Acoustic Characteristics of VOT in Plosive Consonants Produced by Parkinson's Patients

Patricia Argüello-Vélez[1,5], Tomas Arias-Vergara[2,3,4(✉)], María Claudia González-Rátiva[1], Juan Rafael Orozco-Arroyave[2,3], Elmar Nöth[3], and Maria Elke Schuster[4]

[1] Faculty of Communications, Universidad de Antioquia UdeA, Calle 70 N, 52-21 Medellín, Colombia
[2] Faculty of Engineering, Universidad de Antioquia UdeA, Calle 70 N, 52-21 Medellín, Colombia
tomas.arias@udea.edu.co
[3] Pattern Recognition Lab., Friedrich-Alexander University, Erlangen-Nürnberg, Germany
[4] Department of Otorhinolaryngology, Head and Neck Surgery, Ludwig-Maximilians University, Munich, Germany
[5] Facultad de Salud, Universidad Santiago de Cali, Cali, Colombia

Abstract. Voice Onset Time (VOT) has been used as an acoustic measure for a better understanding of the impact of different motor speech disorders in speech production. The purpose of our paper is to present a methodology for the manual measuring of VOT in voiceless plosive sounds and to analyze its suitability to detect specific articulation problems in Parkinson's disease (PD) patients. The experiments are performed with recordings of the diadochokinetic evaluation which consists in the rapid repetition of the syllables /pa-ta-ka/. A total of 50 PD patients and 50 healthy speakers (HC) participated in this study. Manual measurements include VOT values and also duration of the closure phase, duration of the consonant, and the maximum spectral energy during the burst phase. Results indicate that the methodology is consistent and allows the automatic classification between PD patients and healthy speakers with accuracies of up to 77 %.

Keywords: Voice onset time · Acoustic analysis · Speech processing · Diadochokinesis

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive loss of neurons in the mid-brain [6]. Primary motor symptoms include

Supported by Antioquia University.

P. Sojka et al. (Eds.): TSD 2020, LNAI 12284, pp. 303–311, 2020.
https://doi.org/10.1007/978-3-030-58323-1_33

tremor, rigidity, freezing of gait, and postural instability. PD also affects muscles involved in the speech production process, resulting in hypokinectic dysarthria, which is a set of motor speech disorders including bradylalia, lack of articulation accuracy, and dysphonia [4]. Many of the symptoms are controlled with medication, however there is no clear evidence indicating the positive effects of those treatments to reduce motor speech disorders. Proper speech therapy combined with the pharmacological treatment improve the communication ability of PD patients [16]. Thus, it makes sense to develop methodologies based on acoustic analysis to evaluate speech impairments in PD patients. The resources of instrumental phonetics allow acoustic analysis of segmental and supra-segmental characteristics ranging from isolated sounds to spontaneous speech analysis. These tools allow a linguistic and physiological understanding of atypical phenomena in speech and their relationship with the presence of symptoms related with subglottic, glottic or supra-glottic nature. In the case of PD, the diadochokinetic exercises (DDK) are used to study the production of voiceless plosive consonants (VPC) and vocal segments, which allows the analysis of coordination in supra-glottic and glottic components [2].

The pronunciation of the VPCs /p/,/t/, and /k/ involves the production of VOT which is defined as the time between the burst and the beginning of the emission of the next vowel [9]. In hypokinetic dysarthria VOT is useful to assess speech impairments by identifying the increase or decrease in duration. These patterns may be related to glottal adduction, degree of vocal cord tension, and quality of intra-oral pressures [5]. Preliminary results describe a significant VOT reduction in some VPCs, which can be explained by the loss of neuromuscular control during speech production. There is also another hypothesis where increase in VOT is related with loss of laryngeal and supra-laryngeal coordination [10].

VOT is measured by manual syllable-by-syllable labeling. The wide-band spectrogram shows the frequency and its relation with the duration of the signal. Abrupt changes associated to physiological phenomena can also be captured. The burst is visually identified as a short explosion bar with energy distributed over the entire frequency spectrum. The onset of the vowel is identified as the high energy values with the corresponding formant structure in the spectrogram [8]. Manual labeling requires to consider the relationship among these results and the signal represented in the oscillogram to have a better view of the signal's disturbances including abnormal changes in the consonant to vowel segment [1]. There are other measurement parameters to define acoustic integrity of VPCs such as the duration of the closure and the point of maximum spectral energy in the burst. These characteristics are considered in this paper through manual labeling and represent physiological correlations with supra-glottic pressure and articulatory accuracy of the sound.

From the automation/engineering point of view, VOT has been considered in DDK exercises to extract relevant information such as VOT duration and VOT ratio between VOT and vowel length (CV ratio). These parameters were used in [11] to classify between PD patients and healthy speakers. The authors

reported accuracies of 92.2%. In [12] the authors measured VOT and obtained articulatory characteristics in relation to physiological correlates of vocal quality, articulatory accuracy, occlusion quality, and glottal and supra-glottal coordination. In general, automated methods help in reducing costs and time of clinical screenings including those that required to evaluate and monitor motor speech disorder in PD patients [14]. The automatic computational methods should consider the existence of approximate or incomplete productions in DDK tasks. Note that the alternating and rapid repetition of the plosives /p/, /t/, and /k/ generate variations that reveal patterns like debilitation of the burst, presence of voicing and loss of the articulatory tension. Fusion of automatic and manual labeling methods may improve the accuracy in detecting the aforementioned variations and allow the description of the acoustic "correctness" of each consonant segment. The purpose of our paper is to present a methodology for measuring VOT of VPCs in /pa-ta-ka/ using phonetic-acoustic manual methods and the automatic measurement method to evaluate and analyze the accuracy of the detection of speech disorders in PD patients.

2 Methodology

Details of the proposed methodology are provided below. It includes the description of the database and the steps followed in the manual labeling process along with its automatic evaluation.

2.1 Data

Speech recordings of the PC-GITA database are considered [13]. This corpus includes 50 PD patients and 50 healthy speakers. The participants were asked to perform the rapid repetition of /pa-ta-ka/ for at least 3 seconds. The speech signals were captured in a sound-proof booth using a professional audio setting. All of the patients were evaluated by a neurologist expert following the MDS-UPDRS-III scale [7]. Table 1 summarizes demographic and clinical information of the speakers.

Table 1. Clinical and demographic information of the speakers. Values in terms of (Mean ± Standard deviation).

	PD patients		HC speakers	
	Male	Female	Male	Female
Number of speakers	25	25	25	25
Age [Years]	61 ± 11	60 ± 7	60 ± 11	61 ± 7
Years diagnosed	13 ± 11	9 ± 5	–	–
MDS-UPDRS-III	37.4 ± 21.7	37.6 ± 14.0	–	–

2.2 Acoustic Phonetic Analysis of VPCs

Voiceless plosive consonants are characterized by three stages in Spanish: approach, closure, and release. During the approach phase, the articulators move towards each other, creating an obstruction of airflow during the closure phase. Finally, the articulators move away from each other during the release phase producing an explosive burst of air with energy spread across the audible spectrum. Figure 1 shows time and spectral representation of the VPC /p/, followed by the vowel /a/. The shaded regions represent the closure and the release phases. The closure phase is characterized by the absence of speech, which can be observed in the time signal (Fig. 1A) and in the spectrogram (Fig. 1B). The release stage is typically observed by looking at the spectral representation. Precision to produce voiceless plosive sounds can be reduced due to the presence of a motor problem. Loss of pressure in the lips changes /p/ sounds; impaired movement of the tongue affects /t/ sounds, and loss of contact with the soft palate and no burst deviates /k/ sounds. Altered versions of these sounds are perceived as /β/, /δ/, and /γ/, instead of /p/, /t/, and /k/, respectively. These weakened consonants are characterized by the loss of tension, increased voicing and incomplete contact of the articulators. As a result, there is no silence nor burst during the closure and release phases, respectively. Figure 2 shows an example of a weak consonant /β/ (/p/), produced in the transition from one utterance of /pa-ta-ka/ to another.

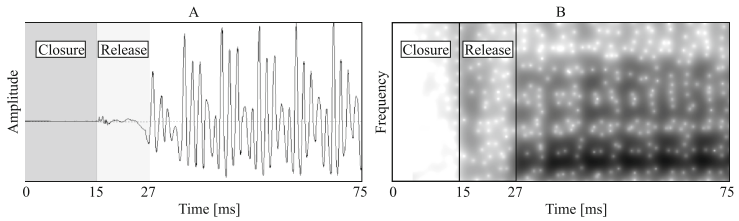

Fig. 1. Time (Figure A) and spectral (Figure B) representations of the voiceless plosive sound /p/. The shaded regions represent the closure and release phases.

2.3 Manual Labeling

Manual labels are found using the software Praat [3], as follows:

1. Determine the start and total duration of the consonant in a syllable by syllable fashion. The total duration of the VPC is measured as the total duration of the approach, closure, and release phases.
2. Identify the closure phase as the time prior to the burst indicated by the point of minimum intensity relative to the surrounding sounds, i.e., tension phase with increased supra-glottic pressure.

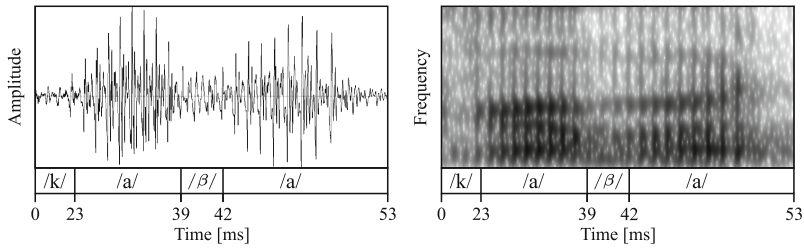

Fig. 2. Weak consonant /β/ (/p/) produced in the syllables /pa-ta-ka/.

3. Measure the maximum spectral energy during the release stage. A spectral slice is extracted in the first energy burst produced in a VPC sound. Then, the point with highest spectral energy is extracted.
4. Measure the VOT by placing labels at the initial burst of the consonant and vowel onset. The time of the initial burst is detected by computing the zero crossing points. The vowel onset is set at the beginning of a periodic-like signal. Formant frequencies and the presence of pitch are used to mark the beginning of voicing in a stop-vowel transition.
5. Identify weak consonants /β/, /δ/, and /γ/, in order to measure negative VOT. In this case, VOT is measured as the time between the end of the previous vowel and the beginning of the next one. The energy between syllables, e.g. /ka/ to /βa/, /pa/ to /δa/, and /ta/ to /γa/, is computed to detect the beginning of the weakened consonant.

These steps are followed to extract the acoustic measures: VOT value, duration of the closure phase, total duration of the consonant, and the frequency point with the maximum spectral energy measured in the burst phase. Mean value, standard deviation, kurtosis, and skewness are computed from the acoustic measures to create 48-dimensional feature vectors (4 measurements * 3 plosives * 4 functionals) per speaker.

2.4 Automatic Classification Between PD Patients and HC Speakers

A radial basis function – Support Vector Machine (rbf–SVM) with margin parameter C and kernel bandwidth γ is considered. Parameters are optimized through a grid search with $10^{-4} < C < 10^4$ and $10^{-6} < \gamma < 10^3$. The selection criterion is based on the performance obtained in the training set following a 10-fold cross validation strategy. The performance of the system is evaluated by means of accuracy (Acc), sensitivity (Sen), specificity (Spe), and the F1-score. Additionally, the Area Under the ROC Curve (AUC) is considered to present results more compactly. The values of the AUC range from 0.0 to 1.0, were 1.0 means perfect classification.

3 Results and Discussion

3.1 Preliminary Observations with Manual Labels

To evaluate the suitability of the proposed method to detect abnormal production of VPCs based on manual labels, the following measures are considered: the VOT, duration of the closure phase, total duration of the consonant, and the frequency point with the maximum spectral energy measured in the burst phase. Figure 3 shows box plots with the four manual acoustic features. Kruskal-Wallis tests were applied and significant differences are found in almost all of the acoustic features except for the duration of the consonants /p/ and /t/ and the frequency of burst in /t/ and /k/. Figure 4A shows the number of positive and negative VOTs measured for the PD and HC groups. In general, the number of positive VOT values is higher than the negative ones. Also, the number of negative VOTs is higher in PD compared with respect to the HC group. Figures 4B and 4C show the number of negative VOT per consonant measured in the HC and PD groups, respectively. The presence of the weak consonants indicates loss of acoustic integrity of VPC, this phenomenon was observed in both HC and PD speakers. As shown in Figs. 4B and 4C, /p/ is more sensible to turn into its voiced version /β/. Finally, it is relatively common to observe an increase in the approximations in /β/ and /γ/ within the HC group. This is because in Spanish the DDK /pa-ta-ka/ has the energy of the accent in /tá/, which maintains the closure and the explosion in this dental sound.

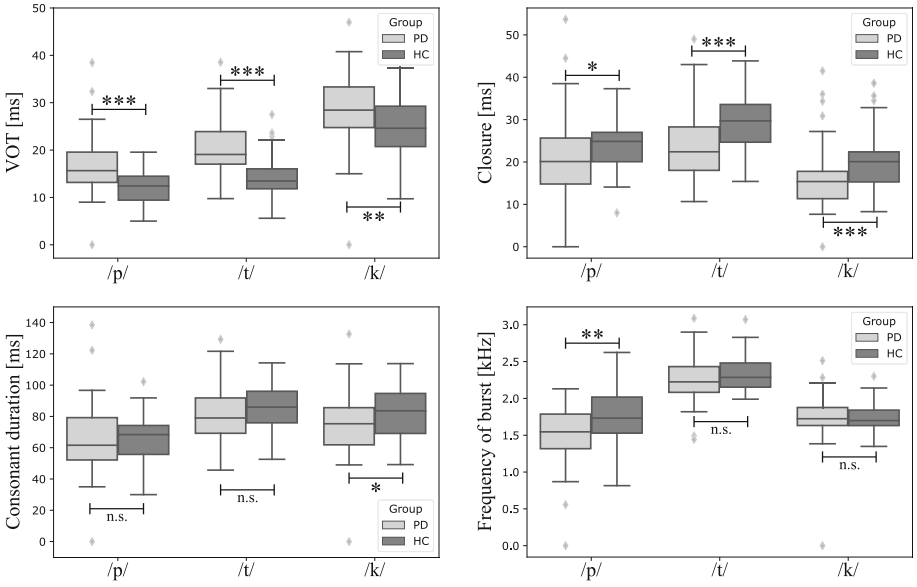

Fig. 3. Box-plots of acoustic measures extracted from each group. Kruskal-Wallis tests were applied with the following significance criteria: p-values: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ and n.s. (non-significant). Light grey diamonds represent outliers.

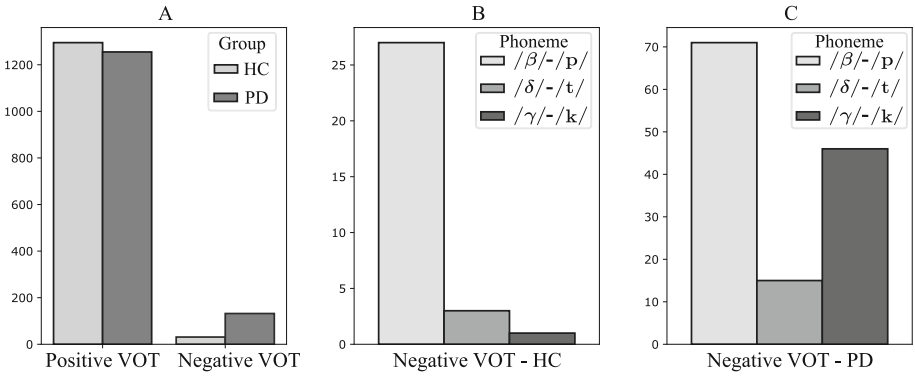

Fig. 4. Number of positive and negative VOT measurements found in our data. Figure 4A shows the number of positive and negative VOTs measured for the PD and HC groups. Figure 4B shows the number of negative VOTs measured in the HC group for each consonant. Figure 4C shows the number of negative VOTs measured in the PD group for each consonant.

3.2 Automatic Classification of PD vs. HC Subjects

Table 2 shows the performance of the rbf–SVM classifier. Four scenarios were considered: feature vectors from consonants /p/, /t/, /k/, and the combination of all. The highest accuracies were obtained with the SVMs trained with features from the consonant /p/ and with features from the three consonants (Acc = 77 %). Furthermore, the lowest accuracy was obtained when the rbf–SVM is trained only with features of the consonant /k/ (Acc = 68 %).

Table 2. Classification results (PD vs. HC) using manually extracted acoustic measures. **Acc:** Accuracy. **Sen:** Sensitivity. **Spe:** Specificity. **AUC:** Area under the ROC curve

Feature set	Acc (%)	Sen (%)	Spe (%)	F1-score	AUC
Consonant /p/	77	80	74	0.77	0.82
Consonant /t/	71	70	72	0.71	0.76
Consonant /k/	68	84	52	0.67	0.74
All consonants	77	76	78	0.77	0.83

The results obtained here confirm that articulatory imprecision is a common characteristic in PD speech and it is exhibited as a slowing down in the transition towards the beginning of the vowel. The decrease in supra-glottic tension, which debilitates the frequency burst in the consonants, is also confirmed by experiments. VOT allows accurate classification of PD vs. HC people. As indicated in [17] there is an increas in VOT of PD patients which is associated with

voicing and aspiration. The method proposed in this paper shows accuracies of up to 77% in the consonant /p/, which is approximated as β in several cases due to the absence of tension before the burst as a decrease in the closure in bilabial sounds [15].

4 Conclusions

Acoustic phonetic analysis with manual labeling allows validation of the acoustic characteristics of consonants and their variations depending on the linguistic context while automatic methods are established as rapid detection tools that together determine the accuracy, sensitivity and specificity of cases. In this way, the fusion of both methodologies makes possible the classification between healthy people and people with PD from the measurement of VOT and closure. In the future, we plan to speed up the labeling process and to automatize the feature extraction.

Acknowledgments. The authors acknowledge to the Training Network on Automatic Processing of PAthological Speech (TAPAS) funded by the Horizon 2020 programme of the European Commission. Tomás Arias-Vergara is under grants of Convocatoria Doctorado Nacional-785 financed by COLCIENCIAS. The authors also thanks to CODI from University of Antioquia (grant Numbers 2018-23541 and 2017-15530).

References

1. Abramson, A.S., Whalen, D.H.: Voice Onset Time (VOT) at 50: theoretical and practical issues in measuring voicing distinctions. J. Phonetics **63**, 75–86 (2017)
2. Ackermann, H., Hertrich, I., Hehr, T.: Oral Diadochokinesis in Neurological Dysarthrias. Folia Phoniatrica et Logopaedica **47**(1), 15–23 (1995). https://doi.org/10.1159/000266338
3. Boersma, P., et al.: Praat, a system for doing phonetics by computer. Glot. Int. **5**, 341–345 (2002)
4. Darley, F., Aronson, A., Brown, J.: Clusters of deviant speech dimensions inthe dysarthrias. J. Speech Hear. Res. **12**(3), 462–496 (1969). https://doi.org/10.1044/jshr.1203.462
5. Forrest, K., et al.: Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. J. Acoust. Soc. Am. **85**(6), 2608–2622 (1989). https://doi.org/10.1121/1.397755
6. Gelb, D.J., Oliver, E., Gilman, S.: Diagnostic criteria for Parkinson disease. Arch. Neurol. **56**(1), 33–39 (1999)
7. Goetz, C.G., et al.: Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Move. Disord. **23**(15), 2129–2170 (2008). https://doi.org/10.1002/mds.22340
8. Klatt, D.H.: Voice onset time, frication, and aspiration in word-initial consonant clusters. J. Speech Hear. Res. **18**(4), 686–706 (1975), http://www.ncbi.nlm.nih.gov/pubmed/1207100

9. Lisker, L., Abramsson, A.: A cross-language study of voicing in initial stops: acoustical measurements. WORD **20**(3), 384–422 (1964). https://doi.org/10.1080/00437956.1964.11659830
10. Martínez-Fernández, R., Gasca-Salas, C., Sanchéz -Ferro, A., Ángel Obeso, J.: Actualización en la enfermedad de Parkinson. Parkinson's disease: a review. Revista Médica Clínica Las Condes **27**(3), 364–376 (2016). https://doi.org/10.1016/j.rmclc.2016.06.010
11. Montaña, D., Campos-Roca, Y., Pérez, C.J.: A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease. Comput. Methods Programs Biomed. **154**, 89–97 (2018). https://doi.org/10.1016/J.CMPB.2017.11.010
12. Novotny, M., Rusz, J., Cmejla, R., Ruzicka, E.: Automatic evaluation of articulatory disorders in Parkinson's Disease. IEEE/ACM Trans. Audio Speech Lang. Process. **22**(9), 1366–1378 (2014). https://doi.org/10.1109/TASLP.2014.2329734, http://ieeexplore.ieee.org/document/6827910/
13. Orozco-Arroyave, J.R., et al.: New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Language Resources and Evaluation Conference, (LREC), pp. 342–347 (2014)
14. Orozco-Arroyave, J.R., et al.: NeuroSpeech: an open-source software for Parkinson's speech analysis. Dig. Signal Process. **77**, 207–221 (2018)
15. Parveen, S., Goberman, A.M.: Presence of stop bursts and multiple bursts in individuals with Parkinson disease. Int. J. Speech-Lang. Pathol. **16**(5), 456–63 (2014). https://doi.org/10.3109/17549507.2013.808702
16. Schulz, G.M., Grant, M.K.: Effects of speech therapy and pharmacologic and surgical treatments on voice and speech in Parkinson's disease: a review of the literature. J. Commun. Disord. **33**(1), 59–88 (2000). https://doi.org/10.1016/s0021-9924(99)00025-8
17. Tykalova, T., et al.: Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Brain Lang. **165**, 1–9 (2017). https://doi.org/10.1016/J.BANDL.2016.11.005