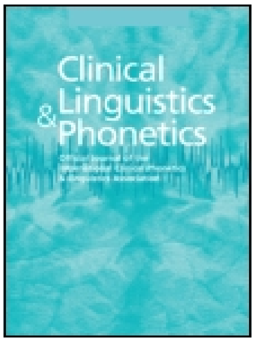

Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: https://www.tandfonline.com/loi/iclp20

Voice quality of psychological origin

António Teixeira, Ana Nunes, Rosa Lídia Coimbra, Rosa Lima & Lurdes Moutinho

To cite this article: António Teixeira, Ana Nunes, Rosa Lídia Coimbra, Rosa Lima & Lurdes Moutinho (2008) Voice quality of psychological origin, Clinical Linguistics & Phonetics, 22:10-11, 906-916, DOI: 10.1080/02699200802175974

To link to this article: https://doi.org/10.1080/02699200802175974

Published online: 09 Jul 2009.

Clinical Linguistics & Phonetics, October–November 2008; 22(10–11): 906–916

Voice quality of psychological origin

ANTÓNIO TEIXEIRA[1], ANA NUNES[2],
ROSA LÍDIA COIMBRA[2], ROSA LIMA[2], & LURDES MOUTINHO[2]

[1]*Departamento Electrónica, Telec. e Informática/IEETA, and* [2]*Centro de Línguas e Culturas (CLC), Universidade de Aveiro, Aveiro, Portugal*

(Received 27 October 2006; accepted 25 September 2007)

Abstract
Variations in voice quality are essentially related to modifications of the glottal source parameters, such as: F_0, jitter, and shimmer. Voice quality is affected by prosody, emotional state, and vocal pathologies. Psychogenic vocal pathology is particularly interesting. In the present case study, the speaker naturally presented a ventricular band voice whereas in a controlled production he was able to use a more normal phonation process. A small corpus was recorded which included sustained vowels and short sentences in both registers. A normal speaker was also recorded in similar tasks. Annotation and extraction of parameters were made using Praat's voice report function. Application of the Hoarseness Diagram to sustained productions situates this case in the pseudo-glottic phonation region. Analysis of several different parameters related to F_0, jitter, shimmer, and harmonicity revealed that the speaker with psychogenic voice was capable of controlling certain parameters (e.g. F_0 maximum) but was unable to correct others such as shimmer.

Keywords: *Voice quality, psychogenic dysphonia, hoarseness diagram, Portuguese*

Introduction

Voice quality is a difficult concept to define, including pertinent issues such as tone, vocal effort, linguistic variety, and its relation with linguistic factors such as prosody. According to Kreiman and Gerratt (2003: 58), 'the speaking voice naturally conveys information about the speaking individual … Voice quality serves as a primary means by which speakers project their identity'. Voice quality is affected by mood, attitude, and emotional state. It changes due to speech prosody and extralinguistic parameters. It is also affected by some voice pathologies that are related to larynx pathologies, articulation problems, and also motor control disorders. In all these cases the differences in voice quality are uncontrolled and do not have a distinctive function.

Some voice qualities depend on the control that a speaker has of their glottis (Gordon and Ladefoged 2001: 383). Speakers are capable, to varying degrees, of producing different voice qualities, even without pathology, according to motivation and effort. This voice quality variation can also be related to social/cultural context.

Correspondence: António Joaquim S. Teixeira, DETI/IEETA, Universidade de Aveiro, 3810-193 Aveiro, Portugal. Tel: +351 234 370520. Fax:+351 234 370545. E-mail: ajst@det.ua.pt

ISSN 0269-9206 print/ISSN 1464-5076 online
DOI: 10.1080/02699200802175974

Voice quality variations are essentially related to modifications of the glottal sourc parameters such as F_0; jitter and shimmer. It is based on these characteristics that Gordon and Ladefoged (2001: 383) identified several voice qualities (e.g. modal, tense, harsh, creaky, and whisper voices).

The frequency perturbation, called jitter, represents cycle-to-cycle variability of the fundamental frequency. It is a measurement 'of how much a given period differs from the period that immediately follows it' (Baken and Orlikoff, 2000: 90). Shimmer represents cycle-to-cycle amplitude variation. If human phonation were absolutely stable shimmer would be zero, but (just like jitter) it presents higher values when laryngeal disorders are present.

> The relationship of amplitude perturbation to specific abnormalities of glottal function, or of more global disorders of speech, remains, at the very best, extremely unclear. It seems clear however that shimmer values tend to normalize as laryngeal pathology abates (Baken and Orlikoff, 2000: 130).

Jitter and shimmer measures are often useful in discriminating between healthy and pathological voices (McAllister, Sederholm, Ternström, and Sundberg, 1996: 252–253). Harmonics-to-Noise Ratio (HNR) is another acoustic feature designed to measure the relative noise component in a speech signal (Michaelis, Fröhlich, and Strube, 1998: 1628). A healthy speaker can produce a sustained [a] or [i] with HNR of ~20 dB, and an [u] at ~40 dB.

Some psychological variables are determinants in the formation of a person's communication standards, including voice quality alterations. According to Baker (1998: 527), in clinical practice and literature 'voice disorders that exist in the absence of organic laryngeal pathology are referred to as functional; muscular tension dysphonias, hysterical dysphonias, conversion reactions or psychogenic dysphonias'. However, 'one common theme, which seems to link' populations with psychogenic voice disorders, 'is a pattern of inhibition around expressing thoughts and feelings' (Baker, 2002: 87). The diversity in terminology for these disorders reflects their complexity and difficulty in having an appropriate definition for those cases 'where no structural or neurological pathology can be identified to account for the dysphonia' (Baker, 2002: 84). According to Rubin and Greenberg (2002: 544–545) the term psychogenic appears interchangeably with *functional* or *non-organic*, the former being preferred due to the ambiguity of the others. This study adopted the following psychogenic designation (Baker, 2002: 84; Rubin and Greenberg, 2002: 544): a psychogenic dysphonia can be a 'manifestation of psychological disequilibrium such as anxiety, depression, personality disorder or conversion reaction to the extent that normal volitional control of the phonation is lost' (Baker, 1998: 527).

According to Baker (2002: 85), these types of dysphonias can be divided into more and less common presentations of Psychogenic Dysphonia. The most common can be presented in different forms: (a) 'Muscle tension dysphonia', usually related with emotional stress, (b) 'Conversion reaction' that 'exists despite normal structure and function of the neurophysiological system, is created by anxiety, stress, depression, or interpersonal conflict', (c) 'Mutational falsetto' that occurs with pubescent males as a resistance to the transition 'from the immature higher voice' to the 'lower pitched voice of maturity', and in mature males that can be a manifestation of a conversion reaction. Vocal quality with psychogenic dysphonia can be very variable, even 'from one second to another' (Baker, 1998: 529), and can sound just like the organic dysphonias (e.g. 'harsh', 'breathy', 'strained', or 'whispery' voice).

The objective with this study is to present a first analysis of a psychogenic dysphonic case where the speaker is also capable of producing a more controlled voice. It investigated differences between this pathological case and a non-pathological voice, how the controlled register compared to the dysphonic and normal voices, and which parameters could be controlled or not.

Material and method

Corpus

A new corpus was created, including several repetitions of sustained vowels (the three corner vowels) due to their stronger stability in phonation terms which obviously contributes to a better evaluation of the stability of the larynx and the vocal folds (Guimarães and Abberton, 2004) and a sentence ('O pato toca no pateta' [u ˈpatu ˈtɔkɐ nu paˈtɛtɐ] (The Duck touches Goofy)) produced in declarative, interrogative, and imperative forms.

Speakers

This study analysed the production of two adult males who were native speakers of European Portuguese with university degrees. The first informant had a diagnosed pathology with psychological origin, which induces the use of false vocal folds (ventricular bands voice) with hyper-function or laryngeal hyperkinesia. This increase of the muscle-skeletal tension induces an aberrant behaviour of the voice, therefore requiring a stronger effort in its production.

The psychogenic cause of this pathology leaves the speaker with the possibility of also producing a more controlled register, using the vocal folds. To produce the second register, the speaker needs both an increased effort and a higher level of concentration. The speaker was asked to repeat the vowels and sentences in his controlled register. The conditions needed for the controlled register (speaker difficulty in simultaneously maintaining concentration and sustaining vowels) combined with some recording problems made it possible to have only sentences for further analysis.

The second speaker, without any diagnosed voice pathologies, produced the same material previously described for the psychogenic speaker.

Acquisition, annotation, and feature extraction

The psychogenic voice speaker was recorded in a non-soundproof room in a hospital environment (in the presence of the therapist). The non-pathological case was recorded at the Universidade of Aveiro Phonetics Lab. In both cases, the signals were recorded directly to a computer hard disk using KAY Elemetrics CSL4400 and SFS software. Both speech and electroglotograph signals were recorded.

SFS and Praat (Boersma, 2001) were used for annotation and speech analysis. The Praat Voice functionality report was particularly useful for analysis, allowing extraction of the relevant parameters needed for a precise voice analysis. For statistical data analyses, SPSS was used.

Results

The following subsections present the results of the analyses regarding the most commonly studied parameters relevant to voice quality: F_0, jitter, shimmer, and harmonic-to-noise ratio. Graphic representation of the 95% Confidence Intervals is used to present the results regarding these parameters, considering two factors: the type of voice (Pathological, Normal, and Controlled, from the left) and the type of speech material (sentences or sustained vowels). Later in the study, the results of the application of the Hoarseness Diagram, combining information about irregularity and noise in a 2D representation, are also presented.

F_0

The present study considered several F_0-related parameters: average value, dispersion, minimum, and maximum. The normal male adult F_0 range is between 80–150 Hz (McAllister et al., 1996: 253). The results are presented in Figure 1.

With mean F_0, a decrease was noticed from an average of 200 Hz for sentences produced with pathological voice, to ~150 Hz when using a controlled register, a smaller value than the one obtained for normal voice samples. Values for the two types of speech material are

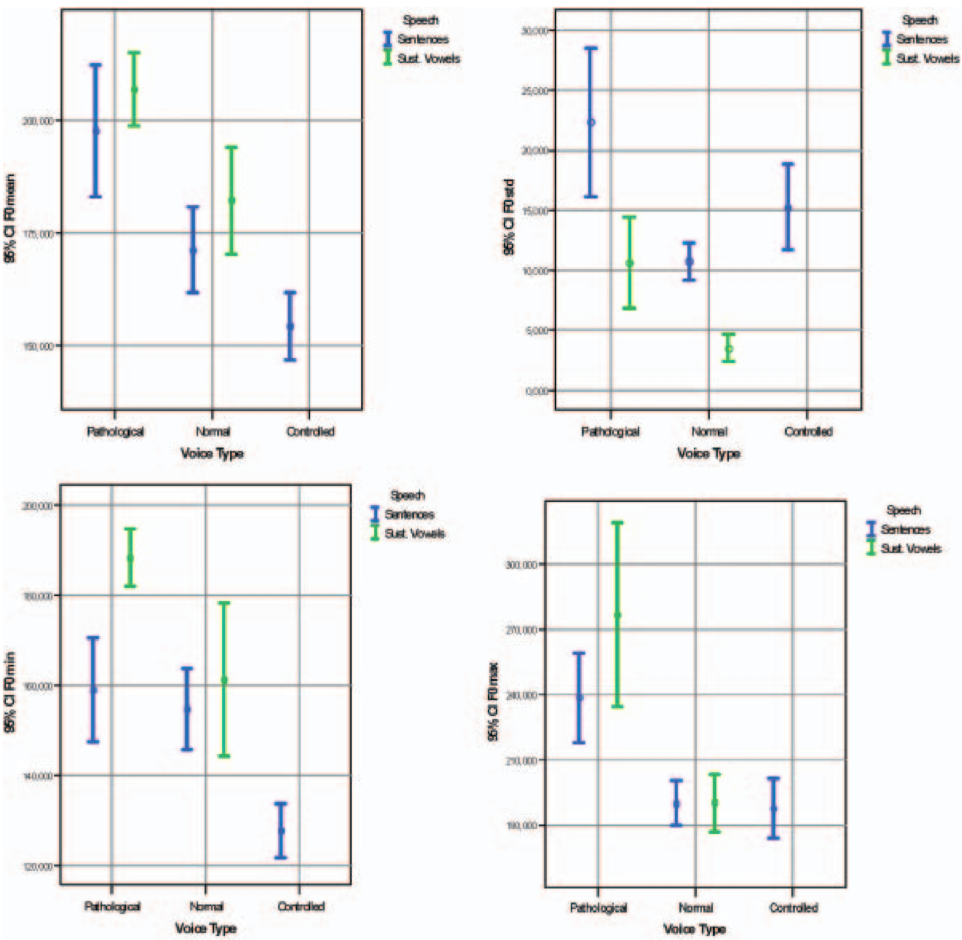

Figure 1. 95% Confidence Interval for four F_0 parameters (mean, SD, min, and max), considering two factors, the type of voice (Pathological, Normal, and Controlled, from the left) and the type of speech material (sentences or sustained vowels).

not very different. For the standard deviation, controlled values overlap both normal and pathologic. Regarding speech material, dispersion in this case was much lower for vowels, as was expected. For the minimum, the controlled voice values are clearly lower, whereas normal and pathological values for sentences are similar. Values for pathological voice are different for vowels and sentences, which does not occur in the normal voice samples. For the F_0 maximum, there is a clear distinction between pathological and the two other voice types. Also, no significant differences between both types of speech material were observed.

Comparing the three situations (pathological voice, normal, and controlled), values were higher in the pathological voice, except for the F_0 minimum in sentences. Controlled voice type generally presents values below or similar to the normal voice. For the maximum and standard deviation parameters, in general terms, the informant with pathological voice was able to approach the values of the normal voice when using a controlled register. For the other two parameters, controlled voice presented values lower than normal.

Because of the atypical distribution of some values, non-parametric tests were used. Separate Kruskal-Wallis tests for each type of speech material confirmed the difference between all four F_0 parameters as significant. Due to the number of cases (i.e. 8), space limitations, and the need for further tests, results were not included. Mann-Whitney tests between each voice type were also performed. The results indicated the difference between all three types as generally significant except for sentences: Normal vs Controlled for F_0 maximum, Pathologic vs Controlled for F_0 std and Normal vs Pathologic for F_0 minimum. Maximum values of Controlled voice are not significantly different from Normal, and standard deviation also shows the difference between Controlled and Pathologic as non-significant. In the Controlled register the speaker is capable of approaching Normal voice maximum values but is unable to do the same for the dispersion of F_0 values.

Jitter

For this study, three different jitter-related parameters were analysed: local, Relative Average Perturbation (RAP), and a five-point Period Perturbation Quotient (PPQ5) (Baken and Orlikoff, 2000: 190). Local jitter parameter is the average absolute difference between consecutive periods, divided by the average period. MDVP gives 1.040% as a threshold for pathology. RAP is the average absolute difference between a period and the average of itself and its two neighbours, divided by the average period. MDVP gives .680% as a threshold for pathology. PPQ5 is the average absolute difference between a period and the average of it and its four closest neighbours, divided by the average period. MDVP calls this parameter PPQ, and gives .840% as a threshold for pathology.

Figure 2 presents 95% confidence intervals for each of the individual parameters considering the type of voice and the type of speech material.

The figures clearly show that with the pathological and normal voice types, sentences present much higher jitter values compared to sustained vowels. This is in agreement with the results for other languages where 'measured perturbation of spoken material' was 'much higher than that of a sustained vowel' (Baken and Orlikoff, 2000: 192).

The 95% confidence intervals for three parameters regarding vowels sustained by normal voice are completely below the threshold values for pathology. For the three different parameters, controlled voice values are close to the normal ones. Both are smaller than those for pathologic voice, being less affected by the local parameter.

The Kruskal-Wallis H-test, which considered voice type as a factor, was made separately for each type of speech material and confirmed the difference for all the three jitter

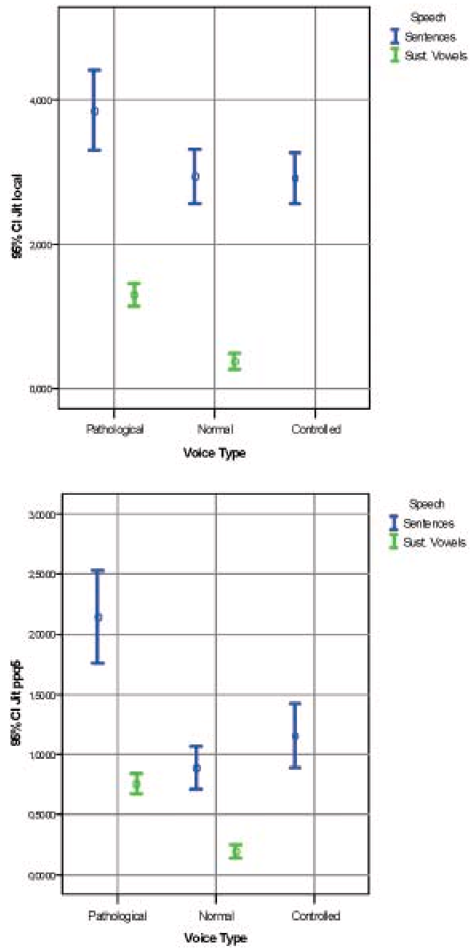

Figure 2. 95% Confidence Interval for three jitter parameters (local, RAP, and PPQ5), considering two factors: the type of voice (Pathological, Normal, and Controlled, from the left) and the type of speech material (sentences or sustained vowels).

parameters as significant. Results for sentences were: [$\chi^2(2)=9.286$, p=.01] for local, [$\chi^2(2)=55.222$, p<.001] for RAP, and [$\chi^2(2)=60.186$, p<.001] for PPQ5. Results for the sustained vowels were: [$\chi^2(1)=13.714$, p<.001] for local; [$\chi^2(1)=13.714$, p<.001] for RAP; and [$\chi^2(1)=13.725$, p<.001] for the remaining PPQ4.

For sentences, Mann-Whitney test results (which are not included) indicated the difference between Normal and Controlled for parameter local as non-significant, as the others were significantly different. That is, in its usual pathologic register, jitter is higher than in a normal voice; when controlling production, values are, at least for one of the jitter measures, not significantly different from normal voice values.

Shimmer

Two parameters related to shimmer, local and Amplitude Perturbation Quotient (APQ3), were analysed and are presented in Figure 3. The shimmer parameter local is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude. MDVP calls this parameter Shim, and gives 3.810% as a threshold for pathology. APQ3 is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude. Normal values for this parameter should not be above 3% (Behlau, Azevedo, and Pontes, 2004: 144).

Analysing the results in Figure 3, the confidence interval for the APQ3 of the normal voice vowels is completely inside the normal region (values below 3%); for local jitter the interval goes a little above the pathologic threshold—approaching 5%—and has a

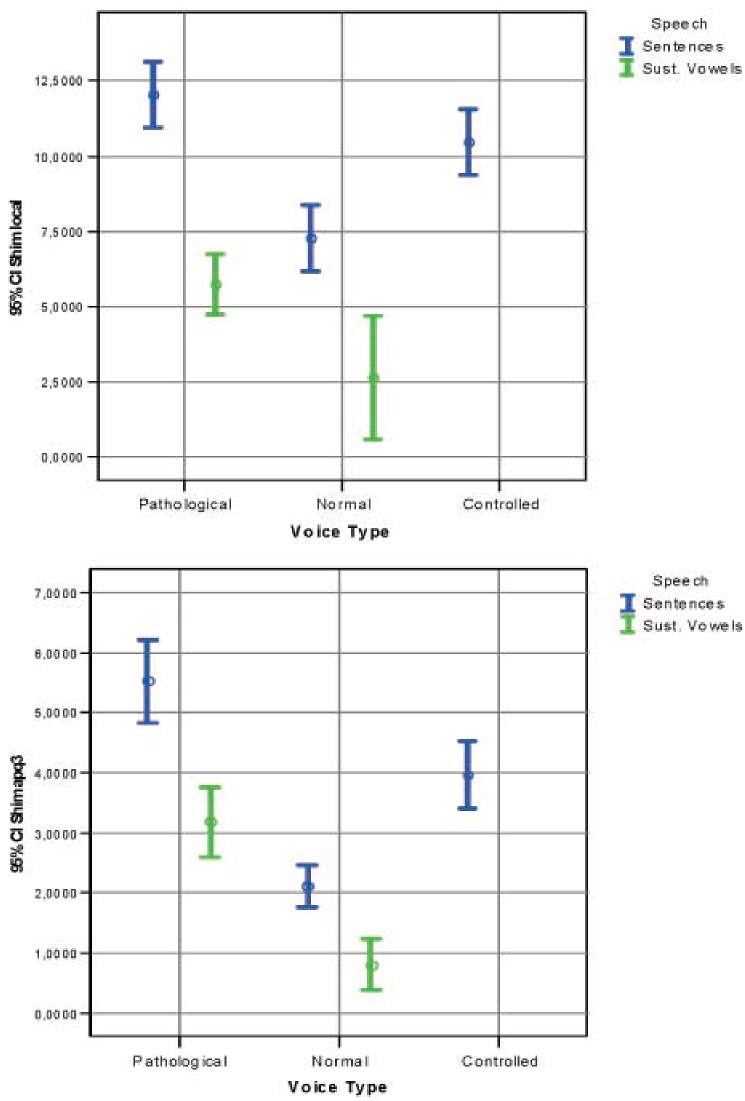

Figure 3. 95% Confidence Interval for two shimmer parameters (local and APQ3), considering two factors: the type of voice (Pathological, Normal, and Controlled, from the left) and the type of speech material (sentences or sustained vowels).

considerable dispersion. Despite this, the two means are inside the normal range. When comparing values for the three different voice types it is clear that the controlled voice presents much higher values than the normal, the values being close to or overlapping (in the case of the local) the pathologic voice. Even with the effort and concentration needed for the controlled voice, the psychogenic voice informant is not able to approach the normal voice shimmer values. For the APQ3, in spite of being able to decrease his average values from ~5.5% on his pathological production to near 4.0% on his controlled register, he cannot approach the normal voice register values, a little higher than 2.0% on the analyses. For the other parameter, which are more sensitive to other perturbations, the scenario is similar: the controlled average values are above 10.0% whereas the normal case presents an average of 2.5%.

Kruskal-Wallis tests were performed separately for each type of speech material resulting in: [$\chi^2(2)=58.421$, $p<.001$] for local parameter in sentences, [$\chi^2(2)=92.302$, $p<.001$] for the APQ3 in sentences, [$\chi^2(1)=8.595$, $p=.003$] for local parameter in sustained vowels, and [$\chi^2(1)=13.714$, $p<.001$] for the APQ3 parameter in sustained vowels. Results confirm the difference between voice types for the two shimmer parameters as significant.

For sentences, Mann-Whitney tests indicated the values of shimmer local and the APQ3 for all three voice types as significantly different, the controlled register values being higher than normal voice values. In the controlled register, the psychogenic voice speaker was not able to produce shimmer with a non-significant difference from the normal voice speaker.

Harmonic-to-noise ratio

Harmonic-to-Noise Ratio (HNR) results are presented in Figure 4. This parameter presents values that are very different and much higher for the two types of speech material analysed. For the normal voice, HNR values for vowels are above 25, placing them in the reported normal range.

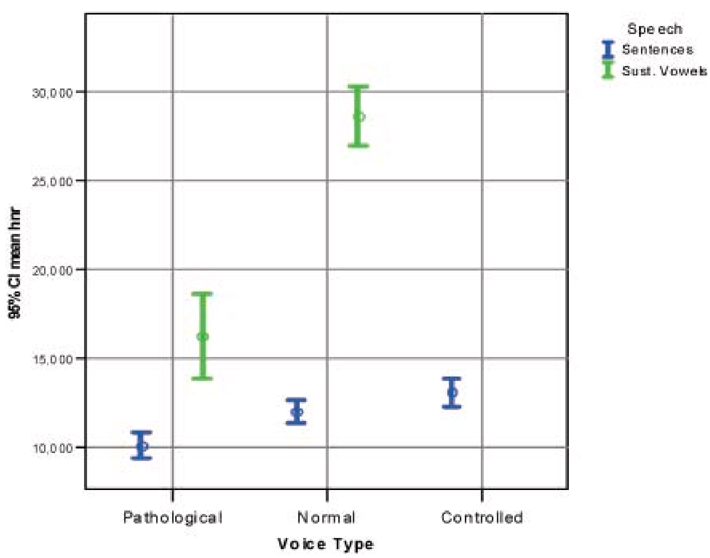

Figure 4. 95% Confidence Interval for HNR, considering two factors: the type of voice (Pathological, Normal, and Controlled, from the left) and the type of speech material (sentences or sustained vowels).

On a first examination of the confidence intervals, HNR does not seem very useful for making voice quality differentiations when using sentences. The Kruskal-Wallis non-parametric test revealed voice type effect on HNR as significant ($\chi^2(2)=26.935$, $p<.001$). The Mann-Whitney tests indicated HNR values for all three voice types as significantly different, with HNR for the Controlled voice being strangely better than the values obtained by the Normal speaker. For sustained vowels, normal voice HNR is much higher than in the pathologic case. The Kruskal-Wallis test, using only data on sustained vowels, confirmed the voice type effect on HNR as significant ($\chi^2(1)=13.714$, $p<.001$).

These results are in agreement with the findings of McAllister et al. (1996): 'pathological voices have a loss of harmonic components and elevated levels of noise components …'.

The hoarseness diagram

The analysis also used the Hoarseness Diagram (Fröhlich, Michaelis, Strube, and Kruse, 2000: 715), a 'new approach to the acoustic analysis of pathological voices in that it combines several acoustic measures'. This diagram separates two independent acoustic measures, the irregularity component (Irreg) and the noise component (Noise). For clinical application, the speakers' data are plotted in the diagram as ellipses representing the mean and standard deviation of the two factors. The programme made available by Michaelis (1998) was used and only sustained vowels were analysed, for which normal and pathologic values have been studied. Also the programme requirements (segments with more than 500 msec duration) made it unsuitable for sentences.

The results are presented in Figure 5. In the diagram, the three vowels for the psychogenic voice appear in the pathologic region according to Fröhlich et al. (2000).

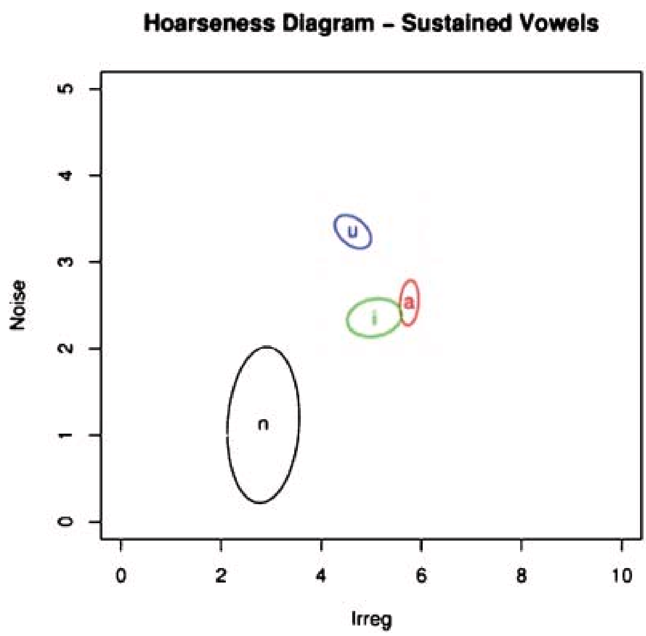

Figure 5. A hoarseness diagram for the two speakers' sustained vowels. Results are presented separately for each of the three vowels in the psychogenic voice while no separation has been made for the normal voice (labelled with an 'n').

Comparing the covered region with the values reported in Fröhlich et al. (2000) for different phonation mechanisms of subjects after resection of laryngeal carcinoma, the pathologic register can be described as 'pseudo-glottic phonation'. Psychogenic speaker vowels also present differences consistent with the previously mentioned study, with increased noise for [u] and higher irregularity for [a].

Conclusion

The present study investigated several different parameters related to voice quality (F_0, jitter, shimmer, and harmonicity) for a psychogenic voice-disordered speaker who was also capable of producing a more controlled type of voice. The study was based on two types of speech material that are commonly used, sustained cardinal vowels and short sentences.

Most parameters presented very different values for measures based on sentences and sustained vowels. By using both types of material one was able to make a comparative evaluation of the controlled voice type for where one considered existing counter indications for using sustained phonation. Also, in this way this study contributed to an improved characterization of European Portuguese normal voice in continuous speech, data for which is either scarce or missing.

Results indicate that the psychogenic voice-disordered speaker is capable of controlling certain parameters (e.g. F_0 maximum), but is incapable of correcting others such as shimmer or F_0 standard deviation. The option for analysing several voice quality related factors, and including different measures for each, revealed a good characterization of the three different voice types contemplated in this study, allowing one to establish differences and similarities between them.

This study was also able to apply the recently proposed Hoarseness Diagram to the case study with interesting and useful results. The Hoarseness Diagram showed that the informant with psychogenic dysphonia has a phonation close to the pseudo-glottal region (between the glottal region and the pseudo-glottic). Also the tendency of increased irregularity for [a] and increased noise for [u] was observed in the pathologic case.

It can also be concluded that the combined use of Praat and the Hoarseness Diagram provided detailed information as far as voice quality characterization is concerned.

Acknowledgements

This research was partially supported by a PhD scholarship from the University of Aveiro and by the Portuguese Science and Technology Foundation (FCT) FACC program. The authors thank the speakers for their help.

***Declaration of interest*:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech & voice*. San Diego: Singular Publishing Group, Inc.

Baker, J. (1998). Psychogenic dysphonia: peeling back the layers. *Journal of Voice*, *12*, 527–535.

Baker, J. (2002). Psychogenic voice disorders—heroes or hysterics? A brief overview with questions and discussion. *Logopedia Phoniatrica Vocology*, *27*, 84–91.

Behlau, M., Azevedo, R., & Pontes, P. (2004). Conceito de voz normal e classificação das disfonias. In M. Behlau (Ed.), *Voz - o livro do especialista* (pp. 53–172). Rio de Janeiro: Revinter.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Fröhlich, M., Michaelis, D., Strube, H., & Kruse, E. (2000). Acoustic voice analysis by means of the hoarseness diagram. *Journal of Speech Language and Hearing Research*, *43*, 706–720.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.

Guimarães, I., & Abberton, E. (2004). Fundamental frequency in speakers of Portuguese for different voice samples. *Journal of Voice*, *19*, 592–606.

Kreiman, J., & Gerratt, B. R. (2003). Jitter, shimmer, and noise in pathological voice quality perception. *VOQUAL'03*, Geneva, Switzerland.

McAllister, A., Sederholm, E., Ternström, S., & Sundberg, J. (1996). Perturbation and hoarseness: a pilot study of six children's voices. *Journal of Voice*, *10*, 252–261.

Michaelis, D. (1998). hd v1.0—Hoarseness diagram. Available online at: http://www.dpi.physik.uni-goettingen.de/~micha/hd.html. Accessed 30 May 2008.

Michaelis, D., Fröhlich, M., & Strube, H. (1998). Selection and combination of acoustic features for the description of pathologic voices. *Journal of Acoustical Society of America*, *103*, 1628–1640.

Rubin, J. S., & Greenberg, M. (2002). Psychogenic voice disorders in performers: a psychodynamic model. *Journal of Voice*, *16*, 544–548.