

## ACOUSTIC STUDY OF VOICING ONSET TIME IN ENGLISH AND PASHTO PLOSIVE SOUNDS

Dr. Bilal Khan<sup>1\*</sup>, Mr. Tariq Amin<sup>2</sup>, Mr. Mudassir Inam<sup>3</sup>

<sup>1\*</sup>Lecturer, Department of English, FATA University, KP, Pakistan

<sup>2&3</sup>Lecturer, Department of English, University of Lakki Marwat, KP, Pakistan

<sup>1\*</sup>[bilal.khan@fu.edu.pk](mailto:bilal.khan@fu.edu.pk), <sup>2</sup>[tariq.ktk733@gmail.com](mailto:tariq.ktk733@gmail.com), <sup>3</sup>[mudassirinam25@gmail.com](mailto:mudassirinam25@gmail.com)

Corresponding Author: \*

Received: February 10, 2024

Revised: March 15 2024

Accepted: March 18, 2024

Published: March 21, 2024

### ABSTRACT

The main aim of this study is to analyze the Voicing Onset Time (VOT) in Pashto and English Plosive sounds through Praat Analysis. Additionally, the researchers intend to highlight the issues in the production of Plosive sounds by EFL Pashto learners due to the differences between the Voicing Onset Time in Pashto and English Plosives. The researchers have collected the data from L1 Pashto Speakers, EFL Pashto Learners and L1 English speakers through voice recordings and then these recordings have been analyzed through Praat Software. The study shows the clear difference in the VOT of the plosive sounds in both the languages. The VOT duration of English plosives are lengthier than the VOT durations of Pashto plosive sounds and thus Pashto EFL learners are unable to produce these sounds like English RP speakers. The study is beneficial for EFL learners of Pashto language and teachers of English as second or foreign language in the field of English Language Teaching. The EFL students and teachers would between realize the differences in the production of L2 English speech sounds and thus would enhance their pronunciation. The Study recommends that further studies should be conducted taking all speech sounds of Pashto and English language in terms of acoustic properties.

**Keywords:** Voicing Onset Time, Acoustic Study, Praat analysis, Pashto Plosives, EFL Learners

### INTRODUCTION

The aim of this study is to identify and compare the Voicing Onset Time in the plosive sounds of English and Pashto languages. A sound system is generally divided in two major types, consonants and vowels, consonant sounds also called stop sounds are those sounds in which there is a clear obstruction in the vocal tract or vocal cavity. On the other hand there is no obstruction the in production of vowel sounds. Plosive or stop sounds are those sounds in which the air is completely blocked in the mouth cavity and then released suddenly, such as /p/, /b/, /t/ etc. Apart from articulatory analysis of speech sound between different languages acoustic analysis of the speech sound is a laborious task as it need much scientific knowledge and analysis. (Chen et al. 2018, Jones et al. 2019, Xu. Et al. 2018). One of the major aspect of acoustic analysis is voicing

on set time (henceforth VOT). ). “Voice onset time (VOT) is an instantaneous acoustic parameter which is characterized as the time between the release of the full oral constriction for plosive production and onset of glottal vibrations” (Korkmaz and aytug, 2018, p.1). In the languages of the world, there is mainly one major distinction in stop sounds and thus there are either voiced stop sounds or voiceless stop sounds. (Osion, 2017). There are voiced and voiceless stop sounds in Pashto and English languages the researcher has limited this study to the stop sounds of these two languages. The researcher intends to record not only the VOT duration of L1 Pashto speakers but also of the EFL Pashto learners. Thus in the first place, the study compares the VOT duration of English plosive sounds produced by L1 Pashto speakers

and the EFL Pashto learners. After this comparison, the study focuses the comparison of the VOT duration of English stops produced by L1 Pashto speakers, EFL Pashto learners and L1 English speakers. Thus the study intends to bring out the difference in the VOT duration of English stops produced by EFL learners and L1 English speakers. The major reason behind highlighting this difference for the researchers is to know the difficulties that are faced by EFL learners in the Pashtoon area of Pakistan. The study is based on the Speech learning Model of Fledge in which he is of the view that identical sounds are produced in the same by two different language speakers but there are major acoustic differences in the similar sounds between languages. There are six stop sounds in English language while nine stop sounds in Pashto language. According to the opinion of the researcher, these differences and similarities help English language teachers in teaching English to Pashto speakers. Thus, they might realize in a better way the difficulties faced by their learners and will help the learners improve their pronunciation. It would lead to enhance the quality of pronunciation of the non-native English speakers, that is, Pashto speakers. The learners might be able to come to know about the problems they face in learning the pronunciation of a second language after reading the study. The study highlights the use of new techniques and scientific procedure in teaching pronunciation to EFL learners.

### Research Objectives

The present study aims:

- To identify the Voicing onset Time in Pashto Plosive sounds.
- To compare the VOT trends In Pashto and English Plosive sounds.

### Literature Review

Acoustic phonetics deals with the acoustic properties of the speech sounds which include loudness, resonance, closure, voicing on time, pitch and the waves through which the speech sounds are transferred from the speaker to the listener (Stevens, 1998). This study mainly focuses on the voicing on time which is a distinct acoustic property of Plosive sounds. VOT is the length of time that passes between the release of

a plosive consonant and the onset of voicing the vocal cords vibration. (Lisker & Abramson 1964; Zlatin 1974; Lieberman & Blumstein 1988). Voice onset time (VOT) is the time period between the release of any stop or plosive of full oral construction to the onset vocal vibration of any plosive sound. [Korkmaz and aytug, 2018]. This VOT is different is the voiceless and voiced plosive sounds. Voiced plosive are produced with a clear vibration of vocal cards while voiceless stops have no vibration in the vocal cards. This VOT is often used to differentiate between voiced and voiceless stops, [Osion, 2017]. This period of voicing onset time can be generally divided in to three different kinds.

Negative VOT: where the plosive release comes after the onset of vocal fold vibrations.

In this situation, which is also described as “voicing lead”, voicing begins before the release

of stop. Zero VOT: where the plosive release and the onset of vocal fold vibrations occurs

approximately at the same time. Positive VOT: where the plosive release precedes the

onset of vocal fold vibration. In this situation, which is described as voicing lag, the

release of stop occurs before the initialization of voicing. (Kaur, 2015)

VOT is the most necessary object in the interpretation of the plosives of a language (Kalita et al. 2018). The present research is also based on the comparative analysis of the acoustic properties of English and Pashto plosives. Thus, here the importance of VOT cannot be neglected. Lisker and Abramson (1964) claims that VOT is an important cue in analyzing plosive sounds. The duration of VOT changes with the change in the place of articulation in different sounds. The findings in this connection are discussed in three sections. The first one is that VOT is longer when the closure is back in the mouth cavity (Peterson and Lehiste, 1960). The second one is the VOT is longer if the contact area is extended (Stevens, Keyser & Kawasaki, 1986). The third finding is the VOT becomes shorter with the movement of the articulators, with faster movement the VOT is shorter (Hardcastle, 1973). These patterns

remained valid for many years. [Lisker and Abramson \(1964\)](#) further states that VOT is longer in velar plosives. Furthermore, they claim that VOT is shorter for both the aspirated and unaspirated bilabial plosives and intermediate for alveolar plosives. The only exception in this regard is those with plosives in Tamil and the aspirated plosives in Cantonese and Eastern Armenian.

[Maddieson \(1997\)](#) states that the duration of VOT in the plosives are different from plosive to plosive due to the differences in the size of the supraglottal cavity. There are two reasons behind the VOT differences. The first one is that there is a smaller volume of cavity in the velar plosives while the alveolar and bilabial plosives have comparatively larger volumes. The second reason is that the velar plosives have larger volume of cavity in the front than that of the alveolar or bilabial plosives ([Beckman et al. 2011](#)).

[Hardcastle \(1973\)](#) finds that VOT is different in the plosives sounds because of the fast movements of the tip of the tongue and lips. He states that the tongue tip moves faster than lower lip and the lower lip moves faster than the tongue body. As [Maddieson \(1997\)](#) notices that the important factor is the physiological movement of the organs of speech which bring variations in voice onset time and thus the faster movements cause shorter VOT.

[Yao Yao \(2007\)](#) mentions different factors which bring variations in the VOT of plosive sounds and he states that the important factor in this regard is the place of articulation. The researcher finds that VOT is longer in velar plosives while shorter for bilabial and alveolar plosive sounds. [Kuan Yi and C. Lei-Mei \(2008\)](#) have studied VOT in the voiceless plosives of Mandarin and English

languages. They came up with a conclusion that the values of VOT in these two languages are similar but not completely identical. Voiceless unaspirated [p, t, k], plosive realization in both languages occupy the same range: the short lag region along the VOT continuum

**Methodology**

The current study follows a mixed approach where the researcher has collected the data from L1 Pashto speakers and EFL Pashto learners using Yousafzai dialects. The data has been collected through audio recorder and then the data has been analyzed through Praat software. The population comprised different categories according to age, gender and education. Therefore the researcher has taken the sample according to these categories. In the current study the researcher has selected samples according to a particular age span, gender and education. The sample comprises two groups group A and B. Group A contains the educated L1 Pashto speakers learning English language at different institutes of Peshawar district. Their educational level is Master in English on the other hand group B contains uneducated L1 Pashto speakers of district Peshawar. 25 participants have been selected in both the groups for data collection.

The data has been collected from the respondents through pronouncing the Pashto and English Plosive sounds in vowel consonant vowel (VCV) format. The vowels before and after the consonant is the same /a/ sound in order to minimize the effect of the vowel on the plosive sounds. Pashto language has nine plosive sounds as shown in the following table. Two of them are bilabial, two dental, two alveopalatal, two velar while one uvular.

**Table 1 Pashto Stimuli list**

Place	VCV	VCV
Bilabial	/apa/	/aba/
Dental	/atta/	/adda/
Alveopalatal	/ata/	/ada/
Velar	/aka/	/aga/
Uvular	/aqa/	

Source: Authors

Unlike Pashto English language has six plosive sounds which are shown in the following figure. Two are bilingual, two alveolar and two velar. These plosive sounds are given in the following table.

**Table 2 English Stimuli list**

Place	VCV	VCV
Bilabial	/apa/	/aba/
Alveopalatal	/ata/	/ada/
Velar	/aka/	/aga/

Source: Authors

### 3.7. Data Collection Procedure

Each participant was made pronounce each word in VCV format of the stimuli list of Pashto two times which is 50 tokens for each word by 25 speakers. Total 450 tokens (50 each speaker multiplied by 9 words of the stimuli list) were collected from group A participants who are L1 Pashto speakers. On the other hand total 300 tokens (50 each speaker multiplied by 6 words of the stimuli list) were collected from the group B participants who are EFL Pashto learners. Total  $450 + 300 = 750$  tokens for each word of the stimuli list were recorded by using Praat software (Weenink & Boersma, 2009). The participant was asked to pronounce the word again in case of mispronunciation. The words were recorded with the help of a microphone through Praat which had a frequency of 8000 Hertz. An HP laptop and a microphone of high quality were used for recordings.

### Results and Discussion

This Chapter consists of two parts. In the first part of the chapter acoustic properties of Pashto and English Plosives produced by Group “A” participants and by group “B” participants are identified respectively. Then the acoustic properties of Pashto plosives by group A, English plosives by group B and English plosives by RP speakers are compared through figures and tables. The main focus is given to the comparison of Pashto plosives with each other first, and then the English plosives produced by Group B participants are analyzed and compared with each other in terms of closure, VOT and plosive duration. Lastly, a comparison between the closure, VOT and Consonant durations of English Plosives by native speakers, Pashto Plosives by group A speakers and English plosives by group B speakers is made. Special focus is given to the

differences and similarities between the different durations of plosives in the same language and between different languages. The first part is mainly concerned with the answers to the questions asked in the current study and the objectives of the study.

#### 4.1.1. Data Analysis of the Group A Speakers

In this section the acoustic properties of Pashto plosives has been identified and compared with each other with the help of tables and figures. At first average is taken of each participants values and then these averages are compared with each other. The following table shows the different averages of the acoustic properties of Pashto plosives by L1 Pashto speakers (group “A”).

There are nine plosives in Pashto language as shown in the above table. The table shows the average measurements of Closure duration, VOT and consonant duration of Pashto plosives produced by group “A” participants who are L1 Pashto monolingual speakers. The data in the table exhibits that maximum closure duration 0.134 m/sec for /p/ plosive and minimum 0.081 m/sec for /d/ plosive in all the plosives. On another side, maximum VOT 0.033 m/sec is noted for voiceless plosive /k/ and minimum VOT 0.007 is noted for voiced plosive /b/, while /p/ plosive has maximum consonant duration and /d/ plosive has minimum. The table is further analyzed below with the help of other tables and Figures.

Total 75 tokens were recorded of each plosive from 25 speakers who were L1 monolingual Pashto speakers. They were listed in group “A”. After the recordings an average is taken out of three tokens by every speakers which resulted in 25 average tokens, again an average is taken out

of each duration which is presented in the above table.

**Table: 3 Average VOT of Pashto Plosives Produced by Group “A” speakers**

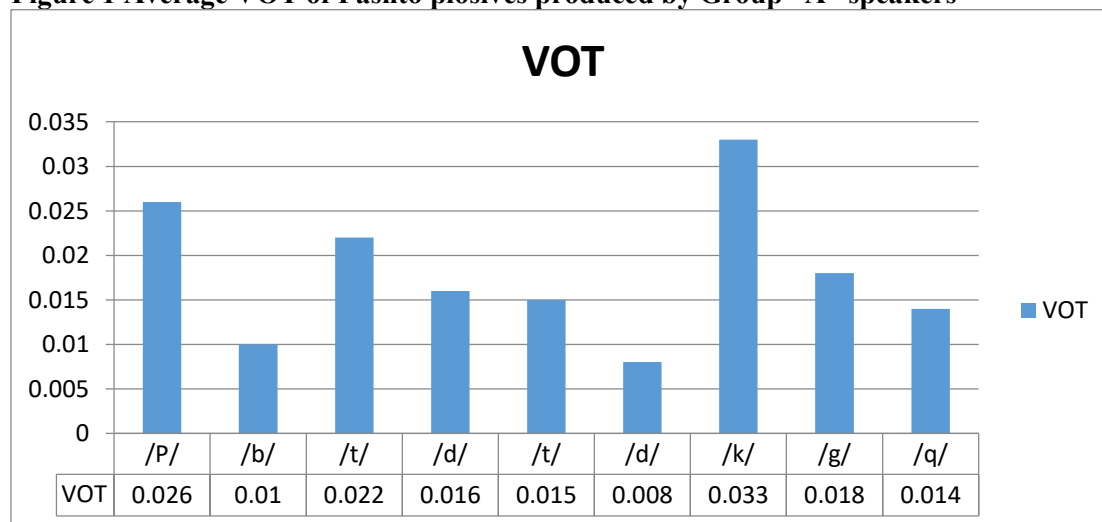
PLOSIVE	AVERAGE VOT
/P/	0.026
/b/	.01
/t/	0.022
/d/	0.016
/tt/	0.015
/dd/	0.008
/k/	0.033
/g/	0.018
/q/	0.014

Source: Authors

The table 3 presents the average VOT of Pashto plosives produced by L1 Pashto monolingual speakers. The highest VOT value is .033 which is noted down for the voiceless plosive /k/ while the

lowest value is .008 which is recorded for voiced plosive /d/. the data is further compared in the following Figure.

**Figure 1 Average VOT of Pashto plosives produced by Group “A” speakers**



Source: Authors

Figure 1 exhibits the differences in sound property of VOT. It shows the highest bar for voiceless consonant /k/ and the lowest bar for voiced plosive /dd/. The Figure clearly shows that voiceless plosives have longer VOT than voiced plosives.

**4.1.2. Data Analysis of Group “B” speakers**

In this section of the analysis, the researcher is mainly concerned with the data taken from Group “B” speakers who are L1 Pashto EFL learners. The section mainly focuses on the acoustic properties of English plosives produced by group “B” participants. These participants are EFL learners and here the focus is given to the identification of the acoustic properties taken from them. The values of closure duration and VOT have been taken averages and then these averages have been analyzed though tables and figures.

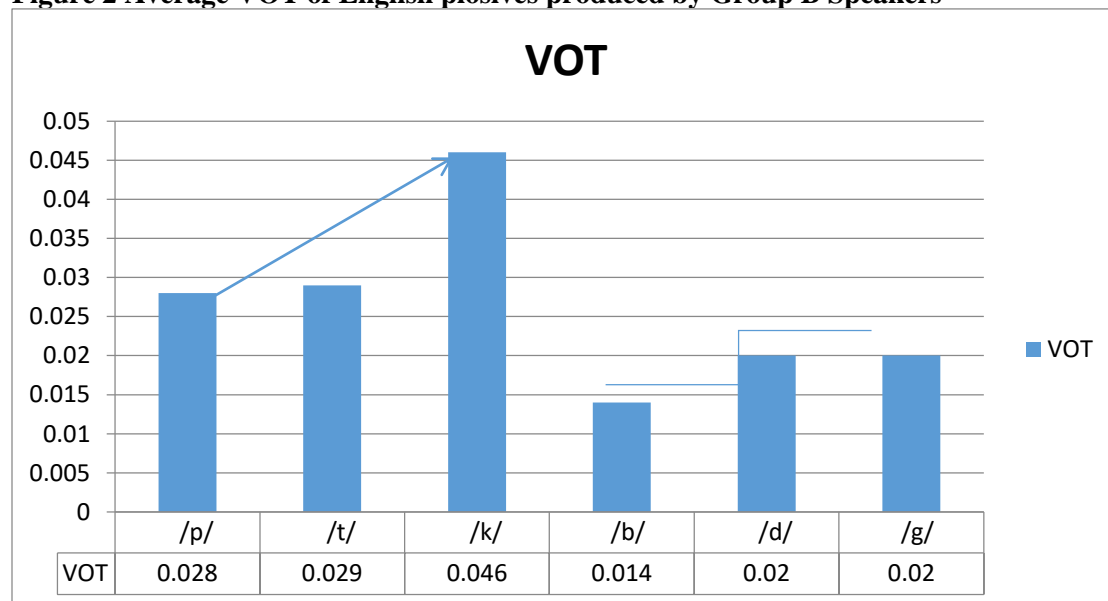
Table: 4 Average VOT of English Plosives Produced by Group “B” speakers

Plosive	Average VOT
/p/	0.028
/t/	0.029
/k/	0.046
/b/	0.014
/d/	0.02
/g/	0.02

Source: Authors

The table presents the average VOT values of English plosives by group B speakers. Voiceless plosives have maximum VOT comparatively voiced plosives. the maximum VOT is noted for velar plosive /k/. the voiced plosives /d/ and /g/ have the same VOT while /p/ and /t/ bears nearly the same values. The difference is quite clear between the bilabial plosives /p/ and /b/ of which the VOT of /p/ is significantly greater than /b/.

Figure 2 Average VOT of English plosives produced by Group B Speakers



Source: Authors

Figure 2 exhibits the average VOT of English plosives. Two quite distinct trends are found the figure. The first trend line shows the increase in the VOT in the voiceless from bilabial to velar. (see blue line on /p/, /t/ and /k/). The second trend line is on the voiced plosives, which is in the decrease of VOT from bilabial /b/ to velar /g/ plosive which is totally opposite to the first trend line. The maximum VOT is noted for the velar

plosive /k/. On the other hand the shortest VOT duration is noted for the bilabial plosive /b/.

**4.1.3. Data Analysis of the RP speakers**

The data in this section is taken from Leigh [Lisker \(1957\)](#). The average values of closure duration and VOT has been taken from the L1 English speakers. These values are first identified and then compared with other with the help of tables and figures.

**Table: 5 Average VOT of English Plosives by RP speakers Adopted from Leigh Lisker (1957, pp. 42-43)**

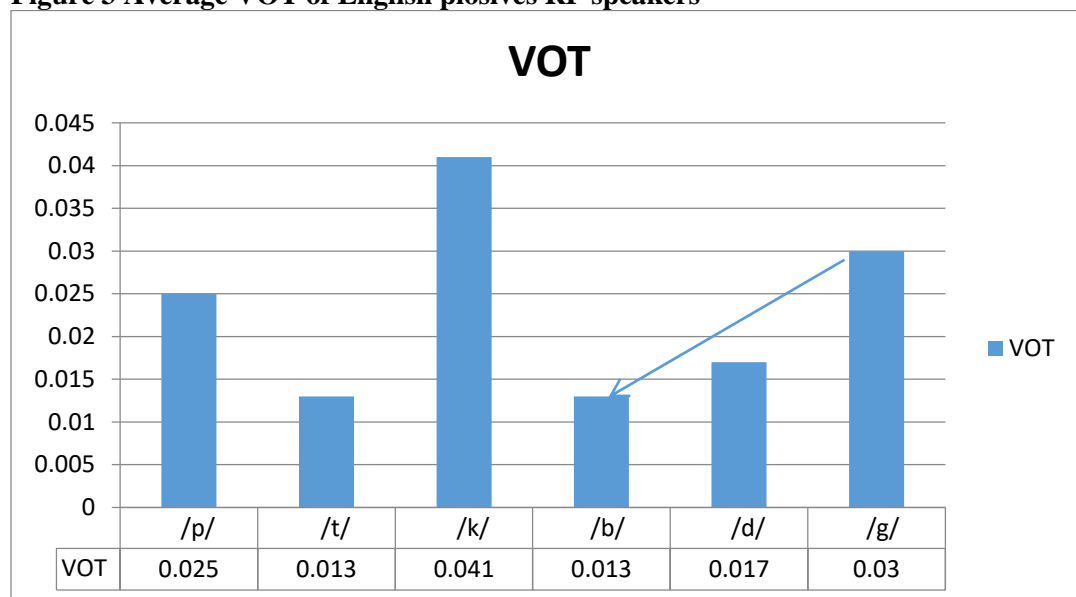
Plosives	Average VOT
/p/	0.025
/t/	0.013
/k/	0.041
/b/	0.013
/d/	0.017
/g/	0.03

Source: Authors

Table 5 exhibits the average VOT of English plosives adopted by Lisker. Velar plosives /t/ and /g/ have maximum VOT than the other plosive of which the maximum VOT is noted for the voiceless velar plosive /k/ which is .041. After

bilabial plosives have longer VOT than Alveolar plosives of which /p/ has maximum VOT .025. The comparison in VOT is further analyzed in the following figure.

**Figure 3 Average VOT of English plosives RP speakers**



Source: Authors

Figure 3 displays the comparison of VOT of English plosives. Voiceless velar plosive shows the longer VOT which is .041 while the minimum duration is shared by two plosives /t/ and /b/. A trend of increase in the VOT is found in the voiced plosives which is from bilabial voiced /b/ to alveolar /d/ and then to velar /g/ (See blue line on the voiced plosives /b/, /d/ and /g/).

**4.1.4. Comparison between the Acoustic properties of Group A, B and RP speakers**

The main focus of this section in the analysis is on the comparison between the different values taken from group “A”, “B” and RP speakers. The comparison is made through tables and figures. At first, the closure durations of group “A”, “B” and RP speakers are compared with each other,

then the VOT of these three groups and at last the plosive durations of these groups are compared.

**Table: 6 Comparison of VOT of Group “A”, Group “B” and RP speakers**

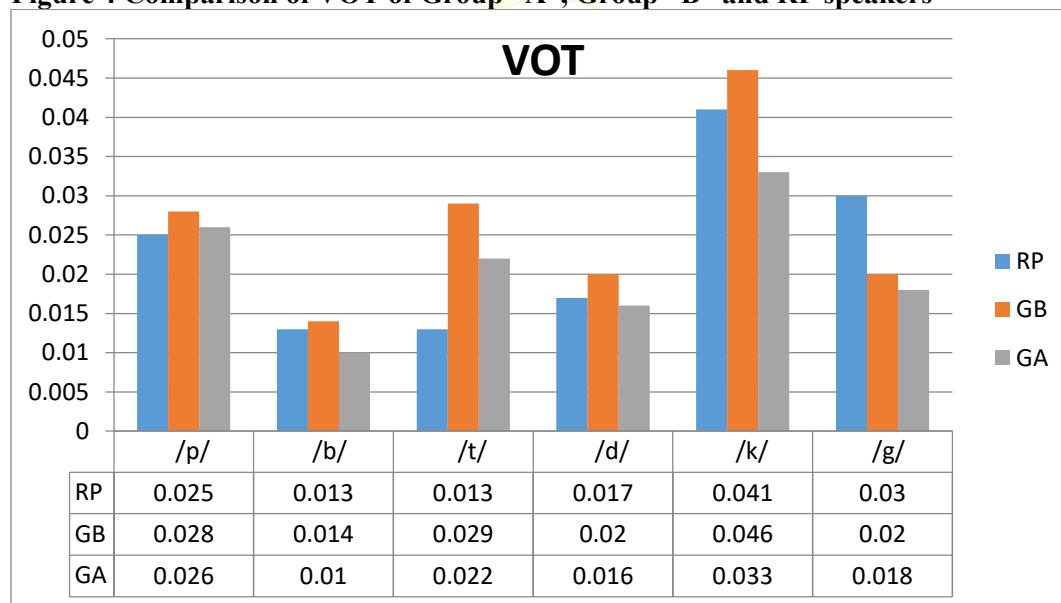
Plosive	RP/Native English	Group B	Group A
/p/	0.025	0.028	0.026
/b/	0.013	0.014	0.01
/t/	0.013	0.029	0.022
/d/	0.017	0.02	0.016
/k/	0.041	0.046	0.033
/g/	0.03	0.02	0.018

Source; Authors

Table presented above demonstrate the different VOT duration of English and Pashto plosives. English plosives are pronounced differently by different sample group participants from each other. The average VOT of bilabial, alveolar and one velar /k/ English plosives by group B speakers is longer than the English bilabial, alveolar plosives and velar plosive /k/ by native

English speakers while all the English plosives by group B speakers are longer than the Pashto plosives by Group A speakers in terms of VOT. The only Native English velar plosive /g/ has longer VOT than the English velar plosive by group B speakers and Pashto velar plosive by group A speakers.

**Figure 4 Comparison of VOT of Group “A”, Group “B” and RP speakers**



Source; Authors

Figure 4 is all about the comparison between the averages VOT of plosives produced by different sample groups. All the English plosives by group B speakers have longer VOT than Pashto and L1 English plosives except /g/ plosive as English /g/ plosive by native English speakers has longer VOT than group A and B plosives. A trend of

decrease is found in the figure in the plosives /b/, /d/ and /g/ which is a decrease in the VOT from the English plosives by group B speakers to native English plosives and then to Pashto plosives by group A speakers. On the other hand /p/ and /t/ English plosives by native English speakers have although shorter than English /p/



and /t/ plosives by group B speakers but longer than Pashto /p/ and /t/ plosives. All the Pashto plosives are short in average VOT than the English plosives by native English and group B speakers.

### Conclusion

The present study focused on the VOT trends in Pashto and English languages. For this purpose, the plosive sounds in these two languages have been analyzed through Praat software. The data was collected from LI Pashto speakers and EFL Pashto learners. After the analysis and comparison it has been identified that the voicing onset time of English plosive sounds are lengthier than the Pashto plosive sounds. This major difference in the acoustic treatment of Plosive creates issues of pronunciation for EFL Pashto learners. Thus EFL Pashto learners are unable to produce English plosive like native English speakers.

### Limitation And Study Forward

The present study has been delimited to the analysis of the plosive sounds in English and Pashto languages in terms of VOT trends. Future researchers can contribute to the field by comparing all the consonant sounds in terms of different acoustic cues.

**Author's Contribution:** The first two authors contributed in the analysis of the data through Praat software. The second author collected the data for the study while the third author contributed in proofreading the article.

### REFERENCES

1. Beckman, J., P. Helgason, B. McMurray, and C. Ringen (2011). "Rate effects on Swedish VOT: Evidence for phonological
2. Lieberman, P., and S. Blumstein (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
3. Lisker, L., and A. S. Abramson (1964). "A cross-language study of voicing in initial stops: Acoustical measurements". *Word* 20, pp. 384-422.
4. Lisker, L., and A. S. Abramson (1967). Some effects of context on voice onset

- time in English stops. *Language and Speech* 10, pp.1-28.
5. Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
6. Zlatin, M. A. (1974). "Voicing contrast: Perceptual and productive voice onset time characteristics of adults". *Journal of the Acoustical Society of America* 56, pp. 981-994.
7. Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics*, 1, 263-271.
8. Huckvale, M. (Ed.). (2015, July 22). *Resources in Speech, Hearing & Phonetics*. Retrieved August 10, 2015, from (<http://www.phon.ucl.ac.uk/home/johnm/siphtra/plostut2/plostut2-4.htm>)
9. Kuan Yi and C. Lei-Mei. (2008). 'A cross-linguistic study of Voice Onset Time in stop consonant production.' *Computational Linguistics and Chinese Language Processing* Vol. 13, No. 2, June 2008, pp. 215-232
10. Ladefoged, P. (2001). *A course in phonetics*. USA: Harcourt College Publishers.
11. Lisker, L. (1957). Closure Duration and the Voiced-voiceless Distinction in English. *International Review of Applied Linguistics*, iii (33), 42-43.
12. Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 3, 693-703.
13. Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426-449). Lawrence Erlbaum: NJ.
14. Y. Yao. (2007). 'Closure Duration and VOT of words-initial Voiceless Stops in English in Spontaneous Connected Speech' U C Bakerly Phonology Report.
15. Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge: MIT Press.
16. Chen, L. M., Oller, D. K., Lee, C. C., & Liu, C. T. (2018, October). LENA

- computerized automatic analysis of speech development from birth to three. In Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018) (pp. 158-168).
17. J Jones, G., Nadjibzadeh, N., Károly, L., & Mohammadpour, M. (2019). An integrated dialect analysis tool using phonetics and acoustics. *Lingua*, 221, 37-48.
  18. Khan, A., Steiner, I., Sugano, Y., Bulling, A., & Macdonald, R. (2017). A multimodal corpus of expert gaze and behavior during phonetic segmentation tasks. arXiv preprint arXiv:1712.04798.
  19. Xu, D., Richards, J. A., & Gilkerson, J. (2014). Automated analysis of child phonetic production using naturalistic recordings. *Journal of Speech, Language, and Hearing Research*, 57(5), 1638-1650.
  20. Kalita, S.; Mahadeva Prasanna, S.R.; Dandapat, S. Importance of glottis landmarks for the assessment of cleft lip and palate speech intelligibility. *J. Acoust. Soc. Am.* 2018, 144, 2656–2661.
  21. Olson, E., “Voice Onset Time in Arabic and English Stop Consonants”, A Senior Honors Thesis, Faculty of The University of Utah, May 2017.
  22. Kaur, J. (2015). Factors Influencing Voice Onset Time (VOT): Voice Recognition. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Elsevier, 19-31.
  23. Korkmaz, Y., & Boyacı, A. (2018, March). Analysis of speaker's gender effects in voice onset time of Turkish stop consonants. In 2018 6th International Symposium on Digital Forensic and Security (ISDFS) (pp. 1-5). IEEE..

