Suboptimal Classifier for Dysarthria Assessment

Eduardo Castillo Guerra[1] and Dennis F. Lovely[2]

[1]Centre for Studies on Electronics and Information Technologies, Central University "Marta Abreu" of Las Villas, Carr. Camajuaní Km 5½, Santa Clara, VC, Cuba, 50100
ecastillo@fie.uclv.edu.cu

[2] Department of Electrical Engineering, University of New Brunswick, Fredericton, NB, E3B5A3, Canada.
lovely@unb.ca

Abstract. This work is focused on the design and evaluation of a suboptimal classifier for dysarthria assessment. The classification relied on self organizing maps to discriminate 8 types of dysarthria and a normal group. The classification technique provided an excellent accuracy for assessment and enabled clinicians with a powerful relevance analysis of the input features. This technique also allows a bi-dimensional map that shows the spatial distribution of the data revealing important information about the different dysarthric groups.

1 Introduction

Dysarthria is a collective name given to a group of neurological diseases that are originated by lesions in the peripheral or central nervous system. The location of the lesions determines the perturbation induced on the speech signal. Therefore, there is a relationship between the speech perturbations observed and the type of dysarthria.

The assessment of dysarthria is often performed based on features extracted from recorded speech which reflect the perturbation patterns described by the different types. The goal for dysarthria assessment is to obtain high sensitivity in the discrimination process while allowing clinicians to perform backward analysis of the feature contribution to the final decision of the classifier. This analysis is necessary to establish a correlation between the most prominent features in each dysarthric group and the neurological damage.

The main limitations encountered in dysarthria assessment can be defines as: the unavailability of objective measures that describe efficiently the speech perturbations, the lack of a gold standard to compare different techniques developed and the few databases available for research. The first limitation mentioned is currently being studied exhaustively where new digital signal processing algorithms are being developed to describe more accurately those perturbations used as clues for assessment [1], [2], [3], [4], [5], [6]. The second limitation is still an unsolved problem in the research community due to the existence of different severity levels of the speech perturbation, the disagreement among researchers about the best set of features to use in the assess-

A. Sanfeliu and J. Ruiz-Shulcloper (Eds.): CIARP 2003, LNCS 2905, pp. 314–321, 2003.

ment and the interrelation between perturbations. However, steps have being given toward this limitation with the workshop on acoustic voice analysis [6]. The last limitation is a critical problem due to the difficulty of recording large population of subjects with these diseases. New pathological speech databases are now available for research but are not focused only on dysarthric patients, therefore, they provide limited information [7]. As a consequence of these limitations, the development of efficient classification techniques for this application is highly desirable and necessary.

2 Experiences in Dysarthria Assessment

Several protocols have been used for dysarthria assessment considering a variety of descriptive features. The diagnosis of the dysarthria has been traditionally performed by the differential diagnosis of dysarthria [8]. This diagnosis method relies on perceptual judgments (PJ) of the pathological speech as the main descriptors. The authors defined 38 features or dimensions that describe more efficiently the speech perturbations. The judgments are grouped into clusters according to the speech mechanism affected and the combination of the clusters exhibited determines the type of dysarthria. The decision is based on minimal distance between the clusters manifested and the combination of clusters that characterize each type of dysarthria.

A survey performed regarding the use of this assessment method showed that more than 60% of clinicians in North America use this system in their clinical practice. However, there are limitations reported for this method in the effectiveness to assess subjects with mixed dysarthrias [9]. The PJ can also be imprecise an inconsistent when certain speech features are analyzed, particularly when they are performed by clinicians which come from different schools and have different reference points. This way of judging often leads to low reliability and repeatability of the process of describing the speech perturbations, causing low assessment rates and difficulties standardizing the results of the research in this area.

Other assessment methods, summarized in [10], describe similar protocols based on linear analyses of different sets of features. The definition of the descriptive features is also different but provides information of similar speech perturbations. Most of these protocols rely also on perceptual judgments of speech and suffer from similar limitations as the traditional method. Some assessment protocols reported perform the discrimination between classes using linear discriminant analysis (LDA). This approach provides better performance than the traditional clustering method. However, the relevancy analysis of the input feature has been found imprecise [3].

More recent studies reported by Callan et al. 1999 [11], implemented the assessment of few dysarthric groups using a small set of objective measures and self-organizing maps (SOM). Despite the small number of subject and groups of dysarthrias assessed the study revealed the effectiveness of the method and opened new options in the assessment of dysarthria.

Nowadays, the trend in dysarthria assessment is the use of different tools to provide objective measures of the speech perturbations and implement non-linear classification techniques to differentiate the different groups [5]. This approach could lead to better and more consistent judgments. However, the backward analysis on the decision of the classification is an important requirement that the classification method has to meet to provide a global picture these diseases.

3 Self-Organizing Maps

A self-organizing map is an unsupervised neural network that learns to recognize regularities and correlations in its input data and adapts future responses according to that input. This network (see block diagram in Fig. 1.) not only learns to recognize groups of similar input vectors but also neighboring neurons learn to recognize neighboring sections of the input space. Therefore, the SOM learns both the distribution and topology of the input vectors on which they are trained.

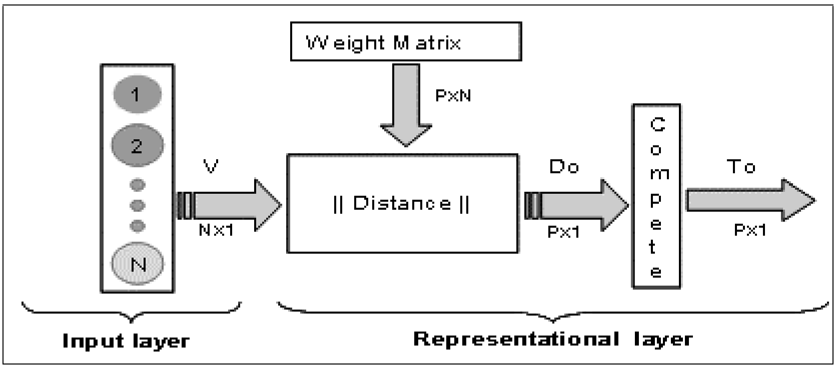

Fig. 1. Diagram of a self-organizing map Network

SOMs are made up of an input layer that interfaces with the multidimensional features and a representational layer. This second layer is a two-dimensional array of nodes which has a weight matrix associated. A distance box is used to estimate the negative distance between the input vectors (V) and the weight matrix. This estimates the neuron that more likely represents the characteristics of the input case (winning neuron). The competitive level produces a one for the output element corresponding to the winning neuron while all other elements are set to zero. However, neurons close to the winning neuron are updated along with the winning neuron using the Kohonen learning rule [12] expressed as:

$$m_i(t+1) = m_i(t) + \alpha(t)[x(t) - m_i(t)] \tag{1}$$

where m_i is the connection weights of node i for time step t, x(t) is the input vector for time step t, and $\alpha(t)$ is learning rate for time step t.

The characteristics described previously allow this type of network not only to learn the distribution of the input vector but also to gain information regarding the neighborhood. The way that neighbors' neurons are updated depends on the topology selected, which can vary between rectangular, hexagonal or random. The weight values are determined

by an unsupervised learning algorithm that offers the advantage that a gold standard target is not required.
Iteratively, the network learns the distribution of the input vector while the data are presented and the weighted connection of the representational layer is corrected. Each winning neuron is obtained through a process of determining which of the nodes in the representational layer is closer to the input vector, according to the distance criteria [12].

3.1 Design of the Classifier

A SOM network designed for this type of application requires two layers, an input layer and the representational layer (Fig. 1). The input layer contains 20 inputs neurons corresponding to the 20 observations obtained from perceptual and acoustic analysis[1], described more precisely in [3]. The representational layer consisted of a 9-by-9-node layer with a hexagonal lattice configuration. A bubble neighborhood function type was used in the training phase as recommended by Kohonen (1995) [12]. This bubble function is reported to provide a configuration that can show better visual information in the map.

Fifteen SOMs were trained using different initial random weights and the configuration that provided the lowest overall quantization error was selected for further studies. The SOMs were trained in two steps as recommended by Kohonen (1995), the ordering phase and the convergence phase. The ordering phase refers to the task of ordering the reference vectors. In this phase, the neighborhood radius is close to the diameter of the map and is decreased during the training. The learning rate is large and decreases toward zero as the network is trained. The initial radius used in the network was 9 with a learning factor of 0.09 and 2000 iterations. This phase established a gross association between the nodes and the input vectors.

The convergence phase is the step in which the reference vector on each node converges to an ‘optimal’ location. The radius and the learning rate are usually smaller in this phase while the number of iterations is usually larger than in the ordering phase. The radius used in this phase of the design decreased from 1 to 0 with a learning rate of 0.008 decreasing to 0 as well. The number of iterations in this phase was 52000 to allow time for the convergence to the optimal position. This phase allows a fine tuning between the vectors and the nodes. The SOM_PAK software [13] was used for both phases of the training.

3.2 Evaluation of the Performance of the Classifier

The performance of the classifier was evaluated using a set of 127 subjects from 9 target classes (AD: Ataxic dysarthria, ALS: Amyotrophic Lateral Sclerosis, FD: Flaccid dysarthria, HC: Chorea, HD: Dystonia, OVT: Organic Voice Tremor, HP: Parkin-

[1] Observations: Pitch level, pitch break, tremor, excess of loudness variation, harsh voice, breathy voice, voice stoppages, audible inspirations, speech rate, short phrases, short rushes of speech, monoloudness, hypernasality, reduced stress, variable rate, prolonged intervals, inappropriate silences, excess or equal stress, articulatory breakdowns and distorted vowels.

son's disease, SD: Spastic dysarthria and NS: Control Group). A cross-validation technique was used to prevent an overly optimistic classification rate. This technique works by omitting one subject's data, then retraining the neural network using the remaining data and finally classifying the omitted observation. In this way, more data samples participate in the network training and a more realistic performance measure is obtained. The main disadvantage of this method is that a set of networks is obtained from the training process. However, the network that provided the lowest overall quantization error was kept as the most representative network. The distribution of the vectors across the SOM map for this network is shown in Fig. 2.

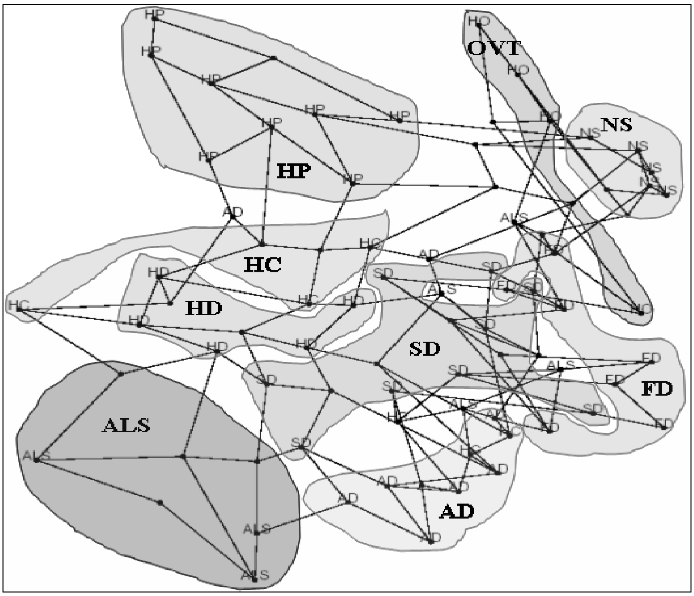

Fig. 2. Distribution of the classified groups across the SOM. The dots represent the nodes of the 9x9 map implemented and the lines represent the Euclidean distances between each node. The colored areas enclose the most probable neurons for each group

It is observed that the SOM performed a good separation between the different groups. The shaded areas represent the zones in which a larger number of nodes of each class tends to concentrate. The normal group is the most concentrated in its area with HP, AD, HD, and OVT having very well defined areas. As expected, the HC group is observed close to HD since both are hyperkinetic dysarthrias sharing common typical dimensions. Similarly, the area of the groups FD and SD are overlapped due to both share many similar dimensions. It is noticeable that there are some ALS neurons in between the area defined for FD and SD. This is explained from the clinical view point based on the fact that ALS is a mixed dysarthria made of a combination of FD and SD.

The confusion matrix observed in Table 1 shows the performance of the classification technique. The lowest classification rates are observed for HD and HC groups due to the great overlap among its speech deviations. It is observed that although ALS class was the most scattered across the map, its percent of correct classification (PCC) was not the worst. The quantization error of each sample of the dataset with respect the selected map was 1.916 for the network that performed worst.

Table 1. Confusion matrix of SOM classifier after cross-validating the dataset

Group	*True Groups*								
	AD	ALS	FD	HC	HD	OVT	HP	NS	SD
AD	11	0	0	1	1	0	1	0	0
ALS	0	10	1	2	1	0	0	0	0
FD	2	0	12	0	1	1	0	0	0
HC	0	0	0	8	0	1	0	0	0
HD	0	0	0	2	10	0	0	0	0
HO	0	1	0	0	0	11	0	0	0
HP	0	0	0	0	0	0	15	0	0
NS	0	0	0	0	0	0	0	19	0
SD	0	2	0	0	1	0	0	0	13
Total	13	13	13	13	14	13	16	19	13
Correct	11	10	12	8	10	11	15	19	13
PCC	0.846	0.769	0.923	0.615	0.714	0.846	0.938	1.00	1.00
N total=127	N Correct=109				Proportion Correct=0.8583				

The total PCC obtained with the SOM method is 0.86, which outperformed the results obtained with the traditional method (0.66) and the LDA method (0.81). The difference in performance between the SOM and the LDA methods is not very significant since some of the input features were measures made of linear combination of different objective algorithms [3]. However, the SOM provided a more reliable relevancy analysis of the input features and provided the bi-dimensional map.

4 Relevance Analysis of the Dataset Features

The contribution of the dimensions to the differentiation of the dysarthric groups can be explained with the use of SOM networks. This has been the main drawback of many non-linear analyses, such as those performed with some types of ANN, in which the relevancy of the observations is not properly understood. The SOM emerged from a vector quantization algorithm that places a number of reference codebooks into a high dimensional input data space which is an organized approximation of the dataset structure. The self-organizing algorithm that shape this structure can be analyzed as a non-linear regression of the reference vectors though the data points [12]. Therefore, the node's weight vectors corresponding to each group can provide information on the relevancy of the input dimensions to the group. Each neuron will have a set of weighs characterizing each target group with a weight associated with each dimension.

Based on the previous explanation, the neurons closer to the centroid of each group will have weights associated with similar characteristics to the mean values of each group. The magnitude and sign of these weights will provide a clinically valuable relevance indicator.

Table 2 shows an example of the most relevant dimensions obtained for the classification of the FD group. The most clinically relevant dimensions found with the traditional and lineal discriminant analyses are also shown for comparison.

Table 2. Relevance analysis for the traditional method of dysarthria assessment (PA), linear discriminant analysis and self-organizing map. '*' indicates coincidence in all studies

	ORDER												
Analysis	1	2	3	4	5	6	7	8	9	10	11	12	13
PA	HN	IC	M	AI	SP	ML	BV	NE	HV	R	-	-	-
LDA	HN	DV	RS	ML	IS	IAB	SP	EES	VR	ELV	VST	R	AI
SOM	HN*	SP	BV	R*	AI*	ML*	PI	PB	HV	EES	-	-	-

The analysis shows, in agreement with the other methods, the dimension hypernasality as the most prominent feature in this group. This is in correspondence with reported studies based on physiological analysis of this type of disease [10]. Short phases (SP), breathy voice (BV), rate (R), audible inspirations (AI), monoloudness (ML), harsh voices (HV), prolonged intervals and excess of loudness variations are speech features also typical in this Dysarthric group. This analysis shows the dimension PB as relevant although it was not found relevant in the previous studies. However, PB is often heard in subjects with FD (i.e. Darley, Aronson & Brown listened PB in 5 of their 30 FD subjects [14]).

The other methods, especially the LDA method, show features that are not commonly seen in this type of disease such as irregular articulatory breakdown (IAB), variable rate (VR) and imprecise consonant (IC). These methods also partially disagree with other physiological studies with respect to the relevancy order [10]. Similar relevancy analyses can be implemented on the rest of the dysarthric groups. In all cases the SOM method performed better than the others method studied demonstrating the feasibility of this technique to perform the relevancy analysis of the input features.

5 Conclusions

The results of the classification process reveal the convenience of using SOM for the assessment of dysarthria. A comparison with the implementation of the traditional and LDA classification methods shows that the SOM classifier outperformed the other two methods nearly by 5% and 20% respectively. The SOM also learned the topology of the data, producing a bidimensional map that provides more complete information about the different dysarthric groups. The map obtained for the dysarthric database can provide information no only about the type of disease or the contribution of the observation, but also about the evolution of the subjects after treatment. This is always an issue in providing objective testimonies of the disease progress.

The SOM technique also bestows a more accurate relevancy analysis than the other methods studied acting as a non-linear regression algorithm. The relevancy analysis implemented in the form explained in Section 4 is simpler and easier to understand than other methods reviewed. This is important to ensure that a system based on this technique is used in regular practice by speech language pathologists. An assessment tool implemented with this classification technique can also avoid exposure to radiation or high magnetic fields in analysis commonly performed on subjects with these diseases.

References

1. Michaelis, D., Strube, H.W.: Empirical study to test the independence of different acoustic voice parameters on a large voice database. Eurospeech proceedings. Vol. 3. (1995) 1891–1894
2. Michaelis, D., Gramss, T. & Strube H.W.: Glottal to noise excitation ratio – A new measure for describe pathological voices. Acust. Acta Acust. Vol. 83 (1997) 700–706
3. Castillo-Guerra E.: A modern approach to dysarthria classification. Ph.D. Thesis at University of New Brunswick. Canada (2002)
4. Klatt D.H., Klatt L.C.: Analysis, synthesis and perception of voice quality variations among female and male talkers. J. Acous. Soc. Am. Vol. 87. No. 2. (1990)
5. Kent, R.D., Weismer, G., Kent, J.F., Houri, K.V., Duffy J.R.: Acoustic studies of dysarthric speech: Methods, progress, and potential. J. Comm. Disord. Vol. 32. (1999) 141–186
6. Titze, I.R.: Recommendation on acoustic voice analysis, summary statement. Workshop on Acoustic Voice Analysis. Denver. Colorado (1994)
7. Kay Elemetrics Corp.: Disordered speech database. Massachusetts Eye and Ear Infirmary. Voice and Speech Lab. Boston. M.A. Ver. 1.03 (1994)
8. Darley F.L., Aronson A.E., Brown J.R.: Differential diagnostic patterns of dysarthrias. J. Speech Hear. Res. Vol. 12. (1969a) 462–496
9. Simmons K.C., Mayo R.: The use of the Mayo Clinic system for differential diagnosis of dysarthria. J. Comm. Disord. Vol. 30 (1997) 117–132
10. Murdoch B.E., Chenery H.J.: Dysarthria, a physiological approach to assessment and treatment. Singular Publishing Group Inc., Aug (1997)
11. Callan, E.D., Roy N., Tasko S.M.: Self organized maps for the classification of normal and disordered female voices. J. Speech Hear. Res. Vol. 42 (1999) 355–366
12. Kohonen, T.: Self-organized maps. Berlin Springer (1995) 77–127
13. Kohonen, T. , Hynninen, J., Kangas J., Laaksonen, J.: SOM_PAK: The Self-Organizing Map Program Package. Technical Report A31. Helsinki University of Technology. Laboratory of Computer and Information Science. FIN-02150 Espoo. Finland (1996)
14. Darley, F.L., Aronson, A.E., Brown, J.R.: Differential diagnostic patterns of dysarthria. J. Speech Hear. Res. Vol. 12 (1969a) 246–249