Original Article

https://doi.org/10.1590/2317-6431-2018-2113

Audiology Communication Research
ISSN 2317-6431

Acoustic voice analysis: comparison between two types of microphones

Análise acústica da voz: comparação entre dois tipos de microfones

Vanessa Mara Krik[1], Vanessa Veis Ribeiro[2], Larissa Thais Donalonso Siqueira[1], Marcelo de Oliveira Rosa[3], Ana Paula Dassie Leite[1]

ABSTRACT

Purpose: To compare acoustic measurements of voice recorded by two types of microphones. **Methods:** The study counted with 103 women from 18 to 54 years old. The recorded sample was the sustained vowel /ε/. The acoustic signal was picked up simultaneously by two unidirectional microphones: the Shure SM58 and the Karsect HT-9. The acoustic analysis was performed on these edited vocal samples. The measured parameters were F_0, Jitter, Shimmer and Glottal to Noise Excitation (GNE) ratio. **Results:** Recurrent differences between the microphones were observed only in Shimmer measurement (p=0.026); the Karsect HT-9 presented higher values. However, the acoustic measures were within the normal range for healthy voices, despite of the microphone used. **Conclusion:** The acoustic analysis results extracted from the voice recording performed with the Shure SM58 and the Karsect HT-9 microphones were similar. Hence, it can be deduced that, connected to a high-quality interface, both microphones can be used in the acoustic analysis to record the sound signal.

Keywords: Acoustics; Speech acoustics; Health evaluation; Sound; Voice

RESUMO

Objetivo: Comparar os valores das medidas acústicas da voz, extraídas de gravações vocais realizadas com dois tipos de microfones. **Métodos:** Participaram da pesquisa 103 pessoas do sexo feminino, com idades entre 18 e 54 anos. Foram coletadas amostras da vogal sustentada /ε/, captadas por dois microfones, simultaneamente: *Shure* SM58 e *Karsect* HT-9. Foi realizada a análise acústica das vozes, com a extração de valores de frequência fundamental (F_0), variação da frequência (*Jitter*) e variação de amplitude dos segmentos fundamentais da voz (*Shimmer*) e proporção *Glottal to Noise Excitation* (GNE). **Resultados:** Houve diferenças entre os microfones apenas na medida de *Shimmer,* com predomínio de valores mais altos captados pelo *Karsect* HT-9. Porém, os resultados de ambos os microfones estiveram dentro dos padrões de normalidade do *software* utilizado. **Conclusão:** Os resultados da análise acústica extraídos a partir da gravação de voz realizada com os microfones *Shure* SM58 e *Karsect* HT-9 foram semelhantes, concluindo-se que ambos os microfones podem ser utilizados para gravação do sinal sonoro na análise acústica.

Palavras-chave: Acústica; Acústica da fala; Avaliação em saúde; Som; Voz

Study carried out at Curso de Fonoaudiologia, Universidade Estadual do Centro-Oeste – UNICENTRO – Irati (PR), Brasil.
[1]Departamento de Fonoaudiologia, Universidade Estadual do Centro-Oeste – UNICENTRO – Irati (PR), Brasil.
[2]Programa de Pós-graduação em Distúrbios da Comunicação Humana, Universidade Federal de São Paulo – UNIFESP – São Paulo (SP), Brasil.
[3]Universidade Tecnológica Federal do Paraná – UTFPR – Curitiba (PR), Brasil.
Conflict of interests: No.
Authors' contribution: VMK study design, data collection, data analysis, writing of the manuscript, final revision of the manuscript before publication; VVR data analysis, writing of the manuscript, review of the study proposal, approval of the manuscript's final version before publication; LTDS review of the study proposal, approval of the manuscript's final version before publication; MOR data collection, review of the study proposal, approval of the manuscript's final version before publication; APDL study design, data analysis, writing of the manuscript, review of the study proposal, approval of the manuscript's final version before publication.
Funding: None.
Corresponding author: Ana Paula Dassie Leite. E-mail: pauladassie@hotmail.com
Received: November 30, 2018; **Accepted:** March 04, 2019

INTRODUCTION

The voice depends of the speaker anatomo-physiological characteristics, psychological traits, physical, social and cultural aspects. It is very important in interpersonal and professional communication skills[1,2].

The human voice is a complex phenomenon; therefore, the voice assessment must be multidimensional. It is recommended that clinical voice evaluation assess the following dimensions: clinical history, assessment of voice quality (perceptual judgement of the voice quality, acoustic and aerodynamic analyses) and patients' self-assessment. These dimensions complement each other and must be analyzed together in order to provide an overview of the voice outcomes. Thus, a diagnosis can be established, a therapeutic approach can be defined and clinical follow-ups can be performed, especially in cases ongoing vocal rehabilitation[1-3].

As mentioned, one of the dimensions of the voice assessment is the acoustic analysis; it aims to objectively quantify the physical parameters of the waveform in the time domain[4]. It is a computerized, objective, non-invasive, low-cost, time-efficient and user-friendly analysis that allows to transform a subjective perception of the voice signal into a quantified objective reality[1,5].

Despite the criticisms regarding the acoustic analysis, specially concerning the lack of precise correlation with the perceptual judgment, its importance is undeniable[6]. Some of the acoustic analysis positive aspects are: help to understand the voice production, generate standardize data, produce voice documentation, follow-up treatment outcome, clinical monitoring of patients ongoing vocal rehabilitation and early detection of voice problems[1,2,6]. Furthermore, it is difficult to determine the etiologies of certain vocal disorders with subjective assessments tools alone, i.e. perceptual judgment. The multidimensional evaluations, considering subjective and objective data, are more reliable[6].

The acoustic analysis is performed by recording a voice sample using a microphone. Hence, the microphone used to record the sound signal can influence in the outcomes of the extracted acoustic parameters[7]. Thus, proper sound recording is essential to guarantee a reliable extraction of acoustic measurements.

For scientific purposes, the voice signal recording should be done with a high-quality, unidirectional microphone with high impedance and flat frequency response. Such characteristics ensure that the acoustic signal is not distorted while picked up by the microphone, avoiding that the software mistakenly process the acoustic signal. However, such equipment has high financial costs. Nowadays, in addition to these equipment's, most of the laboratories pick up the acoustic signal with an external sound card[8,9], this reduces the internal noise of the computer and the environment noise while standardizing the procedure.

International studies comparing several microphones showed different accuracy of acoustic measurements according to the microphone type[7,10], which interfere on the discrimination of pathological from normal voices[7]. However, recent study comparing six different models of microphones connected to a preamplifier observed little differences between them; in addition, a $100 microphone produced the best responses and was suggested for clinical use[11].

These mentioned studies used imported high-cost microphones that does not represent the reality of most institutions and speech language pathology clinics in Brazil. For this reason, it is important to perform a study comparing the acoustic analysis outcomes using two different microphones that are commonly used in Brazilian clinics routine. Thus, influences regarding different sound signal pick up will be verified also in the Brazilian reality.

Hence, the aim of the present study was to compare acoustic measurements of voice recorded by two types of microphones.

METHODS

Study design

Observational, cross-sectional and analytical study design.

Ethical considerations

The present study was approved by the Committee for Ethics in Research of the *Universidade Estadual do Centro-Oeste* under the protocol number 706.335. All subjects agreed to participate and signed an informed consent form.

Sample

Participants were recruited in the local community. To select the participants, inclusion and exclusion criteria were established. Inclusion criteria were: women from 18 to 59 years old; absence of vocal complaint and normal voice quality (nearly periodic - type 1 signals[12]). Individuals who reported smoking and alcoholism were excluded. To apply the selection criteria, the participants answered a questionnaire elaborated by the researchers, with questions regarding: sociodemographic data, general health information and vocal habits. To perform a voice quality screening, the voice quality was perceptually judge using a visual analog scale. The judge was a voice specialist with more than ten years of experience in perceptual judgment and did not participate in the data collection. The judge listened to the voice and evaluated the overall degree of deviation on a visual analogue scale of 100 mm. Voices below the threshold of 35.5mm[13] were considered normal. Data from the questionnaire were used to characterize the sample.

A pilot study was performed to calculate the sample size. The estimation method considered the highest standard deviation of the difference between the means of the acoustic parameters; the highest value was 3.44Hz for the fundamental frequency (F_0). The significance level was set at 5% and the test power at 80% to detect the minimum differences between the microphones of 1 Hz. The minimum sample size was 95 participants. Thus, the study counted with 103 women, between 18 and 54 years old (mean age of 21 years old), without vocal complain and with normal voice quality.

Data collection

All procedures were performed by the same speech language pathologist in a room with acoustic insulation. Data collection was done individually and it lasted between 10-20 minutes per subject.

For the recordings, the participants were comfortably seated on a chair with knees bent at 90º angle and feet flat on the floor. They were instructed to keep their neck relaxed, without cervical inclination and with the chin parallel to the floor.

The individuals were asked to breathe in and phonate the sustained vowel /ε/ in maximum phonation time at habitual pitch and loudness. The acoustic signal was picked up simultaneously by two unidirectional microphones: the Shure SM58 cardioid vocal microphone for professional use, positioned 3 cm distance from the participant's mouth and the Karsect HT-9 head-mounted electret condenser microphone, place at 3 cm distance from the mouth and at a 45° microphone-to-mouth angle.

The samples were recorder in a computer using the external sound card M-Audio Fast Track C400. The software used was the Audacity (GNU GPL™) at stereo audio mode to guarantee that the signals of each microphone could be separated from the same sample once the recording occurred simultaneously. The Audacity program (GNU GPL™) was also used to calibrate the recording system. Thus, the level of the recorded signal was similar. Samples that eventually presented high saturated levels were excluded. For these cases, a new recording was performed. Posteriorly, the beginning and the end of the recordings were excluded to avoid raise and decay instabilities.

The acoustic analysis was performed on these edited vocal samples and extracted using the Voxmetria software (CTS *Informática*™). The measured parameters were F_0, Jitter, Shimmer and Glottal to Noise Excitation (GNE) ratio.

Data analysis

Descriptive and inferential statistical analysis were performed on the outcomes of the acoustic measurements obtained by the two microphones. The Shapiro-Wilk test was used to test for normal distribution which was obtained only for the F_0. The inferential analysis of this variable was performed using the parametric Paired t-test, the other variables were analyzed using the Wilcoxon test. The statistical tests were performed using the Statistics software, version 17.0 (Stat Soft Inc. ™) and the level of significance was set at 5%.

RESULTS

Recurrent differences between the microphones were observed only in Shimmer measurement (p=0.026); the Karsect HT-9 presented higher values. The outcomes for F_0, Jitter and GNE ratio were the same for both microphones, the Karsect HT-9 and the Shure SM58 (Table 1).

DISCUSSION

The type of microphone used to record the acoustic signal interferes in further analysis of the vocal parameters. The reason for this is the lack of recommendations regarding microphones technical specifications for data acquisition in clinical practice and research. Thus, findings among different studies are difficult to be performed and less reliable[14]. Some international studies compared the use of different microphones; however, they were mostly imported high-cost microphones, that does not represent the reality of most institutions and voice clinics in Brazil. Commonly, the Karsect HT-9 microphone is used among Brazilian professionals because it also provides the interface Andrea PureAudio™ USB-SA (Andrea Electronics) [15-18]. International studies commonly use the Shure SM58 microphone[19-22] coupled with high-quality recording interfaces. It is important to obtain scientific evidence to support the use of different types of microphones in the Brazilian clinic routine.

No difference was found between the values of F_0 obtained by the Shure SM58 and the Karsect HT-9 microphones. This outcome, along with previous studies[1,23-25], states that F_0 is one of the most consistent and robust acoustic parameters, despite of different recording conditions. The same outcome was observed for Jitter, a perturbation measurement, and GNE, a noise measurement.

These outcomes from the present research agrees with previous study that used synthesized vowels, with altered systematically acoustic characteristics, to determine which microphones were more reliable in the extraction of vocal measures. Using this acoustic signal with known characteristics, the authors compared six microphones, including the Shure SM58; all microphones presented very similar values to the generated acoustical measures[11].

On the other hand, regarding the perturbation measurement Shimmer, the results were different between the microphones, with higher values in the measurements extracted using the Karsect HT-9 microphone when compared to the Shure SM58. Shimmer, a short-term amplitude perturbation, is an involuntary variability of the waveform amplitude, i.e., it represents irregularities in the amplitude of the glottic cycles, from one cycle to another[2]. Shimmer is not as robust as F_0[1], however it has reasonable reliability[6]. In addition, the authors from this

Table 1. Comparison of acoustic measurements of fundamental frequency, Jitter, Shimmer and Glottal to Noise Excitation ratio in phonations recorded with the *Shure* SM58 and the *Karsect* HT-9

Measure	Microphone	Mean	SD	Median	Q25	Q75	p-value
F_0	*Shure* SM58	221.900	17.700	221.500	210.570	231.840	0.705
	Karsect HT-9	222.000	17.900	221.500	210.570	231.850	
Jitter	*Shure* SM58	0.300	0.193	0.200	0.100	0.400	0.181
	Karsect HT-9	0.268	0.188	0.200	0.130	0.350	
Shimmer	*Shure* SM58	2.595	1.096	2.300	1.790	3.360	0.026*
	Karsect HT-9	2.771	1.274	2.500	1.790	3.530	
GNE ratio	*Shure* SM58	0.839	0.136	0.900	0.780	0.940	0.112
	Karsect HT-9	0.843	0.132	0.900	0.790	0.940	

Paired t-test and Wilcoxon test; *Statistically significant results (p<0.05)

Subtitle: SD = standard deviation; Q25 = first quartile; Q75 = third quartile; F_0 = fundamental frequency; GNE = Glottal to Noise Excitation previous study infer that perturbation measurements are more sensitive to different recording conditions[26].

It is noteworthy that in the present study, the median and quartile values of Shimmer were within the normal range for healthy voices, despite of the microphone used. Thus, the interpretation of the values regarding the normality of the measurement, considering the Voxmetria software standards, would be similar.

A study that compared four different types of microphones showed no difference between them for the perturbation measurement Jitter, however, Shimmer was different. The authors attributed this difference to a combination of microphone frequency response and internal noise characteristics. The amplitude perturbation measurement relies on the cycle-to-cycle variation of the peak value of the glottal pulse, which is susceptible to noise and frequency shaping. To reduce this effect on the classification of a normal voice parameter, the authors suggested the use of an equalization filter to flatten the frequency response of the microphones which improved the classification accuracy. Also, they advise to combine acoustic parameters to achieve a good discrimination between normal and pathological voices, when non-professional microphones are used to extract acoustic measures[7].

The microphone used for the acoustic signal picked up can offer conditions that will contribute to an efficient signal. The Shure SM58 has two types of built-in filters below the metal part: the rolloff filter, a unidirectional (cardioid) pick up pattern that isolates the main sound source while minimizing unwanted background noise, and the built-in high frequency spherical filter that minimizes wind and breath "pop" noise. The Karsect HT-9 contains a foam windscreen filter. Thus, difference between both filters may have influenced the acoustic signal pick up and, consequently, generated different Shimmer values.

Despite the statistical difference, the relevance of these results in clinical practice must be considered. Difference in only one acoustic parameter are not considered to be clinically relevant and do not have enough impact that would suggest not to use the Karsect HT-9 microphone; also, the acoustic analysis is one dimension of several procedures considered to determine a diagnosis[27]. In addition, the analysis of perturbations measurements is more reliable and, as recommended by the literature, should be analyzed with other parameters; moreover, the isolated interpretation of these measurements should be cautious[7,27].

Still regarding the clinical relevance, the additional noise level would be the same for all reevaluations during follow-ups. However, the microphones should be used with a high-quality interface (M-Audio Fast Track C400), as proposed by the present study methodology.

This research recorded normal voices, therefore, further studies should be performed to compare different types of vocal signals and to investigate any influence that the microphone may have in clinical parameters that aim to discriminate normal and altered voice qualities. In addition, one of the present study limitations was to include only women. Hence, it is necessary to perform this research with adult male participants and verify if the findings are similar, regardless of the participant gender.

CONCLUSION

The acoustic analysis results extracted from the voice recording performed with the Shure SM58 and the Karsect HT-9 microphones were similar. Hence, it can be deduced that, connected to a high-quality interface, both microphones can be used in the acoustic analysis to record the sound signal.

REFERENCES

1. Behlau M, Madazio G, Oliveira G. Functional dysphonia: strategies to improve patient outcomes. Patient Relat Outcome Meas. 2015;6:243-53. http://dx.doi.org/10.2147/PROM.S68631. PMid:26664248.
2. Behlau M. Voz: o livro do especialista. 1. ed. Rio de Janeiro: Revinter; 2001.
3. Lopes LW, Silva JD, Simões LB, Evangelista DS, Silva POC, Almeida AA, Lima-Silva MFB. Relationship between acoustic measurements and self-evaluation in patients with voice disorders. J Voice. 2017;31(1):119.e1-10. http://dx.doi.org/10.1016/j.jvoice.2016.02.021. PMid:27049448.
4. Sader RM, Hanayama EM. Considerações teóricas sobre a abordagem acústica da voz infantil. Rev CEFAC. 2004;6(3):312-8.
5. Eadie TL, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. J Voice. 2005;19(1):1-14. http://dx.doi.org/10.1016/j.jvoice.2004.02.002. PMid:15766846.
6. Christmann MK, Brancalioni AR, Freitas CR, Vargas DZ, Keske-Soares M, Mezzomo CL, Mota HB. Uso do programa MDVP em diferentes contextos: revisão de literatura. Rev CEFAC. 2015;17(4):1341-9. http://dx.doi.org/10.1590/1982-021620151742914.
7. Parsa V, Jamieson DG, Pretty BR. Effects of microphone type on acoustic measures of voice. J Voice. 2001;15(3):331-43. http://dx.doi.org/10.1016/S0892-1997(01)00035-2. PMid:11575630.
8. Uloza V, Petrauskas T, Padervinskis E, Ulozaitė N, Barsties B, Maryn Y. Validation of the acoustic voice quality index in the Lithuanian language. J Voice. 2017;31(2):257.e1-11. http://dx.doi.org/10.1016/j.jvoice.2016.06.002. PMid:27427182.
9. Mezzedimi C, Di Francesco M, Livi W, Spinosi MC, De Felice C. Objective evaluation of presbyphonia: spectroacoustic study on 142 patients with Praat. J Voice. 2017;31(2):257.e25-32. http://dx.doi.org/10.1016/j.jvoice.2016.05.022. PMid:27427181.
10. Titze IR, Winholtz WS. Effect of microphone type and placement on voice perturbation measurements. J Speech Hear Res. 1993;36(6):1177-90. http://dx.doi.org/10.1044/jshr.3606.1177. PMid:8114484.
11. Kisenwether JS, Sataloff RT. The effect of microphone type on acoustical measures of synthesized vowels. J Voice. 2015;29(5):548-51. http://dx.doi.org/10.1016/j.jvoice.2014.11.006. PMid:25998411.
12. Titze IR. Workshop on acoustic analysis. Iowa: National Center for Voice and Speech; 1995. Summary statement; p. 26-30.
13. Yamasaki R, Madazio G, Leão SHS, Padovani M, Azevedo R, Behlau M. Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. J Voice. 2017;31(1):67-71. http://dx.doi.org/10.1016/j.jvoice.2016.01.004. PMid:26873420.
14. Vaz Freitas S. A avaliação das alterações vocais: registo e análise áudioperceptual e acústica da voz. In: Peixoto V, Rocha J, organizadores. Metodologias de intervenção em Terapia da Fala. Porto: Edições Universidade Fernando Pessoa; 2009. p. 225-53.
15. Zambão VR, Penteado RZCM, Calçada MLM. Condições de trabalho e uso profissional da voz de cantores de bandas de baile. Rev CEFAC. 2014;16(6):1909-18. http://dx.doi.org/10.1590/1982-0216201417713.
16. Lima AT, Lucena JA, Araújo ANB, Lira ZS, Gomes AOC. Perfil de extensão vocal em coristas após técnica de vibração de língua associada a escalas. Rev CEFAC. 2016;18(3):626-34. http://dx.doi.org/10.1590/1982-0216201618315415.
17. Santos ACM, Borrego MCM, Behlau M. Effect of direct and indirect voice training in Speech-Language Pathology and Audiology

students. CoDAS. 2015;27(4):384-91. http://dx.doi.org/10.1590/2317-1782/20152014232. PMid:26398263.

18. Martins S, Oliveira D. Correlation between the degree of neurogenic oropharyngeal dysphagia with the level of dysphonia in the elderly: analysis related. Rev Pesqui Cuid é Fundam Online. 2014;6(3):1191-201. http://dx.doi.org/10.9789/2175-5361.2014v6n3p1191.

19. Murray ESH, Hands GL, Calabrese CR, Stepp CE. Effects of adventitious acute vocal trauma: relative fundamental frequency and listener perception. J Voice. 2016;30(2):177-85. http://dx.doi.org/10.1016/j.jvoice.2015.04.005. PMid:26028369.

20. Valença EHO, Salvatori R, Souza AHO, Oliveira-Neto LA, Oliveira AHA, Gonçalves MIR, Oliveira CR, D'Ávila JS, Melo VA, Carvalho S, Andrade BM, Nascimento LS, Rocha SB, Ribeiro TR, Prado-Barreto VM, Melo EV, Aguiar-Oliveira MH. Voice formants in individuals with congenital, isolated, lifetime growth hormone deficiency. J Voice. 2016;30(3):281-6. http://dx.doi.org/10.1016/j.jvoice.2015.03.015. PMid:25953587.

21. Zourmand A, Ting H-N, Mirhassani SM. Gender classification in children based on speech characteristics: using fundamental and formant frequencies of Malay vowels. J Voice. 2013;27(2):201-9. http://dx.doi.org/10.1016/j.jvoice.2012.12.006. PMid:23473455.

22. Fadel CBX, Dassie-Leite AP, Santos RS, Rosa MO, Marques JM. Acoustic characteristics of the metallic voice quality. CoDAS. 2015;27(1):97-100. http://dx.doi.org/10.1590/2317-1782/20152014159. PMid:25885203.

23. Morris RJ, Brown WS Jr. Comparison of various automatic means for measuring mean fundamental frequency. J Voice. 1996;10(2):159-65. http://dx.doi.org/10.1016/S0892-1997(96)80043-9. PMid:8734391.

24. Ribeiro VV, Pedrosa V, Silverio KCA, Behlau M. Laryngeal manual therapies for behavioral dysphonia: a systematic review and meta-analysis. J Voice. 2018;32(5):553-63. http://dx.doi.org/10.1016/j.jvoice.2017.06.019. PMid:28844806.

25. Carson CP, Ingrisano DR-S, Eggleston KD. The effect of noise on computer-aided measures of voice: a comparison of CSpeechSP and the multi-dimensional voice program software using the CSL 4300B module and multi-speech for Windows. J Voice. 2003;17(1):12-20. http://dx.doi.org/10.1016/S0892-1997(03)00031-6. PMid:12705815.

26. Karnell MP, Hall KD, Landahl KL. Comparison of fundamental frequency and perturbation measurements among three analysis systems. J Voice. 1995;9(4):383-93. http://dx.doi.org/10.1016/S0892-1997(05)80200-0. PMid:8574304.

27. Brockmann-Bauser M, Drinnan MJ. Routine acoustic voice analysis: time to think again? Curr Opin Otolaryngol Head Neck Surg. 2011;19(3):165-70. http://dx.doi.org/10.1097/MOO.0b013e32834575fe. PMid:21483265.