Abstract: In this study, the voice quality of 40 patients with a variety of clinical diagnoses was assessed, using the Universidade de Aveiro’s Voice Evaluation Protocol. A number of acoustic parameter were extracted including: median F0, mean F0, F0 std deviation, Jitter and Shimmer and HNR. Analysis of the correlation between corresponding parameters of the CAPE-V and GRBAS scales was made. The perceptual parameters grade (global in CAPE-V), roughness and breathiness were also compared individually with the objective acoustic parameters. 

Keywords: Voice, assessment, acoustic parameters

I. INTRODUCTION

Studies in the area of voice assessment still lack objectivity in the description and evaluation of many aspects of vocal quality. In therapeutic practice in Portugal, the assessment of vocal pathologies is not uniform, because each Speech and Language Therapist (SLT) uses individual and diverse assessment procedures to evaluate. The patient is assessed with scales that try to include the most important and appropriate parameters for each patient, based on existing protocols [1-11] which have normally been developed for languages other than European Portuguese (EP) and which, therefore, may not capture or account for language specific features of voice. Although voice quality measures using GRBAS [11] are normally considered to be language independent, the performance of this scale for assessment of EP has not, to date, been systematically tested. It is not the scale we expect to be language dependent, but the material used to carry out the test (e.g., if the text is English some EP vowels will be missing and some non-EP vowels might be included).

The aim of this project was to develop the first standardised and evaluated protocol for subjective voice assessment in EP: Universidade de Aveiro’s Voice Evaluation Protocol [12]. It is intended as a working tool for Speech and Language Therapists (SLTs), which brings together a range of essential information, thus preparing patients for a therapeutic intervention. SLTs involved in future studies will use the same evaluation instrument to acquire data that is comparable, thereby normalising the practice and nomenclature for this area of intervention and allowing also better inter-professional communication. A pilot study to test the reliability of the protocol including the analysis of inter-rater correlation, using a group of patients with various vocal disorders, has been reported elsewhere [12].


Perceptual analysis is based on sustained productions of /a, i, u, O/, CAPE-V sentences and reading the EP version of the “The North Wind and the Sun” passage, recently proposed as a standard text for “Advanced Voice Assessment” [13].

In our previous study [12], the reliability of the voice quality protocol was tested, using two independent raters, who evaluated, coincidentally, a group of patients who exhibited some change in voice quality.

The protocol parameters severity, roughness, breathiness, change of loudness (CAPE-V), grade, breathiness and strain (GRBAS), presented high reliability and were highly correlated (with good inter-rater agreement and a high value of correlation [12], similar results to assessments of other languages [1, 4, 14]). Values for the overall severity and grade were similar to those reported in the literature.

II. METHOD

In this study, the voice quality of 40 patients was assessed, with the Universidade de Aveiro’s Voice Evaluation Protocol [12] (with full ethics committee approval). These patients had been admitted to the Department of Otolaryngology of the Hospital de São João, Porto, Portugal. The sample included several clinical diagnoses: nodules, polyps, hypotonia of the vocal folds, Reinke’s oedema, musculo-skeletal
syndrome and dysfunctional dysphonia. The diagnosis for the sample was made by an experienced SLT and an Otorinolaringology consultant.

The speech tasks were recorded directly onto a PC, using Praat 5.1.10 [15] in a quiet environment.

The majority of recordings used a Sony F-V220 microphone and a SoundMAX Digital Audio internal soundcard (16 bits and 22050 Hz sampling frequency). A small number of recordings were made instead with an external sound card Edirol UA-25, set to 16 bits and 44100 Hz sampling frequency, and a Sennheiser e815S microphone. During the recordings, the microphone was held on a tripod placed 25-30 degrees to the left of the microphone. During the recordings, the microphone was held on a tripod placed 25-30 degrees to the left of the patient's mouth, at a distance of 30-40 cm.

Various acoustic parameters were extracted from the audio signal using Praat 5.1.10 [15] including: F0 Hz (median), F0 Hz (mean), F0 (std deviation), Jitter% (ppq5 – five-point frequency perturbation quotient equivalent to MDVP’s PPQ) and Shimmer% (apq11 – eleven-point amplitude perturbation quotient equivalent to MDVP’s APQ) and HNRdB (mean Harmonics-to-Noise Ratio).

A single sustained sample selected from each speaker’s productions of vowel /a/ was used to extract the acoustic parameters using the following criteria: one token “considered as perceptually closest to the subject’s natural voice” [16, p. 23] and produced with a “comfortable pitch and volume” [16, p. 23]; 100 consecutive cycles taken 200 ms after phonation onset [17, p. 1261] were used for analysis.

Analysis of the correlation between corresponding parameters of the CAPE-V and GRBAS scales was made. The perceptual parameters grade (global in CAPE-V), roughness and breathiness were also compared individually with the objective acoustic parameters.

Acoustic parameter data and scale evaluation scores were compared to find statistically significant differences between males and females using the Mann-Whitney U Test. Correlation analysis (Spearman correlation test) between the perceptual parameters in CAPE-V and GRBAS scales (global in CAPE-V and grade in GRBAS, roughness and breathiness) were also evaluated with the acoustic parameters: median F0, mean F0, standard deviation F0, jitter ppq5, shimmer apq11, and mean HNR. Finally, analysis of the correlation (Spearman correlation test) between corresponding parameters of the CAPE-V and GRBAS scales was made. All statistical analyses were conducted using SPSS 13.0 and a p value of less than 0.05 was considered significant. All data presented are given in mean ± standard deviation (S.D.)

III. RESULTS

In the tables below, the variable total represents the case when male and female data is combined.

Table 1 shows correlation analysis between CAPE-V and GRBAS scales. Statistical significances are found between the perceptual subscale grade from GRBAS and subcales global and roughness from CAPE-V, roughness in GRBAS and global in CAPE-V, and breathiness in GRBAS and in CAPE-V. The correlation values are good, ranging from 0.60 to 0.87, with the exception of the correlation value between the subcale roughness in GRBAS and the subscale global in CAPE-V for the total value. The results found for males alone can be ascribed to the smaller sample size.

Table 2 presents the acoustic parameter data and scale evaluation scores. The sample consists of 9 males (mean age 56.11±3.55) and 31 females (mean age 43.29±2.36). Statistically significance differences between males and females are found in age, median F0, mean F0, jitter ppq5, mean HNR and in several parameters in the CAPE-V scale (global and roughness) and in the GRBAS scale (grade and roughness). Such differences in the acoustic parameters are to be expected. The differences in scales can be explained due to the smaller number of males.

Table 3 shows the correlation analysis between the CAPE-V and GRBAS scales with the selected acoustic parameters. Statistical significances are found for median F0 and mean F0 with the perceptual subcales global and roughness for CAPE-V and grade and roughness for GRBAS. However, the correlations are weak, with values ranging from -0.38 to -0.60. No significant differences are found when we consider either only the male sample or the other acoustic parameters.

IV. DISCUSSION AND CONCLUSIONS

The two scales (GRBAS and CAPE-V) have been previously used simultaneously [4], with results showing a strong correlation between the two rating systems (Spearman’s correlation coefficients from ranging 0.89 to 0.95) for: GRBAS grade vs. CAPE-V global; GRBAS roughness vs. CAPE-V roughness; GRBAS breathiness vs. CAPE-V breathiness. Our results have also shown a good correlation except for roughness, because the term used in EP and Brazilian Portuguese for Grade is “grau de rouquidão”, which (see Table 1), appears to have been erroneously related to the CAPE-V EP term “rouquidão” (roughness).

This issue will be addressed in the future with a further validation of the EP version of CAPE-V that will use the procedures presented in [18], including the production of a CD-ROM with voice samples to be evaluated, voices used for training and samples of voices that represent specified grades of severity.

The lack of success in finding hypothesised correlations between acoustic and perceptual measures have long been known [20, pp. 75-80], and do not seem to be related to language specific characteristics, as our results have shown, even when we limit our set of acoustic and perceptual parameters as in [19]. Different factors contribute to a failure to find consistent
correlations: deficiencies in the theoretical framework; incoherencies in the definitions of parameters; limitations in estimation techniques [20, pp. 75-80].

Ongoing and future work will extend this study to a larger number of patients, especially by increasing the number of males analysed, so the protocol can be used with more confidence. The pilot protocol presented in [12] has a large number of parameters and it is intended to evaluate these further to derive, if possible, a best set of parameters for EP voice quality assessment.

V. ACKNOWLEDGEMENTS

This work was supported by Fundação para a Ciência e a Tecnologia, Portugal (PTDC/SAU-BEB/67384/2006).

REFERENCES


Table 1. Correlation analysis between CAPE-V and GRBAS scales.

<table>
<thead>
<tr>
<th></th>
<th>CAPE-V</th>
<th>GRBAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade</td>
<td>Roughness</td>
</tr>
<tr>
<td>Global</td>
<td>0.00</td>
<td>0.86*</td>
</tr>
<tr>
<td>Female</td>
<td>0.65*</td>
<td>0.20</td>
</tr>
<tr>
<td>Total</td>
<td>0.60*</td>
<td>0.42*</td>
</tr>
<tr>
<td>Roughness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.23</td>
<td>-0.14</td>
</tr>
<tr>
<td>Female</td>
<td>0.79*</td>
<td>0.14</td>
</tr>
<tr>
<td>Total</td>
<td>0.75*</td>
<td>0.26</td>
</tr>
<tr>
<td>Breathness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.18</td>
<td>0.67*</td>
</tr>
<tr>
<td>Female</td>
<td>0.09</td>
<td>-0.10</td>
</tr>
<tr>
<td>Total</td>
<td>0.04</td>
<td>0.08</td>
</tr>
</tbody>
</table>

* Spearman correlation test, statistical significance (p<0.05).
Table 2. Acoustic parameters and scale evaluation scores (Gender: Male=9, Female=31; Age: Male=56.11±3.55*, Female=43.29±2.36*, Total=46.18±13.62).

<table>
<thead>
<tr>
<th>Acoustic parameters</th>
<th>Mean ± S.D.</th>
<th>Scales</th>
<th>Mean ± S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median F0 (Hz)</td>
<td></td>
<td>CAPE-V</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>123.88±7.59*</td>
<td>Global</td>
<td>0.62±0.07*</td>
</tr>
<tr>
<td>Female</td>
<td>184.93±6.11*</td>
<td>Female</td>
<td>0.41±0.04*</td>
</tr>
<tr>
<td>Total</td>
<td>171.19±40.76</td>
<td>Total</td>
<td>0.46±0.21</td>
</tr>
<tr>
<td>Mean F0 (Hz)</td>
<td></td>
<td>Roughness</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>124.09±7.62*</td>
<td>Male</td>
<td>0.63±0.04*</td>
</tr>
<tr>
<td>Female</td>
<td>184.86±6.10*</td>
<td>Female</td>
<td>0.42±0.05*</td>
</tr>
<tr>
<td>Total</td>
<td>171.19±40.65</td>
<td>Total</td>
<td>0.46±0.24</td>
</tr>
<tr>
<td>Std Dev F0 (Hz)</td>
<td></td>
<td>Breathiness</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.50±0.42</td>
<td>Male</td>
<td>0.37±0.10</td>
</tr>
<tr>
<td>Female</td>
<td>2.63±0.51</td>
<td>Female</td>
<td>0.36±0.04</td>
</tr>
<tr>
<td>Total</td>
<td>2.60±2.57</td>
<td>Total</td>
<td>0.36±0.25</td>
</tr>
<tr>
<td>Jitter ppq5 (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.85±0.28*</td>
<td>Male</td>
<td>2.33±0.17*</td>
</tr>
<tr>
<td>Female</td>
<td>0.34±0.04*</td>
<td>Female</td>
<td>1.55±0.15*</td>
</tr>
<tr>
<td>Total</td>
<td>0.45±0.49</td>
<td>Total</td>
<td>1.73±0.85</td>
</tr>
<tr>
<td>Shimmer apq11 (%)</td>
<td></td>
<td>Roughness</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3.68±0.56</td>
<td>Male</td>
<td>0.89±0.31*</td>
</tr>
<tr>
<td>Female</td>
<td>2.99±0.48</td>
<td>Female</td>
<td>0.32±0.13*</td>
</tr>
<tr>
<td>Total</td>
<td>3.15±2.46</td>
<td>Total</td>
<td>0.45±0.81</td>
</tr>
<tr>
<td>Mean HNR (dB)</td>
<td></td>
<td>Breathiness</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>12.28±1.56*</td>
<td>Male</td>
<td>1.22±0.32</td>
</tr>
<tr>
<td>Female</td>
<td>17.88±0.84*</td>
<td>Female</td>
<td>1.22±0.14</td>
</tr>
<tr>
<td>Total</td>
<td>16.63±5.20</td>
<td>Total</td>
<td>1.23±0.83</td>
</tr>
</tbody>
</table>

* Mann-Whitney test, statistical significance (p<0.05).

Table 3. Correlation analysis between the CAPE-V and GRBAS scales with the selected acoustic parameters: median F0, mean F0, standard deviation F0, jitter ppq5, shimmer apq11, and mean HNR.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Median F0 (Hz)</th>
<th>Mean F0 (Hz)</th>
<th>Std Dev F0 (Hz)</th>
<th>Jitter ppq5 (%)</th>
<th>Shimmer apq11 (%)</th>
<th>Mean HNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPE-V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.08</td>
<td>-0.08</td>
<td>0.43</td>
<td>0.34</td>
<td>-0.29</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.41*</td>
<td>-0.39*</td>
<td>-0.09</td>
<td>0.00</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-0.49*</td>
<td>-0.48*</td>
<td>0.00</td>
<td>0.13</td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td>Roughness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.13</td>
<td>0.13</td>
<td>-0.06</td>
<td>0.26</td>
<td>-0.16</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.55*</td>
<td>-0.53*</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-0.60*</td>
<td>-0.58*</td>
<td>0.03</td>
<td>0.12</td>
<td>-0.27</td>
<td></td>
</tr>
<tr>
<td>Breathiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.00</td>
<td>0.00</td>
<td>0.38</td>
<td>0.12</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.17</td>
<td>0.20</td>
<td>0.04</td>
<td>0.11</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.12</td>
<td>0.14</td>
<td>0.12</td>
<td>0.10</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>GRBAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.37</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.52*</td>
<td>-0.48*</td>
<td>0.10</td>
<td>-0.06</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-0.58*</td>
<td>-0.56*</td>
<td>0.09</td>
<td>0.02</td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td>Roughness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.26</td>
<td>-0.26</td>
<td>0.13</td>
<td>-0.16</td>
<td>-0.19</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.22</td>
<td>-0.24</td>
<td>0.19</td>
<td>0.19</td>
<td>-0.30</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-0.38*</td>
<td>-0.39*</td>
<td>0.05</td>
<td>0.19</td>
<td>-0.30</td>
<td></td>
</tr>
<tr>
<td>Breathiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.20</td>
<td>0.20</td>
<td>0.03</td>
<td>0.24</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.20</td>
<td>0.24</td>
<td>-0.07</td>
<td>0.24</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.18</td>
<td>0.20</td>
<td>-0.05</td>
<td>0.13</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

* Spearman correlation test, statistical significance (p<0.05).