# A Preliminary Study on the Acoustic Cues of Final Stop Voicing in Pashto

Le Xuan Chan, Asif Khan
*International Christian University*

## 1 Introduction

Pashto is an Indo-Iranian language spoken mainly in Pakistan and Afghanistan and is estimated to have over 30 million native speakers. Though many languages around the region where Pashto is spoken such as Urdu demonstrate more than three laryngeal categories, Pashto only has a two-way laryngeal contrast for stop consonants (Hussain, 2018). What is of interest to the present study is that voicing is contrastive for final stops in Pashto language. During an elicitation session for a separate project, we noticed that though the acoustic signals for the voiced /d/ and voiceless /t/ in the final position sounded similar to native English and Japanese speakers, the consultant maintained that they were two distinct sounds. In the initial position, however, both native English speakers and Japanese speakers were able to perceive /d/ and /t/ in Pashto as two different sounds. It is this observation that led to this study, in which we ask the question, 'What distinguishes a voiced stop from a voiceless stop in the final position?'

Obstruents in Pashto as listed in Shafeev (1964) and David (2014) consist of the labial stops /p, b/, coronal stops /t, d/, palatal affricates /t͡ʃ, d͡ʒ/, retroflex stops /ʈ,ɖ/, and velar stops /k, g/. Robson & Tegey (2009) lists an additional uvular stop /q/, which we also found in this study. As mentioned above, Pashto stops demonstrate a 2-way laryngeal contrast with the exception of the uvular /q/. Hussain (2018) describes the two categories as voiced unaspirated stops and voiceless unaspirated stops. As will be elaborated on in section 2, Voice Onset Time (VOT) has been reported as the main cue for voicing in Pashto stops. However, these studies focus on initial stops, and to date there has been a lack of studies on voicing correlates of final stops. The definition of VOT is set out in Abramson & Lisker (1964) as the duration of time between the burst release of a stop consonant to the voicing onset of the subsequent vowel. What is of interest is that VOT is unavailable as an acoustic cue for final stops due to the absence of a following vowel within the same word boundary. If VOT is unavailable, then what other acoustic cues become relevant for voicing of final stops in Pashto?

## 2 Literature Review

**2.1** *Voicing Cues of Initial Stops in Pashto* As mentioned, Hussain (2018) looked at voicing contrast of initial stops in Pashto and concluded that Voice Onset Time (VOT) was the main acoustic cue for voicing contrast across all places of articulation, while Syed (2013) provides additional information on the VOT of initial voiceless stops in Pashto. Both Hussain (2018) and Syed (2013) found that voiceless stops in Pashto exhibited short lag VOT of around 6-31 ms, while Hussain (2018) reports that voiced stops experienced lead voicing (voicing that takes place during the closure duration) of about -110 to -128ms. Hence, voiced stops in Pashto are "truly" voiced stops.

**2.2** *Voicing Cues of Final Stops in Other Languages* Several studies have described vowel length of the preceding vowel as the main voicing cue for final stops. Weismer et al. (1981), for example, show that children who omit final stops in English utilise vowel length to preserve voicing contrast. The study in Raphael et al. (1975) looked at voicing of final stops in English CVNC structures, and reported that both vowel duration and nasal duration were both acoustic cues for voicing contrast, where both the preceding vowel and nasal segments were reported to be longer before the voiced /d/ over the voiceless /t/. Nasal duration was found to be a stronger cue than vowel duration in this case, but this is unsurprising for CVNC utterances.

Other studies, however, have argued that preceding vowel duration might not be a substantial voicing cue for final stops, especially in the context of natural speech. The study in Wardrip-Fruin (1982) found that voicing during closure was more significant than preceding vowel duration in disambiguating voiced and voiceless English stops. Hogan & Rozsypal (1980) also found that preceding vowel duration was more important before final fricatives, but the expansion of vowel nucleus before stops failed to produce any categorical change in


Proceedings of *AJL 6*

voicing for the final stop. Instead, the study cites other acoustic cues such as closure duration, voice bar i.e. voicing during closure, and burst duration to be more significant for voicing of final stops. Though not directly related to voicing cues, the study in Berkovits (1993) showed that final stops in Hebrew had a longer closure duration than their non-final counterparts, leading us to believe that closure duration might be significant at the word-final position. The burst release of stop consonants have also been investigated as a voicing cue for stop consonants. Revoile et al. (1982) reported that deletion of the burst in final voiceless stops resulted in a decrease in voicing perception. Chodroff & Wilson's (2014) spectral analysis on bursts as a voicing cue of stop consonants also shows that the burst spectrum of final stops might be relevant in our present study.

Hence, rather than one prevalent cue, there seems to be multiple cues for voicing of final stops. In the present study, we then ask the following questions: In lieu of VOT, which of these acoustic correlates become more salient as a cue for voicing of final stops in Pashto - vowel duration, closure duration, voicing during closure, or burst release?

## 3 Methodology

**3.1** *Stimuli* The data in this study was collected from a male native speaker of Pashto. The speaker was tasked with producing near-minimal pairs of various words in Pashto ending in the labial stops /p, b/, coronal stops /t, d/, and velar stops /k, g/. Retroflex and uvular stops were not elicited in this study. The list of target words used is shown in (1) and (2).

(1) Target words with final voiced stops in Pashto

a. /kabab/ ‘kebab’
b. /kab/ ‘fish’
c. /bad/ ‘bad’
d. /qad/ ‘height’
e. /zag/ ‘foam’
f. /ʁwag/ ‘ear’

(2) Target words with final voiceless stops in Pashto

a. /t͡ʃap/ ‘left’
b. /tap/ ‘weak/frail’
c. /ʁat/ ‘big’
d. /d͡ʒat/ ‘illiterate’
e. /dak/ ‘full’
f. /sak/ ‘back sprain’

The target tokens were first elicited in isolation, and then inserted in carrier sentences to control for the environment in which the final stops occured in. The isolated tokens were to be the neutral standard of reference for voicing of final stops in Pashto, while phrase-final tokens were elicited in carrier sentences to investigate the target tokens in natural speech. The target tokens were then placed before the nasal-initial word /na/ 'not' and the vowel-initial word /ao/ 'and' in carrier sentences to investigate the voicing cues before nasal consonants and vowels, both of which are sonorants. We expected the voicing cues to be similar before nasals and vowels as both are sonorants. However, as nasals are produced with an ensued closure of the oral cavity but with a lowered velum, we also expected to see additional effects on certain voicing cues such as closure duration and burst release. Thus, a total of four environments were used in this test. Some carrier sentences are shown in (3).

(3) Carrier Sentences based on Varying Environments

| | | | |
|---|---|---|---|
| a. | Isolated | /bad/ | ‘bad’ |
| b. | Phrase-final | /Kenta bad, zoɽ, ao badʃakal de/ | ‘Kenta is bad, old, and ugly. |
| c. | Before nasal | /Kenta bad na de/ | ‘Kenta is not bad.’ |
| d. | Before vowel | /Kenta bad ao zoɽ de/ | ‘Kenta is bad and old.’ |

**3.2** *Experiment* Two target words were elicited per target segment, so each target segment had a total of eight utterances based on different environments, and a total of 48 tokens were elicited. Alongside acoustic signals, we also measured the vocal fold vibrations of the speaker using an electroglottograph (EGG) which was worn by the speaker during the recording session. The EGG measures the electrical resistance between two electrodes which are attached non-invasively to the speaker's throat. The recorded signal corresponds to the vocal fold vibrations as the vocal fold movements disrupt the electrical resistance. Meanwhile, the acoustic signal was collected simultaneously using a lavalier microphone which was also worn by the speaker. During the recording session, the speaker was asked to pay attention to commas between words to elicit the phrase-final tokens. Where there were no commas, the speaker was asked to produce the sentence smoothly and without pause. Both the EGG data and acoustic signals were recorded, processed, and annotated using Praat (Boersma & Weenink, 2021). The EGG data was then also processed using Eggnog (Villegas, 2019). In data analysis, the following criteria were looked at: 1) closure duration of the final stop, 2) length of the vowel preceding the final stop, 3) the presence of voicing during closure, and 4) the presence of a burst release. Due to unforeseen circumstances, however, only one repetition was fit for analysis.

## 4 Results

**4.1** *Closure Duration* Where there was a burst release, closure duration was measured from the offset of the preceding vowel to the burst. Where there was no burst release, closure duration was measured from the offset of the preceding vowel to the onset of the following segment. In this regard, the results in the phrase-final position were omitted as closure duration could not be determined accurately where there was no burst.

We found the difference of closure duration between voiced and voiceless stops was to be significant. The length of closure was significantly shorter for voiced stops than voiceless stops at the final position. The results are in Table 1.

Mean closure duration (ms)

| Environment | Isolated | Phrase final | Before /n/ | Before /ao/ | Overall |
|---|---|---|---|---|---|
| **Voiced** | 88.67 | n/a | 97 | 75.5 | 87.6 |
| **Voiceless** | 158.67 | n/a | 118.17 | 140.83 | 139.22 |

**Table 1**: Mean closure duration of Pashto final stops

The mean difference between voiced and voiceless stops is similar in isolation and before vowels, at around 70ms. Interestingly, the difference is much smaller before nasals, at around 20ms. Voiced stops before nasals display longer closure durations over other environments, while voiceless stops before nasals display shorter closure durations over other environments. This suggests that closure duration might not be a strong cue for voicing before nasals, and this might be due to the articulatory influence of the nasal segment.

**4.2** *Preceding Vowel Duration* Duration of the preceding vowel was measured from the onset of the vowel to the offset of the vowel. The preceding vowel duration from /kabab/ 'kabab' was omitted from the data as the second syllable was lengthened due to primary stress, resulting in the vowel length being significantly longer than the other tokens before voiced stops. The remaining results are in Table 2.

Mean preceding vowel duration (ms)

| Environment | Isolated | Phrase final | Before /n/ | Before /ao/ | Overall |
|---|---|---|---|---|---|
| **Voiced** | 85 | 83.4 | 55.6 | 58 | 70.5 |
| **Voiceless** | 67.67 | 70.83 | 49.83 | 60.83 | 62.29 |

**Table 2**: Mean preceding vowel duration of Pashto final stops

The difference between vowel duration preceding voiced and voiceless final stops were not found to be significant. In most cases, vowels preceding voiced stops had a marginally longer duration over voiceless stops, but there were multiple instances where preceding vowel length was longer before voiceless stops as well. The overall difference in mean vowel duration between a voiced stop and a voiceless stop was also found to be only 8 ms, which suggests that preceding vowel duration is not a cue for voicing of final stops in Pashto.

**4.3** *Vocal Fold Vibration during Closure* Voicing during closure was recorded by observing the EGG signal during the closure duration. As expected, voiced stops in the final position displayed strong voicing during closure while voiceless stops do not. This is consistent with previous reports about lead voicing in initial voiced stops. A comparison between the EGG data of final voiced /b/ and voiceless /p/ is shown in Figures 1 and 2.

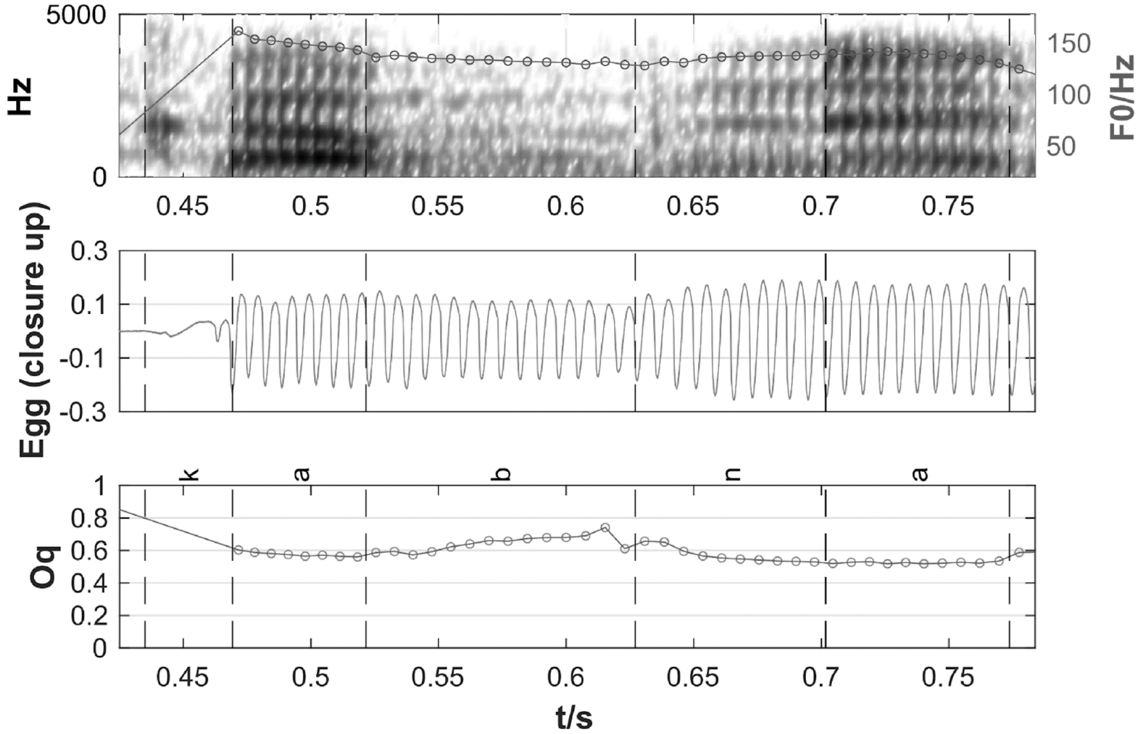


**Figure 1**: /kab/ 'fish' before nasal

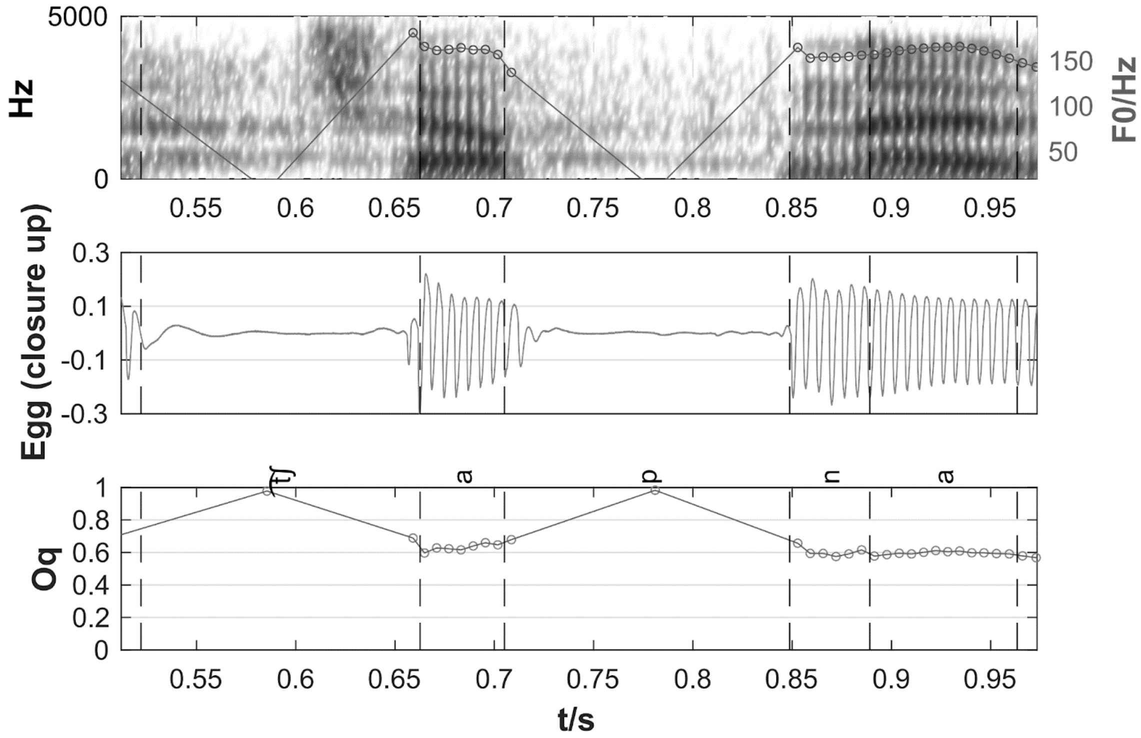


**Figure 2**: /t͡ʃap/ 'left' before nasal

As seen, voiced stops remain fully voiced even in the final position, and vocal fold vibration remains a strong cue for voicing in the final position. Interestingly, final voiced stops produced in isolation and the phrase final position displayed a higher tendency to have a burst release after the closure, which in turn affects vocal fold vibration as voicing cannot take place during a burst. Hence, voicing during closure in these environments were "partial", i.e. it does not last for the entire duration of the closure and is stopped before the burst. Before nasals and vowels, most of the stops were not released, so vocal fold vibration lasts the entirety of the closure duration and ensues into the following voiced segment. A comparison between the environments is shown in Figures 3, 4, and 5

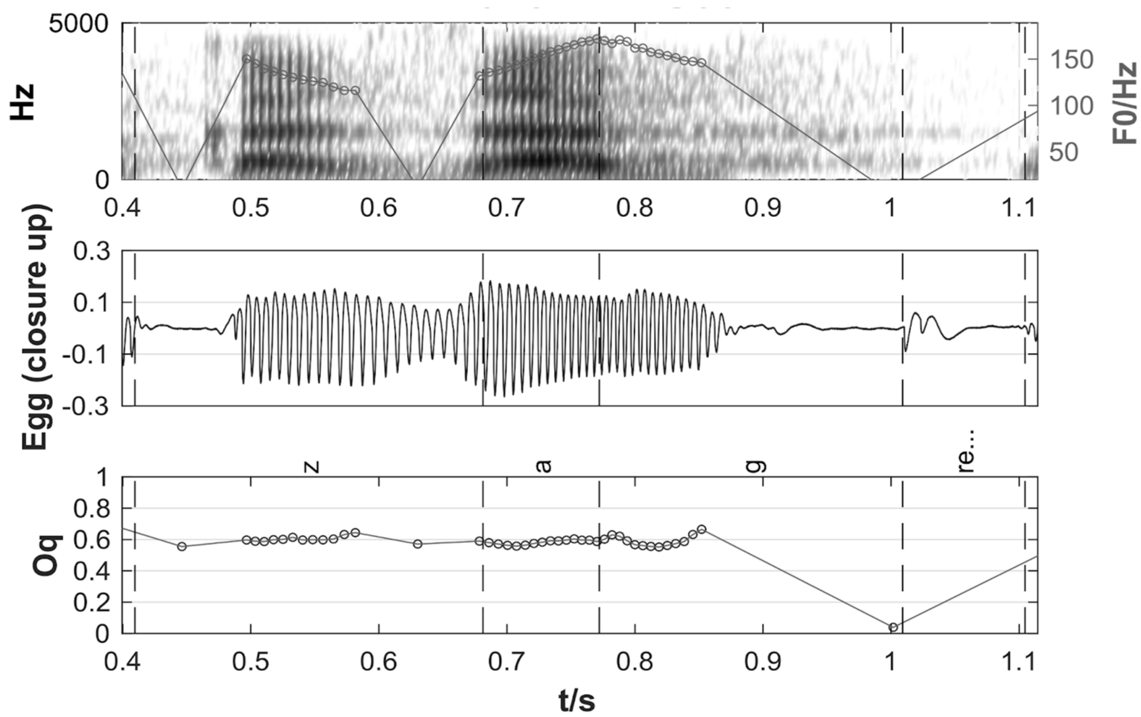


**Figure 3**: /zag/ 'foam' in the phrase-final position

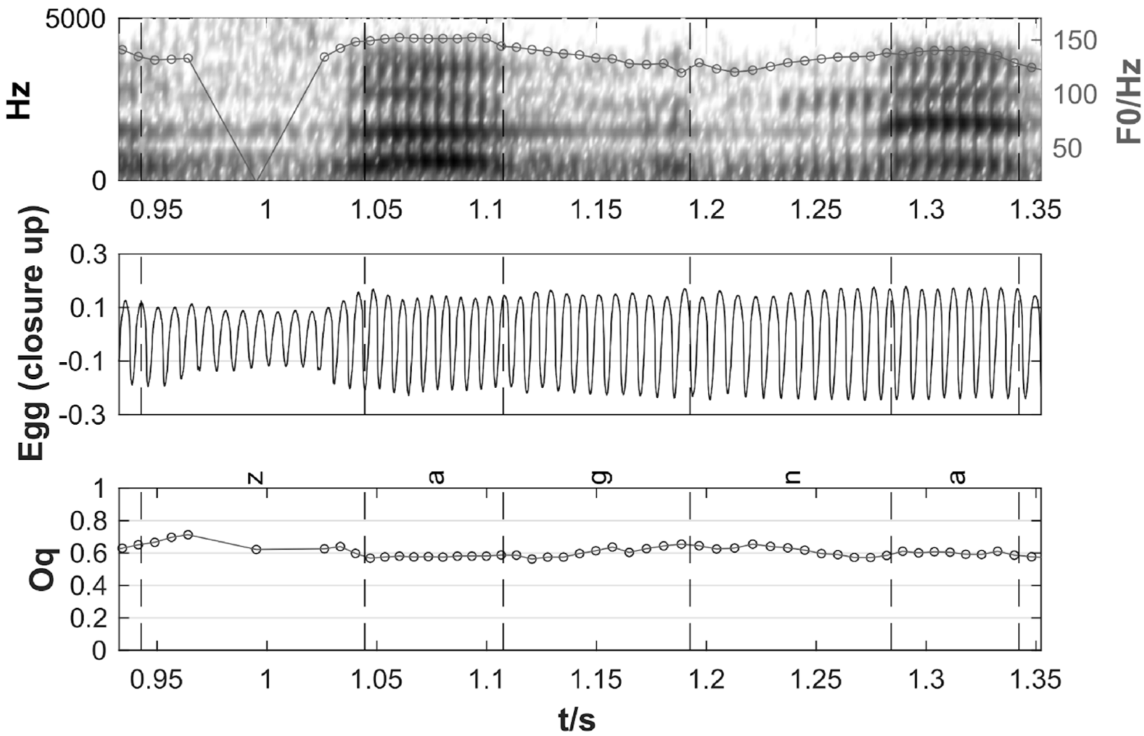


**Figure 4**: /zag/ 'foam' before nasal

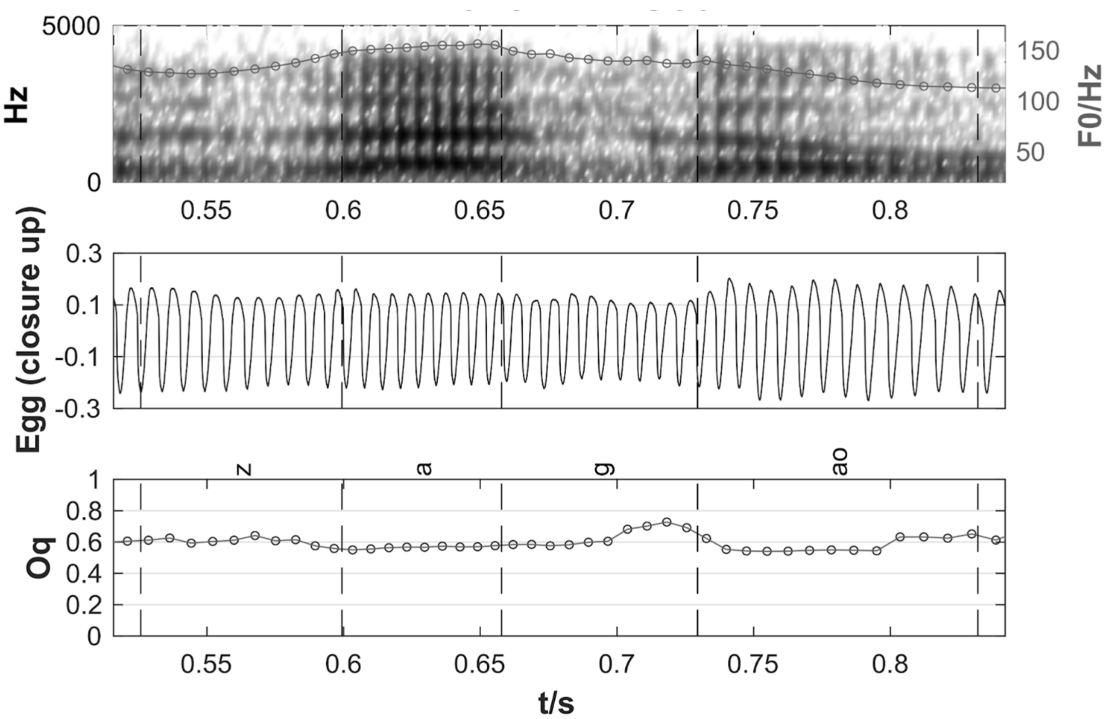


**Figure 5**: /zag/ ‘foam’ before vowel

In some cases, however, tokens produced in isolation had a voiced release. In these cases, vocal fold vibration lasts throughout the entire closure duration as well. This is shown in Figure 6.

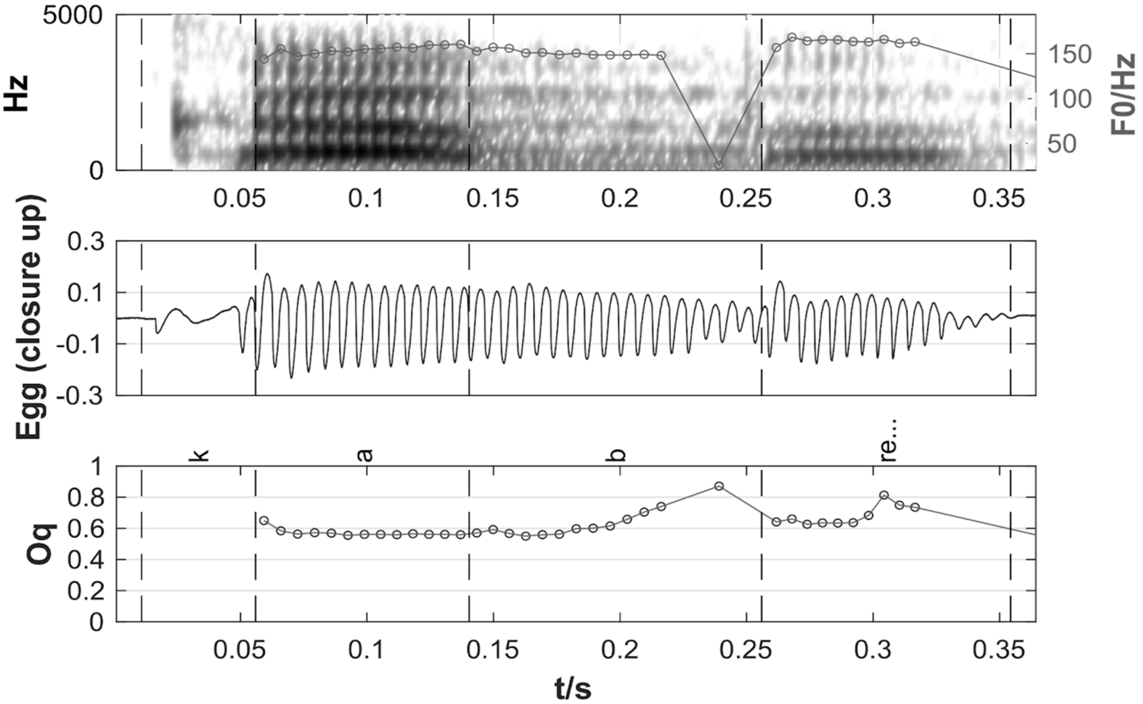


**Figure 6**: /kab/ ‘fish’ in the phrase-final position

The length of the vocal vibration during closure was found to be similar across all environments regardless of whether voicing lasts throughout the closure duration. As expected, length of voicing before nasals and vowels correspond directly to the duration of closure, while mean length of voicing was found to be around 10 ms shorter than the closure duration in the isolated tokens. This is shown in Table 3.

| Environment | Isolated | Phrase final | Before /n/ | Before /ao/ |
|---|---|---|---|---|
| Mean Voicing Duration (ms) | 77.17 | 79 | 97.17 | 75.5 |

**Table 3**: Mean voicing duration of Pashto final voiced stops

**4.4** *Burst Release* The presence of burst release was more sporadic in nature and was seen to be affected by environment as well as the place of articulation of the final stop. The general pattern was that both voiced and voiceless stops had burst releases when produced in isolation or in the phrase-final position. An example of a voiced stop with a burst release is in Figure 7 while a voiceless stop with a burst release is shown in Figure 8.

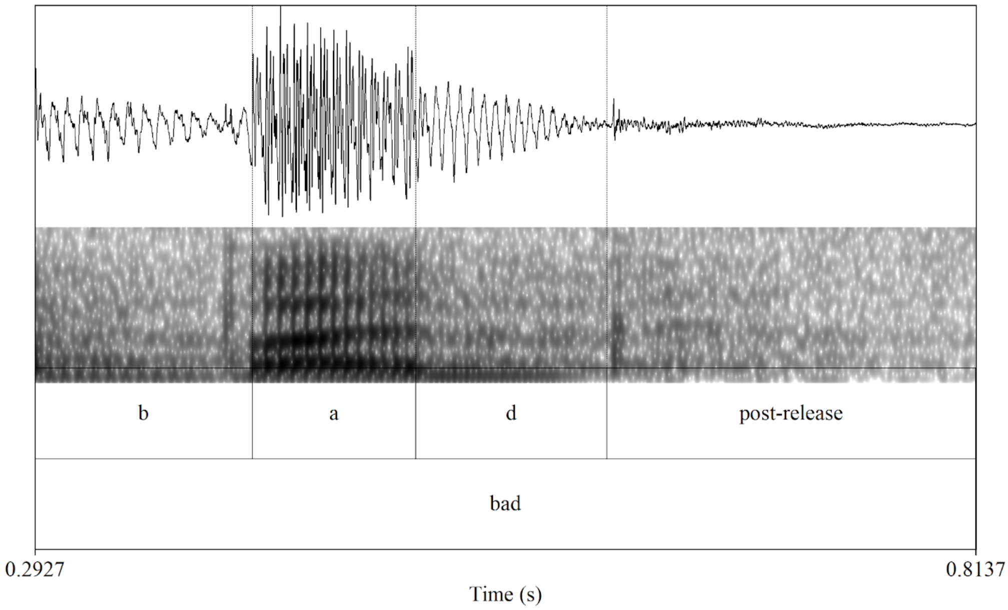


**Figure 7**: /bad/ 'bad' in the phrase-final position

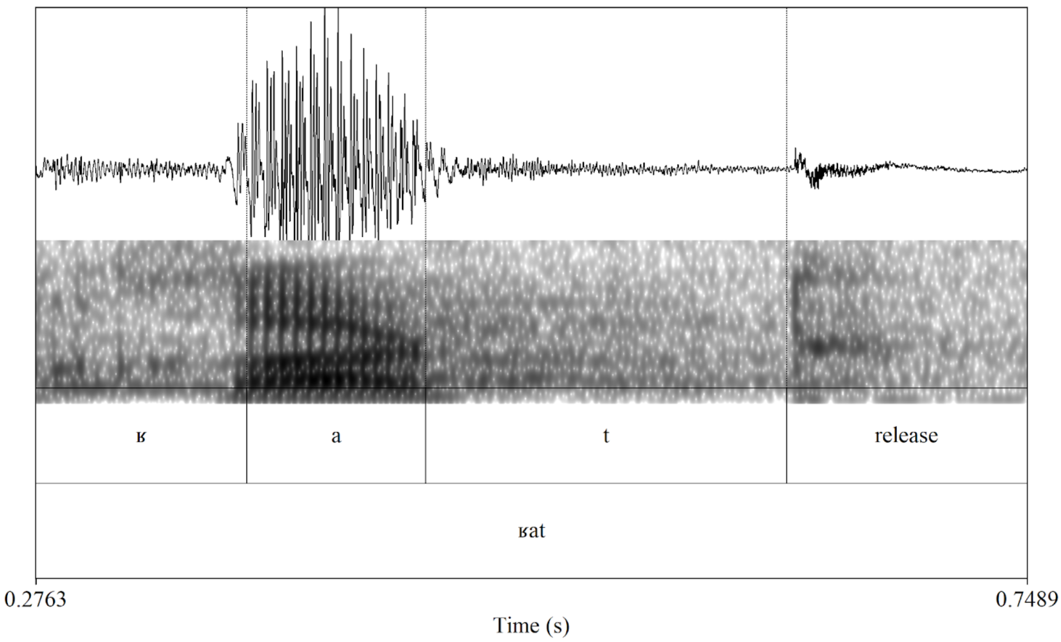


**Figure 8**: /ʁat/ 'big' in the phrase-final position

Additionally, in the case of voiced stops, we also recorded several instances where the release was voiced, resulting in a vowel-like acoustic signal. This is shown in Figure 9.

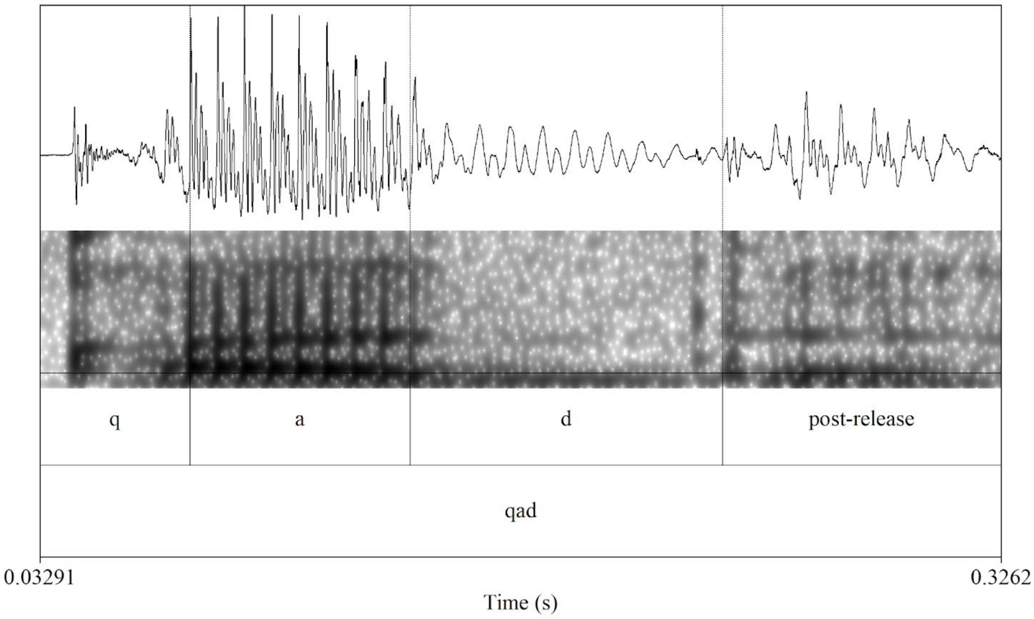


**Figure 9**: /qad/ 'height' in the phrase-final position

Before nasals and vowels, voiced stops generally did not have burst releases. Rather, the onset of the following segment follows the duration. This is shown in Figures 10 and 11.

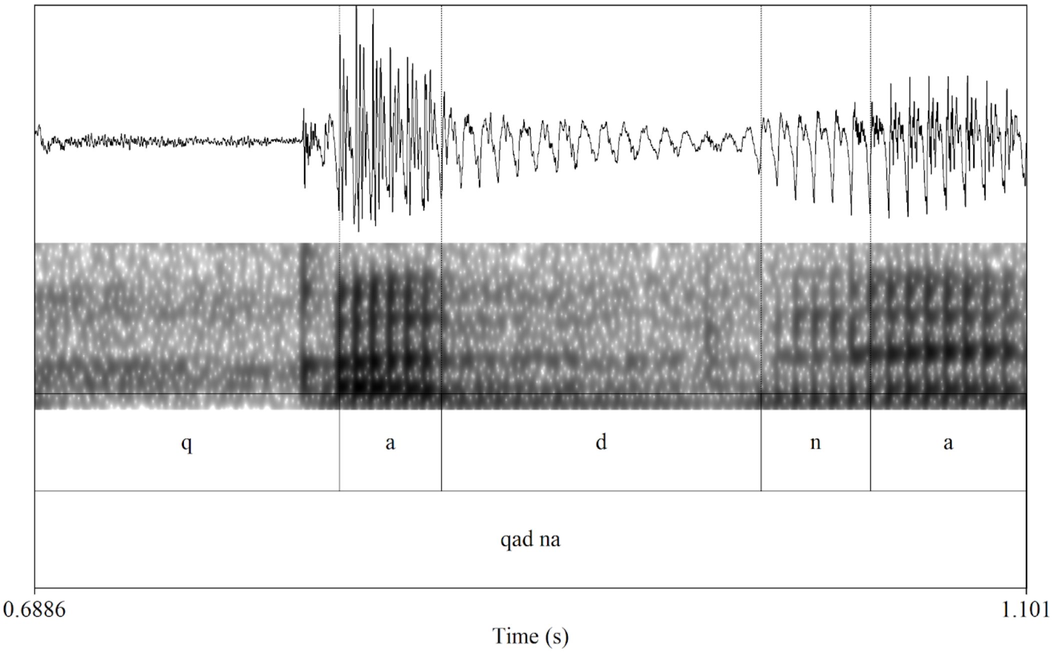


**Figure 10**: /qad/ 'height' before nasal

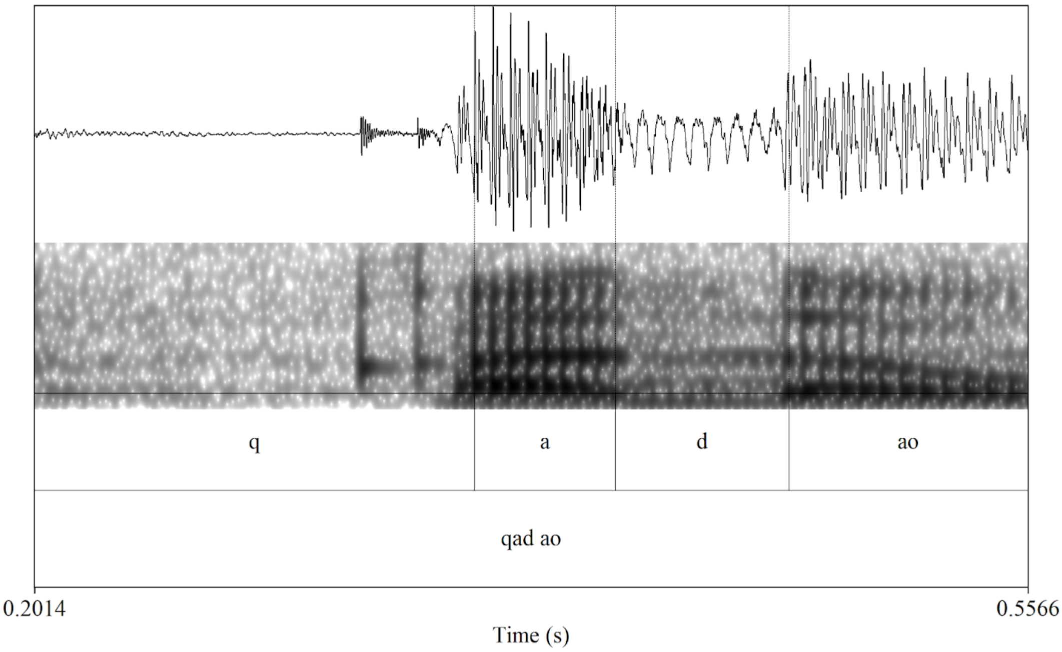


**Figure 11**: /qad/ ‘height’ before vowel

For voiceless stops produced before nasals and vowels, burst releases were more frequent than voiced stops, but not all stops had burst releases. To this end, we found that the voiceless stops displaying burst releases were alveolar and velar stops, while labial stops did not display any burst releases. An example of a voiceless alveolar stop with a burst release is in Figure 12 and voiceless bilabial stop without a burst release is shown in Figure 13.

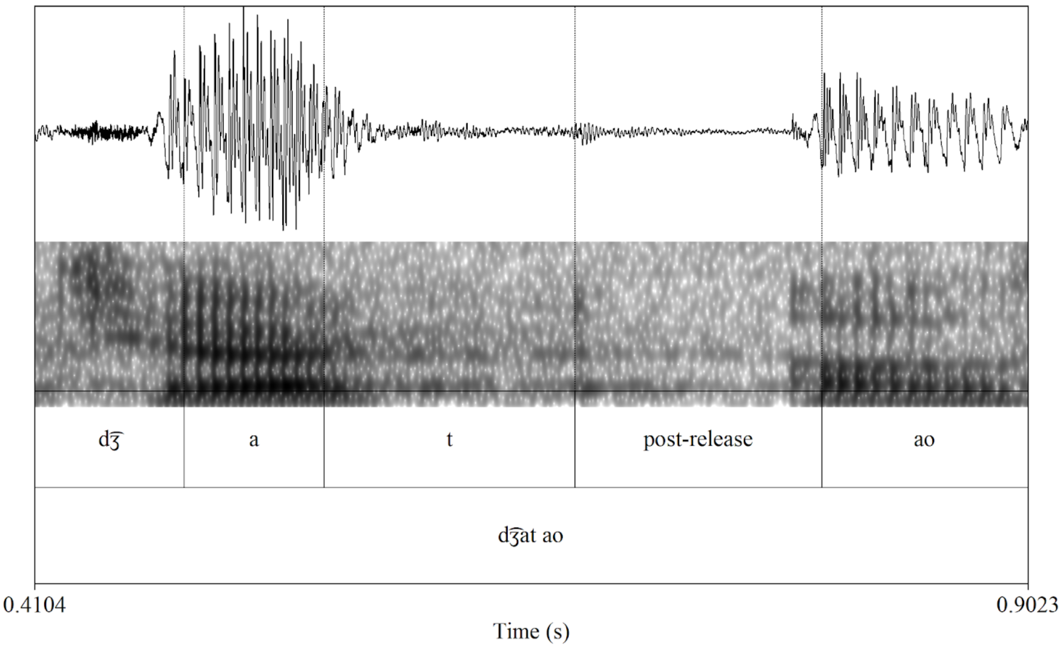


**Figure 12**: /d͡ʒat/ ‘illiterate’ before vowel

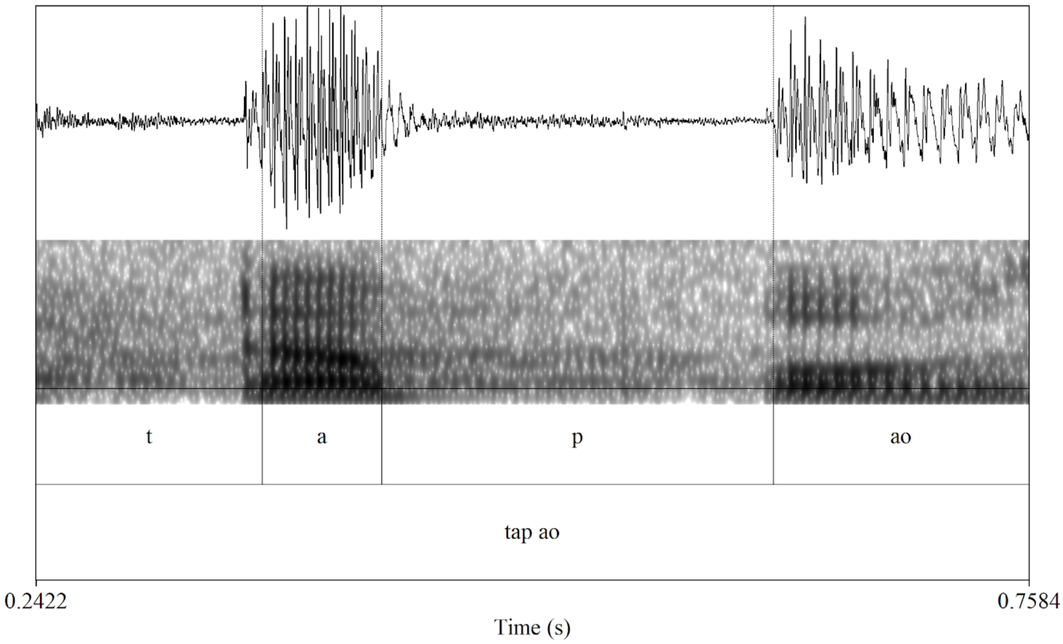


**Figure 13**: /tap/ ‘weak/frail’ before vowel

## 5 Discussion

Due to unavoidable limitations, we were only able to acquire speech samples from one speaker of Pashto in this study. As mentioned in section 3, we were also only able to utilise one repetition of the recorded stimuli. More data across multiple speakers and more repetitions would be needed for a more conclusive analysis. The current data, however, seems to provide some useful insights for voicing cues of final stops in Pashto.

The main acoustic cues for voicing are closure duration and voicing during closure, while vowel length does not seem to be a significant cue for voicing in Pashto. Vocal fold vibration during closure was found in all instances of final voiced stops but not in final voiceless stops. Additionally, final voiceless stops displayed a significantly longer closure duration than voiced stops in all instances. On the other hand, vowels preceding voiced and voiceless stops did not display any significant difference in vowel length. Though the results for burst release were not as clear-cut as either acoustic cues, there is sufficient evidence to say that burst release is also affected by voicing. In this regard, the results of this study seem to be similar with what was found in Wardrip-Fruin (1982) and Hogan & Rozsypal (1980), which both state that vowel duration was less important than other acoustic properties such as voice bar, closure duration, and burst duration.

The study in Wardrip-Fruin (1982) noted that in the final position, the voiceless feature is defined, as per its label, as the lack of voicing rather than some alternative acoustic configuration. What we found in this study also suggests that the voicing feature is still very much defined by vocal fold vibration than other acoustic cues. Given that voiced stops in Pashto are “true” voiced stops, it is not surprising to see speakers rely heavily on vocal fold vibration to disambiguate between voiced and voiceless stops. It is possible that the difference in closure duration is a secondary voicing cue brought about by the presence or absence of vocal fold vibration. The shorter closure duration of voiced stops might be attributed to the physical difficulty of sustaining vocal fold vibration while the oral cavity is closed. As vocal fold vibration is caused by the airflow through the vocal folds, there is a need for the air to be released, leading to an earlier release of the closure. As vocal fold vibration does not occur for voiceless stops, the closure can be sustained for a longer duration before being released. It is possible to test for this through a perception test by controlling for both vocal fold vibration as well as closure duration, to see if vocal fold vibration is enough as a cue for disambiguation, or if closure duration is also needed.

The “partial” and “full” vocal fold vibrations seen in section 4.3 might also be attributed to the difficulty of sustaining vocal fold vibration during closure, which also affects burst releases. The speaker faces a physical dilemma when producing final voiced stops - on one hand, vocal fold vibration needs to occur to realise the voice feature; on the other hand, orally releasing the airflow caused by vocal fold vibration stops vocal fold vibration. This dilemma results in the “partial” voicing during closure in the isolated phrase-final cases, where the voicing stops and airflow from the vocal fold vibration of voiced stops is released as a burst release. Before nasals, however, this dilemma is resolved without the need for a burst release as airflow can be released through the nasal cavity. Hence, we see an absence of burst release for voiced stops before nasals, which in turn allows vocal fold vibration to last the entire duration of the closure. Interestingly, we see the same absence of burst releases before vowels as well, suggesting that vowels following final voiced stops might be nasalized in the onset.

The interaction between vocal fold vibration and burst release also explains why burst releases were observed more in voiceless stops. As vocal fold vibration is absent for voiceless stops, there is no dilemma and burst releases can take place more readily. However, we still have to account for voiceless labial stops which were not released at all in this study. One reason for this is that perhaps the releasing of labial stops can be observed visually by the interlocutor, whereas alveolar and velar stops are usually not able to be seen by the interlocutor. As such, burst releases are necessary to signal the word boundary for final alveolar and velar stops, but not for labial stops.

## 6 Conclusion

What we found in this study is that vocal fold vibration remains as the main cue for voicing of final stops in Pashto - voiced stops are disambiguated from voiceless stops by the presence of vocal fold vibration. No voiced stops were de-voiced in the final position, which seems to suggest that the voice feature in Pashto is defined in terms of vocal fold vibration more than any other acoustic correlates. Closure duration was observed to be a secondary cue for voicing with voiceless stops having a longer closure duration than voiced stops, but this may just be a byproduct of the presence or absence of vocal fold vibration. Similarly, the sporadic nature of burst releases in different environments may also be due to vocal fold vibration. Though voiced stops displayed burst releases in the phrase-final position, burst releases were not observed before nasals and vowels, implying that wherever possible, voicing takes precedence over other acoustic correlates in Pashto. Moving forward, a perception test would certainly be helpful to isolate each acoustic correlate and test for the claims made in this study.

## References


Berkovits, Rochele. 1993. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21(4). 479-489.

Boersma, Paul & Weenink, David. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.50, retrieved 15 June 2021 from http://www.praat.org/

Chodroff, Eleanor & Wilson, Colin. 2014. Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America* 136(5). 2762-2772.

David, Anne Boyle & Brugman, Claudia. 2014. *Descriptive Grammar of Pashto and its Dialects*. Berlin, Boston: De Gruyter Mouton.

Hogan, John T. & Rozsypal, Anton J. 1980. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America* 67(5). 1764-1771.

Hussain, Qandeel. 2018. A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics* 71. 284-305.

Villegas, Julián. 2019. Eggnog: ElectroGlottoGraphy Not Obligatory Guidance. Available from http://onkyo.u-aizu.ac.jp/Eggnog

Raphael, Lawrence J., Dorman, Michael F., Freeman, Frances & Tobin, Charles. 1975. Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. *Journal of Speech and Hearing Research* 18(3). 389-400.

Revoile, Sally, Pickett, James M., Holden, Lisa D., & Talkin, David. 1982. Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. *The Journal of the Acoustical Society of America* 72(4). 1145-1154.

Robson, Barbara, & Tegey, Habibullah. 2009. Pashto. ln Gernot Windfuhr (ed.), *The Iranian Languages*. 721-772.

Shafeev, D. A. 1964. *A short grammatical outline of Pashto*. The Hague: Mouton.

Syed, Nasir Abbas. 2013. Voice Onset Time (VOT) for Voiceless Plosives in Pashto (L1) and English (L2). *The Journal of Humanities and Social Sciences* 21(3). 79.

Wardrip–Fruin, Carolyn. 1982. On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *The Journal of the Acoustical Society of America* 71(1). 187-195.

Weismer, Gary, Dinnsen, Daniel & Elbert, Mary. 1981. A study of the voicing distinction associated with omitted, word-final stops. *Journal of Speech and Hearing Disorders* 46(3). 320-328.