ORIGINAL ARTICLE

Cepstral Analysis of Normal and Pathological Voice in Spanish Adults. Smoothed Cepstral Peak Prominence in Sustained Vowels Versus Connected Speech∗

Jonathan Delgado-Hernández, a Nieves M. León-Gómez, b Laura M. Izquierdo-Arteaga, a Yanira Llanos-Fumero a

a CREN Salud, Departamento de Logopedia, La Laguna, Tenerife, Spain
b Unidad de Logopedia, Servicio de Rehabilitación, Hospital Universitario de Canarias, La Laguna, Tenerife, Spain

Received 27 April 2017; accepted 24 May 2017

KEYWORDS
Acoustic analysis of voice; Dysphonia; Smoothed cepstral peak prominence; Praat

Abstract
Introduction and objectives: In recent years, the use of cepstral measures for acoustic evaluation of voice has increased. One of the most investigated parameters is smoothed cepstral peak prominence (CPPs). The objectives of this paper are to establish the usefulness of this acoustic measure in the objective evaluation of alterations of the voice in Spanish and to determine what type of voice sample (sustained vowels or connected speech) is the most sensitive in evaluating the severity of dysphonia.

Method: Forty subjects participated in this study 40, 20 controls and 20 with dysphonia. Two voice samples were recorded for each subject (one sustained vowel/a/ and four phonetically balanced sentences) and the CPPs was calculated using the Praat programme. Three raters perceptually evaluated the voice sample with the Grade parameter of GRABS scale.

Results: Significantly lower values were found in the dysphonic voices, both for/a/ (t[38]=4.85, P<.000) and for phrases (t[38]=5.75, P<.000). In relation to the type of voice sample most suitable for evaluating the severity of voice alterations, a strong correlation was found with the acoustic-perceptual scale of CPPs calculated from connected speech (r s=−0.73) and moderate correlation with that calculated from the sustained vowel (r s=−0.56).

Conclusion: The results of this preliminary study suggest that CPPs is a good measure to detect dysphonia and to objectively assess the severity of alterations in the voice.

© 2017 Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Published by Elsevier España, S.L.U. All rights reserved.


Corresponding author.
E-mail address: jonathan@cren.es (J. Delgado-Hernández).

2173-5735/© 2017 Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Published by Elsevier España, S.L.U. All rights reserved.
Análisis cepstral de la voz normal y patológica en adultos españoles. Medida de la prominencia del pico cepstral suavizado en vocales sostenidas versus habla conectada

Resumen

Introducción y objetivos: En los últimos años se ha incrementado el uso de medidas cepstrales para la evaluación acústica de la voz. Uno de los parámetros más investigados es la prominencia del pico cepstral suavizado (CPPs). Los objetivos de este trabajo son conocer la utilidad de esta medida acústica en la evaluación objetiva de las alteraciones de la voz en español y determinar qué tipo de muestra de voz (vocales sostenidas o habla conectada) es la más sensible para evaluar la severidad de la disfonía.

Método: En este estudio participaron 40 sujetos, 20 normofónicos y 20 con disfonía. Se graron 2 muestras de voz para cada sujeto (una/a sostenida y 4 frases fonéticamente balanceadas) y se calculó la CPPs con el programa Praat. Tres expertos valoraron perceptivamente la voz de la muestra con el parámetro grado de la escala GRABS.

Resultados: Se encontraron valores significativamente menores en las voces disfónicas, tanto para la /a/ (t(18) = 4.85, p < 0,000) como para las frases (t(18) = 5.75, p < 0,000). En relación con el tipo de muestra de voz más indicado para evaluar la severidad de las alteraciones de la voz se encontró una fuerte correlación con la escala acústico-perceptiva de la CPPs calculada a partir del habla conectada (r = -0.73) y moderada con la calculada a partir de la vocal sostenida (r = -0.56).

Conclusión: Los resultados de este estudio preliminar apuntan a que la CPPs es una buena medida para detectar la disfonía y para evaluar objetivamente el grado de severidad de las alteraciones en la voz.

© 2017 Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

Due to the complexity of vocal production, voice evaluation requires the combination of a battery of tests when passing judgement on an alteration or dysphonia to avoid falling into errors of interpretation. These tests may be classified as objective, such as laryngeal video-stroboscopy and acoustic analysis, and subjective such as the acoustic-perceptive evaluations. With regard to acoustic analysis, the most commonly used measures for evaluation of dysphonia have been based on the parameters of frequency alteration (jitter), on amplitude alteration (shimmer) and the harmonic to noise ratio (HNR). A major limitation of these measures is that they require precise detection of the fundamental frequency cycles ($F_0$). In the most severe cases of dysphonia, the irregularity of phonation makes it almost impossible for automatic detection programmes to determine the $F_0$ cycles with precision and calculation of these parameters in these less harmonious voices is therefore unreliable. Another limitation we found was in the type of voice sample used in analyses. Normally sustained vowels are used, since their production is easier and there is less influence of dialect articulation or varieties, but the drawback is they do not represent the everyday voice of the patient.

As a result, a change of paradigm has been observed in the acoustic analysis of dysphonia, where the parameters from the cepstral metric hold greater relevance than the traditional acoustic parameters. The cepstrum is the result of taking the inverse transform of the logarithm of the estimated spectrum of a signal. In other words, it consists of applying 2 Fourier transforms to the sound. Cepstral analysis enables us to determine $F_0$ from the frequency of the harmonic component of a sound, represented by the cepstral peak of the high region of the cepstrum. The cepstrum peak amplitude offers information about the degree of frequency of the signal of lowest frequency in relation to other periodic components or noises present.

One of the most used cepstral measures in objective evaluation of the voice is the smoothed cepstral peak prominence (CPPs). Although it was initially developed by Hillebrand et al. and Hillebrand and Houde to determine the degree of air loss through the vocal cords measured by the breathiness parameter (B) of the GRABS5 scale, many studies have demonstrated its usefulness in the overall evaluation of the severity of dysphonia. It represents the difference between the most prominent cepstral peak, corresponding to the first harmonics, and the point with the same frequency on the line of regression through the smoothed cepstral and is expressed in dB. Fig. 1 shows a graphical representation of the calculation of the CPPs in the production of the vowel /a/. Amplitude, measured in dB is represented by the vertical axis and frequency in the horizontal axis, which is a measurement of time.

The main aim of this study was to determine the usefulness of the CPPs measure with Praat as the acoustic measure for objective evaluation of the voice alterations in Spanish. We also aimed to compare the different types of samples of voice regularly used in the calculation of CPPs (analysis of sustained vowels and connected speech) and to
Voice Recording

Two different voice samples were obtained from each study subject, with a total of 80 samples: a sustained vowel /a/ for approximately 4s and the reading of 4 phonetically balanced phrases of the Spanish matrix sentence test.21 The sentences used were: Carmen tiene 2 libros grandes (Carmen has 2 large books), Elena compra 12 platos nuevos (Elena buys 12 new plates), Teresa hace 7 regalos pequenos (Teresa makes 7 small gifts), Josefa vende 3 zapatos azules (Josefa sells 3 blue shoes). The voices were recorded with an AKG C544L headset condenser microphone and an alesis ir/oz express audio interface. The Praat v. 5.4.0422 programme was used with a sample frequency of 44.1 kHz and 16 bit resolution. The signal-to-noise ratio (SNR) proposed by Deliyski et al.23,24 was calculated in all audios post hoc. Values above 30 dB23 were considered optimum for the acoustic analysis.

Perceptual Analysis

3 speech therapists who were experts in voice rehabilitation therapy (with a mean experience of 11 years) were used for this analysis and perceptually evaluated voice quality of the sample subjects through the GRABS scale. In this perceptual scale a score from 0 to 3 is given for the severity of 5 parameters: G (grade), indicating to what general and overall level the voice has been affected; R (roughness), showing the degree of hoarseness of the voice; A (asthenics), indicating the degree of asthenia, fatigue or tiredness detected in the voice; B (breathiness), revealing the loss of air through the vocal cords, the breathiness or blowing and S (strain), indicating the degree of tension or stiffness of the voice.

For this perceptual evaluation the recordings of the combination of the sustained vowel and sentences of each subject were used, and were presented with JVC HA-RX300E headphones. Each expert rater evaluated the parameter grade (G) of the GRABS scale for each voice.

Acoustic Analysis

The values of the CPPs (in dB) were obtained, both for sustained vowels and connected speech, of all sample subjects with the Praat v. 5.4.0422 programme. The following configuration was used to calculate the CPPs:

Sound [vowel or sentences] → analyse periodicity → PowerCepstrogram [Pitch floor=60 Hz; Time step=.002 s; maximum frequency=5000 Hz; pre-emphasis=50 Hz] → Query → Get CPPs... [subtract tilt before smoothing=no; time averaging window=.01 s; quefrency averaging window=.001 s; peak search pitch range=60–330 Hz; tolerance (0–1)=.05; interpolation=parabolic; tilt line quefrency range=.001-.05; line type=straight; fit method=robust].

Statistical Analysis

Analysis of results was performed with the Statistical R v. 3.3.1 (R Core Team, Vienna, Austria) programme. To evaluate reliability between the 3 raters, the Fleiss kappa coefficient25 (Fk) was calculated. This is an extension of

Table 1 Diagnosis of Dysphonic Subjects.

<table>
<thead>
<tr>
<th>No. subject</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spasmodic dysphonia</td>
</tr>
<tr>
<td>2</td>
<td>Scar in LVC and in 1/3 middle of RVC</td>
</tr>
<tr>
<td>3</td>
<td>Nodules</td>
</tr>
<tr>
<td>4</td>
<td>Nodules</td>
</tr>
<tr>
<td>5</td>
<td>Right laryngoceles</td>
</tr>
<tr>
<td>6</td>
<td>Bilateral Reinke’s oedema</td>
</tr>
<tr>
<td>7</td>
<td>Fusiform oedema in RVC</td>
</tr>
<tr>
<td>8</td>
<td>Spasmodic dysphonia</td>
</tr>
<tr>
<td>9</td>
<td>Mild oedema in LVC</td>
</tr>
<tr>
<td>10</td>
<td>Cordectomy</td>
</tr>
<tr>
<td>11</td>
<td>Cordectomy</td>
</tr>
<tr>
<td>12</td>
<td>Cystic lesion in RVC and nodule in LVC</td>
</tr>
<tr>
<td>13</td>
<td>Pediculated angiomatous polyp in RVC</td>
</tr>
<tr>
<td>14</td>
<td>Bilateral paralysis</td>
</tr>
<tr>
<td>15</td>
<td>Spasmodic dysphonia</td>
</tr>
<tr>
<td>16</td>
<td>Presihibition</td>
</tr>
<tr>
<td>17</td>
<td>Longitudinal hiatus</td>
</tr>
<tr>
<td>18</td>
<td>Longitudinal scar and hiatus</td>
</tr>
<tr>
<td>19</td>
<td>Bilateral capillary ectasias</td>
</tr>
<tr>
<td>20</td>
<td>Right mucosa enlargement/LVC sulcus vocals</td>
</tr>
</tbody>
</table>

RVC: right vocal cord; LVC: left vocal cord.

determine which was most sensitive for evaluating the severity of dysphonia.

Method

Subjects

40 Spanish speaking volunteers took part in this study: 14 men and 25 women, with a mean age of 41.9 years (SD 17.02). The sample was divided into 20 subjects who presented with dysphonia to varying degrees of severity and different aetiologies assessed according to ENT values (Table 1), with a mean age of 47.8 years (SD 19.8), and 20 normophonic subjects with no history of vocal pathologies and a mean age of 36 years (ST 11.20).
the Cohen kappa coefficient for more than 2 judges/raters. The Fk value=1 would correspond to a perfect agreement between the raters and the Fk=0 would correspond to an agreement explicable by coincidence. The Fk is considered as reasonably reliable with values from 0.41. 20

Two comparisons were made with the Students’ t-test for independent samples, one with the sustained vowels sample and the other with the connected speech. The aim was to study the CPPs values of the normophonic group and that of the dysphonic group. A statistical significance of $P<.05$ was established.

To estimate the relationship between the CPPs values and the mean scale of overall severity (G) of the dysphonia the correlation coefficient was obtained by Spearman ($r_s$) ranges, using the samples of sustained vowels and connected speech. The value of the correlation index varied in the interval $[-1, 1]$, with $r_s=1$ as a perfect positive correlation and $r_s=0$ as the absence of a lineal ratio. The cut-off figure of $r_s=.60$ was chosen from which the degree of correlation was considered strong or high between the cepstral and perceptive values.

### Results

The recordings made in this study show a mean SNR of 42.02 dB (SD 5.05), which is the level recommended by Deliyski et al. 23 for voice measures with a precision of at least 99.

Normality and homoscedasticity criteria were studied prior to calculating the Student’s t test statistic. The Kolmogorov–Smirnov test confirmed the normality of the sample (CPPs[vowel] $P=.60$; CPPs[phrases] $P=.30$) and the Levene test the homogeneity of variances (CPPs[vowel] $P=.10$; CPPs[phrases] $P=.068$).

In Table 2 we can observe the results of the Student’s $t$ test comparisons of the CPPs values between the control group and the dysphonic group, both in sustained vowels and connected speech.

Significantly lower CPPs values in the dysphonic group were observed, both when the sustained /a/ vowel and connected speech was used (Fig. 2).

Regarding the relationship between the CPPs and perceptive acoustic scores, firstly the level of agreement between the 3 raters who carried out evaluation was calculated. An Fk coefficient of .54 was obtained, which means that these scores had good Inter-evaluator reliability. Next, the correlation coefficient was calculated by Spearman ranges for the 2 voice samples used (Fig. 3). Moderate correction is observed in the sustained /a/ vowel ($r_s=-.56$, $P<.000$) and high in the connected speech ($r_s=-.73$, $P<.000$). In the case of sentences 53% ($r_s^2=.53$) of the variance of the G parameter mean was accounted for by the CPPs.

### Discussion

The aims of this study were to discover the usefulness of CPPs as an acoustic measurement for the objective evaluation of voice alterations in Spanish and to compare the different types of voice samples usually used when calculating this cepstral measurement (analysis of sustained vowels and connected speech) to determine which is the most sensitive for evaluating the severity of the dysphonia.

---

**Table 2** Results of the Student’s $t$-Test Comparisons Between Control and Dysphonic Subjects in Sustained Vowels and in Connected Speech.

<table>
<thead>
<tr>
<th>CPPs</th>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>95% CI</th>
<th>$t$</th>
<th>$g$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sustained voice</td>
<td>Control</td>
<td>20</td>
<td>14.90</td>
<td>1.81</td>
<td>14.10–15.70</td>
<td>4.85</td>
<td>38</td>
<td>&lt;.000</td>
</tr>
<tr>
<td></td>
<td>Dysphonia</td>
<td>20</td>
<td>11.52</td>
<td>2.52</td>
<td>10.42–12.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connected speech</td>
<td>Control</td>
<td>20</td>
<td>7.982</td>
<td>0.79</td>
<td>7.634–8.331</td>
<td>5.75</td>
<td>38</td>
<td>&lt;.000</td>
</tr>
<tr>
<td></td>
<td>Dysphonia</td>
<td>20</td>
<td>6.211</td>
<td>1.12</td>
<td>5.720–6.703</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CPPs: smoothed cepstral peak prominence; SD: standard deviation; $g$: grades of freedom; 95% CI confidence interval of 95%; n: number of subjects; $P$: significance; $t$: Student’s $t$ test statistical value.

---

**Figure 2** Comparison of the CPPs values between the control and dysphonic group in the sustained vowel /a/ and in the phonetically balanced sentences. dB: decibels; CPPs: smoothed cepstral peak prominence.
The outcome of this preliminary study coincides with those obtained in other studies and shows the usefulness of this acoustic parameter for differentiating between normal and pathological voice and the correlation with perceptive evaluations of the severity of the dysphonia. As the severity of voice alterations increases, the value of the CPPs decreases and vice versa. The CPPs is a measure of the spectral harmonic frequency, which makes it particularly appropriate for evaluation of voice quality, even in the most dysphonic voices, since it does not require exact determination of the F0 cycles such as jitter, shimmer or HNR. The most periodic voices show a well-defined harmonic configuration within the spectrum and therefore a more prominent cepstral peak. In contrast, the lower vocal quality voices present with a reduction in the CPPs (Fig. 4).

This cepstral measure may be calculated with different acoustic analysis programmes in a reliable manner but not numerically comparable. Maryn and Weenick compared the CPPs values measured with the Speech tool programme with those obtained in the Praat programme in a sample of 289 subjects (28 normophonic and 261 with dysphonia). Results show an almost perfect correlation between the 2 programmes (r=0.961), but higher values are observed in the Praat due to the differences in the signal processing (CPPspraat=6.61 dB, CPPspracht=11.66 dB). In this study the Praat was used because this is one of the most powerful most well used programmes in the clinical environment worldwide and because in previous years the CPPs were included as the main acoustic markers in multivariate indexes, such as the acoustic voice quality.
index (AVQ)\textsuperscript{30-33} and the acoustic breathiness index (ABI),\textsuperscript{34} which is calculated with this acoustic analysis programme.

With regard to this type of sample used for acoustic voice evaluation, Maryn et al.\textsuperscript{17} performed a meta-analysis where they studied 69 acoustic measures in sustained vowels and 26 in connected speech. They found that 4 of these measures in sustained vowels (the Pearson r peak of autocorrelation, the $F_0$ amplitude, the spectral envelope of the residual signal and the CPPs) and 3 in connected speech (the SNR based on the lineal prediction code and the inverse filtrated, the cepstral peak prominence [CPP] and the CPPs) were valid acoustic markers for the evaluation of the severity of dysphonia and, among all of these, the CPPs was the only one with sufficient concurrent validity in the 2 types of voice.

In this study a relationship was found between the values of this cepstral peak in both types of voice sample with the severity of dysphonia, although a greater degree of correlation was observed in the values of the CPPs obtained from connected speech. This result coincides with that found in the English language by Halberstam\textsuperscript{35} and differs from that found by Brinca et al.\textsuperscript{3} in a sample of Portuguese speakers, where it was observed that the values obtained from the sustained vowels /a/ correlated better with the perceptive acoustic evaluations. The results of the cepstral analysis in connected speech are conditioned by the type of sample voice used, by dialectic varieties and the reading fluency of the study sample.\textsuperscript{4} In this study 4 phonetically balanced simple sentences was used, which were read out earlier as entertainment, and as a result all the subjects of the sample presented good reading fluency. Furthermore, the sample is homogeneous with regard to dialectic variety. It is important to pursue more exhaustive studies in Spanish in connected speech since the CPPs may provide information about the everyday voice of the dysphonic patient, thus contributing greater ecological validity to the acoustic evaluations.

There are currently no regulatory studies of this acoustic marker in the Spanish speaking population. In the English language Heman-Ackah et al.\textsuperscript{9} studied the CPPs in a sample of 50 normophonic subjects and in 835 with dysphonia of several degrees of severity and aetiology. To calculate the CPPs they used the Speech toll programme with the Hillenbrand et al.\textsuperscript{6,7} algorithm in samples of connected speech. The results shows high sensitivity and specificity in the detection of pathologic voices. This study establishes normality in 4.0 dB or higher and shows the CPPs as a good acoustic measure of dysphonia. In a subsequent study normalisation will be covered from this cepstral measure with the Praat programme in the Spanish population.

Conclusion

The results of this preliminary study suggest that the CPPs is a good measure for detecting dysphonia and for objectively evaluating the degree of severity of voice alterations, especially if voice samples are used in connected speech.

Conflict of Interests

The authors have no conflict of interests to declare.

References

35. Halberstam B. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. ORL. 2004;66:70–3.