Cepstral and entropy analyses in vowels excerpted from continuous speech of dysphonic and control speakers

Antonella Castellana[1], Andreas Selamtzis[2], Giampiero Salvi[2], Alessio Carullo[1], Arianna Astolfi[3]

[1]Department of Electronics and Telecommunications, Politecnico di Torino, Italy
[2]Department of Speech, Music and Hearing, KTH, Stockholm, Sweden
[3]Department of Energy, Politecnico di Torino, Italy

antonella.castellana@polito.it, selamt@kth.se

Abstract

There is a growing interest in Cepstral and Entropy analyses of voice samples for defining a vocal health indicator, due to their reliability in investigating both regular and irregular voice signals. The purpose of this study is to determine whether the Cepstral Peak Prominence Smoothed (CPPS) and Sample Entropy (SampEn) could differentiate dysphonic speakers from normal speakers in vowels excerpted from readings and to compare their discrimination power. Results are reported for 33 patients and 31 controls, who read a standardized phonetically balanced passage while wearing a head mounted microphone. Vowels were excerpted from recordings using Automatic Speech Recognition and, after obtaining a measure for each vowel, individual distributions and their descriptive statistics were considered for CPPS and SampEn. The Receiver Operating Curve analysis revealed that the mean of the distributions was the parameter with the highest discrimination power for both CPPS and SampEn. CPPS showed a higher diagnostic precision than SampEn, exhibiting an Area Under Curve (AUC) of 0.85 compared to 0.72. A negative correlation between the parameters was found (Spearman; $\rho = -0.61$), with higher SampEn corresponding to lower CPPS. The automatic method used in this study could provide support to voice monitorings in clinic and during individual's daily activities.

Index Terms: Cepstral Peak Prominence Smoothed, Sample Entropy, continuous speech, dysphonia, voice pathology, automatic speech recognition, phonetic alignment

1. Introduction

Vocal pathologies often manifest as noise and irregularity in the voice signal. Consequently, markers that quantify irregularity have attracted considerable interest in clinical voice research (see Buder for an overview [1]). Traditional time-domain parameters, such as *jitter* and *shimmer*, have been the first investigated methods, but their dependence on accurately identifying cycle boundaries, i.e. where vocal fold cycles begin and end, make them unreliable with highly perturbed signals [2]-[3]. Recent studies are therefore considering new approaches for analyzing the voice signal, and among them cepstral-based and nonlinear techniques are the most promising measures of voice quality. Two markers related to these methods will be examined and compared in this paper for their ability to discriminate between healthy and pathological voices: the Cepstral Peak Prominence Smoothed (CPPS) and the Sample Entropy (SampEn).

CPPS is a measure (in dB) of the peak amplitude in the cepstrum, normalized for overall signal amplitude by means of a linear regression line calculated relating quefrency to cepstral magnitude [4]. The cepstrum is a log power spectrum of a log power spectrum: the first power spectrum shows the frequency content of the signal energy, while the second spectrum reflects the periodicity of harmonics in the spectrum. The Smoothed version of the CPP is obtained by applying two smoothing processes before calculating the Cepstral Peak Prominence [4]. Several studies have highlighted the relevance of CPPS as indicator of overall grade of dysphonia and of different types of voice quality in both sustained vowels and continuous speech [5-11]. These studies indicate that the higher the peak in the cepstrum, the better the quality of voice.

Sample Entropy (SampEn) is a metric from the field of nonlinear time series analysis, introduced by [12], as the successor of Approximate Entropy (ApEn) [13]. SampEn quantifies the irregularity of a time series; a low number for SampEn signifies regularity while a higher number signifies increasing degree of irregularity. This means that for a healthy voice signal the expected SampEn value is low while for a pathological voice the presence of irregularity should result in a higher SampEn value. A number of studies have used SampEn and ApEn for assessing voice irregularity from electroglottographic or acoustic signals [14-16].

Routine clinical examination of vocal health is largely based on perceptual evaluation of the voice quality [17], and videoendoscopic images of the larynx. The perceptual evaluation is done using either running speech or sustained vowels. Although sustained vowels are used for videoendoscopic examination, it has been argued that asking subjects to produce sustained vowels is somehow artificial [17-18]. For that reason clinicians tend to prefer running speech when they evaluate voice quality perceptually. In this paper we analyze all the [a] vowels excerpted from running speech, focusing on a specific widely researched pattern taken from its natural context. In this way we also comply with the clinic-friendly paradigm.

This paper aims to address the following questions:
a) How powerful are the CPPS and SampEn metrics in discriminating vocal health in shorter vowels where smoothing in the time dimension is not possible?
b) How do these two metrics correlate with each other?
c) Can the combination of these two metrics improve discrimination of vocal health?

2. Method

2.1. Voice samples

The study sample consisted of 33 voluntary patients, 25 females and 8 males (age range: 20-82 years; mean: 50.0 years; standard deviation SD: 16.5 years) and 31 healthy adults with normal voices, 18 females and 13 males (age range: 19-49 years; mean: 25.4 years; SD: 7.8 years). All subjects were native Italian speakers.

Each participant followed the same procedure, which can be summarized in two steps:

(1) They read aloud an Italian standardized phonetically balanced passage of 300 words at a comfortable pitch and loudness,

(2) Two otolaryngologists performed the clinical practice that included a careful case history and the videolaryngoscopy examination.

Table I summarizes the otolaryngologic diagnoses and their occurrences in the patient group.

Before performing step (1) of the protocol, subjects wore an omni-directional head mounted microphone Mipro MU-55HN, which was placed at a distance of about 2.5 cm from the lips of the speaker, slightly to the side of the mouth. The microphone, which exhibits a flat frequency response (±3 dB) in the range from 40 Hz to 20 kHz, was connected to a bodypack transmitter ACT-30T that transmits to a wireless system Mipro ACT 311. The output signal of this system was recorded with a handy recorder ZOOM H1 (Zoom Corp., Tokyo, Japan) using a sample rate of 44.1 kHz and a resolution of 16 bit.

Voice recordings were performed in a quiet room with an A-weighted equivalent background noise level of 50.0 dB (SD = 2 dB), which was measured with a calibrated class-1 sound level meter (NTi Audio XL2) over a period of 5 minutes in four different days. The background noise level is 10 dB lower than the mean lowest A-weighted levels of 39 dB (60 dB) and 44 dB (65 dB) at 30 cm (2.5 cm), which were respectively found in healthy males and females producing their softest possible vowel [19]. Pathological voice tends to be softer than healthy voice, but in our experiment subjects were asked to read aloud, so an acceptable Signal-to-Noise Ratio was kept.

Table 1: *Diagnoses for the patient group.*

Organic dysphonia	Number
Cyst	7
Edema	9
Sulcus vocalis	3
Polyp	4
Chronic laryngitis	2
Vocal fold hyposthenia	3
Vocal fold paresis	2
Vocal fold nodule	1
Post-surgery dysphonia	2
Overall	*33*

2.2. Data processing

The first 118 words of each reading were considered in our analyses. Phonemic annotations for the speech files were created using automatic speech recognition and the orthographic transcriptions (prompts) by means of forced alignment. The recognizer was trained on the Italian SpeechDat corpus [20] using HTK [20] and the RefRec scripts [22]. The SpeechDat corpus includes more than 3000 speakers recorded over the telephone line with 8-bit, 8 kHz, A-law quality. Rather than resampling our speech files to fit the acoustic models, the feature extraction procedure was modified to limit the mel scale filterbank from 0 to 4 kHz. Monophone models with 32 Gaussian components per HMM state were used. Out-of-vocabulary words and their canonical pronunciation were added to the dictionary. This step does not affect the automatic nature of the method because the prompt text in this application is known in advance. The time-aligned transcriptions were used to extract speech samples belonging to all occurrences of the [a] vowel both for pathological and control speakers.

For computational reasons, signals were down-sampled to 25 kHz the middle 1024 samples (40.96 ms) were extracted from each vowel waveform to consider a segment unaffected by transient onset and offset behavior. For each vowel, a CPPS measure and a SampEn measure were computed, thus obtaining a time series of each parameter per subject, which was treated as a distribution. The size of the time series ranges from 43 to 71 values. For each individual distribution the following descriptive statistics were calculated: *mean*, *median*, standard deviation (*std*), *range*, 5th percentile (*5prc*) and 95th percentile (*95prc*).

2.2.1. Cepstral Peak Prominence Smoothed

A MATLAB (R2014b, version 8.4) script has been developed to implement an adjusted version of the Cepstral Peak Prominence Smoothed (CPPS) definition given by Hillenbrand *et al*. [4]. Since vowel segments excerpted from continuous speech are not long enough, the version of CPPS used here did not include the time-smoothing step of the cepstra.

The method implemented in the algorithm can be summarized as follows. Each selected signal was first Hamming-windowed and then the Fast Fourier Transform algorithm was implemented twice: the first time on the signal in time, the second time on the log power spectrum, obtaining the cepstrum. Before extracting the cepstral peak, a quefrency-smoothing step was performed, where a cepstral-magnitude average was obtained across quefrency with a seven-bin averaging window. A regression line was calculated in the quefrency vs cepstral magnitude domain between 1 ms and the maximum quefrency, since quefrencies below 1 ms are more affected by the spectral envelope than by the regularity of the harmonics [23]. The level difference (in dB) between the peak in the cepstrum and the value of the regression line at the same quefrency represents the CPPS measure, where the peak search is limited to the range from 3.3 ms to 16.7 ms, corresponding to fundamental frequencies of 300 Hz and 60 Hz, respectively.

2.2.2. Sample Entropy

SampEn is calculated [12, 24] by separating a time series into sequences of length of m and $m+1$ points. Then the conditional probability is calculated that the Chebyshev distance between two sequences of length $m+1$ is less than a tolerance r, given that the Chebyshev distance between two sequences of length m is less than a tolerance r. This probability is calculated as the ratio of the number of sequences of length $m+1$ whose Chebyshev distances are less than r over the number of sequences of length m whose

Chebyshev distances are less than r. SampEn is then defined as the negative natural logarithm of this ratio.

For each vowel, SampEn was calculated for a number of overlapping windows of length N, using the MATLAB implementation provided in [24]. The length of each window N was f_0 dependent and equal to four times the length of the glottal cycle, while the overlap was equal to the length of one glottal cycle. The f_0 was estimated using the YIN algorithm [25]. The reason for adapting the window length in this way is the sensitivity of SampEn to fundamental frequency [26] and maintaining consistency of the analyzed pattern across subjects. The reader should keep in mind that this type of windowing is best applicable for Type 1 and Type 2 (non-chaotic) voice signals [27]. The total estimate of SampEn per vowel was taken as the mean of the values obtained from all windows. As suggested in [16], the sequence length m was taken to be equal to round($\log_{10}(N)$) and the matching tolerance r was taken to be 0.1 times the standard deviation of the N points in the analyzed window.

2.3. Statistical Analyses

The two-tailed Mann-Whitney U-test, a non-parametric test based on independent samples [28], was used to evaluate statistical differences of the paired lists of descriptive statistics related to the groups of healthy and pathological subjects. The null hypothesis states that MD = 0, where MD is the median of the population of the differences between the sample data for the two groups. If the null hypothesis is accepted, the two lists of values come from the same population, i.e. it is not possible to distinguish healthy and pathological voice samples. The selection of the Mann-Whitney U-test was made based on the one-sample Kolmogorov-Smirnov test, which verified that the data in each list did not come from a normal distribution.

The Spearman correlation coefficient was used to determine the relationship between CPPS and SampEn values. The above-mentioned tests were performed using a MATLAB script (R2014b, version 8.4).

The Receiver Operating Characteristic (ROC) analysis was used to investigate the discrimination power of each descriptive statistic of CPPS and SampEn distributions in healthy and pathological voices. The area under the Receiver Operating Characteristic, which is named Area Under Curve (AUC), was calculated as an indicator of classification accuracy [29]. The AUC ranges from 0.5 to 1.0: a value higher than 0.9 indicates an outstanding separation between the two groups, an AUC between 0.9 and 0.8 designates an excellent discrimination power, an AUC between 0.8 and 0.7 an acceptable discrimination, while an AUC close to 0.5 shows a poor ability to separate the two groups. Moreover, the ROC analysis was performed to determine the preliminary criteria for positivity, i.e., the threshold value from which a voice could be indicated as pathological, for CPPS and SampEn in [a] vowels excerpted from continuous speech. The optimal threshold between the dysphonic group and controls was evaluated plotting together sensitivity and specificity versus each possible threshold. Sensitivity corresponds to the true positive rate, i.e. the percentage of subjects with voice disorders that are classified as positive. Specificity is the true negative rate, that is the percentage of people with healthy voice who are identified as negative. Instead of conventionally selecting the optimal threshold value where the two curves cross, the authors gave priority to the sensitivity, maximizing in this way the proportion of subjects with voice disorders that are classified as positive. As a final step, a logistic regression

model combining both cepstral and entropy measures as predictive variables and assuming the presence/absence of dysphonia as dependent variable was performed. The Wald test was carried out for assessing the significance of coefficients for each predictive variable, where the null hypothesis states that the coefficient of the variable is equal to zero, i.e., the variable is not contributing to the logistic model [30]. These analyses were performed using the statistical program SPSS (v. 21; SPSS Inc, New 223 York, NY).

3. Results

Table 2 shows the p-values of the Two-tailed Mann-Whitney U-test of the lists of descriptive statistics for CPPS distributions related to the patients and controls. The p-values were lower than 0.05 for the *mean*, *median*, *5prc* and *95prc*, which means the null hypothesis was rejected. The *Std* and the *range*, instead, had p-values higher than 0.05. These outcomes reveal that CPPS distributions are significantly different in central tendency, with an overall average value of 12.2 dB and 14.1 dB for the *mean* in patients and controls, respectively, but not in variance, with an overall average value of 3.0 dB and 2.8 dB for the *std* in patients and controls, respectively.

Table 2 also shows that only the *mean* and *95prc* of the SampEn distributions had p-values of the two-tailed Mann-Whitney U-test lower than 0.05. SampEn distributions are thus significantly different in central tendency, with an overall average value of 0.7 and 0.5 for the *mean* in patients and controls, respectively, but not in variance, having an overall average value of 0.3 and 0.2 for the *std* in patients and controls, respectively.

Table 2: *Analysis results for each paired list of descriptive statistics related to the dysphonic group and the control group. Two-tailed Mann-Whitney U-test p-values: values lower than 0.05 are in bold and indicate the rejection of the null hypothesis. Area Under Curve (AUC) and the relative 95% Confidence Interval (CI)*

Statistics	CPPS		SampEn	
	U-test	AUC (CI)	U-test	AUC (CI)
Mean	**0.001**	0.85 (0.76÷0.95)	**0.002**	0.72 (0.59÷0.85)
Median	**0.001**	0.85 (0.75÷0.94)	0.999	0.70 (0.58÷0.84)
Std	0.116	0.59 (0.45÷0.73)	0.884	0.59 (0.45÷0.73)
Range	0.158	0.61 (0.47÷0.74)	0.509	0.45 (0.31÷0.60)
5prc	**0.001**	0.73 (0.61÷0.85)	0.993	0.67 (0.54÷0.80)
95prc	**0.001**	0.79 (0.68÷0.90)	**0.004**	0.71 (0.58÷0.84)

In Table 2 the discrimination power of each descriptive statistic for CPPS and SampEn distributions is highlighted through AUC values and the respective 95% Confidence Intervals (CI). Generally, all the descriptive statistics from CPPS distributions had higher AUCs than the ones from SampEn distributions, highlighting a better diagnostic precision of the cepstral measure with respect to the entropy

metric. Among the descriptive statistics from CPPS distributions, the *mean* and the *median* had an AUC of 0.85, showing a good discrimination power between dysphonic and healthy subjects. Regarding the descriptive statistics from SampEn distributions instead, the *mean* had the highest AUC of 0.72, highlighting a moderate discrimination power.

Figure 1 shows the boxplots of the *mean* from individual CPPS distributions for the group of patients and controls: as expected, healthy speakers had higher values than patients. The bold line in Figure 1 indicates the optimal threshold obtained from the ROC analysis for the mean, i.e., 14.0 dB. The criterion for positivity thus corresponds to 14.0 dB or lower, with a sensitivity of 79% and a specificity of 71%.

Figure 2 shows the boxplots of the *mean* from SampEn distributions for subjects belonging to the healthy group and the pathologic one: predictably, an opposite trend with respect to Figure 1 is evident, having higher values of entropy for subjects with voice disorders. The criterion for positivity is 0.58 or higher, with a sensitivity of 73% and a specificity of 68%.
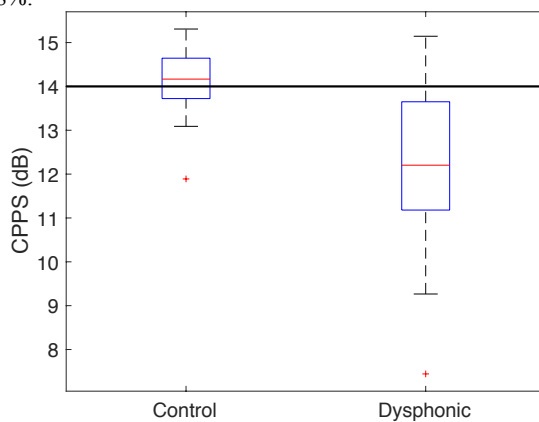

Figure 1: *Boxplots of the mean from individual CPPS distributions for the controls and the dysphonic group. The bold line marks the optimal threshold of 14.0 dB*
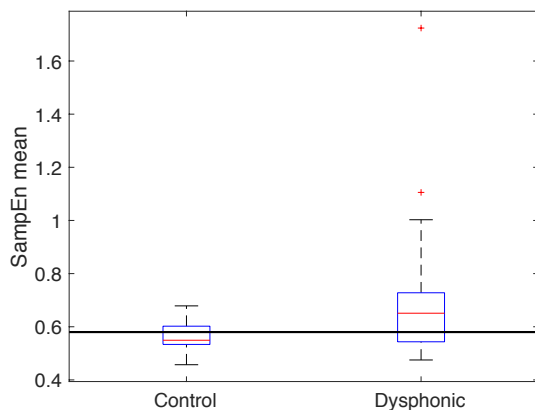

Figure 2: *Boxplots of the mean from individual SampEn distributions for the controls and the dysphonic group. The bold line indicates the optimal threshold of 0.58.*

The limited number of subjects with a given pathology did not allow for a statistical testing that would reveal which pathologies are easier to discriminate using CPPS or SampEn.

However, by checking the CPPS and SampEn results in relation to the diagnosis, it was observed that some pathologies were more successfully discriminated than others. CPPS *mean* and SampEn *mean* discriminated successfully vocal fold paresis (2 of 2), polyp (4 of 4), and post-surgery dysphonia (2 of 2). Moreover, CPPS *mean* discriminated laryngitis (2 of 2) and it partially differentiates edema (5 of 9), cyst (3 of 5) and sulcus (2 of 3).

A Spearman correlation coefficient of -0.61 (p-value<0.001) was found between CPPS *mean* and SampEn *mean*, thus highlighting a strong negative correlation between the two parameters, i.e. when one increases the other decreases (see Figure 3).

The Wald test p-values in the output of the logistic regression model with CPPS *mean* and SampEn *mean* as predictive variables were 0.002 and 0.874, respectively, underlying that only CPPS *mean* is a significant variable in predicting dysphonia for the present database.
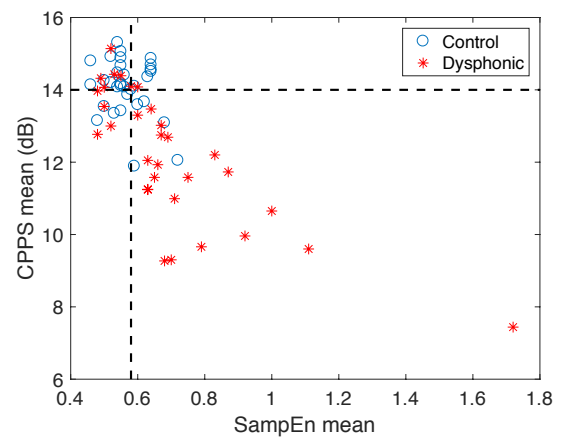

Figure 3: *Scatter plot between CPPS mean and SampEn mean: circles indicate the controls; stars represent the dysphonic group. The dashed lines indicate the optimal thresholds for the two parameters (14 dB for CPPS and 0.58 for SampEn).*

4. Conclusions

This study investigates the efficacy of Cepstral Peak Prominence Smoothed (CPPS) and Sample Entropy (SampEn) in discriminating between dysphonic and healthy subjects using excerpted vowels. The *mean* from both CPPS and SampEn distributions are the best in discriminating the two groups, but CPPS *mean* has higher diagnostic precision than SampEn *mean* (Area Under Curve of 0.85 and 0.72, respectively). A strong negative correlation between the two metrics was also found. Future studies need to consider homogeneous data (in gender and diagnosis) and larger data samples that include the perceptual assessment of voice, in order to investigate correlations between the parameters and the degree of dysphonia. The automatic method and the preliminary results of this study support in-clinic voice monitoring for the characterization of vocal health and self-monitoring of voice during everyday activities to characterize changes in the vocal behavior of professional voice users.

5. Acknowledgements

6. References

[1]   E.H. Buder, Acoustic analysis of voice quality: a tabulation of algorithms 1902–1990. In: R.D. Kent, M.J. Ball, eds. Voice Quality Measurement, San Diego, CA: Singular Publishing Group, 2000, pp.119–244.

[2]   C.R. Rabinov, J. Kreiman, B. Gerratt and S. Bielamowicz, "Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter," J. Speech Hear. Res., vol. 38, pp. 26–32, 1995.

[3]   B. DS. Bielamowicz, J. Kreiman, B. R. Gerratt, M. S. Dauer and G. S. Berke, "Comparison of voice analysis systems for perturbation measurement," J. Speech Hear. Res., vol. 39, 126–134, 1996.

[4]   J. Hillenbrand, R.A. Houde, "Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech," J. Speech Hear. Res., vol. 39, pp. 311-21, 1996.

[5]   V. Wolfe, D. Martin, "Acoustic correlates of dysphonia: type and severity," J. Commun. Disord., vol. 30, pp. 403–415, 1997.

[6]   Y.D. Heman-Ackah, D.D. Michael, G.S. Goding, "The relationship between cepstral peak prominence and selected parameters of dysphonia", J. Voice, vol.16, pp. 20–27, 2002.

[7]   Y.D. Heman-Ackah, R.J. Heuer, D.D. Michael, R. Ostrowski, M. Horman, M.M. Baroody, J. Hillenbrand, R.T. Sataloff, "Cepstral peak prominence: a more reliable measure of dysphonia," Ann. Otol. Rhinol.Laryngol., vol.112, pp. 324–333, 2003.

[8]   B. Halberstam, "Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels," J. Otorhinol. Relat. Spec., vol. 66, pp.70–73, 2004.

[9]   C. Moers, B. Möbius, F. Rosanowski, E. Nöth, U. Eysholdt, T. Haderlein, "Vowel- and text-based cepstral analysis of chronic hoarseness," J. Voice, vol.26, pp. 416–424, 2012.

[10] L.F. Brinca, P.F. Batista, A.I. Tavares, I.C. Goncalves, M.L. Moreno, "Use of Cepstral Analyses for Differentiating Normal From Dysphonic Voices: A Comparative Study of Connected Speech Versus Sustained Vowel in European Portuguese Female Speakers," J. Voice, vol. 28, pp. 282–286, 2014.

[11] A. Castellana, A. Carullo, S. Corbellini, A. Astolfi, M. Spadola Bisetti and J. Colombini, "Cepstral Peak Prominence Smoothed distribution as discriminator of vocal health in sustained vowel," in Proceedings of 2017 I²MTC Conference, May 22–25, Torino, Italy.

[12] J.S. Richman, J.R. Moorman. "Physiological time-series analysis using approximate entropy and sample entropy," Am. J. Physiol. vol. 278, H2039-49, 2000.

[13] S.M. Pincus "Approximate entropy as a measure of system complexity," Proc. Natl. Acad. Sci. USA, vol. 88, pp. 2297-2301 1991.

[14] K. Manickam, C. Moore, T. Willard, N. Slevin, "Quantifying aberrant phonation using approximate entropy in electrolaryngography," Speech Comm., 47(3), pp. 312–321, 2005.

[15] C.M. Douglas, C. Moore, K. Manickam, L. Lee, A. Sykes, A. Carr, S. Jones, J. Jones, R. Swindell, J.J. Homer, N. Slevin, "Electroglottogram approximate entropy: a novel single parameter for objective voice assessment," J. Laryngol. Otol., 124(05), pp.520-528, 2010.

[16] C. Fabris, W. De Colle, G. Sparacino, "Voice disorders assessed by (cross-) Sample Entropy of electroglottogram and microphone signals," Biomed. Sig. Proc. Cont., 8(6), 920–926, 2013.

[17] V. Parsa, D.G. Jamieson, "Acoustic discrimination of pathological voice: sustained vowels versus continuous speech," J. Speech Lang. Hear. Res., vol. 44, pp. 327–339, 2001.

[18] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy and M. De Bodt, "Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels," J. Voice, vol. 24, pp. 540–555, 2010.

[19] H. Šràmkovà, S. Granqvist C.T. Herbst and J.G. Švec , "The softest sound levels of the human voice in normal subjects," J. Acoust. Soc. Am., vol. 137 (1), pp. 407-418, 2015.

[20] Italian SpeechDat (II) FDB-3000L. (Available online at http://catalog.elra.info/product_info.php?products_id=631 last accessed March 2017)

[21] Steve Young, Julian Odell, Dave Ollason, Valtcho Valtchev and Phil Woodland, "The HTK Book", Entropic Cambridge University Laboratory, 1997

[22] B. Lindberg, F. T. Johansen, N. Warakagoda, G. Lehtinen, Z. Kačič, A. Žgank, K. Elenius, G. Salvi: "A Noise Robust Multilingual Reference Recogniser Based On Speechdat(II)," Proc. Int. Conf. Spoken Lang. Proc. (ICSLP), 2000.

[23] G. de Krom, "A cepstrum-based technique for determing a harmonics to noise ratio in speech signals," J. Speech Hear. Res., vol. 36, pp. 254-266, 1993.

[24] D.K. Lake, J.R. Moorman, C. Hanqing, "SampEn for MATLAB 1.1-1", 2012. (Available online at http://www.physionet.org/physiotools/sampen/matlab/1.1-1/ last accessed March 2017).

[25] A. De Cheveigné, H. Kawahara. "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am., vol. 111, pp. 1917–1930, 2002.

[26] M. Aboy, D. Cuesta-Frau, D. Austin, P. Mico-Tormos. "Characterization of sample entropy in the context of biomedical signal analysis," Ann. Int. Conf. IEEE Eng. Med. Bio. Soc., IEEE Eng. Med. and Bio. Soc. Conf., pp. 5943–5946, 2007.

[27] I.R. Titze, "Workshop on acoustic voice analysis: summary statement," National Center for Voice and Speech, 1995.

[28] J.D. Gibbons and S. Chakraborti, "Nonparametric Statistical Inference," Taylor & Francis, 2003, pp. 215-223.

[29] V. Bewick, L. Cheek and J. Ball, "Statistics review 13: Receiver operating characteristic curves," Critical Care, vol. 8, pp. 508-512, 2004

[30] D. Hosmer, S. Lemeshow, R. Sturdivant, "Applied Logistic Regression," third edition, Wiley, 2013.