ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Acoustic and Aerodynamic Clusters Within Primary Muscle Tension Dysphonia

Sarah Rose Bellavance[a] **and Aaron M. Johnson**[b]

[a] Department of Communicative Sciences and Disorders, New York University, NY [b] Department of Otolaryngology—Head and Neck Surgery, New York University Grossman School of Medicine, NY

ARTICLE INFO

Article History:
Received April 11, 2025
Revision received June 26, 2025
Accepted July 9, 2025

Editor-in-Chief: Jessica E. Huber
Editor: Amanda I. Gillespie

https://doi.org/10.1044/2025_JSLHR-25-00270

ABSTRACT

Purpose: Primary muscle tension dysphonia (pMTD) is a form of vocal hyperfunction with no preexisting tissue trauma to the vocal folds. There are no known structural or neurological causes of pMTD, and there is rarely obvious, confirmatory evidence to reliably diagnose individuals accurately. Furthermore, acoustic and aerodynamic measurements taken during voice assessments vary widely within this population. The purpose of this study was to find subgroups within a sample of pMTD patients based on acoustic and aerodynamic measurements. We use a computational approach to elucidate what has largely been observational in the past.

Method: A retrospective chart review was conducted to collect variables of interest for a sample of 72 pMTD patients seen at the NYU Langone Voice and Swallowing Center from January 1, 2021, to October 1, 2023. An exploratory factor analysis was conducted to find simpler structures in the data. Using factor scores from each patient, a *k*-means clustering analysis was conducted.

Results: The exploratory factor analysis grouped together variables across patients, which resulted in three principal axes. These three principal axes separately consisted of aperiodicity, fundamental frequency, and aerodynamic measurements. These principal axes explained 44.7% of the total variance. Four clusters of patients were identified across the three principal axes. These were characterized by (a) a high amount of aperiodicity in the voice, (b) lower fundamental frequency values, (c) higher fundamental frequency values, and (d) high aerodynamic values.

Conclusions: The clusters identified in the current study are reliable and moderately separated. Furthermore, these clusters align with previously identified subgroups in related work. The analysis presented here lays the groundwork for additional clustering analyses with new pMTD samples, as well as future work establishing subtype classifications of pMTD.

Primary muscle tension dysphonia (pMTD) is a voice disorder characterized by excessive tension in the laryngeal and supraglottic areas without the presence of any known structural or neurological causes (Morrison et al., 1983; Verdolini et al., 2006). This contrasts from secondary muscle tension dysphonia, in which some comorbidity is present (Belafsky et al., 2002; Verdolini et al., 2006). pMTD is a form of nonphonotraumatic vocal hyperfunction (Hillman et al., 2020; previously referred to as nonadducted vocal hyperfunction in Hillman et al., 1989), which is a form of laryngeal hyperfunction without tissue trauma. Its etiology is thought to be multifactorial (Desjardins et al., 2022; Van Houtte et al., 2011), with possibilities ranging from acute incidents (e.g., respiratory tract infection; Koufman & Blalock, 1988) to personality factors (e.g., anxiety; Roy & Bless, 2000).

Difficulty in reliably diagnosing pMTD is largely due to its heterogenous presentation (Roy, 2008), and there does not exist an accurate, standardized assessment for pMTD (Khoddami et al., 2015; Kunduk et al., 2016).

Correspondence to Sarah Rose Bellavance: srb664@nyu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Differential diagnosis from similar voice disorders involving excessive laryngeal tension can be difficult to achieve based on auditory-perceptual assessment alone (Roy, 2010), which is the most accessible form of assessment made in-clinic. Furthermore, endoscopic evaluation does not always provide clear evidence for laryngeal hyperfunction (Behrman et al., 2003; Sama et al., 2001; Stager et al., 2000). Differential diagnosis of pMTD from other voice disorders using standard American Speech-Language-Hearing Association (ASHA)–recommended acoustic and aerodynamic measurements has also proven difficult to achieve (Shembel et al., 2021).

The heterogenous nature of pMTD results in a wide range of acoustic measurements in this population (Verdolini et al., 2006), which may explain inconsistent differences between pMTD and vocally healthy groups. For example, smoothed cepstral peak prominence (CPPS), a measure of voice periodicity and an acoustic correlate of dysphonia severity (Murton et al., 2020; Patel et al., 2018), has been found to be different between pMTD and vocally healthy samples (Toles & Shembel, 2024; Van Stan et al., 2021) but also overlapping (Belsky et al., 2021; Moore & Shembel, 2025). Similarly, the difference in amplitudes between the first and second harmonics (H1–H2), an acoustic correlate of glottal closure and spectral tilt, has been found to be both different (Van Stan et al., 2021) and overlapping (Toles & Shembel, 2024) between these populations. Mean sound pressure level (SPL) has also inconsistently differed between these populations (has differed in Belsky et al., 2021; Espinoza et al., 2017; but not in Toles & Shembel, 2024; Zheng et al., 2012). In contrast, no differences have been found in the ratio of low (< 4000 Hz) to high (> 4000 Hz) spectral energy (L/H ratio; Belsky et al., 2021; Toles & Shembel, 2024) or in the cepstral spectral index of dysphonia, a composite acoustic measure that correlates with dysphonia severity (Belsky et al., 2021). When examining MTD samples alone, one study found differences in fundamental frequency ($F0$) between pre- and post-therapy in male patients, suggesting improved tension in the thyrohyoid muscle (de Oliveira Lemos et al., 2018). However, this study included both primary and secondary MTD patients in its sample, therefore conflating possible etiologies. Regarding differences between disordered populations, acoustically disambiguating between adductor laryngeal dystonia (AdLD) and pMTD has received particular attention due to the perceptual similarity between these two disorders (Ludlow et al., 2018; Roy et al., 2007). For example, the average amount of automatically detected creaky voice using the Creak Detector (COVAREP; Degottex et al., 2014) has been shown to differ between these populations (Marks et al., 2023; Roy et al., 2024). However, further specificity of the acoustic heterogeneity of pMTD, particularly regarding metrics commonly taken during vocal assessment, may still be informative in understanding the variability seen in pMTD.

Aerodynamic measurements in the pMTD population also have wide ranges (Belsky et al., 2021; Gillespie et al., 2013; Higgins et al., 1999; Verdolini et al., 2006). In their description of a small pMTD sample, Hillman et al. (1989) identified three aerodynamic combinations: high airflow with normal subglottal pressure, normal airflow with high subglottal pressure, and high airflow with high subglottal pressure. In an expansion of this work, Gillespie et al. (2013) found clusters of these aerodynamic profiles as well as two more: normal airflow with normal subglottal pressure and low airflow with normal subglottal pressure. While this line of work is valuable in showing potential subgroups within pMTD, the maximum number of possible combinations of airflow and subglottal measurements is nearly met (i.e., remaining possible combinations are high airflow with low subglottal pressure, normal airflow with low subglottal pressure, low airflow with high subglottal pressure, low airflow with low subglottal pressure). Therefore, these studies suggest that the defining feature of pMTD may be its wide range of variability, rather than the presence of clearly delineated subgroups. Nevertheless, much more work in the area of aerodynamic measurements has focused on comparisons between pMTD and vocally healthy samples. When comparing mean measurements of pMTD individuals to vocally healthy controls, mean subglottal pressure (Espinoza et al., 2017; Zheng et al., 2012), inspiratory and expiratory airflow durations (Belsky et al., 2021), and open quotient (Espinoza et al., 2017) have been found to differ, but mean airflow during voicing has also not been found to differ between these pMTD and vocally healthy groups (Belsky et al., 2021). While these results are valuable in tracking vocal improvements in pMTD patients, there still exists a lack of specificity regarding how pMTD presents aerodynamically.

Previous work has defined subgroups of pMTD, but without quantitative measurements. Koufman and Blalock (1982) define five types of pMTD, of which three do not include structural abnormalities: hysterical aphonia/dysphonia, habituated hoarseness, and falsetto. These subgroups were based on a retrospective review of onset, laryngoscopic evaluation, associated symptoms, and perceptual voice quality. Of note, the range in sample size across these subgroups is one to 15 patients. In an overview of MTD, Spencer (2015) outlined glottal fry (creaky voice), excessively lowered $F0$, register flipping (alternating chest and higher registers), functional aphonia, dysphonia following phonosurgery, excessively pressed phonation, puberphonia, ventricular phonation unrelated to glottal insufficiencies, and functional hypophonia. These subgroups were generated from clinical observation and with consideration of work done to define subgroups at the time of publication but was not accompanied by any formal analysis.

Given the heterogeneous nature of this population, Shembel et al. (2021) call for movement away from mean-based statistical analyses (e.g., *t* tests, analyses of variance) when examining characteristics of this group. In their analysis, clusters of standard ASHA-recommended acoustic and aerodynamic metrics were used to predict diagnosis between pMTD and other voice disordered samples but do not find differential metrics. The current study acknowledges the heterogenous nature of this population in its design. We extend this line of work by (a) implementing data reduction on voice measurements before clustering, thereby allowing for commonalities between variables to be accounted for prior to clustering, and (b) examining clusters within a pMTD sample only, thereby allowing for intragroup comparisons.

Purpose

Given the multifactorial etiology of pMTD, the aim of this work is to define subgroups within a sample of patients with pMTD in an effort to elucidate the nature of the heterogeneity within this diagnosis. To this end, our first objective was to conduct an exploratory factor analysis (EFA) to find latent structures within a group of acoustic and aerodynamic measurements collected from adult patients diagnosed with pMTD during intake. Our second objective was to conduct a clustering analysis of a sample of patients with a primary diagnosis of muscle tension dysphonia using the factors identified in the first objective.

Method

Data Collection

This study is a retrospective chart review with approval from the NYU Langone Institutional Review Board (#s23-00936). We requested medical records through DataCore, the resource based in the Information Technology Department at NYU Langone. The search criteria for the DataCore request were as follows: (a) patients greater than or equal to 18 years old at the date of encounter and (b) patients who had a vocal function assessment completed by a speech-language pathologist at the NYU Langone Voice and Swallowing Center, Department of Otolaryngology, between January 1, 2021, and October 1, 2023. This patient cohort was subsetted by diagnosis. For each patient, we requested medical record number, first name, last name, age at date of encounter, legal sex, preferred language, visit type, procedure, medical diagnosis, and date of encounter. This data request resulted in a total of 2,775 unique encounters. The data set was then filtered for first encounters only, resulting in 2,713 patients.

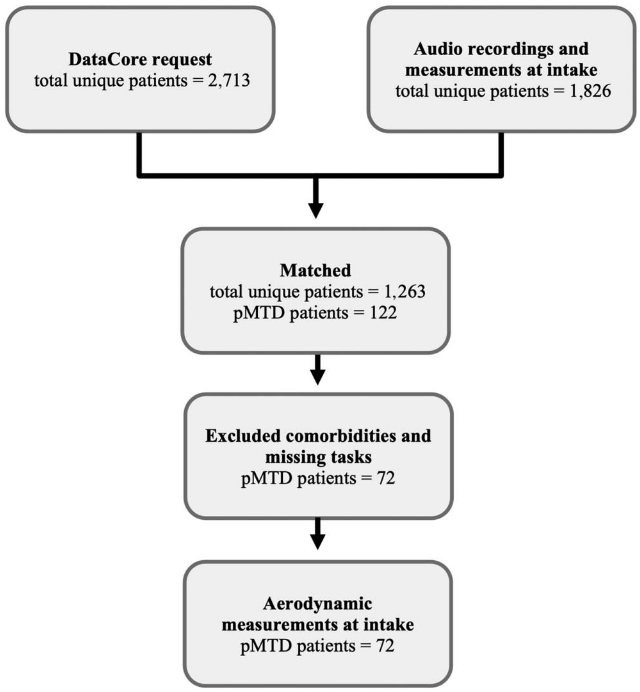
These data were then matched to audio recordings collected in the NYU Voice and Swallowing Center from patients at intake. During a patient's voice assessment, the patient was recorded producing the following tasks: three sustained /a/ vowels, one all-voiced Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) sentence ("We were away a year ago"), spontaneous speech, and two pitch glides (up and down). All recordings were made during standard clinical assessments in a quiet room (approximate signal-to-noise ratio of 30 dB) using a large-diaphragm condenser microphone (P220; AKG) positioned approximately 5 in. in front of the patient. Signals were digitized using a USB audio interface (USBPre 2; Sound Devices) that also provided phantom power to the microphone and pre-amplification of the signal at a fixed gain, allowing for full dynamic range capture of the audio signal without clipping. Recordings were saved using the free lossless audio codec file format for later analysis using Praat (Boersma & Weenink, 2025). Annotated Textgrids for these voice tasks were generated by clinicians at the time of intake using a custom Praat script created by the second author.

Duration of creaky voice was measured in the three sustained vowels, one all-voiced CAPE-V sentence, and the spontaneous speech sample. We ran an automatic creak detection algorithm (COVAREP; Degottex et al., 2014) on the extracted sound files in MATLAB. We used the wrapper developed by Marks et al. (2023) for easy-to-read output of the duration of creak over the duration of voiced segments for a given sound file.

Additional acoustic measurements of these files were also collected at the time of encounter using a custom Praat script created by the second author. These were the mean and standard deviation of CPPS during a sustained vowel production and the all-voiced CAPE-V sentence, the mean and standard deviation of $F0$ during spontaneous speech, and the minimum and maximum $F0$ for two pitch glide tasks (up and down). The database of acoustic measurements from May 2021 through March 2024 totaling 1,826 unique patients was matched to the files from the Creak Detector output and the Datacore request. We matched 1,263 unique encounters between the acoustic data, the Creak Detector output, and Datacore records (see Figure 1). All acoustic measurements for the three vowel productions, including proportion of creak, were averaged for each patient.

Aerodynamic measurements were also collected from patients at the time of encounter using the standard voicing efficiency protocol of the Phonatory Aerodynamic System (PENTAX Medical). Briefly, patients produced

Figure 1. Flowchart of sources for patient data. pMTD = primary muscle tension dysphonia.



five sequential /pa/ syllables at an approximate rate of 1.5 to two syllables per second at a comfortable pitch and loudness, per the recommended protocol from ASHA (Patel et al., 2018). Measures extracted from the middle three syllables included mean SPL during voicing (dB), mean peak air pressure during voicing (cmH20), and mean airflow during voicing (L/s). Additionally, the all-voiced CAPE-V sentence, "We were away a year ago," was captured to estimate average airflow (L/s) during speech. These measurements, along with the CAPE-V overall auditory-perceptual severity score, were only extracted from the medical record for the pMTD subset, which is described in more detail below.

Participants

Of the 1,263 audio-matched patients, 122 individuals received a primary diagnosis of pMTD. There were 27 individuals with a primary diagnosis of pMTD who had the following structural comorbidities: vocal fold polyp, sulcus vocalis, lesion of vocal fold, vocal cord nodules, history of tracheostomy, benign neoplasm of larynx, recurrent glottic respiratory papillomatosis, paradoxical vocal cord motion, glottic insufficiency, vocal fold atrophy, vocal fold paresis or paralysis, laryngeal spasm, laryngospasm, presbyphonia, or vocal cord dysfunction. These individuals were then excluded from analysis. This resulted in a total of 95 patients with a diagnosis of pMTD. We removed an additional 22 patients due to missing or inaccurate aerodynamic measurements and one patient due to missing vocal task recordings. This left a total of 72 patients for analysis. There were 48 patients who had a legal sex of female and 24 who had a legal sex of male. Information for gender identity and sex assigned at birth was not available for all patients. The age ranged from 19 to 79 years with a mean of 41.72 years. The preferred language for all patients was English.

Results

Factor Analysis

Factor analysis is a useful exploratory technique to discover latent structures of a data set. That is, factor analysis helps reduce dimensionality in a data set so that underlying commonalities between variables can be discovered. Principal axes (PAs), rather than factors, are used as the extraction method in this factor analysis due to it producing the smallest residuals of all methods. Full details on the preparation and selection of these PAs can be found in the Appendix. Using a promax oblique rotation for a four-PA EFA yielded two variables that did not meet the threshold of sufficient loading (|0.4|). Standard deviation of CPPS in spontaneous speech did not sufficiently load onto any PA (| < 0.40|) and was therefore excluded from the factor analysis. Rerunning the EFA with the updated set of variables yielded only two variables loaded onto the fourth PA, indicating that the EFA should be run with three, rather than four, PAs. After running this iteration of the factor analysis, "minimum *F*0 of pitch glide" did not sufficiently load onto any PA (| < 0.40|) and was therefore excluded from the factor analysis. The final factor analysis consisted of three PAs, with loadings above the |0.4| threshold on at least one PA (see Table 1). The proportion of variance explained by PA 1, PA 2, and PA 3 was 17.3%, 14.5%, and 12.9%, respectively, for a total variance explained by these factors of 44.7%. Age and aperiodicity measurements loaded onto PA 1 (standard deviation of CPPS in vowel production, severity score, percent creak in the all-voiced CAPE-V sentence, and percent creak in spontaneous speech). Mean CPPS in spontaneous speech and mean CPPS in vowel production contrasted with PA 1. All *F*0 measurements loaded onto PA 2 (mean *F*0 in spontaneous speech, standard deviation of *F*0 in spontaneous speech, max *F*0 of pitch glide). Mean SPL and aerodynamic measurements loaded onto PA 3 (mean peak pressure, mean airflow, average airflow in all-voiced CAPE-V sentence). Average airflow in the all-voiced CAPE-V sentence cross-loaded onto two PAs (i.e., the loadings are greater than |0.4| on both

Table 1. Factor loadings for variables included in exploratory factor analysis.

Variable type	Variable	PA 1	PA 2	PA 3
Demographic	Age	**0.438**	−0.370	
Perceptual	Severity score	**0.627**	−0.322	0.161
Acoustic	Percent creak in spontaneous speech	**0.657**		0.107
	Mean CPPS in spontaneous speech	**−0.615**		
	Mean CPPS in vowel	**−0.602**		0.149
	Percent creak in all-voiced CAPE-V	**0.486**	0.104	
	Standard deviation of CPPS in vowel	**0.484**		
	Max *F*0 of pitch glide	−0.158	**0.751**	0.270
	Mean *F*0 in spontaneous speech	−0.112	**0.652**	−0.120
	Standard deviation of *F*0 in spontaneous speech	0.322	**0.589**	0.217
	Mean SPL		0.178	**0.721**
Aerodynamic	Mean peak pressure		0.249	**0.712**
	Average airflow in all-voiced CAPE-V	−0.165	**−0.437**	**0.556**
	Mean airflow		−0.385	**0.516**

Note. PA = principal axis; CPPS = smoothed cepstral peak prominence; *F*0 = fundamental frequency; SPL = sound pressure level; CAPE-V = Consensus Auditory-Perceptual Evaluation of Voice. Values greater than |0.4| are boldfaced for ease of interpretation.

PAs). Dropping this variable from the analysis led to a worse-fitting PA structure, so it was retained.

Factor scores were then computed for individual patients using the regression method where the standardized observed values are weighted by factor loading coefficients (Estabrook & Neale, 2013; Tabachnick & Fidell, 2013; Thurstone, 1935). These scores were then used as input to a model-based clustering algorithm.

Clustering Analysis

k-means clustering analysis is an unsupervised machine learning technique to group data. Here, we use a nonparametric unsupervised clustering method to explore possible clusters within the data set. First, we computed the total within-cluster sum of squares to determine the number of clusters that minimizes intracluster variation. A within-cluster sum of squared distances plot was generated to visualize the optimal number of clusters using the "elbow" method (see Figure 2, left panel). This plot demonstrated that four clusters was the optimal number. As an additional check, a silhouette plot was generated (see Figure 2, right panel). This plot also demonstrated that four clusters was the optimal number of clusters.

We used a four-cluster solution with 100 initial starts and 1,000,000 iterations (see Figure 3). This yielded a compactness percentage (average sum of squared distances between centroids divided by total sum of squared distances between points and centroid for each cluster plus average sum of squared distances between centroids) of 62.4%, which is an indicator of how similar members are within a cluster and how well-separated each cluster is from another. Cluster means and sizes are given in Table 2.

These four clusters are clearly differentiable along at least one PA (see Figure 4). Cluster 1 had much higher factor scores along PA 1. Clusters 2 and 3 had much lower and higher factor scores along PA 2, respectively. Cluster 4 had much higher factor scores along PA 3. An approximately even number of patients were assigned to each cluster (see Table 2). In the following paragraphs, each cluster will be described more specifically. Descriptive statistics for each cluster can be found in Table 3.

Cluster 1 is characterized by higher factor scores of age and aperiodicity measurements, mid-range factor scores for *F*0 measurements, and low factor scores for aerodynamic measurements. This indicates a clustering of pMTD patients that are older and have higher aperiodicity in their voice and were rated as having higher severity, have mid-range *F*0 values, and have lower airflow. We can call this cluster "Aperiodic," since it is most distinguishable along the first PA, which primarily has to do with aperiodicity measurements.

Cluster 2 is characterized by low factor scores of age and aperiodicity measurements, low factor scores for *F*0 measurements, and mid-range factor scores for aerodynamic measurements. This indicates a clustering of pMTD patients who are younger and have lower levels of aperiodicity in their voice and lower severity, have lower *F*0 values, and have mid-range airflow. We can call this cluster "Low *F*0," since it is most distinguishable along the second PA, which primarily has to do with *F*0 measurements such as mean *F*0 and maximum *F*0.

Cluster 3 is characterized by low factor scores of age and aperiodicity measurements, high factor scores for *F*0 measurements, and mid-range factor scores for aerodynamic measurements. This indicates a clustering of pMTD

Figure 2. Left: Within-cluster sum of squared distances from the centroids is plotted on the *x*-axis given the number of clusters on the *y*-axis. Dotted line indicates the optimal number of clusters given visual inspection of the where the “elbow” bends. Right: Silhouette plot showing the numerical output of the distance of individual points to points in neighboring clusters. The dotted line indicates the optimal number of clusters given the highest silhouette width.

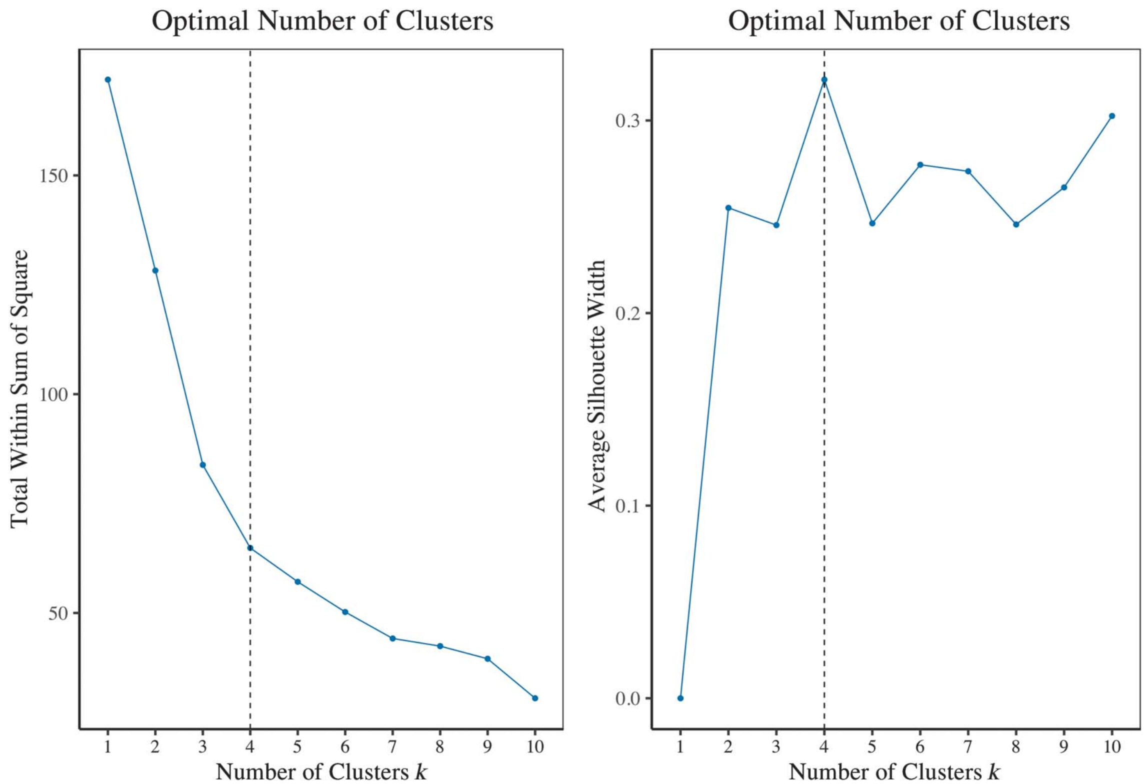


Figure 3. Left: Individual patients’ factor scores as points for three principal axes (PAs) in a three-dimensional scatter plot. Each axis is a PA. Points are colored according to cluster assignments identified by *k*-means analysis. Right: Three plots each showing only two principal axes each. Individual patients’ factor scores are points for three PAs. Each plot axis is a principal axis. Points are colored according to cluster assignments identified by *k*-means analysis. Ellipses are set at a 95% confidence interval.

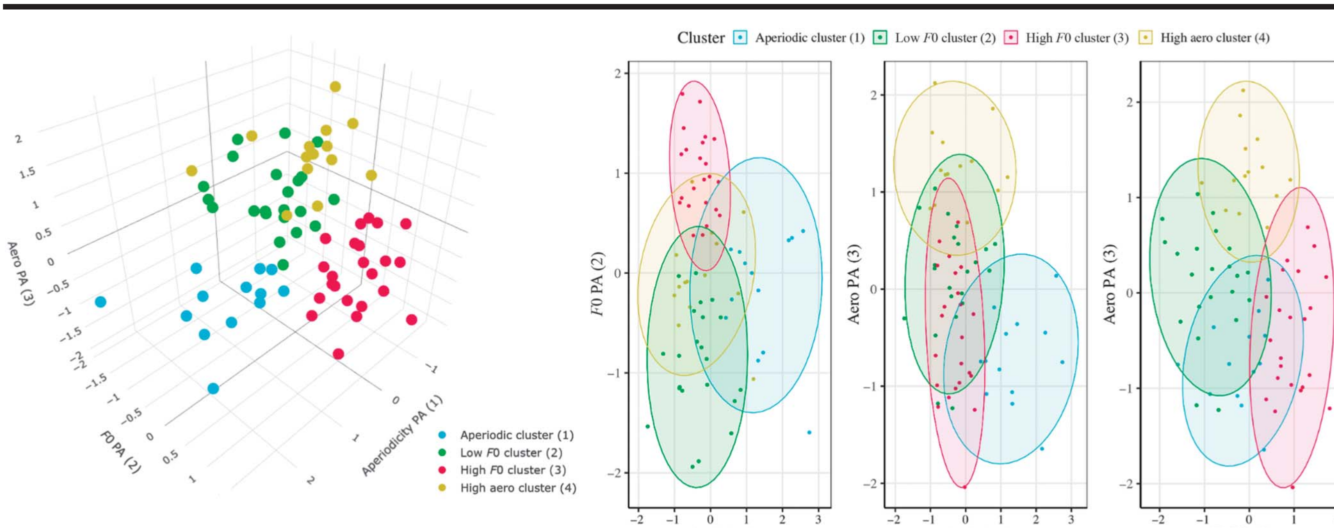


Table 2. Factor loadings for variables included in exploratory factor analysis.

Cluster	Aperiodicity PA (1) mean	*F*0 PA (2) mean	Aero PA (3) mean	*N*
Aperiodic Cluster (1)	1.410	–0.193	–0.719	13
Low *F*0 Cluster (2)	–0.360	–0.899	0.116	22
High *F*0 Cluster (3)	–0.313	0.982	–0.480	23
High Aero Cluster (4)	–0.230	–0.021	1.270	14

Note. PA = principal axis; *F*0 = fundamental frequency.

patients who are younger and have low aperiodicity, have high *F*0 values, and have low airflow. We can call this cluster "High *F*0," since it is most distinguishable along the second PA, which primarily has to do with *F*0 measurements such as mean *F*0 and maximum *F*0.

Cluster 4 is characterized by low factor scores of age and aperiodicity measurements, mid-range factor scores for *F*0 measurements, and high factor scores for aerodynamic measurements. This indicates a clustering of pMTD patients who are younger and have a low amount of aperiodicity in their voice and low severity, have mid *F*0 range, and have high levels of sound pressure. We can call this cluster "High Aerodynamic," since it is most distinguishable along the third PA, which primarily has to do with aerodynamic measurements such as mean peak pressure and mean airflow.

Semitone Conversion

A follow-up analysis was conducted to address concerns that Clusters 2 and 3 may be most differentiated by legal sex (by way of *F*0 values), which is not highly informative for the objectives of the current study. That is, these two clusters were most differentiated by PA 2, which is composed of *F*0 mean and standard deviation during spontaneous speech and maximum *F*0 in the pitch glide task. A post hoc inspection of individual patient cluster assignments revealed both male and female patients in Cluster 2 and all female patients in Cluster 3. While a

Figure 4. Box plots of individual patients' factor scores. Cluster assignments are on the *x*-axis. Factor scores of patients for each principal axis (PA) are given on the *y*-axis. *F*0 = fundamental frequency.

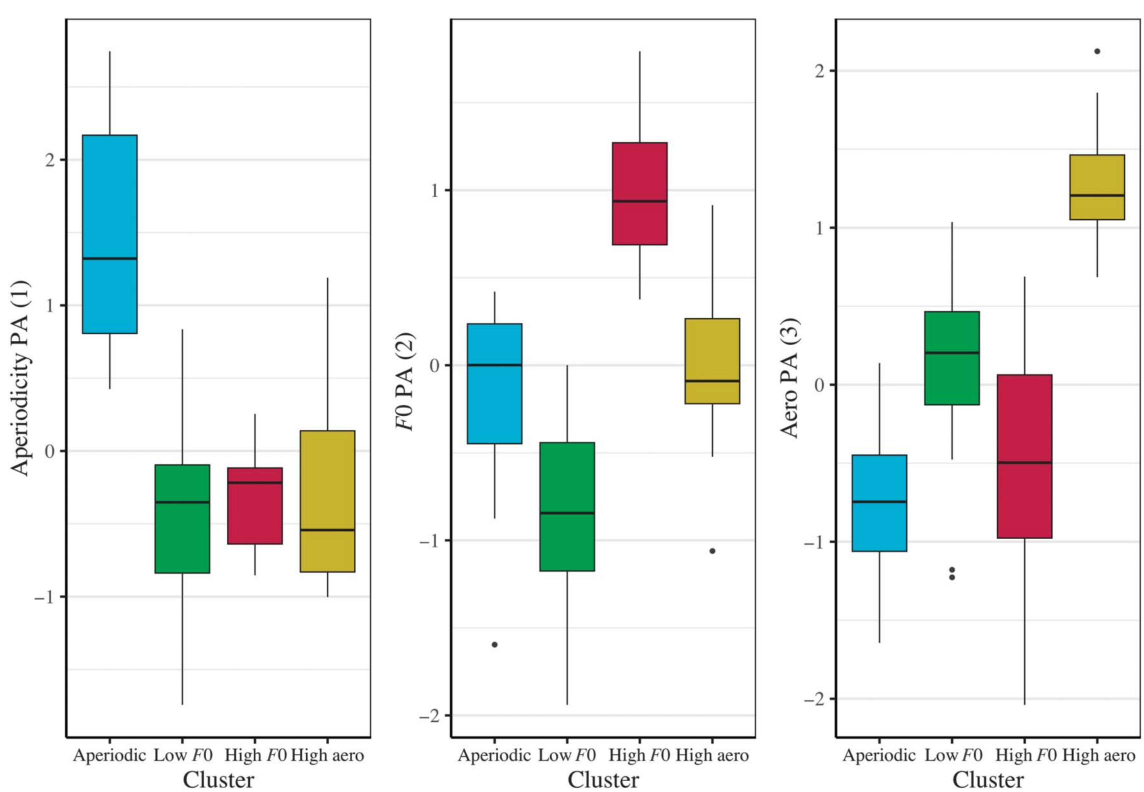


Table 3. Descriptive statistics for all variables for each cluster.

Variable type	Variable	Aperiodic Cluster (1)	Low *F*0 Cluster (2)	High *F*0 Cluster (3)	High Aero Cluster (4)
Demographic	Age	57.4 *(14.1)*	47.6 *(19)*	29.8 *(9.04)*	37.6 *(11.9)*
		26–76	21–79	19–55	23–69
Perceptual	Severity score	34.2 *(20.8)*	21 *(15.7)*	9.17 *(5.11)*	18.9 *(11)*
		8–67	0–72	0–20	5–40
Acoustic	Percent creak in spontaneous speech	18.9 *(11.3)*	3.73 *(4.52)*	10.6 *(10.2)*	10.6 *(10.2)*
		3–34.9	0–16.6	0.136–15.7	0.946–36.4
	Mean CPPS in spontaneous speech	9.14 *(1.84)*	11.6 *(1.81)*	11.5 *(1.71)*	11.3 *(1.35)*
		5.69–13.1	7.8–15.4	8.49–14.8	8.81–13.5
	Mean CPPS in vowel	11 *(2.63)*	15.5 *(3.53)*	14.2 *(2.52)*	16.1 *(2.78)*
		6.23–14.17	9.02–21.9	9.75–18.7	11.7–20.1
	Percent creak in all-voiced CAPE-V	20 *(8.9)*	6.07 *(8.63)*	8.99 *(7.29)*	7.47 *(5.88)*
		8.65–38.7	0–31	0–24.5	0–18.9
	Standard dev. CPPS in vowel	0.869 *(1.63)*	0.227 *(1.48)*	0.21 *(1.66)*	0.626 *(1.45)*
		1.16–3.68	1.04–1.89	1.25–2.16	1.31–3.66
	Max *F*0 of pitch glide	448 *(169)*	413 *(102)*	792 *(230)*	686 *(203)*
		116–758	229–570	460–1199	415–1034
	Mean *F*0 in spontaneous speech	139 *(25.1)*	129 *(27.4)*	192 *(25.8)*	145 *(26)*
		99–184	86.5–185	148–237	89.2–184
	Standard dev. of *F*0 in spontaneous speech	37.3 *(11.7)*	23.5 *(11.3)*	39.1 *(10.2)*	38.1 *(12.5)*
		15–59.9	8.42–61.5	23.3–62	19.9–70.6
	Mean SPL	82.9 *(4.49)*	85.2 *(3.42)*	84.8 *(3.73)*	92.2 *(3.02)*
		76.5–88.9	79.6–92	78.7–91.8	88–99.2
Aerodynamic	Mean peak pressure	6.87 *(1.31)*	7.86 *(2.03)*	8 *(1.93)*	11.2 *(2.51)*
		4.96–9.03	4.32–11.3	5.27–11.5	7.62–15.6
	Average airflow in all-voiced CAPE-V	0.113 *(0.058)*	0.199 *(0.055)*	0.103 *(0.044)*	0.214 *(0.041)*
		0.023–0.194	0.071–0.3	0.019–0.162	0.155–0.279
	Mean airflow	0.113 *(0.064)*	0.2 *(0.078)*	0.12 *(0.048)*	0.218 *(0.047)*
		0.051–0.292	0.093–0.358	0.012–0.194	0.138–0.289

Note. Numbers represent the mean (*standard deviation*) and range. *F*0 = fundamental frequency; CPPS = smoothed cepstral peak prominence; Standard dev. = standard deviation; SPL = sound pressure level; CAPE-V = Consensus Auditory-Perceptual Evaluation of Voice.

heterogenous cluster was encouraging, we converted all *F*0 measurements to semitones for more conservative comparisons and to more closely approximate the perceptual experience of clinicians during voice assessments. Given an octave divided into 12 notes on a logarithmic frequency scale, a semitone is the interval between two adjacent notes. The following equations were used to compute semitone range and semitone mean and standard deviation during spontaneous speech:

$$\text{semitone range} = 12 \times \log_2 \frac{high\ F0}{low\ F0} \quad (1)$$

$$\text{semitone} = \frac{\ln\left(\frac{frequency}{50}\right)}{\ln 2} \quad (2)$$

We conducted a new EFA with *F*0 measurements transformed to semitones using the same procedures outlined above. This new analysis yielded two PAs (see Table 4) with the same composition as PA 1 and PA 3 in the original analysis. However, semitone variables did not sufficiently load onto any PA. The proportion of variance explained for PA 1′ and PA 2′ were 21% and 17.1% of the variance, respectively, for a total variance explained by these factors of 38.1%. This is a smaller number than in the previous factor analysis using *F*0 values. Because less variance was accounted for using these variables, a clustering analysis was not pursued.

Discussion

Due to the heterogenous presentation of pMTD, the current study sought to identify possible subgroups of this voice disorder. We found that there are reliable, moderately separated (initial starts = 100, iterations = 1,000,000, compactness percentage = 62.4%) clusters within our sample of pMTD patients using age, auditory-perceptual

Table 4. Factor loadings for variables included in exploratory factor analysis with semitone conversion.

Variable type	Variable	PA 1′	PA 2′
Demographic	Age	**0.521**	0.146
Perceptual	Severity score	**0.692**	0.245
Acoustic	Percent creak in spontaneous speech	**0.516**	
	Mean CPPS in spontaneous speech	**−0.689**	
	Mean CPPS in vowel	**−0.624**	0.133
	Percent creak in all-voiced CAPE-V	**0.407**	−0.133
	Standard deviation of CPPS in vowel	**0.491**	
Aerodynamic	Mean SPL	−0.106	**0.504**
	Mean peak pressure		**0.488**
	Average airflow in all-voiced CAPE-V		**0.831**
	Mean airflow		**0.759**

Note. PA = principal axis; CPPS = smoothed cepstral peak prominence; SPL = sound pressure level; CAPE-V = Consensus Auditory-Perceptual Evaluation of Voice. Values greater than |0.4| are boldfaced for ease of interpretation.

severity scores, and acoustic and aerodynamic measurements. Specifically, we found one cluster primarily defined by high aperiodicity, one by low *F*0 values, one by high *F*0 values, and one by high aerodynamic values. The clusters of pMTD patients computed using multiparametric analyses in the current study provide evidence for more concrete profiles of pMTD subgroups in the larger pMTD population.

Our EFA reduced a larger set of variables to three PAs from which we were able to see commonalities in age, auditory-perceptual severity scores, and acoustic and aerodynamic measurements. The majority of variables loaded onto PAs in ways that are conceptually coherent. The first PA (PA 1) mostly grouped together variables that measured aperiodicity, the second PA (PA 2) mostly grouped together variables that measured *F*0, and the third PA (PA 3) mostly grouped together aerodynamic variables. There are a few exceptions to note. The first exception is that PA 1 included age and severity score in addition to acoustic variables quantifying aperiodicity in the signal. Given that perceived severity score has been shown to correlate to aperiodicity in the signal (Awan & Roy, 2009; Lee et al., 2020), the loading of severity score onto PA 1 is in line with a coherent interpretation. Correlation between older age and increased noise has also been demonstrated in Stathopoulos et al. (2011). The current study provides further support for age, severity score, and acoustic aperiodicity measurements to explain similar variance in patient data, especially given that this data set included a wide range of adult ages. The second exception is that average airflow in the all-voiced CAPE-V sentence cross-loaded onto both PA 2 and PA 3. Cross-loaded variables are not inherently a problem. Furthermore, this variable did not strongly load onto more than one PA (less than |0.5|; Costello & Osborne, 2005). However, interpretation is more ambiguous. This variable negatively loaded onto PA 2, meaning it was moderately negatively correlated with *F*0 measurements. That is, as average airflow decreases, measures of *F*0 increase. Therefore, this cluster may present with a more pressed or strained voice quality (Grillo & Verdolini, 2008; Netsell et al., 1984). However, average airflow more strongly, and positively, loaded onto PA 3, which also included all other aerodynamic measurements. Therefore, an interpretation of average airflow along PA 3 is also justified. The third exception is that mean SPL loaded onto PA 3, which otherwise comprises aerodynamic variables. However, SPL often accompanies aerodynamic measurements when considering vocal efficiency (Tanaka & Gould, 1985; Titze et al., 2016), which can be thought of as the efficient conversion of energy into acoustic output with minimal airflow loss (Titze, 1992). Therefore, the inclusion of SPL on PA 3 still allows for a coherent interpretation of this PA as one that concerns phonation behavior.

The clustering analysis found one cluster primarily defined by high aperiodicity, one by low *F*0 values, one by high *F*0 values, and one by high aerodynamic values. Due to concerns that Clusters 2 and 3 were largely clustering patients by legal sex, a post hoc inspection of individual patient cluster assignments was performed. This revealed both male and female patients were present in Cluster 2 and all female patients were present in Cluster 3. It is important to note that there were twice as many female patients as there were male patients included in the current sample, so a cluster entirely composed of female patients is not improbable. Nevertheless, a follow-up analysis was performed with *F*0 measurements transformed to semitones to more accurately reflect the perceptual nature of pitch during voice assessments (given that humans perceive pitch logarithmically, rather than linearly). Semitone measurements did not sufficiently load onto PAs, leading to their exclusion. However, this analysis yielded two PAs that had identical variable compositions to those in the original analysis, indicating a degree of robustness for PA 1 and PA 3. This analysis resulted in less variance explained than in the original analysis, indicating a weaker ability for these PAs to account for variance in the sample. For this reason, we recommend the original analysis, including *F*0 measurements, to be used as the primary analysis for interpretation. It is additionally important to note that measurements of *F*0 included maximum *F*0 during a pitch glide and standard deviation of *F*0 during spontaneous speech in addition to mean *F*0 during spontaneous speech, so measures of variability are also included in the PA for which these clusters in the original analysis were most differentiable. Furthermore, patients in Cluster 3 may have exhibited a more strained voice quality as seen by the negative factor loading of mean airflow on PA 3. This may be a more informative interpretation for finding subgroups within a pMTD sample.

The range in aerodynamic values across clusters in the present study supports previous work by Hillman et al. (1989) and Gillespie et al. (2013) in demonstrating subgroups within pMTD based on airflow and air pressure measurements. In particular, we found one cluster to have high values of airflow and air pressure (along PA 3) in comparison to other clusters. However, we are unable to discuss other combinations of airflow and air pressure measurements due to the nature of data reduction in the current analysis. Given that Clusters 2 and 3 may only be differentiating legal sex, this analysis may provide stronger motivation for investigating aperiodicity (amount of creak in spontaneous speech, severity score) and air pressure levels as potential avenues for specifying subgroups of pMTD. That is, relatively high aperiodicity and relatively high aerodynamic measurements were two well-defined subgroups that emerged from the current study. Patients who report excessive muscle tension with these two distinct sets of characteristics may need to receive separate diagnoses to receive more appropriate and efficient care.

Results from the current study show some overlap with findings from Shembel et al. (2021). While their analysis did not show strong evidence for the ability to distinguish pMTD from other voice disorders using standard acoustic and aerodynamic metrics, the most differentiating variables examined were highest *F*0, mean cepstral peak prominence (CPP) of a sustained vowel, mean airflow, and mean peak pressure. These variables all loaded onto PAs in the current study. While not the main goal of Shembel et al. (2021), a difference in pMTD clusters was seen for mean CPP of a sustained vowel. This finding is reflected in the current study, given that the mean CPPS in vowel production somewhat strongly loaded onto PA 2.

The current study also shares some overlap with previous attempts at defining subgroups of pMTD. Specifically, the current study provides evidence for a highly aperiodic subgroup, a low *F*0 subgroup, a high *F*0 (and possibly low airflow) subgroup, and a high aerodynamic subgroup. The first two subgroups map neatly onto two of the subgroups previously proposed by Spencer (2015). The highly aperiodic subgroup and the high *F*0 subgroup may also support the "habituated hoarseness" and "falsetto" subgroups, respectively, as defined by Koufman and Blalock (1982). For the last subgroup, the current study showed a higher weighting of pressure measurements (SPL and mean peak pressure) than airflow measurements on the PA for which this subgroup had a higher mean. The cluster with higher pressure measurements is in line with the "excessively pressed phonation" subgroup previously identified by Spencer (2015).

The current study also provides further support for the inclusion of creaky voice in differentiating types, or possibly subgroups, of voice disorders. Automatic detection of creaky voice using the Creak Detector has shown to be useful in differentiating pMTD from AdLD (Marks et al., 2023; Roy et al., 2024). The current study found the amount of creak in spontaneous speech to strongly load onto PA 2, which includes other measurements of aperiodicity. Several types of creaky voice are characterized by aperiodicity (Keating et al., 2015), and aperiodicity is one component of the Creak Detector in its detection of creaky voice (specifically intraframe periodicity as originally computed in Ishi et al., 2008). Here, we caution that the Creak Detector is an automatic detection method and therefore may be best interpreted as a reliable differentiating measure rather than as a valid detector of true instances of creak.

Clinical Implications

The identification of four distinct pMTD subgroups suggests several directions for clinical practice, though implementation will require further validation. The clustering results indicate that comprehensive assessment should include measurements across all three PAs: aperiodicity (including severity ratings and acoustic noise measures), *F*0 characteristics, and aerodynamic function. Clinicians may benefit from prioritizing these specific measures when evaluating pMTD patients, as they appear to capture meaningful variation within this population. The strong loading of creaky voice detection onto the aperiodicity axis suggests that systematic assessment of vocal fry may help differentiate patient subgroups. However, given the limitations of automatic detection methods noted in our study, clinical judgment should complement these measures.

Ultimately, identifying subgroups of pMTD could lead to distinct therapeutic approaches. For example, patients in the high-aperiodicity cluster may benefit from interventions targeted at improving vocal fold vibration regularity, while those in high-aerodynamic clusters might need approaches focused on respiratory–phonatory coordination. However, our findings represent preliminary evidence for pMTD subgroups and require replication before clinical implementation.

Limitations

Due to the data reduction nature of analyses conducted in this study, more patients would have provided a more robust analysis. Having several hundred observations for an EFA is generally recommended (Goretzko et al., 2021), but good levels of agreement between sample and population data include variables that share a sufficiently high amount of communality and a sufficiently high variables-to-factors ratio (Mundfrom et al., 2005). In the current study, communality between variables and the variable-to-factors ratio were both high. While this provides reassurance for the current results, we encourage future attempts to replicate these results with larger sample sizes.

Conclusion

The aim of the current study was not to diagnose pMTD but rather to characterize possible subgroups within this heterogenous population. We found reliable, moderately separated clusters that differ along axes of aperiodicity, *F*0, and aerodynamic measurements. It may be the case that subgroups of pMTD along these lines could prove useful in further defining this heterogenous population. The results presented here may inform more specific diagnoses and appropriate treatment decisions based on subgroup classification. Furthermore, this study supports previous work demonstrating the utility of machine learning techniques, rather than mean-based statistical tests, when investigating qualities of pMTD samples (Shembel et al., 2021). Finally, this analysis would greatly benefit from replication with different samples within this population to test the robustness of the clusters identified within the current sample and to inform future work establishing subtypes of pMTD.

Data Availability Statement

The script for data analysis can be found at https://osf.io/et6am/.

Acknowledgments

The authors received no financial support for the research, authorship, or publication of this article. We would like to thank our anonymous reviewers for their invaluable feedback and Hamzad Persaud for his assistance in data collection through DataCore at NYU Langone.

References

Awan, S. N., & Roy, N. (2009). Outcomes measurement in voice disorders: Application of an acoustic index of dysphonia severity. *Journal of Speech, Language, and Hearing Research, 52*(2), 482–499. https://doi.org/10.1044/1092-4388(2008/08-0034)

Behrman, A., Dahl, L. D., Abramson, A. L., & Schutte, H. K. (2003). Anterior-posterior and medial compression of the supraglottis: Signs of nonorganic dysphonia or normal postures? *Journal of Voice, 17*(3), 403–410. https://doi.org/10.1067/S0892-1997(03)00018-3

Belafsky, P. C., Postma, G. N., Reulbach, T. R., Holland, B. W., & Koufman, J. A. (2002). Muscle tension dysphonia as a sign of underlying glottal insufficiency. *Otolaryngology—Head and Neck Surgery, 127*(5), 448–451. https://doi.org/10.1067/mhn.2002.128894

Belsky, M. A., Rothenberger, S. D., Gillespie, A. I., & Gartner-Schmidt, J. L. (2021). Do phonatory aerodynamic and acoustic measures in connected speech differ between vocally healthy adults and patients diagnosed with muscle tension dysphonia? *Journal of Voice, 35*(4), 663.e1–663.e7. https://doi.org/10.1016/j.jvoice.2019.12.019

Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.35) [Computer software]. https://www.praat.org

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10,* Article 7. https://doi.org/10.7275/jyj1-4868

Courtney, M. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2 0 to make more judicious estimations. *Practical Assessment, Research, and Evaluation, 18*(1), Article 8. https://doi.org/10.7275/9CF5-2M72

de Oliveira Lemos, I., da Cunha Pereira, G., Druck SantAnna, G., & Cassol, M. (2018). Effects of a voice therapy program for patients with muscle tension dysphonia. *Folia Phoniatrica et Logopaedica, 69*(5–6), 239–245. https://doi.org/10.1159/000487942

Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, S. (2014). COVAREP—A collaborative voice analysis repository for speech technologies. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 960–964. https://doi.org/10.1109/ICASSP.2014.6853739

Desjardins, M., Apfelbach, C., Rubino, M., & Verdolini Abbott, K. (2022). Integrative review and framework of suggested mechanisms in primary muscle tension dysphonia. *Journal of Speech, Language, and Hearing Research, 65*(5), 1867–1893. https://doi.org/10.1044/2022_JSLHR-21-00575

Espinoza, V. M., Zañartu, M., Van Stan, J. H., Mehta, D. D., & Hillman, R. E. (2017). Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 60*(8), 2159–2169. https://doi.org/10.1044/2017_JSLHR-S-16-0337

Estabrook, R., & Neale, M. (2013). A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to nicotine dependence. *Multivariate*

Behavioral Research, 48(1), 1–27. https://doi.org/10.1080/00273171.2012.730072

Gillespie, A. I., Gartner-Schmidt, J., Rubinstein, E. N., & Abbott, K. V. (2013). Aerodynamic profiles of women with muscle tension dysphonia/aphonia. *Journal of Speech, Language, and Hearing Research, 56*(2), 481–488. https://doi.org/10.1044/1092-4388(2012/11-0217)

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology, 40*(7), 3510–3521. https://doi.org/10.1007/s12144-019-00300-2

Grillo, E. U., & Verdolini, K. (2008). Evidence for distinguishing pressed, normal, resonant, and breathy voice qualities by laryngeal resistance and vocal efficiency in vocally trained subjects. *Journal of Voice, 22*(5), 546–552. https://doi.org/10.1016/j.jvoice.2006.12.008

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17*(1), 65–70. https://doi.org/10.1111/j.2044-8317.1964.tb00244.x

Higgins, M. B., Chait, D. H., & Schulte, L. (1999). Phonatory air flow characteristics of adductor spasmodic dysphonia and muscle tension dysphonia. *Journal of Speech, Language, and Hearing Research, 42*(1), 101–111. https://doi.org/10.1044/jslhr.4201.101

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32*(2), 373–392. https://doi.org/10.1044/jshr.3202.373

Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zañartu, M., & Mehta, D. D. (2020). An updated theoretical framework for vocal hyperfunction. *American Journal of Speech-Language Pathology, 29*(4), 2254–2260. https://doi.org/10.1044/2020_AJSLP-20-00104

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Ishi, C. T., Sakakibara, K.-I., Ishiguro, H., & Hagita, N. (2008). A method for automatic detection of vocal fry. *IEEE Transactions on Audio, Speech, and Language Processing, 16*(1), 47–56. https://doi.org/10.1109/TASL.2007.910791

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement, 34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. International Congress of Phonetic Sciences.

Khoddami, S. M., Ansari, N. N., & Jalaie, S. (2015). Review on laryngeal palpation methods in muscle tension dysphonia: Validity and reliability issues. *Journal of Voice, 29*(4), 459–468. https://doi.org/10.1016/j.jvoice.2014.09.023

Koufman, J. A., & Blalock, P. D. (1982). Classification and approach to patients with functional voice disorders. *Annals of Otology, Rhinology & Laryngology, 91*(4), 372–377. https://doi.org/10.1177/000348948209100409

Koufman, J. A., & Blalock, P. D. (1988). Vocal fatigue and dysphonia in the professional voice user: Bogart-Bacall syndrome. *The Laryngoscope, 98*(5), 493–498. https://doi.org/10.1288/00005537-198805000-00003

Kunduk, M., Fink, D. S., & McWhorter, A. J. (2016). Primary muscle tension dysphonia. *Current Otorhinolaryngology Reports, 4*(3), 175–182. https://doi.org/10.1007/s40136-016-0123-3

Lee, Y., Kim, G., & Kwon, S. (2020). The usefulness of auditory perceptual assessment and acoustic analysis for classifying the voice severity. *Journal of Voice, 34*(6), 884–893. https://doi.org/10.1016/j.jvoice.2019.04.013

Ludlow, C. L., Domangue, R., Sharma, D., Jinnah, H. A., Perlmutter, J. S., Berke, G., Sapienza, C., Smith, M. E., Blumin, J. H., Kalata, C. E., Blindauer, K., Johns, M., Hapner, E., Harmon, A., Paniello, R., Adler, C. H., Crujido, L., Lott, D. G., Bansberg, S. F., Barone, N., Drulia, T., & Stebbins, G. (2018). Consensus-based attributes for identifying patients with spasmodic dysphonia and other voice disorders. *JAMA Otolaryngology–Head & Neck Surgery, 144*(8), 657–665. https://doi.org/10.1001/jamaoto.2018.0644

Marks, K. L., Díaz Cádiz, M. E., Toles, L. E., Buckley, D. P., Tracy, L. F., Noordzji, J. P., Grillone, G. A., & Stepp, C. E. (2023). Automated creak differentiates adductor laryngeal dystonia and muscle tension dysphonia. *The Laryngoscope, 133*(10), 2687–2694. https://doi.org/10.1002/lary.30588

Moore, A., & Shembel, A. C. (2025). Salient voice symptoms in primary muscle tension dysphonia. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2024.12.020

Morrison, M. D., Rammage, L. A., Belisle, G. M., Pullan, C. B., & Nichol, H. (1983). Muscular tension dysphonia. *The Journal of Otolaryngology, 12*(5), 302–306.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Murton, O., Hillman, R., & Mehta, D. (2020). Cepstral peak prominence values for clinical voice evaluation. *American Journal of Speech-Language Pathology, 29*(3), 1596–1607. https://doi.org/10.1044/2020_AJSLP-20-00001

Netsell, R., Lotz, W., & Shaughnessy, A. L. (1984). Laryngeal aerodynamics associated with selected voice disorders. *American Journal of Otolaryngology, 5*(6), 397–403. https://doi.org/10.1016/S0196-0709(84)80055-1

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Revelle, W. (2007). *psych: Procedures for psychological, psychometric, and personality research* (p. 2.4.6.26) [Data set]. https://doi.org/10.32614/CRAN.package.psych

Roy, N. (2008). Assessment and treatment of musculoskeletal tension in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology, 10*(4), 195–209. https://doi.org/10.1080/17549500701885577

Roy, N. (2010). Differential diagnosis of muscle tension dysphonia and spasmodic dysphonia. *Current Opinion in Otolaryngology & Head & Neck Surgery, 18*(3), 165–170. https://doi.org/10.1097/MOO.0b013e328339376c

Roy, N., Awan, S. N., Jennings, S., Jensen, J., & Merrill, R. M. (2024). Adductor laryngeal dystonia versus muscle tension dysphonia: Examining the utility of automated acoustic analysis to detect task dependency as a distinguishing feature. *Journal of Speech, Language, and Hearing Research, 67*(10), 3612–3630. https://doi.org/10.1044/2024_JSLHR-24-00104

Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: A foundation for future research. *Journal of Speech, Language, and Hearing Research, 43*(3), 737–748. https://doi.org/10.1044/jslhr.4303.737

Roy, N., Smith, M. E., Allen, B., & Merrill, R. M. (2007). Adductor spasmodic dysphonia versus muscle tension dysphonia: Examining the diagnostic value of recurrent laryngeal nerve lidocaine block. *Annals of Otology, Rhinology & Laryngology, 116*(3), 161–168. https://doi.org/10.1177/000348940711600301

Sama, A., Carding, P. N., Price, S., Kelly, P., & Wilson, J. A. (2001). The clinical features of functional dysphonia. *The Laryngoscope, 111*(3), 458–463. https://doi.org/10.1097/00005537-200103000-00015

Shembel, A. C., Lee, J., Sacher, J. R., & Johnson, A. M. (2021). Characterization of primary muscle tension dysphonia using acoustic and aerodynamic voice metrics. *Journal of Voice, 37*(6), 897–906. https://doi.org/10.1016/j.jvoice.2021.05.019

Spencer, M. L. (2015). Muscle tension dysphonia: A rationale for symptomatic subtypes, expedited treatment, and increased therapy compliance. *Perspectives on Voice and Voice Disorders, 25*(1), 5–15. https://doi.org/10.1044/vvd25.1.5

Stager, S. V., Bielamowicz, S. A., Regnell, J. R., Gupta, A., & Barkmeier, J. M. (2000). Supraglottic activity: Evidence of vocal hyperfunction or laryngeal articulation? *Journal of Speech, Language, and Hearing Research, 43*(1), 229–238. https://doi.org/10.1044/jslhr.4301.229

Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research, 54*(4), 1011–1021. https://doi.org/10.1044/1092-4388(2010/10-0036)

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tanaka, S., & Gould, W. J. (1985). Vocal efficiency and aerodynamic aspects in voice disorders. *Annals of Otology, Rhinology & Laryngology, 94*(1), 29–33. https://doi.org/10.1177/000348948509400107

Thurstone, L. L. (1935). *The vectors of mind*. The University of Chicago Press.

Titze, I. R. (1992). Vocal efficiency. *Journal of Voice, 6*(2), 135–138. https://doi.org/10.1016/S0892-1997(05)80127-4

Titze, I. R., Maxfield, L., & Palaparthi, A. (2016). An oral pressure conversion ratio as a predictor of vocal efficiency. *Journal of Voice, 30*(4), 398–406. https://doi.org/10.1016/j.jvoice.2015.06.002

Toles, L. E., & Shembel, A. C. (2024). Acoustic and physiologic correlates of vocal effort in individuals with and without primary muscle tension dysphonia. *American Journal of Speech-Language Pathology, 33*(1), 237–247. https://doi.org/10.1044/2023_AJSLP-23-00159

Van Houtte, E., Van Lierde, K., & Claeys, S. (2011). Pathophysiology and treatment of muscle tension dysphonia: A review of the current knowledge. *Journal of Voice, 25*(2), 202–207. https://doi.org/10.1016/j.jvoice.2009.10.009

Van Stan, J. H., Ortiz, A. J., Cortes, J. P., Marks, K. L., Toles, L. E., Mehta, D. D., Burns, J. A., Hron, T., Stadelman-Cohen, T., Krusemark, C., Muise, J., Fox-Galalis, A. B., Nudelman, C., Zeitels, S., & Hillman, R. E. (2021). Differences in daily voice use measures between female patients with nonphonotraumatic vocal hyperfunction and matched controls. *Journal of Speech, Language, and Hearing Research, 64*(5), 1457–1470. https://doi.org/10.1044/2021_JSLHR-20-00538

Verdolini, K., Rosen, C. A., & Branski, R. C. (Eds.). (2006). *Classification for voice disorders-I*. Erlbaum.

Zheng, Y.-Q., Zhang, B.-R., Su, W.-Y., Gong, J., Yuan, M.-Q., Ding, Y.-L., & Rao, S.-Q. (2012). Laryngeal aerodynamic analysis in assisting with the diagnosis of muscle tension dysphonia. *Journal of Voice, 26*(2), 177–181. https://doi.org/10.1016/j.jvoice.2010.12.001

Appendix

Details for the Exploratory Factor Analysis

Preparation of factors. A Pearson's correlation was run between variables. No correlations between variables exceeding 0.80 were found, indicating the analysis could proceed. To check for a total lack of correlation between variables (i.e., the correlation coefficients = 0), the Bartlett test was conducted. The null hypothesis that all correlations are 0 was rejected, indicating the analysis could proceed. The Kaiser–Meyer–Olkin (KMO) test was conducted, yielding a measure of sampling adequacy (MSA) value of 0.59, which is "miserable" (Kaiser & Rice, 1974), indicating that low-contribution variables should be eliminated. The averaged percent of creak in vowel productions had an MSA value < 0.40, indicating it was a low-contribution variable, and was therefore excluded. Following this, the KMO test was rerun, yielding an MSA value of 0.62, which is "mediocre." This value is above the 0.60 threshold for conducting an analysis, indicating the analysis could proceed.

Determining number of factors. Parallel analysis (Horn, 1965) is recommended over scree test (Cattell, 1966) for determining number of eigenvalue factors due to the highly subjective nature of scree plot visualization (Courtney, 2013). Parallel analysis compares a matrix of eigenvalues from the data to a matrix of random numbers of the same sample size. A parallel analysis plot was built, recommending four factors. Therefore, four factors were used for the initial loading of factors.

Extraction method and rotation. Principal axes factor analysis consists of successive eigen value decompositions until minimal change in the diagonal of the correlation matrix is achieved. Principal axes extraction method was selected over other methods because it had the same or better root-mean-square of the residuals (0.05) and the smallest Bayesian Information Criterion value (–107.28). Proportion of variation of the factors was the same for all methods.

A promax oblique rotation (Hendrickson & White, 1964) was conducted using the pa() function in the "psych" package (Revelle, 2007) in R given the theoretical motivation that at least some factors would be correlated. A promax oblique rotation (as opposed to an orthogonal rotation) is a transformation of an orthogonal rotation that allows for, and then minimizes, correlation between factors, thereby allowing for a simpler factor structure. Furthermore, this rotation does not force correlations if none exist, making this rotation method the most conservative option.