Study of the automatic detection of Parkison's Disease based on speaker recognition technologies and allophonic distillation

L. Moro-Velazquez, J. A. Gomez-Garcia, J. I. Godino-Llorente, *Senior member, IEEE,* J. Rusz, S. Skodda, J. R. Orozco-Arroyave, E. Nöth, and N. Dehak, *Senior member, IEEE.*

***Abstract*— The use of new tools to detect Parkinson´s Disease (PD) from speech articulatory movements can have a considerable impact in the diagnosis of patients. In this study, a novel approach involving speaker recognition techniques with allophonic distillation is proposed and tested separately in four parkinsonian speech databases (205 patients and 186 controls in total). The results of applying this new scheme in the databases provides up to 94% of accuracy in the automatic detection of PD and improvements up to 9% respect to baseline techniques. Results not only point towards the importance of the segmentation of the speech for the differentiation of parkinsonian and control speakers but confirm previous findings about the relevance of plosives and fricatives in the detection of parkinsonian dysarthria.**

I. Introduction

Idiopathic Parkinson's Disease (PD) is a condition affecting the coordination of movements, and therefore, it might be feasible to monitor and analyze the behavior of patients during the performance of a complex motor task, and employ such assessment for diagnosis purposes. Speech, being a task which is almost universal, involves a large amount of muscles and the coordination of very precise movements, is a good candidate for that evaluation.

One of the earliest works studying parkinsonian dysarthria and its related articulatory aspects from a phonetic point of view is [1], in which authors perceptually analyze the speech dysfunctions of 200 untreated patients. Results show that a 90% of the patients exhibit some type of dysphony or misarticulation and that the articulatory errors are mainly concentrated in the consonants requiring the greatest narrowing or closure during articulation, specially plosives and fricatives, with more errors found in velar articulations, mostly /k/ and /g/ phonemes.

More recent works explore the automatic detection or assessment of PD by analyzing different types of segments of speech, the velocity or acceleration of articulators, specific transitions or the evolution of formants [2]–[4]. As an example, work [2] reports 80% of accuracy in PD detection using acoustic features extracted from only vowels in monologues. In the study [3], authors automatically subdivide to obtain, the initial burst, onset and occlusion of allophones to analyze articulation, obtaining 88% of efficiency in differentiating PD from control subjects.

In the present study, the importance of the distinct allophones attending to the articulatory mode (i.e., plosives, fricatives, nasals, liquids or vowels) and its application in the detection of PD is assessed within a state-of-the-art speaker and speech recognition frameworks: Gaussian Mixture Models-Universal Background Model (GMM-UBM) and speech forced alignment (SFA).

II. Theoretical introduction

A. Phonetics and allophonic distillation

Different types of categorization of the allophones can be found in literature, from which the one proposed by [5] concerning the mode of articulation in Spanish language is considered here. In this categorization, allophones can be divided into vowels and consonants, and consonants can be subdivided into plosives, fricatives, affricates, nasals and liquids. In a direct sense, this categorization is related to the constriction of the articulators found in the

L. M-V, J. A. G-G and J. I. G-L. are with Universidad Politécnica de Madrid, Madrid, 28031, Spain (corresponding author phone: +34 913365527; e-mail: laureano.moro@upm.es).

L. M-V and N. D. are with Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD 21218 USA.

J. R. Author is with the Faculty of Electrical Engineering, Czech Technical University in Prague, 160 00 Prague, Czech Republic

S. S. is with the Department of Neurology, Knappschaftskrankenhaus, Ruhr University Bochum, 44801 Bochum, Germany.

J.R. O-A is with the Faculty of Engineering, Universidad de Antioquia UdeA, 1226 Medellín, Colombia.

E. N. is with the Pattern Recognition Laboratory, Friedrich-Alexander-Universitat Erlangen-Nurnberg, 91054 Erlangen, Germany.

supralaryngeal cavities and to the on/off mechanisms of the glottal source.

The plosive consonants are preceded by a stop or a total obstruction of the articulators while in fricatives the constriction is not complete. Affricates are a mix between plosives and fricatives, and liquids are similar to fricatives but with less approximation of articulators. Nasals are sonorant consonants produced when the soft palate is lowered and the air coming from the glottis passes through the nasal cavities.

The allophonic distillation technique proposed in this paper consists in selecting specific allophones from a certain speech signal attending to the manner of articulation and discarding the rest. The method employed here to identify these specific allophones in a recording is the SFA.

B. Speech forced alignment

SFA technologies are used to automatically detect specific segments (allophones) in a certain utterance knowing its orthographic transcription. In general, these techniques analyze the speech signal, and provide its automatic segmentation into separated allophones with their correspondent labeling. An example of the result of a SFA process is shown in Fig. 1.

III. MATERIALS

A. Parkinsonian speech databases

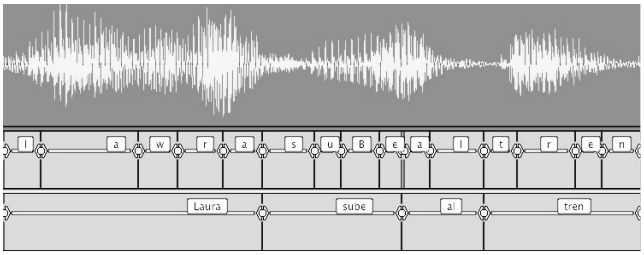
Four databases, *GITA, Neurovoz, German and Czech*, containing utterances from PD patients and age-matched control speakers were used separately in this study to train and test models for automatic detection of PD. These utterances consisted in Diadochokinetic (DDK) tasks (repetition of the sequence /pa-ta-ka/).

The first database, *GITA*, includes speech from 50 parkinsonian and 50 control Colombian speakers (25 men in each group) [6]. The DDK tasks used from *Neurovoz,* a new Spanish corpus, were uttered by 47 parkinsonian and 32 control speakers (18 women in each group). The *German* database contains utterances from 88 parkinsonian (47 men) and 88 control (44 women) German speakers as described in [7]. Finally, the *Czech* database includes DDK tasks from 20 patients and 16 controls (all were men) in Czech language, as described in [2]. In this last one, unlike the rest of databases, all the patients were newly diagnosed and were not under treatment.

B. UBM and SFA databases: Albayzin-FisherSP

The *Albayzin* database is composed by 4800 utterances from 136 different speakers in Spanish language as described in [8]. This database is used to create the UBM. The *FisherSP* database contains 163 hours of telephone speech from 136 speakers in Spanish with their respective orthographic transcripts. FisherSP is used to train a Forced Alignment Model (FAM) with Kaldi development kit [9].

Figure 1. Example of SFA in the utterance “Laura sube al tren”.



IV. METHODS

The main objective of this study was to analyze the influence of certain allophones in GMM-UBM classifiers to detect PD from speech using different databases. This new approach consists in using only a specific type of allophone in the UBM database by applying an allophonic distillation. The idea is to produce GMM-UBM classifiers more influenced by the acoustic characteristics of these allophones to detect PD. This approach was complemented with a group of trials including score fusion, while all results were compared with a baseline in which the UBM was not distilled.

A. Baseline: GMM-UBM

In a first stage, several trials using state-of-the-art speaker recognition technologies, GMM-UBM, were performed. At this stage, all four parkinsonian speech databases were trained and tested separately following a k-folds cross-validation scheme, with k=11. These trials were performed using the configurations leading to best results in [10]: Rasta-PLP+Δ+ΔΔ with number of Rasta-PLP coefficients ranging from 10 to 20 in steps of 2 and 5 coefficients in the FIR filter to calculate derivatives, obtained from 15 ms frames with 50% of overlapping. As the UBM database is

sampled at 16kHz, the four databases were downsampled to this frequency. For all results, in this and further stages, equal error rate from training scores was used as threshold for test scores and Confidence Intervals (CI) were calculated as detailed in [11].

B. GMM-UBM with allophonic distillation

In this second stage, new trials following the same type of processing and classification scheme described for the previous stage were performed, but in this case, an allophonic distillation is applied to the UBM database, *Albayzin,* to obtain, in turns, fricative, liquid, nasal, plosive and vowel segments. Affricates were not used as they were scarce in the UBM database. This new approach allowed to analyze the importance of the initialization of the GMM model, as the distillation in the UBM database produces models more oriented to the acoustic characteristics of fricatives, liquids, nasals, plosives or vowels.

To obtain the distillation, a FAM was applied to the utterances of *Albayzin* and its transcriptions. A diagram of this stage is shown in Fig. 2.

Although the *Albayzin* is in Spanish, the influence of its distinct distillations can be useful for the initialization of the GMM-UBM binary classifiers since the pronunciation of /pa-ta-ka/ is very close in Spanish, German an Czech.

C. Fusion of pairs of scores from GMM-UBM with allophonic distillation

After finishing the stage represented by Fig. 2, the scores are fused in pairs following a logistic regression scheme considering all the possible combinations of pairs coming from different types of distillations (plosive-fricative, plosive-liquid, nasal-vowels, etc.)

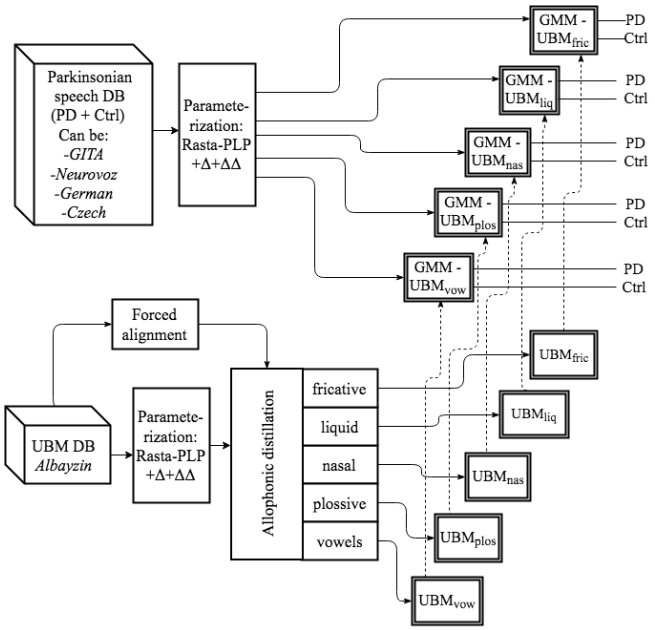
V. RESULTS

Best results of the first stage of trials can be found in Table I. This table includes accuracy, CI, AUC, sensitivity, specificity, number of Rasta-PLP coefficients employed and number of Gaussians.

Results of the second stage of trials can be found in Table II. For each database, best results are marked in bold and results outperforming the baseline are highlighted as a shaded row.

Finally, Table III shows the best results of the pairs of scores fusion, including the type of combination. For the sake of simplicity, only best results for each database are referred.

Figure 2. Diagram of proposed methodology using allophonic distillation.



VI. DISCUSSION

In this work, a novel methodology employing allophonic distillation in the creation of GMM-UBM models was studied for the detection of PD from speech. This methodology was compared with traditional GMM-UBM techniques in four different databases including parkinsonian and normophonic speakers.

The obtained baseline results are in the range of those obtained in [10], except for the *German* database. These differences can be attributable to the use of *Albayzin* as UBM. A comparison of Tables I and II, reveals that the use of allophonic distillation in the UBM produces improvements in the results of all databases and with most of the distillations, suggesting that this technique outperforms in accuracy and reliability to GMM-UBM in the automatic detection of PD.

The highest accuracy obtained in the trials is 94% with AUC of 0.96, sensitivity of 0.90 and specificity of 1.00, in the *Czech* database with fricative allophonic distillation. Considering that this database contains only newly diagnosed patients, results suggest that the new proposed approach is successful in detecting PD in early

stages. Also, although *Czech* and *Neurovoz* are class-unbalanced. The observed sensitivity and specificity never differ more than 0.10 absolute points.

The fusion of scores following a logistic regression scheme produces better results in the trials of *German* and *GITA* and exhibit no changes of accuracy in *Neurovoz* and *Czech*. This suggest that the fusion of scores can be an appropriate complement to the proposed scheme.

TABLE I. BASELINE RESULTS INCLUDING ACCURACY±CI, AUC, SENSIBILITY, SPECIFICITY AND NUMBER OF COEFFICIENTS AND GAUSSIANS

Database	Acc. ±CI(%)	AUC	Sens.	Spec.	# Coeff.	# Gauss
GITA	81±8	0.88	0.82	0.8	18	256
Neurovoz	79±9	0.85	0.87	0.65	14	64
German	70±7	0.75	0.70	0.69	14	8
Czech	88±0	0.94	0.85	0.93	18	32

TABLE II. RESULTS OF GMM-UBM WITH ALLOPHONIC DISTILLATION INCLUDING ACCURACY±CI, AUC, SENSIBILITY, SPECIFICITY AND NUMBER OF COEFFICIENTS AND GAUSSIANS

Database	Distill.	Acc.± CI(%)	AUC	Sens.	Spec.	# Coeff.	# Gauss
GITA	*Fricative*	80±8	0.85	0.78	0.82	14	256
	Liquid	82±8	0.87	0.80	0.84	14	256
	Nasal	**83±7**	**0.89**	**0.86**	**0.80**	**10**	**32**
	Plosive	82±8	0.88	0.86	0.78	10	8
	Vowels	83±7	0.88	0.86	0.80	14	16
Neurovoz	*Fricative*	83±9	0.9	0.89	0.73	10	64
	Liquid	81±9	0.89	0.85	0.73	10	32
	Nasal	82±9	0.87	0.85	0.77	18	64
	Plosive	**86±8**	**0.88**	**0.89**	**0.81**	**14**	**64**
	Vowels	81±9	0.88	0.87	0.69	14	64
German	***Fricative***	**72±7**	**0.8**	**0.72**	**0.73**	**14**	**8**
	Liquid	69±7	0.76	0.69	0.69	14	8
	Nasal	69±7	0.76	0.67	0.70	10	4
	Plosive	70±7	0.74	0.72	0.69	20	256
	Vowels	69±7	0.73	0.67	0.7	18	128
Czech	***Fricative***	**94±0**	**0.96**	**0.90**	**1.00**	**18**	**64**
	Liquid	94±0	0.95	0.90	1.00	20	64
	Nasal	91±0	0.95	0.90	0.93	16	64
	Plosive	91±0	0.96	0.85	1.00	14	16
	Vowels	91±0	0.95	0.90	0.93	14	32

Analyzing the influence of the type of allophone employed in the distillation, plosives provide the best results in *Neurovoz* while fricatives are the ones producing the best accuracies in *German* and *Czech* databases. Nasal and vowel allophones are the most significant in *GITA* database implying that the models tend to be slightly more efficient using sonorant segments with this database, probably due to particularities of regional pronunciation. Observing the accuracies in Tables II and III, fricative distillations produce the best results, in concordance with previous findings [1]. However, the sequence of syllables employed in the recordings (/pa-ta-ka/) contains only plosives. This fact can be explained by the inclination of many patients to replace plosives with fricatives [4], phenomenon (known as spirantization) which can be better detected with a classifier more influenced by the acoustic characteristics of fricatives or plosives.

To conclude with more discernment which types of allophones are more influenced by PD with automatic detection purposes, the allophonic distillation must be performed directly in the parkinsonian databases. Therefore, in future works and attending to the observed results, the study of the allophonic distillation in training-testing databases is recommended as well as inter-database trials.

TABLE III. BEST RESULTS AFTER FUSION OF SCORES, INCLUDING THE BEST COMBINATION, ACCURACY±CI, AUC, SENSIBILITYAND SPECIFICITY

Data Base	Allophonic distillation combination	Acc. ±CI(%)	AUC	Sens.	Spec.
GITA	nasal-liquide	86±7	0.89	0.86	0.86
Neurovoz	plosive-fricative	86±8	0.9	0.89	0.81
German	fricative-nasal	74±6	0.8	0.74	0.74
Czech	plosive-fricative	94±0	0.96	0.9	1

REFERENCES

[1] J. A. Logemann, et al. "Frequency and Cooccurrence of Vocal Tract Dysfunctions in the Speech of a Large Sample of Parkinson Patients," *J. Speech Hear. Disord.*, vol. 43, no. 1, p. 47, Feb. 1978.

[2] J. Rusz, et al. "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 2171–2181, 2013.

[3] M. Novotny and J. Rusz, "Automatic evaluation of articulatory disorders in Parkinson's disease," *IEEE/ACM Trans. audio, speech, Lang. Process.*, vol. 22, no. 9, pp. 1366–1378, 2014.

[4] J. I. Godino-Llorente, et al. "Towards the identification of Idiopathic Parkinson's Disease from the speech. New articulatory kinetic biomarkers," *PLoS One*, vol. 12, no. 12, , Dec. 2017.

[5] A. Quilis, *Tratado de fonología y fonética españolas*, 2nd ed. Editorial Gredos, 1999.

[6] J. R. Orozco, et al. "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," *Lr. 2014. Proc. Ninth Int. Conf. Lang. Resour. Eval.*, pp. 342–347, 2014.

[7] S. Skodda, et al. "Intonation and speech rate in parkinson's disease: General and dynamic aspects and responsiveness to levodopa admission," *J. Voice*, vol. 25, no. 4, pp. e199–e205, 2011.

[8] A. Moreno *et al.*, "Albayzín speech database: Design of the phonetic corpus," *Eurospeech 1993. Proc. 3rd Eur. Conf. Speech Commun. Technol.*, vol. 1, no. JANUARY, pp. 175–178, 1993.

[9] D. Povey *et al.*, "The Kaldi speech recognition toolkit," in *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011, pp. 1–4.

[10] L. Moro-Velázquez, et al. "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's Disease," *Appl. Soft Comput. J.*, vol. 62, pp. 649–666, 2018.

[11] N. Saenz-Lechon, et al. "Methodological issues in the development of automatic systems for voice pathology detection," *Biomed. Signal Process. Control*, vol. 1, no. 2, pp. 120–128, 2006.