PROPERTIES OF THE CEPSTRAL PEAK PROMINENCE AND ITS USEFULNESS IN VOCAL QUALITY MEASUREMENTS

C. A. Ferrer1,2, M. S. de Bodt2, Y. Maryn3, P. Van de Heyning2, M. E. Hernández-Díaz1

1Center for Studies on Electronics and Information Technologies, Central University of Las Villas, Santa Clara, Cuba
2Dept of Otorhinolaryngology, Head & Neck Surgery & Communication Disorders, Antwerp University, Belgium
3Dept of Speech-Language Pathology & Audiology, Sint-Jan General Hospital, Bruges, Belgium

Abstract: Unlike many acoustic measures, Cepstral Peak Prominence (CPP) has shown consistently high correlations with subjective vocal quality ratings. However, this superiority of the CPP index is reported based on empirical results, with its theoretical advantages not always clearly stated. In this paper the properties of the CPP which makes it a good predictor for vocal quality are addressed, as well as how it differs from other measures. The reported experimental setups of the previous studies are analyzed, and reasons for the observed variability in the results are given. After this discussion, the clinical usefulness of CPP is addressed. This paper can be useful for clinicians, willing to interpret the results of the acoustic measures, as well as for researchers, in planning the experimental setup and interpreting the relevance of the results.

Keywords: Cepstrum, vocal quality, breathiness, roughness

I. INTRODUCTION

Many acoustic measures have been proposed to correlate with overall vocal quality or one of its dimensions (i.e. breathiness, roughness, strain, hoarseness, etc.; an extensive tabulation of methods can be found in [1]). In spite of the large number of measures available, there is a lack of consistent results across different studies for most of the measures (e.g. jitter, shimmer, HNR) [2]. Recent work [3][4][5] has shown the Cepstral Peak Prominence (CPP) [6] or its smoothed version (CPPs) [7] to correlate highly with vocal quality dimensions and overall grade. The high correlation in these latter studies has been consistent and notably superior to the other acoustic measures considered.

There are several topics, though, which should be more clearly stated. Most of these works provide experimental data, where CPP results better based on empirical evidence. The theoretical advantages of CPP over the rest of the acoustic measures are not always clearly stated. Some of the experimental setups also favor abnormally high values for the amount of variance explained.

In this paper we address the properties of the CPP that makes it a good predictor for vocal quality, and how it differs from other measures. Besides, the reported experimental setups of the previous studies are analyzed, and reasons for the observed variability in the results are given. After this discussion, the clinical usefulness of CPP is addressed. This paper can be useful for clinicians, willing to interpret the results of the acoustic measures, as well as for researchers, in planning the experimental setup and explaining the relevance of the results.

II. CPP PROPERTIES

The origin of the CPP measure in [6] is, as its companion measure RPK for autocorrelation peak, a basic pitch detector. Both measures were devised to appraise the prominence of the peak that should occur at the pitch value in the cepstrum and autocorrelation functions, respectively. As such, CPP is sometimes erroneously believed to be a measure of signal periodicity, when, in fact, it only measures the periodicity of the signal spectrum. It is precisely this subtle difference (measuring spectral harmonic periodicity instead of strict periodicity) what makes it particularly suited for vocal quality measures, superior to many other measures.

Following is a categorization of measures in five groups, according to signal characteristics, which have been most correlated with different vocal quality dimensions. The first two are the more common amplitude (shimmer) and frequency (jitter) perturbations, absent in the original studies on CPP [6][7], while the other three groups are the ones actually included in those studies. The sensibility of CPP to the signal characteristics is commented, as well as its possible advantages and drawbacks compared to other measures.

1. Amplitude perturbations (shimmer).

Signals with amplitude perturbations have frequently been related to roughness [3], and sometimes with breathiness. The traditional measures of shimmer are obtained in the time domain, relying on a Pitch Detection Algorithm (PDA). CPP is sensitive to shimmer, since shimmer affects the spectral harmonic structure [8]. CPP values diminish as shimmer increases, and can be more robust than time-domain techniques relying on a PDA. It has been shown that shimmer, jitter and time-domain Harmonics-to-Noise Ratios (HNR) are quite sensitive to even small errors in the pulse boundaries [9].
2. Frequency perturbations (jitter)

Jitter shares a similar condition than shimmer, being mostly related to roughness, and less frequently to breathlessness. Jitter affects spectral structure to a greater extent than shimmer [8] and CPP can therefore be also a good measure of this perturbation. The same advantage regarding the sensibility of time-domain measures of jitter to errors in the PDA holds in this case in favor of CPP.

3. Additive Noise

The presence of additive noise has been related mainly with breathiness. The prominence of the cepstral peak is also affected by increasing levels of noise, since it reduces the dip between harmonics. In fact, several studies have focused in this property to develop HNR reduction algorithms. In fact, several studies have focused in this property to develop HNR reduction algorithms. Existing HNR measures have been regarded as overall disperiodicity measures, since they have been shown sensitive also to jitter and shimmer [10][8][11]. CPP also shares this feature, being sensitive to these three groups.

4. First Harmonic Amplitude

A high amplitude of the first harmonic (with respect to the second harmonic [7] or to the first formant [12]) has been related to breathiness. The underlying assumption is that breathy voices do not produce abrupt glottal closures, producing an excitation which is more rounded, almost sinusoidal. First harmonic amplitude prominences are closely related to glottal flow measures like the amplitude quotient or the speed quotient [13]. Here the CPP is superior to its companion RPK in [6] and to other HNR measures. The CPP will produce no prominent peak for a perfect sinusoid, since a sinusoid consists of only one harmonic (no spectral periodic structure). That is the main difference with other periodicity measures: a perfectly periodic signal not necessarily produces a high CPP. This lack of higher harmonics is also typical of nasal voices [1], extending the sensibility of CPP to the nasality dimension.

5. Spectral Tilt

An increment in the energy content in the higher portion of the spectrum has been related to breathiness [14]. CPP is not able to measure spectral tilt changes, which would be reflected in the lower part of the cepstra, discarded for its calculation. However, spectral tilt measures have been reported to have the smallest relevance in breathiness ratings in several other studies [6][7]. CPP inability to follow spectral tilt changes can be of negligible effect on its prediction of breathiness.

As seen, CPP can produce adequate response to most of the signal characteristics which have been related to many vocal quality dimensions (breathiness, roughness, hoarseness, nasality). If an orthogonal representation of the GRBAS scale is accepted [15], the CPP can be expected to be a better predictor of overall Grade, than of any individual dimension. This would occur because selective response of CPP to one particular dimension is affected by its sensitivity to the others.

The next section explains the results of the CPP index in several reported studies in terms of the previous discussion.

III. Reported Studies

The studies covered in this section are the original CPP and CPPs by Hillenbrand et. al. [6] and Hillenbrand & Houde [7], and more recent studies by Heman-Ackah et. al. [3], Awan & Roy [4] and Maryn et. al. [5].

- Hillenbrand et. al. (1994) [6].

This study consisted in the voluntary control of three breathiness phonation levels by 15 normal subjects on four vowels. The number of judges was high (20) and the rating scale was an unrestricted visual-analog (VA). The different acoustic indexes were calculated over three types of signals: the original, a band-pass filtered signal and a high-pass filtered version. CPP emerged as a very good predictor of breathiness ratings (Pearson’s $r$ greater than 0.9, more than 80% of the variance explained by $r^2$) with RPK in the band-pass signal showing similar results.

This study intentionally limited the perturbation to breathiness. This has, according to the discussion in Section II, a twofold consequence. First, breathiness ratings do coincide with “grade” since it is the only deviant dimension, and second, the obtained correlations can be high because CPP is not affected by interference with other distortions. The possible influence of using non-pathological speakers is addressed in the analysis of the next study.

- Hillenbrand & Houde (1996) [7]

Here a broad pathological database was screened to select 20 recordings presenting mainly breathiness, as well as 5 recordings from nonpathological subjects. Recordings included a sustained vowel as well as running speech. The number of judges was 20, and the scale used was unrestricted VA. CPP and CPPs were again the best predictors, with similar results (up to 85% and 92% of the variance explained in the running speech and sustained vowels, respectively).

In this study no version of the RPK could match the performance of its equivalent cepstral measure (a best result of 72% of variance explained). A possible cause is that pathological speakers showed a stronger influence of first harmonic amplitudes and spectral tilt measures in
breathiness ratings than in the previous study, and CPP is better suited than RPK to reflect at least the former factor. Again, the restriction of deviant dimension to breathiness can explain the extremely high correlations obtained.

- Hemman-Ackah et. al. (2002) [3]
  Voices from 19 patients were available, preoperative and postoperative, in both a sustained vowel and running speech. Two judges rated grade, breathiness and roughness in a 120 mm VA scale.
  The results are lower than the ones in [6] and [7] with the cause being the absence of a selective screening of the deviant dimensions, which is more likely the case in the clinical practice. Here the results for grade (65%-75% of the variance explained) are considerably better than the ones for breathiness (50%) or roughness (20%-25%). These results are in complete correspondence with the discussion in Section II.

  Recordings from 83 dysphonic and 51 normal female subjects were rated by 12 judges as belonging to four groups or voice types: normal, breathy, rough and hoarse. The degree of the dimension was not the goal of the study, only the type.
  The study found CPP to be good at discriminating normal from dysphonic voices, but it was not relevant for the separation among the different dysphonic types. A logarithmic shimmer measure was found best suited for the later purpose. This is also in correspondence with our analysis in Section II. CPP is similarly sensitive to breathy and rough signal characteristics, and can not be a reliable separator among them.

- Maryn et. al. (2007) [5]
  This study comprised recordings from both a sustained vowel and running speech, from 229 patients and 22 normal subjects. Samples were rated by five judges in the G, R and B dimensions of the GRBAS scale.
  CPP ranked again the best among all acoustic measures considered, and again the correlation was strong with overall grade and breathiness ratings. Results are the lowest reported (50% of the variance explained for grade) but the size of the database is also the largest, thus including more variability than previous studies.

IV. DISCUSSION AND CONCLUSION

According to the previous sections, CPP can be expected to appraise overall grade better than any other acoustic measure of vocal quality previously reported. If proper screening of samples is performed (i.e. limit signal deviation to a single dimension) CPP can produce extremely high values of correlation with the individual dimensions.

A significant reduction in the percent of variance explained occurs when considering signals with a wide range of variability, but even in that case, CPP can still perform as the best single predictor of overall vocal quality. Another point in favor of CPP is its similar performance on sustained vowels and running speech. The desirability of using running speech for acoustic measures has been pointed out in several studies [7][5], and only a small fraction of the existing measures can work on running speech.

The usefulness of CPP is limited, though, in trying to separate the different dimensions of vocal quality. Its sensitivity to most of the relevant distortions found in pathological voice makes it better suited to predict grade than any individual dimension. Since the later is usually the case in clinical practice, complementary acoustic measures are needed to perform an accurate and exhaustive description of vocal quality in terms of objective measures.

REFERENCES


