Objective discrimination between Progressive Supranuclear Palsy and Multiple System Atrophy using speech analysis

K. Daoudi, N. Brodu, J. Rusz and J. Klempir

***Abstract*— We propose a fully reproducible speech-based technique for objective differential diagnosis between progressive supranuclear palsy (PSP) and multiple system atrophy (MSA). Our technique yields a classification mean accuracy of 86.1% which is a significant improvement as compared to a recent pioneer study on this task.**

I. Introduction

Speech disorder (or *dysarthria*) is frequently an early and prominent clinical feature of Parkinson's disease (PD) as well as atypical parkinsonian syndromes (APS). In this paper, we focus on speech-based differential diagnosis between PSP and MSA. Previous descriptions of speech in PSP and MSA have been mainly limited to the perceptual estimation of dysarthria type [1], where spastic components appear to be more dominant in PSP and hypokinetic components in MSA. Few studies have provided more accurate objective descriptions of dysarthria in APS. In general, these studies have shown that the impairment of specific speech dimensions is more pronounced in APS than in PD. However, little effort has been put into the investigation of complex speech impairment in APS. Recently, a direct and objective comparison between individual speech patterns in PSP and MSA was proposed in a pioneer study [2], where a classification accuracy score of 75% has been achieved. As a continuation of that work, we propose here a technique which yields higher performances in objective discrimination between PSP and MSA.

II. Material and method

From 2011 to 2014, 12 consecutive Czech patients with the clinical diagnosis of probable PSP (10 men, 2 women) and 13 patients with the diagnosis of probable MSA (6 men, 7 women) were recruited for the study. We refer to [2] for clinical characteristics of patients and recording details. Here, we use the speech data consisting in sustained phonations of the vowel /a/ and fast /pa/-/ta/-/ka/ syllable repetitions. In order to extract the acoustic cues of the speech signals, we use the ComParE set which has been the baseline feature set in the Parkinson's Disease Challenge of the INTERSPEECH'2015 conference [3]. It contains 6373 static features as functionals of low-level descriptor contours. ComParE is implemented in the open-source software openSMILE (audeering.com/research/opensmile/). We carry out classification using the Scikit-Learn toolkit (http://scikit-learn.org/) which implements a large variety of state of the art classifiers. Training and test is made using the Leave-One-Speaker-Out (LOSO) method, which is the most appropriate method in our setting. We performed classification experiments using a wide range of classifiers, but we only present the results of linear and radial basis function (RBF) support vector machines (SVM) and the Extremely Randomized Tress (ERT) classifier [4]. Indeed, SVM is the most widely used classifier in this kind of tasks. On the other hand, ERT is not commonly used, still it is the classifier which yielded significantly higher performances than all others, in term of accuracy and robustness.

K. Daoudi is with INRIA Bordeaux Sud-Ouest, France (phone: +33-524-574106; fax: +33-524-574124; e-mail: khalid.daoudi@inria.fr).

N. Brodu s with INRIA Bordeaux Sud-Ouest, France.

J. Rusz is with Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic.

J. Klempir is with Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University in Prague, Czech Republic.

III. Results

Table II presents the mean accuracy scores (in %) and their standard deviation (SD), computed over 10 seeds. We performed a greedy search over the hyper-parameters of each classifier to obtain the best score. The results show that classification using Phonations only is not satisfactory for both classifiers, though ERT is higher than SVM. On the other hand, using Syllables only, the performance of ERT is very good and is significantly better than SVM which is very low. When fusing the features, the performance of ERT increases further. Moreover, the best scores of SVM are obtained using speaker normalization of features which is difficult to obtain in a real-world application. ERT does not require such pre-processing and uses the raw features. This shows that the choice of ERT as a classifier is more appropriate than the traditional SVM. Indeed it achieves a classification score of 86.1%, which is a significant improvement w.r.t the 75% obtained in [2] using an SVM, and it is insensitive (and thus robust) to feature normalization.

TABLE I. Classification accuracy scores

	Best SVM	***ERT***
	% Mean accuracy / SD	% Mean accuracy / SD
Phonations	54.4 / 6.4	**64.8** / 3.2
Syllables	61.4 / 2.6	**82.2** / 2.8
Fusion	66 / 2	**86.1** / 2.1

In conclusion, we presented a fully reproducible speech-based technique which achieves a very good objective differential diagnosis between PSP and MSA. This technique can serve as a reference/baseline for other studies on this subject. We must however analyze carefully the influence of gender-unbalanced nature of the data.

References

[1] Kluin KJ, Gilman S, Lohman M, Junck L (1996) Characteristics of the dysarthria in multiple system atrophy. Arch Neurol 53:545–548

[2] Rusz J, et al. (2015). Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiples system atrophy, Dans : Journal of Neurology, Vol 262, Issue 4 , pp 992-1001.

[3] Sculler B, et al. (2015). The INTERSPEECH 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson's & Eating Condition. Proc. INTERSPEECH'2015.

[4] Geurts D, et al. (2006). Extremely randomized trees. Machine Learning, 63(1), 3-42, 2006.