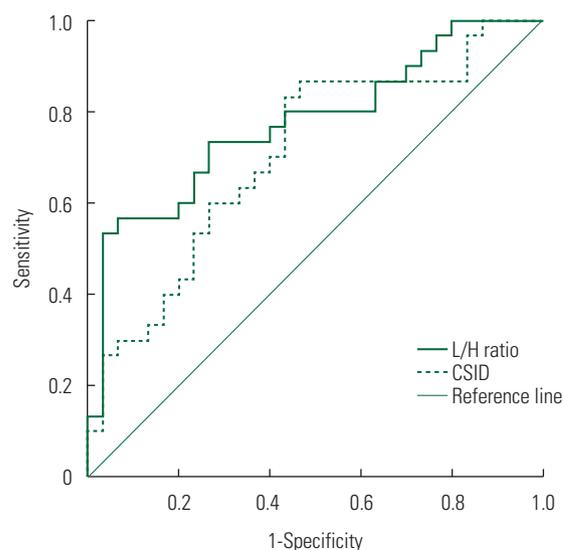

Figure 1. ROC Curve for the discrimination of the muscle tension dysphonia (AUC > 0.5).

Table 4. Pearson correlation between perceptual and acoustic measures

Task	Acoustic measures	G	R	B	A	S
Sustained vowel /a/	CPP	-0.657***	-0.555**	-0.492**	-0.272	-0.428*
	L/H ratio	-0.304	-0.191	-0.021	0.258	-0.031
	CPP F0	0.150	0.203	0.129	0.173	0.162
	CSID	0.762***	0.635***	0.464*	0.010	-0.666***

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 5. Sensitivity, specificity for CPP, L/H ratio, CPP F0, CSID as computed from the receiver operating characteristic (ROC)

Measures	Cut off value	Sensitivity	Specificity	AUC	SEM	Asymptoticp	Asymptotic 95% CI	
							Lower Bound	Upper Bound
CPP	10.350	0.267	0.233	0.131	0.044	0.000	0.044	0.217
L/H ratio	28.960	0.667	0.633	0.707	0.067	0.006	0.574	0.839
CPP F0	216.85	0.333	0.467	0.304	0.069	0.009	0.169	0.440
CSID	11.815	0.733	0.733	0.774	0.061	0.000	0.655	0.894

Cutoff values were determined via Youden's Index. AUC, area under curve; SEM, standard error of measurement.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

with most of the existing studies [4-10].

The L/H ratio was used to diagnose the breathiness of patients with larynx-related dysphonia. The L/H ratio had a strong correlation with the breathiness scale, and is a useful acoustic variable for identifying and diagnosing the breathiness of patients with voice disorders, such as unilateral vocal cord paralysis, vocal cord nodules, and presbyphonia [5,14]. In this study, the L/H ratio of patients with MTD was slightly higher than that of normal speakers, which could reflect the excessive tension in the larynx of patients with MTD. In addition, because hoarseness showed noise energy clearly at 100 to 2600 Hz, the L/H ratio was slightly higher in patients with MTD when compared to that of normal speakers [15].

The CSID is an acoustic index for abnormality that puts weight on cepstral and spectral analysis results to objectively describe the severity of voice disorders, which was previously represented subjectively. This study found that the CSID for patients with MTD was 18.51, whereas that for normal speakers was 8.96. This finding is consistent with previous studies [11-13,16].

With regard to the acoustic features of patients with MTD, grade was the most distinct, followed by roughness, breathiness, and strain. These acoustic features are reported to have high correlations with cepstral and spectral parameters [3,5,8, 10]. In this study, grade, roughness, breathiness, and strain increased, and the CPP decreased, indicating a negative correlation; whereas CSID increased, demonstrating a positive correlation.

In clinical studies, correlations between cepstral and spectral parameters and perceptual assessment results could be considered highly related to the overall severity of voice disorders. According to a longitudinal study that investigated dysphonic patients before and after treatment [17], not only did their CPP increase and their CPP SD decrease after voice therapy, but their problems of voice quality based on a visual analog scale tended to decline, demonstrating the close relationship between cepstral and spectral parameters and the severity of voice disorders [17,18]. This tendency can also be observed in patients with neurologic disease. In a previous study, there were significant correlations between the perceptual parameters of grade, breathiness, and strain in the connected speech of dysarthric patients and the cepstral parameters of smoothed CPP and CPP F0; demonstrating correlations between perceptual parameters and cepstral analysis [2].

Hoarseness and roughness were the most pronounced voice features in patients with MTD, which may demonstrate compensatory vocalization, such as insufficient vocal cord adduction or vocal efforts due to excessive tension in the vocal cord muscles [19]. These irregular vocal cord movements may restrict harmonic formation; thus, as the severity of the voice disorder increases, CPP decreases, while CSID increases.

The CSID objectively evaluates voice disorder severity by placing weight on CPP, L/H ratio, and information about sex by means of multiple regression and is reported to be highly correlated with perceptual assessment results. A previous study that examined the correlation between CSID and GRBAS [16] reported that CSID had a moderate correlation with perceptual assessment results of dysphonic patients. Similarly, another study [12] showed that CSID had a high correlation with visual analog scale results, which perceptually compared patients before and after the intervention. Therefore, these findings show the possibility of objective assessment of voice disorder severity and the clinical usefulness of CSID.

While the clinical researchers in the above studies reported findings that they determined subjectively, CSID is closely related to the level of pitch that patients hear themselves. In a VHI survey of 332 dysphonic patients, a significant positive correlation was found between CSID and the level of discomfort they felt as a result of their voice problems [13]. This presented a new direction with regard to correlations with subjective assessment.

In summary, patients with MTD showed the acoustic features of hoarseness, roughness, and strain. They had a low CPP and CPP F0, and a high L/H ratio and CSID. The CPP and CSID, in particular, showed high correlations with perceptual assessment results. In addition, the CSID was selected as the most important parameter, with more than a 73% sensitivity and specificity to distinguish the voice of patients with MTD from that of normal speakers. As mentioned earlier, CSID is an objective parameter that reflects voice disorder severity, and is calculated based on the CPP, L/H ratio, and sex information. Because there were significant differences between the voice of patients with MTD and that of normal speakers, CSID reflected these features as well.

For detecting whisper, which is one of the vocal disorder symptoms that could appear in dysphonic patients, the CSID showed an area under the curve of 1.0, and was considered a perfect predictive parameter [20]. Another study examined 88 female patients with MTD, before and after intervention, through CSID and perceptual assessment [17] reported that CSID reflected changes arising from MTD interventions fairly

well; and stressed that CSID is a useful variable not only for predicting voice disorders, but also for observing intervention effects. In addition, based on the ROC analysis results in this study, CPP was not able to independently predict MTD acoustic features, and the L/H ratio had a lower accuracy than CSID. These findings confirm the usefulness of CSID.

By applying cepstral and spectral analysis and identifying the acoustic features of patients with MTD, this study has demonstrated the feasibility and usefulness of cepstral parameters in clinical practice. This provides an alternative method to objectively evaluate the effectiveness of treatment for voice disorders. Considering the limitations of sex and voice tasks in this study, further studies are necessary in male patients and they should consider the acoustic features of sentence tasks.

CONFLICT OF INTEREST

The author has no conflict of interests.

REFERENCES

1. Rubin JS, Sataloff RT, Korovin GS. Diagnosis and treatment of voice disorders. Plural Publishing; 2014.
2. Seo IH, Seong CJ. Voice quality of dysarthric speakers in connected speech. Phon Speech Scien. 2013;5(4):33-41.
3. Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: A four factor acoustic model. Clin Linguist Phon. 2006;20(1):35-49.
4. Heman-Ackah Y, Heuer R, Michael D, Ostrowski R, Horman M, Baroody M, et al. Cepstral peak prominence: a more reliable measure of dysphonia. Ann Otol Rhinol Laryngol. 2003;112(4):324-333.
5. Hillenbrand J, Cleveland R, Erickson R. Acoustic Correlates of Breathy Vocal Quality. J Speech Hear Res. 1994;37:769-778.
6. Shim HJ, Jang HR, Shin HB, Ko DH. Cepstral, Spectral and Time-Based Analysis of Voices of Esophageal Speakers. Folia Phoniatr Logop. 2015;67(2):90-96.
7. Kumar B, Bhat J, Prasad N. Cepstral analysis of voice in persons with vocal nudules. J Voice. 2010;24(6):651-653.
8. Heman-Ackah T, Michael D, Goding G. The relationship between cepstral peak prominence and selected parameters of dysphonia, J Voice. 2002;16(1):20-27.
9. Watts C, Awan S. Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. J Speech Hear Res. 2011;54:1525-1537.
10. Lowell SY, Kelley RT, Awan SN, Colton RH, Chan NH. Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. Ann Otol Rhinol Laryngol. 2012;121(8):539-548.
11. Awan SN, Roy N, Zhang D, Cohen SM. Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. J Voice. 2016;30(2): 130-144.
12. Peterson EA, Roy N, Awan SN, Merrill RM, Banks R, Tanner K. Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. J Voice. 2013;27(4): 401-410.
13. Awan SN, Roy N, Cohen SM. Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). J Voice. 2014;28(4):430-439.
14. Hillenbrand J, Houde R. Acoustic correlates of breathy vocal quality dysphonic voices and continuous speech. J Speech Hear Res. 1996;39:311-321.
15. Ferrand C. Speech science: an integrated approach to theory and clinical practice. 2th ed. Pearson Education; 2007.
16. Jalalinajafabadi F, Gadepalli C, Ascott F, Homer J, Luján M, Cheetham B. Perceptual evaluation of voice quality and its correlation with acoustic measurement. EMS. 2014;283-286.
17. Awan S, Roy N. Outcomes measurement in voice disorders: application of an acoustic index of dysphonia severity. J Speech Hear Res. 2009;52(2):482-499.
18. Awan S, Roy N, Jette M, Meltzner G, Hillman R. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V. Clin Linguist Phon. 2010;24(9):742-758.
19. Altman KW, Atkinson C, Lazarus C. Current and emerging concepts in muscle tension dysphonia: a 30-month review. J Voice. 2005;19(2):261-267.
20. Seo IH, Lee OB. Cepstral and spectral analysis of whispery voice by healthy adults: Preliminary study. J Speech Hear Disord. 2015; 24(4):259-266.