Journal of Speech Sciences 7(1): 01-15. 2018.
Available at: http://revistas.iel.unicamp.br/joss

SPEECH TASK AFFECTS THE OBJECTIVE EVALUATION OF DYSPHONIC VOICES

DA SILVA, Wellington[1*]
CONSTANTINI, Ana Carolina[2]

[1]Department of Linguistics, University of Campinas, Campinas, Brazil
[2]Department of Human Development and Rehabilitation, University of Campinas, Campinas, Brazil

---

Abstract: *The acoustic analysis of speech has proved useful in the clinical evaluation of dysphonia, for it allows an objective assessment of the voice. However, the literature has suggested that the type of speech task used to obtain voice samples from patients (sustained vowel or connected speech) may affect both the perceptual and the acoustic evaluation of dysphonic voices. This study aimed at investigating whether the type of speech task significantly influences the acoustic analysis of dysphonic voices. Five acoustic parameters related to voice quality (cepstral peak prominence, difference between the magnitudes of the first and second harmonics, harmonics-to-noise ratio, jitter and shimmer) were automatically computed from voice samples of 5 female and 5 male subjects with and without dysphonia. These recordings consisted of three types of speech task: connected speech, count and sustained vowel. Analyses of variance with repeated measures showed that all five acoustic parameters were significantly affected by speech task. Further analyses through the Duncan's multiple-range test indicated that the type of speech task may also influence the discrimination of dysphonic voices. It is concluded that speech task affects the acoustic assessment of dysphonic voices by significantly raising or reducing the values of the acoustic parameters.*

Keywords: Acoustic analysis; Voice; Dysphonia; Sustained vowel; Connected speech.

---

1 Introduction

The acoustic analysis of speech has been very useful in the clinical evaluation of dysphonia, for it allows the objective assessment of the voice. The perceptual-auditory evaluation, which is frequently used by speech therapists, is by nature a subjective method and thereby it is influenced by the listeners' preferences, training, experience, as well as by socioeconomic and cultural aspects (1). The acoustic analysis consists in measuring physical acoustic parameters of the speech signal such as those related to intensity (perceived amount of energy in the speech signal, responsible for the sensation of volume), temporal characteristics (speech rate and duration of utterances and pauses) and fundamental frequency (acoustic correlate of vocal fold vibration and perceived as the pitch of the voice). These parameters can be extracted directly from the digital recordings of the speech by means of softwares of acoustic analysis. Therefore, in addition to being objective, the acoustic analysis has the advantage of being non-invasive and consistent, since the numerical outcome for a given voice sample does not change as long as the algorithm used in the computation of the parameter remains the same (2).

In the clinical assessment of dysphonia, be it perceptual or instrumental, speech therapists often use two types of speech material: connected speech and sustained vowel. Connected speech consists, for example, in reading an excerpt from a text, sentence or counting numerals. The second type of task, sustained vowel, consists in producing vowels (usually the vowel /a/) in a sustained manner for a few seconds. Both in the perceptual and in the acoustic assessment of dysphonic voices, great emphasis has been placed on the use of sustained vowel (2,3). This preference may arise from several reasons. Sustained vowels are easier to be elicited and analyzed than connected speech, are produced with relatively stable phonation and, unlike samples of connected speech, are not affected by variations brought about by the phonetic context, pauses, fluctuations in fundamental frequency and speech rate, among others (2,4). Therefore, the use of sustained vowel offers the speech therapist a number of practical advantages.

*Corresponding author: ton.wdasilva@gmail.com

The practice of using only samples of sustained vowel in the clinical assessment of dysphonic voices has been criticized by researchers, though. The main criticism concerns the ecological validity of this method, that is, whether this kind of speech sample is representative of the speech used in daily interactions (2). Maryn and colleagues (2), for example, argue that dysphonia symptoms are often noticed by the patients in continuous speech (i.e. connected speech) during their daily conversational routine rather than in sustained vowel productions. Furthermore, some voice disorders, such as adductor spasmodic dysphonia, are more easily perceived in connected speech (2).

Some studies also suggest that stimulus type (sustained vowel or connected speech) may significantly influence the listeners' judgments on the type and the severity of the dysphonia. Wolfe, Cornell, and Fitch (3), for example, studied to what extent the sustained vowel is representative of the voice in connected speech. The authors had 25 listeners rate on bipolar scales speech samples from 20 subjects with no voice disorder and 60 subjects with dysphonia. Based on the results, they argued that, despite being strongly correlated with the judgments for the sentences, perceptual judgments for sustained vowels may not be a reliable clinical index to the severity of dysphonia, since the mean severity ratings for sustained vowels were greater than those for sentences. A similar result was obtained in (4), which found that sustained vowels are judged as more dysphonic than connected speech samples.

Other studies have also shown that some acoustic parameters of speech are more effective in the evaluation of voice disorders if computed for a certain type of speech sample. Halberstam (5), for example, investigated the correlations between acoustic parameters and the perception of hoarseness by voice specialists. The work examined both speech samples of sustained vowel (/a/) and connected speech (the reading of a text), selected from subjects with vocal pathologies and subjects with no vocal pathology. The author concluded that a parameter based on the smoothed prominence of the cepstral peak and computed for the connected speech samples was, among those studied, the one that best correlated with the perceptual judgments of hoarseness. The same parameter, however, presented a lower correlation when computed for the sustained vowel samples. On the other hand, the measures related to the variations in the periodicity of the speech signal (jitter) presented greater correlations with the judgments for the sustained vowels. Sapienza and Stathopoulos (6), however, found only a small effect of the type of speech sample on the values of aerodynamic measures of laryngeal function (maximum flow declination rate, alternating glottal airflow and airflow open quotient), as well as on the values of acoustic parameters (sound pressure level and fundamental frequency). The authors analyzed three types of speaking tasks (sustained vowel, syllable repetition and reading) produced by ten women with bilateral vocal fold nodules and ten women with normal voice. Since the differences found between the tasks were small, they concluded that the choice of the speech task may not have a considerable effect on the objective assessment of dysphonic voices.

It can be observed, therefore, that the scientific knowledge on the influence of the type of speech task (sustained vowel or connected speech) on the (perceptual and acoustic) evaluation of dysphonic voices is still inconclusive. Before moving on to the objectives of the present study, it is useful to examine how some acoustic parameters have been used in the acoustic analysis of dysphonic voices.

1.1 Acoustic parameters in the objective evaluation of dysphonic voices

Some acoustic parameters of speech have been reported to correlate significantly with perceptual judgments of voice quality and thus have proved very useful in the objective evaluation of the voice (7,8). The most frequently investigated parameters have been the cepstral peak prominence (CPP), the difference of amplitude between the first and second harmonics of the voice spectrum (H1-H2), the harmonics-to-noise ratio (HNR), jitter and shimmer.

1.1.1 Cepstral Peak Prominence

The cepstral peak prominence (CPP) is a parameter computed from the cepstrum, a graphical representation of the speech signal in which signal amplitude variation is represented on the vertical axis and a variable called "quefrency" on the horizontal axis (9). The cepstrum allows the separation of the source (which provides the harmonics) from the filter (which generates the formantic regions), so that the slow variations of quefrency are attributed to the formants, and the rapid variations to the fundamental frequency (9). Since the cepstral peaks correspond to the properties of the source, the cepstrum reflects the harmonic organization of the voice (10,11).

CPP is computed as the difference between the amplitude of the cepstral peak and the amplitude of the predicted value on the regression line (applied to the cepstrum to normalize the parameter for the overall amplitude of the signal) for the quefrency at the cepstral peak (12). A highly periodic signal has a well-defined harmonic structure and, therefore, displays a more prominent cepstral peak than a less periodic signal (12). Thus, a breathy voice, for example, should present a lower CPP value than a modal voice (11).

In general, the literature has shown that CPP is one of the most effective parameters for the discrimination or prediction of voice quality types (13,14), as well as the acoustic parameter that best correlates with perceptual judgments of dysphonia (10,14,15,16). Heman-Ackah, Michael and Goding (15), for example, examined speech samples (connected speech and sustained vowel) of 19 patients with unilateral recurrent laryngeal nerve paralysis collected before and after surgical intervention. These speech samples were rated according to overall dysphonia, roughness, breathiness, asthenia, and strain and were also subjected to an acoustic analysis, in which some acoustic parameters, among them a smoothed version of the cepstral peak prominence (CPPS, obtained from the averaging of several cepstra), were computed. CPPS was the parameter that best correlated with the perceptual judgments of overall dysphonia and breathiness. Kumar, Bhat, and Prasad (17) investigated whether there is a significant difference in CPP values between subjects with vocal nodules and subjects with no voice pathology. CPP values were computed for samples of the sustained vowel /a/ produced by these two groups. A t-test of independent variables showed a significant difference between these groups. The subjects in the clinical group (with vocal nodules) presented lower CPP values than the subjects in the control group due to the hoarse and breathy voice of the patients with vocal nodules, which is caused by inadequate closure of the vocal folds. The same result was obtained in a similar study (18), which analyzed samples of the sustained vowel /a/ produced by subjects with unilateral vocal fold paralysis.

1.1.2 H1-H2

H1-H2 corresponds to the intensity difference in dB between the first and second harmonics of the voice spectrum. It is a measure that is mainly correlated with breathy and creaky voice (11). In the creaky voice, the magnitude of the second harmonic is greater than that of the first (the fundamental frequency). In the breathy voice, on the other hand, the magnitude of the second harmonic is substantially lower than that of the first (19). This parameter is also a correlate of the proportion of the glottal cycle during which the glottis is open (the open quotient), so that the greater the difference between the magnitudes of the first and second harmonics, the greater the open quotient (19,20).

Hillenbrand and Houde (13) investigated the effectiveness of six acoustic parameters in predicting breathiness ratings, including H1-H2. The breathiness ratings and the acoustic parameters values were obtained from samples of the sustained vowel /a/ and of connected speech produced by 20 speakers with dysphonia and 5 speakers without dysphonia. H1-H2 presented a moderate correlation with the breathiness ratings for both speech tasks (sustained vowel and connected speech). However, the correlation was stronger for the sustained vowel samples. The parameter that best correlated with the breathiness ratings, for both speech tasks, was the smoothed cepstral peak prominence (CPPS), described above. Kumar, Bhat and Mukhi (21) studied the changes in acoustic parameters related to harmonic amplitude computed for voice samples of the sustained vowel /a/ produced by subjects with vocal nodules and subjects with normal voice. The voice of patients with vocal nodules, according to the authors, is characterized by a hoarse and breathy quality due to inadequate closure of the vocal folds. H1-H2 was significantly greater in the clinical group, that is, there was an abnormal reduction of the amplitude of the second harmonic relative to that of the first. Similar findings were obtained in another study (22), which examined vowel harmonic amplitude differences in subjects with unilateral vocal fold paralysis. This condition is caused by the paralysis of the vagus nerve of the larynx and is also characterized by a breathy voice.

1.1.3 Harmonics-to-Noise Ratio

Harmonics-to-Noise Ratio (HNR) is a measure that quantifies the degree to which noise replaces the harmonic structure of a vowel (23). It is computed as the energy proportion of the noise in the speech signal in relation to that of the harmonics, converted to a decibel scale. The increase in spectral noise (which results in a decrease in HNR) has been linked to an increase in perceived hoarseness (23) and in breathiness (11).

Yumoto, Gould, and Baer (23) analyzed samples of sustained vowel of subjects without voice pathologies and of subjects with a range of laryngeal disorders pre and postoperatively. The authors compared the perceptual judgments of hoarseness based on spectrographic analysis with their HNR measure and obtained high correlations between these variables, concluding, thus, that the HNR is a useful parameter for objectively assessing the results of treatment for hoarseness. In addition, they set a threshold of 7.4 dB for pathology, according to the distribution of the normal subjects. The relevance of HNR in the evaluation of dysphonic voices was also evidenced in (24). The authors applied a k-means nearest neighbor classifier to predict laryngeal pathology in samples of the sustained vowel /a/ from 53 normal and 163 dysphonic subjects. HNR presented high classification accuracy (94%) and thus, according to the authors, is an effective tool for the classification of normal and pathologic voices.

1.1.4 Jitter and shimmer

Jitter and shimmer are parameters used to quantify the cycle-to-cycle variations in the speech signal. In general, jitter can be defined as the cycle-to-cycle variability in period duration, whereas shimmer measures the cycle-to-cycle variability in amplitude (25). Small variations in period and amplitude are a natural characteristic of normal speech. However, large perturbations may reflect laryngeal dysfunction and are associated with a rough voice quality (26).

The relation between measures of jitter and shimmer and voice quality was investigated, for example, in (27). The work examined the effectiveness of four acoustic parameters (average fundamental frequency, jitter, shimmer, and harmonics-to-noise ratio) in predicting dysphonic severity. These parameters were computed for samples of the sustained vowel /a/ produced by normal and dysphonic speakers and the data obtained were correlated with perceptual judgments of breathy, hoarse, rough, and normal voice. For the voice samples classified as rough, both harmonics-to-noise ratio (r = -0.85) and shimmer (r = 0.66) yielded significant correlations. For the breathy samples, the combination of less jitter, more shimmer, and lower harmonics-to-noise ratio yielded a significant R^2 of 0.74. None of the parameters produced significant correlations with the hoarse voice type.

Some studies have also investigated the influence of the speech material type (i.e. sustained vowel or connected speech) on the effectiveness of jitter and shimmer measures in discriminating between normal and dysphonic voices. Schoentgen (28) computed jitter measures in samples of the sustained vowel [a] and of an isolated sentence, produced by French-speaking normal and dysphonic speakers. The author found no significant difference in the discrimination performance of connected speech in comparison with sustained vowels. In a more recent study, Zhang and Jiang (29) investigated the acoustic characteristics of speech samples of sustained vowel (sustained vowel /a/) and connected speech (vowel /a/ extracted from the reading of a text) from normal subjects and patients with laryngeal pathologies. Unlike (28), the speech material type significantly influenced the discrimination performance of jitter and shimmer. Both parameters statistically discriminated between normal and pathological voices for sustained vowels, but not for connected speech.

1.2 The present study

The aim of the present study is to evaluate whether the type of speech material (e.g. sustained vowel and connected speech) may significantly influence the acoustic analysis of dysphonic voices. We examine whether the measured values for the five acoustic parameters described above are significantly affected by the type of speech material used in the analysis and also whether dysphonic subjects are grouped differently for different speech tasks. In this way, this work contributes to advance the knowledge on the influence of the type of speech sample on the acoustic analysis of dysphonic voices. Such knowledge is important to make the objective analysis of pathological voices more reliable and efficient.

2 Material and methods

2.1 Voice samples

The voice samples analyzed in this study were selected from the database of the Speech Therapy Clinic of the School of Medical Sciences of the University of Campinas, which contains recordings from subjects who presented a dysphonic voice at the time of the recording, as well as from subjects who did not present any type of voice disorder. These recordings consist of three types of speech task: connected speech (in which the patient talks about his/her voice condition), count (in which the subject counts from 1 to 20) and sustained vowel (in which the subject produces the vowel [a] in a sustained manner for a few seconds at a comfortable pitch and loudness, a procedure that is repeated three times).

The procedure for obtaining the recordings of this database was as follows: subjects were recorded in a sound treated booth, with a unidirectional microphone Shure (model SM 58) positioned 10 centimeters from the subject’s mouth and connected to a Samsung desktop computer, model PC3200, equipped with a soundcard (Mobile-pre, from M-audio). The recordings were stored on the hard drive with the software Praat (30) with a sampling rate of 44.1 kHz.

The present study was conducted with recordings from 10 speakers of this database, 5 females, mean age of 32 years and 5 males, mean age of 41 years. Additional subject information is presented in Table 1. According to perceptive-auditory analysis, two of these subjects presented no dysphonia at the time of the recording, since they were at the end of the treatment. The remaining subjects presented mild functional dysphonia, which is a type of voice disorder characterized by the absence of structural or neurologic laryngeal pathology (31). Their voices were characterized by different voice qualities, mostly breathy and creaky.

Table 1: Details of the subjects selected for this study

Subject	Sex	Age (years)	Dysphonia level
1	Male	26	Mild dysphonia
2	Female	30	Mild dysphonia
3	Female	41	Mild dysphonia
4	Male	26	Mild dysphonia
5	Male	64	No dysphonia
6	Female	25	Mild dysphonia
7	Female	24	Mild dysphonia
8	Female	41	Mild dysphonia
9	Male	26	No dysphonia
10	Male	63	Mild dysphonia

2.2 Acoustic analysis

In order to perform an acoustic analysis on the voice samples, the oral and nasal vowels and diphthongs of these samples were manually segmented in the software Praat (30). In the case of the sustained vowels, the starting and the end points of the interval corresponding to the vowel were delimited. The criterion used for acoustic segmentation was the visibility in the spectrogram of the second formant of the vowels, following (9).

Then, the Praat script “VQAnalysis” was used to compute automatically from the vowels and diphthongs of the voice samples five acoustic parameters related to voice quality: cepstral peak prominence (CPP), H1-H2, harmonics-to-noise ratio (HNR), jitter and shimmer. To obtain more precision in the measurements, CPP and H1-H2 were computed for the vowels /ɛ a ɔ/ with duration greater than 100ms only. Being open and open-mid vowels, their first formant is further away from their first and second harmonics, which are thus less affected by this formant, reflecting more the effect of voice quality.

CPP was computed here for an analysis window of 100ms and according to the method described by (12). In this method, a linear regression line relating quefrency to cepstral magnitude is applied to normalize the parameter for the overall amplitude of the signal. CPP is then taken as the difference between the amplitude of the cepstral peak and the amplitude of the predicted value on the regression line for the quefrency at the cepstral peak.

H1-H2 was calculated as the difference of amplitude between the first and second harmonics of a FFT spectrum with window length of 25ms, extracted at the midpoint of the vowel. The amplitude of the first harmonic was taken as the maximum intensity in dB within the frequency range from 0 Hz to the frequency corresponding to 1.5 times the fundamental frequency median. Similarly, the amplitude of the second harmonic was taken as the maximum intensity in dB within the frequency range from the frequency corresponding to 1.5 times the fundamental frequency median to the frequency corresponding to 2.5 times the fundamental frequency median.

Shimmer was calculated as the average absolute difference between the amplitudes of consecutive periods of the speech signal, divided by the average amplitude, whereas jitter was computed as the average absolute difference between consecutive periods, divided by the average period. The HNR was taken as the mean value of the Harmonicity object (the object in Praat which represents the degree of acoustic periodicity), expressed in dB.

3 Results

3.1 Effect of speech task on the acoustic parameters

To examine the effect of the factor SPEECH TASK on the acoustic parameters, that is, whether there are significant differences between the mean values of the acoustic parameters in the three levels of this factor (connected speech, count, sustained vowel), a multivariate repeated measures analysis of variance (RM-ANOVA) was conducted. One test of RM-ANOVA was performed for each acoustic parameter, with the mean value of each subject for the acoustic parameter as the dependent variable and the factor SPEECH TASK with three levels as the independent variable. Post-hoc comparisons were performed with the Wilcoxon signed rank test using the Bonferroni method for the p-value correction.

Because the repeated measures designs assume that there is no interaction between subjects and treatments, the subjects that exhibited a different tendency across the levels from the majority of the subjects were excluded from the group and a simple one-way ANOVA was conducted for them. When the assumptions of normality and/or homoscedasticity (tested for with Shapiro-Wilk and Fligner-Killeen tests) were not met, the non-parametric equivalent Kruskal-Wallis test was carried out instead. In this case, the post-hoc analyses were carried out with the Tukey HSD test (after the ANOVA) or with the Wilcoxon rank sum test with the Bonferroni adjustment for the p-value (after the Kruskal-Wallis test).

The statistical analyses were carried out using the R package (32). Significance levels were set to 0.05. Table 2 shows the means and standard deviations of the acoustic parameters according to speech task, computed for the data analyzed through RM-ANOVA. The results of the statistical analysis for the five acoustic parameters are presented as follows.

3.1.1 Cepstral Peak Prominence

The multivariate RM-ANOVA conducted with 7 subjects who exhibited the same tendency across the levels of the factor showed a significant effect of speech task on CPP [Wilks' $\Lambda = 0.25$, $F\ (2, 5) = 7.62$, $p < 0.04$]. The Wilcoxon signed rank post-hoc test revealed statistically significant differences between the count and connected speech tasks ($p < 0.05$) and between sustained vowel and count ($p < 0.05$). In addition, the difference between sustained vowel and connected speech was marginally significant ($p < 0.07$). Sustained vowel was the speech task that presented the greatest mean value of CPP (see Table 2).

The one-way ANOVA performed with the data of the remaining 3 subjects also revealed a statistically significant difference between groups [$F\ (2, 69) = 7.221$, $p < 0.002$]. There was a significant difference of 5.3 dB between sustained vowel and count ($p < 0.001$) as well as a significant difference of 3.2 dB between sustained vowel and connected speech ($p < 0.04$). In both cases, sustained vowel presented greater mean.

3.1.2 H1-H2

There was a significant effect of speech task on H1-H2 as revealed by the multivariate RM-ANOVA [Wilks' $\Lambda = 0.02$, $F\ (2, 4) = 83.14$, $p < 0.0006$]. However, the Wilcoxon signed rank post-hoc test only showed marginally significant differences ($p < 0.1$) for the three pairs of conditions. The reason for this discrepancy may be the high standard deviations presented by the three speech tasks, which can be observed in Table 2. With regard to the means, sustained vowel exhibited a greater value than the other tasks.

For subjects 2 and 7, the one-way ANOVA showed no significant effect of speech task on H1-H2 [F (2, 32) = 1.479, p = 0.24], whereas for subjects 6 and 8 the Kruskal-Wallis test revealed a significant difference between groups [$\chi^2(2) = 12.064$, $p < 0.003$]. In this case, there was a significant difference of 5.2 dB on average between sustained vowel and count ($p < 0.01$), as well as between sustained vowel and connected speech ($p < 0.0003$), as revealed by the Wilcoxon test. Sustained vowel had a greater mean than the other tasks.

3.1.3 Harmonics-to-Noise Ratio

Because there was no interaction between subjects and treatments, the multivariate RM-ANOVA was performed with all the 10 subjects. The test revealed a significant effect of speech task on HNR [Wilks' $\Lambda = 0.079$, F (2, 8) = 46.44, $p < 10^{-04}$]. The Wilcoxon post-hoc test showed that the mean of sustained vowel was significantly greater than the mean of connected speech ($p < 0.006$) and that the mean for count was significantly greater than the mean of connected speech ($p < 0.006$).

3.1.4 Jitter

There was also no interaction between subjects and treatments for this parameter. The multivariate RM-ANOVA carried out with the 10 subjects showed a statistically significant effect of speech task on the jitter values [Wilks' $\Lambda = 0.052$, F (2, 8) = 71.81, $p < 10^{-05}$]. According to the Wilcoxon post-hoc test, the sustained vowel group mean was significantly lower than the count ($p < 0.006$) and the connected speech group means ($p < 0.006$). In addition, connected speech presented a significantly greater group mean than count ($p < 0.006$).

3.1.5 Shimmer

A statistically significant effect of speech task on shimmer was also observed through the multivariate RM-ANOVA, which was conducted with 9 subjects that exhibited the same behavior across the levels of the factor [Wilks' $\Lambda = 0.128$, F (2, 7) = 23.93, $p < 0.0008$]. The Wilcoxon post-hoc test showed that sustained vowel had a significantly lower group mean than connected speech ($p < 0.02$) and count ($p < 0.02$). It also showed that connected speech had a significantly greater group mean when compared to the count task ($p < 0.03$).

The subject 4, who presented a different behavior across the treatments, showed no statistically significant difference between groups, as indicated by the Kruskal-Wallis test [$\chi^2(2) = 0.55$, $p = 0.76$].

Table 2: Means and standard deviations of the five acoustic parameters according to speech task for the data analyzed through RM-ANOVA

Acoustic Parameter	Speech Task	Mean	Std. Deviation
CPP	Count	22.08 dB	1.61 dB
	Connected Speech	20.65 dB	1.66 dB
	Sustained Vowel	24.64 dB	2.34 dB
H1-H2	Count	-0.11 dB	3.75 dB
	Connected Speech	1.78 dB	2.92 dB
	Sustained Vowel	4.29 dB	3.71 dB
HNR	Count	17.62 dB	3.05 dB
	Connected Speech	13.62 dB	2.68 dB
	Sustained Vowel	20.45 dB	4.90 dB
Jitter	Count	0.66 %	0.18 %
	Connected Speech	1.14 %	0.36 %
	Sustained Vowel	0.19 %	0.12 %
Shimmer	Count	6.24 %	1.80 %
	Connected Speech	8.54 %	1.73 %
	Sustained Vowel	3.44 %	1.29 %

3.2 Effect of speech task on the grouping of subjects

The way subjects are grouped together by the acoustic parameters may indicate common voice quality abnormalities between them. Given the significant effect of speech task found in the group means of the acoustic parameters, it is also useful to know whether the grouping of subjects according to their values for the acoustic parameters may differ as a function of the speech task. In order to evaluate this, the Duncan's multiple-range test was performed on the subjects' acoustic data for each speech task, after testing for a significant effect of the factor SUBJECT (independent variable) on the acoustic parameters (dependent variable) by means of a one-way ANOVA. Duncan's multiple-range test is a post-hoc test which groups the levels of the factor according to the probability of means differences and the significance level (33,34).

For the cepstral peak prominence, the ANOVA revealed a significant difference between groups for connected speech [F (9,112) = 2.649, $p < 0.009$] and for sustained vowel [F (9,20) = 2.693, $p < 0.04$]. Table 3 shows the grouping of subjects for this parameter according to the Duncan's test. It can be noticed that subject 06, who presented the greatest mean in both tasks, differed significantly from subjects 04, 01 and 09 in the connected speech task. When considering the sustained vowel, however, subject 06 differed significantly from subjects 05 and 03. It should be recalled that subjects 05 and 09 displayed no dysphonia. Therefore, CPP discriminated one subject with dysphonia (subject 06) from one subject with non-pathological voice in each speech task.

Table 3: Subjects' means for cepstral peak prominence according to speech task

Speech Task			
Connected Speech		**Sustained Vowel**	
Subject	**Mean**	**Subject**	**Mean**
subj06	23.76[a]	subj06	28.80[a]
subj07	22.60[ab]	subj09	28.13[ab]
subj02	22.28[ab]	subj07	27.73[ab]
subj05	21.68[ab]	subj10	24.73[ab]
subj10	21.48[ab]	subj02	24.43[ab]
subj03	21.17[ab]	subj08	24.30[ab]
subj08	21.13[ab]	subj04	23.77[ab]
subj04	19.09[b]	subj01	23.13[ab]
subj01	18.90[b]	subj05	22.90[b]
subj09	18.83[b]	subj03	22.40[b]

Values are expressed in dB. Superscripted letters indicate the groups obtained through the Duncan's test considering the probability of means differences and significance level (α = 0.01). Values in the same column followed by the same letter are not significantly different. The speech task not shown did not present a significant difference between group means. See text for more information.

For H1-H2, the ANOVA showed a significant effect of the factor SUBJECT for the count speech task [F (9,47) = 13.75, $p < 10^{-09}$], connected speech [F (9,112) = 7.423, $p < 10^{-07}$] and for sustained vowel [F (9,20) = 12.74, $p < 10^{-05}$]. The results of the Duncan's test for this parameter are presented in Table 4. Subject 01, who showed the lowest H1-H2 mean in the three speech tasks, was isolated in his group in the count and sustained vowel conditions. Considering the connected speech, however, he was grouped together with subjects 04 and 05. The negative H1-H2 means displayed by subject 01 is compatible with the strong creaky and rough quality that characterizes his voice, since the literature shows that, for creaky voices, the amplitude of the second harmonic is greater than that of the first. Subject 02, who presented the greatest H1-H2 mean in the count and connected speech tasks, was grouped together with subjects 06, 07 and 08 in the count task, and also with subjects 03, 09 and 10 in connected speech. In the sustained vowel task, however, her H1-H2 mean was lower and she was placed in a different group from subject 08, who had the greatest mean for this task. The relatively high H1-H2 values displayed by subject 02 are explained by her breathy voice.

Table 4: Subjects' means for H1-H2 according to speech task

Speech Task					
Count		**Connected Speech**		**Sustained Vowel**	
Subject	**Mean**	**Subject**	**Mean**	**Subject**	**Mean**
subj02	7.90[a]	subj02	7.68[a]	subj08	10.48[a]
subj07	6.62[ab]	subj07	6.76[a]	subj06	8.27[ab]
subj06	4.52[abc]	subj10	5.02[ab]	subj03	7.61[abc]
subj08	3.81[abc]	subj06	4.35[ab]	subj10	7.27[abc]
subj03	3.79[bc]	subj03	4.07[ab]	subj02	5.89[bc]
subj10	3.20[bcd]	subj09	3.33[ab]	subj07	5.40[bc]
subj09	1.27[cde]	subj08	2.58[ab]	subj09	5.12[bc]
subj04	-1.05[de]	subj05	0.96[bc]	subj04	4.71[bc]
subj05	-1.43[e]	subj04	0.16[bc]	subj05	3.67[c]
subj01	-6.42[f]	subj01	-2.84[c]	subj01	-2.62[d]

Values are expressed in dB. Superscripted letters indicate the groups obtained through the Duncan's test considering the probability of means differences and significance level (α =

0.01). Values in the same column followed by the same letter are not significantly different. See text for more information.

Jitter also had a significant effect of the factor SUBJECT for count [F (9,277) = 4.255, $p < 10^{-04}$], connected speech [F (9,923) = 13.26, $p < 10^{-15}$] and for sustained vowel [F (9,20) = 21.44, $p < 10^{-07}$]. According to the results of the Duncan's test, presented in Table 5, subjects 01 and 10 were grouped together in the connected speech and sustained vowel tasks. They were the subjects who exhibited the greatest means in these tasks. The count task, on the other hand, was not efficient at separating these subjects from others. Their relatively high jitter values are a result of a creaky voice, which causes large period perturbations. However, the creaky voice of subject 10 was not reflected in the parameter H1-H2 as much as the voice of subject 01. One possible reason for this is that subject 10 showed an audible air leak in addition to creakiness, which raised his H1-H2 values.

Table 5: Subjects' means for jitter % according to speech task

Speech Task					
Count		**Connected Speech**		**Sustained Vowel**	
Subject	**Mean**	**Subject**	**Mean**	**Subject**	**Mean**
subj08	0.94^{a}	subj01	1.90^{a}	subj01	0.43^{a}
subj09	0.88^{ab}	subj10	1.46^{ab}	subj10	0.37^{a}
subj01	0.82^{ab}	subj08	1.23^{bc}	subj03	0.20^{b}
subj10	0.74^{abc}	subj05	1.14^{bc}	subj05	0.20^{b}
subj02	0.64^{abc}	subj04	1.11^{bc}	subj08	0.20^{bc}
subj04	0.61^{abc}	subj09	1.08^{bc}	subj09	0.13^{bc}
subj05	0.59^{abc}	subj02	0.91^{bc}	subj04	0.10^{bc}
subj06	0.58^{bc}	subj06	0.83^{c}	subj06	0.10^{bc}
subj03	0.44^{c}	subj07	0.82^{c}	subj07	0.10^{bc}
subj07	0.38^{c}	subj03	0.73^{c}	subj02	0.09^{c}

Values are given in %. Superscripted letters indicate the groups obtained through the Duncan's test considering the probability of means differences and significance level ($\alpha = 0.01$). Values in the same column followed by the same letter are not significantly different. See text for more information.

Shimmer had a significant effect of the factor SUBJECT only for count [F (9,277) = 5.522, $p < 10^{-06}$] and for connected speech [F (9,923) = 6.979, $p < 10^{-09}$]. The results of the Duncan's test, presented in Table 6, show that the connected speech task discriminated more groups of subjects than the count task. In the latter, the subjects were divided into two groups, whereas in the former, they were classified according to four groups. In both tasks, subjects 01, 08 and 09 were in the group which had the greatest shimmer means. Subject 04, who was among the greatest means in the count task, presented the lowest mean in the connected speech task. The normal subjects (subjects 05 and 09) were in the same group in the count task, whereas in the connected speech task they were placed in distinct groups.

Table 6: Subjects' means for shimmer % according to speech task

Speech Task			
Count		Connected Speech	
Subject	**Mean**	**Subject**	**Mean**
subj01	9.1[a]	subj09	11.7[a]
subj08	8.9[a]	subj01	10.8[ab]
subj04	7.1[ab]	subj08	9.4[abc]
subj02	7.0[ab]	subj05	8.6[bcd]
subj09	6.1[ab]	subj02	7.8[cd]
subj03	5.8[b]	subj10	7.6[cd]
subj05	5.6[b]	subj06	7.5[cd]
subj10	4.7[b]	subj03	6.9[cd]
subj06	4.6[b]	subj07	6.7[cd]
subj07	4.3[b]	subj04	6.3[d]

Values are given in %. Superscripted letters indicate the groups obtained through the Duncan's test considering the probability of means differences and significance level (α = 0.01). Values in the same column followed by the same letter are not significantly different. The speech task not shown did not present a significant difference between group means. See text for more information.

For HNR, the ANOVA revealed a significant difference between groups for count [F (9,277) = 12.62, $p < 10^{-15}$], connected speech [F (9,923) = 31.81, $p < 10^{-15}$] and for sustained vowel [F (9,20) = 34.06, $p < 10^{-09}$]. The results of the Duncan's test for this parameter are presented in Table 7. Subject 01 presented the lowest mean in the three speech tasks, which indicates more noise in his voice. He differed significantly from all the other subjects in the count and connected speech tasks. In the sustained vowel task, however, he was grouped together with subjects 08 and 10. The greatest mean in the count and connected speech tasks comes from subject 07. In the count task, she was grouped together with subject 03, whereas in the connected speech task she was in the same group as subjects 02, 03 and 04. In the sustained vowel task, however, subject 07 differed significantly from subject 02, who presented the greatest mean in this task. Subjects 05 and 09 were in the same group in the count and connected speech tasks, but they fell into distinct groups in the sustained vowel task.

Table 7: Subjects' means for HNR according to speech task

Speech Task					
Count		Connected Speech		Sustained Vowel	
Subject	**Mean**	**Subject**	**Mean**	**Subject**	**Mean**
subj07	22.65[a]	subj07	16.99[a]	subj02	27.13[a]
subj03	20.80[ab]	subj02	16.03[ab]	subj06	25.07[ab]
subj02	18.94[bc]	subj03	15.48[ab]	subj09	24.07[abc]
subj06	18.46[bc]	subj04	15.15[ab]	subj07	23.23[bc]
subj05	18.32[bc]	subj06	14.34[bc]	subj03	22.87[bc]
subj04	17.69[bc]	subj08	13.93[bc]	subj04	20.43[cd]
subj10	16.86[c]	subj09	12.50[c]	subj05	19.07[d]
subj09	16.28[c]	subj10	12.49[c]	subj10	15.13[e]
subj08	15.52[c]	subj05	12.26[c]	subj08	14.93[e]
subj01	11.10[d]	subj01	7.53[d]	subj01	12.53[e]

Values are expressed in dB. Superscripted letters indicate the groups obtained through the Duncan's test considering the probability of means differences and significance level (α = 0.01). Values in the same column followed by the same letter are not significantly different. See text for more information.

4 Discussion

The study reported in this paper was conducted to investigate whether the type of speech material (e.g. sustained vowel and connected speech) may affect the acoustic assessment of dysphonic voices, contributing, thus, to make the objective evaluation of pathological voices more reliable and effective. The analyses conducted by RM-ANOVA showed that all five acoustic parameters examined in this study, which are among the most used acoustic measures in this type of evaluation, were significantly affected by the type of speech task considered for the analysis. The sustained vowel, which has been the preferred type of speech sample in the voice clinic, differed significantly from the other speech tasks for four acoustic parameters. Even for H1-H2, which was the exception, the sustained vowel task presented a mean value considerably greater than the other tasks (thirty-nine times greater than that of the count task and almost three times greater than the connected speech mean). This finding is in line with that of previous studies which have suggested that the type of speech sample used in the acoustic analysis may produce significant changes in the acoustic parameters (5,29,35).

The direction of the difference in mean value presented by the sustained vowel task depended on the acoustic parameter, though. Sustained vowel showed greater mean values than the other tasks for CPP, H1-H2, and HNR, whereas for the perturbation measures (jitter and shimmer), it presented lower mean values than the other tasks. Results reported in the literature have shown that higher values of CPP, H1-H2, and HNR and lower values of jitter and shimmer are associated with non-pathological voices (11,23,27). Therefore, this result indicates that a subject's voice may seem less dysphonic than it really is in the acoustic analyses performed with samples of sustained vowel. This finding goes against the results of studies which compared the perceptual judgments for sustained vowels with those for connected speech (3,4). These studies have found that sustained vowels are rated as more dysphonic than connected speech samples. One reason for this discrepancy may be the subjective nature of the perceptual evaluation. Listeners may have the impression that a voice is more dysphonic when assessed through sustained vowel simply because sustained vowels are assessed in isolation, out of a linguistic context. On the other hand, connected speech is characterized by variations brought about by prosody, phonetic context and pauses, which can mask the severity of the dysphonia for the listeners and significantly affect the acoustic parameters.

One question that still has to be addressed is whether these observed differences in the acoustic parameters may significantly influence the discrimination or classification of dysphonic voices by means of acoustic analysis. To shed light on this matter, we examined through the Duncan's multiple-range test whether the grouping of subjects according to their acoustic data differ as a function of the speech task. The results revealed some differences in how subjects were grouped across the speech tasks. Because some acoustic parameters correlate better with specific types of voice quality and the subjects evaluated in this study showed different types of voice quality, it is expected that the groups of subjects differ from each other for different parameters, as was the case in this analysis. In addition, the five acoustic parameters did not discriminate the dysphonic subjects from the normal speakers very well. One possible explanation for this is the fact that the dysphonic subjects evaluated in this study presented only a mild dysphonia, which made it difficult to distinguish these two groups of subjects through acoustic analysis. Nevertheless, the parameters that best discriminated subjects were H1-H2 and HNR. The subjects that stood out with extreme values for these parameters were subjects 01 and 02. Subject 01 presented a rough voice with creakiness, which yielded negative H1-H2 means. He also showed relatively low HNR values, which indicates more non-periodic components (noise) in his voice. Subject 02 displayed relatively high H1-H2 values due to a breathy voice and also high HNR means. Given the differences found in the grouping of subjects across speech tasks, it is possible to conclude that the type of speech task may influence the discrimination between normal and dysphonic voices by means of acoustic analysis.

5 Conclusions

The results of the present study show that the type of speech task chosen for analysis affects the acoustic assessment of dysphonic voices by significantly raising or reducing the values of the acoustic parameters. Therefore, to avoid a mistaken assessment, it is advisable to use more than one speech task in the acoustic evaluation of dysphonic patients, so that the results can be compared across the tasks. In addition, the acoustic evaluation should be accompanied by perceptual analysis whenever possible. In this way, the acoustic analysis can be a valuable tool for voice clinicians as it provides a means of complementing the clinical assessment of dysphonia with objective and consistent data.

6 Acknowledgements

The authors thank Plinio A. Barbosa for providing the Praat script used in the acoustic analysis. The first author acknowledges a doctoral fellowship from the National Council of Technological and Scientific Development - CNPq (141567/2015-5).

REFERENCES

1. Master S, Biase N, Pedrosa V, Chiari B. O espectro médio de longo termo na pesquisa e na clínica fonoaudiológica. Pró-Fono Revista de Atualização Científica. 2006;18(1):111-120.

2. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward Improved Ecological Validity in the Acoustic Measurement of Overall Voice Quality: Combining Continuous Speech and Sustained Vowels. Journal of Voice. 2010;24(5):540-555.

3. Wolfe V, Cornell R, Fitch J. Sentence/vowel correlation in the evaluation of dysphonia. Journal of Voice. 1995;9(3):297-303.

4. Maryn Y, Roy N. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. Jornal da Sociedade Brasileira de Fonoaudiologia. 2012;24(2):107-112.

5. 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(2):70-73.

6. Sapienza C, Stathopoulos E. Speech task effects on acoustic and aerodynamic measures of women with vocal nodules. Journal of Voice. 1995;9(4):413-418.

7. Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P. Acoustic measurement of overall voice quality: A meta-analysis. The Journal of the Acoustical Society of America. 2009;126(5):2619-2634.

8. Wolfe V, Martin D. Acoustic correlates of dysphonia: type and severity. Journal of Communication Disorders. 1997;30(5):403-416.

9. Barbosa P, Madureira S. Manual de fonética acústica experimental: aplicações a dados do português. São Paulo: Cortez Editora; 2015.

10. Heman-Ackah Y, Sataloff R, Laureyns G, Lurie D, Michael D, Heuer R et al. Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia. Journal of Voice. 2014;28(6):783-788.

11. Shue Y, Chen G, Alwan A. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. In: Eleventh Annual Conference Of The International Speech Communication Association. 2010. p. 34-37. [accessed 26 Jan 2018] Available from: http://www.isca-speech.org/archive/archive_papers/interspeech_2010/i10_0034.pdf

12. Hillenbrand J, Cleveland R, Erickson R. Acoustic Correlates of Breathy Vocal Quality. Journal of Speech Language and Hearing Research. 1994;37(4):769-778.

13. Hillenbrand J, Houde R. Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech. Journal of Speech Language and Hearing Research. 1996;39(2):311-321.

14. Lowell S, Colton R, Kelley R, Mizia S. Predictive Value and Discriminant Capacity of Cepstral- and Spectral-Based Measures During Continuous Speech. Journal of Voice. 2013;27(4):393-400.

15. Heman-Ackah Y, Michael D, Goding G. The Relationship Between Cepstral Peak Prominence and Selected Parameters of Dysphonia. Journal of Voice. 2002;16(1):20-27.

16. Heman-Ackah Y, Michael D, Baroody M, Ostrowski R, Hillenbrand J, Heuer R et al. Cepstral Peak Prominence: A More Reliable Measure of Dysphonia. Annals of Otology, Rhinology & Laryngology. 2003;112(4):324-333.

17. Radish Kumar B, Bhat J, Prasad N. Cepstral Analysis of Voice in Persons With Vocal Nodules. Journal of Voice. 2010;24(6):651-653.

18. Radish Kumar B, Bhat J, Fahim S, Raju R. Cepstral Analysis of Voice in Unilateral Adductor Vocal Fold Palsy. Journal of Voice. 2011;25(3):326-329.

19. Gordon M, Ladefoged P. Phonation types: a cross-linguistic overview. Journal of Phonetics. 2001;29(4):383-406.

20. Holmberg E, Hillman R, Perkell J, Guiod P, Goldman S. Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice. Journal of Speech Language and Hearing Research. 1995;38(6):1212-1223.

21. Radish Kumar B, Bhat J, Mukhi P. Vowel Harmonic Amplitude Differences in Persons With Vocal Nodules. Journal of Voice. 2011;25(5):559-561.

22. Radish Kumar B, Bhat J, Usman J. Vowel harmonic amplitude differences in individuals with unilateral vocal fold paralysis. Journal of Laryngology and Voice. 2015;5(1):7.

23. Yumoto E, Gould W, Baer T. Harmonics-to-noise ratio as an index of the degree of hoarseness. The Journal of the Acoustical Society of America. 1982;71(6):1544-1550.

24. Shama K, Krishna A, Cholayya N. Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology. EURASIP Journal on Advances in Signal Processing. 2006;2007(1).

25. Teixeira J, Oliveira C, Lopes C. Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. Procedia Technology. 2013;9:1112-1122.

26. Askenfelt A, Hammarberg B. Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. Journal of Speech Language and Hearing Research. 1986;29(1):50-64.

27. Martin D, Fitch J, Wolfe V. Pathologic Voice Type and the Acoustic Prediction of Severity. Journal of Speech Language and Hearing Research. 1995;38(4):765-771.

28. Schoentgen J. Jitter in sustained vowels and isolated sentences produced by dysphonic speakers. Speech Communication. 1989;8(1):61-79.

29. Zhang Y, Jiang J. Acoustic Analyses of Sustained and Running Voices From Patients With Laryngeal Pathologies. Journal of Voice. 2008;22(1):1-9.

30. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program, version 6.0.06]. 2015 [accessed 30 Nov 2015]. Available from: http://www.praat.org/.

31. Roy N. Functional dysphonia. Current Opinion in Otolaryngology & Head and Neck Surgery. 2003;11(3):144-148.

32. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015 [accessed 30 Nov 2015]. Available from: https://www.R-project.org/.

33. Bewick V, Cheek L, Ball J. Statistics review 9: one-way analysis of variance. Critical Care. 2004;8(2):130–136.

34. Machado A, Barbosa P. Uso de técnicas acústicas para verificação de locutor em simulação experimental. Language And Law / Linguagem E Direito. 2014;1(2):100-113.

35. Parsa V, Jamieson D. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. Journal of Speech Language and Hearing Research. 2001;44(2):327-339.