A Pattern Recognition Approach to Spasmodic Dysphonia and Muscle Tension Dysphonia Automatic Classification

*Gastón Schlotthauer, *María Eugenia Torres, and †,‡María Cristina Jackson-Menaldi, *Entre Ríos Argentina, †St. Clair Shores, Michigan, ‡Detroit, Michigan

Summary: Spasmodic dysphonia (SD) and muscle tension dysphonia (MTD) are two voice disorders that present similar characteristics. Usually, they can be differentiated only by experienced voice clinicians. There are many reasons that support the idea that SD is a neurological disease, requiring surgical treatments or, more usually, laryngeal botulinum toxin A injections as a therapeutic option. On the other hand, MTD is a functional disorder correctable with voice therapy. The importance of a correct diagnosis of these two disorders is critical at the treatment-selection moment. In this article, we present and compare the results of neural network and support vector machine-based methods that can help the clinicians to confirm their diagnosis. As a preliminary approach to the problem, we used only a sustained vowel /a/ to extract eight acoustic parameters. Then, a pattern recognition algorithm classifies the voice as normal, SD, or MTD. For comparison with previous works, we also separated the voices into normal and pathological (SD and MTD) voices with the methods proposed here. The results overcome the best classification rates between normal and pathological voices that have been previously reported, and demonstrate that our methods are very effective in distinguishing between MTD and SD.

Key Words: Spasmodic dysphonia–Muscle tension dysphonia–Neural networks–Support vector machines.

INTRODUCTION

The quality of life of patients with voice disorders is seriously affected as a direct consequence of their pathologies, causing psychological and emotional problems.[1] Two pathologies with consequences on the communication-related quality of life of patients are spasmodic dysphonia (SD) and muscle tension dysphonia (MTD). SD is a larynx focal dystonia. A dystonia is a disorder of the central nervous system, in which there is an increased contraction of the muscles.[2,3] There are several types of SDs. The most common SD is called adductor SD (AdSD), and it occurs when the muscles that bring the vocal folds together contract too strongly. This disorder is characterized by a very strained or strangled voice quality, with occasional voice stoppages, or breaks, when the air cannot escape. These breaks are more evident during speech associated with voiced sounds and when initiating sustained phonation. Abductor SD (AbSD) is less common. In this disorder, the spasms occur in the muscles that open the vocal folds. For this reason, it is difficult to bring the vocal folds together to produce voiced sounds.

SD is a rare disorder. For example, it has an incidence of one case per 100 000 in Munich,[4,5] and it is estimated to affect 30 000–50 000 people in North America.[6] The true incidence may be greater, because the diagnosis is often missed.[4] Because of its rarity, many physicians are not familiar with this voice pathology. As a result, individuals with SD often are told that their voice disorder is because of nervousness, laryngitis, reflux, etc. The onset of voice dystonia occurs slowly over a period of several months to a year, and two-thirds of the affected patients are females.[7] SD, a true dystonia, and laryngeal tremor are often confused.[8] Severe cases of vocal tremor may cause speech breaks similar to those of AdSD. Patients with SD may present vocal tremor or MTD associated with it.[9,10] Voice can be normal during laughing, coughing, crying, or other involuntary vocal use or singing.

Patients with MTD exhibit excessive muscular tension while speaking. This voice dysfunction is not associated with abnormalities of the laryngeal structures. When the muscles associated with speech production lose some of their coordination or contract inappropriately, they can produce a hoarse voice, neck pain, neck fatigue, and even complete loss of the voice. Diagnosis of MTD can be difficult, because the vocal folds actually have a relatively normal appearance at rest according to Verdolini et al.[11] The classification MTD represents a persistent, unexplained dysphonia that is behaviorally modifiable.[11] It is only during speech tasks that the abnormal contraction of the muscles is seen. The key treatment for MTD is voice therapy.[2,3,11] However, MTD can mimic the strained, effortful voice characteristics of AdSD, leading to diagnostic confusion and possibly inappropriate management.[12]

Diagnosis of SD is generally based on auditory-perceptual characteristics. It is considered on the basis of history and physical examination. Workup may include magnetic resonance imaging (MRI) of the brain, laryngeal electromyography (EMG), laboratory test necessary for dystonia patients, neurological assessment, voice assessment using protocol that includes continuous and stroboscopic light with flexible and rigid scopes with different repetitive phonatory tasks, and

Accepted for publication October 8, 2008.
From the *Laboratorio de Señales y Dinámicas no Lineales, Facultad de Ingeniería Universidad Nacional de Entre Ríos Oro Verde, Entre Ríos, Argentina; †Lakeshore Professional Voice Center of the Lakeshore Ear, Nose and Throat Center, St. Clair Shores, Michigan; and the ‡Department of Otolaryngology, School of Medicine, Wayne State University, Detroit, Michigan.
Address correspondence and reprint requests to Gastón Schlotthauer, Laboratorio de Señales y Dinámicas no Lineales, Facultad de Ingeniería Universidad Nacional de Entre Ríos Oro Verde, Entre Ríos 3100, Argentina. E-mail: gschlott@bioingenieria.edu.ar
Journal of Voice, Vol. 24, No. 3, pp. 346-353
0892-1997/$36.00

doi:10.1016/j.jvoice.2008.10.007

objective and perceptual voice analysis. The differential diagnosis of AdSD and MTD can often be difficult because of the absence of a well-established set of diagnostic criteria to distinguish the two disorders.[13]

Different maneuvers in the clinical practice with voiced and voiceless tasks may help to distinguish AbSD than ADSD. Patients suffering from AbSD have problems with voiceless onset with sounds, such as /f/, /sh/.ch. /h/, /k/, /p/, /s/, /t/, by counting from 60 to 69, and symptoms also appear when singing. The symptoms of AdSD are more distinguishable during voiced reading. Symptoms are more prominent when counting from 80 to 89 and less when counting from 60 to 69, and they improved when singing or whispering. AdSD is characterized by intermittent voice offsets in the middle of voiced sounds, and the essential symptom is voice breaks. Furthermore, some patients had pitch or phonatory breaks during voiced sounds, uncontrolled rises in their fundamental frequency, or breathy voice quality. SD is clearly not a functional condition; however, as most of the other voice disorders, stress can make SD worse, and voice therapy can make it better. The underlying condition may easily be confused with MTD. It must be emphasized that SD is actually organic (neurological). Traditional objective voice measures for patients with SD may not always be helpful in the differential diagnosis because of the wide variation of findings across subjects. Acoustical analysis may help to diagnose an SD. Fundamental frequency from conversational sample may be useful in identifying each patient's compensatory strategy for managing his or her vocal spasms with extreme muscle tension.[2,14] Rees et al. suggested that visual inspections of voice spectrograms can help to distinguish AdSD from MTD.[15] The auditory-perceptual similarity of AdSD and MTD can lead to misdiagnosis, and ultimately inappropriate/unnecessary medical, behavioral, or surgical intervention. An additional problem is related to the absence of a set of established and validated diagnostic criteria to distinguish between AdSD and MTD in clinical practice, making difficult the differential diagnosis, even among experienced clinicians.[12,16,17]

Patients with AdSD may attempt to prevent their symptoms by increasing the tension in their laryngeal muscles in an effort to compensate the symptoms. The consequence is the appearance of additional physical disturbances similar to MTD along with AdSD. The overriding symptoms of MTD can escalate over time such that the underlying symptoms of AdSD are difficult to discern.[2,3] One of the most used treatments for AdSD is the injection of botulinum toxin A (Botox) into the muscles around the larynx. These injections serve to paralyze the muscles affected by the abnormal contractions, providing some relief from symptoms.[3] Furthermore, different surgeries could be possible options for this kind of dystonia.[7]

Acoustical measures of vocal function are routinely used in the assessments of pathological voices. They are very appealing because of their noninvasive nature. The extraction of such measures from sustained vowel samples is common because of its simpler acoustic structure. In recent years, the use of these measures, in combination with pattern recognition techniques, has motivated the emergence of several works concerning automatic diagnosis.[18–24] These works addressed the problem of normal versus pathological discrimination. The best-reported result reached a 96.5% level of correct classifications.[25]

As can be observed, most of the attempts of automatic classification in this area have been focused on the separation between normal and pathological voices. To our knowledge, there are no previous results concerning automatic discrimination between AdSD and MTD. At present, their differentiation depends solely on the clinical diagnosis skills of highly specialized voice therapists. Therefore, an interesting challenge is to make a contribution to the development of an automatic diagnosis system that, using well-known acoustic parameters of voice, could provide a support for the differential diagnosis between these two pathologies. This is the primary purpose of this study. With this in mind, in the present study, we attempt to separate normal voices from pathological voices and, additionally, obtain a classification as AdSD or MTD for the pathological cases.

It is important to take into account that we have used only one speech sample of each patient. For each of them, we extracted eight well-known acoustic parameters and created an eight-element array. These arrays have been classified into three categories: normal, AdSD, and MTD, using two strategies, one based on neural networks (NNs) and the other one based on support vector machines (SVM). The inclusion of nonlinear techniques allowed us to achieve our goal.

MATERIALS AND METHODS

The analyzed voices have been obtained from 89 speakers divided into 36 dysphonic (15 patients with MTD and 21 with AdSD) and 53 normal speakers. The speech signals correspond to a sustained vowel /a/. It was not possible to differentiate between MTD and AdSD based only on auditory perception. Subjects have been instructed to sustain the vowel /a/ for at least 3 seconds at a comfortable pitch and loudness using a professional digital audiotape recording and a professional microphone in an Industrial Acoustics Company (IAC) sound suite. To confirm our diagnosis of MTD or SD, we used an examination workout, including basis of history and physical examination; MRI of the brain; laryngeal EMG; laboratory test; neurological evaluation; voice assessment using videostroboscopy with flexible and rigid scopes and the routine phonatory task, adding voiced and unvoiced words or sentences; and a trial of voice therapy in each case. A patient diagnosed with MTD voice improved to a normal level with a trial of a short voice therapy using a variety of behavioral approaches. However, voice therapy is assumed to be ineffective for AdSD, and a poor response to voice therapy is often cited as corroborating evidence for the diagnosis. Botox injections alleviate spasmodic overclosure. All patients received a short term of voice treatment to confirm diagnosis. The response of patients with true AdSD to voice treatment was poor, and Botox injections improved their voices. Patients with successful voice treatment results were confirmed as having MTD.

Selected acoustic measures have been extracted from the sustained vowels, including short-term perturbations of fundamental frequency and intensity (termed jitter and shimmer, respectively), and glottal noise measures. Currently, a consensus does not exist on the utility of the independent use of these measures for discriminating between normal and pathological voices.[23] However, here we propose to build an eight-dimensional vector associated with each patient and to use the nonlinear properties of NN and SVM for classification.[26–28] We will show that the inclusion of these measures in combination with nonlinear techniques has allowed us to attain accurate discriminations, according to the proposed goals. Many of the methods previously proposed by other authors rely on an assumption of normality of the data. However, a Lilliefors test of the null hypothesis that the samples come from a normal distribution was performed for each analyzed acoustical parameter.[29] In all cases, the null hypothesis was rejected at the 1% significance level and with p-values less than 0.000025. This assumption is not needed by NN and SVM, which are data-driven based methods.

Because of the small number of available data, we have applied the *Leave-One-Out* (LOO) method in all the cases to have an estimation of the classification error. This means that the classification space is computed with every case in the database except the case that is being classified. In this way, the classification results are more realistic and close to the true classification rates.[26,30]

Features extraction

Here we present the selected parameters to construct the data vector associated with each patient. The estimation of the fundamental frequency (F_0) is of special importance, because the calculation of many parameters depends on an appropriated F_0 estimation. Parsa and Jamieson[31] concluded that, because of its robustness, the waveform matching algorithm is the one of choice for pathological voices or in the presence of moderate levels of background noise. Our experience confirms this assertion and, therefore, we have adopted this algorithm for F_0 extraction. The chosen parameters are degree of voice breaks (unvoiceness), three measures of period perturbation quotient or *jitter* (local jitter ratio [*jitt*], relative average perturbation [*RAP*], and five-point period perturbation quotient [*ppq5*]), three measures of amplitude perturbation quotient or *shimmer* (shimmer [*shimm*], three-point amplitude perturbation quotient [*apq3*], and 11-point amplitude perturbation quotient [*apq11*]) and harmonics-to-noise ratio (HNR).[14,18,32]

This selection is based on the fact that the analyzed disorders show voice breaks, increased jitter and shimmer, and decreased HNR,[11,33] and the physicians recognize these measures as the most relevant ones for this application. The three distinct jitters and shimmers reflect separate phenomena. For example, an individual with a rising F_0 in a monotonic way has a high *jitt* but low *RAP* and *ppq5*.

Neural networks

NNs are arrangements of simple and biologically inspired elements operating in parallel. A NN can be trained to solve a given problem adjusting the values of its inner connections, based on a comparison of the desired and obtained output. In such sense, it is a data-driven method.

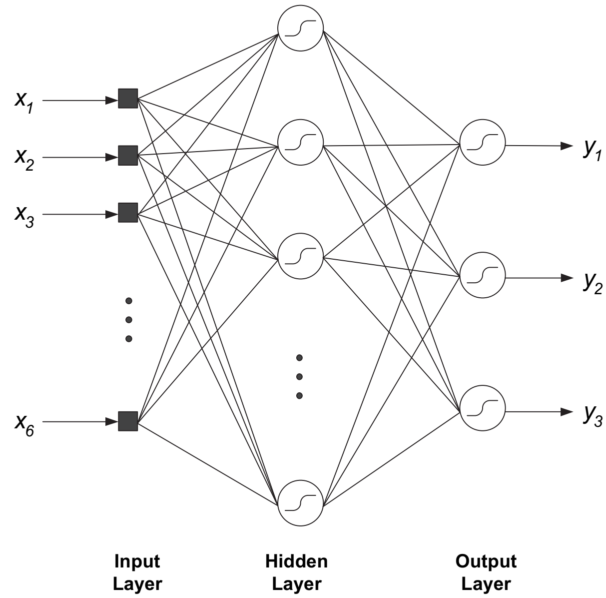

FIGURE 1. Multilayer perceptron configuration. There are six elements in the input layer, where the data (**x**) are presented. The size of hidden layer varied in our work. There are three output neurons, one corresponding to each class.

NNs are broadly used in the field of pattern recognition.[26,27,34] In the present work, a multilayer perceptron (MLP) has been applied. The resilient back propagation algorithm was chosen for training because of its excellent performance in pattern recognition problems.[35,36]

To improve the generalization capability of the classifier and facilitate its design, feature dimensionality reduction is necessary. The classical procedure in statistics is principal components analysis (PCA),[26] which reduces dimensionality by forming linear combinations of the features, reducing data redundance. The purpose of PCA is to find an optimal projection, which can account for a given percentage of the original data variance. Such directions are given by the largest eigenvectors of the covariance matrix of the full data. This optimizes a sum-squared error criterion.

Performing a PCA, the number of input units in the NN can be reduced, and consequently, the number of weights to be adapted can be diminished. In our case, six components contributed with 99.5% of the variance in the data set, meaning that the PCA lowered the size of the NN input vectors from eight to six.

We used hyperbolic tangent activation function both in the hidden layer and in the output layer. To select the best possible number of neurons in the hidden layer, its size was varied from eight to 34, running 100 experiments in each case. In the output layer, there were three neurons, one for each class (AdSD, MTD, and normal). In Figure 1, the MLP configuration is depicted. The six elements x_i $(i = 1, 2, \ldots, 6)$ of the input vectors are presented in the input layer. Each circle represents an

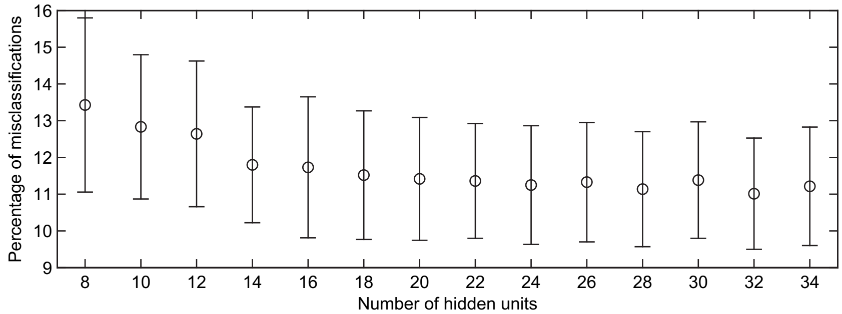

FIGURE 2. MLP NNs, three classes: percentage of misclassifications. The errors are plotted as a function of the number of hidden units of the NN. The length of each bar on each direction is the standard deviation.

individual neuron. The output layer contains three neurons, and the observed values are labeled as $y_j \ (j = 1, 2, 3)$. The "winner" output corresponds to the neuron with the highest value, and the input vector is classified with the class associated with this output.

Support vector machines

For comparison purposes, we considered an SVM approach as an alternative method. SVMs have recently been proposed and adopted as popular tools for learning from experimental data. An advantage of SVMs is that they do not depend on a random initialization, as the NN's weights. They do not need any fine-tuning of parameters and exhibit a great ability to generalize. In many problems, SVMs have been shown to provide better performance than more traditional techniques, such as highly tuned NNs.[37]

The basic idea behind the SVMs is to transform the data into a higher-dimensional space by some mapping fixed in advance and to find a large-margin separating hyperplane in the transformed space.[27,28,38] The output y of an SVM is computed as follows:

$$y(\mathbf{x}) = \sum_{i=1}^{N} \gamma_i K(\mathbf{x}, \mathbf{x}(i)) + b, \tag{1}$$

where $\mathbf{x}$ is the input feature vector, the kernel $K(\mathbf{x},\mathbf{x}(i))$ is a scalar-valued function of the testing sample $\mathbf{x}$ and a training sample $\mathbf{x}(i)$, and b is the bias term.

The coefficient γ_i and the bias b have to be estimated, and a set of support vectors $\{\mathbf{x}(i), i = 1,\ldots,N\}$ that may be a subset of the entire training set of data samples has to be identified. The two most commonly used kernel functions are the polynomial kernel $(\mathbf{x}^T\mathbf{y}+1)^p$ and the Gaussian radial basis function (RBF) $\exp(-\frac{1}{2\sigma^2}\|\mathbf{x}-\mathbf{y}\|^2)$. For details, please refer works by Vapnik[28] and Cortes and Vapnik.[39]

TABLE 1.
MLP NNs: Results of Tukey's Multiple Comparison Test

Group (Number of Hidden Units)	Mean Error ± Standard Deviation (%)	Groups With Means Not Significantly Different
32	11.01 ± 1.52	32 28 34 24 26 22 30 20 18 16 14
28	11.13 ± 1.57	32 28 34 24 26 22 30 20 18 16 14
34	11.21 ± 1.61	32 28 34 24 26 22 30 20 18 16 14
24	11.25 ± 1.62	32 28 34 24 26 22 30 20 18 16 14
26	11.33 ± 1.63	32 28 34 24 26 22 30 20 18 16 14
22	11.36 ± 1.56	32 28 34 24 26 22 30 20 18 16 14
30	11.38 ± 1.59	32 28 34 24 26 22 30 20 18 16 14
20	11.42 ± 1.67	32 28 34 24 26 22 30 20 18 16 14
18	11.52 ± 1.75	32 28 34 24 26 22 30 20 18 16 14
16	11.73 ± 1.92	32 28 34 24 26 22 30 20 18 16 14
14	11.80 ± 1.58	32 28 34 24 26 22 30 20 18 16 14
12	12.64 ± 1.98	12 10 8
10	12.83 ± 1.96	12 10 8
8	13.43 ± 2.37	12 10 8

Each group is labeled using the number of hidden units of the NN. The first column shows the number of hidden units in the NN. The mean errors (mean of the percentage of misclassifications) and the standard deviations are presented in the second column. Finally, in the third column, the groups that are not significantly different from the group in the first column are shown, with a significance level of $\alpha = 0.05$.

RESULTS

In this section, we present the results obtained with the two approaches: NNs and SVMs. By means of a statistical test, we quantify the results obtained in the first case for different number of neurons in the hidden layer.

Neural networks

For the purpose of comparing the performance of several network sizes, we conducted 100 experiments with each

TABLE 2.
MLP NNs: Best Confusion Matrix for 14 Hidden Units

Actual Class	Predicted Class: AdSD	Predicted Class: MTD	Predicted Class: Normal	Correct Classifications (%)
AdSD	20	1	0	95.24
MTD	5	10	0	66.67
Normal	0	0	53	100.00
Total				93.26

TABLE 3.
MLP NNs: Best Confusion Matrix for 16 Hidden Units

Actual Class	Predicted Class: AdSD	MTD	Normal	Correct Classifications (%)
AdSD	17	4	0	80.95
MTD	2	13	0	86.67
Normal	0	0	53	100.00
Total				93.26

configuration, registering the mean classification error for each network size. In each experiment, the weights were randomly initialized. As it was stated in the previous section, errors are measured using the LOO method. This means that the accuracy of the classifier was estimated by training it separately 89 times, using the complete training set, from which a different single pattern has been deleted. Each resulting network was tested on the single deleted point, and the jackknife estimate of the accuracy is just the mean of these LOO accuracies.[26]

The mean percentage of misclassifications obtained and the corresponding standard deviation for each number of hidden units are depicted in Figure 2. As can be seen, the error seems to be stabilized after 14 hidden units.

To contrast these mean errors and to obtain information about which mean errors pairs were significantly different and which ones were not, a multiple comparison test that considers all pairwise comparisons has been applied. We applied the Tukey test with a significance level of $\alpha = 0.05$. This is the recommended method when we want to test the family of all pairwise comparisons.[40]

In Table 1, we show the results of the Tukey multiple comparison test. Each group was labeled using the number of hidden units of the NN. The first column corresponds to this number. The mean error (mean of the percentage of misclassifications) and the corresponding standard deviation are presented in the second column. The third column shows the groups that are not significantly different from the one indicated in the first column, with a significance level of $\alpha = 0.05$. From the analysis of Table 1, we can conclude that, when the number of hidden units in the NN is increased above 14, the error is not significantly reduced.

TABLE 4.
MLP NNs: Best Confusion Matrix for 22 Hidden Units

Actual Class	Predicted Class: AdSD	MTD	Normal	Correct Classifications (%)
AdSD	18	3	0	85.71
MTD	3	12	0	80.00
Normal	0	0	53	100.00
Total				93.26

TABLE 5.
MLP NNs: Average of the 100 Confusion Matrices for 32 Hidden Units

Actual Class	Predicted Class: AdSD	MTD	Normal	Correct Classifications (%)
AdSD	15.78	4.48	0.74	75.14
MTD	4.45	10.44	0.11	69.60
Normal	0.01	0.01	52.98	99.96
Total				88.99

Pathological voices are classified as pathological (AdSD or MTD) in 97.63% of the cases, and normal voices are classified as normal in 99.96% of the cases.

In Tables 2–4, we show the confusion matrices for three different numbers of hidden units: 14, 16, and 22 neurons, respectively. We can appreciate the fact that, in the three cases, the best results reach 93.26% of correct classifications over all the voices and a 100% of correct classifications of normal voices.

Although the total percentages of discrimination obtained in Tables 2–4 are equal, we can see that the values corresponding to each class (pathology) are different. The highest percentage of correct classifications of AdSD voices was obtained with 14 hidden units (Table 2). However, the best result for MTD recognition was achieved using 16 hidden neurons (Table 3). It can be observed that an intermediate result was obtained with 22 hidden units (Table 4).

The minimum mean error (11.01%; see Table 1) along the 100 experiments with the different numbers of hidden neurons was obtained with 32 units. The averaged confusion matrix for this case is shown in Table 5. With this network configuration, the mean of correct classifications was 88.99%. It is important to add that, in this case, the mean of normal voices correctly classified was 99.96%, and by combining AdSD and MTD in a single class (pathological voices), 97.63% of correct classifications was achieved.

To compare our results with previous works, we tested the ability of our classifier for discrimination between pathological and normal voices. For this purpose, we have changed the output layer, leaving now two output neurons. Increasing the number of hidden units from one to 14, and running 100 realizations in each case, we obtained the minimum mean error

TABLE 6.
MLP NNs: Classification in Two Categories

Actual Class	Predicted Class: Pathological	Normal	Correct Classifications (%)
Pathological	35.13	0.87	97.58
Normal	0.07	52.93	99.87
Total			98.94

Average of the 100 confusion matrices for 8 hidden units.

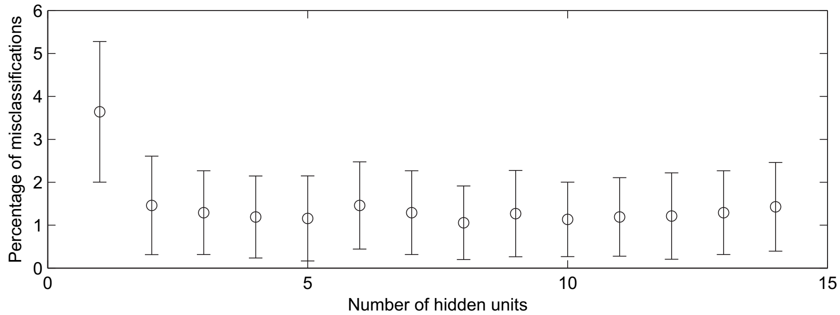

FIGURE 3. MLP NNs, two classes: percentage of misclassifications. The errors are plotted as a function of the number of hidden units of the NN. The length of the bars on each direction represents the standard deviation.

while working with eight hidden neurons. In many realizations (even in those with only one hidden unit), we have observed that the classification was 100% correct. In Table 6, an averaged confusion matrix for the 100 realizations is presented. The percentage of correct classifications reached 98.94% (97.58% of pathological voices and 99.87% of normal voices correctly classified). This result is better than the highest percentage of correct discrimination between normal and pathological voices found in the literature[25] (96.5% of correct detections).

In Figure 3, the mean percentage of misclassifications is depicted. The minimum error was obtained with eight neurons. A Tukey test with a significance level of $\alpha = 0.05$ was applied, indicating that the MLPs with two or more neurons in the hidden layer are not significantly different.

Support vector machines

In the case of SVM, the reduction of the input vector dimension using PCA caused a decrease of the correct classifications. Because of this fact, we used the original eight-dimensional input vectors. As in the previous case, we applied the LOO strategy. First, we tried to classify the voices in three classes (MTD, AdSD, and normal) using different kernels. Tables 7 and 8 show the results obtained with polynomial kernel ($p = 2$) and Gaussian radial basis function ($\sigma = 0.5$), respectively. It can be observed that the polynomial kernel yields the best classification in three classes. To check the ability of SVM for separating normal and pathological voices, we grouped AdSD and MTD together. In this case, our concern was to classify the voices as pathological or normal. Tables 9 and 10 summarize the obtained results. Again, we can appreciate the fact that the polynomial kernel had a better performance than Gaussian radial basis function.

DISCUSSION

In this study, we have compared two different approaches for the automatic classification of pathological voices. In particular, we have focused our attention on two different aspects: (1) to discriminate between normal and pathological voices; and (2) to discriminate between normal, AdSD and MTD, the two pathologies for which a proper indicator for their differential diagnosis does not exist, and misdiagnosis occurs quite often.

According to the results presented in the previous section, AdSD pathology was better recognized using a NN with 14 hidden units, with a 95.24% (Table 2). MTD reached 86.67% of correct classifications by means of a NN with 16 hidden units (Table 3), whereas normal voices were, in most of the cases, 100% recognized. When interested in a three-class classification, the best result was obtained, reaching 93.26% of successful classifications, using NNs of 14, 16, or 22 hidden units (Tables 2–4).

In case of separating pathological and normal voices, it is also possible to reach a very good discrimination and, as it has been shown, these results overcome those published to date (Table 6). The results with a MLP with eight hidden neurons averaged 98.94% of correct classifications in 100 realizations, overcoming the best-reported percentage of correct classifications (96.5%). In addition, comparing NNs and SVMs as automatic classifiers for pathological and normal voices, our results allow to conclude that, in the present application, the MLP NNs with

TABLE 7.
Confusion Matrix for SVMs. Classification in Three Classes with Polynomial Kernel ($p = 2$)

Actual Class	Predicted Class AdSD	MTD	Normal	Correct Classifications (%)
AdSD	18	3	0	85.71
MTD	5	9	1	60.00
Normal	0	1	52	98.11
Total				88.76

TABLE 8.
Confusion Matrix for SVMs. Classification in Three Classes with RBF Kernel, $\sigma = 0.5$

Actual Class	Predicted Class AdSD	MTD	Normal	Correct Classifications (%)
AdSD	16	2	3	76.19
MTD	4	9	2	60.00
Normal	0	0	53	100.00
Total				87.64

TABLE 9.
Confusion Matrix for SVMs. Classification in Two Classes with Polynomial Kernel (p = 2)

Actual Class	Predicted Class: Pathological	Predicted Class: Normal	Correct Classifications (%)
Pathological	35	1	97.22
Normal	1	52	98.11
Total			97.75

eight hidden units are more successful than SVMs. These preliminary results suggest the necessity of future works using the presented approach with larger databases.

CONCLUSIONS

In this work, we apply two machine learning-based approaches—NNs and SVMs—with the purpose of the automatic discrimination between AdSD, MTD, and normal voices, using acoustic parameters of sustained vowel /*a*/.

From this study, it can be seen that an automatic classification is possible between two pathologies that often cause erroneous diagnosis by professionals who are not highly specialized. Additionally, this discrimination is feasible using only acoustical measures that are well known by both the speech physicians and the therapists. This is a very important attribute of our approach because of the knowledge that the specialists have on these parameters. Their properties, significances, and relations with pathologies have been extensively studied in the literature and, additionally, the computational issues involved in their estimation are well established. In this way, researchers can use the approach here presented with no further implementation difficulties. To our knowledge, there is no previous work regarding automatic classification of SD and MTD.

The presented automatic classification tools must be refined and tested on a higher number of normal and pathological voices, including several samples by each subject, before clinical usage as a support for the differential diagnosis. Furthermore, sustained vowels versus running speech may be tested. Nevertheless, the results obtained are very promising and suggest that this goal is well within reach. A reliable system of automatic classification could save many MTD patients from inappropriate Botox injections or surgery as a consequence of a misdiagnosis.

TABLE 10.
Confusion Matrix for SVMs. Classification in Two Classes with RBF Kernel, $\sigma = 0.5$

Actual Class	Predicted Class: Pathological	Predicted Class: Normal	Correct Classifications (%)
Pathological	34	2	94.44
Normal	3	50	94.34
Total			94.38

Acknowledgment

This work is supported by ANPCyT, Universidad Nacional de Entre Ríos, and CONICET under Projects PICT-2008-0433, PID 6107, and PID 6111.

REFERENCES

1. Krischke S, Weigelt S, Hoppe U, Köllner V, Klotz M, Eysholdt U, Rosanowski F. Quality of life in dysphonic patients. *J Voice*. 2005;19:132-137.
2. Barkmeier JM, Case JL, Ludlow CL. Identification of symptoms for spasmodic dysphonia and vocal tremor: a comparison of expert and nonexpert judges. *J Commun Disord*. 2001;34:21-37.
3. Jackson-Menaldi C. La voz patológica. Buenos Aires, Argentina: Editorial Médica Panamericana; 2002. (In Spanish).
4. Schwartz JS, Song P, Blitzer A. Spasmodic dysphonia. In: Cooper G, ed. *Therapeutic Uses of Botulim Toxin*. Totowa, New Jersey: Humana Press; 2007:109-121.
5. Konkiewitz CE, Trender-Gerhard I, Kamm C, et al. Service-based survey of dystonia in Munich. *Neuroepidemiology*. 2002;21:202-206.
6. Duffy J, Yorkston K, Beukelman D, et al. *Medical Interventions for Spasmodic Dysphonia & Some Related Conditions: A Systematic Review*. Minneapolis, MN: Academy of Neurologic Communication Disorders and Sciences; 2001. Technical Report 2.
7. Deems DA, Sataloff RT. Spasmodic dysphonia. In: Sataloff RT, ed. *Professional Voice. The Science and Art Clinical Care*. 2nd ed. San Diego: Singular Publishing Group; 1997:499-505.
8. Case JL. *Clinical Management of Voice Disorders*. Austin, TX: Pro-Ed; 1996.
9. Inamura R, Tsuji DH. Adduction spasmodic dysphonia, vocal tremor and muscular tension dysphonia: it is possible to reach a differential diagnosis? *Braz J Otorhinolaryngol*. 2006;72:434.
10. Barkmeier JM, Case JL. Differential diagnosis of adductor-type spasmodic dysphonia, vocal tremor and muscle tension dysphonia. Speech therapy and rehabilitation. *Curr Opin Otolaryngol Head Neck Surg*. 2000;8:174-179.
11. Verdolini K, Rosen CA, Branski RC, eds. *Special Interest Division 3. Voice and Voice Disorders. American Speech-Language-Hearing Association, Classification Manual for Voice Disorders—I*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2006.
12. Roy N, Mauszycki SC, Merrill RM, Gouse M, Smith ME. Toward improved differential diagnosis of adductor spasmodic dysphonia and muscle tension dysphonia. *Folia Phoniatr Logop*. 2007;59:83-90.
13. Roy R, Gouse M, Mauszycki SC, Merrill RM, Smith ME. Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope*. 2005;115:311-316.
14. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. 2nd ed. San Diego, CA: Singular Thomson Learning; 2000.
15. Rees CJ, Blalock D, Kemp SE, Halum SL, Koufman JA. Differentiation of adductor-type spasmodic dysphonia from muscle tension dysphonia by spectral analysis. *Otolaryngol Head Neck Surg*. 2007;137:576-581.
16. Cannito MP, Woodson G. The spasmodic dysphonias. In: Kent R, Ball M, eds. *Voice Quality Measurement*. San Diego, CA: Singular Thomson Learning; 2000:411-430.
17. Ludlow CL. Management of the spasmodic dysphonias. In: Rubin JS, Sataloff RT, Korovin G, Gould WJ, eds. *Diagnosis and Treatment of Voice Disorders*. New York, NY: Igaku-Shoin; 1995:436-454.
18. Boyanov B, Hadjitodorov S. Acoustic analysis of pathological voices. *IEEE Eng Med Biol Mag*. 1997;16:74-82.
19. Hansen JHL, Gavidia-Ceballos L, Kaiser JF. A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment. *IEEE Trans Biomed Eng*. 1998;45:300-313.
20. de Oliveira Rosa M, Pereira JC, Grellet M. Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans Biomed Eng*. 2000;47:96-104.
21. Ritchings RT, McGillion M, Moore CJ. Pathological voice quality assessment using artificial neural networks. *Med Eng Phys*. 2002;24:561-564.

22. Godino-Llorente JI, Gómez-Vilda P. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Trans Biomed Eng*. 2004;51:380-384.
23. Umapathy K, Krishnan S, Parsa V, Jamieson DG. Discrimination of pathological voices using a time–frequency approach. *IEEE Trans Biomed Eng*. 2005;52:421-430.
24. Moran RJ, Reilly RB, de Chazal P, Lacy PD. Telephony-based voice pathology assessment using automated speech analysis. *IEEE Trans Biomed Eng*. 2006;53:468-477.
25. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*. 2001;44: 327-339.
26. Duda RO, Hart PE, Stork DG. *Pattern Classification*. 2nd ed. New York: Wiley; 2001.
27. Hu YH, Hwang JN. *Handbook of Neural Network Signal Processing*. Boca Raton, FL: CRC Press; 2002.
28. Vapnik VN. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag; 1995.
29. Lilliefors HW. On the Komogorov-Smirnov test for normality with mean and variance unknown. *J Am Stat Assoc*. 1967;62:399-402.
30. Lachenbruch P, Mickey R. Estimation of error rates in discriminant analysis. *Technometrics*. 1968;10:1-11.
31. Parsa V, Jamieson DG. A comparison of high precision F0 extraction algorithms for sustained vowels. *J Speech Lang Hear Res*. 1999;42:112-126.
32. de Krom G. A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals. *J Speech Lang Hear Res*. 1993;36:254-266.
33. Cannito MP, Buder EH, Chorna LB. Spectral amplitude measures of adductor spasmodic dysphonic speech. *J Voice*. 2005;19:391-410.
34. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon; 1995.
35. Demuth H, Beale M. *Neural Network Toolbox for Use with MATLAB*. Natick, MA: The MathWorks Inc; 2004.
36. Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *Proceedings of the IEEE International Conference on Neural Networks*; IEEE, USA:586-591.
37. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press; 2000.
38. Kulkarni SR, Lugosi G, Venkatesh SS. Learning pattern classification—a survey. *IEEE Trans Inf Theory*. 1998;44:2178-2206.
39. Cortes C, Vapnik V. Support vector networks. *Mach Learn*. 1995;20:273-297.
40. Rafter JA, Abell ML, Braselton JP. Multiple comparison methods for means. *SIAM Rev*. 2002;44:259-278.