ORIGINAl ARTICLE

The Effect of Anchor Voices and Visible Speech in Training in the GRABS Scale of Perceptual Evaluation of Dysphonia

Faustino Núñez-Batalla, a,∗ Juan Pablo Díaz-Molina, a Isabel García-López, b Adriana Moreno-Méndez, c María Costales-Marcos, a Carla Moreno-Galindo, a Pablo Martinez-Camblor d

a Servicio de ORL, Hospital Universitario Central de Asturias, Oviedo, Spain
b Servicio de ORL, Hospital Universitario La Paz, Madrid, Spain
c Universidad Manuela Beltrán, Bogotá, Colombia
d CAIBER, Oficina de Investigación Biosanitaria, Oviedo, Asturias, Spain

Received 5 October 2011; accepted 6 October 2011

KEYWORDS
Voice disorders;
Voice analysis;
GRABS method

Abstract
Introduction: Perceptual evaluation of voice quality remains a key standard for judgement of vocal impairment. The GRABS method has become a commonly used scale for rating severity of dysphonia, but it has no published, standardised protocol to follow. Training is important for reaching good interrater agreement for its parameters; however, the references most often cited for the GRABS provide no guidelines for clinical administration, speech material or rating calibration. This study investigated the effect of anchors (standard reference voices) and visible speech (narrow band spectrogram) in training non-expert professionals in the GRABS method.

Material and methods: Four inexperienced listeners evaluated 107 recorded pathological voices using the GRABS scale in 2 separate sessions; at first, without a visible spectrogram and then, 6 months later, with anchors and a narrow band spectrogram as additional information.

Results: The results show that anchors and visible speech helped to improve the reliability of G, B, A and S parameters. Interrater agreement according to k statistics was significantly stronger with the addition of spectrographic information for rating breathiness and strain.

Discussion: This study found that non-expert listeners showed significant improvement after training with external anchors (standard reference voices) and a narrow band spectrogram.

© 2011 Elsevier España, S.L. All rights reserved.
El espectrograma de banda estrecha como ayuda para el aprendizaje del método GRABS de análisis perceptual de la disfonía

Resumen

Introducción: La evaluación perceptual de la calidad vocal sigue siendo un importante método para evaluar los trastornos vocales. El método GRABS se ha consolidado como una escala frecuentemente utilizada para determinar la severidad de una disfonía, pero no se ha publicado un protocolo estándar para guiar su uso. El entrenamiento es importante para alcanzar una buena concordancia en la calificación de sus parámetros entre distintos observadores, sin embargo, las referencias bibliográficas más citadas no describen orientaciones para su uso clínico, muestras a analizar o calibración.

Material y método: Este estudio investigó el efecto de voces patrón y el espectrograma de banda estrecha en el entrenamiento del GRABS de profesionales no expertos. Las voces de 107 pacientes fueron evaluadas por 4 profesionales no expertos utilizando el espectrograma en 2 sesiones, primero sin voces patrón ni espectrograma y 6 meses después con voces patrón y espectrograma de banda estrecha.

Resultados: Los resultados muestran que las voces patrón y el espectrograma ayudaron a mejorar la concordancia de los parámetros G, B, A y S. La concordancia entre los distintos observadores de acuerdo al estadístico k fue significativamente mayor con la adición del espectrograma para los parámetros B y S.

Discusión: Este estudio demuestra que los profesionales no expertos mejoran significativamente sus puntuaciones tras el entrenamiento con voces patrón y la visualización del espectrograma de banda estrecha.

© 2011 Elsevier España, S.L. Todos los derechos reservados.

Introduction

The recommendation of the National Centre for Voice and Speech (www.ncvs.org) is that the evaluation of a pathological voice must begin with the performance of a spectrographic analysis to classify it, in order to determine the most appropriate method to study each particular case. Thus, a voice with sufficient periodicity (type 1) should be analysed using the parameters for short-term disruptions (jitter, shimmer, normalised noise energy [NNE] and harmonics to noise ratio [HNR]). A voice which presents aperiodicity, subharmonics and voice breaks (type 2) cannot be reliably analysed with the aforementioned parameters (a disruption greater than 5% indicates this type of voice) and should instead be studied by a perceptual classification method (GRABS) and a visual method such as a spectrogram. At present, as long as those dimensions studying chaotic phenomena (fractal dimension, Lyapunov exponent, etc.) are not applied in a practical manner, chaotic voices (type 3) can only be studied by perceptual methods. Therefore, the GRABS classification is an indispensable method for the assessment of voice pathology, since it must be employed on a significant number of patients in whom measurements of short-term disruption are not reliable or cannot be calculated. Since a narrow-band spectrogram is a graphical representation of a sound, observing it simultaneously with the perception of dysphonia can improve both the learning process and the reliability of a subjective method for the classification of pathological voices such as the GRABS method. This study analyses the aid of narrow-band spectrograms in learning this method.

Materials and methods

We conducted a retrospective study of 107 voice samples corresponding to as many patients, 29 males and 78 females, diagnosed with Reinke’s oedema through videostroboscopy. The acoustic signal was recorded using the Voice Assessment application of the Dr. Speech 3.0 program for Windows 95. The computer used was a Pentium-100 compatible PC with 16 MB RAM. A Windows-compatible sound card with 16-bit resolution and recording frequency of 44,100 Hz (Sound Blaster 16) was installed for the digitisation of the voice signal. The microphone used was dynamic unidirectional. The sampling frequency was 44,100 Hz. The microphone used had a high-resolution frequency and was placed at 10 cm from the mouth of patients while they pronounced the vowel /e/ at a comfortable intensity and tone within a soundproofed chamber. The computer captured 3 s from each emission. We followed the recommendations of the National Centre for Voice and Speech.¹

Spectrographic Analysis of Voices

Spectrographic analysis consisted of a narrow-band spectrogram created from digitised voices using the Praat³ program and given to each observer. Each participant conducted the spectrographic analysis of voices on their own PC. The Praat program was configured to obtain a narrow-band spectrogram: after loading the voice to be studied the ‘‘View and edit’’ window was opened and the ‘‘spectrum’’ option was selected in the menu. In ‘‘Spectrogram settings’’, the
The subharmonics, 0.045. The presence showed maximum so in the study.

Figure 1 (A) Visual analogue scale for parameter B in females, with 3 degrees of severity, displaying the corresponding voice samples and the narrow-band spectrograms simultaneously. (B) Visual analogue scale for parameter R in females, with 3 degrees of severity, displaying the corresponding voice samples and the narrow-band spectrograms simultaneously.

"Window length (s)" number was changed from 0.005 to 0.045. After completing this step, the corresponding window showed the narrow-band spectrogram of the voice under study. For a better observation of the curve, the image could be displayed using the "Paint visible spectrogram" option in the "Praat picture" window, from which it was possible to copy and print the spectrogram.

Scales and Patterns of Spectrographic Reference

We created scales for each GRABS item (except G) through spectrographic patterns illustrating their severity extremes so that observers could inspect the spectrogram of the voice being analysed. One end showed a normal voice for the parameter and the other end showed an example of maximum severity. The B parameter was illustrated by the presence of noise in the spectrogram following Yanagihara4 grades, the R parameter was illustrated by the presence of subharmonics,5 the A parameter by the absence of both noise and harmonic paths in the spectrogram6 and the S parameter by the presence of high intensity acoustic signals (noise, harmonics and subharmonics). Fig. 1A shows the scale for the B parameter and Fig. 1B shows the scale for the R parameter.

Perceptive Analysis of Voice (GRABS)

Reference Assessment

Each of the following items was scored from 0 to 3 (0=normal, 1=mild, 2=moderate, 3=severe) using the GRABS method by 2 professionals experienced in speech pathology, together and without visual aids, through the reproduction of each vocal sample: G (grade of dysphonia), the overall level of vocal involvement; R (roughness), the voice quality related to irregular glottic pulses, of a low-frequency noise, roughness or vocal fry component; B (breathiness), voice related to the noise caused by turbulences created by an insufficient glottis; A (asthenia), the auditory impression of weakness in spontaneous phonation, hypokinetic or hypofunctional voice; S (strain), the auditory impression of excessive strain or tension associated with spontaneous phonation. This analysis was repeated independently by the 2 observers 6 months after the first assessment, also without visual aids, to compare the initial joint assessment with the 2 subsequent, independent ones.
Assessment by Professionals Untrained in the Use of the GRABS System

The 4 professionals involved did not have specific training in the GRABS system, although they did have experience in the care of voice disorders (1 ENT specialist, 1 speech therapist and 2 otolaryngology residents in their third year).

Assessment Without Visual Aid. This was conducted by each observer listening to the voices on their computer using headphones and with no visual cues.

Assessment While Observing the Spectrogram. In order to use the spectrogram as a visual aid we selected anchor voices (reference patterns) for the parameters R (roughness), A (asthenia), B (breathiness) and S (strain). Each parameter was illustrated by narrow-band spectrograms of voices with minimum severity and maximum severity, thus creating a visual analogue scale which helped to score the voices under study (Fig. 1A and B). The spectrographic curves relating to each GRABS item were: Yanagihara score for air voice or breathiness (B), the presence of subharmonics as a reference for roughness (R), the lack of signal (non-harmonic, no noise) to score asthenia (A) and a prominent signal in harmonics, noise and subharmonics to score strain (S). This assessment with the visual aid of a spectrogram was repeated 6 months after the initial assessment without spectrogram.

Variables and Statistical Analysis

Continuous variables were described through the mean and standard deviation and categorical variables through relative frequencies. We used the Spearman correlation coefficient and kappa index (reflecting the level of agreement between measurements) to study the difference between the raw scores (belonging to each individual and the baseline score). We used the McNemar test to contrast similarity of the initial and final results. The Spearman correlation coefficient was used as a measure of correlation (association or interdependence) between the scores of professionals with no experience in reference scoring (+1 reflected positive perfect correlation, −1 reflected negative perfect correlation and 0 indicated no correlation). The kappa index was used to rule chance coincidences (+1 reflected total agreement, −1 reflected total disagreement and 0 indicated complete independence).

Statistical analysis was performed using the SPSS program, version 15.0 for Windows (SPSS Inc., Chicago, IL).

Results

Reference Assessment

We compared the GRABS scores obtained jointly by the 2 observers with the GRABS scores obtained independently by each observer 6 months later. The Pearson chi-square test, likelihood ratio, Phi coefficient, Cramer’s V and contingency coefficient tests reached statistical significance (P<.05) for all comparisons. Table 1 summarises the P values of chi-square.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Without spectrogram</th>
<th>With spectrogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>1.98 (0.73)</td>
<td>1.89 (0.75)</td>
</tr>
<tr>
<td>R</td>
<td>1.42 (0.90)</td>
<td>1.18 (1.02)</td>
</tr>
<tr>
<td>A</td>
<td>0.69 (0.93)</td>
<td>0.40 (0.72)</td>
</tr>
<tr>
<td>B</td>
<td>1.00 (0.87)</td>
<td>1.35 (0.93)</td>
</tr>
<tr>
<td>S</td>
<td>1.09 (0.86)</td>
<td>0.78 (0.81)</td>
</tr>
</tbody>
</table>

Assessment by Untrained Professionals

The results of spectrogram assessments, aided and unaided by the visual reference provided by the spectrogram are shown in Table 2 and Fig. 2.

The parameters grade of dysphonia (G), roughness (R), asthenia (A) and strain (S) presented lower mean values after use of the spectrogram, while the breathiness parameter (B) increased (Fig. 2C).

Comparison of Results in Untrained Professionals With the Reference Assessment

We compared the scores of untrained professionals with the reference assessments using the percentage of total matches, the κ coefficient and the Spearman correlation coefficient by averaging the total assessment of the 4 reviewers, with and without using the spectrogram (Table 3 and Fig. 3).

(a) The values which showed a higher correlation with the reference assessment without the aid of the spectrogram were G and R, with Spearman values of 0.52 and 0.47, respectively, followed by B and A with values of 0.39 and 0.37, respectively. The parameter with the lowest correlation was S, with a value of 0.34. After using the spectrogram, the G and R parameters remained without major changes and their Spearman correlation values decreased by only 0.01, A presented a slight improvement (increase of 0.03), and B presented the most significant improvement increasing from 0.39 to 0.50. The correlation of S decreased slightly after use of the spectrogram, going from 0.34 to 0.27.

(b) The percentage of success without the use of the spectrogram was highest when assessing the grade of
Table 3  Influence of the Spectrogram in the Perceptual Evaluation of Pathological Voices According to the GRABS Method.

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th></th>
<th></th>
<th>R</th>
<th></th>
<th></th>
<th>A</th>
<th></th>
<th></th>
<th>B</th>
<th></th>
<th></th>
<th>S</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>S</td>
<td>K</td>
<td>%</td>
<td>S</td>
<td>K</td>
<td>%</td>
<td>S</td>
<td>K</td>
<td>%</td>
<td>S</td>
<td>K</td>
<td>%</td>
<td>S</td>
<td>K</td>
</tr>
<tr>
<td>Initial evaluation</td>
<td>55.2</td>
<td>0.52</td>
<td>0.28</td>
<td>43.7</td>
<td>0.47</td>
<td>0.22</td>
<td>48.1</td>
<td>0.37</td>
<td>0.18</td>
<td>38.2</td>
<td>0.39</td>
<td>0.14</td>
<td>39.0</td>
<td>0.34</td>
<td>0.08</td>
</tr>
<tr>
<td>Final evaluation</td>
<td>56.4</td>
<td>0.51</td>
<td>0.30</td>
<td>42.3</td>
<td>0.46</td>
<td>0.21</td>
<td>49.3</td>
<td>0.40</td>
<td>0.14</td>
<td>44.7</td>
<td>0.50</td>
<td>0.21</td>
<td>45.8</td>
<td>0.27</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Values are reported as mean values of the evaluation by 4 professionals without training regarding the percentage of matches (%), Spearman correlation coefficient (S) and kappa coefficient (K).

dysphonia (G) with 55.2%, followed by asthenia (A) with 48.1%, roughness (R) with 43.7% and strain (S) with 39%, while the breathiness parameter (B) presented the lowest number of matches with 38.2%. Use of the spectrogram improved the percentages of the G, A, B and S parameters. This improvement was discrete in parameters G and A, with a marked improvement in the breathiness (6.5%) and strain (6.8%) parameters. The parameters displaying the best concordance with the reference assessment without using the spectrogram were G and R, with kappa values of 0.28 and 0.22, respectively, followed by A and B with values of 0.18 and 0.14, respectively. The S parameter was the most different with respect to the reference assessment, with a kappa value of 0.08. Correlation improved significantly after the use of the spectrogram in the breathiness parameter (κ changed from 0.14 to 0.21 after using the spectrogram) and the strain parameter (κ changed from 0.08 to 0.16 after using the spectrogram). We performed the McNemar test in order to verify if these improvements were statistically significant, comparing the assessments with and without spectrogram with the reference assessment and obtaining statistically significant differences for both B (P=.001) and S (P=.007).

Figure 2  Distribution of the mean GRABS scores (grade of dysphonia, roughness, asthenia, breathiness and strain) without the aid of the spectrogram (A) and with the aid of the spectrogram (B) in the 107 voices assessed by 4 professional without experience. Representation of the differences found in the evaluation of voices with and without the aid of the spectrogram (C).

Figure 3  Graphical representation of the comparisons between the assessment by professionals without experience and the reference assessment, with and without the aid of the spectrogram. (A) Representation of the results of the kappa value. (B) Representation of the results of the Spearman correlation coefficient. Results which reached statistical significance in the McNemar test are marked with an asterisk.
Discussion

The psychoacoustic sensation resulting from hearing a dysphonic voice remains the primary tool for judging vocal disorders, both for patients who suffer it and for physicians who treat it. Perceptual speech quality scores define the presence or absence of a voice disorder; therefore, it is necessary to establish a method to carry out these perceptual analyses in a consistent and clinically significant manner.\(^7\),\(^8\)

At present, there is no method of rating the perceived quality of voice that is clearly better than the rest. All efforts to obtain a perceptual rating tool for dysphonia have encountered a host of problems related to reliability, usefulness, and validity. However, the development of guidance protocols and the design of methods to document the auditory-perceptual characteristics of an abnormal vocal quality have helped enormously in this task by improving communication and consistency among professionals. In this context, a series of guiding principles for perceptual voice analysis systems have been recommended: (1) the perceptual dimensions should reflect a minimum set of perceptual and clinically relevant voice parameters; (2) the procedures and results should be obtained conveniently; (3) the procedures and results should be applicable to a wide range of vocal pathologies and clinical settings; (4) assessments should clearly enhance intra- and interobserver reliability in validation studies, and (5) voice samples should be used as a reference for assessments and training.\(^7\)

The GRABS method is one of the most widespread methods for perceptual voice assessment. However, it lacks a standard protocol, references to calibrate the scores and a training method with which to improve the consistency of clinical assessment. Examiners with expertise in the perceptual voice assessment generally use their own references for orientation. However, it has been observed that the use of external references improves scoring reliability.\(^9\),\(^10\) This work analyses the influence of visual, spectrographic references combined with acoustic references (anchor voices or reference patterns) as a training method, in order to improve the consistency of perceptual voice assessment using the GRABS system.

The main finding of this study was that, except for the R parameter (roughness), the remaining items in the GRABS system improved in the assessment of matching scores by untrained observers with the reference. The B and S parameters of the GRABS system showed the greatest improvement when aided by the spectrogram. This improvement in coincidence can be explained by the specific characteristics observed in the spectrograms. These characteristics include the presence of noise, which is indicative of breathiness and the presence of a strong acoustic energy in the high-frequency region of the spectrum associated with strained voice.\(^1\)

Another important observation was the worsening of the coincidence of the R parameter in the presence of subharmonics in the spectrogram. In fact, the score without the aid of the spectrogram showed higher coincidence values. This is related with the fact that subharmonics are not the only features of a spectrogram which are associated with rough voice. This observation has already been noted by several authors.\(^1\),\(^11\),\(^12\) This finding confirms the notion that perceptual assessment cannot be replaced by acoustic parameters or by spectrographic parameters, at least in regard to the roughness quality of dysphonia. The study by Martens et al.\(^13\) used a wide-band spectrogram to determine if there were irregularities in the glottic pulses which determined the perception of a rough voice. These irregularities are independent of short-term disruption values and there is no objective way to measure them.

The present study was based on the assumption that subharmonics are related to rough voice, based on previous studies on this subject which established that configurations with subharmonics were related to frequency disruptions perceived as broken and rough voices.\(^14\) The perception of roughness depends on the combined disruption features. Thus, if the disruption parameters originate a voice with 2 or more frequencies it is perceived as dysphonia, but if a pulsatile element is added then it is perceived as a broken voice or vocal fry.\(^15\) Physiological and anatomical studies using high-speed image recordings revealed that there are different vibrational frequencies between the 2 cords or phase differences that are repeated every 6–9 cycles in patients with voices described as rough, without the presence of asymmetries in rigidity or cord mass being necessary.\(^16\) Smith et al.\(^17\) created a model which confirmed that different vibrational frequencies between the 2 cords which gave rise to subharmonics depended on very specific conditions such as a given subglottic pressure, a sufficiently broad glottic area and a specific rigidity of the vocal cords. Other authors have also confirmed that subharmonics give rise to vocal roughness.\(^18\),\(^19\)

The use of a sustained vowel compared to normal speech as recorded samples for subsequent playback and perceptual assessment is controversial. Physicians have a constant exposure to spoken language in their relationship with dysphonic patients and this will undoubtedly influence the assessment obtained with the GRABS method. The basic voice assessment protocol of the European Laryngological Society recommends conducting the perceptual classification of dysphonia during the anamnesis of patients.\(^20\) Some studies have reported that sustained vowels enable a more accurate assessment than continuous speech recordings,\(^21\) whilst others have found the opposite.\(^22\) In either case, the advantage of speech over sustained vowels was not significant. What does seem clear is that sustained vowels are not influenced by the articulators and, therefore, represent a purer product of vocal vibration.

Conclusions

Acoustic references (anchor voices) combined with visual references (narrow-band spectrograms) help improve the consistency of the GRABS perceptual classification system in all parameters except for roughness (R).

The air quality or breathiness of a dysphonic voice is the parameter best predicted by observing the presence of noise and the disappearance of harmonics in a spectrogram.

The roughness quality of a dysphonic voice is not determined solely by the presence of subharmonics in the spectrum.

The use of spectrographic patterns together with the corresponding voice, as employed in this study, is a valid tool
for training in the perceptual assessment of dysphonia using the GRABS scoring system.

**Conflict of Interests**

The authors have no conflicts of interest to declare.

**References**


