



Predictive Value of Auditory Perceptual Assessment of Voice in Diagnosis of Functional Voice Disorders and Minimal Associated Pathological Lesions

Marwa S.M. Saleh, Emad K. Abdel-Haleem, Eman S. Hassan, Reham A. Ibrahim

Phoniatrics Unit, ENT Department, Assiut University, Assiut, Egypt

Abstract:

Background: Auditory perceptual assessment is considered the preliminary assessment tool of disordered voice quality. The validity of perceptual measurement depends on characteristics of the raters and the stimuli. It is associated with minimal cost and no specialized instrumentation is required. Because auditory-perceptual evaluation of dysphonia has been heavily criticized on the basis of its reliability and validity, we conducted this study to assess the reliability of auditory perceptual assessment as one of the objective measures for assessment of voice disorders and to determine the predictive relationships between instrumental measures as endoscopic examination and perceptual ratings of impaired voice quality.

Subjects and Methods: The study included 50 patients complaining of dysphonia with age ranges from 5 -61 years. Auditory perceptual assessment (APA) of patients' voices was done and voice recordings were assessed by three experienced phoniatricians in a double blind, randomized fashion. The examiners (Ex1.2.3) assessed the recorded voice quality by using the modified GRBAS scale. Laryngoscopic examination was performed using indirect laryngoscope, flexible fibrotic laryngoscopy as well as rigid laryngoscope.

Results: Results showed considerable similarities among raters in assessing the disordered voice quality especially on assessing the overall grade of dysphonia and the "Strain" voice quality with strong inter-rater reliability (Cronbach's Alpha) was 0.780 and 0.710 respectively. However, data showed some differences with poor inter-rater reliability especially for ratings of glottal attack and associated laryngeal function as cough (Cronbach's Alpha) was 0.383 and 0.531 respectively. The study also showed significant relations on comparing between instrumental (video endoscopic) diagnosis and perceptual predictive diagnosis of patients' disordered voice quality according to two of the raters with $P < 0.03$ & 0.04 .

Conclusion: This study showed considerable variability of the inter-rater reliability across parameters of disordered voice quality with high inter-rater reliability for some parameters and low for others. Yet, there was a strong association between instrumental diagnostic measures and the perceptual predictive diagnosis denoting a clear relationship between listener's judgments of disordered voice quality and the underlying vocal pathology; a thing that is in favor for the validity of APA as a popular and fundamental tool for the assessment of vocal impairment.

Key words: Auditory perceptual assessment, Functional voice disorders, minimal associated pathological lesions

Introduction

Auditory perceptual assessment is considered the preliminary assessment tool of disordered voice quality. In fact, it is considered the most popular

method of examination in clinical practice. Even among highly sophisticated voice clinics that have access to advanced vocal imaging instrumentation auditory perceptual

assessment is still the gold standard for investigation and documentation of disordered voice quality.

Because voice is perceptual in nature, perceptual features of voice quality are also likely to have greater shared reality among a wide range of listeners including clinicians, clients, employers and other associates of those clients.¹ As a consequence of this shared reality perceptual description of disordered voice quality is easier to be communicated to patients or other health professionals than many instrumental measures of voice quality.

Objective measures of voice quality are susceptible to multiple confounding influences that can be difficult to control in a clinical environment. Such confounding factors include environmental recording conditions, the specific characteristics of hardware and software systems, the testing and analysis protocols used individual variability in acoustic and aerodynamic parameters, and the degree of severity of the dysphonia.²

The validity of perceptual measurement depends also on characteristics of the raters and the stimuli. It is associated with minimal cost. Apart from a good quality microphone and audio-recording device, no specialized instrumentation is required.³

Perceptual evaluation is also quick, unobtrusive and comfortable for the client.⁴ The auditory-perceptual evaluations are the GRBAS (Grade, Roughness, Breathiness, Asthenia, and Strain) scale⁵ and modified GRBAS scale which was described by Kotby et al., 1989⁶: it consists of: -Overall grade of dysphonia either normal (0), slight (1), moderate (2) or severe (3). Character (quality): strained (S), leaky (L), breathy (B) or irregular "rough"(I). Pitch: overall increase, decrease or diplophonia. Register:

which either habitual register (modal or falsetto), tendency of vocal fry at the end of phrase, register break. Loudness: This may be excessively loud, excessively soft or fluctuating. Glottal attack: This may be normal, soft or hard. Associated laryngeal functions: include cough, whisper and laughter which may be affected. The rating is made by assessing current conversational speech or when reading a passage. Because auditory-perceptual evaluation of dysphonia has been heavily criticized on the basis of its reliability, instrumental measures have sometimes been promoted as a better alternative. However, despite decades of intensive research, an ideal set of objective instrumental measures of disordered voice quality remains elusive.⁷ Nevertheless, there are a lot of factors that affect the reliability of auditory perceptual assessment such as listeners' background training, the type of voice sample, environmental recording conditions, the specific characters of hardware and software systems and the listening conditions. Such confounding factors combined auditory perceptual assessment to its inherent subjectivity. However, this subjectivity is not sufficient to reject this approach of assessment. These factors could be overcome by determining perceptual inter-rater reliability and correlating auditory perceptual assessment of disordered voice quality to the underlying vocal fold pathology. Acknowledging these limitations in the validity and reliability of instrumental measures of voice quality; we conducted this study to assess the reliability of auditory perceptual assessment as one of the objective measures for assessment of voice disorders and to determine the predictive relationships between instrumental measures as endoscopic examination and perceptual ratings of impaired voice quality.

Patients and Methods:

Subjects:

The study included 50 patients complaining of dysphonia with age ranges from 5-61 years and a mean age of 32.0 ± 15.69 . Patients were selected from the outpatient clinic of Phoniatics Unit, Assuit University Hospital from October 2016 to October 2017.

Method:

Patients were given a clear explanation of the study objectives and the plan of the study. Informed consent was obtained from patients to participate in the study. All patients completed the following assessment procedures:

1. Elementary diagnostic procedures included:

- Personal history, complaint, analysis of symptoms, full ENT, head and neck examination.
- **Auditory perceptual assessment (APA) of patients' voice:** voice recording by using a sensitive microphone (Karaokeaiwa DM-H19) connected to the computer in sound treated room with an optimal distance was obtained. The patients followed a protocol in voice recording information: ⁸ - Reading a standardized Arabic text in Arabic summary followed by automatic speech e.g. counting from 1 to 10 and uttered a sustained vowel /a/. Voice recordings were assessed by three experienced phoniaticians in a double blind, randomized fashion. Audio recordings of subjects' voices were assigned to each examiner randomly. Each examiner completed separately the auditory perceptual assessment. To ensure blindness, none of the examiners or the data collector has heard the audio recordings before the assessment. The examiners (Ex1.2.3) assessed the recorded

voice quality by using the modified GRBAS scale.

Modified GRBAS scale ⁶: which consisted of:

- Overall grade of dysphonia either normal (0), slight (1), moderate (2) or severe (3).
- Character (quality): either strained (S), leaky (L), breathy (B) or irregular "rough"(R) and each character was classified in terms of severity into (I), (II), (III) where (I) was mild, (II) was moderate and (III) was severe.
- Pitch: overall increase, decrease or diplophonia.
- Register: Either habitual register (modal or falsetto), tendency of vocal fry at the end of phrase or register break.
- Loudness: excessively loud, excessively soft or fluctuating.
- Glottal attack: normal, soft or hard.
- Associated laryngeal functions: included cough, whisper and laughter.

2. Clinical diagnostic aids included:

- Laryngoscopic examination was performed using indirect laryngoscope, flexible fibrotic laryngoscopy, rigid 90-degree laryngoscope (R. wolf 445057, and explorent 620090) and video Rhinolaryngoscope (KARL STORZ-ENDOSKOPE model TP100).

Statistical analysis:

Date entry and data analysis were done using SPSS version 23 (Statistical Package for Social Science). Data were presented as number, percentage, mean, standard deviation. Chi-square test and Fisher Exact test were used to compare between qualitative variables. Independent sample t-test was used to compare quantitative variables

between groups. P-value considered statistically significant when $P < 0.05$. For Reliability statistics using Cronbach's Alpha: the resulting of a coefficient of reliability range from 0 to 1.

Cronbach's Alpha	Inter-rater reliability
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

Results:

The study included 50 patients complaining of dysphonia with age ranges from 5 -61 years and a mean age of 32.0 ± 15.69 . Of these 50 patients; 25 patients were females and 25 were males. 28 (56%) of the patients had positive history of voice abuse.

1. Perceptual ratings of the different parameters of vocal impairment

Table (1) shows perceptual ratings for severity of vocal impairment parameters. It's clear, however that there were considerable differences among raters. For example, when assessing the "Strain" quality; one rater classified 18 (36%) of patients as having no strain at all, 2 (4%) of patients as having only mild strain, 21 (42%) of patients with moderate strain and 9 (18%) as having severe strain. On the other hand, another rater classified 6 (12%) of patients as having no strain, 16 (32%) as having mild strain, 22 (44%) had moderate strain and 6(2%) of patients with severe strain. The third rater classified 2 (4%) of patients as having no strain, 30 (60%) with mild strain, 11 (22%) with moderate and only 7(14%) of patients with severe strain. Yet, the results also showed considerable similarities among raters especially on assessing "Breathiness" quality as one rater classified 44 (88%) of patients as having

no breathiness, only 1 (2%) of patients having mild breathiness, 4(8%) of patients as having moderate and only 1(2%) of patients as having severe breathiness. Also, another rater came up with a close classification; categorizing all patients as having no breathiness. The last rater also classified 48(96%) of patients as having no breathiness, only 1 (2%) of patients as having mild breathiness, 1 (2%) of patients as having moderate and none of the patients had severe breathiness.

2. Overall Voice ratings:

Other similarities and differences among raters can be detected in their judgments of overall voice ratings (table 2).

3. Estimates of inter-rater reliability

On calculating the inter-rater reliability (Cronbach's Alpha) ; estimates of the overall grade of dysphonia and the "Strain" voice quality were the highest 0.780 and 0.710 respectively, which meant that there was an acceptable reliability between the three examiners, whereas, other parameters were ranging between unacceptable and poor inter rater reliability (Table 3).

4. Raters' perceptual predicted diagnosis of patients' disordered voice quality

According to table (4); there were some differences among raters in perceptual predicted diagnosis and estimates of inter-rater reliability were unacceptable (Cronbach's Alpha: 0.279).

5. Comparison between instrumental (video endoscopic) diagnosis and perceptual predictive diagnosis of patients' disordered voice quality for each rater;

a non-significant relation between both assessments was found according to rater 1 ($P=0.264$). However, there was a significant relation between the two assessments according to rater 2 and rater 3 with $P < 0.03$ & 0.04 respectively (table 6).

Table (1): The number (percentage) of subjects according to perceptual ratings of vocal impairment parameters:

Vocal Impairment parameter	Severity	Rater 1 “n=50”	Rater 2 “n=50”	Rater 3 “n=50”
Overall grade of dysphonia	• None	2(4%)	1(2%)	0.0
	• Mild	2(4%)	12(24%)	11(22%)
	• Moderate	23(46%)	30(60%)	24(48%)
	• Sever	23(46%)	7(14%)	15(30%)
Strained	• None	18(36%)	6(12%)	2(4%)
	• Mild	2(4%)	16(32%)	30(60%)
	• Moderate	21(42%)	22(44%)	11(22%)
	• Sever	9(18%)	6(12%)	7(14%)
Leaky	• None	29(58%)	17(34%)	20(40%)
	• Mild	2(4%)	19(38%)	12(24%)
	• Moderate	12(24%)	14(28%)	14(28%)
	• Sever	7(14%)	0.0	4(8%)
Breathiness	• None	44(88%)	50 (100%)	48(96%)
	• Mild	1(2%)	0	1(2%)
	• Moderate	4(8%)	0	1(2%)
	• Sever	1(2%)	0	0.0
Roughness	• None	27(54%)	26(52%)	20(40%)
	• Mild	1(2%)	14(28%)	18(36%)
	• Moderate	15(30%)	10(20%)	8(16%)
	• Sever	7(14%)	0.0	4(8%)

Table (2): Overall voice ratings:

Vocal impairment parameter	Overall voice rating	Rater 1	Rater 2	Rater 3
Vocal pitch	• Overall increase	18(36%)	7(14%)	20(40%)
	• Decrease	8(16%)	17(34%)	12(24%)
	• Diplophonia	0.0	0.0	7(14%)
	• Normal	24(48%)	26(52%)	11(22%)
Vocal loudness	• Excessively loud	1(2%)	7(14%)	0.0
	• Soft	26(52%)	20(40%)	18(36%)
	• Fluctuation	6(12%)	3(6%)	23(46%)
	• None	17(34%)	20(40%)	9(18%)
Associated laryngeal function “Cough”	• Normal	13(26%)	38(76%)	3(6%)
	• Soft	18(36%)	6(12%)	1(2%)
	• Hard	19(38%)	6(12%)	46(92%)
Associated laryngeal function “Cough”	• Affected	43(86%)	25(50%)	45(90%)
	• Preserved	7(14%)	25(50%)	5(10%)
Total		50	50	50

Table (3) Estimates of inter-rater reliability:

Parameter	Cronbach's Alpha	Inter- rater Reliability
Overall grade of dysphonia	0.780	Acceptable
Character (quality):		
Strain	0.710	Acceptable
Leaky	0.656	Questionable
Breathy	0.380	Unacceptable
Irregular	0.678	Questionable
Pitch	0.376	Unacceptable
Loudness	0.441	Unacceptable
Glottal attack	- 0.383	No reliability
Cough	0.531	Poor

Table (4): Raters' perceptual predicted diagnosis of disordered voice quality:

Item	Descriptive "n=50"
1-Rater 1:	
• Functional	22(44.0%)
• MAPLs	26(52.0%)
• Normal	2(4.0%)
2-Rater 2:	
• Functional	28(56.0%)
• MAPLs	22(44.0%)
• Normal	0.0
3-Rater 3:	
• Functional	20(40.0%)
• MAPLs	30(60.0%)
• Normal	0.0

Table (5): Instrumental (Video endoscopic) assessment of the patients' disordered voice quality:

Item	Descriptive "n=50"
• Functional	22(44.0%)
• MAPLs	28(56.0%)

Table (6): Comparison between instrumental (video endoscopic) diagnosis and perceptual predictive diagnosis of patients' disordered voice quality

Raters' perceptual predictive diagnosis	Video endoscopic diagnosis "n=50"		
	MAPLs "n=22"	Functional "n=28"	P-value
Rater 1:			
• Functional	9(40.9%)	13(46.4%)	P=0. 264 n.s
• MAPLs	11(50.0%)	15(53.6%)	
• Normal	2(9.1%)	0.0	
Rater 2:			
• Functional	16(72.7%)	12(42.9%)	P<0.03*
• MAPLs	6(27.3%)	16(57.0%)	
• Normal	0.0	0.0	
Rater 3:			
• Functional	12(54.5%)	8(28.6%)	P<0.04*
• MAPLs	10(45.5%)	20(71.4%)	
• Normal	0.0	0.0	

Chi-square Test

Discussion :

Clinical voice evaluation starts with a case history, interview and then proceeds to perceptual and instrumental assessments.⁹ Auditory-perceptual evaluation is a primary part of the routine clinical voice assessment because of the ease and competency of the method.¹⁰⁻¹¹

Auditory perceptual assessment has become the most widely used tool of assessment of disordered voice quality probably because of its minimal cost, limited time required for the voice evaluation, patient comfort and minimal requirements for technical knowledge and skills. Moreover, it doesn't require too much sophisticated instrumentations and it's suitable for everyday clinical practice.

The most commonly used perceptual evaluation systems have many similarities in terms of the voice features evaluated and definitions of those features. The GRBAS (Grade, Roughness, Breathiness, Asthenia and Strain)⁸, CAPE-V (Consensus Auditory-Perceptual Evaluation of Voice)¹², (Stockholm Voice Evaluation Approach)¹³ and the Perceptual Voice Profile,¹⁴ for example, all incorporate the perceptual features of breathy, rough and strained.¹⁵ And modified GRBAS scale which was described by **Kotby1986**:⁶ it consists of: -Overall grade of dysphonia. Character (quality): strained (S), leaky (L), breathy (B) or irregular "rough"(I), pitch, register, loudness, glottal attack and associated laryngeal functions: include cough, whisper and laughter which may be affected.

In our study, we used the modified GRBAS scale for perceptual assessment of patients with disordered voice quality; subjects were evaluated by three examiners in a double blinded randomized fashion. Inter-rater

reliability (Cronbach's Alpha) for each parameter of disordered voice quality was calculated. In addition, calculating the estimates of inter-rater reliability for associated laryngeal functions as cough was done.

Estimates of inter-rater reliability were strongest for perceptual ratings of the overall grade of dysphonia (Cronbach's Alpha =0.78) and strain quality of disordered voice (Cronbach's Alpha = 0.71). Our results supported the findings of a study by **Chen et al., 2018**¹⁶ who used the GRBAS scale for assessment of vocal impairment and showed high reliability for the overall grading of dysphonia and strain quality of disordered voice. A fact confirms the reliability of the GRBAS scale of perceptual voice evaluation.

While our results showed lowest reliability for ratings of breathiness (Cronbach's Alpha = 0.38), Chen et al., 2018¹⁶ showed high reliability for ratings of this particular parameter, this could be explained by variation of listeners' agreement when they are asked to isolate specific perceptual features in voices with complex mix of perceptual characters especially with mildly or moderately impaired voice quality. Similarly, our results showed fair inter-rater reliability for ratings of leaky and roughness qualities (Cronbach's Alpha = 0.65 and 0.67) respectively. This came in agreement with those of **Chen et al., 2018**¹⁶ and **Webb, 2004**³ who also showed the same results for ratings of these voice parameters.

Our inter-rater reliability results reflected considerable variability across parameters, with a low reliability for ratings of pitch and loudness (Cronbach's Alpha = 0.44 and 0.37) respectively. In contrast to the studies done by Webb, 2004³ and **Chen et al, 2018**¹⁶ who used GRBAS scale in assessment of voice quality, their estimates of inter-rater reliability

were good, the variations between our study and their studies may be related to the degree of listeners' agreement and reliability which appears to be reduced when judges were asked to isolate specific perceptual features in voices with a complex mix of perceptual characteristics, particularly if those voices are mild–moderately impaired.

Similarly, a study by **Özcebe et al, 2017**¹⁷ who used CAPE-V in assessment of voice quality, their estimates of inter-rater reliability for pitch and loudness were actually high. In their case, they used CAPE-V in the assessment of vocal impairment, a fact that might explain the differences between their results and ours.

We used the modified GRBAS scale for the assessment of vocal impairment accordingly we evaluated additionally the glottal attack and associated laryngeal functions as cough; estimates of inter-rater reliability for the glottal attack were very poor (Cronbach's Alpha = 0.38) and ratings for cough were poor (Cronbach's Alpha = 0.53). A thing to add that there isn't a standard scale assessing the glottal attack and the associated laryngeal function and these poor ratings of inter-rater reliability particularly for these parameters probably due to the nature of the rating task itself since auditory perceptual assessment of impaired voice quality entails a listener's judging of voice sample consisting of connected speech; a thing that's not existing in glottal attack or the associated laryngeal functions as cough.

On analysis of the raters' perceptual predictive diagnosis of patients' disordered voice quality, we found that 44% of cases diagnosed as functional voice disorder, while 52% of patients diagnosed as MAPLS and only 4% of them were normal according to rater 1. As regards rater 2; 56% of cases

diagnosed as functional disorders and the other 44% diagnosed as MAPLS. For rater 3; 40% of cases were diagnosed as functional and the remaining 60% of patients were diagnosed as MAPLS. However, estimates of inter-rater reliability for perceptual predictive diagnosis were poor (Cronbach's Alpha =0.279). From these data, it's clear that there were considerable differences between raters in predicting the diagnosis of impaired vocal quality which may be attributed to a functional origin or minimal associated pathological lesions and since some of the functional vocal impairment can lead to disordered vocal qualities which are quite similar to those resulting from minimal associated pathological lesions so, that might be the cause of some misinterpretation by the raters in predicting the diagnosis perceptually. Furthermore, **Kreiman and Gerratt, (2000)**¹⁸ suggested that experienced listeners may introduce more variability into judgments of voice quality because they use a flexible strategy to determine salient perceptual features, making continual adjustments as they fine-tune their decisions. Consequently, this clinical experience may actually lower the reliability of judgments. Also, while the raters were all voice specialists, their diverse background, training, and clinical settings may reflect inconsistencies. Moreover, we averaged our data across raters, which may have obscured notable consistencies in individual perceptual decision making. Finally, we didn't include normal voices in our listening samples, to reflect the clinical spectrum of voice severity, which includes recovery to normal or near-normal vocal quality.

Our analysis of the video endoscopic findings of patients' disordered voice quality we found out that 22 (44%) of patients were diagnosed as functional

voice disorders and 28 (56%) of patients were diagnosed as MAPLs. On comparing between these findings and the predicted perceptual diagnosis for each rater; a statistically significant relation between both tools of assessment was found for rater 2 and 3 with $P < 0.03$ and < 0.04 respectively. However, our data showed a non-significant relation for rater 1. The reliability of perceptual judgments of disordered voice quality is not the only determinant of the adequacy of this type of voice evaluation method. The validity of this method should also be considered and in this context we compared between the perceptual predictive diagnosis and the video endoscopic findings accordingly, strong association between objective measurement which is the video endoscopic evaluation and auditory perceptual assessment was found for two of the raters which confirms the validity of auditory perceptual assessment.

Although GRBAS is an easy, reliable, and valid auditory-perceptual assessment method, the narrow rating range is limited to 0–3 (normal, mild, moderate, and severe), making it difficult to rate subtle voice changes. In addition, because GRBAS does not have a specific protocol for data collection in terms of the variability of speech samples and the possible effects of the task order, it is difficult to compare different raters' results across different studies.¹⁹⁻²⁰ Although these factors can limit the reliability of listener's judgments of voice quality, additional studies showed that coupling auditory perception with instrumental measures produced higher accuracy in determining the presence/absence and extent of voice disorders than either approach alone.²¹

Our data showed some strong inter-rater reliability for some parameters of vocal impairment such as overall grade

of dysphonia severity, strain, they also showed fair reliability for ratings of other parameters as leaky and roughness voice qualities, yet, they showed low reliability for ratings of pitch and loudness. The most valuable outcomes of our study are the strong association between instrumental diagnostic measures as shown by the video endoscopic examination and the perceptual predictive diagnosis denoting a clear relationship between listener's judgments of disordered voice quality and the underlying vocal pathology; a thing that is in favor for the validity of APA as a popular and fundamental tool for the assessment of vocal impairment. So, after all, we can diagnose voice disorders by our ears.

Conclusion:

Auditory perceptual assessment is used as an easy, non-invasive and subjective method for the routine clinical voice assessment. This study showed considerable variability of the inter-rater reliability across parameters of disordered voice quality with acceptable inter-rater reliability for some parameters and unacceptable or poor for others. Yet, the most valuable outcomes of our study are the strong association between instrumental diagnostic measures as shown by the video endoscopic examination and the perceptual predictive diagnosis denoting a clear relationship between listener's judgments of disordered voice quality and the underlying vocal pathology; a thing that is in favor for the validity of APA as a popular and fundamental tool for the assessment of vocal impairment. So, after all, we can diagnose voice disorders by our ears.

Reference:

- 1- Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *Journal of Voice*. 1999 Dec 1;13(4):508-17.
- 2- Behrman A. Common practices of voice therapists in the evaluation of patients. *Journal of Voice*. 2005 Sep 1;19(3):454-69.
- 3- Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *European Archives of Oto-Rhino-Laryngology and Head & Neck*. 2004 Sep 1;261(8):429-34.
- 4- Yamaguchi H, Shrivastav R, Andrews ML, Niimi S. A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatrica et Logopaedica*. 2003;55(3):147-57.
- 5- Omori K. Diagnosis of voice disorders. *JMAJ*. 2011 Jul;54(4):248-53.
- 6- Kotby MN. Voice disorders: recent diagnostic advances. *Egypt J Otolaryngol*. 1986;3(10):69-98.
- 7- Deliyski DD, Shaw HS, Evans MK, Vesselinov R. Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatrica et Logopaedica*. 2006;58(4):274-88.
- 8- Hirano M. Clinical examination of voice. *Disorders of human communication*. 1981; 5:1-99.
- 9- Colton RH, Casper JK, Leonard R. *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Lippincott Williams & Wilkins; 2011.
- 10- Hirano M. Objective evaluation of the human voice: clinical aspects. *Folia Phoniatrica et Logopaedica*. 1989;41(2-3):89-144.
- 11- Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx*. 2015 Jun 1;42(3):183-8.
- 12- American– Speech– Language Hearing Association: *Ethics in research and professional practice*. ASHA Supplement, 2002, 1 (7), 63-65.
- 13- Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*, 1980 ,90:441–451.
- 14- Oates J, Russell A. Learning voice analysis using an interactive multi-media package: development and preliminary evaluation. *J Voice*; 1998,12: 500–512.
- 15- Fex S. Perceptual evaluation. *J Voice*, 1992, 6 (2), 155-158.
- 16- Chen Z., Fang R., Zhang Y., Ge P., Zhuang P., Chou A. and Jiang J. (2018): The Mandarin Version of the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) and its reliability. *Journal of Speech, Language, and Hearing Research*. Oct 26; 61 (10): 2451-7.
- 17- Özcebe E, Esen F A, Tuğ çe Karahan Tigra, Önal İncebay and Taner Yılmaz (2017): Reliability and validity of the Turkish Version of the Consensus Auditory- Perceptual Evaluation of Voice (CAPE-V), *J of Voice on press*.
- 18- Kreiman J, Gerratt B. (2000): *Measuring Vocal Quality*. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement*. San Diego, CA: Singular: 73–101.
- 19- Kreiman J, Gerratt BR, Ito M. (2007): When and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am.*; 122:2354– 2364.
- 20- Mozzanica F, Ginocchio D, Borghi E, Bachmann C, Schindler A. Reliability and validity of the Italian version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Folia Phoniatrica et Logopaedica*. 2013;65(5):257-65.
- 21- Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, Hillman R. Evidence-based clinical voice assessment: a systematic review. *American Journal of Speech-Language Pathology*. 2013.