Acoustic and Perceptual Indicators of Normal and Pathological Voice

J. Muñoz  E. Mendoza  M.D. Fresneda  G. Carballo  P. López

Department of Personality, Evaluation and Psychological Treatment, University of Granada, Granada, Spain

Key Words
Vocal quality - Acoustic correlates

Abstract
The aim of this study was to obtain acoustic correlates to vocal quality of a group of men and women with and without voice disorders, based on the evaluation of normal, hoarse and rough voice qualities. Vocal quality was related to the following acoustic features: frequency perturbation measures (JITA, RAP, and SPPO), amplitude perturbation (SAPQ and VAM), soft phonation index (SPI), and fundamental frequency tremor intensity (FTIRI). While these measures presented normal values for normal voice, hoarseness showed some deviations in perturbation frequency variables and very high SPI values, while rough voice showed deviations in all the measures. Qualities of female voices were perceived as normal, breathy and hoarse, but the acoustic correlates of these qualities were less conclusive.

Introduction
Clinicians’ evaluations of patients’ vocal quality are fundamental to clinical considerations as well as to validating criteria of voice-measuring instruments, whether aerodynamic, acoustic, etc. [1–4]. Conflicting studies have recently appeared concerning the relation between certain vocal qualities (e.g., hoarseness, breathiness and roughness) and the various acoustic measures. This inconsistency derives from the application of different methodologies in research projects, and also from factors such as judges’ experience, voice sample, type and severity of pathology, and different software used in acoustic analyses. Such factors make it difficult to compare
studies and render external validation of data impossible.

Inconsistency is evident in many studies; for example, evaluations of hoarseness have been variously related to measures of vocal jitter [3, 5, 6], shimmer [7], noise level [8–12] and hoarseness diagrams. This last is based on four acoustic measures, three of which assess aspects of signal irregularity and the fourth the additive noise content [13, 14].

Perceived roughness has been linked with high fundamental frequency (F0) perturbation [15–17], and contradictorily, with the absence of such perturbation [18–20]. It has been shown that perception of this vocal quality is affected by the relation between frequency perturbation (jitter) and F0; that is to say, in cases where there are two voices or stimuli with high perturbation frequency values, the one with lower F0 will be perceived as being rougher [1, 3, 18, 21, 22]. Finally, in other studies, roughness has been related to high spectral noise ratings [5, 23–27], and a good correlation between roughness and low-frequency noise components has been shown [28].

The vocal quality ‘breathy’ is known to be a very common symptom in both organic and functional disorders [23, 29–31]. Previous work has shown that perception of this quality is affected by the amplitude of the first harmonic [32], aspiration noise [22, 23, 33], spectral tilt [34] and degree of signal periodicity [2, 35]. Breathiness has also been related to pitch amplitude and harmonic-to-noise ratio as the best predictions for breathy voice quality [5].

Some studies have used multidimensional scales with the aim of establishing relevant dimensions for perception of voice quality [31, 36–38]. These studies have incorporated different stimuli and measures for the interpretation of results, showing F0 to be a consistent acoustic correlate. Similarly, Kempster et al. [39] using two multidimensional scale procedures, found that listeners use three-dimensional information when evaluating similarity of dysphonic voices: intensity, F0 and perturbation.

Bearing in mind the work carried out in this field and the diversity of findings, our objective in this study has been to determine whether some kind of correspondence can be established between normal and pathological voice quality and certain acoustic correlates; in other words, whether a voice judged to be of a certain quality according to criteria of skilled listeners can be related to a specific acoustic profile. In order to obtain valid results careful consideration was given to methodology.

We took as our starting point a study by Muñoz et al. [40], in which 5 judges with experience in the field of vocal rehabilitation (between 2 and over 8 years) evaluate voice quality of Spanish speakers using the ‘Buffalo III Voice Screening Profile’ [41]. Reliability was measured by means of the Cronbach coefficient alpha and presented values of between 0.64 and 0.90.

A further important aspect of the study is that listeners’ perceptions of different voice qualities were based on the same segment of the sustained vowel /a/, which we had evaluated acoustically to achieve higher refinement acoustic correlates to different voice qualities.

**Methods**

**Subjects**

The study was carried out with the participation of 69 subjects (37 males and 32 females) selected at random and matched by gender from a database of 233 subjects (47 males without vocal pathologies, 36 males with vocal pathologies, 49 females without vocal pathologies and 101 females with vocal pathologies) with an age range between 20 and 50 years. All diagnoses were made in accordance with indirect laryngoscope
criteria. Subject distribution was as follows: 14 subjects (7 males and 7 females) without problems or history of vocal alterations, 14 subjects (7 males and 7 females) with laryngeal mucous lesions (nodules, polyps, etc.), 14 subjects (7 males and 7 females) with laryngeal motility lesions or alterations (paralysis, paresis, etc.), 13 subjects (9 males and 4 females) with functional dysphonia, and 14 subjects (7 males and 7 females) who had received laryngeal micorsurgery. The group of 13 subjects did not present a defined laryngeal pathology, only slight swelling of the vocal cords. This group had an unequal number of subjects as there were no more subjects available in our database. The last group had presented laryngeal mucous lesions prior to surgery.

All subjects presented only one vocal pathology and had not received previous treatment, except for laryngeal micorsurgery in some cases.

**Equipment**

Recording was performed using an AKG D 222 ED flat-response microphone, and a Sony 77 ES digital audiotape recorder. Voice samples were analyzed by Multidimensional Voice Program (MDVP; model 4300, Kay Elemtrics Corp.), using Computerized Speech Lab (CSL) hardware for signal acquisition, analysis and interpretation of data. The MDVP uses analog/digital hardware to sample speech at 50 kHz prior to analysis of sustained phonation. Speech segment for acoustic analysis in our study was the sustained vowel /a/. This was the same listening material as that presented to the 5 female listeners in the earlier study on reliability of different voice qualities [40]. The segment was presented three times consecutively through Fonestar FA-950 headphones.

**Procedure**

Voice Samples

Voice samples were recorded in the Voice Laboratory of the Psychology Department, University of Granada (Spain). Environmental noise levels were kept to a minimum. All recordings were made in the morning, and digital audiotape intensity was maintained between 20 and 30 dB [33]. Each subject was asked to sustain the vowel /a/ at a comfortable pitch and loudness, for a duration of about 5–6 s. The microphone was positioned at 15 cm from the subject's mouth. Finally, analysis of the signal proceeding from the recorder and stored on digital tapes was captured by the MDVP. The initial and final part of the vowel were eliminated, capturing 2.750 s for analysis (the central part of the vowel).

**Acoustic Measures**

A wide range of acoustic measures were analyzed so as not to limit the number of acoustic measures that may influence perceptual evaluation of voice quality.

- F0 information measurements:
  1. Average F0 (Hz); (2) standard deviation (SD, Hz) of the F0 within the analyzed voice sample.

- Short- and long-term frequency perturbation measurements:
  1. Absolute jitter (JITA): evaluation of the period-to-period variability of the pitch period within the analyzed voice sample. Voice break areas are excluded.
  2. Relative average perturbation (RAP, %): evaluation of the period-to-period variability of the pitch within the analyzed voice sample with smoothing factor of 3 periods. Voice break areas are excluded.
  3. Smoothed pitch period perturbation quotient (SPPQ, %): evaluation of the short- or long-term variability of the pitch period within the analyzed voice sample. The factory setup for the smoothing factor is 55 periods. Voice break areas are excluded.
  4. Coefficient of F0 variation (VF0, %): this represents the relative standard deviation of the F0. It reflects in general the variation of F0 (short-term to long-term) within the analyzed voice sample. Voice break areas are excluded.

- Short- and long-term amplitude perturbation measurements:
  1. Shimmer in decibel: evaluation of the period-to-period (very short-term) variability of the peak-to-peak amplitude within the analyzed voice sample. Voice break areas are excluded.
  2. Amplitude perturbation quotient (APQ, %): evaluation of the period-to-period variability of the peak-to-peak amplitude within the analyzed voice sample at smoothing of 11 periods. Voice break areas are excluded.
  3. Smoothed amplitude perturbation quotient (SAPQ, %): evaluation of the short- or long-term variability of the peak-to-peak amplitude within the analyzed voice sample. The factory setup for the smoothing factor is 55 periods. Voice break areas are excluded.
  4. Coefficient amplitude variation (VAM, %): relative standard deviation of the peak-to-peak amplitude. This reflects in general the peak-to-peak amplitude variations (short- to long-term) within the analyzed voice sample. Voice break areas are excluded.

- Noise-related measurements:
  1. Noise-to-harmonic ratio (%); (2) voice turbulence index (%); (3) soft phonation index (SPI, %).

**Tremor measurements**
(1) F₀ tremor intensity index (FTRI, %); (2) amplitude tremor intensity (%). For more detailed information on each variable, see the MDVP manual [42].

Perceptual Evaluation

Perceptual evaluations in the earlier study [40] used the same segment of the sustained vowel /a/ which we had previously analyzed acoustically. Evaluations of voice quality were made using the ‘Buffalo III Voice Screening Profile’ [41], which consists of 10 ratings: breathy, rough, hoarse (evaluating laryngeal tone); high pitch, low pitch, high loudness, low loudness, hypernasal resonance, hyponasal resonance and overall voice rating. These ratings are assessed on a 5-point scale (1 = normal; 2 = mild; 3 = moderate; 4 = severe; 5 = very severe).

Results

Classification of Subjects with and without Voice Disorders according to Perceptual Vocal Qualities

The aim here was to classify all the participants in separate groups according to their voice quality. Those appearing in the same group should have similar voice qualities, and these should be distinct from the other groups. To achieve this, we performed two cluster analyses [43, 44], one for the group of males and the other for the female group, with three groups per analysis.

Cluster Analysis of Female Samples

Figure 1 shows mean values for all perceptual variables corresponding to clusters 1, 2 and 3 for the female group. Standard deviations and F ratios corresponding to these variables appear in table 1.

Cluster 1 (normal voices) includes 16 cases with the following characteristics: all women in the sample with no vocal disorder (7 women); 3 (out of a total of 4) presenting functional dysphonia, and 6 women belonging to other groups in the study (laryngeal microsurgery, mucous lesions and motility lesions). As may be seen in figure 1, cluster 1 shows very homogeneous values in the various perceptual attributes, and generally speaking, values are lower than in the other clusters. Bearing in mind the characteristics of the sample and results obtained, we conclude that the vocal quality defining the female in this cluster is ‘normal quality’.

Cluster 2 (hoarse voices) comprised 9 women presenting various vocal disorders (microsurgery, mucous lesions and laryngeal motility lesions). As may be seen in figure 1, the highest values in laryngeal tone variables correspond to the quality ‘hoarse’. The highest or most prominent ratings for other attributes are: low pitch, low loudness, hypnasal resonance and overall voice rating. Given the correspondence shown between hoarse quality and these prominent attributes, we have classified the women in this cluster as ‘hoarse vocal quality’.

Cluster 3 (breathy voices) comprised 7 women with various vocal disorders: microsurgery, mucous lesions, motility lesions and 1 subject with functional dysphonia. Figure 1 shows the highest values for the quality ‘breathy’ and for the attributes high pitch, low loudness, hypnasal resonance and overall voice rating. Again, on the basis of the correspondence shown between breathiness and the attributes mentioned, women in this cluster are defined as ‘breathy vocal quality’.

Cluster Analysis of Male Samples

Figure 2 shows mean values for all perceptual variables corresponding to clusters 1, 2 and 3 for the male group. Standard deviations and F ratios for these variables are shown in table 2.

Cluster 1 (normal voices) includes 17 cases with the following characteristics: all males from the sample with no vocal disorders (7 men); 6 (out of a total of 9) men with functional dysphonia, and 4 men belonging to subject
groups presenting mucous lesions, lesions in motility or laryngeal musculature. As may be seen from figure 2, this cluster shows very homogeneous values for the different perceptual attributes. Furthermore, as was the case with the female group, the males in this cluster show lower values than those in the other clusters. Bearing in mind this consideration and the characteristics of the subjects concerned, we conclude that the vocal quality defining this group is ‘normal quality’.

Included in cluster 2 (hoarse voices) are 14 males presenting various vocal disorders (microsurgery, mucous lesions, laryngeal motility

---

Table 1. F ratios and standard deviations (SD) of perceptual variables in females’ clusters 1 (normal), 2 (hoarse) and 3 (breathy)

<table>
<thead>
<tr>
<th>Variable</th>
<th>F ratios</th>
<th>Cluster 1 (normal)</th>
<th>Cluster 2 (hoarse)</th>
<th>Cluster 3 (breathy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(2, 29)</td>
<td>SD</td>
<td>SD</td>
<td>SD</td>
</tr>
<tr>
<td>Breathy</td>
<td>26.50</td>
<td>0.57</td>
<td>0.74</td>
<td>0.64</td>
</tr>
<tr>
<td>Rough</td>
<td>5.48</td>
<td>0.23</td>
<td>0.53</td>
<td>0.19</td>
</tr>
<tr>
<td>Hoarse</td>
<td>17.54</td>
<td>0.15</td>
<td>0.91</td>
<td>1.11</td>
</tr>
<tr>
<td>High pitch</td>
<td>10.49</td>
<td>0.47</td>
<td>0.5</td>
<td>0.83</td>
</tr>
<tr>
<td>Low pitch</td>
<td>42.70</td>
<td>0.18</td>
<td>0.55</td>
<td>0.43</td>
</tr>
<tr>
<td>High loudness</td>
<td>2.09</td>
<td>0.42</td>
<td>0.36</td>
<td>0.15</td>
</tr>
<tr>
<td>Low loudness</td>
<td>44.51</td>
<td>0.36</td>
<td>0.75</td>
<td>0.94</td>
</tr>
<tr>
<td>Hypernasal resonance</td>
<td>3.18</td>
<td>0.4</td>
<td>0.57</td>
<td>0.31</td>
</tr>
<tr>
<td>Hypophasal resonance</td>
<td>19.49</td>
<td>0.23</td>
<td>0.36</td>
<td>0.54</td>
</tr>
<tr>
<td>Overall voice rating</td>
<td>41.51</td>
<td>0.43</td>
<td>0.61</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Fig. 2. Normal, hoarse and rough vocal qualities in males’ clusters 1, 2, and 3.

Table 2. F ratios and standard deviations (SD) of perceptual variables in males’ clusters 1 (normal), 2 (hoarse) and 3 (rough)

<table>
<thead>
<tr>
<th>Variable</th>
<th>F ratios $F_{(2, 34)}$</th>
<th>Cluster 1 (normal) SD</th>
<th>Cluster 2 (hoarse) SD</th>
<th>Cluster 3 (rough) SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breathy</td>
<td>8.58</td>
<td>0.23</td>
<td>0.62</td>
<td>0</td>
</tr>
<tr>
<td>Rough</td>
<td>22.18</td>
<td>0.39</td>
<td>0.39</td>
<td>0.65</td>
</tr>
<tr>
<td>Hoarse</td>
<td>26.46</td>
<td>0.56</td>
<td>0.8</td>
<td>0.66</td>
</tr>
<tr>
<td>High pitch</td>
<td>0.52</td>
<td>0.37</td>
<td>0.47</td>
<td>0.24</td>
</tr>
<tr>
<td>Low pitch</td>
<td>20.86</td>
<td>0.47</td>
<td>0.71</td>
<td>1</td>
</tr>
<tr>
<td>High loudness</td>
<td>38.38</td>
<td>0.3</td>
<td>0.37</td>
<td>0.47</td>
</tr>
<tr>
<td>Low loudness</td>
<td>35.02</td>
<td>0.35</td>
<td>0.77</td>
<td>0.52</td>
</tr>
<tr>
<td>Hypernasal resonance</td>
<td>1.29</td>
<td>0.41</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Hyponasal resonance</td>
<td>14.37</td>
<td>0.23</td>
<td>0.52</td>
<td>0.4</td>
</tr>
<tr>
<td>Overall voice rating</td>
<td>42.03</td>
<td>0.41</td>
<td>0.65</td>
<td>0.47</td>
</tr>
</tbody>
</table>

lesions and 1 subject with functional dysphonia). As may be seen from figure 2, the highest value in perceptions of laryngeal tone (breathy, rough, hoarse) corresponds to the quality ‘hoarse’. High values are also observed in the following attributes: low pitch, low loudness, hypernasal resonance and overall voice rating. On the basis of the correspondence between hoarseness and these marked attributes, the vocal quality of the males in this cluster was defined as ‘hoarse quality’.

Cluster 3 (rough voices) comprises 7 males with various vocal disorders (microsurgery, mucous lesions, motility lesions and 2 sub-
<table>
<thead>
<tr>
<th>Variable</th>
<th>Normal mean</th>
<th>Normal SD</th>
<th>Hoarse mean</th>
<th>Hoarse SD</th>
<th>Rough mean</th>
<th>Rough SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>JITA</td>
<td>60.36</td>
<td>26.83</td>
<td>105.16</td>
<td>38.99</td>
<td>365.98</td>
<td>256.17</td>
</tr>
<tr>
<td>RAP</td>
<td>0.48</td>
<td>0.23</td>
<td>0.67</td>
<td>0.21</td>
<td>2.41</td>
<td>1.73</td>
</tr>
<tr>
<td>SPPQ</td>
<td>0.67</td>
<td>0.21</td>
<td>0.86</td>
<td>0.27</td>
<td>3.95</td>
<td>3.94</td>
</tr>
<tr>
<td>SAPQ</td>
<td>4.95</td>
<td>1.36</td>
<td>5.27</td>
<td>1.89</td>
<td>9.16</td>
<td>3.43</td>
</tr>
<tr>
<td>VAM</td>
<td>15.55</td>
<td>5.21</td>
<td>11.28</td>
<td>3.48</td>
<td>19.66</td>
<td>5.75</td>
</tr>
<tr>
<td>SPI</td>
<td>11.52</td>
<td>6.25</td>
<td>15.81</td>
<td>8.46</td>
<td>13.89</td>
<td>5.59</td>
</tr>
<tr>
<td>FTRI</td>
<td>0.36</td>
<td>0.13</td>
<td>0.46</td>
<td>0.32</td>
<td>1.28</td>
<td>1.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>F0 mean</th>
<th>F0 SD</th>
<th>F0 mean</th>
<th>F0 SD</th>
<th>F0 mean</th>
<th>F0 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>208.72</td>
<td>26.82</td>
<td>204.34</td>
<td>37.41</td>
<td>151.69</td>
<td>56.77</td>
</tr>
<tr>
<td>RAP</td>
<td>0.65</td>
<td>0.22</td>
<td>3.13</td>
<td>2.37</td>
<td>2.41</td>
<td>1.24</td>
</tr>
<tr>
<td>SPPQ</td>
<td>0.71</td>
<td>0.21</td>
<td>4.48</td>
<td>4.24</td>
<td>6.57</td>
<td>9.34</td>
</tr>
<tr>
<td>V0</td>
<td>1.18</td>
<td>0.36</td>
<td>8.65</td>
<td>9.24</td>
<td>8.88</td>
<td>10.88</td>
</tr>
<tr>
<td>FTRI</td>
<td>0.34</td>
<td>0.24</td>
<td>1.51</td>
<td>1.67</td>
<td>2.76</td>
<td>5.23</td>
</tr>
</tbody>
</table>

**Table 3.** Discriminant analysis means and standard deviations (SD) of acoustic variables: males

**Table 4.** Discriminant analysis means and standard deviations (SD) of acoustic variables: females

Projects with unspecified or functional lesions. Figure 2 shows the highest ratings corresponding to the quality ‘rough’. For other perceptual attributes the highest values are low pitch, loud loudness, hypernasal resonance and overall voice rating. On the basis of the correspondence between roughness and these marked attributes, the vocal quality of the males in this cluster was defined as ‘rough quality’.

*Acoustic Correlates of Voice Quality of Male and Female with and without Vocal Disorders*

The aim here was to determine which acoustic measures are related to the three voice qualities obtained in the clusters earlier. Accordingly, we assigned a code for the cluster number to which each subject belonged and performed two discriminant analyses (one for each sex group), using the MDVP acoustic measures as dependent variables. In both analyses we adopted F to enter value = 3 and F removal value = 2.99, to obtain the variable or set of variables from the model. Tables 3 and 4 show means and standard deviations for males and females, respectively.

Using these criteria, measures obtained from the model for the male group were as follows: perturbation frequency measures (JITA, RAP and SPPQ), amplitude perturbation measures (SAPQ and VAM), SPI and FTRI.
Table 5. Means and standard deviations (SD) of perceptual variables in males and females without vocal disorders [45]

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>variable</td>
<td>mean</td>
<td>SD</td>
<td>F0</td>
<td>0.39</td>
<td>0.17</td>
<td>RAP</td>
<td>0.55</td>
<td>0.26</td>
<td>SPPQ</td>
<td>0.69</td>
</tr>
<tr>
<td>JITA</td>
<td>56.74</td>
<td>28.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAP</td>
<td>0.69</td>
<td>2.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAPQ</td>
<td>16.73</td>
<td>8.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPI</td>
<td>10.42</td>
<td>6.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTRI</td>
<td>0.49</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The classification matrix shows 86% correct classifications for the total male group, with a breakdown as follows: normal voice quality males, 88% correct; hoarse quality group, 82% correct; rough quality group, 89% correct.

In the discriminant analysis for the female group, the measures obtained from the model were: frequency perturbation measures (RAP, SPPQ, VF0), FTRI and F0. The classification matrix shows 85% correct classifications for the total female group. Percentage of correct classifications for normal, hoarse and breathy voice were 94, 78 and 71%, respectively.

Since the MDVP normative data were derived from a small number of subjects, we compared measures from the two discriminant analyses with normative data (table 5) obtained from the sample of 47 males and 49 females without vocal disorders [45].

Discussion

Perception of Vocal Quality in Subjects with and without Vocal Disorders

Our results show that the judges grouped the voice quality types in a different way for female and male subjects.

Broadly speaking, we may say that laryngeal tone qualities used to evaluate the female subjects were ‘normal’, ‘hoarse’ and ‘breathy’ (clusters 1, 2 and 3, respectively). The group of females rated as normal included all the women in the sample presenting no vocal problems, 3 out of 4 who belonged to the functional dysphonia group, and some presenting vocal disorders. The fact that the judges included some subjects with vocal disorders seems to indicate that the existence of a laryngeal pathology does not necessarily aurally distinguish a pathological voice from a normal one [46]. Some talkers with no appreciable pathology in the larynx may have a hoarse voice, presenting considerable variations in acoustic measurements such as pitch perturbation [46]. By the same token, there are also patients with certain laryngeal pathologies whose voice quality is not affected until in advanced state. The key factor in such considerations is the way in which physiological changes occasioned by vocal pathologies affect the mechanical properties of the larynx. In summary, we may say that the judges in our study placed the female voice at a point along a scale of three voice qualities, guiding their evaluations principally by vocal pitch and volume attributes, with resonance characteristics in second place. As a matter of fact, qualities of pitch and volume have also been analyzed in studies focusing on the multidimensional nature of voice quality perception [6, 37, 47, 48].
We can see from figure 1 that, as in other studies, the judges associated breathy voices with high pitch [49–51] and low loudness [4, 52], this last attribute showing higher values than for hoarseness. Hoarse voice (cluster 2) was characterized by low pitch, which corresponds with the fact that in clinical practice a voice with low pitch is likely to be judged as more severely dysphonic [39, 52, 53].

We should add that breathiness and hoarseness in females were characterized by lack of nasal resonance (hyponasality), as severely dysphonic voices or with more disagreeable perceptual attributes than the normal voices. Attributes of tone, volume and resonance associated with breathy and hoarse voice are related to a fundamental aspect of the voice rehabilitation process, since speech-language pathologists normally concentrate on the perceptual attributes of pitch and volume, as well as molding resonance cavities, to achieve an audible improvement in the vocal quality of their patients [54–56].

Turning to the male voices, the judges based their ratings on the laryngeal tone qualities normal, hoarse and rough (clusters 1, 2 and 3, respectively). In this case the breathy attribute was not used, since it has been principally associated with female voices [2, 33, 57]. Included in the ‘normal’ cluster 1 are all the normal subjects, together with most of the functional dysphonia subjects and some with vocal disorders. This again supports the argument by Koike et al. [46] that the presence of a lesion is not necessarily related to the perception of vocal distortion. Once again, the judges placed the male voices at a point along a scale of three laryngeal tone qualities, guiding their evaluations chiefly by attributes of pitch (low-high) and loudness (low-high), with resonance characteristics in second place (hyper-hyponasal) [39, 58]. It may also be observed that the judges related both roughness and hoarseness to low pitch, the distinguishing factor being perceived intensity. This accords with the finding of a previous study [59] that male rough voice basically correlates with high intensity, while hoarseness correlates with low pitch and low voice intensity. On the basis of these data, we may conclude that the most prominent perceptual characteristic of rough voice is higher intensity, while that of hoarseness is low pitch and low intensity. Additionally, both vocal qualities are perceived as lacking nasal resonance and are evaluated as more severely dysphonic or disordered than normal-rated voices [58].

Acoustic Correlates of Vocal Quality

Our results show that perception of male voice quality (normal, hoarse and rough) is related to measures of perturbation frequency (JITA, RAP and SPPQ), amplitude perturbation (SAPQ and VAM), SPI and FTRI.

Comparing mean values obtained in the discriminant analysis with normative values from the previous study [45], it may be observed that the male subjects with normal vocal quality presented normal acoustic values, except in the SPI. This may be explained by the fact that some subjects with vocal disorders were included in the cluster, and may have augmented values of SPI. This also prompts us to speculate that those subjects did not present disorders in perturbation frequency and amplitude (jitter and shimmer), whose values would otherwise also have increased, since MDVP measures are more reliable when voices are moderately aperiodic [17]. If this is correct, we may conclude that the human ear and acoustic measurements were equally sensitive in their evaluation of ‘normal’ voice.

Males with hoarse voices presented alterations in perturbation frequency measures (JITA, RAP and SPPQ) and in SPI. This last measure reflects difficulty in closure of vocal cords during phonation. As may be observed,
RAP and SPPQ deviate slightly from normal values and by contrast, JITA shows much higher values. However, this does not mean that vocal jitter is disordered, as the MDVP manual [42] indicates that the best variable to measure the jitter parameter is RAP. Similarly, although the SPI measure is associated with noise components, we cannot relate it to other studies in which the harmonic-to-noise ratio is shown to be the chief prediction for hoarseness [5, 9, 12, 60], since SPI has not been sufficiently studied.

Males with rough vocal quality presented deviant values for all measures in the model: $F_0$, frequency perturbation measures (JITA, RAP, SPPQ), amplitude perturbation measures (SAPQ and VAM), SPI and FTRI. These subjects therefore present difficulties in sustaining periodic vibration in the vocal cords over a specific time, perturbation in period of sustained pitch and sustained amplitude, difficulties in cord closure during phonation and greater intensity of tremor in $F_0$.

Regarding the parameters of frequency and amplitude perturbation, rough vocal quality is principally related to vocal jitter percentages (RAP, %), but not to shimmer, since the MDVP manual [42] indicates that the amplitude perturbation measure which best evaluates shimmer is APQ (not included in the discriminant analysis). Our findings coincide with other studies linking roughness with frequency perturbation [15, 17, 22, 51], although it has also been associated with both frequency and amplitude perturbation [7, 18, 61]. Conflicting with our results, other authors [19, 20] report that roughness may be perceived in the absence of such perturbation.

It should be emphasized that the judges perceived rough vocal quality exclusively in the male group, owing to the fact that they present a low $F_0$. Tables 1 and 2 show that females and males in the third cluster obtained the same measure for RAP (mean = 2.41); however, the voices emitted with lower pitch, i.e. the male voices ($F_0$ mean = 131) were evaluated as rough. Some authors [21, 22, 50, 62] have shown that stimuli with the same perturbation quantity (jitter) are perceived as rougher in low frequencies than in high frequencies; thus, given two voices of different $F_0$ with the same aperiodicity in vocal production, the voice with lower $F_0$ will be judged more rough.

Rough voices have also been correlated with low values for noise-to-harmonic ratio [5] and with spectral noise quantity [5, 11, 23, 24, 26, 63]. Similarly, Wolfe et al. [64] posit a certain relation between predictive importance of noise-to-harmonic ratio in rough voices and low $F_0$ values, claiming that the parameter’s efficiency may be partly the result of a tendency in favor of male subjects and/or low $F_0$. This may explain why the variable did not arise in the rating of the women’s voices, with their high $F_0$ values. Again, as noted in our results, SPI was included as a noise measure but cannot be related to noise measures described in previous studies.

On the basis of our discussion so far, we may speculate that the human ear and acoustic measures were equally sensitive in the evaluation of different qualities of male voice, a phenomenon which may be due to the fact that some analysis and synthesis systems have been developed using male voices [65].

Perceptual evaluation of female voice qualities (normal, hoarse and breathy) are associated with $F_0$, frequency perturbation measures (RAP, SPPQ and VF0), and $F_0$ tremor intensity measures (FTRI). As may be observed (table 4), the group of women rated as normal vocal quality show mean values for each variable falling within the normative values (table 5).
Female hoarse voice is related to high values in frequency perturbation measures (RAP, SPPQ, VF₀), and in F₀ tremor intensity. These results coincide with findings of previous studies in which perceptions of hoarseness are related to measures of vocal jitter [3, 6]. However, results concerning F₀ measures are rather inconclusive, since hoarse voice presents similar F₀ values to normal voice.

In the light of these observations, we may consider that the human ear and acoustic measures were equally sensitive in evaluating perturbation of hoarse female voices, but that they differed in their evaluation of F₀ and pitch.

Turning to women rated as breathy-voiced, we may observe that these present high jitter values (RAP), higher tremors and variations in F₀, and a lower F₀ in comparison with normal and hoarse voices. Previous studies do not seem to support these findings, since most authors consider the principal predictions of breathy voice to be presence of aspiration noise [2, 25, 33, 66], together with energy losses and gains in various spectral zones [9, 67].

The fact that acoustic correlates of vocal quality for women are not as conclusive as those for men may be due to different factors: (a) First, we know that judge reliability increases in line with the severity of pathology, whereas analysis packages (CSL, Cspeech and SoundScope) are less reliable than judges when the pathology is severe, and more reliable with moderately aperiodic voices [68, 69]. Thus it may be that the vocal disorders of the women in our study were more severe than those of the male group (for instance greater extent of lesion), with the result that our acoustic measures lost sensitivity and committed larger rating errors. (b) A second hypothesis relates to the fact that, as noted above, some analysis and synthesis systems have been developed using male voices, losing resolution in high frequencies [2, 65]. In consequence the vocal spectrum obtained with women’s voices is not adjusted to the full model, since analyses should take into account sex differences in phonation and articulation mechanisms. On this point, Titze [65] has questioned whether the generation-filter theory, developed exclusively with men’s voices, would have had the same results if it had been based on women’s voices.

In conclusion, we regard the findings of this study as important for the clinical field, as long as the same work methodology and even the same software (MDVP) are used to obtain acoustic analyses. Many voice analysis systems use algorithms giving results in different units, whose transformation into other comparable measures can be problematic [70].

In the future it would be interesting to extrapolate the acoustic-perceptual profile for the male group, using the data obtained in this study as a starting point, since it is important to test the clinical implications of our results. While it is clear that acoustic-evaluation instruments are becoming common in clinical practice, in our view not all speech therapists understand the significance of the parameters measured by different systems and, above all, they lack precise information concerning which parameters are most important when distinguishing between different qualities of voice, whether normal or pathological.

Acknowledgment

The author would like to thank Julian Bourne of the Department of Translation and Interpreting, University of Granada, for translating this article from Spanish.
References


Acoustic and Perceptual Indicators of Normal and Pathological Voice

52 Mendoza E: Tratamiento conductual de disfonías profesionales; unpublished doc diss University of Granada, Granada, 1989.