Rearranging Voice Disorders: Refining the New Two-Dimensional Continuous Model

[1]Ofer Amir, Rachel Blais, and Ilan Roziner, *Tel-Aviv, Israel*

Summary: Background. Voice and laryngeal pathologies are traditionally arranged in categories. A recent study has suggested an alternative approach, using two continuous scales, Organicity and Tonicity, which form a two-dimensional plane on which all pathologies/conditions can be arranged.

Objectives. This study was designed to examine the validity and reliability of the new continuous 2D model and to learn how it is affected by experts' background characteristics.

Methods. Ninety-three international experts from 16 countries participated in the study and rated a comprehensive list of 35 laryngeal and voice pathologies/conditions on two continuous scales. On the Organicity scale, 0 represented "nonorganic," and 10 defined "organic." Similarly, on the Tonicity scale, 0 described "hypotonic" and 10 represented "hypertonic."

Results. Cronbach's alphas were high for Organicity and Tonicity (0.99), with varied interexpert agreement scores. Pathologies/conditions populated all four quadrants of the constructed two-dimensional plane, with a majority of 21 of the 35 located in the first quadrant (high-Tonicity, high-Organicity). Results showed strong replicability when compared to the preliminary study. In addition, ratings on the two scales were found consistent and statistically unaffected by the experts' background characteristics.

Conclusions. The new two-dimensional model is valid and reliable, and it provides a simple yet comprehensive approach for arranging voice disorders using a continuous perspective rather than a categorical one. Furthermore, the new model offers a framework that facilitates examination of the differences in how professional experts view different laryngeal pathologies and conditions.

Key Words: Voice disorders–Laryngeal disorders–Pathology–Categories–Continuous scales.

INTRODUCTION

Voice and laryngeal pathologies are traditionally arranged into categories. These categories represent theoretical and clinical knowledge of the different pathologies and conditions. Over the years, several categorical schemes have been proposed. The most fundamental and elementary scheme defines voice pathologies based on their etiology as either "organic" or "functional."[1] However, acknowledging the limitations and rigidity of this binary scheme, later approaches suggested arranging voice disorders into different categories and using more categories and subcategories. Three recent scoping reviews provide a systematic, comprehensive, and enlightening overview of the diversity of the different categorical approaches,[2–4] which were shown to define 2 to 11 different categories. The reviews conclude that no categorical approach provides a holistic and comprehensive system that covers all pathologies and encompasses the full range of voice disorders.

One essential and inherent limitation of all categorical approaches is that each laryngeal pathology or condition is expected to fit into one of the categories, but none should fit more than a single category.[5] Nonetheless, all categorical classification systems in the field of voice disorders present some overlap between categories. This is evident, for example, in the case of vocal nodules, typically viewed as resulting from a combination of organic and nonorganic factors.[1,6–8] Another example of this overlap can be seen when considering "psychogenic aphonia" and "muscle tension dysphonia (MTD)," which are usually pooled together into a "functional" category despite the crucial differences in their origin and clinical symptoms. Hence, while separating voice and laryngeal pathologies into categories might provide a simplified view of the field, it fails to provide a naturalistic overview that comprehends the multifactorial nature of many pathologies and their combined etiologies.[1,2,9]

Considering the theoretical and clinical limitations of the traditional categorical approaches and the multifactorial nature of many voice disorders, a new approach was recently presented for arranging the field of voice disorders/pathologies.[10] This approach arranges all pathologies and conditions using two continuous scales: Organicity and Tonicity. As noted, the "classic" definition of pathologies as either "organic" or "functional" fails to capture the multidimensionality of many laryngeal pathologies and voice disorders. Therefore, the new approach defines an "Organicity" continuum, which ranges between these two extremes ("organic" and "nonorganic") rather than regarding them as two mutually exclusive descriptions. Similarly, categorizing, for example, "functional aphonia" and MTD as "functional" (or "nonorganic") is also clinically erroneous. This is due to the first being a hypo-tension disorder and the latter being a hyper-tension disorder.

Accepted for publication April 8, 2024.

From the Department of Communication Disorders, Faculty of Medical and Health Sciences, Tel-Aviv University, Tel-Aviv, Israel

[1] ORCHID ID: 0000-0002-2873-954X.

Address correspondence and reprint requests to Ofer Amir, Department of Communication Disorders, Faculty of Medical and Health Sciences, Tel-Aviv University, Tel-Aviv, Israel. E-mail: oferamir@tauex.tau.ac.il

Journal of Voice, Vol xx, No xx, pp. xxx–xxx

0892-1997

https://doi.org/10.1016/j.jvoice.2024.04.011

Therefore, the new model defined a second continuum, "Tonicity," which ranges between "hypo-tension" and "hyper-tension." Within this framework, Organicity and Tonicity pertain to etiology rather than compensatory behaviors or secondary characteristics. The two continuous scales are then arranged as the axes of a two-dimensional plane that facilitates the arrangement of all voice/laryngeal pathologies/conditions. The preliminary development of this model was performed by asking a group of 39 international experts to rate a list of pathologies and conditions. Results demonstrated that the new two-dimensional model provides a valid and reliable framework for arranging all pathologies/conditions without categorizing. Accordingly, it was suggested that the new model eradicates the major limitation of the traditional categorical approaches, that is, intergroup overlap, while providing a flexible and dynamic representation of the field.[10]

Due to the novelty of this model, the present study was designed as an advanced step to validate and refine the new two-dimensional model. Specifically, it examined how experts' background characteristics (ie, profession, gender, age, experience, and origin) might affect or bias ratings on the two scales. In addition, because preliminary data suggested that interexpert agreement could vary between pathologies, the present study also aimed to identify specific pathologies/conditions on which experts exhibit weaker agreement. To examine these questions and thus enhance the model validity, the number of participating experts was increased.

METHODS

This study was approved by the Tel Aviv University ethics committee. Then, 123 experts (certified laryngologists or speech-language pathologists) who did not participate in the previous study on the new 2D model were approached and asked to participate. The initial list of experts was prepared from lists of members of international professional laryngology and voice associations [eg, The Voice Foundation, International Association of Communication Science and Disorders (IALP), Collegium Medicorum Thearti (CoMeT), Union of the European Phoniatricians (EUP), and British Voice Association (BVA)]. Of them, 93 (75.6%) accepted the invitation, signed an informed consent and participated in the study. Participants included 56 (60.2%) women and 37 (39.8%) men. All participants were practicing professionals, 51 (54.8%) laryngologists or phoniatricians, and 42 (45.2%) speech-language pathologists specializing in voice disorders, from 16 different countries (Australia, Belgium, Brazil, Finland, France, Germany, India, Israel, Italy, Japan, Mexico, Spain, Turkey, UAE, UK, and USA). Thirty-four (36.6%) of the experts had more than 30 years of experience, 25 (26.9%) had 21–30 years, 28 (30.1%) had 11–20 years, and 6 (6.4%) had 10 years or fewer.

All experts completed an online survey, which was prepared and distributed using the Qualtrics^{XM} platform. A list of 35 laryngeal and voice pathologies or conditions, adopted from the Classification Manual for Voice Disorders,[9] was presented to the experts, who rated them independently on the Organicity and Tonicity scales. No further description or definition of the pathologies/conditions was given to the experts. On the 11-point *Organicity* scale, 0 represented "nonorganic" and 10 represented "organic." Similarly, on the *Tonicity* scale, 0 represented hypotonic, and 10 represented "hyper-tonic." To reduce a possible order effect, the two scales and the different pathologies were presented in a random order that was changed between participants. The experts were instructed to rate the pathologies/conditions based on their etiology but not by their possible outcome or by possible compensatory behaviors patients may show. The survey was conducted in English, and all participants fully completed it.

Statistical analyses

SAS/STAT, ver. 9.4 (SAS Institute) was used for preliminary analyses. Two coefficients were calculated for each scale: Cronbach's α, a measure of internal consistency, and inter-class correlation 1 (ICC1), a measure of inter-rater reliability. Interrater agreement regarding each pathology was estimated using the r_{WG} coefficient.[11] While Cronbach's α coefficient reflects the mean correlation between judges' ratings across different conditions, the r_{WG} quantifies the extent to which they assign the same rating to each pathology. The distribution of each scale and each pathology was described in terms of its mean and standard deviation. A Pearson correlation coefficient was calculated between the two scales.

The present study may be viewed as a follow-up study that elaborates on the preliminary study, which introduced the Organicity and Tonicity scales as a two-dimensional scheme for arranging voice disorders.[10] Therefore, the present data were also compared to the previous study. To that end, a series of independent-sample t tests were conducted, in which the mean Tonicity and Organicity scores obtained in the two studies were compared.

In the next step of the analyses, latent profile analysis (LPA) was used to determine possible classes (clusters) of pathologies/conditions with similar values on the two scales. The LPA was done with Mplus 8.6 software,[12] fitting models with 1 to 5 latent classes. Following recommended criteria for determining the number of classes,[13] we relied on model interpretability, a lower value of log-likelihood statistic, a higher value of entropy index, the smallest value of Bayesian information criteria, and a significant result of bootstrap likelihood ratio test (BLRT). We also preferred models with a substantial proportion of cases in the smallest class based on estimated posterior probabilities. Statistical significance was set at $P < 0.05$.

RESULTS

Reliability and distribution

Internal consistency of the different pathologies/conditions' ratings on the two scales was high. Specifically, Cronbach's

TABLE 1.
Interrater Agreement Coefficients (r_{WG}) and Distribution of Organicity and Tonicity Ratings

	Organicity			Tonicity		
Pathology/Condition	r_{WG}*	Mean	SD	r_{WG}*	Mean	SD
Transgender voice	0.07	3.55	3.04	0.69	6.07	1.75
Nodules	0.13	5.63	2.95	0.73	7.30	1.64
Presbyphonia	0.25	6.90	2.75	0.54	2.92	2.15
Muscle tension dysphonia	0.25	2.65	2.73	0.79	8.24	1.43
Paradoxical motion	0.26	3.15	2.72	0.57	6.47	2.08
Abductor spasmodic dysphonia	0.31	7.77	2.63	0.12	6.24	2.97
Tremor	0.34	7.66	2.57	0.52	5.53	2.19
Reactive lesion	0.37	6.20	2.51	0.80	6.68	1.43
Polyp	0.37	7.69	2.51	0.63	6.30	1.93
Adductor spasmodic dysphonia	0.38	7.94	2.50	0.50	7.95	2.23
Parkinson's disease	0.41	8.48	2.44	0.39	3.28	2.47
Granuloma	0.42	7.63	2.42	0.66	6.76	1.84
Gap	0.43	4.84	2.39	0.60	4.09	2.01
Ventricular phonation	0.43	3.18	2.38	0.57	7.78	2.08
Laryngitis	0.44	7.81	2.36	0.81	5.80	1.36
Reinke's edema	0.45	7.94	2.34	0.66	5.40	1.83
Puberphonia	0.46	1.84	2.32	0.48	6.38	2.28
Reflux	0.47	7.27	2.29	0.84	5.84	1.27
Recurrent nerve paralysis	0.47	9.04	2.29	0.57	2.16	2.08
Hyperemia	0.50	6.84	2.25	0.73	5.96	1.64
Ectasia	0.50	7.73	2.23	0.78	5.75	1.48
Web	0.51	8.77	2.22	0.72	5.83	1.67
Myasthenia gravis	0.51	8.81	2.22	0.63	2.24	1.92
Cyst	0.51	8.27	2.21	0.74	5.93	1.62
Paresis	0.52	8.61	2.20	0.60	2.54	2.00
Superior laryngeal nerve paralysis	0.54	8.99	2.14	0.46	2.93	2.33
Hypertension dysphonia	0.56	2.02	2.10	0.81	8.63	1.36
Scarring	0.59	8.35	2.03	0.72	6.41	1.66
Sulcus	0.63	8.63	1.93	0.65	5.92	1.88
Erythroplasia	0.69	8.48	1.76	0.83	5.59	1.30
Keratosis	0.74	8.90	1.60	0.78	6.09	1.49
Psychogenic aphonia	0.77	0.76	1.53	0.06	4.39	3.07
Leukoplakia	0.83	9.16	1.32	0.83	5.64	1.29
Papilloma	0.85	9.51	1.22	0.74	5.66	1.60
Carcinoma	0.86	9.54	1.19	0.73	5.74	1.63

*r_{WG} coefficient ranges between 0 and 1. Its values are interpreted as follows: 0.00-0.30 = lack of agreement, 0.31-0.50 = weak agreement, 0.51-0.70 = moderate agreement, 0.71-0.90 = strong agreement, 0.91-1.00 = very strong agreement.[14]

α for both Organicity and Tonicity scores were 0.99. Inter-rater reliability was moderate for Organicity, with ICC(1) = 0.59, and close to moderate for Tonicity, with ICC(1) = 0.46. Table 1 summarizes variability in inter-rater agreement, which varied between pathologies, with an average r_{WG} value of 0.48 for Organicity and 0.64 for Tonicity. Inter-rater agreement on the Organicity scale was moderate-to-strong ($r_{WG} \geq 0.50$) for 14 of the 35 pathologies/conditions, and weak for 14. There was a lack of agreement for five of the 35 pathologists rated. In contrast, inter-rater agreement on Tonicity was moderate-to-strong for most pathologies, weak for three, and there was a lack of agreement for two pathologies.

The distribution of the scores obtained for each pathology on the two scales is also presented in Table 1. The different pathologies were spread across nearly the full range of both scales. Hence, all score combinations on the two scales were evident. For example, while adductor spasmodic dysphonia was rated high on Organicity (7.94) and high on Tonicity (7.95), recurrent laryngeal nerve paralysis was characterized as high on Organicity (9.04) but low on Tonicity (2.16). Substantial differences were observed in the *Standard Deviations* of both scales, with smaller *SD*s corresponding to higher levels of interrater agreement. The correlation between the scores obtained on the two scales was moderate and negative, $r = -0.40$, 95% C.I. = [− 0.644, − 0.072], demonstrating their divergent validity.

Figure 1 illustrates the distribution of pathologies/conditions on the two scales arranged on a two-dimensional plane. The ratings of the pathologies were scattered across the Organicity scale and, to a slightly lesser degree - for the

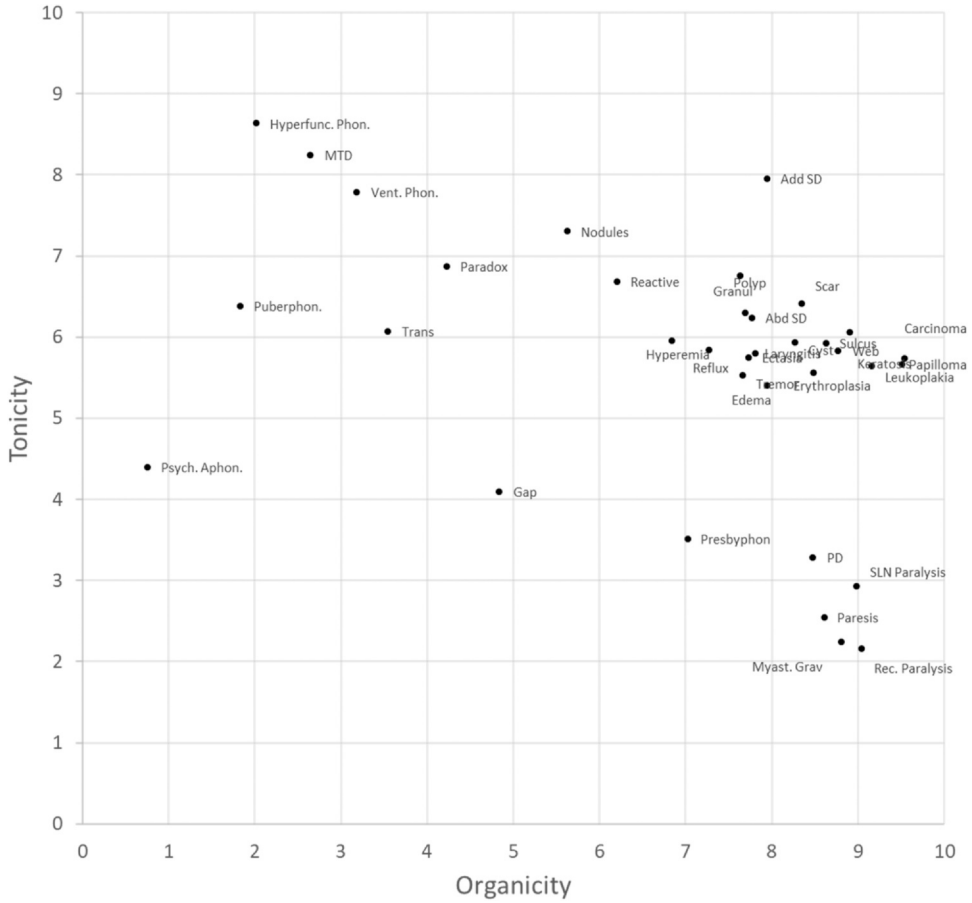

FIGURE 1. Distribution of the voice pathologies on the Organicity and Tonicity continuous scales arranged on a two-dimensional plane. Abd SD, abductor spasmodic dysphonia; Add SD, adductor spasmodic dysphonia; Gap, glottal gap; Hyperfunc Phon, hyperfunction phonation; LPR, laryngo-pharyngeal reflux; MTD, muscle tension dysphonia; Paradox., paradoxical vocal folds' movement; PD, Parkinson's disease; Psych. Aphon, psychogenic aphonia; Rec Paralysis, recurrent laryngeal nerve paralysis; SLN paralysis, superior laryngeal nerve paralysis; Tremor, vocal tremor; Vent. Phon., ventricular phonation; Web, glottal web.

Tonicity scale. Yet, Figure 1 also demonstrates an unbalanced distribution of the pathologies. Specifically, 21 of the 35 pathologies/conditions were concentrated on the first quadrant of the 2D plane (ie, high Organicity, high Tonicity). In comparison, only two pathologies populated the third quadrant (ie, low Organicity, low Tonicity).

Stability and replicability

The data from this study were also compared to the previous study, using a series of independent-sample *t* tests. Of the 70 comparisons between the two studies (35 pathologies × 2 scales), only four (Organicity: LPR and Paresis; Tonicity: Trans Voice and Keratosis) were statistically significant ($0.008 < P < 0.045$). This result is expected, given the set error rate of 5% (ie, 5% of 70 = 4). In addition, the correlation between mean scores obtained for the 35 pathologies in the two studies was $r = 0.93$ for Organicity and $r = 0.98$ for Tonicity (both P's < 0.001). These findings demonstrate high stability and replicability of the Organicity and Tonicity scales.

Arranging pathologies/conditions into clusters

Data show that the 35 voice pathologies/conditions are scattered continuously across all areas of the two-dimensional plane rather than arranged in categories. Nevertheless, Figure 1 suggests that the pathologies could also be viewed as arranged in four clusters, roughly matching the quadrants of the two-dimensional plane. A LPA was performed to examine this view. Results for five alternative models (1- to 5-cluster solutions) are summarized in Table 2.

Using the criteria of low log-likelihood and BIC values, results suggest that either the 3- or 4-cluster solutions should be preferred. The ***P*** BLRT values indicate that the 2-, 3-, and 4-cluster solutions significantly or near-significantly differ from solutions with fewer clusters. The entropy criterion had satisfactory values in the 2-, 3-, 4- and 5-cluster solutions. However, the proportion of the smallest class in the 5-cluster solution was 3%, that is, included only one pathology/condition. Therefore, in light of these statistical analyses, jointly with the fact that the four-cluster model is in agreement with the visual presentation

TABLE 2.
Comparison of the Different Latent Profile Models

# of classes	Log-likelihood	BIC	*P* BLRT	Entropy	Proportion of the smallest class
1	− 147.33	308.87	NA	NA	NA
2	− 132.01	288.90	< 0.001	0.95	0.23
3	− 126.38	288.31	0.069	0.96	0.17
4	− 120.02	286.26	0.066	0.98	0.06
5	− 116.02	289.72	0.375	0.99	0.03

TABLE 3.
Arrangement of the Pathologies/Conditions into Four Clusters

		Organicity		Tonicity	
Cluster	Pathologies in cluster	Mean	Mean r_{WG}	Mean	Mean r_{WG}
I	Abductor Spasmodic Dysphonia; Adductor Spasmodic Dysphonia; Carcinoma; Cyst; Ectasia; Edema; Erythroplasia; Granuloma; Hyperemia; Keratosis; Laryngitis; Leukoplakia; Laryngo-Pharyngeal Reflux; Papilloma; Polyp; Reactive Lesion; Scarring; Sulcus; Vocal Tremor; Glottal Web	8.06	0.54	6.07	0.69
II	Hyperfunction Phonation; Muscle Tension Dysphonia; Paradoxical Vocal Fold Movement; Puberphonia; Transgender Voice Ventricular Phonation; Vocal Nodules	2.82	0.31	4.30	0.66
III	Glottal Gap; Psychogenic Aphonia	2.99	0.60	7.29	0.33
IV	Myasthenia Gravis; Vocal Folds' Paresis; Parkinson's Disease; Presbyphonia; Recurrent Laryngeal Nerve Paralysis; Superior Laryngeal Nerve Paralysis	8.46	0.44	2.70	0.53

of the data (Figure 1), and as this solution agrees with the one proposed previously,[10] the 4-cluster solution was selected.

The arrangement of the pathologies/conditions into the four clusters, based on their Organicity and Tonicity scores, is presented in Table 3. As shown, cluster I (high Organicity, high Tonicity) consists of 20 pathologies; cluster II (low Organicity, low Tonicity) consists of seven pathologies; cluster III (low Organicity, high Tonicity) consists of two pathologies; and cluster IV (high Organicity, low Tonicity) of six. A comparison of the current data with the findings of Amir et al[10] reveals strong similarities in the cluster composition. Specifically, the correlation between cluster membership in the two studies, calculated over the 35 pathologies/conditions, yielded a Cramer's V value of 0.85 ($P < 0.0001$). Only minor adjustments were made based on the current findings. These included a single pathology from the original cluster I (ie, vocal nodules), which shifted to cluster II, and two other pathologies (puberphonia and transgender voice) that migrated from cluster III to cluster II.

Background characteristics

Finally, data were examined for the possible effect of the experts' background characteristics on the ratings obtained on the two scales. Specifically, differences were examined between genders, age groups, professions, years of experience, and continent of residence. Table 4 summarizes mean values and standard deviations of the mean ratings on the two scales, arranged by the different background characteristics, alongside the analyses of the statistical results.

Data show no significant differences for any background characteristic, demonstrating that ratings on the two scales are not affected by the experts' gender, age, profession, experience, or place of residency.

DISCUSSION

This study was carried out to examine the new two-dimensional continuous model for arranging voice disorders.[10] This model represents a new approach, inherently different from the available traditional categorical approaches.[2–4] While all categorical models attempt to arrange voice/laryngeal pathologies into a set number of groups (categories), the new model moves away from this approach and arranges pathologies using two continuous scales (Organicity and Tonicity), which form a two-dimensional plane on which all pathologies are placed, without the need for defining categories. This approach eradicates overlaps between categories and demonstrates that voice pathologies can be arranged on a continuum. Because the initial presentation of the new model[10] was based on ratings of the different pathologies performed by 39 experts, it was impossible to examine potential biases

TABLE 4.
Mean Values and Standard Deviations (in Parentheses) of the Ratings on the Two Scales, Arranged by Background Characteristics, and Summary of the Statistical Analyses

Background	Value	N (%)	Organicity		Tonicity	
Characteristic			Mean (SD)	Statistic	Mean (SD)	Statistic
Gender	Female	57 (61.3)	8.30 (1.18)	$t(91) = 2.01$,	5.76 (0.91)	$t(55.6) = 1.54$,
	Male	36 (38.7)	7.79 (1.17)	$P = 0.057$	5.52 (0.62)	$P = 0.162$
Age group	≥50	26 (29.2)	8.04 (1.14)		5.52 (0.48)	
(years)	51-60	26 (29.2)	8.11 (1.16)	$F(2,86) = 0.09$,	5.49 (0.60)	$F(2,86) = 1.76$,
	61≤	37 (41.6)	7.98 (1.24)	$P = 0.910$	5.81 (0.97)	$P = 0.179$
Profession	SLP	43 (46.2)	7.78 (1.29)	$t(91) = 1.57$,	5.74 (0.79)	$t(90) = 1.83$,
	MD	50 (53.8)	8.17 (1.09)	$P = 0.119$	5.46 (0.68)	$P = 0.070$
Experience	≤20	34 (37.4)	8.14 (1.27)		5.59 (0.66)	
(years)	21-29	23 (25.3)	7.89 (1.07)	$F(2,88) = 0.31$,	5.57 (0.50)	$F(2,88) = 0.17$,
	30≥	34 (37.4)	8.01 (1.16)	$P = 0.732$	5.67 (0.98)	$P = 0.843$
Continent	Europe	35 (37.6)	8.15 (1.06)		5.77 (0.87)	
	Americas	23 (24.7)	8.27 (1.33)		5.69 (0.83)	
	Asia	33 (35.5)	7.53 (1.10)	$F(3,89) = 3.62$,	5.41 (0.49)	$F(3,88) = 1.47$,
	Australia	4 (2.2)	9.49 (0.73)	$P = 0.016$*	5.31 (0.65)	$P = 0.229$

*Following the significant effect of continent obtained for Organicity, follow-up specific comparisons were performed with Tukey's correction. None of the pairwise comparisons between continents reached a significance of $P < 0.05$.

related to the experts' background characteristics. Therefore, the present study was set to re-examine the model using a large group of 93 international experts.

Reliability and validity

Results confirmed that the ratings of the different pathologies/conditions were highly reliable. Moreover, as shown in Figure 1, the pathologies are spread across the full range of the Organicity scale and slightly less on the Tonicity scale. While this was not a primary aim of this study, we found that the nominal values obtained for each pathology on the two scales were highly consistent with those reported in the preliminary research. This stability was evident by the lack of significant differences between the studies and the strong correlations between the values obtained for each pathology ($r = 0.93$ and 0.98 for Organicity and Tonicity, respectively).

As noted, this study also examined whether experts' background characteristics could bias responses on the two scales. As the present study included a large group of 93 professional experts (compared to 39 in the previous study), these comparisons could be conducted. To that end, comparisons were made within our cohort between the responses of men versus women, between age groups, the two professions (SLP and MD), years of professional experience, and continent of residence (see Table 4). Results demonstrated that the experts' responses on the two scales were consistent and impervious to all background characteristics. This is taken as additional support for the validity and stability of the two-dimensional model. Therefore, the combined results described here provide empirical evidence for the validity and reliability of the new two-dimensional continuous model.

Identifying clusters within the continuous model

Inspection of Figure 1 primarily demonstrates that the different pathologies/conditions are indeed arranged continuously rather than categorically. Yet, data also show that the full spectrum of all voice/laryngeal pathologies may be viewed as arranged in four clusters. It should be clarified that these clusters do not necessarily represent common etiologies. Thus, they are – by no means – categories. Instead, these clusters indicate that different pathologies, which are closely located on the two-dimensional plain, are similar in their level of "Tonicity" and "Organicity." Hence, they are rated similarly on both scales. It is evident, for example, that cluster I (ie, high Tonicity, high Organicity) comprises the largest number of pathologies. This is expected considering laryngeal anatomy and physiology. As most intrinsic laryngeal muscles are adductors, it is not surprising that most laryngeal pathologies are characterized by high tonicity, whether the pathology is mainly organic or not (ie, high or low Organicity). Similarly, it is unsurprising that the pathologies in cluster IV (high Organicity, low Tonicity) might be considered "neurogenic" using traditional categorical taxonomies. Therefore, it is imperative to note that the new model does not arrange pathologies based on their etiology but clinical characteristics. This point should be further explained, as patients with recurrent laryngeal nerve paralysis, for example, could show significant laryngeal hypertension as compensatory behaviors or secondary characteristics. Therefore, it should be clarified that the ratings on the model represent the pathologies' "stereotypical" characteristics but not further clinical manifestations.

Another example of the clinical merit of using clusters that emerge from this continuous model is the difference

between "psychogenic aphonia" and "hyper-functional phonation." While most categorical schemes place both conditions in the "functional"[1,9] or "malregulative"[15] category, the new model clearly shows that the two pathologies/conditions are different by nature. Specifically, while one is characterized by high Tonicity, the other is characterized by low Tonicity. Therefore, it is apparent that the new model provides a holistic overview of the field of voice disorders and encompasses all pathologies.

Interexpert (dis)agreement: opening "Pandora's box"

As this study included a large group of 93 international experts, it has provided an intriguing opportunity to examine differences in the way experts view and describe specific laryngeal and voice pathologies/conditions. As shown in Table 1, five laryngeal conditions received r_{WG} scores on the Organicity scale that were lower than 0.30, representing a lack of agreement between the experts, and 14 conditions received r_{WG} scores between 0.30–0.50, representing low agreement. This demonstrates substantial differences in how experts view the etiology of these conditions. For example, the common clinical diagnosis of "vocal nodules" received a low r_{WG} score of 0.13. Inspection of the raw data that yielded this value reveals that 23 experts (24.7%) rated nodules as low on the Organicity scale, with values of 0–3. In contrast, 37 experts (39.8%) rated it as high on Organicity, with values of 7–10. The remaining 33 experts (35.5%) rated nodules as having a balanced contribution of organic and nonorganic etiologies. This means that even in this case of a seemingly simple diagnosis, international experts do not necessarily agree on its underlying etiology. Another example of this professional disagreement can be shown in the case of "Presbyphonia." Data show that while a majority of 60 experts (64.5%) rated it high on Organicity, 21 (22.5%) considered it as having mixed etiologies, and 12 (12.9%) rated it low on Organicity, that is, viewed it as a mainly nonorganic (ie, "functional") condition.

These differences in professional experts' opinions are of great interest, theoretically and clinically. They represent diversity in professional beliefs and understanding of the various laryngeal pathologies and conditions. These conceptual differences shape professional clinical decision-making, which could lead to different treatment recommendations. Such differences cannot be attributed to differences in the background characteristics of the participating experts, as shown above. Instead, this could suggest that despite years of accumulating research and clinical work in the field, there is still a need to define better the nature and the etiology of many laryngeal pathologies and voice conditions and to reach a broader and international agreement. While this might be viewed as "opening Pandora's box," we suggest that further examination of these professional differences of view would enhance the exchange and dissemination of knowledge, reduce miscommunication, and eventually advance our theoretical understanding and clinical practice.

CONCLUSION

This study examined the new continuous two-dimensional model for arranging voice disorders. Compared to the preliminary research, it nearly tripled the number of participating experts, which confirmed and enhanced the validity and reliability of the new model. Mean ratings of the different pathologies/conditions on both continuous scales were consistent, irrespective of all experts' background characteristics. Data show that while experts strongly agree on the clinical description of many pathologies, they present divergent descriptions of other pathologies, as manifested in interexpert differences in ratings on the scales. These findings warrant further research and discussion to resolve these professional inconsistencies.

Limitations and future directions

While the new two-dimensional model is valid and highly reliable, it does not intend to eradicate the use of the existing categorical terminology. Instead, it provides a holistic overview of the field, demonstrating that pathologies and conditions defined differently or assigned to different categories could be either "close" or "distant" on the continuous two-dimensional plane.

It should be noted that the comprehensive list of pathologies/conditions used in our study was adopted from a previous study.[9] Hence, this list includes voice and laryngeal pathologies, but also specific terms that might not be described as pathologies but as "clinical descriptions" (eg, "gap" or "erythroplasia"). Whether or not these diagnoses should be included in the model, as it further evolves, should be discussed and examined in the future. Furthermore, the term "transgender voice" is also included in this list for the same reason. However, we acknowledge that it might not be relevant to this model as a voice disorder per se. Therefore, its inclusion in future developments of this model should be reconsidered.

Finally, after establishing the consistency and repeatability of the model in this study, it is now possible to examine the associations between ratings of the pathologies on the model and various domains of voice. For example, future research could examine how the magnitude of the differences between pathologies in acoustic or perceptual measures are associated with the Euclidian distances between the pathologies on the two-dimensional plane. Moreover, the fact that many pathologies are concentrated on one quadrant (high Organicity, high Tonicity) might indicate that adding a third dimension to the model could provide further insight. Finally, as this model offers a theoretical framework for arranging pathologies in the field of voice disorders, other continuous scales could be defined in the future, providing an even more accurate representation of the field. These possibilities should be examined in future research.

Declaration of Competing Interest

None.

Acknowledgments

The first author wishes to thank our esteemed colleagues and professional friends from around the world, who participated as experts in this study and enabled the development of the new model.

References

1. Aronson A. *Clinical Voice Disorders.* New York, NY: Thieme Stratton; 1985.
2. A. Constantini, V. Riberio and M. Behlau. Voice disorders classifications: a scoping review - part A. *J Voice*.
3. A. Constantini, V. Riberio and M. Behlau. Voice disorders classifications: a scoping review - part B. *J Voice*.
4. C. Payten, G. Chiapello, K. Weir, et al. Frameworks, terminology and definitions used for the classification of voice disorders: a scoping review. *J Voice*.
5. Agresti A. *An Introduction to Categorical Data Analysis*. 2nd Ed. Hoboken, New Jersey: Wiley-Interscience; 2007.
6. Boone D, McFarlane S, Von Berg S, et al. *The Voice and Voice Therapy.* Upper Saddle River, NJ: Pearson Education; 2005.
7. Pannbacker M. Classification systems of voice disorders: a review of the literature. *Lang Speech Hear Ser Schools.* 1984;15:169–174.
8. Stemple J. *Voice Therapy: Clinical Studies*. 2nd ed San Diego, CA: Singular; 2000.
9. Verdolini K, Rosen C, Branski R. *Classification Manual for Voice Disorders-I.* Mahwah, NJ: Psychology Press; 2014.
10. Amir O, Gutman I, Baken R, et al. A two-dimensional scheme for arranging voice disorders. *JAMA Otolaryngol Head Neck Surg.* 2022;148:139–144.
11. James L, Demaree R, Wolf G. Estimating within-group interrater reliability with and without response bias. *J Appl Psychol.* 1984;69:85.
12. Muthén L, Muthén B. *Mplus User's Guide*. 8th ed Los Angeles, CA: Muthén & Muthén; 2017.
13. Nylund K, Asparouhov T, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Equ Model Multidiscip J.* 2007;14:535–569.
14. LeBreton J, Senter J. Answers to 20 questions about interrater reliability and interrater agreement. *Organ Res Methods.* 2008;11:815–852.
15. Hacki T, Moerman M, Rubin J. 'Malregulative' rather than 'Functional' dysphonia: a new etiological terminology framework for phonation disorders—a position paper by the Union of European Phoniatricians (UEP). *J Voice.* 2022;36:50–53.