e-ISSN: 0975-5160, p-ISSN: 2820-2651

Available online on www.ijtpr.com

International Journal of Toxicological and Pharmacological Research 2022; 12 (5); 189-196

Original Research Article

Development of a Tool to Objectively Identify the Normal Human Voice

Krishna Murari Bansal[1], Prithvi Raj Singh[2]

[1]Associate Professor, Department of ENT, Krishna Mohan Medical College & Hospital, Pali Dungra, Sonkh Road, Mathura U.P.

[2]Associate Professor, Department of ENT, Krishna Mohan Medical College & Hospital, Pali Dungra, Sonkh Road, Mathura U.P.

Received: 25-03-2022 / Revised: 25-04-2022 / Accepted: 30-05-2022

Corresponding author: Dr. Prithvi Raj Singh

Conflict of interest: Nil.

Abstract

Background: Acoustic analysis is used to assist in differential diagnosis, documentation, and evaluation of treatment for voice disorders. Clinical data has shown that Jitter, Shimmer, Mean Pitch, and Harmonic Noise Ratio are the indices of voice pathology. A voice with some periodicity can now be analyzed with a computerized acoustic analyzer, a relatively newer technique that can be widely used in clinical practice. Voice is an acoustic output of the vibrations of the vocal folds and is the basic source of speech. In contrast, speech is a meaningful acoustic output created by the modulation of voice by organs of articulation into basic building blocks, the 'phonemes'. Phonemes help in distinguishing one word from another in a particular language. Some sounds like clicks, whistling, and whispering can be produced by organs of articulations without voice. Because of anatomical, physiological, racial, cultural, and social factors, every human voice is unique and fingerprints is the signature of each individual.

Aim: To create a database of normal voices, analyze and identify different parameters of these voices and hence identify benchmarks of normal voices.

Material and Method: The study were conducted in the department of Otolaryngology. Voice samples of 250 normal males and 250 normal females aged between 19 to 30 years were collected using a sustained vowel /a/ which was recorded and analyzed using a freely downloadable software "PRAAT". The parameters like Jitter, Shimmer, and Pitch were derived, and mean, SD, and range of voice parameters were calculated. The Microphone was held at a distance of 5cm in front of the lips and 3 cm above the breath stream. Each person was first trained to produce sustained vowel /a/ by the examiner herself through utterance of the voice at comfortable loudness and pitch.

Results: In males, the value of parameters was mean pitch (124.05), jitter (0.011), and shimmer (0.04). In females the parameters were mean pitch (212.27), jitter (0.01), shimmer (0.06). The Pitch in females is consistently higher than in males. Jitter Range is almost the same in both males and females. There is not much difference in Shimmer between Males and Females.

Conclusion: Voice can be objectively analyzed using Acoustic Parameters like mean Pitch, Jitter, and Shimmer. Most of the studies show that normal voice parameters depend on gender, region, methodology of the voice collection, software and hardware used, different algorithms used for calculations and the setup, etc. Hence every institution should standardize the method of acoustic analysis for its own consumption. For the purpose, we have created a huge database of the voices of normal young adults. Voices can be objectively analyzed using

acoustic parameters like mean pitch, jitter, shimmer, and harmonic-to-noise ratio. A large database yields more reliable normative parameters. Institutions should develop their own standard protocol for the selection of subjects, recording of voices, and their analysis.

Keywords: Voice Database, Acoustic Analysis, Parameters Of Normal Voice, Hoarseness Of Voice, Voice Assessment, Voice Disorder.

Introduction

Voice is an acoustic output of the vibrations of the vocal folds and is the basic source of speech. [1] In contrast, speech is a meaningful acoustic output created by the modulation of voice by organs of articulation into basic building blocks, the 'phonemes'. Phonemes help in distinguishing one word from another in a particular language. Some sounds like clicks, whistling, and whispering can be produced by organs of articulations without voice. Because of anatomical, physiological, racial, cultural, and social factors, every human voice is unique and fingerprints are the signature of each individual. It has a vital role in both emotional and linguistic communication. For earning one's livelihood and to express feelings in one's occupation and personal interactions voice production skill is a must. Voice can give information about the speaker's age, sex, personality, emotional, and health status.

A normal voice is coherent i.e., it is well-planned, clear, and sensible. [2] It results from the coordinated, intricate movements of the muscles larynx. This is the most advanced sensorimotor system to be found in the human body. A harsh voice is a serious handicap to the speaker. Social contacts and professional and familial relationships suffer because of changes in voice. Voice disorders in lecturers, teachers, salesmen, actors, or singers cause problems in the profession. As said earlier every voice is unique. Even in a given individual, voice can change according to his /her physical and emotional status. Hence defining a normal voice and measuring its parameters for an individual or a population is a difficult task.

It is logical to assume that the parameters of voice can be different for different disease entities. Before trying to correlate the parameters of voice with such entities it becomes obligatory to define what is normal. The methods being used now by the scoring of voice analysis by perception are fairly common. But they are subjective and won't yield to documentation. The location of vocal folds makes it difficult to physically measure their movements. X-rays and ultrasounds are not of much use because cartilages surround the vocal folds. Video evaluation does not help since the movements of the vocal folds are quite rapid (80 to 300 Hz). [3,4] Stroboscopy and high-speed videography need the instruments to be placed in the throat and cause gagging and laryngeal spasm and stop the production of voice. Moreover, all these procedures are invasive and cause discomfort to the patient. Non-invasive physical measurements have a greater advantage in that they yield an objective and documentable approach and can allow reliable comparison of voice samples (e.g., before and after treatment), therapeutic methods (e.g., microsurgery versus laser), etc.5

In the majority of ENT and Speech Therapy Units clinical evaluation of voice is still perceptive. But the modern era demands a reliable, documentable technique to quantify and standardize voice characteristics. [6] There are

objective methods like computerized voice analysis which are noninvasive, fast, and reliable and clinicians can easily set up a system for the purpose. Though computerized analysis is being extensively used throughout the world no world standards have been established for normative parameters. [7] Since voice may be affected by physical, emotional, professional, and social factors the Clinician, ENT surgeon, or speech pathologist has to evaluate the patient in its entirety. The inquiry should be exhaustive and should consider all these factors. In a few attempts the data analyzed is not large enough to proclaim the universality of these parameters. It is needed to test on a larger database to make the parameters more universal. In the Indian context, the research carried out is limited and the size of the data tested is quite small. The purpose of the current work is a) to create a large normal voice database of the local population and b) to evaluate this data and create standardized values of normative parameters.

Material and Methods

The study were conducted in the department of Otolaryngology. Young healthy adults, between the age of 19 and 30 years, both male and female were selected for the study. All of them gave informed consent. They underwent thorough clinical evaluation and those having any pathological condition or even suspicion of it were excluded. They were subjected to an assessment of their voices by a speech therapist and an ENT surgeon. Only those who were certified as having normal voices were selected for the study. Finally, there were 500 test subjects left.

Type of Study: Prospective Study

Source of Data: Young adults between 19 and 30 years of age, both males and females were selected from nearby colleges. Informed consent was taken

Inclusion Criteria: Detailed history specifically about systemic diseases, upper respiratory tract infection, tobacco and alcohol consumption, and voice abuse was taken. They underwent detailed clinical examinations. Those who were screened by the above-said procedure were selected for the study.

Exclusion Criteria: Voice abuse Hearing loss Upper or lower respiratory tract infection Any chronic systemic illness like tuberculosis, diabetes, hyper or hypothyroidism, Neurological disease, etc.

Sample Collection Setup

1. The voice samples were recorded in a sound-treated room.
2. A regular Windows Desktop Computer was used.
3. A microphone, unidirectional microphone (Sony Audio-Technica 250XL) was used

The Microphone was held at a distance of 5cm in front of the lips and 3 cm above the breath stream. Each person was first trained to produce sustained vowel /a/ by the examiner herself through the utterance of the voice at comfortable loudness and pitch. The sustained vowel, /a/ was recorded for a minimum of 3 seconds using PRAAT8 software. A total of 500 vowel samples were recorded. Only 1000 quality samples in terms of uniformity of volume and pitch were available because of technical reasons. The 3-sec sustained vowels were then extracted in the spectrogram of PRAAT to get the most stable and uniform middle 1-sec segment.

Extraction of Parameters

The extracted audio clip was submitted to PRAAT8 to get the parameters

To speed up and automate the extraction of parameters, a PRAAT Script was written and utilized.

The Script extracted, in a single go, parameters of batches of 20 voice clips in each from a folder in 1-2 minutes and pushed them to an excel sheet.

Statistical Analysis

The extracted parameters were descriptively summarised using SPSS 20 Software2. Mean +/- Standard Deviation, and Standard Deviation were used for continuous variables. The number of percentages was used in the data summaries for categorical data. Wherever necessary graphs and diagrams were represented. A normality test was used for the determination of distribution and range.

Result

Table 1: Are shown the computed values of Pitch.

Statistical Parameter	Male	Female
Mean	124.0761416	212.1341253
SD	25.24413511	21.22014861
Mean+SD	153.324421	156.4485892
Mean-SD	99.6800866	197.9669098
Max Values	268.182	436.422
Min Values	66.231	75.594

The Pitch in females is consistently higher than in males.

Table 2: Shows the values of jitter (ddp) in males and females are depicted.

Statistical Parameter	Male	Female
Mean	0.0112	0.011695
SD	0.011552	0.009269
Mean+SD	0.02115	0.018764
Mean-SD	0.00109	0.0099
Range of Jitter	0.00106-0.02312	0.001012-0.020752
Max Values	0.1432	0.12344
Min Values	0.00150	0.00180

Jitter Range is almost the same in both males and females

Table 3: Shimmer values males and females

Statistical Parameter	Male	Female
Mean	0.042649	0.065705
SD	0.062237	0.0020
Mean+SD	0.121116	0.120225
Mean-SD	0.012273	0.015471
Max Values	0.43502	0.58255
Min Values	0.01164	0.01693
Range of Shimmer	0.013141-0.13141	0.015262-0.120224

There is not much difference in Shimmer between Males and Females

Discussion

Our objective was to identify and standardize the parameters of normal voice by a simple, easier, and non-invasive method so that this becomes a handy tool for day-to-day use to the clinical practitioner who addresses vocal disorders. The perceptive methods have been subjective, difficult to quantify, and document, and not reproducible. [9] Other non-subjective methods wherein instruments are used are invasive, time-consuming, and need expensive equipment. [10,11] With stroboscopy analysis alone it is difficult to diagnose disorders like spasmodic dysphonia. [12] As a non-invasive, objective, easier method acoustic analysis proves valuable in the diagnosis and management of voice disorders. [13] Acoustic analysis requires a simple computer, a microphone, and voice

analysis software. [14] Worldwide acoustic analysis became a standard practice and many researchers began to analyze normal and abnormal voices and soon came to know that voice is multi-dimensional.

When we tried to compare our values with that of other studies, it was found that most of the data were calculated in different ways. For e.g., Jitter can be measured as Jitter%, Jitter ddp, absolute Jitter, local jitter, etc. Similarly, Shimmer can be measured as dda, Shimmer %, Shimmer Db, absolute Shimmer etc. Hence comparison becomes difficult. There are some studies given below which we tried to compare with our study. For example, Ana Clara Felippe et al.2006 [15], Kirt Aries 2012 [16] and Grahaam Williamson 2017 [17] considered jitter in terms of % jitter. Deqhan et al 2010 [18] considered Average jitter. Simone [19] took the jitter factor for calculation. The author of this study considered Jitter(ddp).

Di Niccola et Al2001 [20] worked to check the possibilities, reliability, and also limitations of a procedure that was objective and parametric and evaluated the normal and pathological voices. Here, HNR was analyzed for determining the relationship between noise and harmonics using 208 subjects of which 60 were normal and 148 were abnormal. The results concluded that voice analysis is not only simple but highly sensitive. They concluded in so many words - "For the data obtained to be valid, a necessary condition is the application of a strict, precise, correct sampling and analysis method following well-defined rules. Finally, the values obtained can serve as a basis for the construction of an objective instrumental voice measurement protocol that can be used in the forensic evaluation of dysphonia. Standardization of the regulations is essential to such a project."

As per the opinion of Bonzi et Al2014 [21] acoustic analysis is one of the major advances in the study of voice, increasing the accuracy of diagnosis in this area. They felt that normal values as standards are important and necessary to guide voice professionals. Using Praat they analyzed 72 voices of female and male voices among the Argentinian Spanish-speaking population. Kirt Delovino et al2012 [16] opine that "despite the accuracy and reliability of each machine, authors have agreed to standardize normative data individually due to a number of factors that may cause variations among each system. These possibilities include the type of programming of the acoustic analysis software, the use of recording criteria, the type of microphone, and other devices used in voice recording."

The effect of recording quality on the analysis of voice and speech was studied by Vogel et al2009 [22]. They opine that acoustical parameters depend on the environment, the expertise of the clinician, and the parameters extracted. According to them, the best quality can be obtained using a sound-treated room, recording on a hard disc, an dedicated mixer, and a good microphone. Graham Williamson [17] in his article states: "It is difficult to be precise about norms for acoustic measures such as jitter, Shimmer, noise-to-harmonics ratio, and fundamental frequency. There are many factors that militate against declaring all-encompassing norms. Some of these are person-specific, gender and age differences), cultural (e.g. what north Americans may consider being within normal limits may be different from what north Koreans consider to be typical), and related to the testing environment (e.g. variation in the equipment used, and, importantly the use of different algorithms in the software programs which are used to make the measurements). Measures of jitter and Shimmer using one software program cannot always be compared directly with measures made by another software program."

Wang et al2004 [23] used a computer speech lab Aerophone system manufactured by Kay Elemetrics Corp to establish parameters for normal individuals(both male and female, 45 each). The fundamental frequency was 118 Hz for males and 213 Hz for females. However, there was no gender difference in jitter though shimmer and H/N ratio were variable in different genders and age groups. Haldun Oguz2011 [24] compared Praat and MDVP (Multi-Dimensional Voice Program) and found significant differences in Jitter and HNR. Maryn et al2007 [25] compared jitter and Shimmer measures using both MDVP and Praat programs. The authors noted that MDVP yielded higher values than Praat. They conclude that one cannot compare jitter and shimmer outcomes across systems and programs. Steven Bielamowicz et al1996 [26] have compared perturbation measures from different systems like CSpeech, SoundScope, Hand Marking Voice Analysis system and Computerized Speech Laboratory. The results were different for different systems and made the authors to opine that different systems yield different units.

Toran et al2010 [14] analyzed the voices of vocal polyp patients pre and post-operatively. There were changes in parameters that indicated improvement in the voice post-operatively. Authors opined that objective analysis can provide a tool to clinicians for a better understanding of the quality of voice. A review was done by Hartl et al2005 [27], for providing an update on current techniques of dysphonia evaluation in routine clinical practice. They concluded that the objective evaluation of the voice parameters like fundamental frequency and the spectral characteristics of voice has the advantage of being simple to perform, reproducible and quantifiable. But, during the evaluation of severe dysphonia, automatic measurements need to be analyzed with caution as the computer algorithms being designed for voices retain a certain periodicity. [28] They concluded that all of these types of analysis are complementary, informing as to different aspects of vocal quality and laryngeal function. No one measurement alone can diagnose or characterize dysphonia.

Conclusion:

Voice can be objectively analyzed using Acoustic Parameters like mean Pitch, Jitter, and Shimmer. Most of the studies show that normal voice parameters depend on gender, region, methodology of the voice collection, software and hardware used, different algorithms used for calculations and the setup, etc. Hence every institution should standardize the method of acoustic analysis for its own consumption. For the purpose, we have created a huge database of the voices of normal young adults. Voices can be objectively analyzed using acoustic parameters like mean pitch, jitter, shimmer, and harmonic-to-noise ratio. A large database yields more reliable normative parameters. Institutions should develop their own standard protocol for the selection of subjects, recording of voices, and their analysis.

References:

1. Mathieson L, Carding P. The Voice and speech production. In: Michael Gleeson. (Ed). Scott-Brown's Otorhinolaryngology and Head and Neck Surgery. London: Hodder Arnold, 2008;7:2166-72.
2. Phoneme -Wikipedia. (n.d). Retrieved September 26, 2018. from https://en. wikipedia.org/wiki/Phoneme
3. Voice Analysis - Wikipedia. (n.d) Retrieved Jan 30, 2017 from http:// www.factindex.com/v/vo/voice_analysis.html
4. Voice Analysis - Wikipedia. (n.d) Retrieved Jan 29, 2017 from https:// wikivisually.com/wiki/Voice_analysis
5. Toran KC, Lal BK. Objective voice analysis for vocal polyps following

micro-laryngeal phono surgery. Kathmandu University Medical Journal. 2010; (8) 30: 185-189.
6. Mehta DD, Hillman RE. Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. Current opinion in otolaryngology & head and neck surgery. 2008;16(3):211.
7. De Felippe AC, Grillo MH, Grechi TH. Standardization of acoustic measures for normal voice patterns. Brazilian Journal of Otorhinolary ngology. 2006;1;72 (5):65 9-64.
8. Boersma P., Weenink D., "PRAAT: Doing phonetics by computer", Version 4.3.14, Computer program, 2005. IBM Support. IBM, Retreived on, 2018.
9. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. Journal of Speech, Language, and Hearing Research. 1993;36(1):21-40.
10. Boone DR, McFarlane SC, Berg SLV, Zraik RI. The voice and Voice Therapy. 8th edition, Boston: Allyn and Bacon; 2010.
11. Patel RR, Harris SM, Halum SL. Objective voice Assessment. In: Benninger M. (Ed). Sataloff's Comprehensive textbook of Otolaryngology-Head and Neck surgery. New Delhi: The Health Sciences Publisher, 2016;1:155-66.
12. Woo P. Stroboscopy and High-speed video examination of Larynx. In: Benninger. MS. (Ed). Sataloff's Comprehensive textbook of Otolaryngology-Head and Neck surgery. New Delhi, The Health Sciences, 2016;1:23-33.
13. Baken R, Olikoff R. Clinical measurement of speech and voice. Singular Thomson learning: 2000;2.
14. Toran KC, Lal BK. Objective voice analysis for vocal polyps following micro-laryngeal phono surgery. Kathmandu University Medical Journal. 2010; (8) 30: 185-189.
15. De Felippe AC, Grillo MH, Grechi TH. Standardization of acoustic measures for normal voice patterns. Brazilian Journal of Otorhinolar - yngology. 2006;1;72(5):659-64.
16. Delovino KA, Casile RU, Hawson F. Vocal Acoustic Measures of Asymptomatic Filipino Young Adults at a Private Tertiary Hospital in Quezon City-A Pilot Study. Philippine Journal of Otolaryngology-Head and Neck Surgery. 2012;1;27(2):7-11.
17. Williamson G. Acoustic Measures (Norms). Voice. [Internet] 01 February 2014. Viewed on 2020,5.
18. Dehqan A, Ansari H, Bakhtiar M. Objective voice analysis of Iranian speakers with normal voices. Journal of Voice. 2010;1;24(2):161-7.
19. Simone AA, Jose CP, Marcelo OR. Standardization of acoustic measures of the normal voice. Brazilian J Otolaryngol. 2002; 68:4(14) :540-44.
20. Di VN, Fiorella ML, Luperto P, Staffieri A, Fiorella R. Objective evaluation of dysphonia. Possibilities and limitations. Acta Otorhinolaryngologica Italica. 2001;21 (1):10-21.
21. Bonzi EV, Grad GB, Maggi AM, Muñóz MR. Study of the characteristic parameters of the normal voices of Argentinian speakers. 2014;6 (1):23-25.
22. Vogel AP, Morgan AT. Factors affecting the quality of sound recording for speech and voice analysis. International journal of speech-language pathology. 2009 ;1;11 (6):431-7.
23. Wang CC, Huang HT. Voice acoustic analysis of normal Taiwanese adults. J Chinese Med Ass. 2004;67(4):179-84.
24. Oğuz H, Kiliç MA, ŞAFAK MA. Comparison of results in two acoustic analysis programs: Praat and MDVP. Turkish Journal of Medical Sciences. 2011;41(5):835-41.

25. Maryn, P, Corthals J, Vanderbeke, E. Vanwynsberge M, De Bodt, P, Cauwenberge V. Intersystem and interprogram difference in acoustic perturbation measurement: a comparison between MDVP and Praat. In: Paper presented in the7th Pan European Voice Conference (PEVOC7), Groningen, The Netherlands, 2007.
26. Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. Journal of Speech, Language, and Hearing Research. 1996;39(1):126-34.
27. Hartl DM, Hans S, Crevier LB, Laccourreye O, Vaissiere J, Brasnu D. Dysphonia: current methods of evaluation. InAnnales d'oto-laryngologie et de chirurgie cervico faciale: bulletin de la Societe d'oto-laryngologie des hopitaux de Paris 2005; 122 (4):163-172.
28. Votta, M., Nicoletti, T., & Cardillo, M. (2020). The relevance of pain management & relief in the COVID-19 outbreak: an evaluation of the ten years implementation of the dedicated law in Italy. Journal of Medical Research and Health Sciences, 2020;3(12): 1130–1134.