Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (DPM)

Everthon Silva Fonseca [a,b], Rodrigo Capobianco Guido [a,*], Sylvio Barbon Junior [c], Henrique Dezani [d], Rodrigo Rosseto Gati [c], Denis César Mosconi Pereira [c]

[a] Instituto de Biociências, Letras e Ciências Exatas, Unesp - Univ Estadual Paulista (São Paulo State University), Rua Cristóvão Colombo 2265, Jd Nazareth, 15054-000, São José do Rio Preto, SP, Brazil
[b] IFSP - São Paulo Federal Institute, Department of Industry and Automation, Catanduva, SP, Brazil
[c] UEL - Londrina State University, Computer Science Department, Londrina, PR, Brazil
[d] FATEC - São Paulo State Technology College, S. J. Rio Preto, SP, Brazil

A B S T R A C T

Background: Voice disorders are related to both modest and severe health problems, including discomfort, pain, difficulty speaking, dysphagia and also cancer. Widely adopted worldwide, the combined invasive and subjective diagnosis of voice disorders is troublesome and error-prone. Contrarily, acoustic-based digital assessment allows for a non-intrusive and objective examination, stimulating the applications of computer-based tools.
Objective: Consequently, this work describes a novel algorithm to investigate speech pathologies from the sounds of sustained vowels, particularly exploring a potential gap: the classification of co-existent issues for which the major phonic symptom is the same, implying in similar inter-class features.
Method: By using the concepts of signal energy (SE), zero-crossing rates (ZCRs) and signal entropy (SH), which provide a joint time-frequency-information map, the proposed approach classifies voice signals based on the discriminative paraconsistent machine (DPM), allowing for the application of paraconsistency to treat indefinitions and contradictions.
Results: An accuracy level of 95% was obtained under a subset of voices from the *Saarbrucken voice database* (SVD), with just a modest training. In complement, the proposed approach offers wider possibilities in contrast to current state-of-the-art systems, allowing for the inputs to be mapped into the paraconsistent plane in such a way that intermediary states can be found.
Conclusion: Different from current algorithms, our technique focuses on a particular problem in the field of speech pathology detection (SPD), not yet explored in detail, proposing a way to successfully solve it. Furthermore, the results we obtained stimulate broaden studies involving speech data inconsistencies whilst providing a valid contribution.

Keywords:
Co-existent voice disorders
Overlapped inter-class features
Discriminative paraconsistent machine (DPM)
Signal energy (SE)
Zero-crossing rate (ZCR)
Signal entropy (SH)

1. Introduction

Most of the times, voice disorders are identified by speech pathologists through acoustic perceptual standards such as breathiness, hoarseness and harshness [1,2]. Nevertheless, these diagnoses depend on the professionals' audition and require a subjective identification. Pathologies with rigid characteristics may be confused with those perceptually defined as hoarse [3], for instance. The same problem occurs even when modern technologies such as video-laryngoscopy and stroboscopic light are used as some speakers may present a reflex action in their supra-glottal cavity, causing wrong assessments. Therefore, non-invasive acoustic analysis of human voices represents an important tool for physicians during the pre-diagnosis of larynx diseases, implying that discrete-time processing of recorded voice signals [4,5] is useful to detect intrinsic characteristics that define normal and pathologically-affected voices, as argued in articles [6–8].

Most of the algorithms for speech pathology detection (SPD) are based on artificial neural networks and related approaches, such as Support Vector Machines (SVMs) [9] and Gaussian Mixture Models [10], for instance. They are either designed to provide binary outputs, distinguishing between "healthy" and "pathologically-affected" voices [11], or multi-class results, identifying and possibly quantifying specific issues [12–15]. Nonetheless, the latter algo-

* Corresponding author.
E-mail address: guido@ieee.org (R.C. Guido).

rithms have not been applied to solve cases in which different pathologies co-exist and present the same major phonic symptom. The main reason for that, as we have noted, is the significant similarity among the inter-class features extracted from speech data, producing inconsistencies and hindering the classifications. Thus, this is just the gap we explore in this article: the automatic detection of multiple pathologies presenting the same major phonic symptom.

To handle the cases we want to focus on, as just mentioned, the Discriminative Paraconsistent Machine (DPM), described in [8], was adopted. Receiving inputs drawn from signal energy (SE) [16], zero-crossing rates (ZCRs) [17] and signal entropy (SH) [18], DPM allows for the application of paraconsistency [19–21] to treat indefinitions and contradictions coming from similar inter-class features. Consequently, it offers expanded possibilities in contrast to those provided by ordinary logic and regular algorithms. Thus, the proposed approach for SPD provides a contribution to the field of multi-class speech pathology classification.

For a better understanding, the remainder of this article is organized as follows. Section 2 presents a basic review on non-invasive diagnosis of voice disorders, focusing on multi-class speech pathology classification, and essential comments on DPM, SE, ZCR and SH. Proceeding, Section 3 describes the proposed approach; Section 4 shows the tests and their corresponding results; Section 5 presents a discussion and, lastly, Section 6 presents the conclusions followed by the respective references.

2. Theoretical background

2.1. Related works on multi-class speech pathology classification

Much literature has recently appeared describing different techniques for SPD. On one hand, most of them, such as those documented in references [22,24–43], are related to binary classifiers that distinguish between normal and pathologically-affected voices. On the other hand, a few articles present multi-class algorithms capable of identifying specific issues, such as those documented in references [15,44–46], as follows.

Article [15] presents a wavelet-based and biologically-inspired method to distinguish among six voice disorders, providing a 100% accuracy value in differentiating normal from pathologically-affected cases. In article [44], the authors evaluate different methods to detect multiple speech disorders based on noise content measures, spectral-cepstral modeling, nonlinear features and stability of the fundamental frequency, obtaining accuracy values varying from 81% to 99%. The authors of article [45] proposed a method based on mel-frequency cepstral coefficients, linear prediction, relative spectra combined with perceptual linear prediction, Gaussian mixture models and SVMs. The highest accuracy value of 99.98%, with a standard deviation of 0.0263, was achieved in case of detection. For multi-class pathology classification, those values become 99.85% and 0.1657, respectively. Complementarily, article [46] shows a strategy based on local discriminant bases and genetic algorithm to separate voices of subjects affected by fold polyps from those with spasmodic dysphonia and, accordingly, nodules from keratosis leukoplakia, with a mean accuracy value around 85%.

Consulting the articles reviewed and numerous records available in *Web of Science*, we found no multi-class strategy specifically designed to handle co-existing speech pathologies for which the major phonic symptom is the same. Furthermore, article [8] is the only existing reference which applies DPMs to detect voice disorders. At that time, just after introducing the new machine, the authors presented an experiment for differentiating normal from pathologically-affected speech, i.e., co-existing pathologies were not considered. Using a 60-signal database, corresponding to the sustained Brazilian Portuguese phoneme /a/, the main interest was the wavelet-transformed input visualization in the paraconsistent plane so that one could examine their characteristics in a more profound manner and successfully label them as “normal”, “pathological”, “normal tending to pathological”, or “pathological tending to normal”.

Therefore, this article provides an original contribution to the state-of-the-art analysis, extending the experiments shown in [8] while contrasting with all the others available in the literature.

2.2. DPMs

The DPM is a mechanism, based on paraconsistent logic, that allows for wider analyses in contrast to binary classifiers [8]. It properly handles two types of inconsistencies from features, i.e., indefinitions and contradictions, by projecting input data into a point in the paraconsistent plane, as shown in Fig. 1. Its limits represent one of the following cases:

- the input belongs exclusively to one class, i.e., C_1;
- the input belongs exclusively to another class, i.e., C_2;
- there is an indefinition, i.e., neither C_1 nor C_2 properly characterizes the input;
- a contradiction is characterized, i.e., the input is likely to belong both to C_1 and C_2.

Furthermore, when not exactly at the diamond limits of Fig. 1, the projected point defines intermediary situations. The axes G_1 and G_2 represent the coefficients of certainty and contradiction ($-1 \leqslant G_1$, $G_2, \leqslant 1$), respectively, on a certain input vector whilst considering a specific statement. Taking advantage of such an ability, the proposed approach uses DPM to label an input voice as impressed by one specific pathology, i.e., C_1; another, i.e., C_2; neither or, lastly; both, i.e., C_1 and C_2, allowing for indefinitions and contradictions to be quantified.

Since DPM is fully described in [8], detailing its architecture and training procedures, no further information is included in this section, except for one important aspect: the capacity to treat inconsistencies efficiently comes from the need of only a very modest set for training in comparison with other algorithms such as SVMs and deep neural networks [47]. Thus, finding just a small subset of consistent features inside a big inconsistent group, as happens in the proposed approach, allows for the required training.

2.3. SE, ZCR and SH

As shown in article [16], which serves as a tutorial on energy and defines three methods to take advantage of such a concept, namely A_1, A_2 and A_3, the quantity SE reflects the potential to perform work [48]. Thus, if $s[\cdot]$ of length M is the digital signal which stores a voice signal,

$$\mathrm{SE}(s[\cdot]) = \sum_{i=0}^{M-1} (s_i)^2 \quad .$$

In the proposed approach, which uses method A_3, we are interested in analysing how the subjects' acoustic mechanisms sustain their total capacity while producing sounds over time. To use this method, $\mathcal{C}$ is defined as the critical base-level of energy, ($0 < \mathcal{C} < 100\%$), and, then, the feature vector $f[\cdot]$, of size

$$T = \begin{cases} \dfrac{100}{\mathcal{C}} - 1 & \text{if } \mathcal{C} \text{ is multiple of } 100, \\ \lfloor \dfrac{100}{\mathcal{C}} \rfloor & \text{otherwise,} \end{cases}$$

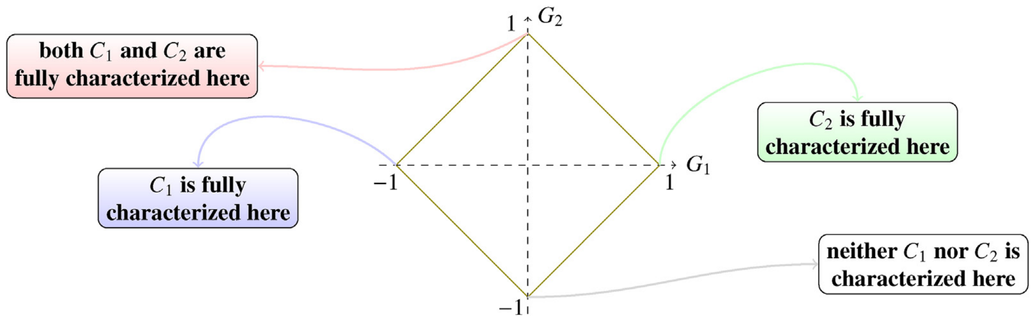


Fig. 1. The output plotted in the paraconsistent plane. Particularly, $(G_1, G_2) = (-1, 0)$, $(G_1, G_2) = (1, 0)$, $(G_1, G_2) = (0, -1)$ and $(G_1, G_2) = (0, 1)$ represent an input purely matching class C_1, solely corresponding to class C_2, an uncertainty, i.e., neither class, and a contradiction, i.e., both classes, respectively.

is determined as follows:

- f_0 is the percentage of $s[\cdot]$ length, starting from its beginning, required to reach $\mathcal{C}$% of the total signal energy;
- f_1 is the percentage of $s[\cdot]$ length, starting from its beginning, required to reach $2\mathcal{C}$% of the total signal energy;
- f_2 is the percentage of $s[\cdot]$ length, starting from its beginning, required to reach $3\mathcal{C}$% of the total signal energy;
- ...
- f_{T-1} is the percentage of $s[\cdot]$ length, starting from its beginning, required to reach $(T\mathcal{C})$% of the total signal energy, so that $(T\mathcal{C}) <$ 100%;

Essentially, A_3 works in the following manner: the more stable the level of energy along $s[\cdot]$ is, the more aligned the points in $f[\cdot]$ are, and vice-versa, as shown in the hypothetical example of Fig. 2. Thus, A_3 is used in this work as the basis to investigate the stability of phonation intensity.

Proceeding, reference [17] consists of a tutorial on ZCR and presents B_1, B_2 and B_3 as being time-domain methods defined to get a basic description of the fundamental frequencies which compose $s[\cdot]$ so that

$$\mathrm{ZCR}(s[\cdot]) = \frac{1}{2}\sum_{j=0}^{M-2} |sign(s_j) - sign(s_{j+1})| \quad ,$$

where $\mathrm{ZCR}(s[\cdot]) \geqslant 0$ and $sign(y) = \begin{cases} 1 & \text{if } y \geqslant 0; \\ -1 & \text{otherwise} \end{cases}$. Specifically, the proposed approach adopts B_3 to convert from $s[\cdot]$ of size M to $f[\cdot]$ of size T. It works exactly as A_3 but considers ZCRs instead of SEs. Thus, B_3 is used here to analyse the spectral stability of phonation.

Lastly, we reviewed entropy, focusing on the concepts explained in [18,49]. It consists of a signal predictability measure originally defined as

$$\mathrm{SH} = -\sum_{i=0}^{K-1} p_i \cdot \log_\beta(p_i) \quad ,$$

where p_i is the probability of the i^{th} symbol, or amplitude, in a set of K different symbols, or signal amplitudes, and β is the basis used. In this work, we adopted an alternative definition for SH in which p_i becomes the proportional energy of the i^{th} sample contained in the input signal [50], being $(0 \leqslant p_i \leqslant 1)$. Specifically, $s[\cdot]$ of length M is investigated over a sliding window of length L, with a possible overlap of V% between its consecutive placements, so that the feature vector, $f[\cdot]$, of length $T = \lfloor \frac{100 \cdot M - L \cdot V}{(100-V) \cdot L} \rfloor$ contains the corresponding entropies. Hereafter and respecting the analogy, this method is referred to as C_3. It is particularly used in this work to calculate the information required to codify the speaker's voice over time, as previously shown [24].

3. The proposed approach

The proposed approach consists of a feature extraction (FE) step followed by a paraconsistent evaluation (PE), as shown in Fig. 3. FE corresponds to the extraction of three features from each signal under analysis to form the feature vector $X[\cdot] = \{x_1, x_2, x_3\}$, as shown in Fig. 4. On one hand, method C_3 is carried out over half a second of the input signal $s[\cdot]$, particularly at its central part due to stability during vocalisation. On the other hand, A_3 and B_3 are performed considering only 50 milliseconds only over the same region. The specific algorithm is:

- BEGINNING
- STEP 1: obtain the standard deviation [23] over $\partial(f[\cdot])$, i.e.,

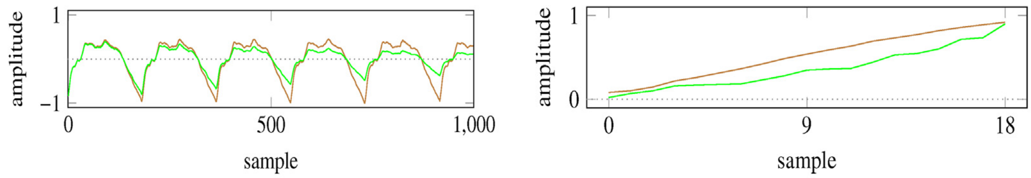
$$x_1 = \sigma(\partial(f[\cdot])) = \sqrt{\frac{\sum_{i=0}^{T-1}\left(\partial(f_i) - \frac{1}{T}\sum_{j=0}^{T-1}\partial(f_j)\right)^2}{T}},$$

  where vector $\partial(f[\cdot])$ of size $T-1$ constitute the first-order derivative of the T-sample long vector $f[\cdot]$, which was primarily calculated based on A_3 with $\mathcal{C} = 1\%$ and using the voice signal $s[\cdot]$ of size M as input. Thus, as explained in section 2 and in article [16], A_3 converts from $s[\cdot] = \{s_0, s_1, .., s_{M-1}\}$ to $f[\cdot] = \{f_0, f_1, f_2, .., f_{T-1}\}$ and, then, we use the latter vector to calculate x_1, where the derivative of $f[\cdot]$ can be obtained just by computing the differences between its consecutive samples, i.e., $\{f_1 - f_0, f_2 - f_1, f_3 - f_2, \ldots\}$;
- STEP 2: repeat exactly the same procedure to obtain x_2, but adopting B_3 instead of A_3;
- STEP 3: obtain the standard deviation calculated directly over $f[\cdot]$, i.e.,

$$x_3 = \sigma(f[\cdot]) = \sqrt{\frac{\sum_{i=0}^{T-1}\left(f_i - \frac{1}{T}\sum_{j=0}^{T-1} f_j\right)^2}{T}},$$

  where the feature vector was first attained based on C_3 with parameters $V = 50\%$, $\beta = 10$ and L chosen to cover about 50 ms of the input signal at each placement, which is a proper option [51].
- END.

Features x_1, serving as an alternative measure of signal shimmer [5], and x_2, used alternatively to express signal jitter [5], quantify the degree of linearity with which SE and ZCRs vary along $s[\cdot]$, reflecting the speakers' ability to maintain their acoustic power and to keep their vocal folds vibration pattern, respectively, as time runs. Complementarily, x_3, obtained on the basis of SH, expresses how stable the information required to codify the speaker's voice over time is. Furthermore, as explained in articles [16–18], such features encompass, jointly, temporal and spectral details about $s[\cdot]$, simi-



(a) original signals under analysis (b) corresponding feature vectors obtained based on A_3 using $C = 5\%$

Fig. 2. (a): Two speech signals superimposed, where the brown and green curves represent a stable and a decaying intensities, respectively. (b): the corresponding feature vectors, plotted with the same colors, obtained by using A_3, where the points in the brown curve are practically aligned due to the stability of the corresponding speech signal in (a).

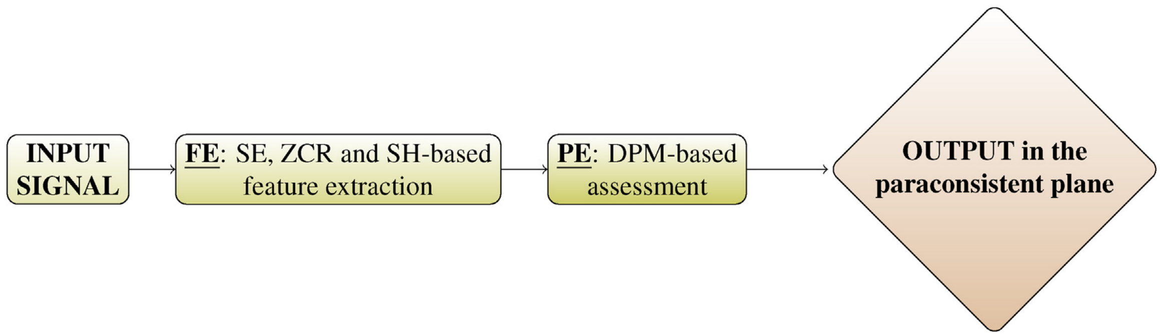


Fig. 3. The proposed approach representation, where FE and PE are the active parts.

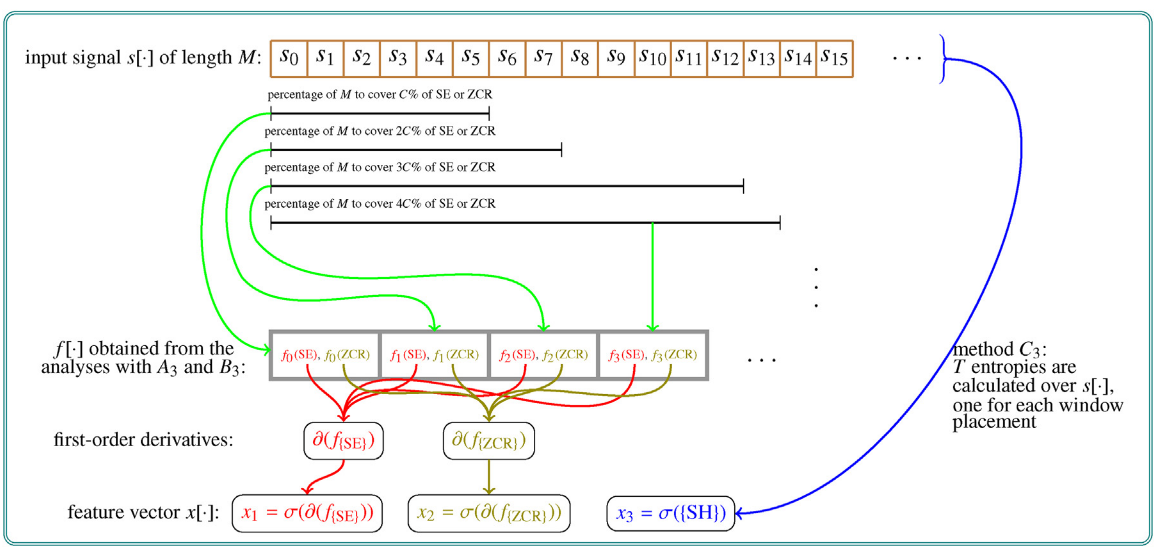


Fig. 4. Overview of the proposed approach for FE.

larly to the Discrete Wavelet Transform [51–53]. And so, as shown ahead, the set {SE,ZCR,SH} is relevant for the proposed approach, justifying our choice.

Once FE is over, PE takes place. Specifically, R feature vectors representing each one of the classes, C_1 and C_2, as discussed ahead, are used to train a DPM according to the procedures detailed in [8]. Since training is supervised [54], there is a need to use two labels for each input: $labC_1$ and $labC_2$. The former is set either to 1 or −1 to denote whether the input belongs, or not, to class C_1, respectively. Accordingly, 1 or −1 means whether the input matches, or not, class C_2, respectively. Notably, if the input belongs to both classes, both $labC_1$ and $labC_2$ are set to 1. Oppositely, if neither C_1 nor C_2 matches the input, the labels are set to −1. Thus, if C_1 and C_2 represent the pathologies P_1 and P_2, respectively, assigning 1 to both labels indicates the subject is influenced by both P_1 and P_2. In contrast, assigning −1 to both indicates that neither P_1 nor P_2 affects the subject. In the proposed approach, we particularly adopted $labC_1 = labC_2 = -1$ to label the healthy subjects.

Just after training, by using a personal computer running for a few milliseconds to carry out this task, the remaining vectors not used in such a procedure can be assessed in a way that, for each input, the output corresponds to a point (G_1, G_2) in the paraconsistent plane, interpreted as previously explained and in article [8]. All the procedures were implemented using MatLab 7.1 and complemented with tests in C/C++ programming language.

4. Tests and results

4.1. Corpus and feature behaviour

To assess the corresponding experiments using DPM, the Saarbrucken voice database (SVD) [56], organized by speech pathologists from the *Institut Für Fonetik* of the *Universität des Saarlandes*, was adopted because it contains data from subjects with multiple pathologies. Initially, our investigation involved four groups: class C_1 with 34 voices from subjects affected only by Reinke edema, class C_2 containing 82 files from patients exclusively with laryngitis, 687 utterances from a group with healthy subjects and, lastly, 10 recording from people with both Reinke edema and laryngitis. During phonation, the speakers sustained, among others, a phoneme sounding as /a/ in the American English word *dogma* for a few seconds at a normal level. That utterance was selected due to its higher information content needed to catalogue voice disorders, according to previous reports [51]. The sampling of all signals were carried out at 50000 samples per second, 16-bit, mono-channel, and stored in wave files, implying that they contain frequencies within the range [0–25000] Hz, according to Nyquist's theorem [52].

Notably, the major phonic symptom of Reinke edema and laryngitis is the same: hoarseness caused by vocal folds swelling. Thus, their acoustic-based distinction is a challenge, justifying the use of paraconsistency. This is what Fig. 5-(i,ii and iii) confirms. Its subfigures indicate significant overlaps among the values of x_1, x_2 and x_3, considering all the $34+82+687+10=813$ vocalizations from the four groups. Interestingly, Fig. 5-(iii) shows a quite relevant aspect: the more affected the subject is, the greatest the standard deviation over SH. In other words, healthy subjects are much more capable of maintaining the information required to codify their voices along phonation than those pathologically-affected, considering the following ascending scale of health: both pathologies, i.e., the worst case, followed by Reinke edema, then laryngitis and, lastly, normality as the best condition.

4.2. The experiment

$R=40$ speakers' sound clips were randomly chosen to be used in conjunction with the DPM: 10 from the 34 subjects affected by Reinke edema to represent class C_1, 10 from the 82 individuals affected by laryngitis to represent class C_2, 10 from the 687 healthy people and, lastly, 10 from patients affected by both pathologies. Although 40 is a modest number, the absence of speakers affected by multiple pathologies does not allow for a wider experiment without unbalancing the classes during training. And since the database contains only 10 subjects affected both by Reinke edema and laryngitis, this is the limit adopted.

Proceeding, we carried out the tests on the basis of a hold-out cross-validation procedure [54], repeating the above-described casual selection ten thousand times. From each repetition, five feature vectors from C_1, five from C_2, five from the speakers with no pathology and five from both classes, totaling 20, were used to train the DPM. Then, the remaining 20 vectors were used to test the system. For training, each input was assigned to the proper labels, informing whether or not a match with C_1, a match with C_2, no match or both matches, occurs.

Fig. 6, containing 20 plotted marks where some of them are superimposed, shows both the worst and the best results. Considering the type of analysis we desire, interesting outcomes were obtained, i.e., the closer (G_1, G_2) is to the corners $(-1, 0)$, $(1, 0)$, $(0, -1)$ and $(0, 1)$, the more the input is likely to match C_1, C_2, none or both, respectively. In other words, the distances from (G_1, G_2) to the corners reveal how likely the input matches each one of the four states. For both the subfigures in Fig. 6, which are very similar and thus justify the absence of more specific statistical analyses such as confidence intervals and confusion matrices, the results are in accordance with the dataset remarks, i.e., stronger issues place the point (G_1, G_2) closer to the corresponding axes than modest problems do.

Although there is one Ω closer to the Δs than to its own group in Fig. 6-(a,b), implying in one subject affected by both pathologies misclassified as affected only by Reinke edema and resulting in a value of accuracy of $\frac{19}{20}=95\%$, we consider an excellent outcome was obtained because inter-class features overlap significantly, as previously shown. The misclassified sample corresponds to a severe case of Reinke edema, much more significant than the laryngitis, justifying the distribution of the symbols in the paraconsistent plane. Furthermore, we observe that no pathologically-affected subject was misclassified as being healthy, reassuring the efficacy of the proposed approach.

5. Discussion and comparisons

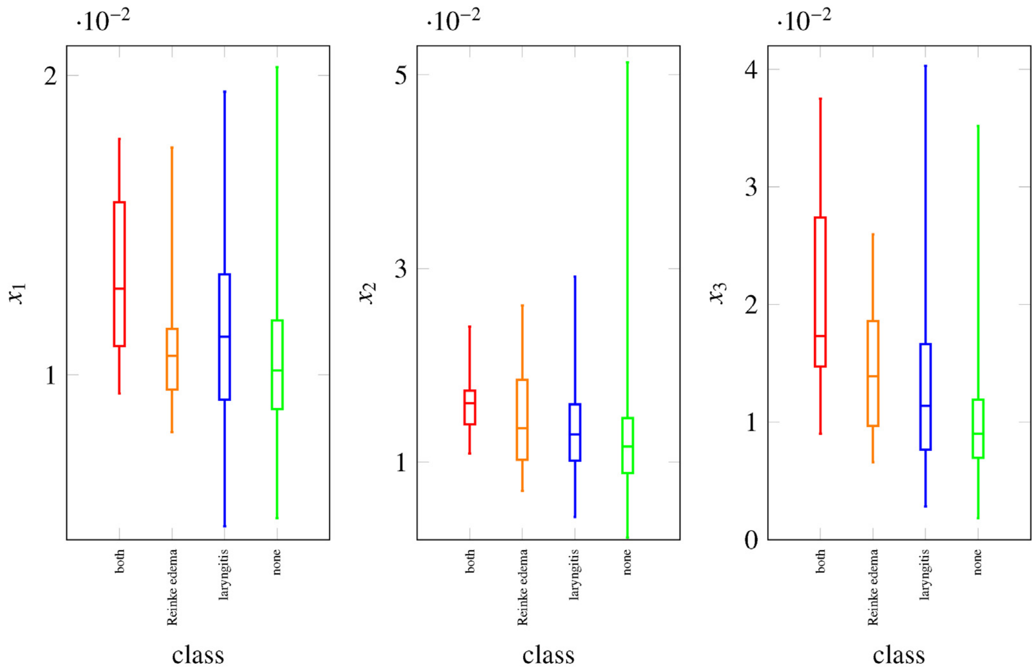
Speech pathologists state that voice disorders affect a considerable number of people, with significant lifetime prevalence [55]. From discomfort and pain to swallowing malfunction and cancer, a diversity of problems might appear in connection to larynx pathologies. Oncologists agree that vocal folds cancer, for instance, appears as small areas of abnormal cells that multiply and consequently undergo subsequent changes, ultimately leading to the development of a significant problem [57,58]. Thus, we argue this study to be of paramount importance.

Focusing on a non-intrusive and acoustic inspection, we have concentrated our efforts on the analysis of co-existent issues for which the major phonic symptom is the same, as previously mentioned. With this in mind, we refrain from establishing a detailed comparison between the proposed approach and current algorithms, because their objectives are relatively different, i.e., although state-of-the-art articles such as [59–61] report accuracies between 94.93% and 100% over SVD, no literature has been found with a specific focus on the problem considered herein.

Nevertheless, to allow for just a basic comparison, engaging readers into a discussion, a structure consisting of two SVMs [62] with radial basis function (RBF) [63] kernels was created, as shown in Fig. 7. They received the same feature vectors used with DPM as input and, adopting the same hold-out cross-validation procedure, were trained in a supervised manner so that each machine should output either 1 or -1 to represent whether or not the input belongs to a certain class, respectively. In the best scenario, a value of accuracy of $\frac{18}{20}=90\%$ was obtained, where two individuals affected both by Reinke edema and laryngitis were misclassified as a case of Reinke edema only.

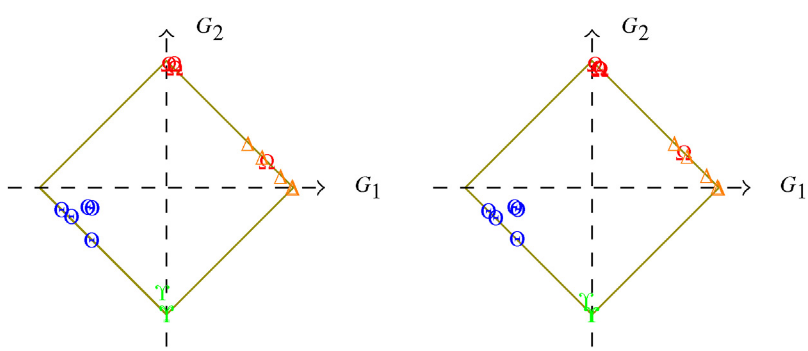
Notably, in Fig. 7, each SVM consists of a three-layer structure: three passive neurons receiving the feature vectors as inputs, 20 active RBF neurons in the hidden layer and, lastly, one linear neuron at the output. The dimension of the hidden layer is equal to the number of training examples to allow for an one-step convergence while adjusting the weights of the final layer in an supervised manner. The hidden layer, which has no weight, was trained by adjusting its kernels using an unsupervised procedure in such a way that its *i*th unit outputs 1 to the *i*th training example. The outputs $\{SVM_1, SVM_2\} \approx \{-1,-1\}, \{-1,1\}, \{1,-1\}$ and $\{1,1\}$ indicate a healthy subject, laryngitis, Reinke edema and both pathologies, respectively.

Despite the fact that we almost obtained the same accuracy value with both approaches, i.e., 95% for the DPM-based and 90% for the SVM-based, different aspects of input data are explored by each one of them: the latter does not allow for inconsistencies to be clearly visualized in the paraconsistent domain while the former was just designed to do so [8]. Furthermore, advanced state-of-the-art multi-class algorithms, such as those based on deep learning [47,65,66], can not be successfully used to solve the problem we presented in this article due to a lack of training examples, i.e., five feature vectors representing each class are not enough for an effective learning curve. Even SVMs, used here for basic comparisons, do not generalize adequately with such an insignificant number of fea-



(a) box plots for x_1, related to SE (b) box plots for x_2, related to ZCR (c) box plots for x_3, related to SH

Fig. 5. (i), (ii) and (iii): The way x_1, x_2 and x_3 vary for the four classes, respectively.



(a) worst scenario (b) best scenario

Fig. 6. The outputs plotted in the paraconsistent plane for the practically-indistinguishable worst and best scenarios. The symbols Ω, Δ, Θ, Υ represent subjects affected by both pathologies, with just with Reinke edema, with only with laryngitis and, lastly, healthy individuals, respectively. Please, recall that there are many superimposed marks.

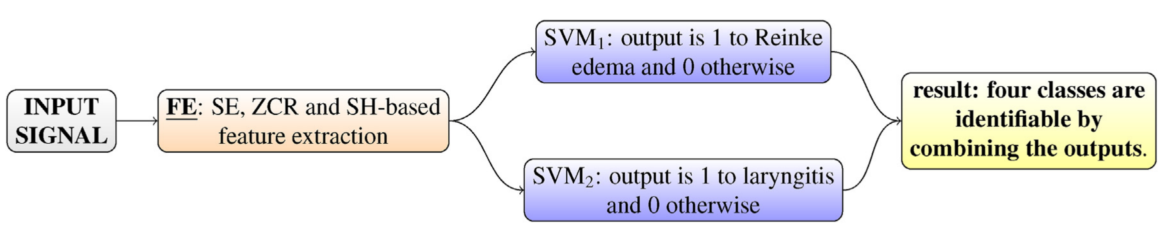


Fig. 7. SVM-based strategy used for comparisons with the proposed approach.

ture vectors [51]. The same holds true for the techniques presented in articles just mentioned, i.e., [59–61].

In contrast, using a modest training set, as shown in the previous section and exemplified in [8]-pp.399, DPM learned and allowed for their inputs to be visualized in the paraconsistent plane so that four extreme and numerous intermediary outcomes could be detected. Additionally, indefinitions and contradictions caused by overlapped inter-class features were successfully handled [8].

Lastly, it is interesting to note that most of the standard multi-label classification algorithms, such as those applied in the field of speech pathology [15,44–46] and reviewed in section 2, were not particularly designed to consider the inter-class correlations associated with the feature vectors. To deal with this important information, a few complex and sophisticated strategies such as model stacking and chaining have been used [64], although with a considerable computational complexity. In contrast, DPM is a straightforward tool capable of treating such an issue, supported by the whole domain, without a complex structure [8].

6. Conclusions

Just after a brief study on recent strategies for cataloguing voice disorders, as well as a brief review on important related concepts, a novel algorithm to inspect human vocal apparatus was presented. It focuses on the detection of co-existent laryngeal issues for which the major phonic symptom is the same, producing features with significant inter-class overlaps. Based on the combination of SE, ZCR and SH, all used for feature extraction, associated with DPM, specifically adopted for classification, the proposed approach was successfully concluded, efficiently handling indefinitions and contradictions with a value of accuracy of 95%. One limitation of this work, imposed by the only public database available with the characteristics needed, consists of the modest number of speech files used for testing it. Therefore, future research will focus on experimenting DPM using a larger corpus, as soon as it becomes available, particularly containing a more relevant number of speech files from subjects with co-existent issues for which the major phonic symptoms are the same.

Conflict of interest

None declared.

Acknowledgements

We are very grateful to "Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)" and "The State of São Paulo Research Foundation (FAPESP)", in Brazil, for the grants provided, respectively through the processes 306811/2014-6 and 2019/04475-0, to support this research.

References

[1] D.M. Bless, Measurement of vocal function, Otolaryngol. Clin. N. Am. 24 (1991) 1023–1033.
[2] R.J. Baken, Clinical Measument of Speech and Voice, Singular Publishing Group Inc, San Diego, 1996.
[3] N. Isshiki, et al., Differential diagnosis of hoarseness, Folia Phoniatrica 21 (1969) 9–19.
[4] J.R. Deller Jr., J.G. Proakis, J.H.L. Hansen, Discrete-Time Processing of Speech Signals, MacMillan Publishing Co, New York, 1993.
[5] L. Deng, D. O'Shaughnessy, Speech Processing: A Dynamic and Optimization-Oriented Approach, Dekker, New York, 2003.
[6] P. Mitev, S. Hadjitodorov, Fundamental frequency estimation of voice of patients with laryngeal disorders, Inform. Sci. 156 (1-2) (2003) 3–19.
[7] J.M. Rees, et al., Adaptation of wavelet transform analysis to the investigation of biological variations in speech signals, Med. Eng. Phys. 30 (2008) 865–871.
[8] R.C. Guido, et al., Introducing the Discriminative Paraconsistent Machine (DPM), Inform. Sci. 221 (2012) 389–402.
[9] C. Cortes, V. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.
[10] J.I. Godino-Llorente, P. Gómez-Vilda, M. Blanco-Velasco, Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters, IEEE Trans. Biomed. Eng. 53 (2006) 1943–1953.
[11] E.S. Fonseca, J.C. Pereira, Normal versus Pathological Voice Signals: Using Wavelet Analysis and Support Vector Machines, IEEE Eng. Med. Biol. Mag. 28 (5) (2008) 44–48.
[12] J. Godino-Llorente, The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders, J. Voice 24 (2010) 47–56.
[13] J.I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, S. Aguilera-Navarro, P. Gómez-Vilda, An integrated tool for the diagnosis of voice disorders, Med. Eng. Phys. 28 (2006) 276–289.
[14] J.I. Godino-Llorente, P. Gomez-Vilda, F. Cruz-Roldan, M. Blanco-Velasco, R. Fraile, Pathological likelihood index as a measurement of the degree of voice normality and perceived hoarseness, J. Voice 24 (2010) 667–677.
[15] P. Salehi, The separation of multi-class pathological speech signals related to vocal cords disorders using adaptation wavelet transform based on lifting scheme, Cumhuriyet Sci. J. 36 (3) (2015) 2371–2382.
[16] R.C. Guido, A Tutorial on Signal Energy and its Applications, Neurocomputing 179 (2016) 264–282.
[17] R.C. Guido, ZCR-aided Neurocomputing: a study with applications, Knowledge-based Systems 105 (2016) 248–269.
[18] R.C. Guido, A tutorial-review on entropy-based handcrafted feature extraction for information fusion, Inform. Fusion 41 (2018) 161–175.
[19] B. Brown, On paraconsistency, in: D. Jacquette (Ed.), A Companion to Philosophical Logic, Blackwell, Oxford, 2002, pp. 628–650.
[20] N.C.A. Costa, D. Krause, O. Bueno, Paraconsistent logics and paraconsistency, In: Handbook of the Philosophy of Science (2007) 791–911.
[21] J.M. Abe, J.C.A. Prado, K. Nakamatsu, Paraconsistent artificial neural network: applicability in computer analysis of speech productions, Lecture Notes Comput. Sci. 4252 (2006) 844–850.
[22] A. Akbari, M.K. Arjmandi, An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features, Biomed. Signal Process. Control 10 (2014) 209–223.
[23] G.L. Shevlyakov, H. Oja, Robust Correlation: Theory and Applications, in: Wiley Series in Probability and Statistics (Book 3), 2016.
[24] A. Al-Nasheri, et al., Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions, IEEE Access 6 (2017) 6969–6974.
[25] Z. Ali, et al., Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model, J. Voice 30 (6) (2016), 757.e7-757.e19.
[26] M.K. Arjmandi, et al., Identification of voice disorders using long-time features and support vector machine with different feature reduction methods, J. Voice 25 (6) (2011) e275–e289.
[27] A. Castellana, et al., Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel, IEEE Trans. Instrument. Measure. 67 (3) (2018) 646–654.
[28] H. Dahmaniae, et al., Assessment of dysarthric speech through rhythm metrics, Comput. Inform. Sci. 25 (2013) 43–49.
[29] E.S. Fonseca, et al., Linear prediction and discrete wavelet transform to identify pathologies in voice signals, IEEE Signal Process. Symposium Proc. (2017) 1–4.
[30] O. Lachhab, J. Di Martino, E.H.I. Elhaj, A. Hammouch, Improving the recognition of pathological voice using the discriminant HLDA Transformation., 3rd IEEE International Coloquium in Information Science and Technology (CIST) (2014) 370–373.
[31] M. Markaki, Y. Stylianou, Voice pathology detection and discrimination based on modulation spectral features, IEEE Trans. Audio Speech Lang. Process. 19 (7) (2011) 1938–1948.
[32] D. Martinez, et al., Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit, in: Advances in Speech and Language Technologies for Iberian Languages, Springer, Berlin, Germany, 2012, pp. 99–109.
[33] J. Mekyska, et al., Robust and complex approach of pathological speech signal analysis, Neurocomputing 167 (2015) 94–111.
[34] G. Muhammad, M. Melhem, Pathological voice detection and binary classification using MPEG-7 audio features, Biomed. Signal Process. Control 11 (2014) 1–9.
[35] S.D. Pranav, M. Sabarimalai, Effective glottal instant detection and electroglottographic parameter extraction for automated voice pathology assessment, IEEE J. Biomed. Health Inform. 22 (2) (2017) 398–408.
[36] N.E. Saeedi, F. Almasganj, Wavelet adaptation for automatic voice disorder sorting, Comput. Biol. Med. 43 (6) (2013) 699–704.
[37] A. Sasou, Automatic identification of pathological voice quality based on the GRBAS categorization., Asia-Pacific and Information Processing Association Annual Summit and Conference (APSIPA ASC), Malaysia (2017) 1243–1247.
[38] S.E. Shia, T. Jayasree, Detection of pathological voices using discrete wavelet transform and artificial neural networks., IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (2017) 1–6.
[39] N. Sripriya, et al., Non-intrusive technique for pathological voice classification using jitter and shimmer., International Conference on Computer, Communication and Signal Processing (ICCCSP) (2017) 1–6.
[40] R.K. Sudarsana, B. Yegnanarayana, Epoch extraction from emotional speech using single frequency filtering approach, Speech Commun. J. 86 (2017) 52–63.
[41] L. Verde, G. Pietro, G. Sannino, A methodology for voice classification based on the personalized fundamental frequency estimation, Biomed. Signal Process. Control 42 (2018) 134–144.
[42] C.M. Vikram, K. Umarani, Phoneme independent pathological voice detection using wavelet bases, MFCCs and GMM-SVM hybrid classifier, In: International Conference on Advances in Computing, Communications and Informatics (ICACCI) (2013) 153–156.
[43] Z. Zhong, et al., Nonlinear signal processing for vocal folds damage detection based on heterogeneous sensor network, Signal Process. 126 (2016) 125–133.
[44] J.R. Orozco-Arroyave, et al., Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases, IEEE J. Biomed. Health Inform. 19 (6) (2015) 1820–1828.
[45] M.A. Mojaly, G. Muhammad, M. Alsulaiman, Detection and classification of voice pathology using feature selection, In: IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), Doha (2014) 571–577.

[46] P.T. Hosseini, F. Almasganj, M.R. Darabad, Pathological Voice Classifcation Using Local Discriminant Basis and Genetic Algorithm, In: 16th Mediterranean Conference on Control and Automation Congress Centre, France (2008) 872–876.
[47] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press, 2016.
[48] E. Sejdic, I. Djurovic, J. Jiang, Time-frequency feature representation using energy concentration: an overview of recent advances, Digital Signal Process. 19 (2009) 153–183.
[49] J.V. Stone, Information Theory: A Tutorial Introduction, Sebtel Press, 2015.
[50] P.S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, 1.ed., CRC Press, 2002.
[51] E.S. Fonseca, et al., Wavelet time-frequency analysis and least squares support vector machine for the identification of voice disorders, Comput. Biol. Med. 37 (4) (2007) 571–578.
[52] D. Sundararajan, Discrete Wavelet Transform: A Signal Processing Approach, Wiley, 2016.
[53] R.C. Guido, Practical and Useful Tips on Discrete Wavelet Transforms, IEEE Signal Process. Mag. 32 (3) (2015) 162–166.
[54] R.O. Duda, P.E. Hart, D.G. Stork, Pattern classification, John Wiley & Sons, 2012.
[55] N. Roy, et al., Voice disorders in the general population: prevalence, risk factors, and occupational impact, Laryngoscope 115-11 (2005) 1988–1995.
[56] Saarbruecken voice database. Available at http://stimmdb.coli.uni-saarland.de/.
[57] https://www.hopkinsmedicine.org/healthlibrary/conditions/otolaryngology/vocal_cord_cancer_22,vocalcordcancer. Accessed on August, 2018.
[58] https://voicefoundation.org/health-science/voice-disorders/voice-disorders/laryngeal-advanced-cancer/symptoms-advanced-laryngeal-cancer/. Accessed on August, 2018.
[59] A. Al-Nasheri, et al., Voice pathology detection using auto-correlation of different filters bank., in: 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar IEEE, 2014, pp. 50–55.
[60] D. Hemmerling, A. Skalski, J. Gajda, Voice data mining for laryngeal pathology assessment, Comput. Biol. Med. 69 (2016) 270–276.
[61] A. Tsanas, Acoustic analysis toolkit for biomedical speech signal processing: concepts and algorithms, Models Anal. Vocal Emissions Biomed. Appl (2013) 37.
[62] R.F. Mello, M.A. Ponti, Machine Learning: A Practical Approach on the Statistical Learning Theory, Springer, 2018.
[63] M.D. Buhmann, Radial Basis Functions: Theory and Implementations, Cambridge University Press, 2009.
[64] S.M. Mastelini, et al., Multi-output tree chaining: an interpretative modelling and lightweight multi-target approach, J. Signal Process. Syst. (2018) 1–25.
[65] P. Harar, et al., Voice pathology detection using deep-learning: a preliminary studying., in: IEEE International Conference and Workshop on Bioinspired Intelligence (IWOB), Funchal - Portugal, 2017, pp. 1–4.
[66] X. Zhang, et al., Pathological voice recognition by deep neural network., IEEE 4th International Conference of Systems and Informatics (ICSAI) (2017) 464–468.